Laboratoire de Recherche en Informatique

2000-2004 report

November 2004 – Acknowledgments: Jean-Baptiste Labrune, Wendy Mackay, Claude Marché, François Goasdoué. Infodesign : Atelier Қ Nicolas Taffin – www.polylogue.com – printed at CNRS Laboratoire de Recherche en Informatique 2000-2004 report



### Table of Contents

Laboratoire de Recherche en Informatique 2000-2004 report

Introduction	A / 1
Overview	A.1 / <b>4</b>
Current and forth-coming four-year	plans 4
Summary of scientific activity	5
Report outline	8
Personnel	A.2 / 9
Researchers and Faculty	10
Administrative and technical staff	11
Doctoral students and post-docs	12
Infrastructure	A.3 / 13
Laboratory space	13
Network and computing resources	14
Library	16
Financial summary	A.4 / 18
Current four-year plan	A.5 / 20
PCRI	20
Development of new areas	21
Development of existing areas	22
Long-range plan for 2006-2009	A.6 / 24
Development of existing areas	24
Development of new areas	24

#### Graph Theory and Fundamental Aspects of Communications

Research Group Members	B.1.1 / 30
Research description	B.1.2 / 32
Graph Theory Team	32
Hamiltonian graph theory	33
Factor, covering and decomposit	ion 34
MinMax parameters	34
Algebraic aspects	35
Communication and	
Network Design Team	36
Optimisation	37
Information dissemination	38
Research Perspective	41
Highlights	B.1.3 / <b>42</b>
Honors	B.1.4 / <b>43</b>
Evaluation of research	B.1.5 / <b>45</b>
Volunteer professional service	B.1.6 / <b>47</b>
Contracts and grants	B.1.7 / <b>48</b>
ROCOCO	48
ROM	48
DAVID	49
Circulant Networks	49
Dynamo	49
Grands graphes	49
PairAPair	50
Progress in graphs	50
Frequency Assignment	50
Routage robuste	50
Other projects	50
Collaborations	B.1.8 / 51
Dissemination and	
technology transfer	B.1.9 / <b>52</b>
Training and education	
(doctoral and post-doctoral)	B.1.10 / 53
Publications	R 1 11 / <b>5</b> 4
	D.1.11/J4

Research Group Members	B.2.1 / 68
Research Description	B.2.2 / <b>70</b>
Algorithms Research team	70
Coding	70
Approximation Algorithms	70
Learning Theory and Stochastic M	Modelling 71
Scheduling	71
Cellular Automata	72
Combinatorics Research team	73
Discrete Tomography	73
Enumerative combinatorics	74
Combinatorics on words	74
Complexity Research team	/5
Logic	/5
Approximate Model Checking	eis 73
Structural Complexity	75
Kolmogorov complexity	76
Ouantum Computation Research	n team 77
Algorithms	77
Complexity	79
Physical Aspects	80
Perspectives	80
Research Perspective	81
Highlights	B.2.3 / 82
Honors	B.2.4 / 83
Evaluation of research	B.2.5 / 84
Volunteer professional service	B.2.6 / 85
Contracts and grants	B.2.7 / 86
QAIP	86
RESQ	87
RAND-APX	87
WEB TV	87
Collaborations	B.2.8 / 88
Dissemination	
and technology transfer	B.2.9 / <b>89</b>
Training and education	
(doctoral and post-doctoral)	B.2.10 / 90
Publications	B.2.11 / <b>91</b>

B.1 / 27 Algorithms and Complexity B.2 / 65

Introduction	
p. 1	
GraFComm	
p. 27	
Algo	
p. 65	
Prog	
p. 101	
BD p. 147	
<b>P</b> • • •	
Demons	
p. 177	
Archi	
p. 207	
Parall	
p. 241	
IASI	
p. 275	
letA	
p. 307	
Bioinfo	
p. 347	

v

#### Programming and B.3 / 101 Software Engineering

Research group members	B.3.1	/ 104
Research description	B.3.2	/ 106
The Software Engineering team		106
The Human Computer Interaction	on team	107
The Mobile Networks team		107
Software Engineering team		108
Testing based on formal specific	ations	108
Probabilistic and Statistical App	oaches	109
		109
In Situ PCRI project		110
New Interaction Paradigms		111
Engineering of interactive system	me	117
Mobile Networks team	115	116
UMTS		116
Mobility in IP		117
4G and Vertical handoff		117
Ad hoc networks (QoS and Secu	urity)	117
Network mobility		118
Research Perspective		118
Highlights	B.3.3	/ 119
Honors	B.3.4	/ 120
Evaluation of research	B.3.5	/ 121
Volunteer professional service	B.3.6	/ 123
Contracts and grants	B.3.7	/ 125
DSoS		125
InterLiving		126
EUD-Net		126
INDIGO		126
SAFARI		126
Jeune équipe CNRS		127
NEMO		127
Micromegas		12/
CEA Other contracts		12/
		120
SAMU		120
13		120
Collaborations	B 3 8	/ 128
Dissemination and	2.3.0	, .20
technology transfer	B.39	/ 129
Training and education	0.0.7	, . 27
(doctoral and post doctoral)	R 3 10	/ 127
	0.0.10	/ 1.52
Publications	Б.3.II	/ 133

Databases	B.4 / 147
Research Group members	B.4.1 / 150
Research description	B.4.2 / 151
Information Integration	152
Digital Libraries	153
Data warehouses	154
Languages and Systems for	
Semi-Structured Data	155
Query languages for web data:	
a case study	156
Polymorphic type inference	
for query languages	156
Logic-based query languages	
for XML-data	157
Pattern-matching based languag	Jes
for XML data: CDuce	157
lemporal databases	159
Databases for biology	161
Research Perspective	101
Highlights	B.4.3 / 162
Honors	B.4.4 / 163
Evaluation of research	B.4.5 / 163
Volunteer professional service	B.4.6 / 165
Contracts and grants	B.4.7 / 166
Selene	166
Delos	166
GraphDuce	167
CASC	167
TRALALA	168
Other Projects	168
Collaborations	B.4.8 / 169
Dissemination and	
technology transfer	B.4.9 / 170
Training and education	
(doctoral and post-doctoral)	B.4.10 / <b>170</b>
Publications	B.4.11 / <b>171</b>

Proofs and Programs	B.5 / 177	Parallel Architectures	B.6 / 207	Introduction
Research Group members	B.5.1 / 180	Research group members	B.6.1 / 210	p. 1
Research description	B.5.2 / 182	Research description:		
PCRI Project LogiCal	182	micro-architecture team	B.6.2 / 212	
Theme 1: Rewriting and		PCRI Project ALCHEMY	212	GraFComm
automated deduction	183	Scientific Context	213	p. 27
Theme 2: Higher-order formalis	ms 186	General Approach	213	
Theme 3: Proof of programs	188	Theme 1: A practical approach		
Highlights	B.5.3 / 192	to program optimizations for		Alma
Honors	B.5.4 / 193	complex architectures	214	Algo
Evaluation of research	B.5.5 / <b>193</b>	Theme 2: Revisiting the process	or	p. 65
Volunteer professional service	B 5 6 / <b>194</b>	Theme 3: A transversal research	roach 215	
Contracts and grants	D.5.0 / 104	direction: methodology		
	D.J./ / 193	and simulation	216	Prog
AVERBOES	195	Theme 4. Beyond performance	218	p. 101
CALIEF	196	Research description:		
GECCOO	196	cluster and grid team	B.6.2 / 218	
Other contracts	196	Theme 1: Cluster Computing	219	RD
VerifiCard	196	Theme 2 : Large Scale Distribute	ed and Parallel	n 147
Collaborations	B.5.8 / 197	Computing.	220	p. 147
Dissemination and		Highlights	B.6.3 / 222	
technology transfer	B.5.9 / <b>197</b>	Honors	B.6.4 / 223	_
Training and education		Evaluation of research	B 6 5 / <b>224</b>	Demons
(doctoral and post-doctoral)	B.5.10 / <b>199</b>	Volunteer professional service	B.6.6 / 221	р. 177
Publications	B 5 11 / 200		D.0.0 / 223	
T ublications	0.0.11 / 200	Contracts and grants	B.6./ / 22/	
		Processor architecture team:	227	Archi
			227	p. 207
		COMPAO / HP Erance	220	
		ACI Blanche	228	
		MAHOTEU	228	Desell
		Cluster and Grid research team:	229	Parall
		CASPer	229	p. 241
		CGP2P	229	
		Data Grid Explorer	229	
		eGrid5000	229	IASI
		AGIR	230	p. 275
		Collaborations	B.6.8 / 230	-

Dissemination and technology transfer

Publications

Training and education

(doctoral and post-doctoral) B.6.10 / 233

#### letA p. 307

B.6.9 / 231

B.6.11 / 235

#### Bioinfo

p. 347

Parallelism	B.7 / 241
Research group members	B.7.1 / 244
Research description	B.7.2 / 245
PCRI Project Grand Large	246
Building a Large Scale Distribute	d
System (LSDS) for Computing	247
Security and reliability of networ	k
control protocols	247
End-User Tools for Computationa	al
Science and Engineering	248
Modelling and Verification	248
Distributed Algorithms	252
Self-stabilizing Routing	253
Scalable self-stabilization	253
Stabilization with mobile agents	255
Observation of self-stabilization	256
Global Computing	258
Fault-tolerance in global comput	ing 258
Global computing emulation	258
Research Perspective	259
Highlights	B.7.3 / 260
Honors	B.7.4 / 261
Evaluation of research	B.7.5 / 261
Volunteer professional service	B.7.6 / 262
Contracts and grants	B.7.7 / 263
Mobicoop	263
STAR	263
CGP2P	264
SR2I	264
FRAGILE	264
AIRCAST	264
Algorithmique Répartie	265
Collaborations	B.7.8 / 265
Dissemination and	
technology transfer	B.7.9 / 265
Training and education	
(doctoral and post-doctoral)	3.7.10 / <b>267</b>
Publications	3.7.11 / <b>268</b>

# Artificial Intelligence<br/>and Inference SystemsB.8 / 275Research Group Members<br/>Research DescriptionB.8.1 / 278B.8.2 / 279B.8.2 / 279

Research Description	B.8.2 / 279
PCRI Project Gemo	280
Theme 1: Knowledge Represen	tation
and Algorithmic	280
Inference algorithms for	
propositional logic	280
Inference algorithms for	
the CARIN fragment of first ord	er logic 282
Theme 2: Information extraction	n
from Semi-structured Data	283
Extraction of frequent trees from	n
an heterogeneous XML collection	on 283
Flexible clustering with Galois la	attices 283
Semi-automated construction	
of an ontology	283
Theme 3: Mediation for	
the Semantic Web	284
Semantic portals	284
Peer-to-peer data management	systems 285
Mediation with end-users	286
Theme 4: Thematic warehouse	s 287
Acquisition of data from the We	eb 288
Extraction and integration of we	eb data driven
by an ontology	288
Mappings between ontologies	288
Research Perspective	289
Highlights	B.8.3 / <b>290</b>
Honors	B.8.4 / 291
Evaluation of research	B.8.5 / 291
Volunteer professional service	B.8.6 / 294
Contracts and grants	B.8.7 / 295
PICSEL1	295
PICSEL2	296
E.DOT	296
ACI MDD	296
PPF Supelec	296
Collaborations	B.8.8 / 297
Dissemination and	
technology transfer	B 8 9 / <b>298</b>
Training and education	
(doctoral and post doctoral)	R 8 10/ 200
	D.0.10/ 299
Publications	B.8.11 / 300

Inference and Learning	B.9 / 307	Bioinformatics	B.10 / 347	Introduction
December Crewer Marria	D 0 1 / 210		D 10 1 / 250	p. 1
Research Group Members	B.9.1 / 310	Research group members	B.10.1 / 350	F
Research Description	B.9.2 / 312	Research description	B.10.2 / 351	
PCRI Project TAO:		Theme 1: A multifield approa	ach	6 56
Thème Apprentissage & Opti	misation 312	for the prediction of a recodi	ng event 352	GraFComm
Fundamentals of Machine Lea	irning,	Theme 2: Analysis of Biologic	cal	р. 27
Knowledge Extraction and	212	Sequences and Structures	355	
	313		255	
AUC-based Learning	313	Models for sequence analysis	300	ΔΙαο
Feature Selection	313	visualisation and algorithm	257	Algo
Human Computer Interaction	212	design for RINA structures	330	p. 65
And Visual Data Mining	313		257	
and Compotence Maps	-iearning	Modelling and simulation of	537	
	514	biological processor	250	Prog
Constraint Satisfaction	1	Theme 3: Analysis and Stora	000	p. 101
and Phase Transition	215	of Massivo Biological Data	250	P
Bounded Relational Reasoning	316	Integration of biomedical dat	360	
Phase Transitions	316	Data Analysis	261	
lunction with Constraint Satisf	faction 316	Research Perspective	364	BD
Software	318		D 10 2 / 265	p. 147
Fundamentals of Evolutionary	Computation	Highlights	B.10.3 / 303	
318	computation	Evaluation of research	B.10.5 / 366	
Self-adaptive operators in con	tinuous	Volunteer professional servi	ce B.10.6 / 367	Demons
search spaces	318	Contracts and grants	B.10.7 / 369	
Estimation of Distribution Algo	orithms 319	НКІЅ	369	p. 177
Genetic Programming and		BIOINGENIERIE Plateforme "E	Biogen 74″ 370	
Context-Free Grammars	319	Phénomène de recodage	370	
Text Mining	320	RAFALE	370	Archi
Inverse Problems	322	Other projects	370	p. 207
Robotics	323	Collaborations	B.10.8 / 371	
Research Perspective	325	Dissemination and		
Highlights	B.9.3 / 327	technology transfer	B.10.9 / 372	Desell
Honors	B.9.4 / 328	Training and education	,	Parall
Evaluation of research	B.9.5 / 329	(doctoral and post-doctoral)	) B.10.10 / <b>373</b>	p. 241
Volunteer professional service	ce B.9.6 / 331	Publications	B.10.11 / 375	
Contracts and grants	B.9.7 / 332			
Collaborations	B.9.8 / 333			
Dissemination and				p. 275
technology transfer	B.9.9 / 334		201	
Training and education		Glossary	381	
(doctoral and post-doctoral)	) B.9.10 / 335			letA
Publications	B.9.11 / 336			p. 307

p. 1	
GraFComm	
Algo p. 65	
Prog	
p. 101	
BD p. 147	
Demons	
р. 177	
Archi p. 207	
Parall	
p. 241	
IASI p. 275	
-	
letA p. 307	
Bioinfo	
p. 347	

INFORMATIQUE

X



Laboratoire de Recherche en Informatique / 2000-2004 report

A / 1

## Le LRI

Directeur : Michel Beaudouin-Lafon

e laboratoire de Recherche en Informatique (LRI), créé il y a plus de 30 ans à l'Université Paris-Sud, est l'un des plus grands et des plus prestigieux laboratoires français de recherche en informatique. Au 1er janvier 2004, le laboratoire comptait 177 membres : 78 chercheurs et enseignants-chercheurs permanents, 63 doctorants, 22 personnels nonpermanents et 14 personnels techniques et administratifs (section A.2). Le LRI est constitué de 10 équipes de recherche, assistées d'une équipe de support système et réseau et d'une équipe administrative. Le laboratoire occupe 4500m<sup>2</sup> dans le bâtiment 490 sur le campus d'Orsay. Ses principaux moyens de recherche sont un réseau haute performance, une grappe de calcul et une bibliothèque (section A.3).

Le LRI est une unité mixte de recherche (UMR 8623) du CNRS et de l'Université Paris-Sud. Les personnels permanents sont issus de ces deux organismes (14 chercheurs et 10 ITAs CNRS, 64 enseignants-chercheurs et 4 IATOS Université Paris-Sud). Le budget annuel, hors salaires des permanents, est de l'ordre de 1.6M€. Moins d'un tiers de ce budget provient des dotations de base du CNRS et de l'Université ; le reste provient de contrats et subventions obtenus par les membres du laboratoire (section A.4).

Les thèmes de recherche abordés par le LRI couvrent un large spectre de l'informatique : algorithmique, complexité, calcul quantique, théorie des graphes, fondements des communications, micro-architecture, clusters et grilles de calcul, génie logiciel, programmation, interaction homme-machine, bases de données, systèmes d'inférence, fouille de données, apprentissage par machine, et bioinformatique. Cette diversité est l'une des forces du laboratoire, car elle favorise les recherches aux frontières des thématiques, là où le potentiel d'innovation est le plus grand. Ainsi, il y a deux ans, des chercheurs de quatre équipes ont décidé de créer un axe transversal sur la bioinformatique. Cet axe est désormais une équipe à part entière qui développe une approche originale en combinant les concepts de différents domaines, comme l'apprentissage par machine et les algorithmes randomisés pour prédire l'information biologique pertinente dans les données génomiques.

> Le reste de ce rapport présente l'activité du laboratoire pour la période 2000-2004 (section A.5) et son projet pour les quatre prochaines années (section A.6).

Introduction

Overview

## LRI Director : Michel Beaudouin-Lafon

RI (Laboratoire de Recherche en Informatique) is the Laboratory for Computer Science at the Université Paris-Sud. Founded more than 30 years ago, LRI is one of the top academic research laboratories in Computer Science in France, both in terms of size and scientific excellence. As of January 1st, 2004, the laboratory had 177 members: 78 permanent researchers, 63 doctoral students, 22 non-permanent research staff and 14 administrative and technical staff (section A.2). LRI consists of ten research groups, supported by systems and network staff and an administrative team. The laboratory occupies 4500 square meters located in building 490 on the Orsay Campus of the University. The primary infrastructure assets include a highperformance network with a shared cluster and an extensive library (section A.3).

Administratively, LRI is a joint laboratory (*Unité Mixte de Recherche UMR8623*) between Université Paris-Sud and CNRS with funding from both. The laboratory includes faculty and staff from the University as well as permanent researchers and staff from CNRS. LRI's annual operating budget (not including University or CNRS salaries) is approximately 1.6 M€. Less than one third of the operating budget is recurring funding from the University and CNRS; the rest is derived from contracts and grants (section A.4).

The research themes addressed by LRI cover a wide spectrum of computer science: algorithms and complexity, quantum computing, graph theory, fundamental aspects of communication, micro-architecture, clusters and grid computing, software engineering, programming, human-computer interaction, databases, inference systems, data mining, machine learning, and bioinformatics. Such diversity is one of the strengths of the laboratory, fostering research at the boundaries across areas, where the potential for innovation is highest. For example, two years ago, researchers from four groups decided to create a transversal team in BioInformatics. They have now formed the new BioInformatics group with an original research approach that combines concepts from several different domains, e.g, machine learning and randomized algorithms in order to predict relevant biological information in genomic data.

The rest of this report summarizes the activity of the laboratory over the 2000-2004 period (section A.5) and presents its research plan for the next four years (section A.6).

A.1 / Introduction

#### Overview

#### Organisation chart



A.1 / 4

#### Current and forth-coming four-year plans

French academic laboratories affiliated with universities are funded on a four-year cycle, called the *Contrat Quadriennal*. The current four-year plan covers the period from 2002 to 2005 and we are now planning for the next period, from 2006 to 2009. This report serves an essential role in the evaluation process, providing an opportunity to assess current research areas while announcing those for the next four-year period. It covers the period from January 2000 to the fall of 2004. This period was chosen as it is adjacent to the period covered by the previous report (January 1995 to summer 1999) and has the same length.

The 2002-2005 plan proposed two major research initiatives: to create a new research federation and to launch or increase research in specified areas (section A.5). LRI played a major role in the establishment of the PCRI (*Pôle Commun de Recherche en Informatique*) that includes LRI, INRIA Futurs and LIX (the Computer Science laboratory at Ecole Polytechnique, joint with CNRS). Between 2002 and 2004, LRI co-founded six joint research projects within the PCRI and approximately half of the laboratory's researchers are now involved in PCRI projects.

LRI also successfully launched research in the five key areas specified in the plan:

- grid computing,
- human-computer interaction,
- bioinformatics,
- quantum computing, and
- mobile networks.

The first two areas, Grid Computing and Human-Computer Interaction, were established as new PCRI projects. A new BioInformatics group was formed and two new teams, Quantum Computing and Mobile Networks, were created within existing groups in the laboratory.

The 2006-2009 plan also proposes two research initiatives. First, the laboratory plans to take advantage of the high concentration of research centers and high-tech industry surrounding the Orsay campus. LRI will play a key role in a new large-scale initiative (*Pôle de compétence*) in software-based and digital technologies. Second, LRI proposes to target the following new research areas (section A.6):

- embedded systems,
- masses of data,
- dynamic networks, and
- robotics.

The first three involve collaborations among existing groups at LRI while the fourth will be developed in collaboration with other laboratories at Paris-Sud. All four areas have the potential to develop into full-fledged LRI research groups or PCRI projects and to take advantage of external collaborations within and beyond the University.

#### Summary of scientific activity

The 2000-2004 period has produced an impressive amount of research. Specific activities are described in detail in section B; here, we provide a statistical summary. The radar views on the next page show five measures for each group:

#### Introduction

Overview

- Group size as of January first, 2004,
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international),
- Number of defended Ph.D.s and habilitations.

Statistics for each group are displayed as a percentage of the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show belowaverage performance, relative to the laboratory. These diagrams also appear in the description of each research group together with a similar diagram for the 1995-1999 period for comparison.



As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures. For example, some research areas are more amenable to contract funding, while some research communities publish in journals rather than conferences and therefore have fewer program committees. Overall, these statistics show an outstanding level of activity across the groups. The table below summarizes the absolute numbers and compares them with the total for the 1995-1999 period.

Intr	no	116	F1/	n
IIICI	JUU	u	LIV	

Overview

	JI UIE 1993-199	9 penou.			
		Wild	utions . S	grants ised)	mmittees
2000-2004	Members	Major pub.	Contracs annu	Program C	ph.C
GraFComm	23	150	175	32	3
Algo	23	157	359	20	10
Prog	25	125	378	43	10
BD	10	68	157	29	6
Démons	11	57	137	12	13
Archi	18	59	410	30	10
Parall	13	59	103	16	8
IASI	18	76	185	54	4
I&A	17	86	200	58	7
Bioinfo	11	27	153	17	1
Total	169	864	2257	311	72
Total (1995-1999)	153	636	1160	201	86
Evolution 1999->2004	+10%	+36%	+94%	+55%	-16%

#### Publications.

Over the 2000-2004 period, members of the laboratory authored or co-authored 864 major publications: 332 articles in international peer-reviewed journals, 28 articles in national, peer-reviewed journals, 27 books or edited books, 46 book chapters and 431 articles in major international conferences. This represents 1.2 major publications per researcher per year<sup>1</sup>. If we also count other publications such as national conferences, other conferences and workshops, reports, doctoral dissertations and dissemination of scientific knowledge, the laboratory produced a total of 1266 publications or 1.7 publications per researcher per year.

#### Contracts and grants.

Total funding from contracts and grants secured during the 2000-2004 period has almost doubled when compared with the previous period. The above table lists annualised funding, to make comparisons across groups easier. The total, non-anualised amount was 2.5M€ for 1995-1999 and 4.7M€ for 2000-2004, representing an 83% increase. Over the 2000-2004 period, LRI was awarded 15 European IST projects (including 3 networks of excellence), 12 joint projects with industry in the RNTL and RNRT national programs, 9 contracts with industry partners and a record 34 projects in the national ACI research programs, most in the programs on masses of data (ACI Masses de Données) and computer security (ACI Sécurité Informatique). LRI members also participated in 23 national CNRS research groups (Actions Spécifiques) and 26 other research projects, most publicly funded.

#### Editorial boards and program committees.

Over the 2000-2004 period, LRI members participated in 47 journal editorial boards, chaired 18 program committees of national or international conferences, participated in

<sup>1 -</sup> This number assumes equal weights for full-time researchers, faculty and Ph.D. students, which is somewhat misleading if we consider that faculty members spend only about half their time on research and that Ph.D. students rarely publish in their first year or two.

191 program committees for international conferences and 55 program committees for national conferences<sup>2</sup>. Lab members are also actively involved in evaluating research, e.g., serving as experts for international programs, steering and scientific committees, and volunteer their professional services in both research and scientific organisations (the individual group descriptions for details).

#### Ph.D.s and habilitations.

62 Ph.D.s and 10 *habilitations* were defended during the 2000-2004 period. This is 15 fewer Ph.D.s than during the previous period, due to the drop in Computer Science doctoral students during the «Internet bubble» (section A.2 below). Lab members remain heavily involved in Masters and graduate programs at Paris-Sud as well as other Universities. The new Master in Computer Science program at Paris-Sud attracts a high number of students and provides a steady stream of new doctoral students.

#### Collaborations.

Although collaborations are not counted in the above statistics they represent an important part of the scientific activity of the laboratory. As seen in the group descriptions, lab members collaborate both formally and informally at local, national, European and international levels, often leading to joint publications. Such collaborations are thoroughly encouraged: over the 2000-2004 period, the laboratory has hosted 90 long-term visitors. While most collaborations involve academic institutions, a significant number are with industrial research centers and industrial partners.

As these statistics indicate, LRI is a dynamic laboratory, actively involved in the research community. More details about each research group can be found in section B of this report, while the rest of this section presents the laboratory as a whole.

#### Report outline

The first part of this report covers the evolution of laboratory personnel (A.2), the infrastructure made available to all laboratory members (A.3), the budget and financial overview (A.4), the summary of the current four-year plan (A.5) and an outline for the next four-year plan (A.5). The second part of the report covers each of the ten research groups in the laboratory, according to the following outline:

- 1 Research group members and long-term visitors;
- 2 Detailed research description, including participation in PCRI projects when applicable;
- 3 Highlights of the past four years and a radar view summarizing research activity;
- 4 Honours, including prizes, awards and keynote addresses;
- 5 Evaluation of research, including participation on editorial boards and program committees;
- 6 Volunteer professional service, including management positions in scientific organisations, organisation of conferences and scientific events, and participation in working groups;
- 7 Contracts and grants;
- 8 Collaborations;
- 9 Dissemination and technology transfer, including patents and licences, summer schools, and invited seminars;
- 10 Training and education (doctoral and post-doctoral), including the list of defended doctoral dissertations;
- 11 Publications.

<sup>2 -</sup> Some communities confer "honourary" membership on program committees. To mitigate this effect, we count the participation of a single person in a particular conference series at most twice in the fouryear period. Without this rule, the participation numbers climb to 238 for international conferences and 80 for national ones.

#### Personnel

The following table summarizes the evolution of the laboratory personnel between 2000 and 2004. The laboratory currently has nearly 180 members, up from about 150 in 2000. These numbers do not include the 50 to 80 visitors and interns who visit the laboratory each year.

CNRS personnel						
	Position	2000	2001	2002	2003	2004
Researchers						
Research directors	DR	2	2	4	4	5
Research scientists	CR	12	10	11	11	9
Total researchers		14	12	15	15	14
Technical and administrative staff (ITA)						
Research engineers	IR	3	3	3	3	3
Engineers	IE	2	2	3	3	2
Associate engineers	AAR	1	1			
Assistant engineers	Al	1	1	1	1	1
Technicians	Т	3	3	4	4	4
Total staff		10	10	11	11	10
University personnel						
Permanent faculty						
Professors	PR	23	22	23	24	22
Assistant professors	MC	35	37	36	35	42
Lecturers	А	1	1			
Total permanent faculty		59	60	59	59	64
Temporary faculty	ATER	4	8	3	13	14
Technical and administrative staff (ITARF	)					
Engineers	IGE	1				
Assistant engineers	ASI			1	1	1
Technicians	Т	2	2	2	1	1
Administrative assistants	ADA	1	1	1		
Associate technicians	AST	1	1		1	1
Assistant technicians	ADT				1	1
Total staff		5	4	4	4	4
Doctoral students		54	49	67	57	63
Post-doctoral fellows		1	6	9	7	5
Visiting professors						3
Total personnel		147	149	168	166	177

#### Introduction

Personnel

The following graphs show the above data in graph format as well as an age pyramid as of January first, 2004.



#### **LRI Personnel**

#### **Age Pyramid**



#### **Researchers and Faculty**

The total number of permanent faculty has grown from 73 in 2000 (14 CNRS researchers and 59 University faculty) to 78 in 2004 (14 CNRS researchers and 64 University faculty). The total number of CNRS researchers was stable (14), however the number of research directors grew from 2 to 5 while the number of research scientists dropped from 12 to 9. This is due to the promotion of two research scientists as research directors and several research scientists obtaining professor positions at other universities. Over the past four years, only three CNRS research scientists have joined the lab. However, given the very low number of positions available nationwide, this is, in fact, a good result.

The number of professors is almost stable (22 vs. 23), although one internal promotion and two external recruitments did not fully compensate for four departures. The overall increase in faculty is due to the significant increase in assistant professors (42 vs. 35). These numbers hide an even more significant movement: 13 assistant professors left the laboratory, mostly for professor positions at other universities, while 20 new assistant professors were recruited. These numbers show that the laboratory is very dynamic and that our assistant professors are heavily sought by other universities. The Computer Science department (which includes LRI and another laboratory, LIMSI), has also contributed by convincing the University to transfer faculty positions to Computer Science when professors in other departments retire (*BQR emploi*). During the 2000-2004 period, positions for two professors and one assistant professor were obtained in this way.

The number of temporary faculty personnel (*ATER*) has been growing steadily from four in 2000 to 14 in 2004. These positions are held mostly by doctoral students in their final year. The relatively high turnaround of assistant professor positions described above directly affects the number of available temporary positions: when a professor or assistant professor leaves the laboratory, either temporarily or permanently, his or her position is made available for one year, sometimes two, for recruiting temporary faculty. The high number of such positions is an indication of the dynamic character of the laboratory.

Finally, the above numbers count only members of the laboratory and do not include permanent researchers involved in PCRI projects who are not members of LRI. Fully 50 non-LRI researchers, faculty and doctoral students work in the six PCRI projects that involve the laboratory: eight permanent senior positions (professors and research directors), nine permanent junior positions (assistant professors and research scientists), 23 doctoral students and ten support staff and temporary personnel.

#### Administrative and technical staff



The administrative staff consists of six members under the supervision of a chief administrator. The technical staff consists of a network and systems support group of five people, as well as one person in charge of building maintenance and one person in charge of reception for the whole building (not just LRI). Each group is organized with individuals specialised in a type of task. Tasks for the administrative staff include: personnel, contracts and

Laboratoire de Recherche en Informatique / 2000-2004 report

#### Introduction

Personnel

grants, travel, library, general administration. Tasks for the technical staff include: network management, computer security, shared cluster management and general equipment. Most of these tasks are highly specialised due to the evolution of administrative rules on one hand and the evolution of the technical infrastructure and the diversity of computer systems on the other. Nevertheless both the administrative and technical staff are able to address any issue even if the person in charge is not available. This enables the staff to handle requests at any time and for the laboratory to remain open every weekday throughout the year.

Running a laboratory of 180 people with an administrative and technical staff of 14 is a major challenge, and our staff do an excellent job with the resources they have. However, while not uncommon in France, this low level of support hurts our international competitiveness. Only one new person has joined the administrative staff over the past five years and one new person will join the technical staff in January, 2005. This is far from sufficient to provide the research groups with the support they need and deserve. A laboratory the size of LRI could readily use two additional administrative staff to develop better contacts with industry, to support researchers when writing grant proposals and to improve communication and public relations. Two additional technical staff would allow those research groups involved in substantial software development to better maintain and distribute their software.

#### Doctoral students and post-docs

The number of doctoral students increased from 58 in 2000 to 76 in 2004 and includes both doctoral students and temporary faculty on *ATER* positions who are in their final year of a Ph.D. This sharp rise follows a significant decline toward the end of the 1990's when highly-skilled computer jobs industry became very attractive. Now that the «Internet bubble» has settled, more and more Masters students are continuing into the doctoral program. The University Master's program in Computer Science is also a great success, with close to 100 students. The current number of doctoral students is similar to that of 1998, with the fewest in 2001 (53 doctoral students). Note that approximately two thirds of doctoral students obtain government grants (*allocation de recherche*), while the other third are funded through research contracts. The number of *CIFRE* contracts, in which a company hires a doctoral student who then splits his or her time between the company and the laboratory, is also increasing.

The number of doctoral dissertations defended over the past four years reflects the dip in new doctoral students at the end of the 1990's. Nevertheless 62 Ph.D.s were awarded during this period, compared with 76 Ph.D.s over the 1995-1999 period. Ten *habilitations* were defended in 2000-2004, compared to nine in 1995-1999. (Note: an *habilitation* is defended six to 10 years after one's Ph.D. in order to apply for a professor position.)

The number of post-doctoral fellows has been slowly increasing but is still too low. We would normally expect their number to be about one quarter of the number of doctoral students, i.e. about 20. In good years, we have a third to half this number. This does not reflect low interest, but rather the scarcity of post-doctoral fellowships available at local and national levels. Our budget and accounting rules make it virtually impossible to use our resources to support post-doctoral fellows unless funding is obtained explicitly for this purpose. We can only hope that funding sources will increase and the rules will evolve to make it easier to welcome post-doctoral fellows.

#### Infrastructure

#### Laboratory space



The laboratory occupies approximately two thirds of building 490, or 4500 square meters. Most is office space: professors and research directors usually have individual offices, while other researchers, faculty and doctoral students share offices. The laboratory also has a library, a coffee room, a large machine room, several smaller rooms for technical support, two archive rooms and four meeting rooms of various sizes that the Masters program also uses for teaching.

The current space situation is such that LRI is unable to expand further. The University has already allocated additional space in building 490 in 2003 to house the In Situ PCRI project. In addition, LRI members from the ALCHEMY PCRI project moved to the INRIA building on another part of the campus, so that INRIA members of the TAO PCRI project could join their LRI colleagues. In 2004, the head of the PCRI, Marie-Claude Gaudel, moved her Software Engineering team to the INRIA building to help absorb growth in the rest of LRI and the remaining PCRI projects housed in building 490.



New space for the In Situ PCRI project

#### Introduction

Infrastructure

The available space in both buildings is limited and we require a long-term solution. The PCRI is planning a new building that will house both INRIA Futurs and LRI, to be ready in late 2006. The move to this new building will be a major event, enabling all members of LRI to be reunited and all joint LRI-INRIA PCRI projects to be co-located. However, given current growth trends, we anticipate that space in this new building will be tight and hope that discussions for constructing an extension, called PCRI2, will converge quickly.

#### Network and computing resources

The computer network is the most vital part of the laboratory infrastructure. Until early 2001, the backbone of the LRI network was made of a CISCO 7204 router with an ATM card and two ATM switches connected at 155 Mbps hooked up to 8 Ethernet switches for a total of about 400 ports. In 2001 it was decided to evolve towards a Giga-Ethernet backbone in order to reduce its cost and increase its bandwidth. The first step was to extend the existing backbone by adding CISCO Catalyst 3508 Giga-Ethernet switches and hooking them up with the ATM backbone through an ATM/Giga-Ethernet switch (CISCO Catalyst 2924M). The subsequent steps consisted of:

- connecting up to 13 switches with 48 ports each on the Giga-Ethernet backbone,
- migrating the workstations from the ATM to the Giga-Ethernet section of the network,
- adding a CISCO router/switch with 12 Giga-Ethernet ports,
- connecting the backbone to the campus network at 1 Gbps.



The cluster, the main switchboard with over 600 ports and some switches

This new infrastructure resulted in a network with over 600 ports. Individual workstations are hooked up at 100 Mpbs while servers use 1 Gbps links. The machines are grouped using virtual networks (VLANs) that correspond to research groups, specific departments or functions. For example, a wireless network has been established as a separate VLAN and covers all the meeting rooms. It is easily accessible for visitors without jeopardizing the security of the rest of the network. The laboratory also manages the computers for the Computer Science Masters program on a separate VLAN. In 2003, when the members of one of the PCRI project physically moved to INRIA's building, the LRI and campus networks were re-configured to provide them with the same access to LRI resources as if they were on-site.



#### Introduction

Infrastructure



Since the LRI network must be accessible from outside the campus, a strict security policy was established. Services controlled from outside are selected and controlled by means of network filters and a firewall. All connections from outside use encrypted channels and are logged for regular analysis. This security policy has allowed the network support staff to detect and defeat several attacks over the last few years.

More than 350 computers are hooked up to the LRI network: workstations, servers, including at least one server per research group and specialized servers, e.g. for Web or mail, printers, and a cluster platform. Many lab members have two computers, typically a desk-top machine and a laptop. The computers used in the lab are characterized by diverse hardware and software platforms, including Linux and Windows PCs as well as Apple Macintoshes. This diversity is encouraged even though it entails higher maintenance costs and greater expertise by the systems support staff. Daily backups of users' accounts are made automatically every night using both hard drives and DLT tapes, representing about 500 Gb of data. More thorough backups are made on a weekly and monthly basis.

The cluster platform has now 100 single or dual-processor PC nodes running Linux and is used for simulation and modelling by several teams: the clusters and grid research team for grid computing, the micro-architecture team for simulating processors, the artificial intelligence group for testing SAT solvers, etc. Finally, two rooms with special equipment are available to all lab members: a desktop publishing room with several multimedia work-stations, scanners and colour printers including an A0 printer for posters and a video-conferencing room with a PictureTel system for 384 Kbps ISDN video-conferencing.

Workstations, servers, cluster nodes, printers and network equipment are renewed regularly under the supervision of the equipment committee in order to keep an up-to-date, top-quality computing environment. The main evolutions in the short and mid-term are the development of the wireless network in order to provide better coverage, bandwidth and security, and IP-based video-conferencing and telephony.

#### Library



The main room of the library

Unlike most French academic research laboratories, LRI is fortunate to have a library and a full-time librarian. The library occupies 230 square meters including the librarian's office, reading areas, bookshelves and storage space. The librarian manages subscriptions, acquisitions and archiving, as well as exchanges with other libraries and updating of on-line catalogues.



Library budget (left) and number of journal subscriptions (right)

The annual library budget for acquisitions has been increasing steadily from  $60k\in$  in 2000 to  $84k\in$  in 2004. The library subscribes to 131 journals and 24 proceedings series, and has 5200 books; 105 books were acquired in 2003. Most of the budget is devoted to journal subscriptions, e.g.  $65k\in$  out of a budget of  $74k\in$  in 2003. The increasing cost of journal subscriptions, especially those of commercial publishers, is a constant concern that has forced us to cancel a number of subscriptions. We are now working with the Mathematics library on campus to reduce the number of redundant subscriptions.



Some journal subscriptions

The library catalogue is available to all lab members through a Web interface and the library is registered with the French network of University libraries. Over the past few years, we have allocated a yearly budget to bind older journals and thus conserve them for the long term.

Electronic access to on-line bibliographic resources has expanded greatly over the past few years. We send our research reports to INRIA as part of a national project to collect grey literature (*Griseli, programme de collecte, traitement et communication de la littérature grise française*). Many on-line catalogues are accessible to members of the laboratory. Unfortunately, full access to the ACM and IEEE Digital Libraries, two of the most important on-line resources for computer science, are too expensive for the laboratory. Discussions are underway with CNRS and INRIA to create a nationwide consortium of research laboratories that would make the subscription rate more affordable.

Even with the development of electronic publications, the library remains a major asset for researchers and students. This is particularly true for books, since they are rarely accessible in electronic form, and for old issues of journals that have not been (and may never be) digitized. In 2003, over a hundred access cards were given to outside visitors, demonstrating the value of our library not only to LRI members but to the wider community as well.



*Some of the books authored by lab members* 

Introduction

#### **Financial summary**

LRI's total annual budget is almost  $9M \in$ . However, a major portion of this budget (about 7.2  $M \in$ ) includes salaries not under the control of the laboratory, including full-time researchers (paid for by CNRS), faculty members (paid for by the University) and some doctoral students (those paid for by the Ministry of Research). Note also that this budget does not include several infrastructure costs that are covered by the University.

	2003	2004
Ministry (salaries)	983	1 074
University (salaries)	4 445	4 355
University (operating)	257	203
CNRS (salaries)	1 308	1 389
CNRS (operating)	138	138
Contracts & Grants (salaries)	450	409
Contracts & Grants (operating)	1 315	1 184
Total	8 896	8 752

Overall annual budget for 2003 and 2004 (in  $k \in$ ).





The *operating budget*, defined as under direct control of the laboratory, includes annual funding from CNRS and the University as well as the contracts and grants earned by the research groups. This represents less than 25% of the overall budget and corresponds to the upper-left quarter of the above pie charts. The table and graph below show that the amount of annual funding has been slowly decreasing over the past eight years, while contracts and grants have increased significantly, more than doubling between 2000 and 2003. The figures for 2004 are not final since the budget will not be closed until the end of the year, however we expect little additional income in the last quarter. 2003 was an exceptional year for grants, with the two national programs for masses of data and computer security (*ACI Masses de Données, ACI Sécurité Informatique*).

	1996	1997	1998	1999	2000	2001	2002	2003	2004
Université Paris-Sud									
& Ministry of Research	203	265	206	313	227	237	172	257	203
CNRS	127	162	148	136	155	178	176	138	138
Contracts and Grants	347	512	385	309	570	628	852	1315	1184

#### Introduction

Financial Summary



The expense budget consists of seven primary budget lines:

- Equipment
- Computer supplies and maintenance
- Other consumables
- Library
- Travel
- Contract salaries
- Development fund

The graph below shows the evolution of expenses over the 2000-2004 period. (The figures for 2004 are not plotted since the budget will be closed at the end of the year and many expenses occur in the final quarter.) Annual funding from the University and CNRS finance basic equipment, supplies and consumables, the library and maintenance of key equipment. Contracts and grants fund extra equipment, travel and contract salaries. As a consequence, variations in contract funding are reflected directly in these budget lines.

	2000	2001	2002	2003	2004
Equipment	257	244	334	321	163
Computer supplies & maintenance	75	68	88	88	62
Other consumables	100	78	110	93	62
Library	70	72	84	125	50
Travel	183	143	316	308	262
Contract salaries	219	211	264	299	234
Development fund	3	16	4	48	16



*Expenses since 2000 (in k \in). Note: 2004 numbers are not yet final and do not appear in the graph.* 

The development fund was formally established in 2002. We put aside 10% of all resources, excepting contract salaries and grants that fund networking activities, e.g., working groups. The fund finances activities that foster research and helps research groups that have fewer contracts and grants. Thus far, the fund has been used primarily for equipment for newcomers and for travel, e.g. external members of doctoral committees and junior researchers with papers accepted at major conferences. As the fund increases, it will be possible to fund other research activities.

The laboratory is financially sound. However, the growing dependence upon contract funding increases pressure on researchers to find contracts and complicates financial management, due to the diversity of rules applicable to multiple funding sources. We believe that maintaining a balance of two thirds contract funding and one third government funding is satisfactory for the operating budget. We do not plan to increase our efforts in finding additional contract funding, at least as long as the current general funding scheme for academic research remains in place.

#### A.5 / Introduction

#### Summary of current four-year plan

The 2002-2005 plan proposed two major research initiatives. The first announced the creation of the PCRI, which fosters joint research projects between LRI, INRIA and LIX. The second identified five key new areas to be established and seven existing areas that needed additional support.

#### PCRI

The *Pôle Commun de Recherche en Informatique* (PCRI) was formally established in January 2002 when CNRS, École Polytechnique, INRIA and Université Paris-Sud signed a joint agreement to create a federation of three laboratories: LRI (joint laboratory of Université Paris-Sud and CNRS), INRIA Futurs (a new research centre of INRIA) and LIX (joint laboratory of École Polytechnique and CNRS). The main goal of PCRI is to establish joint projects among

these laboratories as well as with external laboratories. The research centers on the following topics:

- Software quality and efficiency;
- Data mining and information integration in a distributed environment;
- Interaction with computer systems;
- Global computing and computer security over the Internet.



en informatique

As of 2004, the PCRI has formed eight projects, seven of which involve LRI. Of the 285 members in the three laboratories, 110 are involved in a PCRI project. The projects involving LRI are, in chronological order of their creation:

- LOGICAL: Logic and computing. INRIA Futurs, LRI Demons and LIX (Gilles Dowek, head).
- GEMO: Integration of data and knowledge over the web. INRIA Futurs and LRI IASI (Serge Abiteboul and Marie-Christine Rousset, heads).
- In Situ: Situated Interaction. LRI Prog and INRIA Futurs (Wendy Mackay, head).
- Grand Large: Global parallel and distributed computing. LRI Parall, LRI Archi and INRIA Futurs (Franck Cappello, head).
- TAO: Machine learning and optimization. LRI I&A and INRIA Futurs (Marc Schoenauer and Michèle Sebag, heads).
- ALCHEMY: Architectures, Languages and Compilers to Harness the End of Moore's Years. INRIA Futurs and LRI Archi (Olivier Temam, head).
- PROVAL: Proofs of programs. LRI Demons and INRIA Futurs (Christine Paulin and Claude Marché, heads).

Detailed descriptions of each project appear within the descriptions of their associated research group in Section B of this report.

LRI is proud of its role in the creation of the PCRI and the successful launch of these new projects. The PCRI steering committee is now examining additional project proposals that involve LRI and we anticipate even greater cooperation among the three founding laboratories and external laboratories.

#### Introduction

Four-year Plan

#### Development of new areas

The following five areas were selected as the main focus of the current four-year plan: Global computing, Human-computer interaction, Bioinformatics, Quantum computing and Mobile networks.

#### **Global Computing**

A new PCRI project, Grand Large, was created in 2003 jointly with INRIA. INRIA recruited a research director and a research scientist, while LRI recruited an assistant professor. Grand Large is heavily involved in both national and international projects, in particular the GridExplorer and Grid5000 projects that seek to establish a nation-wide GRID platform. The Cluster and Grid team in the Archi group participated in a multi-disciplinary research project within the university funded by the Ministry of Research (*PPF - Programme PluriFormation*) and an assistant professor spent two years at LAL (*Laboratoire de l'Accélerateur Linéaire*), a Physics laboratory, to work on this project. Finally, a collaboration with the Parall group has been formed to study the use of self-stabilizing algorithms for GRID computing and the Cluster and Grid team is planning to join this group.

#### **Human-Computer Interaction**

The In Situ PCRI project was created in 2002 jointly with INRIA. INRIA recruited a research director and a research scientist that year, and another research scientist in 2004. A CNRS researcher also joined the group in 2004. The group has grown very quickly from 5 original members to over 20 in two years. The group develops original research with a multi-disciplinary team including computer scientists, psychologists, and designers. In Situ is involved in several projects at the national and European levels (*RNTL, ACI Masses de Données, IST Disappearing Computer*). The group is one of the largest academic research groups in Human-Computer Interaction in France and has excellent international visibility.

#### **Bioinformatics**

As described at the beginning of this report, a new team was first established with researchers from four different research groups (Algo, IASI, I&A, BD). LRI recruited a professor in 2002 and an assistant professor in 2003. The team was converted into a research group in 2002 and now has 11 members. This group collaborates with Biology laboratories both on and off campus. The group participated in a multi-disciplinary program in Bioinformatics, funded by the Ministry of Research (*PPF - Programme Pluri-Formation*). The group is very active at the national level and is acquiring visibility at the international level.

#### **Quantum Computing**

An active team has been created within the Algorithms and Complexity group. LRI recruited an assistant professor and two CNRS researchers in 2000/2001. The team studies the computational potential of quantum computing and has contributed complexity measures as well as new quantum algorithms. It also collaborates with physicists whose objective is to study the practical problems raised by the implementation of quantum computers. The team has acquired an international reputation in a short period of time, and is involved in numerous projects, especially at the European level.

#### **Mobile Networks**

The Software Engineering group is hosting a new team working on mobile networks. LRI recruited an assistant professor in 2002 and a professor in 2004. Although the team has not yet reached the critical mass necessary to become an independant group, it is actively collaborating with INRIA's Hipercom project, which will soon become a PCRI project. The group works on mobile and ad hoc networking, specifically protocols and algorithms that optimize communication. It is involved in several national projects and participates in international standardisation efforts.

#### Development of existing areas

The following existing areas in LRI were identified in the current plan as deserving additional support.

#### Software certification

The Logical PCRI project and the Demons group have continued their work on the certification of software systems. In addition to the *Coq Proof Assistant*, new software tools were developed to certify *C* and *Java* programs. A new PCRI project, called Proval, is being established to focus on program verification.

#### Information integration

The IASI group focused its research on the hot new topic of information integration, leading to the creation of the GEMO PCRI project. GEMO is one of the first Artificial Intelligence and Database research groups in France. The group is well known internationally and has a long-standing collaboration with France Telecom R&D. It participates actively in the world-wide collective initiative towards a Semantic Web, which seeks to improve retrieval of data and resources via semantic mark-up based on ontologies. The database group also works on information integration from a slightly different perspective, with an emphasis on digital libraries.

#### Data mining

The change of the head of the I&A group resulted in a re-orientation of some of its research. The TAO PCRI project was established to foster the cross-fertilization of machine learning, knowledge discovery and evolutionary computation. Data mining is addressed from several perspectives, with applications in such diverse domains as health risk assessment, human resource management and road traffic.

#### Self-stabilization

The Parall group has developed the self-stabilisation area with a focus on proving the properties of self-stabilizing algorithms and designing algorithms that scale to large distributed systems. This applies in particular to grid computing and is being developed in the context of the Grand Large PCRI project.

#### Software testing

The software engineering team has recruited two assistant professors on the topic of software testing, to compensate for the departure of a CNRS researcher. The team focuses on deriving test cases from specifications and has introduced probabilistic and statistical approaches to uniform generation of combinatorial structures.

#### Network communication

The GraFComm group suffered several departures but a new team on network optimisation was created with the recruitment of a professor. Research on information dissemination and routing has evolved to address loosely coupled networks, such as the Internet or peer-to-peer networks, rather than highly coupled systems, such as parallel computers. The team now focuses on dynamic networks with applications to complex systems.

#### **Probabilistic algorithms**

This area has suffered with the departure of a professor and no new recruitment. The team is still very productive but needs to be reinforced. Probabilistic algorithms are an important area of research since they offer the only known approach to certain complex problems.

#### Introduction

Four-year Plan

Overall, with the exception of probabilistic algorithms, the objectives of the current fouryear plan have been reached. They have led to reinforced collaborations among the groups as well as with outside groups, especially through the PCRI joint projects. Other areas that were not in the plan have emerged as well, including robotics, evolutionary computing, spatio-temporal databases, embedded systems, languages and systems for semi-structured data (XML). This is a natural phenomenon in a laboratory the size of LRI, and one that we encourage since novel research directions cannot, by definition, be planned in advance.

#### A.6 / Introduction

#### Long-range plan for 2006-2009

The 2006-2009 plan proposes two major research initiatives: to expand its role in regional research networks and to target specified research areas. First, LRI will increase its participation in joint PCRI projects and begin to include external laboratories, outside the PCRI federation. The laboratory also plans to participate actively in the Num@tec innovation platform for software technologies. Num@tec is a joint initiative between CEA and CNRS for scientific research, technological innovation and technology transfer, with a plan to build new buildings (17000 square meters) for hosting research, R&D projects and technology transfer activities. The PCRI and Num@tec together will be at the core of a larger regional pole on digital technology; LRI plans to play a key role in this large-scale initiative.

The 2006-2009 plan will also target specific research areas. LRI will actively develop three existing areas: grid computing, bioinformatics and quantum computing. The laboratory will also create four new areas: embedded systems, masses of data, dynamic networks and robotics. We will, of course, continue to support existing areas of excellence and will continue to identify and explore emerging areas.

#### Development of existing areas

#### Grid computing

The Grand Large PCRI project is coordinating a major national effort in grid computing, the Grid5000 project. This area requires new recruits in order to maintain its leadership at the national level and to leverage its influence at the international level. Since the main users of grid systems come from other disciplines, we will seek additional collaborations with, for example, Physicists and Biologists.

#### **Bioinformatics**

The new Bioinformatics team needs new recruits in order to reach a critical mass. Also, the multi-disciplinary nature of this project requires partners from different fields who can collaborate closely together. We plan to propose a joint group between LRI and a Biology laboratory in order to expand the group size and strengthen the cross-disciplinary research approach.

#### Quantum computing

LRI is now a major European actor in quantum computing, with research in quantum algorithms, quantum complexity and quantum protocols. In addition to supporting this team, we plan to develop collaborations with Physicists so as to bridge the gap between the models and the creation of quantum computers.

#### Development of new areas

#### **Embedded systems**

Embedded systems are hardware+software systems that are integrated into products and subject to environmental constraints. These systems, although increasingly built into the fabric of everyday life, also present new challenges. We will focus on two key problems: how to validate and certify embedded systems, especially those involved in critical systems, and how to optimise these systems under environmental constraints such as power consumption, real time and low cost. Three LRI groups will cooperate on this area: Proofs and Programs (certification), Software Engineering (validation) and Parallel Architectures (optimization). Embedded systems are of direct interest to the Num@tec initiative and will probably become a key area of the European seventh framework program.

#### Masses of data

The advent of distributed sources of semi-structured and unstructured data, especially visible with the explosion of the Web, has dramatically redefined the fundamental problems of data management. Specifically, we seek to address issues related to data warehouses, data mining, knowledge acquisition, semantic Web, distributed computing, and information visualisation. Several LRI groups work in this domain, including Databases, Inference & Machine Learning, Artificial Intelligence & Inference Systems, Parallelism, Software Engineering. This area has started to emerge with multiple grants secured by LRI from the French Ministry of Research program, the *ACI Masses de Données* launched in 2002, and we will open a professor position in this area in 2005.

#### **Dynamic networks**

Dynamic networks, whose topologies change, sometimes radically, over time, are becoming a unifying research area not only for telecommunications, but also for transportation and social networks. Fundamental research in dynamic networks has applications in many areas, including mobile, satellite, ad hoc, peer-to-peer and sensor networks, as well as grid computing, the Web, and even robotics. LRI has recognized expertise in the necessary areas from several groups: Graph Theory and Fundamental Aspects of Communication, Software Engineering (mobile networks team), Parallelism, Databases, and Inference & Machine Learning (robotics team). We plan to create a new focus on dynamic networks, particularly with respect to security and access to information.

#### Robotics

Robotics is a recent research area for LRI. Our focus is on using machine learning and evolutionary computing for robot controllers (Inference and Machine Learning group). We plan to develop this area in collaboration with other laboratories on campus, in particular LIMSI and IEF (*Institut d'Electronique Fondamentale*). By creating a critical mass and taking advantage of the complementary expertise of these three laboratories, we will be able to address real-world problems such as the interactions between control and behaviour. This area is of particular interest to Num@tec and will allow us to collaborate with CEA, which has recognized expertise in robotics.

Long-range Plan


Laboratoire de Recherche en Informatique / 2000-2004 report

B.1 / 27

# équipe Théorie des Graphes et Fondements des Communications

## **Responsable : Pierre Fraigniaud**

équipe fut l'une des composantes initiales du LRI, lors de sa création il y a plus de trente ans. Ses principaux domaines d'études sont la théorie structurelle des graphes et la recherche opérationnelle. Ses domaines d'applications sont la conception de réseaux de télécommunication et d'algorithmes de communication. Grâce à ses compétences en théorie des graphes, l'équipe s'intéresse également à des problèmes provenant d'autres domaines de l'informatique, comme l'informatique distribuée, l'algorithmique discrète, et le calcul parallèle. L'équipe est très active et très productive, comme en témoigne sa longue liste de publications scientifiques dans des journaux ainsi que dans les actes de conférences, la participation de ses membres à de nombreux comités de programmes de conférences, et leurs nombreuses invitations dans des rencontres et séminaires internationaux.

L'équipe a une approche plutôt fondamentale des problèmes scientifiques. Cependant, ceci ne se fait pas au détriment de collaborations industrielles, comme en témoignent plusieurs participations à des projets RNRT, ainsi qu'une participation à un projet européen IST. Le recrutement comme professeur d'Abdel Lisser (en provenance de France Telecom R&D), et celui de David Forge comme maître de conférences, a permis à l'équipe d'investir de nouveaux champs d'études comme la théorie des matroïdes et la programmation mathématique. Ces recrutements ne font cependant que compenser les départs de deux membres actifs de l'équipe en 1999 et 2000 : celui de Dominique Sotteau, actuellement attachée scientifique à l'ambassade de France au Canada, et celui de Dominique Barth, actuellement professeur à l'université de Versailles. De plus, trois membres de l'équipe ont pris leur retraite ces deux dernières années : Odile Favaron, Marie-Claude Heydemann, et Maryvonne Mahéo. L'équipe leur souhaite de profiter au mieux de leur nouvelle vie. Cependant, lorsque l'on constate qu'O. Favaron a, à elle seule, cosigné plus de 30 articles parus dans des revues ces quatre dernières années, on mesure l'impact de ces départs sur le fonctionnement scientifique de l'équipe.



Research groups
GraFComm
Introduction

# Graph Theory and Fundamental Aspects of Communications

Head: Pierre Fraigniaud

s one of the founding members of LRI, the group goes back more than 30 years. Today, it tackles problems from structural Graph Theory and Operational Research, with applications to network design and network provisioning. Thanks to its expertise in graph theory, the group also addresses problems coming from other fields of computer science, such as distributed computing, discrete algorithms, and parallel computing. The group is very active and productive, as witnessed by its long list of publications in international journals and conference proceedings, by the participation of its members in many conference program committees, and by numerous invitations to international conferences and seminars.

The group addresses problems that are mostly of fundamental flavor. Nevertheless, we also collaborate with industry, as illustrated by several participations in national RNRT projects and one european IST project. Over the 2000-2004 period, the appointments of Abdel Lisser, who left France Telecom R&D, as professor, and of David Forge as assistant professor, enabled the group to investigate new domains: matroid theory, mathematical programming, etc. These appointments came however after two productive group members left in 1999-2000: Dominique Sotteau, now Scientific Officer at the French Embassy in Canada, and Dominique Barth, now professor at Versailles university. In addition, three group members retired in 2003 and 2004: Odile Favaron, Marie-Claude Heydemann, and Maryvonne Mahéo. The group wishes them happiness in their new life. However, considering that Odile Favaron alone co-authored more than 30 journal articles over the past four years, one can easily understand that these departures will create a real overthrow in the scientific life of the group.

Laboratoire de Recherche en Informatique / 2000-2004 report

# **Research Group Members**

## Personnel as of 01/01/2004

Full time faculty			
Name	First Name	Position*	Institution
BERTHOME	Pascal	MC	IUT Orsay
DELORME	Charles	MCHC	Université Paris XI
DJELLOUL	Selma	MC	IUT Fontainebleau
FAYARD	Didier	PR1	IUT Orsay
FLANDRIN	Evelyne	PR2	IUT Paris V
FORGE	David	MC	Université Paris XI
FRAIGNIAUD	Pierre	DR2	CNRS
HEYDEMANN	Marie-Claude	PR1	IUT Orsay
KOUIDER	Mekkia	MCHC	Université Paris XI
LI	Нао	CR1	CNRS
LISSER	Abdel	PR2	Université Paris XI
MAHEO	Maryvonne	MCHC	Université Paris XI
SACLE	Jean-François	MC	Université Paris XI

Doctoral students			
BENAJAM	Wadie	ATER	Université Paris XI
DANG NGOC	Frédéric	D	Grant France Telecom R&D
FAIK	Taoufik	ATER	Université Paris XI
GANCARZEWICZ	Grzegorz	D	Grant France/Poland
GASTAL	Lynda	AM	IUT Sceaux
GAURON	Philippe	AM	IUT Orsay
ILCINKAS	David	AM	Supélec
NGUYEN	Le Huy	D	Grant CNRS

Temporary personnel			
CADA	Roman	Post-doc	Grant Université Paris XI
FAVARON	Odile	Associated	Université Paris XI
MENG	Jixiang	Invited Prof.	Grant Xinjiang University

\* See the glossary for acronyms.

Long term vi	sitors					
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
BACSÓ	Gabor	Hungarian	Acad. of Sciences (Budapest)	01/07/02	31/07/02	Invited Prof. Paris XI
BARRIÈRE	Eulalia	Spanish	UPC (Barcelona)	26/06/00	31/07/00	PICASSO
CADA	Roman	Czech	Univ. of Plzen	02/07/00	15/07/00	BARRANDE
DUCKWORTH	William	Australian	Macquarie Univ. (Sydney)	15/09/02	29/09/02	ARC/CNRS
EL SAHILI	Amine	Lebanese	Univ. of Beyrouth	03/06/02	08/07/02	Assistant Prof.
FOMIN	Fedor	Norwegian	Univ. of Bergen	10/01/04	17/01/04	ACI PairAPair
GAIVORONSKI	Alexei	Norwegian	Univ. of Trondheim	05/05/04	05/06/04	Invited Prof. Paris XI
GENEST	François	Canadian	Univ. of Montréal	15/11/03	30/11/03	EGIDE
HENNING	Michael	South Africa	n Univ. of Natal	01/07/02	31/07/02	Invited Prof. Paris XI
HOANG	Chinh	Canadian	Waterloo Univ. (Ontario)	05/07/01	29/07/01	Invited Prof. Paris XI
HU	Zhiquan	Chinese	Central China Normal Univ.	24/06/02	30/06/03	MAE
KRISTIANSEN	Petter	Norwegian	Univ. of Bergen	17/02/03	21/02/03	
MANS	Bernard	Australian	Univ. Macquarie (Sydney)	31/01/00	30/06/00	CNRS/ARC
				15/09/02	29/09/02	CNRS/ARC
MARCZYK	Antoni	Polish	Acad. of Science (Krakow)	01/10/00	30/10/00	POLONIUM
ORDAZ	Oscar	Venezuelian	Univ. Centrale	01/08/01	31/08/01	
PELC	Andrzej	Canadian	UQO (Gatineau)	20/03/04	26/03/04	ACI PairAPair
RASJBAUM	Sergio	Mexican	Univ. of Mexico	13/05/04	20/05/04	INRIA Grand Large
ROSENBERG	Arnold	American	UMass (Amherst)	29/01/00	31/05/00	CNRS "Poste rouge"
RYJÁČEK	Zdenek	Czech	Univ. of Plzen	02/07/00	15/07/00	BARRANDE
SHPARLINSKI	Igor	Australian	Univ. Mcquarie (Sydney)	17/01/01	31/01/01	CNRS/ARC
SHU	Jinlong	Chinese	Eastern China Normal Univ.	25/10/01	25/10/02	Grant Chinese Gov.
TUZA	Zsolt	Hungarian	Acad. of Sciences (Budapest)	14/06/00	30/06/00	CNRS/Acad Sc.
TVRDIK	Pavel	Czech	Tech. Univ. (Prague)	15/02/00	04/03/00	
VESTERGAARD	Preben	Danish	Aalborg Univ.	02/07/00	18/07/00	
				01/06/03	30/06/03	Invited Prof. Paris XI
WOŹNIAK	Mariusz	Polish	Acad. of Science (Krakow)	01/10/00	30/10/00	POLONIUM
				01/10/01	25/10/01	POLONIUM
				01/04/02	30/06/02	Invited Prof. Paris V

## **Group evolution**

Some remarkable facts about the evolution of the research group:

- In July 2000, Nicolas Hanusse was appointed as Assistant Professor (MC). He left LRI one year later to become CNRS Research Scientist (CR) at LaBRI (Université de Bordeaux).
- In July 2001, Abdel Lisser was appointed as Professor (PR), after leaving France Telecom R&D. Since then, he is the leader of the "Optimization" sub-team.
- In July 2003, David Forge was appointed as Assistant Professor (MC), and joined the "Graph Theory" team.

The latter two arrivals balance two departures at the end of the 90's: Dominique Sotteau (DR CNRS), now scientific officer at the French embassy in Canada, and Dominique Barth (MC), now Professor at Université Versailles Saint-Quentin (PRiSM laboratory).

Three group members retired in 2003 and 2004: Odile Favaron (MC), Marie-Claude Heydemann (PR), and Maryvonne Mahéo (MC). As one can notice from the list of publications, these retirements will eventually have a significant impact on the scientific production of the group (Odile Favaron was still *collaborateur bénévole* in 2003-2004). Therefore,

the group expects to appoint new members in the coming years, for each of its teams: Graph Theory, Optimization, and Information Dissemination (e.g., in the domain of Dynamic Networks).

The number of Ph.D. defenses in the group was relatively low in 2000-2004. However, the number of Ph.D. students in the group increased in recent years (and two more will start in 2004), and several theses will be defended during the next two years. Finally, no Habilitation has been defended during 2000-2004, but two Assistant Professors will defend their Habilitation in 2005 or 2006.

## B.1.2 / GraFComm

## **Research description**

GraFComm is organized in two research teams: *Graph Theory* and *Communication and Network Design*. The former is devoted to the study of fundamental problems in graph theory, while the latter focuses on various types of applications in network design and network provisioning (including information dissemination, routing, protocol for mobile agents, etc.).

The cross-fertilization between these two teams is ensured by our group seminar (supervised by Mekkia Kouider), occurring once a week, and during which members of GraFComm, or their visitors, present their most recent results.

## Graph Theory Team

- Participants: Hao Li (leader), Charles Delorme, Selma Djelloul, Odile Favaron, Evelyne Flandrin, David Forge, Marie-Claude Heydemann, Mekkia Kouider, Maryvonne Mahéo, Jean-François Saclé.
- Ph.D. Students: Taoufik Faik, Grzegor Gancarzewicz.

As one of the founding members of LRI, the Graph Theory team goes back more than 30 years. The long list of results obtained by its members has earned the team an international reputation in several topics in graph theory. Motivated by the fact that a graph is a discrete topological structure that can be used to model and study large collections of problems in various areas of computer science (algorithms, logic, telecommunication, computer architecture, combinatorial optimization, etc.), the Graph Theory team conducts advanced research in the structural properties of graphs.

During the 2000-2004 period, the team carried out research in fundamental problems of graph theory. The appointment of David Forge in 2003 allowed the team to extend its competence to include matroid theory. The team focused on several core areas of extremal graph theory, and graph structure theory. Its work is very fruitful, as illustrated by almost 100 publications in international journals during the 2000-2004 period. Members of the team were able to solve important conjectures that were suggested by graph theorists of international renown, and that had been widely studied for many years. In particular, the team achieved important advances in various problems on cycles, colourings, factors, algebraic graph theory, matroids, domination theory, etc. The team collaborated with colleagues from all over the world, leading to many joint publications.

In the following, the description of the main contributions of the team are grouped in four research areas:

- Hamiltonian graph theory;
- Factoring, covering, and decomposition;
- MinMax parameters;
- Algebraic aspects;

#### Hamiltonian graph theory

The *Hamilton* problem is one of the core problems in graph theory. The *circumference* of a graph, i.e., the length of a longest cycle, is a natural and important generalization of Hamilton cycles. It is NP-hard to compute the circumference of a graph, and thus much work in the literature has focused on obtaining lower bounds for arbitrary graphs, or for some specific classes of graphs. One of the most basic and important foundation of extremal Hamiltonian theory is the famous result by Dirac (1952): the circumference of a graph is at least twice its minimum degree. As a consequence, if the degree of every vertex is at least half the total number of vertices, then the graph is Hamiltonian. Woodall conjectured in 1975 that if a 2-connected graph of order n has at least n/2+k vertices of degree at least k, then its circumference is at least  $min\{n,2k\}$ . This conjecture is one of the fifty open problems listed in the famous book *Graph Theory with Applications* (Bondy and Murty, 1976). One important contribution of the team was to prove Woodall's conjecture. In collaboration with Roland Häggkvist, Cristina Bazgan and Mariusz Wozniak, the team also obtained several results about the length of the longest cycles, and the length of the longest paths, in graphs having half their vertices with a "large" degree.

Along another way to generalize Dirac's result, Ore (1960) derived a famous sufficient condition for hamiltonicity, based on the sum of the degrees of any pair of independent vertices. This condition was generalized by members of the team in 1991 (in collaboration with Jung) into a condition on the sum of the degrees of any triple of independent vertices. Under various conditions on the sum of the degrees of any four independent vertices, the team also derived many results about hamiltonicity and about the existence of a maximum dominating cycle, i.e., a maximum cycle C such that every edge has at least one extremity in C.

The team also focused on the existence of cycles of various lengths. In particular, a graph is *pancyclic* if it contains cycles of all lengths, from 3 to the order *n* of the graph. We derived several results about pancyclicity, in particular sufficient conditions for a bipartite graph to contain all even cycles. *Cyclability* and *pancyclability* are two important generalizations of hamiltonicity and pancyclicity, respectively. The team investigated the stability, under Bondy-Chvátal's closure and Ryjáček's closure, of cyclability and pancyclability properties, for a given subset of vertices. In particular, the team derived refinements of the closure concept by considering local structures of subsets of vertices. This allowed us to obtain new sufficient degree conditions for cyclability.

## Key references:

[55] R. Faudree, E. Flandrin, M. Jacobson, J. Lehel, and R. Schelp. Even cycles in graphs with many odd cycles. *Graphs and Combinatorics*, 16:399-410, 2000.

[108] H. Li. On a conjecture of Woodall. Journal of Combinatorial Theory B, 86:172-185, 2002.

Research groups

GraFComm Research Description: Graph Theory team

#### Factor, covering and decomposition

A vast research area in graph theory is to decompose or to cover a graph using simpler structures. In several contexts (e.g., network design) such decompositions allow simplifying algorithms. In this domain, the team focused on the existence of spanning subgraphs, either regular (k-factors), or of bounded degrees ([a,b]-factors). Additional conditions, such as connectivity or degree parity, may even be required. An important contribution of the team is solving a conjecture by Kano about connected [a,b]-factors, and their minimum degree. On the other hand, the team disproved another conjecture by Kano involving the stability and the connectivity of a graph. However, by reformulating this latter conjecture, the team extended sufficient conditions implying the existence of a b-tree (i.e. a [1,b]-factor) to sufficient conditions implying the existence of connected [a,b]-factors.

The team also derived sufficient conditions for the existence of partitions into elementary cycles. Alternatively, the team has studied the minimum number of cycles c(G) required to cover the vertices. In particular, by considering the size of the neighbourhood of pairs or triples of vertices the team has obtained bounds on c(G).

It is well known that every strong tournament contains a Hamiltonian cycle. An important contribution of the team in the framework of tournaments, in collaboration with Chen and Gould, is to answer to a question asked by Bollobás in the 80's: Given a positive integer k, what is the least integer g(k) so that all but a finite number of g(k)-connected tournaments contain k vertex-disjoint cycles that span the vertices?

Last but not least, the team focuses on Gallai's famous conjecture (1960), about the decomposition of the edges of a graph into a minimum number of paths. For about 30 years, the only known bound was 3n/4 (due to Lovász), where n is the order of the graph. The team improved this bound to 2n/3 for any graph, and proved that this bound is sharp for disconnected graphs consisting of triangles.

Key references:

- [20] G. Chen, R. Gould, and H. Li. Partitionning vertices of a tournament into independent cycles. *Journal of Combinatorial Theory B*, 83:213-220, 2001.
- [38] N. Dean and M. Kouider. Gallai's conjecture in disconnected graphs. *Discrete Mathematics*, 213:43-54, 2000.
- [98] M. Kouider and M. Mahéo. 2-edge-connected [2,k]-factors in graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 35:89-95, 2000.

#### **MinMax parameters**

MinMax parameters are often involved in computing whether a graph satisfies certain properties. They are also often considered for searching specific structures, for partitioning a graph, or for assigning weights to the vertices or the edges of a graph in order to satisfy some specific properties. In fact, many maximum and minimum graph parameters have *minimum maximal* and *maximum minimal* counterparts. Most often the terms "minimum" and "maximum" refer to the cardinality of a subset, or a partition, of the vertices, and the notions of maximality and minimality are related to partial orders on subsets of vertices. For instance, the independent domination number is the minimaximal counterpart of the independence number, where the partial order is set-inclusion.

MinMax parameters are hard to compute. Thus, the team essentially focuses on:

- Deriving sharp bounds,
- Establishing connections between different MinMax parameters,
- Characterizing families of graphs for which two parameters are equal (or close), and
- Analyzing the impact of adding or deleting an edge.

The general concept of *domination* intuitively refers to the ability to reach all the vertices of the graph from a subset of vertices (the dominating set). This concept has a large number of variants, depending on various types of motivations. For instance, one can impose conditions on the structure of the dominating set (total domination, paired domination, etc.), on the way the vertices must be dominated (multiple domination, distance-*k* domination, secure domination, etc.), or simultaneously on both (double domination). The team widely contributed to this field, with roughly 25 references in the attached bibliography. In particular, members of the team have proved a relation between a specific MinMax parameter (namely the *upper irredundancy*) and both independence and chromatic numbers.

The *achromatic number* is the maximinimal counterpart of the *chromatic number* for a certain partial order defined on the set of partitions of the vertices. This parameter has been studied for more than forty years. In 1998, a refinement of this partial order gave rise to a new parameter, called *b-chromatic number*. Members of the team have studied this new parameter. In particular, they characterized classes of graphs for which the chromatic and the b-chromatic numbers are equal for every induced subgraph.

#### Key references:

- [5] G. Bacsó and O. Favaron. Independence, irredundance, degrees and chromatic number in graphs. *Discrete Mathematics*, 259:257-262, 2002.
- [68] O. Favaron and M. A. Henning. Paired domination in claw-free cubic graphs. *Graphs and Combinatorics,* to appear.
- [203] C. Hoang and M. Kouider. b-coloring of sparse graphs. Technical report, W.Laurier University, 2002.

#### Algebraic aspects

The team studied several algebraic aspects of graph theory, such as matroids, graph spectrum, Cayley graphs, etc.

*Matroid theory* was introduced in the 30's as a common generalization of linear algebra and graph theory. Since then, this theory has found applications in many branches of theoretical computer science, as well as inapplied and pure mathematics. In their study of the topology of hyperplane arrangements, and of their complements, Orlik and Solomon (1980) proved that the matroid of an arrangement catches some information on their topology via a specific structure, called now the Orlik-Solomon algebra. For these reasons, these algebras have been intensively studied in the last twenty years, using both algebraic and combinatorial methods. An important contribution of the team is related to the study of the Orlik-Solomon algebra, as well as some other related algebras (e.g., the Orlik-Terao algebra, the Cordovil algebra, etc.). In particular, the team provided several very simple combinatorial proofs of various results that were already known, but proved using non-combinatorial arguments. Matroid theory is also very central in combinatorial studies, like graph theory and optimization (see the recent book "Combinatorial Optimization" by A. Schrijver). Hence, not surprisingly, the team also focused on combinatorial problems, e.g., orientaResearch groups

GraFComm Research Description: Graph Theory team tions on graphs, by using matroids and oriented matroids as a tool.

The *spectrum* of graphs is a concept that has been used since the 50's (in particular by A. Hoffman) as a tool for proving the non-existence of graphs with too stringent conditions on regularity and diameter. In combination with optimization techniques, e.g., semidefinite programming, this concept provides rather efficient methods for bounding graph parameters such as independence number, chromatic number, max-cut, and, to a lesser extend, connectivity. Graphs with different spectra are certainly not isomorphic, but the converse is a source of problems that have been deeply investigated by the team.

*Cayley* graphs are graphs constructed from a group. These graphs have highly symmetrical properties, and, for this reason, they have found many applications in several areas of computer science, for they include hypercubes, toroidal meshes, wrapped FFT-graphs (or Butterflies), etc. Based on the notion of Cayley graphs, and using other algebraic objects, e.g., association schemes, the team investigated the construction of large graphs, e.g., voltage graphs, with specific symmetry properties. In fact, graph symmetry is by itself a wide field of investigations, e.g., determining a non-trivial symmetry in a given graph, with applications to algorithm design, e.g., routing.

## Key references:

- [33] R. Cordovil and D. Forge. Diagonal bases in Orlik-Solomon type algebras. Annals of Combinatorics, 7:247-257, 2003.
- [40] C. Delorme. Spectra and cuts. Australasian Journal of Combinatorics, 26:183-191, 2002.
- [41] C. Delorme. Laplacian eigenvalues and fixed size multisections. *Discrete Mathematics*, 276:149-159, 2004.
- [45] C. Delorme and J. Gómez. Some new large compound graphs. *European Journal of Combinatorics*, 23(5):539-547, 2002.

## **Communication and Network Design Team**

- Participants: Pierre Fraigniaud (leader), Pascal Berthomé, Didier Fayard, Nicolas Hanusse (2000-2001), Abdel Lisser.
- Ph.D. Students: Ignacio Alvarez-Hamelin, Wadie Benajam, Moaiz Ben Dhaou, Lynda Gastal, Philippe Gauron, David Ilcinkas.

The *Communication and Network Design* team tackles problems occurring in network design and network provisioning. These activities are split into two sub-teams, characterized by the fundamental tools used for solving problems rather than by the nature of the considered problems. The *Optimization* sub-team focuses its activities on the design of sophisticated mathematical optimization techniques. The *information dissemination* sub-team focuses its activities on the design and analysis of discrete and/or distributed algorithms, using tools borrowed from graph theory and combinatorics.

Obviously, these two sub-teams have lots to share. The evolution of the research activities of one of the members of the group (Pascal Berthomé), moving from the information dissemination sub-team to the optimization sub-team illustrates the close interactions between the two parts of the communication and network design team.

### Optimization

- Participants: Abdel Lisser (leader), Pascal Berthomé, Didier Fayard.
- Ph.D. Students: Wadie Benajam, Moaiz Ben Dhaou, Lynda Gastal.

The research topics tackled by the Optimization sub-team were strongly influenced by Prof. Abdel Lisser, who joined the GraFComm research group in 2001, from France Telecom R&D. Since 2001, the topics studied by the sub-team are closely related to recent advances in telecommunication network, especially concerning synchronous and optical wired backbone networks. Recently, the team initiated a new research direction, on Frequency Assignment problems in wireless GSM networks. Telecommunication network design problems vary according to the technologies, and to the numerous strategies of the telecommunication companies. This gives rise to a variety of new combinatorial optimization problems, in particular routing and survivability. In order to tackle such problems, the Optimization sub-team considered models and methods in the deterministic as well as stochastic frameworks. Specifically, the team focused on the following problems:

- 1/ Multicommodity flow models,
- 2/ SDP relaxations,
- 3/ Robust and Stochastic combinatorial optimization, and
- 4/ Resource assignment and multi-terminal flow problems.

#### Multicommodity flow models

The sub-team studied different formulations of *multicommodity netflow* problems, e.g., node-arc formulation and path formulation. Our main contributions are the design and implementation of several algorithms for solving minimum cost multicommodity netflow (MCNF) problems. MCNF problems are widely used in different areas of optimization, especially, as far as telecommunication is concerned, with applications to routing, survivability, availability, etc. In the domain of telecommunication network optimization, MCNF problems take the form of large integer linear programming problems. The Optimization sub-team considered both 0-1 integer variables and continuous formulations of these problems. In the two cases, different relaxations were studied and implemented, such as linear relaxations and Lagrangian relaxations. The sub-team solved large MCNF problems, with several million variables and constraints. This is considered among the largest solved instances of the problem. The sub-team also studied direct methods, as well as decomposition approaches. Using several decomposition methods, the sub-team extended the models to capture survivability issues, in order to solve real-world problems. In particular, specific decomposition methods, such as Kelly's or Analytic Center Cutting Plane, were used to reduce the size of the linear problems. This is considered pioneering work for deriving lower bounds on survivability optimization problems.

#### **SDP** relaxations

We have shown that semi-definite programming relaxations (SDP) provide tight lower bounds for several combinatorial optimization problems. We studied several SDP relaxations for several combinatorial problems, among which telecommunication clustering problems, frequency assignment problems, and quadratic assignment problems. For clustering problems, we proposed both linear and SDP relaxations for the equipartition problem, and presented numerical results for large real-world instances. In parallel with our research on wired network, we studied different relaxations for the frequency assignment problem. We proposed several semi-definite programming relaxations using advanced cutting-plane techniques from combinatorial optimization, and we solved medium-sized, realworld problems using algorithms based on SDP cutting plane.

# Research groups

Research Description: Communication and Network Design team

#### Robust and Stochastic combinatorial optimization

The MCNF problems studied above are deterministic. We extended our work on network design problems by introducing uncertainty. We mostly studied two approaches. First, we studied recent robust optimization techniques, and we tackled the robust shortest path problem. Uncertainty was modelled by the underlying graph, and we solved the problem for some specific families of graphs (we proposed theoretical results, as well as numerical ones). Second, modelling uncertainty by means of stochastic programming, we studied variants of stochastic MCNF problems. Uncertainty was there expressed by demand (or commodity) random variables. We studied different distribution functions, and we modelled the problem by a two-stage mixed integer program that we solved by combining different methods from combinatorial optimization and stochastic programming. Combinatorial stochastic programming is a very recent research area, and little has been done from a computational point of view. Our work is in fact among the very first to deal with network design problems.

#### Resource assignment and multi-terminal flow problems

The Optimization sub-team also considered other types of problems such as task mapping and resource allocation for multi-point communications. For these problems, mathematical programming approaches have been used successfully, and approximation algorithms and heuristics have been proposed. Some polynomial solutions have even been exhibited in specific cases. The sub-team also investigated flow theory. In particular, we have studied multi-terminal flow problems, where the capacity of the edges is subject to variations. A deep understanding of the so-called "Gomory-Hu cut tree method" allowed us to provide new (simple) algorithms.

#### Key references:

- [4] C. Andrade, A. Lisser, G. Plateau, and N. Maculan. Telecommunication network capacity design for uncertain demand. *Computational Optimization and Applications*, 29:129-147, 2004.
- [19] P. Chardaire and A. Lisser. Simplex and interior point specialization algorithms for solving non-oriented multicommodity flow problems. *Operations Research*, 2:260-276, 2002.
- [113] A. Lisser and F. Rendl. Telecommunication clustering using linear and semidefinite programming. *Mathematical Programming*, 95(1):91-101, 2003.
- [118] M. Bendhaou and D. Fayard. *Optimisation Combinatoire*, chapter Assignment Problems. Hermes, to appear.
- [134] P. Berthomé, M. Diallo, and A. Ferreira. Generalized parametric multi-terminal flows problem. In *Graph Theoretical Concepts in Computer Science (WG)*, LNCS 2880, pages 71-80, 2003.

#### Information dissemination

- Participants: Pierre Fraigniaud (leader), Pascal Berthomé, Nicolas Hanusse.
- Ph.D. Students: José Ignacio. Alvarez-Hamelin, Philippe Gauron, David Ilcinkas.

Although of fundamental nature, the problems addressed by the Information Dissemination sub-team followed the evolution of network technologies at the turn of the new century. While during most the 90's, very high computational power was explored in the framework of parallel multi-computer systems, it is now believed that decentralized, loosely coupled systems are much better candidates in terms of storage, computation, and cost. In addition, the evolution of telecommunications in terms of both technology (high speed networks, wireless links, ad hoc networks, etc.) and services (Web, peer-to-peer, glob-

al computing, etc.) introduced new problems, and/or required new solutions adapted to the decentralized nature of both the systems and the applications. As a consequence, the activities of the Information Dissemination sub-team are now mostly concentrated on understanding the fundamental complexity of problems that naturally occur in large decentralized systems of various types, from Internet and the Web, to ad hoc and peer-to-peer networks. The Information Dissemination sub-team mostly focused its activities on the following problems:

- 1/ Routing problems,
- 2/ Peer-to-peer network design and Small Worlds properties,
- 3/ Protocol design for mobile agents, and
- 4/ Group-communications in arbitrary topologies.

In each of these topics, our research is *not* directly driven by the technology. Instead, the goal of the sub-team is to identify the limits of different network technologies (optical, radio, etc.) and systems (Internet, peer-to-peer, mobile agents, etc.), and to propose new applications that may in turn motivate the need for new hardware or software technologies.

#### **Monsieur Jourdain does routing too!**

#### Molière, The Bourgeois Gentleman, Act 2, Scene 4:

- M. Jourdain: Oh, really? So when I say: "Nicole bring me my slippers and fetch my nightcap" is that routing?
- Philosophy Master: Most clearly.

While this may be an exaggeration, routing is definitely at the core of today's approach to Information Dissemination. It has relationships with fundamental notions in theoretical computer science and discrete mathematics as well as numerous implications in various scientific fields, from networking to sociology. We present three examples to illustrate this claim.

Routing obviously plays a central role in data communication. Perhaps more surprisingly, it also has applications to key problems in distributed computing. For instance, we have shown that a network supporting shortest paths interval routing allows broadcast and leader election to be performed in linear message complexity [84]. This link between compact routing and distributed computing answers a question raised by D. Peleg (Weizmann Institute) and is of growing interest in the community.

. . .

Routing is also related to central notions in graph theory, such as minor theory. For instance, minimizing the header size for oblivious routing in unreliable networks is a minor-closed problem. In this context, we have shown that deterministic routing in planar networks requires headers that are at least log(tw) bits long, where tw is the tree width of the network [83]. An important consequence of this result is to close a conjecture on meshes by F. Fich (University of Toronto).

Finally, routing provides a framework for investigations in social sciences. For instance, Kleinberg's model is widely accepted as a good model for the "six degrees of separation" among individuals found by the social psychologist Milgram in the 1960's. However, this model does not capture the impact of the number of routing criteria, e.g., occupation or location, which has been shown by sociologists to have a strong empirical effect. We have recently designed a new model taking the number of criteria into account [146]. Under this model, the performance of greedy routing better fits experimental observations.

# Research groups

Research Description: Communication and Network Design team

#### Routing

Routing encompasses such diverse problems as analyzing message transfers between endpoints in a communication network, interactions between social entities, lookup strategies in fully decentralized peer-to-peer systems and data transfers between facilities in a transportation network (see sidebar). The sub-team focuses mostly on *compact routing*, i.e., the design of routing strategies requiring a limited amount of memory-space at the routers. Other fields of investigations are routing in ad hoc networks (802.11, Bluetooth, etc.), and routing in all-optical networks (cf. the European project DAVID).

#### Peer-to-peer network design and Small Worlds properties

The peer-to-peer (P2P) paradigm is opposed to the client-server paradigm. In P2P systems, all users play the same role and are both clients and servers. Fully decentralized P2P systems require sophisticated lookup strategies based on dynamically maintained overlay networks. The sub-team significantly contributes to this scientific domain, in particular via the design of the so-called D2B network based on the *de Bruijn* graph. Closely related to this field, the sub-team also studies social networks, experimenting with "small world" properties. In particular, the team recently proposed a new model enhancing Kleinberg's model, and capturing the "six degrees of separation" phenomenon in a more accurate way.

#### Protocol design for mobile entities

The team investigates the behaviour (in terms of capabilities and limits) of software agents and physical robots moving in networks. Specifically, the team focuses on problems such as *graph exploration* and *graph searching*, motivated by various problems occurring in network security, e.g., capturing of an intruder, or resource discovery. The contributions of the team are either impossibility results, e.g., the impossibility to explore all networks with a finite group of finite automata, even if the network is planar and has a small degree, or bounds on the size of a group of agents searching for an intruder under specific constraints. The group also investigates the impact of anonymity on the capabilities of software agents, for complex tasks such as rendezvous, election, or labelling.

#### Group-communications in arbitrary topologies

The sub-team carried on its traditional activities on protocol design for group communications, including one-to-all broadcasting and all-to-all broadcasting. However, the subteam followed the recent evolution of this topic, either guided by the advances in the design of approximation algorithms, or by the evolution of the technology. The sub-team mainly focused on (1) the design of approximation algorithms for broadcasting in distanceinvariant models, e.g., one-hop optical networks, (2) the design of multicast protocols taking the self-similar nature of the Internet traffic into account, and (3) the design of group communication protocols for a cluster of clusters of...of clusters of workstations.

## Key references:

- [7] L. Barrière, P. Fraigniaud, L. Narayanan, and J. Opatrny. Robust position-based routing in wireless ad hoc networks with irregular transmission ranges. *Wireless Communications and Mobile Computing*, 3(2):141-153, 2003.
- [49] K. Diks, P. Fraigniaud, E. Kranakis, and A. Pelc. Tree exploration with little memory. *Journal of Algorithms*, 51(1):38-63, 2004.
- [139] P. Fraigniaud. Approximation algorithms for minimum-time broadcast under the vertex-disjoint paths mode. In 9th Annual European Symposium on Algorithms (ESA), LNCS 2161, pages 440-451, 2001.

- [142] P. Fraigniaud and C. Gavoille. Routing in trees. In 28th International Colloquium on Automata, Languages and Programming (ICALP), LNCS 2076, pages 757-772. Springer, 2001.
- [146] P. Fraigniaud, C. Gavoille, and C. Paul. Eclecticism shrinks even small worlds. In 23rd ACM Symp. on Principles of Distributed Computing (PODC), pages 169-178, 2004.

## **Research Perspective**

In the short and mid-term, the GraFComm group will continue its activities in Graph Theory and Optimization, with specific emphasis on the following topics.

The group aims at investigating the theory of *dynamic networks*. Such networks appear in several scientific domains, including computer science (Web, peer-to-peer, etc.), network engineering (ad hoc, satellite, etc.), sociology, biology, etc. They are characterized by the continuous and possibly brutal evolution of their topologies: nodes can join and leave at any time, links can be established or removed dynamically, etc. After having led the CNRS working group (Action Spécifique) DYNAMO, the group now leads a European COST Action proposal on Algorithmic and Structural Aspects of Dynamic Networks.

The appointment of Prof. Abdel Lisser has reinforced the group's activities in optimization and network design. We anticipate future growth, not only in GraFComm but also throughout LRI, since the number of formal and informal inter-group collaborations on this topic is growing in the lab.

Finally, the group will address new problems in Graph Theory. Some group members are currently focusing on aspects of minor theory, e.g., graph-decomposition, for its links to logic and algorithmic. A GraFComm member is currently preparing her habilitation on this topic, and the group strongly supports the development of such new activities in the graph theory team.

Research groups
GraFComm
Research Perspective

## B.1.3 / GraFComm

## Highlights

The group covers a vast area of Computer Science, as witnessed by the variety of its scientific contributions in both journals and conferences: from Graph Theory and Combinatorics (J. of Graph Theory, J.C.T. B, Discrete Maths, WG, etc.) to Parallel and Distributed Computing (PODC, SPAA, DISC, etc.), from Algorithms (SODA, ESA, etc.) to Operational Research (Mathematical Programming, Operations Research, etc.), from fundamental aspects (STACS, ICALP, MFCS, etc.) to networks (CCCN, IFIP Networking, etc.).

The important contributions of the group to the scientific fields mentioned above, both quantitatively (roughly 150 journal and conference publications) and qualitatively (the journals and conferences listed above are among the very best in their respective domains) have earned the group an international reputation, as shown by the large number of invited presentations at conferences and seminars. The number of participations in conference program committees (as members or as chair) is significant, especially when considering that conferences in graph theory and optimisation often have small committees that select submissions based on short (half a page) abstracts.

The two radar views below show that despite a significant improvement in the amount of contract funding, it remains only 50% of what could be expected according to the size of the group. The Graph Theory team had difficulties finding contracts in 2000-2004 because very few institutional support was eligible for structural graph theory. The most important difference between the two plots is the number of Ph.D.s defended. The group suffered greatly from the decay in the number of students in the masters program (DEA) at the end of the 90s. The number of Ph.D. students in the group is however getting back to normal, with 10 students at the end of 2004.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international), and,
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

## B.1.4 / GraFComm

## Honors

## Prizes and awards

• Mekkia Kouider was distinguished by the Journal of Graph Theory for her paper *Stability and* (*a*,*b*)-*factors*, for which she received a 1-year free subscription to the journal.

## **Keynote addresses**

- Odile Favaron, invited keynote speaker, *South African International Graph Theory Conference*, Itala, South Africa, June 18-22, 2001. Title: "Domination related parameters in claw-free graphs".
- Odile Favaron, invited keynote speaker, *18th miniconference on Discrete Mathematics and Operational Research*, Clemson, USA, October 15-18, 2003. Title: "t-colorings and s-complete t-colorings of graphs".
- Pierre Fraigniaud, invited keynote speaker, 9th Colloquium on Structural Information and Communication Complexity (SIROCCO), Andros, Greece, June 10-12, 2002. Title: "Mobile Entities in Networks".
- Pierre Fraigniaud, invited keynote speaker, 5<sup>th</sup> Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel), Banyuls-sur-mer, May 12-14, 2003. Title: "Réseaux dynamiques pour les sytèmes pair-à-pair".
- Hao Li, invited keynote speaker, *CAS International Conference on Graph Theory and Combinatorics, Kunming*, China, June 2001. Title: "On Woodall's conjecture".
- Hao Li, invited keynote speaker, *International Conference on Graph Theory and Combinatorics*, Taipei, Taiwan, June 2002. Title: "Pancyclic and Pancyclability".
- Hao Li, invited keynote speaker, *First Chinese National Conference on Graph Theory and Combinatorics*, Ulmuqi, China, August 2004. Title: "Vertex-distinguishing proper edge colorings of graphs".

Research groups

GraFComm Highlights Honors

## **Steering Committees**

- Pierre Fraigniaud is member of the Steering Committee of the series of *Symposia on Theoretical Aspects of Computer Science* (STACS), since 2003. STACS 2005 will be the 22nd edition.
- Pierre Fraigniaud is member of the Steering Committee of the series of *ACM Symposia on Parallelism in Algorithms and Architectures* (SPAA), since September 2002. SPAA 2005 will be the 17th edition.
- Pierre Fraigniaud is member of the Steering Committee of the series of *Colloquia on Structural Information and Communication Complexity* (SIROCCO), since June 2001. SIROC-CO 2005 will be the 12<sup>th</sup> edition.

## **Other honors**

• Hao Li was acknowledged in the *Math Reviews* journal for his paper "On a conjecture of Wooddall" in JCT B, 2002 [108]. Quoting from the editorial: "This paper is the most beautiful that has appeared in the last 3 years in the Journal of Combinatorial Theory, Series B".

The following GraFComm members were invited to foreign universities (with their local and/or travel expenses fully covered):

- Odile Favaron: Univ. of Victoria, Canada, 2 weeks July 2001; UNISA, South Africa, 2 weeks April 2002; Univ. of Clemson, USA, 10 days Oct 2003.
- Pierre Fraigniaud: Carleton Univ., Ottawa, Canada, 3 months Oct-Dec 2000; Concordia Univ., Montreal, Canada, 6 months Jan-June 2001; Weizmann Inst., Israel, 1 week April 2004; ETH Zurich, Swiss, 2 weeks Feb 2000, and 1 week Feb 2004; UQO, Canada, 2 weeks Sept 2004, 2 weeks Jan 2003, and 2 weeks Jan 2002; Univ. of Bergen, 1 week Nov 2003, and 1 week June 2004.
- Hao Li: Hungarian Academy of Sciences, 6 weeks, 2003; Chinese Academy of Sciences, 1 month in 2003, 2 months in 2004; City University of Hong Kong, 10 days in 2002; Lanzhou University, China, 1 week in 2001, 2003, and 2004; Eastern China Normal University, 2 weeks in 2004.
- David Ilcinkas: Weizmann Inst., Israel, 1 week April 2004; UQO, Canada, 3 weeks Sept 2004, and 2 weeks Jan 2004.
- Mekkia Kouider: AGH, Krakow, Poland, 2 weeks, 2000; Univ. of Aalborg, Denmark, 3 weeks 2002 and 1 week 2004; Univ. W. Laurier, Waterloo, Canada, 2 weeks, 2004; Univ. of Rio, Brazil, 3 weeks, 2002.

# **Evaluation of research**

## **Editorial boards**

- AKCE International Journal of Graphs and Combinatorics: Hao Li and Odile Favaron, members since 2004.
- TOCS, Theory of Computing Systems: Pierre Fraigniaud, member since Jan 1st, 2003.
- JOIN, *Journal of Interconnection Networks*: Pierre Fraigniaud, member since March 2002.
- DM&TCS, *Discrete Mathematics and Theoretical Computer Science*: Pierre Fraigniaud, member until 2001.
- 4OR, *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*: Abdel Lisser, member since 2002.

## **Program committees**

#### Chair:

- DISC, 19th Conference on Distributed Computing, 2005: Pierre Fraigniaud, Program Chair.
- SPAA, 13th ACM Symp. on Parallel Algorithms and Architectures, July 4-6, 2001, Greece: Pierre Fraigniaud, Program Chair.
- SIROCCO, *8th Colloquium on Structural Information and Communication Complexity*, June 27-29, 2001, Vall de Nuria, Spain: Pierre Fraigniaud, Program co-Chair (with F. Comellas, UPC, Barcelona).

#### Member (international events):

- WAE, *4th Int. Workshop on Efficient and Experimental Algorithms*, May 10-13, 2005, Santorini Island, Greece: Pierre Fraigniaud.
- WG, 31st Int. Workshop on Graph-Theoretical Concepts in Computer Science, Metz, France, June 2005: Pierre Fraigniaud.
- Graph Theory 2004, Conference in Memory of Claude Berge, Paris, July 5-9, 2004: Charles Delorme.
- DISC, 18th Int. Symposium on Distributed Computing, Amsterdam, Netherlands, October 4-8, 2004. Pierre Fraigniaud.
- MASS, 1st IEEE Conference on Mobile Ad-hoc and Sensor Systems, October 24-27, 2004, Fort Lauderdale, Florida, USA. Pierre Fraigniaud.
- DIALM-POMC, *1st and 2nd Joint Workshop on Foundations of Mobile Computing*, San Diego, USA, Sept. 19, 2003, and Philadelphia, USA, Oct. 1, 2004. Pierre Fraigniaud (twice).
- DIWANS, 1st Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks, Florence, Italy, June 29, 2004. Pierre Fraigniaud.
- FUN, 3rd Int. Conference on Fun with Algorithms, May 26-28, 2004, Isola d'Elba, Italy. Pierre Fraigniaud.
- WMAN, *4th Int. Workshop on Algorithms for Wireless, Mobile, Ad Hoc and Sensor Networks*, April 30, 2004, Santa Fe, New Mexico, USA. Pierre Fraigniaud.
- EuroPar, 9th European Conference on Parallel and Distributed Computing, August 26-29, 2003, Klagenfurt, Austria. Pierre Fraigniaud.
- STACS, 20th Symposium on Theoretical Aspects of Computer Science, Berlin, Germany, Feb. 27-March 1, 2003. Pierre Fraigniaud.
- WADS, *7th and 8th Workshop on Algorithms and Data Structures*, Providence, USA, August 8-10, 2001, and Ottawa, Canada, July 30-August 1, 2003. Pierre Fraigniaud (twice).
- SIROCCO, *10th Colloquium on Structural Information and Communication Complexity*, June 20-22, 2000, L'Aquila, Italy and June 18-20, 2003, Umea, Sweden. Pierre Fraigniaud (twice).
- IPDPS, 1st and 3rd Int. Parallel and Distributed Processing Symposium, Cancun, Mexico, May 01-04, 2000, and Fort Lauderdale, USA, April 15-19, 2002. Pierre Fraigniaud.

Research groups

GraFComm

Honors Evaluation of research

- SPAA, *12th ACM Symposium on Parallel Algorithms and Architectures*, Bar Harbor, USA, July 09-13, 2000. Pierre Fraigniaud.
- ICPP, 16th Int. Conference on Parallel Processing, Toronto, Canada, August 21-24, 2000. Pierre Fraigniaud.
- ISPAN, 5th Int. Symposium on Parallel Architectures, Algorithms and Networks, Dallas, USA, Dec. 7-9, 2000. Pierre Fraigniaud.

#### Member (national events):

- AlgoTel, 2nd and 6th Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications, May 26-28, 2004, Batz-sur-mer, France: Pierre Fraigniaud.
- RenPar, 14th Rencontres Francophones du Parallélisme, Hammamet, Tunisie, April 9-13, 2002: Pierre Fraigniaud.

## Evaluation committees and invited expertise

#### International:

- National Natural Science Foundation of China (NSFC): Hao Li, invited expert.
- Science and Engineering Research Canada (NSERC): Marie-Claude Heydemann and Pierre Fraigniaud, referees.

#### National:

- National research program on global computing, *ACI GRID, Globalisation des Ressources Informatiques et des Données*: Pierre Fraigniaud, member of the Scientific Committee (2001-2003).
- National research program on global computing, *GRID 5000*: Pierre Fraigniaud, member of the Scientific Committee since April 2004.
- Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes (LIMOS), Clermont-Ferrand: Pierre Fraigniaud, member of its last Evaluation committee.

## Other evaluation activities

- Reviewer for Ph.D. dissertation: Charles Delorme (7), Pierre Fraigniaud (8), Marie-Claude Heydemann (1), Abdel Lisser (4);
- AMS, American Mathematical Society: Odile Favaron, referee.
- SAMS, South African Mathematical Society: Odile Favaron, referee.
- French-Algerian bilateral projects: Mekkia Kouider, referee.
- Eastern China Normal University: Hao Li, consultant professor since July 2003.

# Volunteer professional service

## Management positions in scientific organisations

- National network on Architecture, Networks, Systems and Parallelism (GdR ARP), Transversal Action TAROT *Techniques Algorithmiques, Réseaux et d'Optimisation pour les Télécommunications*: Pierre Fraigniaud, chair until June 2001. Under his supervision was created the AlgoTel series of conferences (*Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications*).
- National network on Algorithms, Languages and Programming (GdR ALP), Action *Algorithmique de Graphes*: Odile Favaron, member of the selecting committee until 2003.
- CNRS Working Group on dynamic networks (*Action Spécifique Dynamo*), in collaboration with LIAFA (Paris), LIP (Lyon), INRIA project Mascotte (Sophia-Antipolis), INSA Lyon, and UTC (Compiègne): Pierre Fraigniaud, chair (http://www.lri.fr/~pierre/dynamo/).

## Working groups

- The "Communication and network design" team is heavily involved in all activities of the Transversal Action *TAROT, Techniques Algorithmiques, Réseaux et d'Optimisation pour les Télécommunications* of the national network *GdR ARP*.
- The "Graph Theory" research team participates in the activities of the Action *Algorithmique de Graphes* of the national network *GdR ALP*.

In addition, GraFComm was involved in two CNRS working groups (*Actions Spécifiques*):

- Algorithmique pour les grands graphes, in collaboration with LaBRI (Bordeaux, leader), LIX (Palaiseau), LIRMM (Montpellier), LIP (Lyon), Leibniz (Grenoble), IRIN (Nantes), and INRIA project Hipercom (Rocquencourt).
- Algorithmes distribués et applications, in collaboration with LIAFA (Paris, leader), IRISA (Rennes), LIP6 (Paris), ID (Grenoble), and EPFL (Lausanne).

## Other professional service

- IWIN, *International bi-annual Workshop on Interconnection Networks*: Pierre Fraigniaud, member of the Steering Committee since 1995.
- AlgoTel, *Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications*: Pierre Fraigniaud, member of the Steering Committee since 1999.
- *Robert Faure Prize*, a French Operations Research award from ROADEF: Abdel Lisser, member of the selecting committee in 2000.
- Every year, the "Graph Theory" team organizes a *Graph theory day* under the supervision of Mekkia Kouider.

Research groups

Evaluation of research Volunteer professional service

# **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Duration (months)	Total €
RNRT	D. Barth	ROCOCO	MENRT	Paris XI	2000-01		34266
RNRT	D. Barth	ROM	MENRT	Paris XI	2000-01		34193
IST	P. Berthomé	DAVID	UE	CEPHYTEN	2000-03		69654
PICS	P. Fraigniaud	Circulant Networks	CNRS DRI	CNRS	2000-02		14329
AS	P. Fraigniaud	Dynamo	CNRS STIC	CNRS	2002-03		7 500
AS	P. Fraigniaud	Grands graphes	CNRS STIC	CNRS	2002-03		7350
ACI	P. Fraigniaud	PairAPair	MENRT	CNRS	2003-06		46154
PICS	Hao Li	Progress in graphs	EGIDE	CNRS	2000-01		4573
Industry	A. Lisser	Frequency Assignment	FT R&D	CEPHYTEN	2002-04		105000
BQR	A. Lisser	Routage Robuste	Paris XI	Paris XI	2004		11706
BQR	E. Flandrin		Paris XI	Paris XI	2001		4573

Note: See the glossary for acronyms.

## Summary of each scientific project and main results

## ROCOCO

*Partners: ILOG, France Telecom R&D (Sophia Antipolis), INRIA (Grenoble), and LRI (Paris XI).* 

The objective of this project was to study the interactions between two optimization techniques: constraint programming and operational research.

http://www.telecom.gouv.fr/rnrt/projets/pres\_d107\_ap99.htm

## ROM

Partners: Alcatel CIT (Marcoussis), France Telecom R&D (Lannion), Institut National des Télécommunications (Evry), and LRI (Paris XI).

The objective of this project was to study the feasibility of a high-speed WDM all-optical router for service differentiation. Main publications: [88, 133]. ROM was part of the kernel for the constitution of the IST European project "DAVID".

http://www.telecom.gouv.fr/rnrt/projets/prom.htm

Research project funded by national network on telecommunications research (*RNRT*), 1999-2001.

Scientific directors for LRI: Dominique Barth and Pascal Berthomé.

Research project funded by national network on telecommunications research (*RNRT*), 1999-2001.

Scientific directors for LRI: Dominique Barth and Pascal Berthomé

#### DAVID

Partners: 15 industrial and academic partners, among which Alcatel (France and Germany), Institut National des Télécommunications (Evry), University of Bologna (Italy), Research Center COM (Denmark).

The objective of the project was to propose a packet-over-WDM network solution, covering the entire spectrum from MAN to WAN, and including traffic engineering capabilities, as well as network management. Main publications: [175].

http://david.com.dtu.dk.

### **Circulant Networks**

Partners: Macquarie University (Sidney), LaBRI (Bordeaux), and LRI (Paris XI).

The objective of this project was the analysis of the algorithmic and structural properties of circulant networks. This academic project led to several publications, among which [84] and [128].

#### Dynamo

Partners: LIAFA (Paris VII), LIP (ENS Lyon), LRI (Paris XI), Mascotte (I3S/INRIA), Ares (INSA-Lyon/INRIA), and Heudiasyc (UTC).

The objective of this action was the algorithmic and struc-

tural analysis of dynamic networks (Web graph, Internet, ad hoc networks, peer-to-peer systems, etc.). This project was the basis for the constitution of a European COST project on this topic, led by LRI, and which should be formally accepted in the fall of 2004. http://www.lri.fr/~pierre/dynamo/

#### **Grands** graphes

Partners: Hipercom (INRIA), LRI (Paris XI), LIX (Polytechnique), LIRMM (Montpellier), LIP (ENS Lyon), Leibniz (IMAG), LaBRI (Bordeaux), IRIN (Nantes).

The objective of this action, led by André Raspaud (LaBRI),

was the design of algorithms for large complex networks such as the Web or Internet. The "PairAPair" project (see below) and the Dynamo project (see above) came out of this group. http://www.labri.fr/Perso/~hanusse/Grandsgraphes/

CNRS Program for international scientific cooperation (*PICS*), 2000-2002. Scientific director for LRI: Pierre Fraigniaud

European project (*IST Program*), 2000-2003.

Scientific director for LRI: Pascal Berthomé.

CNRS working group (*Action Spécifique*), 2002-2003. Scientific director: Pierre Fraigniaud. Research groups GraFComm Contracts and grants

(*Action Spécifique*), 2002-2003. Scientific director for LRI: Pierre Fraigniaud.

CNRS working group

#### PairAPair

Partners: LRI (Paris XI), LaBRI (Bordeaux), Hipercom (INRIA), and Gyroweb (INRIA).

The objective of this on-going project, led by Laurent Viennot (Gyroweb), is to analyse peer-to-peer systems from

different points of view: from the models and the algorithms, to the design of the protocols and their analysis.

http://gyroweb.inria.fr/pairapair/

#### **Progress in graphs**

Partners: LRI (Paris XI), Univ. of Memphis, and Vanderbilt Univ.

The main purpose of this project was to build a data-base about the recent progress in graph theory, in particular as far as cycle problems are concerned. For that purpose, members of

the project visited the main research centers in this field to collect information.

#### **Frequency Assignment**

Partners: France Telecom R&D, Belfort.

The objective of this project was to model the Frequency Assignment problem, to study and to compute tight lower bounds using semidefinite programming relaxations. This industrial project yields several ongoing publications and conference presentations, among which [170, 172], and [171].

#### **Routage robuste**

Partners: LRI (GraFComm and I&A), France Telecom R&D (Belfort).

This project addressed the design of robust routing protocols in telecommunication networks. Its objectives were to study and model "uncertainty" for routing problems. Theoritical

studies as well as numerical experiments were performed using different approaches.

## **Other projects**

Abdel Lisser kept the responsability of the RNRT project *PORTO* until 2001. This project began in 1999 when he was still at France Telecom R&D. Partners: Alcatel CIT (Marcoussis), France Telecom R&D (Issy Les Moulineaux), INRIA (Nice, mascotte team). The objective of the project was to study and design methods and planning strategies for SDH and WDW backbone networks. The main result is an optical network optimization tool. http://www.telecom.gouv.fr/rnrt/projets/pporto.htm

Industrial Project funded by France Telecom, 2002-2003.

Scientific director: Abdel Lisser.

University grant (BQR, Bonus Qualité Recherche), 2004.

Scientific director: Abdel Lisser.



CNRS Program for international scientific cooperation (*PICS*),

Scientific director for

1998-2001.

LRI: Hao Li.

# Collaborations

## **Cooperation agreements**

- Laboratory of Graph Theory, Combinatorics and Networks, Academy of Mathematics and System Sciences, Chinese Academy of Sciences: Hao Li, co-director, in charge of international cooperation, since 2004. The lab. director is Prof. Xiaodong Hu.
- CNRS Program for internation scientific cooperation (PICS), *Database design for graph theoretical results*, between Université Paris-Sud, CNRS (H. Li) and the universities of Memphis (B. Bollabas) and Vandebilt (M. Plummer), USA, 1998-2001.
- POLONIUM Integrated action, *Extremal problems in graphs and networks*, between Université Paris-Sud (E. Flandrin) and Akademia Gorniczo-Hutnicza, Poland (M. Wozniak). Grant: 3.200 euros, 2004.
- BARRANDE Integrated action, *Structural properties of graphs and hamiltonicity*, between Université Paris-Sud (E. Flandrin) and Zapadoceska Universita, Czech Republic (Z. Ryjacek). Grant: 2.990 euros, 2004.
- PICASSO Integrated action, *Algorithms for systems of mobile agents* between Université Paris-Sud (P. Fraigniaud) and UPC Barcelona (E. Barrière). Grant: 2.847 euros/year, 2002-2003.
- ARC-CNRS Collaboration, *Algorithmic and structural properties of Circulant networks* between LRI (P. Fraigniaud), LaBRI (C. Gavoille), and Macquarie University, Sydney (B. Mans). Grant: 36.000 FF/year for the French side, 2000-2002.

## Collaborations leading to joint publications

GraFComm actively collaborates with the following foreign universities:

#### Canada

- Université du Québec en Outaouais, Canada (A. Pelc);
- University of Ottawa, Canada (P. Flocchini);
- Carleton University, Ottawa, Canada (N. Santoro and E. Kranakis);
- Concordia University, Montreal, Canada (L. Narayanan and J. Opatrny);
- University of Victoria, Canada (E. Cockayne, C. Mynhardt);
- University W. Laurier, Waterloo, Canada (C. Hoang).

## USA

- Clemson University, USA (W. Goddard, S. Hedetniemi, R. Laskar);
- Vandebilt University, USA (M. Plummer);
- University of Memphis, USA (D.Schelp);
- Univ. of Massachussets at Amherst, USA (A. Rosenberg).

#### Western Europe

- University of Bergen, Norway (F. Fomin);
- UPC Barcelona, Spain (J. Gómez, D. Thilikos);
- RWTH, Aachen, Germany (D. Rautenbach);
- Discrete Maths Institute, Bonn, Germany (D. Rautenbach);
- University of Kagenfurt, Austria (F. Rendl);
- University of East Anglia, UK (P. Chardaire);
- University of Geneva, Switzerland (J.-P. Vial);
- Instituto Superior Tecnico, Lisbon, Portugal (R. Cordovil);
- University of Aalborg, Denmark (P. Vestergaard);

#### **Central and Eastern Europe**

- Hungarian Academy of Sciences (E. Gyori);
- AGH, Krakow, Poland (M. Wozniak and A. Marczyk);
- University of West Bohemia, Czech Republic (Z. Ryjacek);
- University of Varsaw, Poland (Z. Lonc).

## Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

Contracts and grants Collaborations

## South America

- Universidad Centroccidental Lisandro Alvarado, Venezuela (I. Marquez);
- Universidad Central de Venezuela, Venezuela (O. Ordaz);
- Universidad Simón Bolívar, Venezuela (D. Quiroz and S. Gonzalez);
- Instituto Pedagógico de Caracas, Venezuela (M.-T. Varela);
- University of Rio de Janeiro, Brazil (S. Klein, N. Maculan);
- University of Fortaleza, Brazil (R. Castro de Andrade).

#### China

- Chinese Academy of Science (Q.-Y. Yan and F. Tian);
- Eastern China Normal University (J.-L. Shu);
- Yunan University, China (J.-P. Li).

#### Other

- Natal University and UNISA, South Africa (C. Mynhardt, P. Grobler, M. Henning);
- Macquarie University, Sydney, Australia (B. Mans);
- University of Beyrouth, Lebanon (A. El Sahili);
- Weizmann Institute, Israel (D. Peleg);
- University of Blida, Algeria (M. Blidia, M. Chellali).

## B.1.9 / GraFComm

# Dissemination and technology transfer

## Summer schools, tutorials, invited seminars

### **Invited seminars:**

- Odile Favaron, *Meeting in honor of Lutz Volkmann 60th birthday*, AWTH Aachen, January 9, 2004. Title: "Total and paired domination in graphs and claw-free graphs".
- Odile Favaron, *Journée scientifique en l'honneur de C. Benzaken*, Grenoble, France, Sept. 6, 2002. Title: "Domination totale dans les graphes".
- Odile Favaron, *series of seminars in South Africa*, Stellenbosch, Pretoria, April 2002. Title: "Bounds on the upper irredundance parameter".
- David Forge, *AMS Spring conference*, Louisiana State University, Baton Rouge, USA, 2003. Title: "Some results on OS algebras".
- David Forge, Conference *Combinatorics in Oporto*, Porto, Portugal, 2003. Title: "A walk through OS algebras".
- Pierre Fraigniaud, *Franco-Japanese Workshop on Computer Science*, Tokyo, Dec 2001. Title: "Graph exploration".
- Pierre Fraigniaud, *University of Massachusetts*, Amherst, USA, Feb 2001. Title: "Approximation algorithms for broadcasting".
- Hao Li gave more than 24 invited seminars in various universities from China and USA (Purdue, Memphis, etc.).
- Mekkia Kouider, University of Rio, Brazil, 2002. Title: "On factorization of graphs".
- Mekkia Kouider, University of Aalborg, Denmark, 2004. Title: "Survey on (a,b)-factors of graphs".
- Abdel Lisser, 9es journées d'optimisation, 2000. Title: "Capacitated Fixe charge multicommodity flow problems".

### **Tutorials**:

- Pierre Fraigniaud, tutorial at *31st LIAFA Spring School on Theoretical Computer Science*, May 4-8, 2003, Porquerolles, France. Title: "Dynamic Networks for Peer-to-Peer Systems".
- Pierre Fraigniaud, tutorial at the *LIX Fall School on Optimisation and Telecom*, Sept. 25-26, 2003, France. Title: "Network exploration".

## Other dissemination and technology transfer

• Charles Delorme maintains a table of the current largest known ( $\Delta$ ,D)-graphs.

## B.1.10 / GraFComm

# Training and education (doctoral and post-doctoral)

Defended doctorates		
Name	Date defended	Current position
VALENCIA PABON Mario	01/12/00	Assistant Professor at
		Univ. de los Andes, Bogotá, Colombia
ALVAREZ-HAMELIN José Ignacio	20/12/02	Post-doc Institut de physique théorique (PXI)
BEN DHAOU Moaiz	18/12/03	Manager Thalensys S.A.

#### **Graduate courses**

- DEA Informatique Distribuée: *Dissémination de l'Information*, Pascal Berthomé (1999-2001), Pierre Fraigniaud (1999-present).
- DEA Informatique Distribuée: *Optimisation de Réseaux de Télécommunications à Très Haut Débit,* Pascal Berthomé (2002-present), Abdel Lisser (2002-present).
- DEA Paris VI: *Optimisation des Réseaux et Applications aux Télécommunications,* Abdel Lisser (2000-present).

GraFComm Dissemination and technology transfer Training and Education

## Publications

## International peer-reviewed journals

- [1] A. Ainouche and M. Kouider. Cycles in partially squared graphs. *Graphs and Combinatorics*, 17(1-9), 2001.
- [2] J. I. Alvarez-Hamelin and N. Schabanel. An internet graph model based on trade-off optimization. *European Physical Journal B, special issue on "Applications of networks"*, 38(2):231-237, 2004.
- [3] D. Amar, M. El Kadi Abderrezzak, and E. Flandrin. Cyclability and pancyclability in bipartite graphs. *Discrete Mathematics*, 236:3-11, 2001.
- [4] C. Andrade, A. Lisser, G. Plateau, and N. Maculan. Telecommunication network capacity design for uncertain demand. *Computational Optimization and Applications*, 29:129-147, 2004.
- [5] G. Bacsó and O. Favaron. Independence, irredundance, degrees and chromatic number in graphs. *Discrete Mathematics*, 259:257-262, 2002.
- [6] P. Balister, A. Kostochka, H. Li, and R. Schelp. Balanced edge colorings. *Journal of Combinatorial Theory B*, 90:3-20, 2004.
- [7] L. Barrière, P. Fraigniaud, L. Narayanan, and J. Opatrny. Robust position-based routing in wireless ad hoc networks with irregular transmission ranges. *Wireless Communications and Mobile Computing*, 3(2):141-153, 2003.
- [8] D. Barth and P. Berthomé. Periodic gossiping in commuted networks. *Theory of Computing Systems*, to appear.
- [9] D. Barth, P. Berthomé, and J. Cohen. The eulerian stretch of a network topology and the ending guarantee of a convergence routing. *Journal of Interconnection Networks*, 5(2):93-109, 2004.
- [10] D. Barth, P. Fragopoulou, and M.-C. Heydemann. Uniform emulations of cartesian-product and Cayley graphs. *Discrete Applied Mathematics*, 116:37-54, 2002.
- [11] C. Bazgan, A. Harkat-Benhamdine, H. Li, and M. Woźniak. A note on the vertex-distinguishing proper colorings of graphs with large minimum degree. *Discrete Mathematics*, 236:37-42, 2001.
- [12] C. Bazgan, A. Harkat-Benhamdine, H. Li, and M. Woźniak. Partitionning vertices of 1tough graph into paths. *Theoritical Computer Sciences*, 263:255-261, 2001.
- [13] C. Bazgan, H. Li, and M. Woźniak. On the Loebl-Komlós-Sós conjecture. *Journal of Graph Theory*, 34:269-276, 2000.
- [14] P. Berthomé, A. Ferreira, B. Maggs, S. Perennes, and C. Plaxton. Sorting-based selection algorithms for hypercubic networks. *Algorithmica*, 26:237-254, 2000.
- [15] P. Berthomé, T. Hagerup, I. Newman, and A. Schuster. Self-simulation for the passive optical star. *Journal of Algorithms*, 34:128-147, 2000.
- [16] S. Brandt, O. Favaron, and Z. Ryjáček. Closure and stable hamiltonian properties in clawfree graphs. *Journal of Graph Theory*, 34(1):30-41, 2000.
- [17] R. Cada, E. Flandrin, H. Li, and Z. Ryjáček. Cycles through given vertices and closures. *Discrete Mathematics*, 276:65-80, 2004.
- [18] M. Cai, E. Flandrin, and H. Li. On quasifactorability in graphs. *Discrete Mathematics*, 233:85-92, 2001.

- [19] P. Chardaire and A. Lisser. Simplex and interior point specialization algorithms for solving non-oriented multicommodity flow problems. *Operations Research*, 2:260-276, 2002.
- [20] G. Chen, R. Gould, and H. Li. Partitionning vertices of a tournament into independent cycles. *Journal of Combinatorial Theory B*, 83:213-220, 2001.
- [21] T. Chich, J. Cohen, and P. Fraigniaud. Unslotted deflection routing: a practical and efficient protocole for multi-hop optical networks. *IEEE/ACM Transaction on Networking*, 9(1):47-59, 2001.
- [22] E. J. Cockayne, O. Favaron, W. Goddard, P. J. Grobler, and C. M. Mynhardt. Changing upper irredundance by edge addition. *Discrete Mathematics*, 266:185-193, 2003.
- [23] E. J. Cockayne, O. Favaron, P. J. Grobler, C. M. Mynhardt, and J. Puech. Generalised Ramsey numbers with respect to classes of graphs. *Ars Combinatoria*, 59:279-288, 2001.
- [24] E. J. Cockayne, O. Favaron, P. J. Grobler, C. M. Mynhardt, and J. Puech. Ramsey properties of generalised irredundant sets in graphs. *Discrete Mathematics*, 231:123-134, 2001.
- [25] E. J. Cockayne, O. Favaron, and C. M. Mynhardt. Irredundance-edge-removal-critical graphs. *Utilitas Math.*, 60:219-228, 2001.
- [26] E. J. Cockayne, O. Favaron, and C. M. Mynhardt. Total domination in claw-free cubic graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 43:219-225, 2002.
- [27] E. J. Cockayne, O. Favaron, and C. M. Mynhardt. Secure domination, weak roman domination and forbidden subgraphs. *Bull. Inst. Combin. Appl.*, 39:87-100, 2003.
- [28] E. J. Cockayne, O. Favaron, and C. M. Mynhardt. On (i-)-edge-removal-critical graphs. Discrete Mathematics, 276:111-125, 2004.
- [29] E. J. Cockayne, O. Favaron, and C. M. Mynhardt. Total domination in K<sub>r</sub>-covered graphs. *Ars Combinatoria*, 71:289-303, 2004.
- [**30**] E. J. Cockayne, O. Favaron, C. M. Mynhardt, and J. Puech. A characterisation of  $(\gamma$ -i)-trees. *Journal of Graph Theory*, 34(4):277-292, 2000.
- [31] J. Cohen, P. Fraigniaud, and C. Gavoille. Recognizing Knödel and Fibonacci graphs. *Discrete Mathematics*, 250:41-62, 2002.
- [32] J. Cohen, P. Fraigniaud, and M. Mitjana. Polynomial time algorithms for minimumtime broadcast in trees. *Theory of Computing Systems*, 35(6):641-665, 2002.
- [33] R. Cordovil and D. Forge. Diagonal bases in Orlik-Solomon type algebras. *Annals of Combinatorics*, 7:247-257, 2003.
- [34] R. Cordovil and D. Forge. A note on tutte polynomials and Orlik-Solomon algebras. *European Journal of Combinatorics*, 24:1081-1087, 2003.
- [35] R. Cordovil and D. Forge. Quadratic Orlik-Solomon algebras of graphic matroids. *Matematica Contemporanea*, 25:25-32, 2003.
- [36] R. Cordovil and D. Forge. Gröbner and diagonal bases in Orlik-Solomon type algebras. *Cubo*, to appear.
- [37] R. Cordovil, D. Forge, and S. Klein. How is a chordal graph like a supersolvable binary matroid? *Discrete Mathematics*, to appear.
- [38] N. Dean and M. Kouider. Gallai's conjecture in disconnected graphs. *Discrete Mathematics*, 213:43-54, 2000.
- [**39**] C. Delorme. Presentations of groups generated by transpositions. *Discrete Mathematics*, 236:59-64, 2001.
- [40] C. Delorme. Spectra and cuts. Australasian Journal of Combinatorics, 26:183-191, 2002.
- [41] C. Delorme. Laplacian eigenvalues and fixed size multisections. *Discrete Mathematics*, 276:149-159, 2004.
- [42] C. Delorme, O. Favaron, and D. Rautenbach. On the Randic index. *Discrete Mathematics*, 257(1):29-38, 2002.

Research groups
GraFComm
Publications

- [43] C. Delorme, O. Favaron, and D. Rautenbach. On the reconstruction of the degree sequence. *Discrete Mathematics*, 259:293-300, 2002.
- [44] C. Delorme, O. Favaron, and D. Rautenbach. Closed formulas for the numbers of small independent sets and matchings and an extremal problem for trees. *Discrete Applied Mathematics*, 130:503-512, 2003.
- [45] C. Delorme and J. Gómez. Some new large compound graphs. *European Journal of Combinatorics*, 23(5):539-547, 2002.
- [46] C. Delorme, S. Gonzales, O. Ordaz, and M. T. Varela. Barycentric sequences and barycentric Ramsey numbers-stars. *Discrete Mathematics*, 277:45-56, 2004.
- [47] C. Delorme, I. Marquez, O. Ordaz, and A. Ortuño. Existence conditions for barycentric sequences. *Discrete Mathematics*, 281:163-172, 2004.
- [48] C. Delorme, O. Ordaz, and D. Quiroz. Some remarks on Davenport constant. *Discrete Mathematics*, 237:119-128, 2001.
- [49] K. Diks, P. Fraigniaud, E. Kranakis, and A. Pelc. Tree exploration with little memory. *Journal of Algorithms*, 51(1):38-63, 2004.
- [50] S. Djelloul and M. Kouider. On weighted mean distance. *Australasian Journal of Comb.*, 23:181-195, 2001.
- [51] S. Djelloul and M. Kouider. Minimum survivable graphs with bounded distance increase. *Discrete Mathematics and Theoretical Computer Science*, 6(1):123-132, 2003.
- [52] Y. Egawa, H. Enomoto, R. Faudree, H. Li, and I. Schiermeyer. Two-factors each component of which contains a specified vertex. *Journal of Graph Theory*, 43(3):188-198, 2003.
- [53] M. El Kadi Abderrezzak and E. Flandrin. New sufficient conditions for bipancyclic bipartite graphs. *Discrete Mathematics*, to appear.
- [54] H. Enomoto and H. Li. Partition of a graph into cycles and degenerated cycles. *Discrete Mathematics*, 276:177-181, 2004.
- [55] R. Faudree, E. Flandrin, M. Jacobson, J. Lehel, and R. Schelp. Even cycles in graphs with many odd cycles. *Graphs and Combinatorics*, 16:399-410, 2000.
- [56] O. Favaron. Extendability and factor-criticality. Discrete Mathematics, 213:115-122, 2000.
- [57] O. Favaron. Inflated graphs with equal independence number and upper irredundance number. *Discrete Mathematics*, 236:81-94, 2001.
- [58] O. Favaron. Independence and upper irredundance in claw-free graphs. *Discrete Applied Mathematics*, 132:85-95, 2003.
- [59] O. Favaron, E. Flandrin, H. Li, and Z. Ryjáček. Clique covering and degree conditions for hamiltonicity in claw-free graphs. *Discrete Mathematics*, 236:65-80, 2001.
- [60] O. Favaron and P. Fraisse. Hamiltonicity and minimum degree in 3-connected claw-free graphs. *Journal of Combinatorial Theory Series B*, 82(2):297-305, 2001.
- [61] O. Favaron, G. Fricke, W. Goddard, S. M. Hedetniemi, S. T. Hedetniemi, P. Kristiansen, R. C. Laskar, and D. Skaggs. Offensive alliances in graphs. *Discussiones Mathematicae-Graph Theory*, 24:263-275, 2004.
- [62] O. Favaron, G. H. Fricke, D. Pritikin, and J. Puech. Irredundance and domination in kings graphs. *Discrete Mathematics*, 262:131-147, 2003.
- [63] O. Favaron, T. W. Haynes, and S. T. Hedetniemi. Domination subdivision numbers in graphs. *Utilitas Math.*, to appear.
- [64] O. Favaron, T. W. Haynes, S. T. Hedetniemi, M. A. Henning, and D. J. Knisley. Total irredundance in graphs. *Discrete Mathematics*, 256(1-2):115-127, 2002.
- [65] O. Favaron, T. W. Haynes, and P. J. Slater. Distance-k independent domination sequences. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 33:225-237, 2000.
- [66] O. Favaron, S. T. Hedetniemi, S. M. Hedetniemi, and D. F. Rall. On k-dependent domination. *Discrete Mathematics*, 249:83-94, 2002.

- [67] O. Favaron and M. A. Henning. Upper total domination in claw-free graphs. *Journal of Graph Theory*, 44(2):148-158, 2003.
- [68] O. Favaron and M. A. Henning. Paired domination in claw-free cubic graphs. *Graphs and Combinatorics*, to appear.
- [69] O. Favaron, M. A. Henning, C. M. Mynhardt, and J. Puech. Total domination in graphs with minimum degree three. *Journal of Graph Theory*, 34(1):9-19, 2000.
- [70] O. Favaron, M. A. Henning, J. Puech, and D. Rautenbach. On domination and annihilation in graphs with claw-free blocks. *Discrete Mathematics*, 231:143-151, 2001.
- [71] O. Favaron, M. Mahéo, and J.-F. Saclé. The Randic index and other Graffiti parameters of graphs. *MATCH - Commun. Math. Comput. Chem.*, 47:7-23, 2003.
- [72] O. Favaron and Y. Redouane. Neighborhood unions and regularity in graphs. *Theoretical Computer Science*, 263:247-254, 2001.
- [73] E. Flandrin, H. Li, A. Marczik, and M. Wozniak. A note on a new condition implying pancyclism. *Discussiones Mathematicae Graph Theory*, 21:137-144, 2001.
- [74] E. Flandrin, H. Li, A. Marczik, and M. Wozniak. A Chvátal-Erdös type condition for pancyclability. *Discrete Mathematics*, to appear.
- [75] E. Flandrin, H. Li, and J. Shu. A sufficient condition for cyclability in directed graphs. *Discrete Mathematics*, to appear.
- [76] D. Forge. Bases in Orlik-Solomon type algebras. *European Journal of Combinatorics*, 23:567-572, 2002.
- [77] D. Forge and M. LasVergnas. Orlik-Solomon type algebras. *European Journal of Combinatorics*, 22:705-708, 2001.
- [78] D. Forge, M. LasVergnas, and P. Schuchert. A set of 10 points not projectively equivalent to the vertices of a polytope. *European Journal of Combinatorics*, 22:705-708, 2001.
- [79] D. Forge and J. Ramirez-Alfonsin. On counting the k-face cells of cyclic arrangements. *European Journal of Combinatorics*, 22:307-312, 2001.
- [80] D. Forge and J. Ramirez-Alfonsin. On reconstructing arrangements from their sets of simplices. *Discrete Mathematics*, 226:175-190, 2001.
- [81] D. Forge, J. Ramirez-Alfonsin, and H. Yeun. Disconnected coverings for oriented matroids via simultaneous mutations. *Discrete Mathematics*, 258:353-359, 2002.
- [82] P. Fraigniaud. A note on line broadcast in digraphs under the edge-disjoint paths mode. *Discrete Applied Mathematics,* to appear.
- [83] P. Fraigniaud and C. Gavoille. Header-size lower bounds for end-to-end communication in memoryless networks. *Computer Network*, to appear.
- [84] P. Fraigniaud, C. Gavoille, and B. Mans. Interval routing schemes allow broadcasting with linear message-complexity. *Distributed Computing*, 14(4):217-229, 2001.
- [85] P. Fraigniaud, J.-C. König, and E. Lazard. Oriented hypercubes. *Networks*, 39(2):98-106, 2002.
- [86] P. Fraigniaud, A. Pelc, S. Perennes, and D. Peleg. Assigning labels in an unknown anonymous network with a leader. *Distributed Computing*, 14:163-183, 2001.
- [87] P. Fraigniaud and J. Peters. Minimum linear gossip graphs and maximal linear (Δ,k)-gossip graphs. *Networks*, 38(3):150-163, 2001.
- [88] P. Gravey, S. Gosselin, C. Guillemot, D. Chiaroni, N. Le Sauze, A. Jourdan, E. Dotaro, D. Barth, P. Berthomé, C. Laforest, S. Vial, T. Atmaca, G. Hébuterne, H. ElBiaze, R. Laalaoua, E. Gangloff, and I. Kotuliak. Multiservice optical network: main concepts and first achievements of the ROM program. *Journal of Lightwave Technology*, 19:23-31, 2001.
- [89] R. Häggkvist and M. Kouider. A lower bound on the circumference of 3-connected k-regular 4k-order graphs. *Graphs and combinatorics*, to appear.

Research groups GraFComm Publications

- [90] G. Hahn, J. Kratochvil, D. Sotteau, and J. Širáň. On injective colourings. *Discrete Mathematics*, 256:179-192, 2002.
- [91] R. Harbane and M.-C. Heydemann. Efficient reconfiguration algorithms of de Bruijn and Kautz networks into linear arrays. *Theoretical Computer Science*, 263:173-189, 2001.
- [92] A. Harkat-Benhamdine, H. Li, and F. Tian. Cyclability of 3-connected graphs. *Journal of Graph Theory*, 34:191-203, 2000.
- [93] M.-C. Heydemann, N. Marlin, and S. Perennes. Complete rotations in cayley graphs. *Europ. J. Combinatorics*, 22:179-196, 2001.
- [94] H. Kheddouci and M. Kouider. Hamilton cycle decomposition of kronecker products of some cubic graphs by cycles. *Journal of Comb. Theory and Comb. Computing*, 32:3-22, 2000.
- [95] M. Kouider. Covering vertices by cycles and neighborhoods. *Combinatorica*, 20(2):219-226, 2000.
- [96] M. Kouider and Z. Lonc. Covering cycles and k-term degree sums. *Combinatorica*, 16(3):407-412, 2000.
- [97] M. Kouider and Z. Lonc. Stability number and (a,b) factors in graphs. *Journal of Graph Theory*, 46(4):254-264, 2004.
- [98] M. Kouider and M. Mahéo. 2-edge-connected [2,k]-factors in graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 35:89-95, 2000.
- [99] M. Kouider and M. Mahéo. Connected [a,b]-factors in graphs. *Combinatorica*, 22:71-82, 2002.
- [100] M. Kouider and M. Mahéo. Some bounds for the b-chromatic number of a graph. *Discrete Mathematics*, pages 267-277, 2002.
- [101] M. Kouider and M. Mahéo. The b-chromatic number of products of graphs. *Combinatorica,* to appear.
- [102] M. Kouider and A. Marczyk. On pancyclism of hamiltonian graphs. *Discrete Mathematicss*, 251:119-127, 2002.
- [103] M. Kouider and P. Vestergaard. On even [2,b] factors in graphs. Australasian Journal of Combinatorics, pages 139-147, 2003.
- [104] M. Kouider and P. Vestergaard. Connected factors in graphs -a survey. *Graphs and combinatorics,* to appear.
- [105] M. Kouider and P. Vestergaard. On even [a,b] factors in graphs. *Discussiones Mathematica*, to appear.
- [106] H. Li. On cycles in 3-connected graphs. Graphs and Combinatorics, 16(3):319-335, 2000.
- [107] H. Li. A generalization of the Gallai-Roy theorem. *Graphs and Combinatorics*, 17(4):681-685, 2001.
- [108] H. Li. On a conjecture of Woodall. Journal of Combinatorial Theory B, 86:172-185, 2002.
- [109] H. Li and J. Li. Independent triangles covering given subset in a graph. *Theoretical Computer Science*, 263:333-344, 2001.
- [110] H. Li and J. Shu. Cyclic partition of strong tournaments. OR transactions, 8(1):53-61, 2004.
- [111] H. Li and J. Shu. Partitioning a strong tournament into k cycles. *Ars Combinatoria*, 72:203-212, 2004.
- [112] A. Lisser, N. Maculan, and G. Plateau. Integer linear models with a polynomial number of variables and constraints for some classical combinatorial problems. *Pesquisa Operacional*, 23(1):161-168, 2003.
- [113] A. Lisser and F. Rendl. Telecommunication clustering using linear and semidefinite programming. *Mathematical Programming*, 95(1):91-101, 2003.
- [114] A. Lisser, R. Sarkissian, and J. Vial. Lower bounds for survivability in telecommunica-

tion networks. Investigation Operativa, 9(1-2):21-47, 2000.

- [115] L. Narayanan, J. Opatrny, and D. Sotteau. All-to-all optical routing in optimal chordal rings of degree 4. *Algorithmica*, 31(2):155-178, 2001.
- [116] J. Opatrny and D. Sotteau. Embeddings of complete binary trees into extended grids with total congestion 1. *Discrete Applied Mathematics*, 98:237-254, 2000.

## National peer-reviewed journals

[117] O. Favaron. From irredundance to annihilation: a brief overview of some domination parameters of graphs. *Saber (Venezuela)*, 32(2):64-69, 2000.

## **Book chapters**

- [118] M. Ben Dhaou and D. Fayard. *Optimisation Combinatoire*, chapter Assignment Problems. Hermes, to appear.
- [119] P. Chardaire and A. Lisser. *Encyclopedia of Optimization*, volume 4, chapter Non-oriented multicommodity flow problems, pages 92-97. Kluwer Academic Publisher, 2001.
- [120] P. Chardaire and A. Lisser. *Handbook on Applied Optimization,* chapter Minimum cost multicommodity flow, pages 404-422. Oxford University Press, 2002.
- [121] A. Lisser. *[title to be announced]*, chapter Multicommodity flow in telecommunication networks. Kluwer Academic Publisher, to appear.
- [122] N. Maculan, M. Passini, J. Brito, and A.Lisser. *Transportation and Network Analysis: Current Trends*, chapter Mathematical model for designing telecommunication networks, pages 165-179. Kluwer Academic Publisher, 2002.

## Major international peer-reviewed conferences

- [123] I. Alvarez-Hamelin and P. Fraigniaud. MLT: A multicast protocol with Qos support. In *12th IEEE Int. Conference on Computer Communications and Networks (ICCCN)*, pages 264-269, 2003.
- [124] I. Alvarez-Hamelin and P. Fraigniaud. Reducing packet-loss by taking long-range dependences into account. In *3rd IFIP Networking Conference*, volume LNCS 3042, pages 1096-1107. Springer-Verlag, 2004.
- [125] L. Barrière, P. Flocchini, P. Fraigniaud, and N. Santoro. Capture of an intruder by mobile agents. In *14th ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pages 200-209, 2002.
- [126] L. Barrière, P. Flocchini, P. Fraigniaud, and N. Santoro. Can we elect if we cannot compare? In 15th ACM Symposium on Parallel Algorithms and Architectures (SPAA), pages 324-332, 2003.
- [127] L. Barrière, P. Flocchini, P. Fraigniaud, and N. Santoro. Election and rendezvous in fully anonymous systems with sense of direction. In *10th Colloquium on Structural Information and Communication Complexity (SIROCCO)*, pages 17-32. Carleton Scientific, 2003.
- [128] L. Barrière, P. Fraigniaud, C. Gavoille, B. Mans, and M. Robson. Recognizing abelian

Research groups

GraFComm Publications cayley colored digraphs. In 8th Annual European Symposium on Algorithms (ESA), LNCS 1879, pages 76-87. Springer, 2000.

- [129] L. Barrière, P. Fraigniaud, E. Kranakis, and D. Krizanc. Efficient routing in networks with long range contacts. In *15th International Symposium on Distributed Computing (DISC)*, LNCS 2180, pages 270-284. Springer, 2001.
- [130] L. Barrière, P. Fraigniaud, L. Narayanan, and J. Opatrny. Robust routing in wireless networks with unstable transmission ranges. In *5th ACM International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM)*, pages 19-27, 2001.
- [131] L. Barrière, P. Fraigniaud, L. Narayanan, and J. Opatrny. Dynamic construction of bluetooth scatternets of fixed degree and low diameter. In 14th ACM-SIAM Symp. on Discrete Algorithms (SODA), pages 781-790, 2003.
- [132] L. Barrière, P. Fraigniaud, N. Santoro, and D. Thilikos. Searching is not jumping. In 29th Workshop on Graph Theoretic Concepts in Computer Science (WG), LNCS 2880, pages 34-45. Springer, 2003.
- [133] D. Barth, P. Berthomé, T. Czarchoski, J.-M. Fourneau, C. Laforest, and S. Vial. A mixed deflection and convergence routing algorithm: Design and performance. In *EuroPar*, LNCS 2400, pages 767-774, 2002.
- [134] P. Berthomé, M. Diallo, and A. Ferreira. Generalized parametric multi-terminal flows problem. In *Graph Theoretical Concepts in Computer Science (WG)*, LNCS 2880, pages 71-80, 2003.
- [135] F. Cappello, P. Fraigniaud, B. Mans, and A. Rosenberg. Hihcohp-toward a realistic communication model for hierarchical hyperclusters of heterogeneous processors. In *Int. Parallel and Distributed Processing Symp. (IPDPS)*, 2001.
- [136] J. Cohen, P. Fraigniaud, and M. Mitjana. Minimal contention-free matrices with application to multicasting. In *DIMACS Workshop on Robust Communication Networks*, volume 53 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 17-33, 2000.
- [137] A. Dessmark, P. Fraigniaud, and A. Pelc. Deterministic rendezvous in graphs. In 11<sup>th</sup> Annual European Symposium on Algorithms (ESA), LNCS 2832, pages 184-195, 2003.
- [138] K. Diks, P. Fraigniaud, E. Kranakis, and A. Pelc. Tree exploration with little memory. In *13th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 588-597, 2002.
- [139] P. Fraigniaud. Approximation algorithms for minimum-time broadcast under the vertex-disjoint paths mode. In 9th Annual European Symposium on Algorithms (ESA), LNCS 2161, pages 440-451, 2001.
- [140] P. Fraigniaud. Minimum-time broadcast under edge-disjoint paths modes. In 2nd International Conference on Fun with Algorithms (FUN), pages 133-148. Carleton Scientific, 2001.
- [141] P. Fraigniaud, L. Gasieniec, D. Kowalski, and A. Pelc. Collective tree exploration. In 6<sup>th</sup> Latin American Theoretical Informatics Symposium (LATIN), LNCS 2976, pages 141-141. Springer, 2004.
- [142] P. Fraigniaud and C. Gavoille. Routing in trees. In 28th International Colloquium on Automata, Languages and Programming (ICALP), LNCS 2076, pages 757-772. Springer, 2001.
- [143] P. Fraigniaud and C. Gavoille. A space lower bound for routing in trees. In 19th Int. Symposium on Theoretical Aspects of Computer Science (STACS), LNCS 2285, pages 65-75. Springer, 2002.
- [144] P. Fraigniaud and C. Gavoille. Lowers bounds for oblivious single-message end-to-end communication. In 17th Symposium on Distributed Computing (DISC), LNCS 2848, pages 211-223, 2003.

- [145] P. Fraigniaud, C. Gavoille, and B. Mans. Interval routing schemes allow broadcasting with linear message-complexity. In 19th ACM Symp. on Principles of Distributed Computing (PODC), pages 11-20, 2000.
- [146] P. Fraigniaud, C. Gavoille, and C. Paul. Eclecticism shrinks even small worlds. In 23<sup>rd</sup> ACM Symp. on Principles of Distributed Computing (PODC), pages 169-178, 2004.
- [147] P. Fraigniaud and D. Ilcinkas. Digraphs exploration with little memory. In 21st Symposium on Theoretical Aspects of Computer Science (STACS), LNCS 2996, pages 246-257. Springer, 2004.
- [148] P. Fraigniaud, D. Ilcinkas, G. Peer, A. Pelc, and D. Peleg. Graph exploration by a finite automaton. In 29th Symposium on Mathematical Foundations of Computer Science (MFCS), LNCS 3151, pages 451-462. Springer, 2004.
- [149] P. Fraigniaud, A. Pelc, S. Perennes, and D. Peleg. Assigning labels in unknown networks. In 19th ACM Symp. on Principles of Distributed Computing (PODC), pages 101-112, 2000.
- [150] N. Hanusse, D. Kavvadias, E. Kranakis, and D. Krizanc. Memoryless search algorithm in networks with faulty advice. In *IFIP World Computer Congress (WCC), Track: International Conference on Theoretical Computer Science*, pages 206-216. Kluwer Academic, 2002.

## National peer-reviewed conferences

- [151] I. Alvarez-Hamelin and P. Fraigniaud. A short survey of multicast protocols. In 2<sup>es</sup> Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel), pages 161-166. INRIA, 2000.
- [152] I. Alvarez-Hamelin and P. Fraigniaud. Multicast tree with minimum congestion. In *10th Conference on Electronics, Combinatorics, and Computers (CONIELECOMP)*, pages 247-250. IEEE, 2001.
- [153] I. Alvarez-Hamelin, P. Fraigniaud, A. Dams, and S. Tressens. Multicast tree with minimum congestion. In *2èmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel)*. INRIA, 2001.
- [154] J. I. Alvarez-Hamelin and N. Schabanel. Optimisation et modélisation du graphe des systèmes autonomes d'Internet. In *Sèmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications(AlgoTel)*, pages 75-82. INRIA, 2003.
- [155] L. Barrière, P. Flocchini, P. Fraigniaud, and N. Santoro. Capture d'un agent hostile dans un réseau. In *4èmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel)*, pages 171-176. INRIA, 2002.
- [156] M. Ben Dhaou, D. Fayard, and V. Zissimopoulos. Placement de tâches communicantes sur des processeurs hétérogènes avec minimisation du temps de complétude. In 7<sup>es</sup> journées Nationales sur la Résolution Pratique de Problèmes NP-complets (JNPC), pages 37-48, 2001.
- [157] P. Fraigniaud and C. Gavoille. Comment router dans un arbre ? In *3èmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel)*. INRIA, 2001.
- [158] P. Fraigniaud, C. Gavoille, and C. Paul. Eclectisme dans les petits mondes. In 6èmes Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel). INRIA, 2004.
- [159] P. Fraigniaud, D. Ilcinkas, G. Peer, A. Pelc, and D. Peleg. Exploration de réseaux par un robot. In 6<sup>es</sup> Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel). INRIA, 2004.
- [160] N. Hanusse, D. Kavvadias, E. Kranakis, and D. Krizanc. Algorithme sans mémoire pour la recherche d'information dans un réseau avec menteurs. In *3èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel)*. INRIA, 2001.

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups
GraFComm
Publications

## Other conferences and workshops

- [161] R. Andrade, A. Lisser, G. Plateau, and N. Maculan. Simulation on the integer capacity planning under uncertain demand problem in telecommunication networks. In *Annals of EUROSIM*, 2001.
- [162] D. Barth, P. Berthomé, A. Borrero, J.-M. Fourneau, C. Laforest, F. Quessette, and S. Vial. Performance comparisons of Eulerian routing and deflexion routing in a 2D-mesh all optical network. In *15th European Simulation Multiconference*, pages 887-891, 2001.
- [163] D. Barth, P. Berthomé, and B. Cohen-Boulakia. Flots avec des fonctions de coût en escalier. In ROADEF, pages 415-416, 2003.
- [164] D. Barth, P. Berthomé, and P. Fragopoulou. About algorithms to allocate subtrees to many multipoint communication requests in a network. In *EURO*, page 209, 2001.
- [165] D. Barth, P. Berthomé, and T. Mautor. Etudes de bornes inférieures pour un problème de dimensionnement de réseau. In *ROADEF*, pages 214-215, 2002.
- [166] M. Ben Dhaou and P. Berthomé. Some results about the maximal request satisfaction problem. In *EURO 2001*, page 209, Rotterdam, 2001.
- [167] M. Ben Dhaou and P. Berthomé. Autour du problème d'affectation de requêtes multipoints. In ROADEF 2002, page 50, 2002.
- [168] M. Ben Dhaou and D. Fayard. Une heuristique pour un problème de placement de tâches. In *Rencontres francophones de recherche opérationnelle (FRANCORO)*, 2001.
- [169] W. Benajam and M. Ben Dhaou. Relaxations semidéfinies pour des problèmes de placement de tâches. In *ROADEF*, pages 205-206, 2003.
- [170] W. Benajam, A. Caminada, and A. Lisser. Relaxations semidéfinies pour le problème d'allocation de fréquence. In *ROADEF*, pages 252-253, 2003.
- [171] W. Benajam, A. Gaivoronski, and A. Lisser. Stochastic frequency assignment problem. In *10th Int. Conference on Stochastic Programming*, to appear.
- [172] W. Benajam, A. Lisser, and M. Minoux. SDP relaxation for the quadratic assignment problem. In EURO XX conference on Operational Research, page 114, 2004.
- [173] P. Berthomé, J. Cohen, and T. Mautor. Optimisation des ressources utilisées pour une diffusion. In ROADEF 2003, pages 51-52, 2003.
- [174] P. Berthomé, L. Gastal, and A. Lisser. Robust shortest paths. In *Optimization*, page 42. APDIO Portuguese OR Society, 2004.
- [175] P. Berthomé, A. Gueye, D. Nott, F. Quessette, and A. Troubnikoff. Managing QoS requests in the optical DAVID network. In *Photonics in Switching*, pages 83-85, Sept. 2003.
- [176] T. Faik. About the b-continuity of graphs. In *CTW on Graphs and Combinatorial Optimization (CTW),* 2004. to appear in the Electronic Notes in Discrete Mathematics.
- [177] P. Fraigniaud and P. Gauron. Brief announcement: An overview of the content-addressable network D2B. In 22nd ACM Symp. on Principles of Distributed Computing (PODC), page 151, 2003.

## **Doctoral dissertations and Habilitations**

- [178] I. Alvarez-Hamelin. Routage dans Internet : trafic autosimilaire, multicast et modèles de topologie. PhD thesis, Université Paris Sud, 2002.
- [179] M. Ben-Dhaou. Optimisation du placement de tâches dans les sphères distribuées. PhD thesis, Université Paris Sud, 2003.
- [180] M. Valencia-Pabon. *Complexité du problème du routage dans les réseaux de télécommunications : anneaux, arbres et grilles*. PhD thesis, Université Paris Sud, 2000.
## Reports

- [181] D. Amar, E. Flandrin, and G. Gancarzewicz. Cyclability in bipartite graphs. Technical report, Wydzialn Matematyki Stosowanej, AGH (Krakow), 2003.
- [182] D. Amar, E. Flandrin, G. Gancarzewicz, and A. Wojda. Bipartite graphs with every matching in a cycle. Technical Report 1348, LRI, Université Paris Sud, 2003.
- [183] R. Andrade, A. Lisser, G. Plateau, and N. Maculan. A branch and bound algorithm for the stochastic network design problem. Technical Report 3, LIPN, 2002.
- [184] R. Andrade, A. Lisser, G. Plateau, and N. Maculan. Planning network design under uncertainty with fixed charge. Technical Report 5, LIPN, 2002.
- [185] R. Andrade, A. Lisser, G. Plateau, and N. Maculan. Branch and bound strategies for the design of backbone telecommunications networks. Technical Report 6, LIPN, 2003.
- [186] D. Barth, P. Berthomé, C. Laforest, and S. Vial. Some Eulerian parameters about performances of a convergence routing in the 2D-mesh network. Technical Report 1346, LRI, Université Paris Sud, 2003.
- [187] D. Barth, J. Cohen, and T. Faik. Complexity of determining the b-continuity property of graphs. Technical Report 37, PRiSM, Université de Versailles St-Quentin en Yvelines, 2003.
- [188] M. Blidia, M. Chellali, and O. Favaron. Independence and domination in trees. Technical Report 1383, LRI, Université Paris Sud, 2004.
- [189] M. Blidia, M. Chellali, and O. Favaron. Ratios of some domination parameters in graphs and claw-free graphs. Technical Report 1393, LRI, Université Paris Sud, 2004.
- [190] F. Bonnans, M. Haddou, A. Lisser, and R. Rebai. Interior point methods with decomposition for multicommodity flow problems. Technical Report 3852, INRIA, 2000.
- [191] F. Bonnans, M. Haddou, A. Lisser, and R. Rebai. Proximal decomposition method for solving global survivability problem in telecommunication networks. Technical Report 4055, INRIA, 2002.
- [192] C. Delorme and J. Shu. Upper bounds on the length of the longest induced cycle in graphs. Technical Report 1321, LRI, Université Paris Sud, 2002.
- [193] T. Faik and J. F. Saclé. Some b-continuous classes of graph. Technical Report 1350, LRI, Université Paris Sud, 2003.
- [194] O. Favaron and M. A. Henning. Total domination in claw-free graphs with minimum degree two. Technical Report 1351, LRI, Université Paris Sud, 2003.
- [195] O. Favaron, R. Laskar, and D. Rautenbach. t-partitions and s-complete t-partitions of a graph. Technical Report 1394, LRI, Université Paris Sud, 2004.
- [196] E. Flandrin, H. Li, A. Marczik, I. Schiermeyer, and M. Wozniak. Chvátal-Erdös condition and pancyclism. Technical Report 1294, LRI, Université Paris Sud, 2001.
- [197] E. Flandrin, H. Li, A. Marczik, and M. Wozniak. A note on Jackson-Ordaz conjecture. Technical Report 1279, LRI, Université Paris Sud, 2001.
- [198] E. Flandrin, H. Li, A. Marczik, and M. Wozniak. A note on a generalization of Ore's condition. Technical Report 1322, LRI, Université Paris Sud, 2002.
- [199] E. Flandrin, H. Li, and B. Wei. A sufficient condition for pancyclability of graphs. Technical Report 1345, LRI, Université Paris Sud, 2002.
- [200] D. Forge and M. Kouider. Coverings of the vertices of a graph by small cycles. Technical Report 1374, LRI, Université Paris Sud, 2003.
- [201] R. Häggkvist and H. Li. Perfect matchings in bridgeless cubic graphs. Technical Report 1257, LRI, Univ. Paris Sud, 2000.

Research groups
GraFComm
Publications

- [202] R. Häggkvist and H. Li. Long cycles in graphs with some large degree vertices. Technical Report 1339, LRI, Univ. Paris Sud, 2002.
- [203] C. Hoang and M. Kouider. b-coloring of sparse graphs. Technical report, W.Laurier University, 2002.
- [204] Z. Hu and H. Li. Partition of a graph into cycles and vertices. Technical Report 1342, LRI, Univ. Paris Sud, 2002.
- [205] Z. Hu and H. Li. A note on Ore condition and cycle structure. Technical Report 1360, LRI, Univ. Paris Sud, 2003.
- [206] Z. Hu and H. Li. Weak cycle partition involving degree sum conditions. Technical Report 1352, LRI, Univ. Paris Sud, 2003.
- [207] M. Kouider. b-chromatic number: subgraphs and degrees. Technical Report 1392, LRI, Université Paris Sud, 2004.
- [208] M. Kouider and P. Vestergaard. Generalized connected domination in graphs. Technical Report 1377, LRI, Université Paris Sud, 2003.
- [209] H. Li. Woodall's conjecture on long cycles. Technical Report 1296, LRI, Univ. Paris Sud, 2001.
- [210] H. Li and J. Shu. The partitions of a strong tournament. Technical Report 1306, LRI, Univ. Paris Sud, 2002.
- [211] H. Li and M. Woźniak. A note on graphs containing all trees of given size. Technical Report 1336, LRI, Univ. Paris Sud, 2002.

## Other publications

[212] P. Fraigniaud. Structural information and communication complexity: A survey of sirocco 2001. SIGACT News 3(4), 2001.



Laboratoire de Recherche en Informatique / 2000-2004 report

B.2 / 65

# équipe Algorithmique & Complexité

**Responsable : Miklos Santha** 

ans les quatre dernières années des changements radicaux ont eu lieu dans l'équipe. Avec l'arrivée de Sophie Laplante et Frédéric Magniez en 2000 et celle de Julia Kempe en 2001, elle a atteint sa taille historiquement la plus élevée : quatorze personnes permanentes. Ces arrivées ont renforcé en particulier le thème calcul quantique, et grâce à cette politique courageuse de recrutement, le LRI est reconnu aujourd'hui comme un des principaux pôles dans ce domaine.

Par la suite la taille du groupe a diminué, et en 2004 il n'y avait plus que dix membres permanents. Jean-Pierre Tillich a rejoint l'INRIA comme chargé de recherche en 2001, et trois de nos membres sont devenus professeurs d'université : Alain Denise à Paris-Sud en 2002, Claire Kenyon à l'École Polytechnique en 2002 (en détachement de son poste à Orsay) et finalement Stéphane Boucheron à Paris 7 en 2004. Nous sommes fiers des promotions de nos anciens collègues, nous les remercions pour leurs apports à la vie de notre groupe, et nous leur souhaitons beaucoup de succès dans leurs nouveaux emplois.

Quant à la production scientifique, l'équipe a maintenu sa place en toute première ligne en Europe dans le domaine de l'informatique théorique. Elle a publié quasiment dans chacun des meilleurs colloques du domaine : FOCS, STOC, ICALP, STACS, MFCS, FSTTCS, SODA, SWAT, LICS, FPSAC, COMPLEXI-TY. En particulier, elle a eu 14 publications à FOCS et STOC, les deux colloques généralistes reconnus comme les plus prestigieux. Elle a eu également un grand nombre de publications dans diverses revues internationales d'informatique théorique, de combinatoire, de mathématiques discrètes, de théorie des nombres, de probabilités discrètes et de physique. Les thèmes de recherches principaux développés durant ces années sont restés l'algorithmique, la combinatoire, la complexité et le calcul quantique.



Research groups
Algo
Introduction

# Algorithms and Complexity

Head: Miklos Santha

n the last four years, the composition of our research group has gone through radical changes. With the arrival of Sophie Laplante and Frédéric Magniez in 2000, and of Julia Kempe in 2001, it has reached its historically largest size: fourteen permanents members. These arrivals have reinforced in particular the quantum computing theme of the group, and thanks to this courageous scientific policy and to the subsequent work, LRI is recognized today as a leading site in this field.

From this peak, the size of the group went down to ten permanent members by 2004. First Jean-Pierre Tillich joined INRIA as a researcher in 2001, and then three of our members became university professors: Alain Denise at Paris-Sud in 2002, Claire Kenyon at École Polytechnique also in 2002, and finally Stéphane Boucheron at Paris 7 in 2004. We are proud of the promotions of our former colleagues, thank them for their contributions to the life of the group, and wish them great success in their new scientific positions.

With respect to scientific production the group has maintained its status among the very best in Europe in theoretical computer science. We have published in almost all the high quality conferences in the field: FOCS, STOC, ICALP, STACS, MFCS, FSTTCS, SODA, SWAT, LICS, FPSAC, COMPLEXITY. In particular, we have had 14 publications in FOCS and STOC, the two most prestigious general conferences. We have also had a large number of publications in international journals in theoretical computer science, combinatorics, discrete mathematics, number theory, discrete probabilities and physics. The main research themes of the group over these years have remained algorithms, combinatorics, complexity and quantum computing.

# **Research Group Members**

## Personnel as of 01/01/2004

Full time faculty			
Name	First Name	Position*	Institution
ALLOUCHE	Jean-Paul	DR2	CNRS
BOUCHERON	Stéphane	CR1	CNRS
DE ROUGEMONT	Michel	PR2	Université Paris II
DURR	Christoph	MC	IUT d'Orsay
FERNANDEZ DE LA VEGA	Wenceslas	CR1	CNRS
GOUYOU-BEAUCHAMPS	Dominique	PR1	IUT d'Orsay
KEMPE	Julia	CR2	CNRS
LAPLANTE	Sophie	MC	IUT d'Orsay
MAGNIEZ	Frédéric	CR2	CNRS
MANOUSSAKIS	Yannis	PR1	IUT d'Orsay
SANTHA	Miklos	DR2	CNRS

Doctoral students			
ABOUELAOUALIM	Abdelfattah	D	Université Paris XI
ALBERT	Julien	А	Université Paris XI
DEGORRE	Julien	А	Université Paris XI
GOULARAS	Dionysis	D	Sté ASTERION Productions
MOHAMMAD-NOORI	Morteza	D	CROUS Versailles
NADEAU	Philippe	AC	Université Paris XI
PHILIPPS	Pierre	D	ENS Lyon
VERHOEVEN	Yves	D	TELECOMS
VERT	Régis	D	Sté MASA

Temporary personnel					
ADAMCZEWSKI	Boris	Post-doc	CNRS		
MILMAN	Pérola	Post-doc	EGIDE		
PEYRONNET	Sylvain	ATER	Université Paris VII		

\* See the glossary for acronyms.

Long term vis	sitors					
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
BACSO	Gabor	Hungarian	Hungarian Academy of Sciences	1/09/00	30/09/00	CNRS-HAS
				17/11/01	6/12/01	CNRS-HAS
BAHADUR	Abhinav	Indian	IIT Bombay	15/5/04	31/7/04	CNRS, ACI
BEREND	Daniel	Israelian	Univ. Ben-Gourion, Beer Sheva	14/06/00	12/07/00	
CHROBAK	Marek	American	UC, Riverside	17/06/01	13/07/01	Ministère de la Recherche
CURRIE	James	Canadian	University of Winnipeg, Canada	20/07/04	30/06/05	sabbatical
DESHPANDE	Amit	Indian	Chennai Mathematical Institute	22/04/02	22/06/02	ENS
FRIEDL	Katalin	Hungarian	Hungarian Academy of Sciences	4/02/02	4/07/02	EGIDE, QAIP grant
HARRISS	Edmund	British	Imperial College Londres	8/10/02	8/12/02	
HAUPTMANN	German	Mathias		5/03/01	10/03/01	
HOYER	Peter	Danish	University of Aarhus, BRICS	31/07/00	6/08/00	
IVANYOS	Gabor	Hungarian	Hungarian Academy of Sciences	5/11/00	4/12/00	CEPHYTEN, QAIP grant
				12/11/01	12/12/01	EGIDE, QAIP grant
				22/01/02	22/02/02	EGIDE, QAIP grant
				8/03/04	2/04/04	CEPHYTEN, RESQ grant
KARLOFF	Howard	American	Georgia Institute of Technology	4/10/00	16/10/00	CNRS-NSF
KHOSROVSHAHI	Gholamreza	Iranian	School of Math. Tehran	19/07/01	2/08/01	
KIWI	Marcos	Chilian	University of Chile	4/02/01	18/02/01	Visiting professor, I.G.M.
LAPLANTE	Sophie	Canadian	University of Chicago	1/09/99	31/08/00	CRSNG
LEI	Yaohui	Chinese	Université de Montréal	1/09/01	31/08/02	CREPUQ fellowship
LOZANO	Antonio	Spanish	LITEC Zaragoza	18/02/02	18/07/02	EGIDE, QAIP grant
LUGOSI	Gabor	Hungarian	University Pompeu Fabra, Barcelone	12/01/00	11/02/00	Visiting professor, UPS
NDOUNDAM	René	Camerounia	nUniversité de Yaoundé	15/01/03	14/07/03	EGIDE, ONU grant
RANDALL	Dana	American	Georgia Institute of Technology Atlanta	24/06/01	5/07/01	CNRS-NS
RAPAPORT	Ivan	Chilian	University of Chile	4/01/01	4/02/01	
				15/06/03	14/07/03	Visiting professor, UPS
REMPE	Lasse	German	University of Kiel	12/09/01	28/02/02	German fellowship
SALVADOR	Liliana	Portugese	University of Porto	27/07/03	25/08/03	EGIDE
				29/09/03	28/11/03	EGIDE
SCHULMAN	Leonard	American	Caltech, Pasadena	6/09/00	10/09/00	
SEN	Pranab	Indian	Tata Institute of Fondamental Research	1/07/01	1/07/02	EGIDE, QAIP grant
SZEGEDY	Mario	German	Rutgers University, NJ USA	3/09/03	2/11/03	EGIDE, RESQ grant
TUZA	Zsolt	Hungarian	Hungarian Academy of Sciences	14/06/00	30/06/00	CNRS-HAS
				10/10/00	17/10/00	CNRS-HAS
				12/10/01	30/10/01	CNRS-HAS
VAKHANIA	Nodari	Georgian	Autom. Univ. of the state Morelos, Mexico	1/10/03	31/12/03	CNRS
DE WOLF	Ronald	Dutch	CWI Amsterdam	31/07/00	6/08/00	
ZWISSIG	Thierry	Swiss	Université de Genève	4/02/02	15/02/02	
				1/04/02	12/04/02	

Algo Members

# **Group evolution**

#### Arrivals

- Sophie Laplante (MC, IUT), 2000
- Frédéric Magniez (CR, CNRS), 2000
  Julia Kempe (CR, CNRS), 2001

## Departures

- Jean-Pierre Tillich (CR, INRIA), 2001
- Alain Denise (Prof UPS, LRI, BioInfo group), 2002
  Claire Kenyon (Prof, Polytechnique), 2002
- Stéphane Boucheron (Prof, Paris-7), 2004

## **Research Description**

We organize our research activity into four areas: algorithms, combinatorics, complexity and quantum computation.

## **Algorithms Research team**

- Permanent members: Christoph Dürr, Wenceslas Fernandez De La Vega, Yannis Manoussakis, Miklos Santha.
- Non-permanent members: Jérémy Barbay, Jean-Christophe Dubacq, David Gross-Amblard, Claire Kenyon, Grégory Olocco, Emmanuel Prouff, Vlady Ravelomanana, Julien Stern, Jean-Pierre Tillich, Yann Verhoeven.

## Coding

Given a set of *n* words with associated probabilities  $p_1,...,p_n$  and some alphabet  $\Sigma$ , the Huffman code is a prefix free code over  $\Sigma$  minimizing *sum*  $|c_i| p_i$ . In optical storage devices or even in the telegraph, each letter of  $\Sigma$  has a different length. We gave an approximation algorithm scheme for constructing the best Huffman code in that setting [130].

Due to the increasing amount of stolen music, the music industry proposed a watermarking system for audio-files and opened a contest. In [113] we succeded in removing the marks from the audio-files, showing that the SDMI system is not secure. We also developed a system to detect watermarkings in [156].

Using ideas of Carlach and Vervoux (Symposium on Applied Algebra, Algebraic Algorithms, and Error-Correcting Codes, 1999) for block turbo-codes, we were able to construct specific *Tanner graphs*, which lead to codes with interesting properties [117, 151].

#### **Approximation Algorithms**

When an optimization problem is NP-complete, one might want to approximate it. Approximation schemes are algorithms which together with the problem description get a number  $\boldsymbol{\varepsilon}$  and output a solution which is  $\boldsymbol{\varepsilon}$  - close to the optimum. If the running time is not only polynomial in the size of the input but also in  $\frac{1}{\boldsymbol{\varepsilon}}$ , then it is a *fully* polynomial time approximation scheme.

One example is the constraint satisfaction problem, where are given m clauses on n boolean variables. Depending on the variant, we might want to minimize or maximize the number of satisfied clauses, which contain literals combined with OR or with XOR. The nearest codeword problem can be encoded in that setting. We gave approximation schemes for various variants [109, 26, 34].

Answering a problem of Woeginger and Yu (Information Processing Letters, 1992) in the affirmative, we gave a fully polynomial-time approximation scheme for the problem of finding two disjoint non-empty subsets out of n integers, with a ratio between the total sums closest to 1 [35]. On the other hand, when the difference between the total sums meant to be minimized, the problem is 2nk –approximable in polynomial time unless P=NP, for any constant k (see fig 1).

Figure 1: Finding two subsets of closest total sum.



A cut of an edge-weighted graph G is defined by a nontrivial vertex set, and its weight is the total weight of edges crossing the cut. Determining the maximal cut of a graph is a Max-SNP-hard problem, meaning there is no approximation scheme unless P=NP. We studied the special case when the graph has constant degree 3 [47], and we gave a randomized approximation scheme when the weights satisfy the triangle inequality [60].

We have obtained polynomial time approximation scheme for Min-Bisection [123] and Min-Sum Clustering [124]. We also studied the Bin-Covering problem [121], the Bin-Packing problem [140] and various other problems motivated by networks, in particular the Internet [66, 136, 72, 73].

## **Learning Theory and Stochastic Modelling**

Statistical learning consists of infering dependences from emprirical data, which are supposed to be collected by sampling from an unknown probability distribution. In this nonstandard statistical setting, it is crucial to develop data-driven risk estimates. In [33] (independent of Koltchinskii and Panchenko) we proposed to use the so-called Rademacher complexity to design data-dependent risk estimates. Thanks to the fact that in [42], we showed, among other properties, that those Rademacher complexities enjoy sub-Poissonian tail behavior, Rademacher complexities have become the standard tool to develop datadriven penalization techniques in model selection.

In [43, 38] we studied the Entropy method, which allows us to develop concentration inequalities for general functions of independent random variables.

In [64], a composite hypothesis testing problem was investigated in an information theoretical perspective, using large deviations theory.

We also applied techniques for analyzing random structures from interactive particle systems or queuing systems to the problems of studying the size of giant component in random graphs or to the classical allocation problems [32, 39, 40, 41].

### Scheduling

In a scheduling problem we are given n tasks and wish to assign them to machines, usually in order to minimize the maximal completion time over all jobs. An incredible number of variants of this problem exist depending on the choice of machines, restrictions to the schedule or objective function (see the webpage of P. Brucker and S. Knust). The goal is to identify which are tractable in polynomial time, which are NP-complete, and to come up with the most efficient algorithm in the first case and with approximation algorithms in the second case.

Research groups

Algo

Research description: Algorithms team In [28] we considered the problem of scheduling n jobs on m machines, where each job must be scheduled in parallel on a given set of machines, and each machine can execute at most one job at a time. In the case where jobs cannot be preempted, the problem is known to be NP-complete even for 3 machines, and we give an approximation scheme, improving on a constant approximation by Goemans (Discrete Applied Mathematics, 1995). In the preemptive case, the problem is NP-complete when the number of machines is unbounded, and for fixed m there was an algorithm with m in the exponent of the running time (Ph.D. thesis by A. Krämer, 1995) which we beat with a linear time algorithm.

When jobs have the same processing time, it allows exchanges on a schedule; an interesting structural property which can be exploited in algorithms. In [31] we improved a polynomial time algorithm for a preemptive scheduling problem with equal processing time, on a single machine, where the goal is to maximize the number of jobs scheduled inside an interval defined by a release time and a deadline. In [105] we consider a scheduling problem with equal processing time and given time windows in which scheduling is allowed. This problem appears in air traffic control, and we give results for different variants (see fig 2).



Figure 2: Aircrafts arrive at different moments at the airport and can wait some bounded time in a waiting loop before approaching the runway. The goal is to maximize the minimal distance between two landings.

We also studied the scheduling problem where each job has a given *execution cost* for being executed on a specific processor and any pair of jobs assigned to different processors have a given *communication cost*. For various variants we given exact polynomial time algorithms or an approximation scheme [61].

#### **Cellular Automata**

In order to understand the dynamics of cellular automata we studied their relationship with other models. In [125] we focused on signals encoded in higher dimensional cellular automata. In [52] we studied one-dimensional cellular automata under the theory of communication complexity (see fig 3).



Figure 3: A communication matrix associated to the cellular automata 94 in Wolfram's numbering.

#### **Key references**

- [31] P. Baptiste, M. Chrobak, C. Dürr, W. Jawor, and N. Vakhania. Preemptive scheduling of equal-length jobs to maximize weighted throughput. *Oper. Res. Lett.*, 32(3):258-264, 2004.
- [34] C. Bazgan, W. Fernandez de la Vega, and M. Karpinski. Polynomial time approximation schemes for dense instances of minimum constraint satisfaction. *Random Structures Algorithms*, 23(1):73-91, 2003.
- [35] C. Bazgan, M. Santha, and Z. Tuza. Efficient approximation algorithms for the subsetsums equality problem. *J. Comput. System Sci.*, 64(2):160-170, 2002.
- [42] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Random Structures Algorithms*, 16(3):277-292, 2000.
- [60] W. Fernandez de la Vega and C. Kenyon. A randomized approximation scheme for metric MAX-CUT. *J. Comput. System Sci.*, 63(4):531-541, 2001. Special issue on FOCS 98 (Palo Alto, CA).

## **Combinatorics Research team**

- Permanent members: Jean-Paul Allouche (leader), Christoph Dürr, Dominique Gouyou-Beauchamps.
- Non-permanent members: Boris Adamczewski, Julien Albert, Morteza Mohammad-Noori, Philippe Nadeau, Jia-Yan Yao.

The group works on combinatorics and related areas. Three directions can be roughly underlined:

## **Discrete Tomography**

Tomography is the area of reconstructing *k*-dimensional objects from (*k*-1)-dimensional projections. In *discrete tomography* we usually want to reconstruct a 2-dimensional matrix from projections on the rows and columns. A well-known and standard problem consists of constructing a 0-1 matrix from which we only know the number of 1's in each column and in each row. (The matrix may not be unique.) This problem can be solved by a greedy polynomial-time algorithm. Now think of 1's representing pebbles, and suppose they are colored. The projections now count different colors separately. We succeeded to show that the problem is NP-complete for 3 colors [49], while previously this was known only for 6 colors. Unfortunately the 2 color case remains open. A generalization of this is the reconstructing problem of tilings, for a fixed set of tiles. First, attention was brought to the domino tiling case, which we could solve only partially [54]. In [48] we classify as best as we can the tile-sets for which the problem is NP-complete and the tile-sets for which it is polynomially solvable.



Figure 1: A domino tiling of a grid and its projections. Combinatorics team

Algo

Research groups

Research Description:

### **Enumerative combinatorics**

Part of our work addresses questions in enumerative combinatorics, in particular the study of certain classes of polyominoes [132] and directed animals [50], as well as the study of sandpiles [51]. These combinatorial structures are related to models with phase transitions in statistical physics, or to discrete dynamical systems.

We have obtained enumeration results for several classes of convex polyominoes on the hexagonal lattice. These polyominoes are considered up to rotation and reflection, as objects living freely in space. We give explicit formulas or implicit functional equations for the generating series, which are convenient for computer algebra. Thus computation can be carried out up to area 70. We are also interested in enumeration of FPL configurations (Fully Packed Loops Configurations). There is a simple one-to-one correspondence between FPL and alternating-sign matrices. Recently, conjectures by J.-B. Zuber (CEA Saclay) have been proven in our subgroup, in conjunction with researchers at Claude Bernard University (Lyon-I).

## **Combinatorics on words**

We focus in particular on (infinite) words generated by finite automata, also called "automatic sequences". The study of these sequences is at the frontier between combinatorics, theoretical computer science, number theory, harmonic analysis, fractals...

Although several results on the structure of automatic sequences as well as links and differences with cellular automata sequences have been obtained, a large part of our study deals with transcendence results for which the tools provided by finite automata prove very useful and give "simple" proofs. Transcendence of formal power series with nonzero characteristic has first been a natural domain to apply the theory of automatic sequences. More recent results deal with transcendence of real numbers: if the expansion of a real number in a given integer basis is "too regular", e.g., generated by a finite automaton, this number must be either rational or transcendental (it cannot be algebraic irrational). If the continued fraction expansion of a real number is "too regular", e.g., generated by a finite automaton, this number must be either quadratic or transcendental (it cannot be algebraic of degree larger than 3). Results in this direction can be found in [16] and [11] for example. A major achievement has been obtained by Adamczewski who proved the following result [5]: an irrational real number having block complexity in O(n) must be transcendental (where the block complexity of an infinite sequence is given by the number of different subblocks of a given length occurring in the sequence). This result contains all previous results on the subject, although it is still far away of an old and still open conjecture by Borel, a weak form of which asserts that irrational numbers having a block complexity (in base *b*) strictly less than *b<sup>n</sup>* must be transcendental.

### **Key references**

- [5] B. Adamczewski, Y. Bugeaud, and F. Luca. Sur la complexité des nombres algébriques. *Comptes Rendus de l'Académie des Sciences, Serie I*, 339(11-14), 2004.
- [16] J.-P. Allouche, J. L. Davison, M. Queffélec, and L. Q. Zamboni. Transcendence of Sturmian or morphic continued fractions. J. Number Theory, 91(1):39-66, 2001.
- [48] M. Chrobak, P. Couperus, C. Dürr, and G. Woeginger. On tiling under tomographic constraints. *Theoret. Comput. Sci.*, 290(3):2125-2136, 2003.
- [50] S. Corteel, A. Denise, and D. Gouyou-Beauchamps. Bijections for directed animals on infinite families of lattices. *Ann. Comb.*, 4(3-4):269-284, 2000. Conference on Combinatorics and Physics (Los Alamos, NM, 1998).

[51] S. Corteel and D. Gouyou-Beauchamps. Enumeration of sand piles. *Discrete Math.*, 256(3):625-643, 2002. LaCIM 2000 Conference on Combinatorics, Computer Science and Applications (Montreal, QC).

## **Complexity Research team**

- Permanent members: Michel de Rougemont, Sophie Laplante, Frédéric Magniez, Miklos Santha.
- Non-permanent members: David Gross, Sylvain Peyronnet, Julien Stern.

Logic brings some light on the inherent computational difficulty of problems, produces efficient methods for verification and a general framework where definability, provability are closely linked to complexity issues. In descriptive and algebraic complexity, one can produce upper bounds for decision, optimization and counting problems, as well as for their approximations.

Our work concerns Logic on finite models, Structural Complexity and Kolmogorov Complexity.

## Logic

Model Checking is an important area of Logic in Computer Science. It proposes efficient techniques to verify a property of a transition system. Classical techniques use a compressed representation of relations such as OBDD or automata. We generalize this approach in two different ways: we first study definability question on compressed models and then study notions of approximate Model Checking.

### **Definability on compressed models**

The Finite Model Theory tradition emphasizes methods of Model theory where complexity is measured by the number of variables, the number of alternations of the quantifiers or the length of the formulas, when a model is explicitly given. We consider models given in a compressed form and try to efficiently verify a property of the original structure. In [100], one studies finite compressed structures, for example the binary words compressed with various compression schemes. On the Run-length schema, any 1st order property on the words is also 1st order definable on the compressed words. On the Lempel-Ziv schema on the other hand, one shows that a 1st order property on the words is not necessarily 1st order definable on the compressed words but is definable in the FO(TC) logic. A property of words definable in the FO(TC) logic will be definable in the same logic on the compressed words, but with a double fixed-point arity. In [98] , we study the structure of automata for LZ-78, i.e. automata reading a compressed word and deciding a regular property. This type of result clarifies the role of the compression schemes for the verification of properties. It is important to know if specifications can be efficiently verified on compressed structures.

#### Approximate Model Checking

If exact Model Checking is too hard, from the complexity point of view, we study how to define "approximate verification". In [144], we introduce an approximate verification based on property testing. If a specification has a tester, we show how to define a probabilistic abstraction of a syntactic program such that many errors will be detected with high probability. In the case of classes of regular trees, we develop testers in [149] and correctors in [114], particularly adapted to the correction and ranking of XML documents.

Research groups

#### Algo

Research description: Complexity team

### **Structural Complexity**

Testers have been used in numerical computation, structural complexity, and approximation algorithms. Nevertheless, only models for exact or approximate computations with absolute error had been considered. We showed that it was possible to generalize the testers in two important cases : when the authorized error is sublinear in the input size [77] ; and when the error is relative [148], i.e. proportional to the output size. We validated this extension by constructing testers for the linear or polynomial functions in the first case, and for the linear or multilinear functions in the second case. In [149], we studied property testing on the class of ordered trees and showed that regular tree properties have testers.



*Figure 1: Operations on XML-trees to make them DTD-compliant (matching the document type definition).* 

## **Kolmogorov complexity**

The work on Kolmogorov complexity concerns upper and lower bounds. It is a measure of the amount of randomness contained in the string. We have generalized this measure to quantum computing and have proven that it satisfies a certain number of important properties similar to its classical analog [110]. In [46], better bounds were given for a basic problem which has interesting consequences for the complexity of probabilistic classes, and for the comparison of counting and decision complexities. In [145], the link with quantum Kolmogorov complexity has been studied and allows to prove lower bounds for models of quantum computations.

## **Key references**

- [77] M. Kiwi, F. Magniez, and M. Santha. Approximate testing with error relative to input size. *J. Comput. System Sci.*, 66(2):371-392, 2003.
- [100] F. Afrati, H. Leiß, and M. de Rougemont. Definability and compression. In 15th Annual IEEE Symposium on Logic in Computer Science (Santa Barbara, CA, 2000), pages 63-73. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.
- [144] S. Laplante, R. Lassaigne, F. Magniez, S. Peyronnet, and M. Rougemont. Probabilistic

abstraction for model checking: An approach based on property testing. In *Proceedings* of 17th IEEE Symposium on Logic in Computer Science, pages 30-39, 2002.

- [145] S. Laplante and F. Magniez. Lower bounds for randomized and quantum query complexity using Kolmogorov arguments. In *Proceedings of 19th IEEE Conference on Computational Complexity*, 2004, pages 294-304.
- [149] F. Magniez and M. Rougemont. Property testing of regular tree languages. In *Proceedings* of 31st International Colloquium on Automata, Languages and Programming, volume 3142 of Lecture Notes in Comput. Sci., pages 932-944. Springer, 2004.

## Quantum Computation Research team

- Permanent members: Miklos Santha (head), Christoph Dürr, Julia Kempe, Sophie Laplante, Frédéric Magniez.
- Non-permanent members: Julien Degorre, Pierre Philipps, Yves Verhoeven, Perola Milman, Pranab Sen.

Quantum computation has been a very active research area in the past 10 years. Feynman (Nobel prize 1965) initiated the idea of constructing a computer exploiting the laws of quantum mechanics, in order to solve problems that are intractable on classical machines. Today, two major results make the quantum model extremely attractive. Shor (Gödel prize 1999) found a quantum algorithm that factors any integer in polynomial time, rendering cryptosystems such as RSA vulnerable. Grover then found an optimal algorithm that finds a distinguished element in an unstructured space quadratically faster than any randomized algorithm.

Our research activity is first and foremost centered around studying the computational potential of quantum computing. We have contributed to the elaboration of the model, the development of complexity measures, and the construction of new quantum algorithms. We also have collaborations with physicists whose objective is to study the practical problems raised by the implementation of quantum computers.

## Algorithms

#### Hidden Subgroup problem

Shor's Quantum algorithms for factoring and computing discrete logarithm come from the partial (ie, for abelian groups) resolution of a more general problem which is the Hidden Subgroup problem (HSP): Given a periodic function f on a group G, HSP is to determine efficiently the subgroup H of periods under the promise that f is injective on G/H. However, few results are known for non abelian groups. One of the most famous non abelian instances of HSP, where G is the symmetric group, is the Graph Isomorphism Problem which is one of the most important challenge of quantum computing.

We gave contributions to HSP twice. In [68], we showed how to combine Shor's results with the elegant theory of black-box groups of Babai and Beals. In particular, we generalized and simplified the previous works of Watrous (STOC'01) and Hallgren, Russell and Ta-Shma (STOC'00).

More recently [128], we solved HSP when the group G is solvable with constant exponent and constant length. This is proven by induction on the length of the composition series of G. Our induction is the first one to succeed for HSP. The base case is a partial solution of a largely studied problem (Ettinger, Høyer and Knill, STACS'99, Regev FOCS'02).

Research groups

Algo

Research description: Quantum Computation team Another contribution is [129, 62], where we improved and extended a work of Buhrman, Fortnow, Newman, and Röhrig (SODA'03) constructing efficient quantum property testers for properties connected to HSP.

#### Collision problem

We also gave a new algorithm comparable to Grover's for the Collision problem: Given a non injective function f defined on a domain of size N, the problem is to find a collision pair (x,y) such that  $x \neq y$  and f(x)=f(y). This problem, which is purportedly hard classically, is important, because it is the basis of many cryptanalysis strategies against secret key cryptosystems. Moreover, the problem can be solved in time  $O(N^{1/3})$  when the function has a linear number of collision pairs (Brassard, Høyer, and Tapp, ACM SIGACT News'97). This algorithm has been proven optimal (Aaronson, STOC'02 and Shi, FOCS'02).

We gave [115] a quantum algorithm in  $O(N^{3/4})$  that finds a collision pair in the general case. The optimality of this algorithm was an open problem until this year. Recently, Ambainis showed that  $N^{2/3}$  is the correct answer using a quantum walk based algorithm (Ambainis, FOCS'04), which is optimal due to Aaronson and Shi.

#### Local Search

A way of relaxing the minimum/maximum finding problem is to look for a solution which is optimal only in some neighborhood structure. Studying these new problems has lead Johnson, Papadimitriou and Yannakakis to introduce the Polynomial Local Search (PLS) class.



The class PLS is a subset of TFNP, the family of total function problems from NP. One can consider this class as a potential source of problems which might admit efficient quantum algorithms. Another important subclass of TFNP is PPP (Polynomial Pigeonhole Principle) which contains the Collision problem. Lower bounds by Aaronson and Shi imply that the deterministic and quantum query complexities of PPP problems are polynomially related.

We [154] proved an analogous result for PLS. As a consequence, if an efficient quantum algorithm is ever to be found for a problem in PLS, it must exploit its specific structure. Another consequence is the previously unknown relation between deterministic and randomized query complexities.

#### *Graph properties*

We considered quantum algorithms for natural graph problems both in the adjacency matrix model and in an adjacency list-like array model. We gave almost tight lower and upper bounds for the bounded error quantum query complexity of Connectivity, Strong Connectivity, Minimum Spanning Tree, and Single Source Shortest Paths. All our results improve classical ones.

### Quantum Walks

In [87] we introduced a new quantum walk based algorithm to simulate Grover Search algorithm on an unstructured database. Our approach has been successfully applied and improved by Ambainis (FOCS'04) to give a tight algorithm for finding a collision pair when the number of such pairs is unknown. It has also been shown by Szegedy (FOCS'04) that quantum walks on graphs search is quadratically faster than classical walks.

In [137] we studied the hitting time of a discrete quantum walk and showed that its behaviour differs strikingly form its classical counterpart. We showed that on the hypercube of dimension n the quantum walk needs only time poly(n) to hit the opposite corner from the starting position, whereas a classical work needs exponential time to hit that corner.

## Complexity

## Complexity classes

A seminal result in classical complexity theory is the celebrated Cook-Levin theorem which states that 3-SAT is NP-complete. Kitaev gave a quantum analogue of the above results where NP is now the quantum analogue of the class MA, that is QMA. Kitaev found a natural analogue of the SAT problem, the Local Hamiltonian problem, whose *k*-local version is NP-hard for all  $k \ge 2$ , and QMA-complete for  $k \ge 5$ .

We [70] showed that already 3-local Hamiltonian is QMA-complete, by using a strong projection on the "legal" space of states. Moreover, we have recently extended this result to 2-local Hamiltonian (in submission).

## Models

We have formally defined a quantum version of cellular automata, which is both a model of computation and a physical model [55]. We then solved the problem of detecting feasible instances. This was important to verify the modern version of Church-Turing Thesis in the quantum world. Independent of us, similar research on quantum cellular was done by van Dam and Watrous.

We have recently shown [102] that any standard quantum algorithm, i.e. based on the circuit model, can be equivalently represented as an adiabatic algorithm. This equivalence gives a strong indication that the adiabatic technique is a powerful technique that could lead to the discovery of new quantum algorithms. Adiabatic computing has been introduced by Farhi (Science'01).

#### Communication Complexity

We also have contributions [86, 135] in quantum Communication Complexity. Here two participants can use a quantum channel for communicating more efficiently than classically.

Research groups

## Algo

Research description: Quantum Computation team

## **Physical Aspects**

#### Quantum testing

The goal of quantum testing is to develop procedures to test whether a quantum computing device correctly computes what it purports to do. Mayers and Yao (FOCS'98) described how to carry out testing if a photon source is good enough to be used in the quantum key distribution protocol of Bennett and Brassard. We [157] showed that it is possible to test quantum gates classically.

#### Implementations

It has been shown earlier (Kempe et al. 2000) that the exchange interaction alone is universal, provided the quantum state is encoded in a specific fashion. In [67] we give the explicit encoding of one qubit into four qubits that achieves this.

In [71] we detail the proof that the XY-interaction (anisotropic exchange) is universal with encoding.

In [88] we study the robustness of several entangled states in the following model: each qubit is subject to a partially depolarizing channel with some noise rate.

We also have other contributions in the geometry of 2-qubit systems [82], the implementation of quantum algoritms using cavity QED [83], and the manipulation of phase gates using selective interaction [89].

#### Perspectives

We will continue to look for efficient algorithms. One of the algorithmic technique that we intend to explore is quantum walks. Recently, vast progress has been made in the application of quantum walks to improve quantum algorithms. Another important algorithmic technique that we intend to pursue is adiabatic computing. The equivalence of this model to the standard one motivates us to discover new algorithms based on adiabatic computing.

In classical cryptography, many cryptographic tasks require hardness assumptions. In the quantum world, there are provably secure protocols for key distribution based only on validity of quantum mechanics, with no hardness assumptions. A major open question, one which we would like to contribute to, is whether similar quantum protocols exist for other cryptographic tasks. Probably one of those protocols that has generated the most exciting developments recently is Coin flipping.

It may be possible to combine our approach of testing quantum gates with Mayers and Yao's for testing quantum sources so that the tester is not required to trust the measurement devices nor the source of classical and quantum states. This has both theoretical and practical interest since it is motivated by recent implementations of quantum computing based on NMR for which even classical states are hard to prepare.

#### **Key references**

- [102] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev. Adiabatic quantum computation is equivalent to standard quantum computation. In *Proc. 45th Annual IEEE Symp. on Foundations of Computer Science (FOCS)*, 2004.
- [126] C. Dürr, M. Heiligman, P. Høyer, and M. Mhalla. Quantum query complexity of some graph problems. In Proceedings of the 31st International Colloquium on Automata, Languages

and Programming (ICALP), pages 481-493, 2004.

- [128] K. Friedl, G. Ivanyos, F. Magniez, M. Santha, and P. Sen. Hidden translation and orbit coset in quantum computing. In *Proceedings of the 35th Annual ACM Symposium on Theory* of Computing (STOC), pages 1-9, 2003.
- [154] M. Santha and M. Szegedy. Quantum and classical query complexities of local search are polynomially related. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, pages 494-501, 2004.
- [157] W. van Dam, F. Magniez, M. Mosca, and M. Santha. Self-testing of universal and faulttolerant sets of quantum gates. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (STOC)*, pages 688-696, 2000.

## **Research Perspective**

Our group is a federation of autonomous researchers, and is well integrated in the international scientific community. The number of high level publications and research grants witnesses the quality of its members. Scientific excellence is and will remain in the future as our highest aspiration.

Algorithmics remains our federating field. Nonetheless, with the departure of four senior members, this theme has suffered serious losses, and it needs reinforcing in the coming years. This could come from hiring new persons in the group and from promotion of our assistant professors.

We consider that teaching and advising of doctoral students remains a central piece of the healthy life of our group. We take our contributions to the new master programs very seriously, including the MPRI where seven team members give lectures, and one course is entirely taught by us. We feel that the fact that currently none of our members has a teaching position at the University Paris-Sud, Centre d'Orsay, is a serious drawback, and will try to correct it.

Between January and April 2006 we will co-organize the program "Quantum information, computation and complexity" at the Institut Henri Poincaré in Paris. During this semester will take place the ninth Workshop on Quantum Information Processing, organized by our group. This workshop is the single most important annual event in quantum computation.

Finally we can only hope that our distinguished team member W. Fernandez de la Vega, praised by the international community for his high quality and innovative research work covering several fields over more than thirty years, will soon be promoted by the CNRS to the DR position.

Research groups

## Algo

**Research Perspective** 

B.2.3 / Algo

# Highlights

During this period our group has produced 10 STOC papers (ACM Symposium on the Theory of Computing), 4 FOCS papers (IEEE Symposium on Foundations of Computer Science) which are the two best general conferences in theoretical computer science, together with 4 SODA papers (Symposium on Discrete Algorithms) and 3 CCC papers (IEEE Conferences on Computational Complexity). Jean-Paul Allouche has published an impressive list of 14 journal papers.

Our team is involved in few program committees but these are from prestigious conferences. The evolution over the years is quite smooth; the increase in contracts over the past four years is due to our research in quantum computing.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international),
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

## Honors

## Prizes and awards

- Christoph Dürr, best paper award at the *International Colloquium on Automata, Languages and Programming Conference track* A (ICALP'04), Turku, Finland, July 2004.
- Julia Kempe, Bernard Friedman Memorial Prize for the best Ph.D. dissertation in the applied sciences, Department of Mathematics, UC Berkeley in May, 2002.
- Frédéric Magniez, AFIT (*Association Française d'Informatique Théorique*) prize for best Ph.D. thesis in Theoretical Computer Science, 2000.

## **Keynote addresses**

- Jean-Paul Allouche, Aperiodic Order Workshop, Edmonton, Canada, 2000.
- Jean-Paul Allouche, Aperiodic Order Workshop, Oberwohlfach, Allemagne, 2001.
- Jean-Paul Allouche, 29<sup>e</sup> École de Printemps d'Informatique Théorique, Cessac, 2001.
- Jean-Paul Allouche, *Discrete Models. Combinatorics, Computation and Geometry*, IHP Paris, 2001.
- Jean-Paul Allouche, Journées Montoises d'Informatique Théorique, Montpellier, 2002.
- Jean-Paul Allouche, Journées Arithmétiques, Graz, Autriche, 2003.
- Julia Kempe, International Conference on Quantum Information: Conceptual Foundations, Developments and Perspectives, Oviedo, Spain, 2002.
- Julia Kempe, Conference on Inhomogeneous Random Systems, Cergy-Pontoise, France, 2003.
- Julia Kempe, invited talk in the *Workshop on Quantum Information Processing*, Waterloo, 2004.
- Frédéric Magniez and Miklos Santha, 1<sup>st</sup> Conference on Theoretical Aspects of Computer Science, Tehran, 2000. Exact and Approximate Testing/Correcting of Algebraic Functions: A Survey, Lecture Notes in Comput. Science, volume 2292, pages 30-83.
- Frédéric Magniez, invited talk in the *Workshop on Quantum Information Processing*, Waterloo, 2004.
- Miklos Santha, 2<sup>nd</sup> European Quantum Information Processing and Communication Workshop, Torino, 2001.
- Miklos Santha, 6th Workshop on Quantum Information Processing, Berkeley, 2002.
- Miklos Santha, *Gordon Research Conference on Quantum Information Science*, Ventura, California, 2004.
- W. Fernandez De La Vega, Journée de l'Informatique Messine, 2000.

## Other honors

• Miklos Santha, member of the *STACS steering committee (International Symposium on Theoretical Aspects of Computer Science)* since 2002.

Research groups

Highlights Honors

# **Evaluation of research**

## **Editorial boards**

- Advances in Applied Mathematics: Jean-Paul Allouche
- Journal de Théorie des Nombres de Bordeaux: Jean-Paul Allouche
- Journal of Integer Sequences: Jean-Paul Allouche
- Mathématiques et Sciences Humaines: Jean-Paul Allouche
- Special issues of *Theoretical Computer Science*, 2001: Yannis Manoussakis, co-editors: G. Chang, M. Deza and J-M. Steyaert.

## **Program committees**

#### **International events**

- 31<sup>st</sup> Annual Conference on Current Trends in Theory and Practice of Informatics, to be held at Slovak Republic, January 2005: Frédéric Magniez
- RNC6, Real Numbers and Computers, Schloss Dagstuhl, Germany, 2004; RNC5, Ecole Normale Supérieure de Lyon, 2003; RNC4, Schloss Dagstuhl, Germany, 2000: Jean-Paul Allouche
- 17th Conference on Learning Theory, Banff, Canada, 2004: Stéphane Boucheron
- Conference on graph theory in memory of Claude Bergé, 2004: W. Fernandez De La Vega
- ICALP, 29th and 31st International Colloquium on Automata, Languages and Programming, Malaga 2002, Turku, 2004: Claire Kenyon (2002), Miklos Santha (2004).
- 45th Annual IEEE Symposium on Foundations of Computer Science, Rome, 2004: Miklos Santha
- 14th International Symposium on Fundamentals of Computation Theory, Malmö, 2003: Miklos Santha
- IEEE Conference on Computational Complexity, 2002: Sophie Laplante
- FOCS, 43rd Annual IEEE Symposium on Foundations of Computer Science, Vancouver, 2002: Claire Kenyon.
- FPSAC, Formal Power Series and Algebraic Combinatorics, Scottsdale, Arizona, 2001: Alain Denise
- LICS, IEEE Logic in Computer Science, 2001: Michel de Rougemont
- CIAC, 4th Italian Conference on Algorithms and Complexity, Rome 2000: Yannis Manoussakis

## National events

- EA01, *Cinquième Conférence Internationale sur l'évolution Artificielle*, Université de Bourgogne, 2001: Jean-Paul Allouche
- Huitièmes Journées Montoises d'Informatique Théorique, Université Marne-la-Vallée, 2000: Jean-Paul Allouche

## Evaluation committees and invited expertise

- National committee for scientific research, *Comité National de la Recherche Scientifique*, section 01 (mathematics) and section 44 (bioinformatics): Jean-Paul Allouche, member
- EU IST Programme, FET ALCOM-FT Project, 2001, 2002, 2004: Miklos Santha, reviewer
- National Science Foundation ITR proposals, Washington D.C., 2003: Julia Kempe, panel member

## **Other evaluation activities**

- Reviewer for several EU IST Programme: Miklos Santha
- Reviewer for Ph.D. dissertation: Yannis Manoussakis (1), Michel de Rougemont (1), Miklos Santha (1)

# Volunteer professional service

## Management positions in scientific organisations

- International Symposium on Fundamentals of Computation Theory, FCT: Miklos Santha, member of the steering committee since 1999.
- Workshop on Quantum Information Processing, QIP: Miklos Santha, member of the steering committee since 2002.
- European annual workshop on Kolmogorov complexity, TAI: Sophie Laplante, member of steering committee, 1998-2002.
- National network on Quantum Information and Communication (*"GdR Information et Communication Quantique"*): Miklos Santha, member of the board, 2001-2004.
- CNRS Working group ("Action Spécifique") Nouveaux Modèles de Calcul: Algorithmes et Complexité: Miklos Santha, co-chair with Etienne Grandjean, 2001-2002.
- Jeune Equipe Informatique Quantique, funded by national program ("Action incitative") for junior researchers: Miklos Santha, director, 2002-2003.

## Organisation of conferences and scientific events

- Special day in honor of Kolmogorov's 100<sup>th</sup> birthday, Complexity, Information, and Randomness: The Legacy of Andrei Kolmogorov held in conjunction with IEEE Conference on Computational Complexity 2003: Sophie Laplante, organizer.
- Meeting on the technical challenges for the developpement of web TV platforms for pupils, December 2003, Orsay, Paris: Yannis Manoussakis, organizer.
- Meeting on the use of new media and internet TV as educational tools, March 2003, Orsay, Paris: Yannis Manoussakis, organizer.
- *Journées de l'Action spécifique Nouveaux modèles de calcul,* November 2002, Paris: Sophie Laplante and Frédéric Magniez, organizers.
- RAND-APX workshop, April 2002, Paris: Christoph Dürr, Sophie Laplante, Frédéric Magniez and Miklos Santha, organizers.
- Workshop ALEA 2002, CIRM Marseille, March 2002: Dominique Gouyou-Beauchamps, co-organizer.
- Worskhop, "Complexity and finite models", University Paris II, 2002: Michel de Rougemont, organizer.
- Workshop *Information Quantique: Aspects Théoriques* at the Institut Poincaré, Paris, November 2001: Miklos Santha, co-organizer.
- European annual workshop on Kolmogorov complexity, TAI, September 2001, Porquerolles, France: Sophie Laplante, local arrangements.
- *Journée de veille et prospective de l'ASTI: Le traitement quantique de l'information,* Paris, October 2000: Miklos Santha, organizer.
- Conference in the Honor of Michel Mendès France, Bordeaux, 2000: Jean-Paul Allouche, co-organizer.

Research groups

## Algo

Evaluation of research Volunteer professional service

# **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Duration (months)	Total €
IST	M. Santha	QAIP	EU	CEPHYTEN	01/02/00- 31/12/02	36	159921
IST	M. Santha	RESQ	EU	CEPHYTEN	20/12/02- 19/12/05	36	283000
IST	M. Santha	RAND APX	EU	CEPHYTEN	1/07/00- 31/12/03	42	24420
SOCRATE	Y. Manoussakis	WEB TV	EU	UPS/SAIC	01/10/02- 30/09/04	24	25000
IST	W. Fernandez de la Vega	APPOL	EU	CEPHYTEN	1/05/00- 30/04/01	12	9000
IST	W. Fernandez de la Vega	APPOL 2	EU	CEPHYTEN	01/11/01- 31/10/04	36	28440
IST	C. Kenyon	Approximation on online algo.	EU	CEPHYTEN	08/10/01- 07/11/04	36	13739
ACI	S. Boucheron	GAP (NIDM)	MENRT	CNRS	20/08/03- 19/08/06	36	7 5 2 5
ACI	F. Magniez	Applications of quantum computation	MENRT	CNRS	16/08/02- 15/08/05	36	62708
ACI	Julia Kempe	Computer networks (SI)	MENRT	CNRS	25/07/03- 10/07/06	36	41 806
ACI (SI)	M. de Rougement	Approximate verification	MENRT	CNRS	15/09/03- 04/09/06	36	45150
ACI	J.P. Tillich	Error correcting codes and cryptography	MENRT	CNRS	2001-2003		75000
AI (JC)	M. Santha	Quantum computer science	CNRS STIC	CNRS	22/10/01- 21/10/03	24	25958
AS	M. Santha	New models of computation	CNRS STIC	CNRS	01/10/01- 31/09/02	12	16007
CIFRE	S. Boucheron	STE MASA		UPS/SAIC	13/01/03 - 12/01/06	36	12501

Note: See the glossary for acronyms.

## Summary of main projects

## QAIP

Partners: CWI, Amsterdam, the Netherlands; University of Latvia, Riga, Latvia; Oxford University, UK; University of Bristol, UK; University of Aarhus, Denmark; Hebrew University, Jerusalem, Israel; Weizmann Institute, Rehovot, Israel; Technion, Haifa, Israel; University of Waterloo, Canada; McGill University, Montreal, Canada; Université de Montréal, Canada; University of Calgary, Canada; University of California at Berkeley, USA.

QAIP's main objective was to study quantum computing

from a computer science angle. The project has focused on showing that quantum computers are actually useful, by exhibiting tasks which they can do significantly better/faster than classical computers. Its goals were to develop new algorithms and analyse the complexity of relevant problems on a quantum computer; to develop new forms of cryptography that make use of quantum mechanics; and to develop and improve methods for error correction in quantum computers.

#### European Commission grant IST-1999-11234

QAIP (Quantum Algorithms and Information Processing), 01/01/2000-31/12/2002. Scientific director for LRI: Miklos Santha.

## RESQ

Partners: Université Libre de Bruxelles, Belgium; CWI, Amsterdam, the Netherlands; University of Bristol, UK; Max Planck Institute for Quantum Optics, Germany; University of Utrecht, Netherlands; MTA SZTAKI, Hungary; University of Geneva, Switzerland; University of Cambridge, United Kingdom; University of Gdansk, Poland; University of Waterloo, Canada.

RESQ is an interdisciplinary project grouping together physicists, computer scientists, mathematicians and statisticians. One

European Commission grant IST-2001-37559, RESQ (Resources for quantum information), 01/01/2003-31/12/2005. Scientific director for LRI: Miklos Santha.

Research groups

## Algo

Contracts and grants

of the main objectives of the project is to bridge the cultural gap between these different disciplines and to develop a community of scientists from these disciplines that can work together and communicate together. More specific objectives of the project are to understand how quantum information can be manipulated in small scale systems; to improve our understanding of the nature of quantum information, both at a fundamental level and from the pragmatic point of view of testing quantum systems; to understand how information can be processed in distributed quantum systems, both from the point of view of algorithms and from the point of view of security and cryptography; and to design new quantum algorithms, study the power of quantum property testers, and explore the ingredients of a basic toolkit for designing quantum algorithms.

#### **RAND-APX**

Partners: University of Oxford, UK; University of Bonn, Germany; University of Edinburgh, UK; University of Leeds, UK; University of Lund, Sweden; Weizmann Institute, Israel.

The project's aim was to pursue studies in the areas of randomised, approximate, and quantum computation. It has covered novel and enhanced methods for the design and analysis of efficient randomised and quantum algorithms for problems of measurement and communication theory, which are totally intractable by existing methods.

#### WEB TV

In the field of education, the use of the technology possibilities requires research and development to ensure the efficiency of these techniques. An important challenge is how to create and experiment with new learning paradigms that take into account this new technology. The WebTV project focuses on the growth and development of webcasting and streaming media over the Internet and their use in educational contexts where students are not simple consumers but creators and producers. The partners include European schools, private companies and educational and computer science university departments. European Commission grant IST-1999-14036 RAND-APX (Randomised and Approximate Computation), 01/07/2000-31/12/2003. Scientific director for LRI: Miklos Santha.

European SOCRATE grant for the development of an Internet TV platform for pupils, 2002-2004. Scientifc director: Yannis Manoussakis.

Laboratoire de Recherche en Informatique / 2000-2004 report

# Collaborations

## **Cooperation agreements**

- CNRS/NSF grant between Christoph Dürr (LRI), Philippe Baptiste (Polytechnique) and Marek Chrobak (UC Riverside), 2003-2006.
- EGIDE Pessoa bilateral grant between Sophie Laplante (LRI) and Luis Antunes (Universidade do Porto, Portugal), 2002-2004.
- EGIDE Procope bilateral grant between Michel de Rougemont (LRI) and Hans Leiss (Munich University), 2002.
- EGIDE Alliance bilateral grant between Michel de Rougemont (LRI) and Marta Kwiatkowksa (University of Birmingham), 2003-2004.

## Collaborations leading to joint publications

We list the affiliations of some of our co-authors.

## France

- Ecole Polytechnique (K. Artiouchine, P. Baptiste, C. Léonard)
- ENS Paris (J. Stern), ENS Lyon (E. Rémila, G. Theyssier)
- ENST, Paris (G. Zémor)
- INRIA Rocq. (P. Flajolet, H. Ollivier, M. Golin)
- INRIA Sophia Ant. (M. Cosnard)
- Lab. Leibniz, Grenoble (M. Mhalla)
- Math. Lab. of Univ. Paris-11 (E. Gassiat, P. Univ. of Aarhus (P. Bro Miltersen) Massart)
- Univ. of Bordeaux-1 (M. Bousquet-Mélou)
- Univ. of Caen (J.-C. Carlach)
- Univ. of Grenoble (C. Simon)
- Univ. of Montpelier-2 (V. Berthé)
- Univ. of Paris-13 (C. Banderier)
- Univ. of Paris-6 (J. Cassaigne, I. Guessarian)
- Univ. of Paris-7 (M. Courbage, R. Lassaigne, R. Mosseri)
- Univ. of Paris-8 (C. Carlet)
- Univ, of Toulouse-3 (F. Gamboa)
- Univ. of Versailles (D. Barth, D. Gardy)

#### Europe

- CWI, Amsterdam (H. Buhrman, R. de Wolf)
- IBM Research, Zurich (G. Cheliotis)
- MTA SZTAKI, Budapest (K. Friedl, G. Ivanyos, Z. Tuza)
- National Technical Univ. of Athens (F. Afrati)
- Max-Planck Institut Tübingen (O. Bousquet)
- Pompeu Fabra Univ., Barcelone (G. Lugosi)
- TU Graz (R. F. Tichy)
- Univ. of Bielefeld (M. Baake)
- Univ. of Athens (I. Milis)
- Univ. of Bonn (M. Karpinski)
- Univ. of Bremen (G. Skordev)
- Univ. of Cambridge, UK (R. R. Weber)
- Univ. of Kiel (A.V. Fishkin)
- Univ. of Munich (H. Leiss, K. Jansen)
- Univ. of Paderborn (S. L. Bezrukov, R. Elsässer, B. Monien, R. Preis)
- Univ. of Rome-1 (T. Calamoneri, I. Finocchi, R. Petreschi)
- Univ. of Szeged (J. Csirik)
- Univ. of Twente (G. Woeginger)

## USA

- AT&T Labs (D. S. Johnson)
- Caltech (D. Bacon, D. Damanik, A. Kitaev, L. J. Schulman)
- dePaul Univ., Chicago (A. Berthiaume)
- Georgia Institute of Technology, Atlanta (D. Randall)
- Harvard Univ. (M. Mitzenmacher)
- IBM Watson Research Center (D. DiVincenzo)
- MIT (S. Lloyd, J. B. Orlin, P. W. Shor)
- NSA (M. Heiligman)
- Rutgers Univ. (M. Szegedy)
- UC Berkeley (D. Aharonov, P. Bartlett, R Jain, E. Mossel, Y. Peres, S. Myrgren, N. Shenvi, U. V. Vazirani, K.B. Whaley)
- UC Riverside (M. Chrobak, N. E. Young)
- UC Santa Barbara (W. van Dam)
- Univ. of Chicago (L. Fortnow)
- ${\ensuremath{\bullet}}$  Univ. of Washington (A.R. Karlin)
- Yale Univ. (R. Kannan)

## Other countries

- Hebrew University of Jerusalem (D. Malkhi, A. Shalev)
- Santiago de Chile (I. Rapaport, E. Goles, M. A. Kiwi)
- Tata Institute of Fundamental Research, India (J. Radhakrishnan)
- Technion (Y. Rabani)
- Tel Aviv Univ. (N. Alon, H. Kaplan, M. Krivelevich, O. Regev)
- Univ. Mexico (F. Luca, N. Vakhania)
- Univ. of British Columbia (W. Evans, J. Friedman)
- Univ. of Calgary (P. Høyer)
- Univ. of Québec, Montréal (P. Leroux)
- Univ. of Waterloo (A. Ambainis, M. Mosca, J. Shallit)
- Univ. of Toronto (D. Lidar)

Research groups

## Algo

Collaborations Dissemination and Technology transfer

## B.2.9 / Algo

## Dissemination and technology transfer

## Summer schools, tutorials, invited seminars

- Stéphane Boucheron gave a course on Concentration Inequalities at the Machine Learning Summer School at Max-Planck Institut Tübingen, Germany, august 2003.
- Christoph Dürr gave a course on the quantum search algorithm at the Winter School *Logique et interaction* (Logic and interaction) in the section *algorithmic complexity* at Luminy, France, january 2002.
- Christoph Dürr gave an introduction to quantum computation at the *journées informatiques X-UPS* in the Ecole Polytechnique, may 2004.
- W. Fernandez De La Vega was "Mercator" invited professor for one year in Univ. of Bonn, Germany.
- Miklos Santha gave a course on *Exact and approximate testing/correcting of algebraic functions: A survey* at the 1st Summer School on Theoretical Aspects of Computer Science, Tehran, july 2000. A corresponding survey with Marcos Kiwi and Frédéric Magniez was published in LNCS, volume 2292, pages 30-83.
- Miklos Santha gave introductory seminars to the students of the ENS Cachan in september 2001 and in september 2003 on quantum computing.
- Miklos Santha gave an introductory seminar on quantum computing for the students of the *Ecole Doctorale* at Paris-Sud in computer science, march 2003.

# Training and education (doctoral and post-doctoral)

Defended habilitations		
Name	Date defended	Current Position
DENISE Alain	10/12/01	Prof. LRI, Bioinformatics
<b>BOUCHERON</b> Stéphane	3/04/02	Prof. Paris-7
YAO Jia-Yan	26/06/03	Wuhan University, China
Defended doctorates		
MAGNIEZ Frédéric	27/01/00	CR, CNRS
CORTEEL Sylvie	31/01/00	CR, CNRS
GROSS David	12/12/00	MC, Arts et Métiers, Paris
STERN Julien	23/03/01	created a company
VERHOEVEN Yann	11/06/01	INRIA, Sophia-Antipolis
BARBAY Jérémy	24/09/02	U. of Waterloo, Canada
OLOCCO Grégory	4/04/03	Air Liquide
PEYRONNET Sylvain	17/12/03	EPITA

#### **Graduate courses**

- DEA *Modélisation stochastique* 2000-2003, Information theory, compression and coding: Stéphane Boucheron, Jean-Pierre Tillich
- DEA *Algorithmique* 2000/2001, algorithms: Jean-Pierre Tillich
- DEA *Algorithmique* 2000/2002, enumerative combinatorics and random generation: Dominique Gouyou-Beauchamps
- DEA *Algorithmique* 2001/2003, quantum computation and quantum information theory: Frédéric Magniez, Miklos Santha
- DEA Algorithmique 2002/2003, quantum computation and quantum information theory: Christoph Dürr, Julia Kempe, Frédéric Magniez, Miklos Santha

- DEA *Algorithmique* 2002/2003, approximation algorithms and property testing: Sophie Laplante, Frédéric Magniez
- DEA *Algorithmique* 2003/2004, quantum computation and quantum information theory: Christoph Dürr, Miklos Santha
- DEA *Algorithmique* 2003/2004, property testing and communication complexity: Sophie Laplante, Frédéric Magniez
- DEA *Logique* de Paris VII 2003-2005, Logic and games, Michel de Rougemont
- Master of Computer Sciences, University of Crete, 2002-2004, Computational geometry/ Cryptography: Yannis Manoussakis

## **Publications**

## International peer-reviewed journals

- B. Adamczewski. Codages de rotations et phénomènes d'autosimilarité. J. Théor. Nombres Bordeaux, 14(2):351-386, 2002.
- [2] B. Adamczewski. Balances for fixed points of primitive substitutions. *Theoret. Comput. Sci.*, 307(1):47-75, 2003. Words.
- [3] B. Adamczewski. Répartition des suites (*n*α) *n* in *N* et substitutions. *Acta Arith.*, 112(1):1-22, 2004.
- [4] B. Adamczewski. Transcendance "à la Liouville" de certains nombres réels. *C. R. Math. Acad. Sci. Paris*, 338(7):511-514, 2004.
- [5] B. Adamczewski, Y. Bugeaud, and F. Luca. Sur la complexité des nombres algébriques. *Comptes Rendus de l'Académie des Sciences, Serie I*, 339(11-14), 2004.
- [6] B. Adamczewski and J. Cassaigne. On the transcendence of real numbers with a regular expansion. *J. Number Theory*, 103(1):27-37, 2003.
- [7] B. Adamczewski and D. Damanik. Linearly recurrent circle map subshifts and an application to Schrödinger operators. Ann. Henri Poincaré, 3(5):1019-1047, 2002.
- [8] F. Afrati, E. Bampis, C. Kenyon, and I. Milis. A PTAS for the average weighted completion time problem on unrelated machines. *J. Sched.*, 3(6):323-332, 2000. Approximation algorithms, Part 2.
- [9] F. Afrati, I. Guessarian, and M. de Rougemont. The expressiveness of DAC. *Theoret. Comput. Sci.*, 286(1):3-32, 2002. Mathematical foundations of computer science (Bratislava, 1997).
- [10] F. Afrati, H. Leiß, and M. de Rougemont. Definability and compression. *Fund. Inform.*, 56(1-2):155-180, 2003. Special issue on computing patterns in strings.
- [11] J.-P. Allouche. Nouveaux résultats de transcendance de réels à développement non aléatoire. *Gaz. Math.*, 84:19-34, 2000.
- [12] J.-P. Allouche, M. Baake, J. Cassaigne, and D. Damanik. Palindrome complexity. *Theoret. Comput. Sci.*, 292(1):9-31, 2003. Selected papers in honor of Jean Berstel.
- [13] J.-P. Allouche and M. Cosnard. The Komornik-Loreti constant is transcendental. *Amer. Math. Monthly*, 107(5):448-449, 2000.
- [14] J.-P. Allouche and M. Cosnard. Non-integer bases, iteration of continuous real maps, and an arithmetic self-similar set. *Acta Math. Hungar.*, 91(4):325-332, 2001.
- [15] J.-P. Allouche, M. Courbage, and G. Skordev. Notes on cellular automata. *Cubo Mat. Educ.*, 3(2):213-244, 2001.
- [16] J.-P. Allouche, J. L. Davison, M. Queffélec, and L. Q. Zamboni. Transcendence of Sturmian or morphic continued fractions. J. Number Theory, 91(1):39-66, 2001.
- [17] J.-P. Allouche, J.-M. Deshouillers, T. Kamae, and T. Koyanagi. Automata, algebraicity and distribution of sequences of powers. *Ann. Inst. Fourier (Grenoble)*, 51(3):687-705, 2001.
- [18] J.-P. Allouche, M. Mendès France, and J. Peyrière. Automatic Dirichlet series. J. Number Theory, 81(2):359-373, 2000.
- [19] J.-P. Allouche, N. Rampersad, and J. Shallit. On integer sequences whose first iterates are linear. *Aequ. Math.*, 2004. To appear.

Research groups

Algo

Training and education Publications

- [20] J.-P. Allouche, K. Scheicher, and R. F. Tichy. Regular maps in generalized number systems. *Math. Slovaca*, 50(1):41-58, 2000.
- [21] J.-P. Allouche and J. Shallit. Sums of digits, overlaps, and palindromes. *Discrete Math. Theor. Comput. Sci.*, 4(1):1-10 (electronic), 2000.
- [22] J.-P. Allouche and J. Shallit. The ring of *k*-regular sequences. II. *Words*, 307(1):3-29, 2003, Elsevier Science Publishers.
- [23] J.-P. Allouche and G. Skordev. Schur congruences, Carlitz sequences of polynomials and automaticity. *Discrete Math.*, 214(1-3):21-49, 2000.
- [24] J.-P. Allouche and G. Skordev. Remarks on permutive cellular automata. *J. Comput. System Sci.*, 67(1):174-182, 2003.
- [25] N. Alon, W. F. de la Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of max-csp problems. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, pages 232-239, 2002.
- [26] N. Alon, W. F. de la Vega, R. Kannan, and M. Karpinski. Random sampling and approximation of MAX-CSPs. *Journal of Computer and System Sciences*, 67(2):212-243, 2003, Elsevier Science Publishers. Special issue on STOC2002 (Montreal, QC).
- [27] N. Alon, H. Kaplan, M. Krivelevich, D. Malkhi, and J. P. Stern. Scalable secure storage when half the system is faulty. *Inf. Comput.*, 174(2):203-213, 2002.
- [28] A. K. Amoura, E. Bampis, C. Kenyon, and Y. Manoussakis. Scheduling independent multiprocessor tasks. *Algorithmica*, 32(2):247-261, 2002, Springer-Verlag.
- [29] E. Bampis, A. Giannakos, A. Karzanov, Y. Manoussakis, and I. Milis. Perfect matching in general vs. cubic graphs: a note on the planar and bipartite cases. *Theor. Inform. Appl.*, 34(2):87-97, 2000.
- [30] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, and D. Gouyou-Beauchamps. Generating functions for generating trees. *Discrete Math.*, 246(1-3):29-55, 2002. Formal power series and algebraic combinatorics (Barcelona, 1999).
- [31] P. Baptiste, M. Chrobak, C. Dürr, W. Jawor, and N. Vakhania. Preemptive scheduling of equal-length jobs to maximize weighted throughput. *Oper. Res. Lett.*, 32(3):258-264, 2004.
- [32] D. Barraez, S. Boucheron, and W. Fernandez de la Vega. On the fluctuations of the giant component. *Combin. Probab. Comput.*, 9(4):287-304, 2000.
- [33] P. Bartlett, S. Boucheron, and G. Lugosi. Model selection and error estimation. *Machine Learning*, 48:85-113, 2002.
- [34] C. Bazgan, W. Fernandez de la Vega, and M. Karpinski. Polynomial time approximation schemes for dense instances of minimum constraint satisfaction. *Random Structures Algorithms*, 23(1):73-91, 2003.
- [35] C. Bazgan, M. Santha, and Z. Tuza. Efficient approximation algorithms for the subsetsums equality problem. J. Comput. System Sci., 64(2):160-170, 2002.
- [36] A. Benkouar, Y. Manoussakis, and R. Saad. The number of edge-colored complete graphs with unique alternating hamiltonian cycles. *Discrete Mathematics*, 263:1-10, 2003, Elsevier Science Publishers.
- [37] A. Berthiaume, W. van Dam, and S. Laplante. Quantum Kolmogorov complexity. *Journal of Computer and System Sciences*, 63(2):201-221, 2001, Elsevier Science Publishers. Special Issue on Complexity 2000.
- [38] S. Boucheron, O. Bousquet, G. Lugosi, and P. Massart. Moment inequalities for functions of independent random variables. *Annals of Probability*, 2004.
- [**39**] S. Boucheron and W. Fernandez de la Vega. On the independence number of random interval graphs. *Combin. Probab. Comput.*, 10(5):385-396, 2001.
- [40] S. Boucheron and W. Fernandez de la Vega. On a square packing problem. *Combin. Probab. Comput.*, 11(2):113-127, 2002.

- [41] S. Boucheron, F. Gamboa, and C. Léonard. Bins and balls: large deviations of the empirical occupancy process. Ann. Appl. Probab., 12(2):607-636, 2002.
- [42] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Random Structures Algorithms*, 16(3):277-292, 2000.
- [43] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583-1614, 2003.
- [44] S. Boucheron and M. R. Salamatian. About priority encoding transmission. *IEEE Trans. Inform. Theory*, 46(2):699-705, 2000.
- [45] H. Buhrman, C. Dürr, M. Heiligman, P. Høyer, F. Magniez, M. Santha, and R. de Wolf. Quantum algorithms for element distinctness. *SIAM J. Comput.*, inpress.
- [46] H. Buhrman, L. Fortnow, and S. Laplante. Resource-bounded Kolmogorov complexity revisited. *SIAM J. Comput.*, 31(3):887-905 (electronic), 2001/02.
- [47] T. Calamoneri, I. Finocchi, Y. Manoussakis, and R. Petreschi. On Max Cut in cubic graphs. *Parallel Algorithms Appl.*, 17(3):165-183, 2002.
- [48] M. Chrobak, P. Couperus, C. Dürr, and G. Woeginger. On tiling under tomographic constraints. *Theoret. Comput. Sci.*, 290(3):2125-2136, 2003.
- [49] M. Chrobak and C. Dürr. Reconstructing polyatomic structures from discrete X-rays: NP-completeness proof for three atoms. *Theoret. Comput. Sci.*, 259(1-2):81-98, 2001.
- [50] S. Corteel, A. Denise, and D. Gouyou-Beauchamps. Bijections for directed animals on infinite families of lattices. *Ann. Comb.*, 4(3-4):269-284, 2000. Conference on Combinatorics and Physics (Los Alamos, NM, 1998).
- [51] S. Corteel and D. Gouyou-Beauchamps. Enumeration of sand piles. *Discrete Math.*, 256(3):625-643, 2002. LaCIM 2000 Conference on Combinatorics, Computer Science and Applications (Montreal, QC).
- [52] C. Dürr, I. Rapaport, and G. Theyssier. Cellular automata and communication complexity. *Theoretical Computer Science*, 322:355-368, 2004, Elsevier Science Publishers.
- [53] J.-C. Dubacq, B. Durand, and E. Formenti. Kolmogorov complexity and cellular automata classification. *Theoret. Comput. Sci.*, 259(1-2):271-285, 2001.
- [54] C. Dürr, E. Goles, I. Rapaport, and E. Rémila. Tiling with bars under tomographic constraints. *Theoret. Comput. Sci.*, 290(3):1317-1329, 2003.
- [55] C. Dürr and M. Santha. A decision procedure for unitary linear quantum cellular automata. *SIAM J. Comput.*, 31(4):1076-1089 (electronic), 2002.
- [56] M. El Haddad, Y. Manoussakis, and R. Saad. Upper bounds for the forwarding indices of communication networks. *Discrete Mathematics*, inpress, Elsevier Science Publishers.
- [57] W. Evans, C. Kenyon, Y. Peres, and L. J. Schulman. Broadcasting on trees and the Ising model. *Ann. Appl. Probab.*, 10(2):410-433, 2000.
- [58] W. Fernandez de la Vega. Random 2-SAT: results and problems. *Theoret. Comput. Sci.*, 265(1-2):131-146, 2001. Phase transitions in combinatorial problems (Trieste, 1999).
- [59] W. Fernandez de la Vega and M. Karpinski. Polynomial time approximation of dense weighted instances of MAX-CUT. *Random Structures Algorithms*, 16(4):314-332, 2000.
- [60] W. Fernandez de la Vega and C. Kenyon. A randomized approximation scheme for metric MAX-CUT. *J. Comput. System Sci.*, 63(4):531-541, 2001. Special issue on FOCS 98 (Palo Alto, CA).
- [61] W. Fernandez de la Vega and M. Lamari. The task allocation problem with constant communication. *Discrete Appl. Math.*, 131(1):169-177, 2003. The Second International Colloquium "Journées de l'Informatique Messine" (Metz, 2000).
- [62] K. Friedl, F. Magniez, M. Santha, and P. Sen. Quantum testers for hidden group properties. *Theoretical Computer Science, Special Issue on the 28th Symposium on Mathematical Foundations of Computer Science*, in press.

Research groups

Publications

- [63] J. Friedman and J.-P. Tillich. Laplacian eigenvalues and distances between subsets of a manifold. J. Differential Geom., 56(2):285-299, 2000.
- [64] E. Gassiat and S. Boucheron. Optimal error exponents in hidden Markov models order estimation. *IEEE Trans. Inform. Theory*, 49(4):964-980, 2003.
- [65] D. Gouyou-Beauchamps and P. Leroux. Enumeration of symmetry classes of convex polyominoes on the honeycomb lattice. *Theoretical Computer Science*, inpress, Elsevier Science Publishers.
- [66] M. E. Haddad, Y. Manoussakis, and R. Saad. Upper bounds for the forwarding indices of communication networks. *Discrete Mathematics*, inpress, Elsevier Science Publishers.
- [67] M. Hsieh, J. Kempe, S. Myrgren, and K. B. Whaley. An explicit universal gate-set for exchange-only quantum computation. *Quantum Information Processing*, 2(4):289-307, 2003.
- [68] G. Ivanyos, F. Magniez, and M. Santha. Efficient quantum algorithms for some instances of the non-abelian hidden subgroup problem. *Internat. J. Found. Comput. Sci.*, 14(5):723-739, 2003. Quantum computing.
- [69] J. Kempe. Quantum random walks an introductory overview. *Contemporary Physics*, 44(4):302-327, 2003. lanl-arXiv quant-ph/0303081.
- [70] J. Kempe and O. Regev. 3-local Hamiltonian is QMA-complete. *Quantum Information and Computation*, 3(3):258-264, 2003.
- [71] J. Kempe and K. Whaley. Exact gate-sequences for universal quantum computation using the XY-interaction alone. *Phys. Rev. A*, 65 (5):052330, 2002.
- [72] C. Kenyon. The evolution of Web-caching markets. Computer, 34(11):128-130, Nov. 2001.
- [73] C. Kenyon and G. Cheliotis. Stochastic models for telecom commodity prices. Computer Networks (Amsterdam, Netherlands: 1999), 36(5-6):533-555, Aug. 2001.
- [74] C. Kenyon and M. Mitzenmacher. Linear waste of best fit bin packing on skewed distributions. *Random Structures and Algorithms*, 20(3):441-464, 2002.
- [75] C. Kenyon and E. Rémila. A near-optimal solution to a two-dimensional cutting stock problem. *Math. Oper. Res.*, 25(4):645-656, 2000.
- [76] C. Kenyon and S. Tompaidis. Options in leasing: the effect of idle time. *Operations Research*, 49(5):675-689, 2001.
- [77] M. Kiwi, F. Magniez, and M. Santha. Approximate testing with error relative to input size. *J. Comput. System Sci.*, 66(2):371-392, 2003.
- [78] F. Magniez. Multi-linearity self-testing with relative error. *Theory of Computing Systems*, 2004. inpress.
- [79] Y. Manoussakis and H. Patil. Bipartite graphs and their degree sets. 15, 2003, Elsevier.
- [80] Y. Manoussakis, H. Patil, and V. Sankar. Further results on degree sets for graphs. 1, 2001.
- [81] Y. Manoussakis and Z. Tuza. Ramsey numbers for tournaments. *Theoret. Comput. Sci.*, 263(1-2):75-85, 2001. Combinatorics and computer science (Palaiseau, 1997).
- [82] P. Milman and R. Mosseri. Topological phase for entangled two-qubit states. *Phys. Rev. Lett.*, 90(23):230403, 4, 2003.
- [83] P. Milman, H. Ollivier, F. Yamaguchi, M. Brune, J. M. Raimond, and S. Haroche. Simple quantum information algorithms in cavity QED. *J. Modern Opt.*, 50(6-7):901-913, 2003. International Conference on Quantum Information, Conceptual Foundations, Developments and Perspectives (Oviedo, 2002).
- [84] F. Noilhan and M. Santha. Semantical counting circuits. *Theory Comput. Syst.*, 36(3):217-229, 2003.
- [85] H. Ollivier and P. Milman. Proposal for realization of a toffoli gate via cavity-assisted atomic collision. *Quantum Information & Computation*, 43:603, 2003.
- [86] J. Radhakrishnan, P. Sen, and S. Venkatesh. The quantum complexity of set membership. *Algorithmica*, 34(4):462-479, 2002, Springer-Verlag.

- [87] N. Shenvi, J. Kempe, and K. Whaley. A quantum random walk search algorithm. *Phys. Rev. A*, 67(5):052307, 2003. lanl-arXiv quant-ph/0210064.
- [88] C. Simon and J. Kempe. Robustness of multiparty entanglement. *Phys. Rev. A*, 65 (5):052327, 2002.
- [89] E. Solano, M. França Santos, and P. Milman. Quantum phase gate with a selective interaction. *Phys. Rev. A* (3), 64(2):024304, 4, 2001.
- [90] J.-P. Tillich. Edge isoperimetric inequalities for product graphs. *Discrete Math.*, 213(1-3):291-320, 2000. Selected topics in discrete mathematics (Warsaw, 1996).
- [91] J.-P. Tillich and G. Zémor. Discrete isoperimetric inequalities and the probability of a decoding error. *Combin. Probab. Comput.*, 9(5):465-479, 2000.

## **Books**

- [92] J.-P. Allouche and V. Berthé. Some applications of combinatorics on words in number theory in "Applied Combinatorics on Words", Lothaire. Cambridge University Press, Cambridge, 2005. à paraître.
- [93] J.-P. Allouche and J. Shallit. *Automatic sequences*. Cambridge University Press, Cambridge, 2003. Theory, applications, generalizations.
- [94] R. Lassaigne and M. de Rougemont. *Logic and complexity*. Discrete Mathematics and Theoretical Computer Science (London). Springer-Verlag London Ltd., London, 2004.

## **Book chapters**

- [95] M. de Rougemont. Sécurité des services informatiques. In *Le pilotage du changement par les cybertechnologies*. Hermès, 2003.
- [96] M. de Rougemont. Logic, randomness and cognition. In *Logic, Thought and Action*. D. Vanderveken, Kluwer Academic Publisher, 2004.
- [97] M. Kiwi, F. Magniez, and M. Santha. Exact and approximate testing/correcting of algebraic functions: a survey. In *Theoretical aspects of computer science (Tehran, 2000)*, volume 2292 of *Lecture Notes in Comput. Sci.*, pages 30-83. Springer, Berlin, 2002.
- [98] H. Leiß and M. de Rougemont. Automata on Lempel-Ziv compressed strings. In Computer science logic, volume 2803 of Lecture Notes in Comput. Sci., pages 384-396. Springer, Berlin, 2003.

## Major international peer-reviewed conferences

- [99] F. Afrati, E. Bampis, A. V. Fishkin, K. Jansen, and C. Kenyon. Scheduling to minimize the average completion time of dedicated tasks. In *FST TCS 2000: Foundations of software technology and theoretical computer science (New Delhi)*, volume 1974 of *Lecture Notes in Comput. Sci.*, pages 454-464. Springer, Berlin, 2000.
- [100] F. Afrati, H. Leiß, and M. de Rougemont. Definability and compression. In 15th Annual IEEE Symposium on Logic in Computer Science (Santa Barbara, CA, 2000), pages 63-73. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.
- [101] D. Aharonov, A. Ambainis, J. Kempe, and U. V. Vazirani. Quantum walks on graphs. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, pages 50-59, 2001.
- [102] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev. Adiabatic quan-

Research groups

#### Algo

Publications

tum computation is equivalent to standard quantum computation. In *Proc.* 45<sup>th</sup> Annual *IEEE Symp. on Foundations of Computer Science (FOCS)*, 2004.

- [103] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev. Adiabatic quantum computation is equivalent to standard quantum computation. In *Proceedings of the* 45th Annual IEEE Symp. on Foundations of Computer Science (FOCS), 2004.
- [104] J.-P. Allouche. Algebraic and analytic randomness. In Noise, oscillators and algebraic randomness (Chapelle des Bois, 1999), volume 550 of Lecture Notes in Phys., pages 345-356. Springer, Berlin, 2000.
- [105] K. Artiouchine, P. Baptiste, and C. Dürr. Runway scheduling with holding loop. In *Discrete Optimization Methods in Production and Logistics (DOM)*, pages 96-101, Omsk-Irkutsk, Russia, 2004.
- [106] J. Barbay and C. Kenyon. On the discrete Bak-Sneppen model of self-organized criticality. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA-01), pages 928-933, New York, Jan. 7-9 2001. ACM Press.
- [107] J. Barbay and C. Kenyon. Adaptive intersection and t-Threshold problems. In *Proceedings of the 13th Annual ACM-SIAM Symposium On Discrete Mathematics (SODA-02)*, pages 390-399, New York, Jan. 6-8 2002. ACM Press.
- [108] D. Barth, S. Corteel, A. Denise, D. Gardy, and M. Valencia-Pabon. On the complexity of routing permutations on trees by arc-disjoint paths. In *Proceedings of Latin American Theoretical INformatics (LATIN)*, volume 1776 of *Lecture Notes in Computer Science*, pages 308-317, Punta del Este, 2000. Springer.
- [109] C. Bazgan, W. F. de la Vega, and M. Karpinski. Approximability of dense instances of nearest codeword problem. In M. Penttonen and E. M. Schmidt, editors, *SWAT*, volume 2368 of *Lecture Notes in Computer Science*, pages 298-307. Springer, 2002.
- [110] A. Berthiaume, W. van Dam, and S. Laplante. Quantum Kolmogorov complexity. In 15<sup>th</sup> Annual IEEE Conference on Computational Complexity (Florence, 2000), pages 240-249. IEEE Computer Soc., Los Alamitos, CA, 2000.
- [111] S. L. Bezrukov, R. Elsässer, B. Monien, R. Preis, and J.-P. Tillich. New spectral lower bounds on the bisection width of graphs. In *Graph-theoretic concepts in computer science (Konstanz,* 2000), volume 1928 of *Lecture Notes in Comput. Sci.*, pages 23-34. Springer, Berlin, 2000.
- [112] S. L. Bezrukov, R. Elsässer, B. Monien, R. Preis, and J.-P. Tillich. New spectral lower bounds on the bisection width. In U. Brandes and D. Wagner, editors, *Graph-Theoretic Concepts in Computer Science, 26th International Workshop, WG 2000, Konstanz, Germany, June 15-17, 2000, Proceedings*, volume 1928 of *Lecture Notes in Computer Science*, pages 155-174. Springer, 2000.
- [113] J. Boeuf and J. Stern. An analysis of one of the SDMI candidates. In I. S. Moskowitz, editor, *Information Hiding*, volume 2137 of *Lecture Notes in Computer Science*, pages 395-410. Springer, 2001.
- [114] U. Boobna and M. de Rougemont. Correctors for XML data. In Z. Bellahsene, T. Milo, M. Rys, D. Suciu, and R. Unland, editors, *Database and XML Technologies, Second International XML Database Symposium, XSym 2004, Toronto, Canada, August 29-30*, volume 3186 of *Lecture Notes in Computer Science*, pages 97-111. Springer, 2004.
- [115] H. Buhrman, C. Dürr, M. Heiligman, P. Høyer, F. Magniez, M. Santha, and R. de Wolf. Quantum algorithms for element distinctness. In *IEEE Conference on Computational Complexity*, pages 131-137, 2001.
- [116] H. Buhrman, S. Laplante, and P. B. Miltersen. New bounds for the language compression problem. In 15th Annual IEEE Conference on Computational Complexity (Florence, 2000), pages 126-130. IEEE Computer Soc., Los Alamitos, CA, 2000.
- [117] E. Cadic, J.-C. Carlach, G. Olocco, A. Otmani, and J.-P. Tillich. Low complexity tailbiting trellises of self-dual codes of length 24, 32 and 40 over GF(2) and *z*<sub>4</sub> of large

minimum distance. In S. Boztas and I. Shparlinski, editors, *AAECC*, volume 2227 of *Lecture Notes in Computer Science*, pages 57-66. Springer, 2001.

- [118] C. Carlet and E. Prouff. On plateaued functions and their constructions. In T. Johansson, editor, *FSE*, volume 2887 of *Lecture Notes in Computer Science*, pages 54-73. Springer, 2003.
- [119] C. Carlet and E. Prouff. On a new notion of nonlinearity relevant to multi-output pseudorandom generators. In M. Matsui and R. J. Zuccherato, editors, *Selected Areas in Cryptography*, volume 3006 of *Lecture Notes in Computer Science*, pages 291-305. Springer, 2004.
- [120] C. Carlet and E. Prouff. Vectorial functions and covering sequences. In G. L. Mullen, A. Poli, and H. Stichtenoth, editors, *International Conference on Finite Fields and Applications*, volume 2948 of *Lecture Notes in Computer Science*, pages 215-248. Springer, 2004.
- [121] J. Csirik, D. S. Johnson, and C. Kenyon. Better approximation algorithms for bin covering. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA-01), pages 557-566, New York, Jan. 7-9 2001. ACM Press.
- [122] J. Csirik, D. S. Johnson, C. Kenyon, J. B. Orlin, P. W. Shor, and R. R. Weber. On the sumof-squares algorithm for bin packing. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (STOC)*, pages 208-217, 2000.
- [123] W. F. de la Vega, M. Karpinski, and C. Kenyon. Approximation schemes for metric bisection and partitioning. In J. I. Munro, editor, SODA, pages 506-515. SIAM, 2004.
- [124] W. F. de la Vega, M. Karpinski, C. Kenyon, and Y. Rabani. Approximation schemes for clustering problems. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pages 50-58, 2003.
- [125] J.-C. Dubacq and V. Terrier. Signals for cellular automata in dimension 2 or higher. In LATIN 2002: Theoretical informatics (Cancun), volume 2286 of Lecture Notes in Comput. Sci., pages 451-464. Springer, Berlin, 2002.
- [126] C. Dürr, M. Heiligman, P. Høyer, and M. Mhalla. Quantum query complexity of some graph problems. In *Proceedings of the 31st International Colloquium on Automata, Languages and Programming (ICALP)*, pages 481-493, 2004.
- [127] W. Fernandez de la Vega. The independence number of random interval graphs. In *Algorithms and complexity (Rome, 2000)*, volume 1767 of *Lecture Notes in Comput. Sci.*, pages 59-62. Springer, Berlin, 2000.
- [128] K. Friedl, G. Ivanyos, F. Magniez, M. Santha, and P. Sen. Hidden translation and orbit coset in quantum computing. In *Proceedings of the 35th Annual ACM Symposium on Theory* of Computing (STOC), pages 1-9, 2003.
- [129] K. Friedl, F. Magniez, M. Santha, and P. Sen. Quantum testers for hidden group properties. In LNCS, editor, *Proceedings of the 28th International Symposium on Mathematical Foundations of Computer Science (MFCS)*, volume 2747, pages 419-428, 2003.
- [130] M. Golin, C. Kenyon, and N. E. Young. Huffman coding with unequal letter costs. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, pages 785-791, May 2002.
- [131] S.-D. Gouraud, A. Denise, M.-C. Gaudel, and B. Marre. A new way for automating statistical testing methods. In *Proceedings of IEEE Automated Software Engineering (ASE)*, San Diego, november 2001.
- [132] D. Gouyou-Beauchamps and P. Leroux. Dénombrement des classes de symétries des polyominos hexagonaux convexes. In 16th Annual International Conference on Formal Power Series and Algebraic Combinatorics (FPSAC04), June 28 - July 2 2004.
- [133] D. Gross and M. de Rougemont. Uniform generation in spatial constraint databases and applications. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS)*, pages 254-259, 2000.
- [134] T. Hérault, R. Lassaigne, F. Magniette, and S. Peyronnet. Approximate probabilistic model checking. In B. Steffen and G. Levi, editors, VMCAI, volume 2937 of Lecture Notes in Computer Science, pages 73-84. Springer, 2004.

Research groups Algo

Publications

- [135] R. Jain, J. Radhakrishnan, and P. Sen. Privacy and interaction in quantum communication complexity and a theorem about the relative entropy of quantum states. In *Proc.* 43<sup>rd</sup> Annual IEEE Symp. on Foundations of Computer Science (FOCS), pages 429-438, 2002.
- [136] A. R. Karlin, C. Kenyon, and D. Randall. Dynamic TCP acknoledgement and other stories about e/(e-1). In Proceedings of the ACM Symposium on Theory of Computing (STOC), pages 502-509, 2001.
- [137] J. Kempe. Quantum walks hit exponentially faster. In RANDOM-APPROX 2003, Lecture Notes in Computer Science, pages 354-369, Heidelberg, 2003. Springer.
- [138] J. Kempe, A. Kitaev, and O. Regev. The complexity of the local hamiltonian problem. In *Proc. of Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*, Lecture Notes in Computer Science. Springer-Verlag, 2004. to appear.
- [139] J. Kempe, A. Kitaev, and O. Regev. The complexity of the local hamiltonian problem. In Proceedings of the 24th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS), 2004.
- [140] C. Kenyon and M. Mitzenmacher. Linear waste of best fit bin packing on skewed distributions. In 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), pages 582-589. IEEE Comput. Soc. Press, Los Alamitos, CA, 2000.
- [141] C. Kenyon, E. Mossel, and Y. Peres. Glauber dynamics on trees and hyperbolic graphs. In *FOCS*, pages 568-578, 2001.
- [142] C. Kenyon, N. Schabanel, and N. E. Young. Polynomial-time approximation scheme for data broadcast. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (STOC)*, pages 659-666, 2000.
- [143] M. A. Kiwi, F. Magniez, and M. Santha. Exact and approximate testing/correcting of algebraic functions: A survey. In G. B. Khosrovshahi, A. Shokoufandeh, and M. A. Shokrollahi, editors, *Theoretical Aspects of Computer Science 2000*, volume 2292 of *Lecture Notes in Computer Science*, pages 30-83. Springer, 2000.
- [144] S. Laplante, R. Lassaigne, F. Magniez, S. Peyronnet, and M. Rougemont. Probabilistic abstraction for model checking: An approach based on property testing. In *Proceedings* of 17th IEEE Symposium on Logic in Computer Science, pages 30-39, 2002.
- [145] S. Laplante and F. Magniez. Lower bounds for randomized and quantum query complexity using Kolmogorov arguments. In *Proceedings of 19th IEEE Conference on Computational Complexity*, 2004, pages 294-304.
- [146] R. Lassaigne and S. Peyronnet. Approximate verification of probabilistic systems. In H. Hermanns and R. Segala, editors, *PAPM-PROBMIV*, volume 2399 of *Lecture Notes in Computer Science*, pages 213-214. Springer, 2002.
- [147] H. Leiss and M. de Rougemont. Automata on lempel-ziv structures. In M. Baaz and J. A. Makowsky, editors, Computer Science Logic, 17<sup>th</sup> International Workshop, CSL 2003, 12<sup>th</sup> Annual Conference of the EACSL, and 8<sup>th</sup> Kurt Güdel Colloquium, KGC 2003, Vienna, Austria, August 25-30, 2003, Proceedings, volume 2803 of Lecture Notes in Computer Science, Vienne, 2003. Springer.
- [148] F. Magniez. Multi-linearity self-testing with relative error. In STACS 2000 (Lille), volume 1770 of Lecture Notes in Comput. Sci., pages 302-313. Springer, Berlin, 2000.
- [149] F. Magniez and M. Rougemont. Property testing of regular tree languages. In *Proceedings* of 31st International Colloquium on Automata, Languages and Programming, volume 3142 of Lecture Notes in Comput. Sci., pages 932-944. Springer, 2004.
- [150] F. Noilhan and M. Santha. Semantical counting circuits. In *Algorithms and complexity* (*Rome, 2000*), volume 1767 of *Lecture Notes in Comput. Sci.*, pages 87-101. Springer, Berlin, 2000.
- [151] G. Olocco and A. Otmani. Low complexity tail-biting trellises for some extremal self-dual codes. In *Eight International Workshop On Algebraic and Combinatorial Coding Theory*, 2002.
- [152] G. Olocco and J.-P. Tillich. Iterative decoding of a new family of block turbo-codes. In Proceedings of 2<sup>nd</sup> International Symposium on Turbo Codes and Related Topics, Brest, pages 302-306, 2000.
- [153] G. Olocco and J.-P. Tillich. A family of self-dual codes which behave in many respects like random linear codes of rate 1/2. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT)*, page 15, 2001.
- [154] M. Santha and M. Szegedy. Quantum and classical query complexities of local search are polynomially related. In *Proceedings of the Annual ACM Symposium on Theory of Computing (STOC)*, pages 494-501, 2004.
- [155] J. Stern and J. P. Stern. Cryptanalysis of the otm signature scheme from fc'02. In *Financial Cryptography*, pages 138-148, 2003.
- [156] J. P. Stern and J.-P. Tillich. Automatic detection of a watermarked document using a private key. In I. S. Moskowitz, editor, *Information Hiding*, volume 2137 of *Lecture Notes in Computer Science*, pages 258-272. Springer, 2001.
- [157] W. van Dam, F. Magniez, M. Mosca, and M. Santha. Self-testing of universal and faulttolerant sets of quantum gates. In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (STOC)*, pages 688-696, 2000.

## **Doctoral dissertations and Habilitations**

- [158] J. Barbay. Analyse fine: bornes inférieures et algorithmes de calculs d'intersection pour moteurs de recherche. Ph.D. thesis, Université Paris-Sud, 2002.
- [159] S. Boucheron. *Problèmes d'informatique, techniques probabilistes*. Habilitation, Université Paris-Sud, 2002.
- [160] S. Corteel. *Problèmes énumératifs issus de l'informatique, de la physique et de la combinatoire*. Ph.D. thesis, Université Paris-Sud, 2000.
- [161] A. Denise. *Structures aléatoires, modèles et analyse des génomes*. Habilitation, Université Paris-Sud, 2001.
- [162] D. Gross. *Approximation dans les bases de données contraintes*. Ph.D. thesis, Université Paris-Sud, 2000.
- [163] F. Magniez. *Auto-test pour les calculs approché et quantique*. Ph.D. thesis, Université Paris-Sud, 2000.
- [164] S. Peyronnet. Model checking et vérification probabiliste. Ph.D. thesis, Université Paris-Sud, 2003.
- [165] J. Stern. *Contribution à une théorie de la protection de l'information*. Ph.D. thesis, Université Paris-Sud, 2001.
- [166] Y. Verhoeven. *Quelques utilisations des arbres en combinatoire*. Ph.D. thesis, Université Paris-Sud, 2001.
- [167] J.-Y. Yao. *Théorie des automates finis et applications*. Habilitation, Université Paris-Sud, 2003.

Research groups

Publications



Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Programmation et génie logiciel

**Responsable: Marie-Claude Gaudel** 

L'équipe Programmation et Génie Logiciel est composée de trois groupes :

• Le groupe Génie Logiciel est connu depuis longtemps pour ses travaux sur le rôle des méthodes formelles dans le développement de logiciel. Actuellement, l'essentiel de ses activités concerne le test basé sur différentes méthodes de spécification formelle, avec une nouvelle orientation vers les méthodes probabilistes et statistiques.

• Le groupe Interaction Homme-Machine travaille sur plusieurs sujets liés aux systèmes interactifs, parmi lesquels les nouvelles techniques d'interaction, les systèmes de travail collaboratif, la visualisation d'informations complexes, la conception participative et la réalité augmentée.

• Le groupe Réseaux Mobiles est récent. Il a été créé en septembre 2000 avec l'objectif de développer au LRI une compétence sur les systèmes mobiles et sans fil. Les travaux de ce groupe portent sur les systèmes cellulaires (UMTS), la mobilité sur internet, les réseaux adhoc et les réseaux mobiles.

La synergie entre ces trois groupes est assurée par de nombreuses réunions d'équipes et plusieurs activités communes, comme la visualisation de logiciels complexes, le test de protocoles, l'application de techniques de travail collaboratif à la composition robuste de services Web. Durant ces quatre dernières années, plusieurs événements ont animé la vie de l'équipe et ont contribué à faire évoluer son activité.

Tout d'abord le départ de Bruno Marre vers le CEA, où il conduit le développement d'outils de test pour la certification de logiciels critiques : ce départ a été compensé par le recrutement de deux jeunes maîtres de conférences qui s'intéressent respectivement au test de propriétés temporelles et au test d'intégration. Ensuite,

la mise en place du projet PCRI In Situ (Interaction Située), en



commun avec l'INRIA, ce qui a permis un fort développement des activités en Interaction Homme-machine. Enfin la création du groupe Réseaux, fort maintenant d'un jeune pro-

fesseur, d'une maître de conférences et de 4 doctorants, qui a vocation à continuer à se renforcer.





Research groups
Prog
Introduction

# Programming and Software Engineering

## Head: Marie-Claude Gaudel

The Programming and Software Engineering group is organized in three research teams:

• The Software Engineering team has been known for years for its work on the use of formal methods in software development. Currently, there is a focus on testing based on various kinds of formal specifications with a new orientation towards probabilistic and statistical methods.

• The Human-Computer Interaction team addresses various topics related to interactive systems, including novel interaction techniques, computer-supported cooperative work, information visualization, participatory design and augmented (or mixed) reality.

• The Mobile Network team was created in September 2000 with the goal to develop at LRI some research in wireless and mobile systems. The topics of the team are cellular systems (UMTS), mobility in IP networks, ad hoc networks and network mobility.

The cross-fertilisation among these teams is ensured by a group seminar and by several common activities, such as the visualisation of complex programs or specifications, the test of communication protocols, the use of techniques from computer-supported cooperative work for composing Web services in a robust way. Over the last four years, several important changes have occurred in our group: Bruno Marre, a former CNRS researcher from the Software Engineering team, has left LRI for CEA-LIST where he leads the development of testing tools used for the certification of critical systems. Two assistant professors have been recruited recently to reinforce the team.

A joint project was established between the Human-Computer Interaction team and INRIA Futurs within the framework of PCRI (*Pôle Commun de Recherche en Informatique*). This new project, called In Situ (*Interaction Située* or Situated Interaction) has grown from 5 members in 2002 to about 20 by the end of 2004. As mentioned above, the



Mobile Network team was created, and is now very active, with one professor, one assistant professor and numerous PhD students. However, there it needs further recruiting in order to reach critical mass.

Laboratoire de Recherche en Informatique / 2000-2004 report

# Research group members

## Personnel as of 01/01/2004

Full time faculty			
Name	First Name	Position*	Institution
AL AGHA	Khaldoun	MC	Université Paris XI
<b>BEAUDOUIN-LAFON</b>	Michel	PR1	Université Paris XI
BOUKHATEM	Lila	MC	Université Paris XI
GAUDEL	Marie-Claude	PRCE	Université Paris XI
PETITJEAN	Eric	MC	IUT d'Orsay
ROUSSEL	Nicolas	MC	Université Paris XI
SCHLIENGER	Françoise	MC	IUT d'Orsay
VOISIN	Frédéric	MC	Université Paris XI
ZAIDI	Fatiha	MC	Université Paris XI

Doctoral students			
APPERT	Caroline	AM	Université Paris XI
BADIS	Hakim	D	INRIA
BEAUDOUX	Olivier	D	AETS-ESEO
BEN HAMIDA	Ines	AM	Université Paris XI
BENZAID	Mounir	ATER	IUT Marne la Vallée
BLANCH	Renaud	AM	Université Paris XI
GOURAUD	Sandrine	ATER	Université Paris XI
LARD	Jérôme	D	THALES
LESTIENNES	Grégory	AM	Université Paris XI
MALES	Davor	D	Université Paris VI
NARS	Emmanuel	A	Université Paris XI

Temporary personnel					
AL-MESHHADANY	Thamer	ATER	Université Paris XI		
PERATO	Laurent	ATER	Université Paris XI		
COURTOIS	Jean-René	Engineer	Université Paris XI		
FOURATI	Alia	Engineer	Université Paris XI		
LANGET	Matthieu	Engineer	CNRS		

Non-LRI personnel of PCRI project In Situ					
Name	First Name	Position	Institution	Dates	
FEKETE	Jean-Daniel	CR	INRIA	since 01/09/02	
MACKAY	Wendy	DR	INRIA	since 01/09/02	
COSTA-CUNHA	Pascal	AM		since 01/09/03	
DACHARY	Loic	Ingénieur expert	INRIA	01/03/03 - 31/12/03	
EVANS	Helen	Engineer	INRIA	15/10/03 - 31/12/03	
GAUDRON	Nicolas	Engineer	INRIA	01/03/03 - 30/09/04	
HANSEN	Heiko	Engineer	INRIA	15/10/03 - 31/12/03	

\*See the glossary for acronyms.

Long term visitors						
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
ZAMBRANO	Nancy	Venezuelian	Central University of Venezuela	08/01/01	30/09/01	Sabbatical
ZHAI	Shumin	Canadian	IBM Almaden Research Center, USA	01/06/02	31/08/02	Sabbatical
QUARFORDT	Pernilla	Swedish	University of Linkoping	17/06/02	31/08/02	PhD student
LYU	Michael R.	Chinese	Chinese University of Hong Kong	17/03/03	21/03/03	
GALVES	Antonio	Bresilian	University of Sao Paulo	01/06/03	30/06/03	Visiting professor, Paris XI
GREENBERG	Saul	Canadian	University of Calgary	22/09/03	21/10/03	Visiting professor, Paris XI

## **Group evolution**

Since 2000, three permanent researchers have left the group, while four new assistant professors have been recruited :

## Departures

# • Bruno Marre, CNRS researcher, left in September 2002, is now CEA-LIST;

- Guy Vidal-Naquet, Professor, left in March 2001, is now at Supelec;
- Mountaz Hascoët, Assistant Professor, left in September 2001, is now at Université de Montpellier (LIRMM).
- Arrivals
- Nicolas Roussel, Assistant Professor, 2001
- ${\ensuremath{\bullet}}$  Eric Petitjean, Assistant Professor, 2002
- Lila Boukhatem, Assistant Professor, 2002
- Fatiha Zaidi, Assistant Professor, 2003
- Michel Beaudouin-Lafon returned in September 2000 from a two-year leave (*détachement*) at University of Aarhus, Denmark.

Research groups

Prog Members

## **Research description**

The *Programming and Software Engineering group* is organized in three research teams: *Software Engineering, Human Computer Interaction,* and *Mobile Network.* 

The first team has been working for a long time on the role of formal specifications in software development. The second team studies novel interaction techniques, cooperative work environments, visualisation and the design of interactive systems. The last team is devoted to mobility and communication via wireless interfaces. It was created with the goal of creating an autonomous research group on these topics.

The cross-fertilisation between these teams is ensured by a group seminar and by several common activities. The most recent ones are *visualisation of complex programs or specifications* (a master thesis was defended in september 2004 and will continue into a Ph.D. thesis, see Figure 1), and *test of communication protocols* based on formal descriptions, which is starting now with the arrival of a young assistant professor from INT (Evry).



Figure 1: Example of the visualisation of a complex piece of software by the SEQUOIA tool. The size of the red boxes represent the a priori importance of the corresponding part of the software according to a semantic fisheye (Furnas, 1986). The SEQUOIA tool is developed with the InfoVis toolkit [91], and will provide the user interface for the AUGUST tool [86].

#### The Software Engineering team

Deriving test cases from specifications is now recognised as a major application of formal methods to software development. It is the current focus of this team. Different methods have been proposed for various formalisms: behavioural descriptions such as transition systems, model-based specifications, algebraic specifications, etc.

Based on previous work on testing from algebraic specifications, the team has proposed a general framework for test data selection from formal specifications. A notion of "exhaustive test set" is derived from the semantics of the formal notation and from the definition of a correct implementation (given as a satisfaction relation or a conformance relation). Then a finite test set is selected via some "selection hypotheses". This approach had been first illustrated by its application to algebraic specifications with the LOFT system and objectoriented Petri nets (CO-OPN2) at EPFL in Switzerland. During the last four years it has been successfully applied to LUSTRE with the GATEL system (in collaboration with CEA-LIST), full LOTOS, and Input/Output automata.

Recently, probabilistic and statistical approaches have been introduced, taking advantage of competences of colleagues in the BioInfo and Algo groups. This work resulted in an innovative way of implementing statistical testing, using uniform generation of combinatorial structures.

There are around ten groups working on software testing in France. The LRI team has been the first one to study testing based on formal specifications. It keeps a leader role, with frequent meetings of the community in Orsay. At the international level, the team is present in almost all of the program committees in the domain and publishes in the best conferences.

#### The Human Computer Interaction team

The human-computer interaction (HCI) theme was originally developed to address the design and development of interactive tools for software development. Since then it has evolved to encompass a wider range of topics, including novel interaction techniques, computer-supported cooperative work, information visualization, participatory design and augmented (or mixed) reality. In 2002, a joint project was established between this group and INRIA Futurs within the framework of the PCRI (Pôle Commun de Recherche en Informatique). This new project, called In Situ (Interaction Située or Situated Interaction) has grown from 5 members in 2002 to about 20 by the end of 2004, including the permanent researchers and contract personnel recruited by INRIA and the external collaborators. This makes it one of the largest academic research group in HCI in France and the only one with such a wide coverage of HCI topics.

Over the past 4 years, the main accomplishments of this team are:

- The development of an interaction model called instrumental interaction and its application to the development of novel post-WIMP interfaces;
- The development of tools and techniques for video-based collaboration,
- Fundamental work on evaluation and optimization of pointing performance,
- New research methods for developing new interactive technologies for non-professional users,
- The creation of a novel interface for non-technical users to manage 'communication appliances' small-group networks,
- The development of paper-based applications, in biology and air traffic control, that provide users with the benefits of both physical paper and on-line information, and
- Novel methods of visualising information and benchmarks for comparing them.

An original aspect of In Situ is to practice multi-disciplinary research by associating industrial designers, psychologists and computer scientists. This mix of backgrounds and corresponding research methods has led to several types of cross-fertilization and original research that is well recognized, with participations in many program committees and publications in the best conferences. Such multi-disciplinary teams are mostly found in Scandinavian countries (KTH Stockholm in Sweden, University of Aarhus in Denmark), in the U.K. (see the 6-years Equator project funded by EPSRC) and the U.S.A. (CMU, MIT Media Lab, IBM, Microsoft Research). We have several collaborations with some of these research centres, demonstrating that our research is at the forefront of the field of HCI.

#### **The Mobile Networks team**

The third and most recent team in the group was created in September 2000 to develop some research in wireless and mobile systems at LRI. The team works on cellular systems (UMTS), mobility in IP networks, ad hoc networks and network mobility.

Research groups

Prog Research description The team produces high quality research in this domain. Four Ph.D. theses have already been defended and four students are preparing theses. The team publishes in top-level conferences and journals (IEEE, ACM) and the members are very active in normalization and standardization associations. The team collaborates nationally and internationally with several laboratories and industrial companies. Team members are present in the program committees of conferences in the mobile and wireless areas. Khaldoun Al Agha was the chair of three program committees and a guest editor for ACM WINET.

The team has developed algorithms for routing in ad hoc networks based on quality of service. These solutions have been implemented and are heavily downloaded and used by academic and industrial labs. These solutions will soon be presented to standardization. Today, we are considered as the experts in France in this area.

## Software Engineering team

- Permanent members: Marie-Claude Gaudel (head), Bruno Marre (until september 2002), Éric Petitjean (since september 2002), Françoise Schlienger, Frédéric Voisin, Fatiha Zaidi (since september 2003).
- Ph.D. students: Sandrine-Dominique Gouraud, Gregory Lestiennes.

The Software Engineering team mainly develops activities in the area of software testing based on formal specifications [110, 106]. For several years, probabilistic and statistical approaches to software testing have taken some importance in our work [99, 86]. Another novelty is the study of interoperability testing [82]. We also address related research topics such as software reliability, robustness, atomicity and actions compensation. They are not reported here, but they appear in our list of publications.

#### **Testing based on formal specifications**

The team has developed for several years a theory of software testing that has led to a general framework for black-box testing based on formal specifications, applicable to arbitrary formalisms. It has been recognized as a pioneering work and a fundamental result in the area of software testing. Among the formalisms considered during the last four years, we present LUSTRE [110] and Input/Output automatas [106].

Testing based on LUSTRE descriptions. LUSTRE is a synchronous language used for the description of control-command embedded systems. The development of a method and a tool for deriving test cases from such descriptions has been realised in the framework of a collaboration with the CEA-List laboratory (*Commissariat à l'Energie Atomique, Laboratoire d'Intégration des Systèmes et des Technologies*, Saclay). Bruno Marre is currently on leave at CEA-List where he manages the development and use of the GATEL tool. Gatel was used by IRSN (*Institut de Radioprotection et de Sûreté Nucléaire*) for assessing the test of critical systems. It is currently experimented by ONERA-Toulouse for Airbus-industrie.

Input/Output Automata with data types. I/O automata or I/O transition systems are popular in the area of testing since they are claimed to provide a realistic basis for dealing with concurrent systems and communication protocols. In [106] we have improved the existing theory for testing based on finite I/O transition systems. As a consequence we have given a more efficient algorithm for test generation. Moreover, we have extended this approach by introducing data types (i.e. guarded transitions and transmissions of values). This leads to infinite I/O transition systems, and raises difficult new problems, among which reachability. Based on our previous work on full Lotos and data types, we have proposed testing strategies that take into account both the topology of the automata and the properties of the data types. I/O automata are not very realistic, since they only allow the modelling of "inputenabled" systems, i.e. systems that accept any input at any moment. We are currently studying how to enrich the model in order to describe and test systems that do not fulfil this strong requirement.

## **Probabilistic and Statistical Approaches**

In recent years, the general problem of studying and simulating random processes has particularly benefited from progresses in the area of random generation of combinatorial structures. This work constitutes the basis for powerful tools for uniform random generation of complex entities, such as graphs, trees, words, paths... We have started some research for using such concepts and tools for random software testing in collaboration with the bio-informatics group at LRI [99, 86].

We have developed a method for using these tools as soon as there is a graph description of the behaviour of the system under test. It may be the control graph of the program, or some specification, and the graph may be either directly available (Statecharts, Petri nets) or indirectly via some semantics in terms of transition systems, automata, state machines, Kripke structures, etc. Such behavioural graphs can be described as labelled combinatorial structures. Therefore, uniform generation can be used for drawing paths from the set of execution paths or traces of the system under test, or, more efficiently, among some subsets satisfying some coverage conditions. This method is similar to the statistical testing methods developed at LAAS (Toulouse) several years ago. It has the advantage of being fully automated: A tool, AUGUST, has been developed and experimented on the same set of C programs as LAAS. It provides a way to perform intensive testing with an excellent experimental detection power.

We are interested in finding a formal way for assessing this detection power, for instance by estimating the probability of remaining undetected faults. We have started to explore the relationship between these kinds of testing methods and probabilistic verification methods in the area of model checking. We have regular brainstorming meetings on this topic with colleagues from the Complexity team (Algo group).

#### Interoperability Testing

Interoperability testing of new protocols and services becomes a strategic activity in the telecom industry. Indeed, new trends require to design new protocols and services that combine wireless networks and the Internet such as the WAP system. We have adapted previous work on component testing based on formal specification, namely the hit-or-jump method, to address interoperability and layer protocol testing. We have obtained promising results when experimenting it on WAP protocols. This testing approach, especially the real platform, can be updated to test other protocols such as the GPRS and UMTS protocols.

We are also leading experiments on how to use and adapt the SPIN model-checker for implementing this hit-or-jump method. Interoperability models correspond to huge transition systems; SPIN's algorithms (and probably several other model checker kernels as well) seem to provide relevant and optimised implementation techniques for this method. In the context of interoperability and integration of services, scalability is a key issue and the know-how of the model-checker is probably worth looking at.

#### **Key references**

[82] A. Cavalli, A. Mederreg, and F. Zaïdi. Application of formal testing methodology to wireless telephony Networks. In 2<sup>nd</sup> International Information and Telecommunications Technologies Symposium, I2TS'2003, Florianopolis, Brazil, october 2003. Research groups

Prog Research description: Software engineering team

- [86] A. Denise, M.-C. Gaudel, and S. Gouraud. A generic method for statistical testing. In Proceedings of 15th Int. Symposium on Software Reliability Engineering (ISSRE 2004), Saint-Malo, F, Nov. 2004. IEEE. 15 pages, in press.
- [99] S.-D. Gouraud, A. Denise, M.-C. Gaudel, and B. Marre. A new way of automating statistical testing methods. In *Automated Software Engineering Conference (A.S.E.), I.E.E.E*, pages 5-12, 2001.
- [106] G. Lestiennes and M.-C. Gaudel. Testing Processes from Formal Specification with Inputs, Outputs and Data Types. In Proceedings of 13th I.E.E.E. International Symposium on Software Reliability Engineering, Annapolis, Maryland, pages 3-14, 2002.
- [110] B. Marre and A. Arnould. Test sequences generation from lustre descriptions: Gatel. In Proc. Fifteenth IEEE Int. Conf. on Automated Software Engineering (ASE 2000), pages 229-237. IEEE, Sept. 2000.

## In Situ PCRI project



- Permanent members: Wendy Mackay (head), Michel Beaudouin-Lafon, Jean-Daniel Fekete, Emmanuel Pietriga (since October 2004), Nicolas Roussel, Olivier Chapuis (since March 2004).
- Ph.D. students: Caroline Appert, Olivier Beaudoux (until July 2004), Renaud Blanch, Pascal Costa-Cunha, Nathalie Henry (since October 2004), Jean-Baptiste Labrune (since October 2004), Jérôme Lard (since June 2004), Emmanuel Nars, Yann Riche (since October 2004).
- Non-permanent members: Stéphane Conversy (until September 2003), Jean-René Courtois, Loïc Dachary (2003), Yangzhu Du, Helen Evans (2002-2003), Nicolas Gaudron (2003-2004), Heiko Hansen (2002-2003), Fabrice Haiiy, Jean-Christophe Latsis, Matthieu Langet.
- External collaborators: Yves Guiard (CNRS & Université de la Mediterranée), Catherine Letondal (Institut Pasteur).

The In Situ project was created in 2002 as a PCRI project between INRIA Futurs and LRI. Two INRIA researchers (Wendy Mackay and Jean-Daniel Fekete) were recruited in 2002 to start the project together with the Human-Computer Interaction team, which had five members at the time. Within two years the team has grown to 21 people, with two new permanent researchers (one from INRIA, one from CNRS), seven new Ph.D. students, five temporary members and two external collaborators who work closely with us.

The In Situ project addresses the problem of how to create effective interactive systems, considering the role of context, both human and technical, as fundamental to the problem. Creating *situated interfaces*, i.e. interfaces that are adapted (or adaptable) to their contexts of use, takes advantage of the complementary features of humans and computers.

The research is organized along three main themes: new interaction paradigms, participatory design and engineering of interactive systems. However these themes are often intermixed in actual projects, which we see as indispensable in order to address real-world problems. More specifically, the In Situ project develops novel interaction techniques, new tools (in the form of toolkits) to develop these techniques and new methods to enhance the design process of interactive systems. In the long run, the project seeks to create a new generation of interactive environments as an alternative to the current generation of desktop environments.

Our group is multi-disciplinary, including computer scientists, psychologists and designers. Working together requires an understanding of each other's methods. Much of computer science relies on formal theory, which, like mathematics, is evaluated with respect to its internal consistency. The social sciences are based more on descriptive theory, attempting to explain observed behaviour, without necessarily being able to predict it.

G

The natural sciences seek predictive theory, using quantitative laws and models to not only explain, but also to anticipate and control naturally occurring phenomena. Finally design is based on a corpus of accumulated knowledge, which is captured in design practice rather than scientific facts but is nevertheless very effective.

Combining these approaches is a major challenge. We are exploring an integrative approach that we call *generative theory*, which builds upon existing knowledge in order to create new categories of artefacts and explore their characteristics. Our goal is to produce prototypes, research methods and software tools that facilitate the design, development and evaluation of interactive systems [67].

#### **New interaction Paradigms**

*Post-WIMP interfaces.* WIMP (Windows, Icons, Menus, Pointing) interfaces have become the de-facto standard of personal computing. Post-WIMP interfaces build on this standard by developing and integrating novel interaction techniques that are better adapted to the tasks at hand. We have developed *instrumental interaction*, an interaction model based on the metaphor of the tools and instruments that we use to mediate our actions in the real world [65, 69]. We have applied this model to various post-WIMP applications, including a document-based environment where the primary objects are documents rather than applications, and where instruments are independent from the object they operate on [71, 19, 173]. We have also explored end-user development, an approach where users can easily tailor and extend their environment to better suit their needs and work practices [178, 107]. Finally we are working on modelling post-WIMP interfaces in order to better predict their performance in the context of use [52].

*Pointing*. Pointing is a fundamental action in graphical user interfaces. We have conducted fundamental work, based on Fitts' law (a robust empirical law in Psychology that predicts the minimum time to reach a target of a specified size at a specified distance [30]), to improve pointing. We have extended Fitts' law to zoomable interfaces [102, 16], where pointing requires a complex coordination of panning and zooming, and shown that multiscale pointing performance was inversely proportional to the size of the view used to display a zoomable document, up to a certain size [100, 16]. We have also developed two novel pointing techniques, semantic pointing [78] (Figure 2) and object pointing [101], that exploit the information known to the system in terms of possible targets to improve pointing performance and shown significant gains over traditional pointing.



Figure 2: With Semantic Pointing, objects can have a visual size different from their size in motor space, i.e. they can "feel" smaller (or larger) than they look. This is achieved by dynamically controlling the mapping between mouse movements and cursor movements. Here it is applied to a dialog box: (a) visual appearance, (b) motor space. The default button is made bigger in motor space and therefore easier to select, while non-interactive elements such as labels or dangerous ones such as the Don't Save button are made smaller in motor space and therefore harder to select. Research groups

Prog Research description: In Situ PCRI project *Video-mediated communication.* Over the past few years, audiovisual systems have greatly improved in their ability to support formal business meeting, e.g., videoconferencing, but they are still hardly usable for casual interactions. Following up on our previous work on mediaspaces in work settings [40], we have prototyped and patented a new communication device called the well [122] that combines audio and video in an original design. The well supports the emergence of a relaxed atmosphere well adapted to the informal discussions that take place before and after a meeting or during breaks. We are now mostly interested in video-mediated communication in domestic settings. In particular, as part of the interLiving project (see below), we have recently explored asynchronous image-based communication [135] as well as proximity-driven video interaction [17, 124, 154].

*Interactive paper.* Desktop computers have revolutionised the workplace, but have also generated problems. They often disrupt effective work practices and require many users to juggle both paper and on-line information. Humans like paper because the sensory-motor system is well suited to managing physical objects in the world. Today's computers overload the human visual system, while ignoring most other senses. Yet we need not choose between the physical and the virtual: paper can be an effective input device to a computer, as well as an object in its own right. We have been working with users who face complex tasks involving both off-line and on-line information, including air traffic controllers, research biologists and research historians. We have developed a diverse set of prototypes that allow users to interact with paper as one of their primary interfaces to on-line information [147]. Prototypes include augmented paper flight strips, augmented laboratory notebooks [109, 84] (Figure 3) and tools for interacting with ancient documents from the National Archives.



Figure 3: The A-Book uses a PDA to interact with paper. Information written on paper is captured digitally and displayed on the PDA according to its current position, giving the illusion that the PDA is a transparent window onto the sheet of paper.



*Familiar data*. Computer users are confronted to increasing amounts of files, email, photos and digital information on their hard drives, that they find increasingly difficult to manage. Since computer screen size and human cognitive abilities are limited, interfaces must offer more powerful tools for efficient navigation of this familiar information. In the context of the Micromegas project funded by the French Ministry of Research *(ACI Masses de*  *Données)*, we develop new visualization and navigation techniques that take advantage of the users' knowledge of their on-line data. So far, we have focused on collecting event-logs describing the users' activities with documents, applications and window system. We plan to use these logs to provide users with better navigation cues and to develop multiscale (or zoomable) interfaces to help users manage complexity.

## **Participatory design**

Participatory design involves users during all stages of the design process. Users are not simply consumers of systems, but active participants in their design and interpretation. Sometimes, users can be major innovators, especially with respect to understanding the situated nature of the users' activity with an interactive system. The In Situ project explores and tests novel participatory design methods and techniques that emphasize the role of context and make it explicit in the design process. Wendy Mackay has developed teaching materials, including a DVD on video techniques for participatory design distributed via ACM/SIGCHI and INRIA, and a Participatory Design Toolkit [108]. We have conducted workshops and research design activities in a wide variety of domains including biology (laboratory notebooks, on-line tools, end-user programming), animation (tools for colouring animation), air traffic control (interactive paper flight strips and a touch-output system for managing strips), office work (use of browsers and on-line tools) and archives (historical analysis of prominent authors' drafts, analysis of how researchers examine ancient documents, from the French National archives and Ethiopian manuscripts).

*InterLiving*, a three-year participatory design project funded by the European IST program, explored the design space of technologies that support communication among family members living in different households [187, 188, 189, 190]. We worked closely with six distributed French and Swedish families (totalling over 80 people) to collaboratively design, prototype and test a series of peripheral communication technologies that help them stay in touch in lightweight ways [126]. The project resulted in new research methods (see technology probes below), a series of implemented and tested prototypes (see communication appliances below) and the design of tangible interface that enables non-technical users to effectively manage small groups (patent applied for).

*Technology probes* address the problem of how to effectively develop innovative technologies for non-technical users in home settings [104]. Working with users, we collaboratively design and then implement extremely simple, single-function communication devices, which are then linked between households. The approach combines elements of social science, engineering and design. From a social science perspective, we gather data both about the participants and their use of the system. From an engineering perspective, we test the system in a real-world setting and learn about the technical requirements of such systems. From a design perspective, we involve the users in the design process and let these systems serve as a source of inspiration for future ideas. Our experience developing, installing and testing technology probes led us to the concept of *communication appliances* and a strategy for how to create them.

*Communication Appliances.* The interLiving project produced a series of communication appliances, many of which were tested *in situ* in the interLiving families' homes. Communication appliances are defined as: simple-to-use, single-function devices that let people communicate, passively or actively, with one or more remotely located friends or family. Connections are pre-determined (though changeable) and stay open, allowing communication without explicit dialling or addressing. Communication appliances may share sound, images, video, text or even touch and connections may range from focused, synchronous contact to peripheral awareness of one another.

Research groups

Prog Research description: In Situ PCRI project We collaboratively developed a series of communication appliances, and have filed a patent on an infrastructure component that provides a tangible interface for managing small social networks by non-technical users. The prototypes include VideoProbe [135], for sharing spontaneous snapshots among family members, TableProbe and StoryTable, for collaboratively editing video via a tangible interface, Tok-i-Tok, for sharing sound patterns, and MessageProbe, for sharing zoomable, post-it like messages. We also created the MirrorSpace [17, 124, 154] (Figure 4), which has been exhibited in numerous venues, including the Pompidou Center art museum and the La Villette science museum in Paris.



Figure 4: Participants in different locations see each other via a screen; when far away, the images of both participants are blurred together. Walking toward the screen clarifies the images; allowing the participants to control their level of visual intimacy.

#### **Engineering of interactive systems**

WIMP interfaces have been with us for twenty years, and little innovation has made it into products despite the many interaction techniques developed in the labs. The lack of appropriate tools to help developers integrate these advanced techniques into commercial software is probably a major reason for this situation. We have developed experimental toolkits in three areas that we believe are essential for moving towards post-WIMP interfaces: distributed interaction, video and information visualization.

*Distributed Interaction*. Today, more and more applications are distributed and users access them through a variety of platforms. The Web has become the de facto standard for deploying such applications. Unfortunately, Web browsers support a limited style of interaction: primarily clicking links and filling out forms. We study component-based architectures that leverage novel interaction techniques in distributed applications over multiple platforms. Individual components that implement, e.g. interaction techniques, may be added, removed or substituted dynamically. In the INDIGO project, Object Servers provide application data, together with rendering and interaction rules, to Interaction Servers that manage rendering and user interaction. This architecture natively supports real-time shared editing of application data across multiple platforms.

*Video.* Supporting communication in collaborative work applications often requires video links, yet it is well known that integrating video into applications is difficult. As a result, the use of video in interactive systems has always been very limited. VideoSpace and its successor Núcleo were designed explicitly to explore new uses of video and new interaction techniques based on video. Rather than focusing on technological aspects such as compression, transport or synchronization, they focus on supporting rapid prototyping and

incremental development of video applications. These toolkits treat video as a first-class object in interactive systems [181].

For example, they allow users to display image streams in HTML documents in place of ordinary static images or to embed these streams into existing X Windows applications. Creating a video link requires only a few lines of code; managing multiple sources and including video processing is not much more complicated. Most of our recent video-based systems, including the *well*, VideoProbe and MirrorSpace, have been implemented with videoSpace or Núcleo. Núcleo has also been used to develop a series of innovative window manager prototypes [123, 153].

Information Visualization. Creating new information visualization techniques using traditional GUI toolkits is cumbersome. We have designed a toolkit to experiment with new techniques in a much simpler and general way than before: the Infovis Toolkit [91, 22]. It is based on specific data structures to achieve a fast action/feedback loop as required by dynamic queries and comes with a large set of components such as range sliders to control and configure the visualizations. InfoViz supports scatter plots, time series, Treemaps, nodelink diagrams for trees and graphs and adjacency matrix for graphs, and all visualizations can use fisheye lenses and dynamic labelling (Figure 5). The toolkit, written in Java, supports OpenGL-based hardware acceleration in order to provide high performance. We are now exploring new visualisation techniques for large graphs and new interaction techniques to filter large datasets since existing techniques do not scale well.

We are currently applying the toolkit to several research projects such as real-time program monitoring [98] and large graph visualisation using adjacency matrices [144] [97, 145] [96]. See http://insitu.lri.fr/~fekete/InfovisToolkit for more details.



*Figure 5: Visualization of scatter plots, treemaps and adjacency matrix (with a fish-eye view) using the InfoVis toolkit.* 

#### **Key references**

- [17] N. Roussel, H. Evans, and H. Hansen. Proximity as an interface for video communication. *IEEE Multimedia*, 11(3):12-16, July-September 2004.
- [69] M. Beaudouin-Lafon and W. Mackay. Reification, polymorphism and reuse: Three principles for designing visual interfaces. In *Proc. International Conference on Advanced Visual Interfaces (AVI 2000)*, pages 102-109, Palermo, Italie, May 2000. ACM Press.
- [78] R. Blanch, Y. Guiard, and M. Beaudouin-Lafon. Semantic pointing: Improving target acquisition with control-display ratio adaptation. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI2004)*, volume 6(1) of *CHI Letters*, pages 519-526. ACM Press, Apr. 2004.
- [91] J.-D. Fekete. The infovis toolkit. In *Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis 04)*, page in press, Austin, TX, Oct. 2004. IEEE Press.

Research groups

Prog Research description: In Situ PCRI project

- [104] H. Hutchinson, W. Mackay, B. Westerlund, B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, B. Eiderbäck, S. Lindquist, and Y. Sundblad. Technology probes: Inspiring design for and with families. In *Proceedings of ACM CHI 2003 Conference on Human Factors in Computing Systems*, volume 5(1) of *CHI Letters*, pages 17-24. ACM Press, 2003.
- [109] W. Mackay, C. Letondal, G. Pothier, K. Bøegh, and H. Sørensen. The missing link: Augmenting biologist's laboratory notebooks. In *Proc. ACM Symposium on User Interface Software and Technology (UIST 2002)*, volume 4(2) of *CHI Letters*, pages 41-50, Paris, France, Oct. 2002. ACM Press.

## Mobile Networks team

- Permanent members: Khaldoun Al Agha (head), Lila Boukhatem.
- Ph.D. students: Hakim Badis, Inès Ben Hamida, Mounir Benzaid, Davor Males.
- Non-permanent members: Thamer Al-Meshhadany, Alia Fourati, Ignacy Gawedzki, Tinku Mohamed-Rasheed, Laurent Perato.
- External member: Anelise Munaretto.
- Temporary members who left: Marcelo Dias de Amorim, Nadjib Achir, Dany Zebiane, Hayder Saleh.

Research work in the Mobile Networks team addresses problems directly related to user mobility and communication on wireless interfaces. Our goal is to design protocols and algorithms that optimize communication system functionalities. We study algorithms in several ways: theoretically (convergence, performance evaluation, worst case calculation, etc.), by simulation (scalability, dynamic system behaviour, mobility consideration, etc.) and by implementations. The combination of these studies allows us to provide robust and high-performance solutions and has proved very competitive at the international level.

#### UMTS

UMTS (Universal Mobile Telecommunication System) is the European proposition for the third generation consortium created by the International Telecommunication Union (ITU). The UMTS must eventually replace all second-generation systems, such as GSM and PDC. It must also offer more bandwidth and quality of service. UMTS provides a bandwidth of up to 2Mbps.

The Code Division Multiple Access (CDMA) is the heart of third generation systems in terms of users and services multiplexing. The flexibility of this method provides an adaptive code allocation depending on user's requirements. UMTS features two major modes: FDD-WCDMA and TDD-TD/CDMA. FDD-WCDMA is a common contribution by the Japanese community.

We are working on optimizing code allocation in a cell while using the WCDMA. We have studied the interference generated by orthogonal codes and Pseudo-Noise sequences. This interference can be reduced in order to increase the number of users or services. We proposed and analyzed a scheme that consists in reversing the allocation process. In the UMTS process and in the uplink direction, the services of each user are multiplexed on one scrambling code while our scheme proposes to multiplex all users of the same service on a unique scrambling code.

We have also worked on reducing the quantity of signalling and control information on the radio interface so that more channels are available for data. Our idea consists in using mobile agents as a proxy for users. Each agent follows the user mobility and controls tasks related to the user. Multiple schemes of handover prediction were studied to optimize the UMTS resource allocation.

#### **Mobility in IP**

The macro-mobility in IP systems should be handled by the Mobile IP protocol. This protocol provides a nomadic user with the same IP address through the network. However, Mobile IP presents some weaknesses in the case of high user mobility. Consequently, the robustness of Mobile IP is valid only in nomadic contexts. On the other hand, the MANET group of the IETF works on creating standards for ad-hoc networking by using the IP protocol. These protocols offer the possibility of routing packets from host to host via the air interface and only in a restricted area.

Our goal is to combine Mobile IP with ad-hoc networking. In fact, we propose an architecture for mobility where the mobile host can roam in the IP networks with its permanent IP address. When macro-mobility occurs in a different IP domain, Mobile IP must provide a temporary address to the user while for micro-mobility, an ad-hoc protocol has to deliver data to the user by wireless routing. The ad-hoc protocol used in our architecture is OLSR (Optimized Link State Routing). We also develop a fast moving extension OLSR in order to guarantee a robust handoff for high-speed users.

#### 4G and Vertical handoff

The fourth generation mobile systems attempts to federate under the IP label all heterogonous wireless interfaces. This way, users can easily roam from one system to another without any specific operation. This vertical handover must provide transparent mobility through these incompliant systems. For example, this technique can offer session continuity when moving from UMTS to WI-FI.

Our work consists in reducing the time latency when roaming. Nowadays, transferring a session from one system to another takes several seconds, which is unacceptable for a user in communication. We created the concept of a virtual cell by using multicast facilities in order to reduce the latency and the packet loss.

#### Ad hoc networks (QoS and Security)

Ad hoc networks are systems that can establish communication sessions between users without using any kind of infrastructure. Hence, the network must be self-organized and dynamically adaptive.

Nodes in ad hoc networks are in constant movement. Delivering packets to a specific destination becomes a complex task. Today, four routing protocols were standardized by the IETF (Internet Engineering Task Force) to offer a solution for routing problems. These solutions are divided into two categories: proactive (routes are always available) and reactive (routes are discovered on demand). OLSR (Optimized Link State Routing Protocol), developed at INRIA, is a proactive protocol based on link state routing mechanisms.

In our work we have addressed issues of quality of service and security. All standardized protocols try to find a route from a source to a destination by considering only one metric: the number of hops. Also, these protocols do not take into account any security consideration and still are vulnerable to attacks. We consider up to four metrics to establish routes, including delay, available bandwidth, jitter, loss probability, etc. This should allow flows

Research groups

Prog Research description: Mobile Networks team to reach their destination with an increased quality of service. We also consider the multipath routing and path oscillation problems.

Concerning security, we are working on a solution to secure OLSR. Our mechanism must enumerate the most important attacks and provides mechanisms that eliminate all negative effects on routing information. The solution will be based on distributing secure keys to nodes without using any centralized mechanism since this is not feasible in an ad hoc context.

#### **Network mobility**

A MONET (Mobile NETwork) is an entire network that changes, as a unit, its point of attachment to the Internet and thus its reachability within the topology. Possible applications of MONETs are access networks deployed in public transportation (buses, trains, taxis, aircrafts) or networks attached to people Personnel Area Networks or PANs) made of cellular interfaces (cell phone, electronic diary). NEMO (NEtwork MObility) is concerned with managing the mobility of MONET networks and providing a continuous network access to all nodes within it in order to maintain session continuity while moving. Our work aims to show that the Mobile-IPv4 and Mobile-IPv6 protocols, which manage node mobility in the Internet, cannot support efficiently a MONET MObile NETwork. Thus, we have proposed an extension to Mobile-IPv6 that uses a multicast solution to support MONET mobility and to reduce the control traffic. We are also interested in resource reservation aspects in this kind of networks. In fact, users' collective movement in NEMOs applications can be quite predictable, e.g., users in a train. Thus, we plan to use resource reservation techniques to provide users with a better quality of service in terms of parameters such as bandwidth and blocking probability.

## **Research Perspective**

With respect to formal methods, testing and reliability, we plan to continue our work on test derivation from infinite models, i.e. symbolic transition systems, with the use of constraint solving techniques. In addition, our new interests in interoperability-integration techniques and in the use of probabilistic and statistical methods will be emphasized.

Regarding human-computer interaction and the In Situ project, we envision a slower growth of the team now that it has reached a critical mass. We will continue our efforts to create a stable, multidisciplinary team, which is a challenge in France where disciplinary barriers are still high, especially with non-scientific fields such as art and design. Our research will continue to address fundamental as well as methodological and technical issues and will seek to cover a wider range of interactive systems, in particular in the context of "ubiquitous", "pervasive" or "ambient" computing.

In the area of Mobile Networks, we will continue to work on wireless and mobile networking, with a special emphasis on security and quality of service. We also plan to develop a new activity on user localization using routing information. B.3.3 / Prog

## Highlights

The two major events over the last four years were the creation of the Mobile Network team and the In Situ PCRI project. The group is also very active in the development and management of Computer Science research in Orsay: Michel Beaudouin-Lafon is director of LRI, Marie-Claude Gaudel has been chair of the Computer Science Department and is now director of PCRI. Other members of the group are in charge of significant responsibilities.

The group is known and recognised for its research at the best international level. This is reflected by the number of international program committees, the number of invited lectures, and the number and quality of publications. In parallel, we have an important contractual activity, both at the national and european levels.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,

Laboratoire de Recherche en Informatique / 2000-2004 report

- Number of participations in program committees (national and international), and,
- *Number of defended Ph.D.s and habilitations.*

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

Research groups

Prog Research perspective Highlights

## Honors

## Prizes and awards

• Michel Beaudouin-Lafon, co-author of Best Long Paper, *Joint HCI-IHM conference on Human-Computer Interaction*, Lille, sept. 2001: «Beyond the 10-bit barrier: Fitts' Law in Multi-Scale Electronic Worlds» [102].

## **Keynote addresses**

## International

- Marie-Claude Gaudel, *Dagstuhl seminar on Perspectives in Model Based Testing*, Sept. 6-10, 2004: "Unfolding Symbolic Transition Systems".
- Michel Beaudouin-Lafon, opening keynote at *International Conference on Advanced Visual Interfaces* (AVI '04, Gallipoli, Italy, may 2004): "Designing Interaction, not Interfaces" [67].
- Khaldoun Al Agha, Panelist in the panel on Mobile Ad-hoc Networks and Dependable Computing, *Mediterranean Ad Hoc Networking Workshop* (Med-Hoc-Net 2003), Mahdia, Tunisia, 27 June 2003.
- Marie-Claude Gaudel, *6th Workshop on Formal Methods* (WMF'2003), Campina Grande Brasil, Oct. 2003: Problems and Methods for Testing Infinite State Machines.
- Marie-Claude Gaudel, *The 2002 ARO/ONR/NSF/DARPA Monterey Workshop on Radical Innovations of Software and Systems Engineering in the Future*, Venise, Oct. 2002: "How to Reach Software Ultra-Reliability?"
- Michel Beaudouin-Lafon and Wendy Mackay, invited conference at *Design Forum, ACM Human Factors in Computing Systems* (CHI '02), Minneapolis, April 2002: "Power and Simplicity".
- Fatiha Zaïdi, *Worshop on protocol and distributed systems testing*, Campinas University, Brazil, Oct 2001: "Open Problems in protocol testing".
- Marie-Claude Gaudel, *Workshop on Concurrency in Dependable Computing*, June 26, 2001 in Newcastle upon Tyne, UK: "Testing Systems from Formal Specifications".
- Marie-Claude Gaudel, *ADA-Europe Conference*, Louvain, May 2001: "Testing from Formal Specifications, a Generic Approach".

#### France

- Khaldoun Al Agha, Journée STIC, Université Paris-Sud XI, 2002: "Réseaux ad hoc MANET".
- Khaldoun Al Agha, *Journée Aristote, École Polytechnique*, 2001: "L'intégration d'IP dans les réseaux mobiles et sans fil de la future génération".

## Other honors

• CSCW Journal (Kluwer Academic): Michel Beaudouin-Lafon, member of advisory board.

## **Evaluation of research**

## **Editorial boards**

## International

- Science of Computer Programming (North-Holland): Marie-Claude Gaudel, since its creation in 1982.
- Formal Aspects of Computing (Springer International): Marie-Claude Gaudel, since its creation in 1989.
- Journal of Software Testing, Verification and Reliability (John Wiley): Marie-Claude Gaudel, since its creation in 1991.
- WINET, ACM Wireless Networks: Khaldoun Al Agha, Guest co-editor, July 2004.
- IJHCS, *International Journal of Human-Computer Studies* (Elsevier): Michel Beaudouin-Lafon since 2004, and guest co-editor with Yves Guiard of special issue "Fitts' law 50 years later, applications and contributions from HCI", Dec. 2004.

#### National

- Revue I3 (Cépaduès-Editions): Michel Beaudouin-Lafon, since its creation in 1998.
- *Revue IHM* (EuropIA): Michel Beaudouin-Lafon, since its creation in 2001.

## Program committees

#### Chair

- MWCN, *Fifth IEEE/IFIP International Conference on Mobile and Wireless Communications Networks*, Singapore, 2003: Khaldoun Al Agha, chair.
- Algotel, 5<sup>e</sup> Rencontres Francophones sur les aspects Algorithmiques des Télécommunications, Banyuls-surmer, France, 2003: Khaldoun Al Agha, co-chair.
- Med-Hoc-Net, *IFIP first Annual Mediterranean Ad Hoc Networking Workshop*, Sardinia, Italy, 2002: Khaldoun Al Agha, chair.
- CHI, *18th ACM Conference on Human Factors in Computing Systems*, Seattle, USA, April 2001: Michel Beaudouin-Lafon, co-chair for long papers track (350 submissions, 20% acceptance rate).
- Ergo-IHM, *12<sup>e</sup> Conférence Francophone sur l'Interaction Homme-Machine*, Biarritz, October 2000: Michel Beaudouin-Lafon, co-chair.

#### **Member (international events)**

- WCC, IFIP World Computing Congress, Toulouse, August 2004: Marie-Claude Gaudel.
- SBMF, *Brazilian Symposium on Formal Methods*, Recife, Pernambuco, Brazil, November 2004: Marie-Claude Gaudel.
- *Workshops on Architecting Dependable Systems*, 2 twin satellite workshops of ICSE 2004 (International Conference of Software Engineering) and DSN -2004 (Dependable Systems and Networks Symposium): Marie-Claude Gaudel.
- WOCN, IFIP International Conference on wireless and Optical Communications Networks (WOCN): Lila Boukhatem (2004).
- FATES, *Formal Approaches To Testing*, Satellite Workshop of ASE, Montreal, Oct. 2003, Linz, Sept. 2004: Marie-Claude Gaudel.

Research groups

Prog Honors Evaluation of Research

- CHI, 19th and 21st ACM Conference on Human Factors in Computing Systems: Michel Beaudouin-Lafon (2002, 2004).
- UIST, 17th ACM Symposium on User Interface Software and Technology: Michel Beaudouin-Lafon (2004).
- Multimedia, 12th ACM International Conference on Multimedia: Nicolas Roussel (2004)
- MWCN, IEEE/IFIP International Conference on Mobile and Wireless Communications Networks: Khaldoun Al Agha (2002), Lila Boukhatem (2003).
- CARI, Colloque Africain de Recherche en Informatique: Frédéric Voisin (2000 2002 2004).
- Med-Hoc-Net, IFIP Mediterranean Ad Hoc Networking Workshop: Khaldoun Al Agha (2003, 2004).
- WOCN, *IFIP International Conference on wireless and Optical Communications Networks*: Khaldoun Al Agha (2004).
- ISWCS, International Symposium on Wireless Communication Systems: Khaldoun Al Agha (2004).
- ECSCW, 8<sup>th</sup> European Conference on Computer-Supported Cooperative Work: Michel Beaudouin-Lafon (2003).
- WMF, 6th Workshop on Formal Methods, Campina Grande Brazil, October 2003: Marie-Claude Gaudel.
- IDTP, Sixth IEEE International Conference on Integrated Design and Process Technology, Pasadena, California, June 2002: Marie-Claude Gaudel.
- FASE, *Fundamental Approaches to Software Engineering*, ETAPS, Grenoble, April 2002: Marie-Claude Gaudel.
- FATES, *Formal Approaches To Testing*, Satellite Workshop of CONCUR'02, Brno, August 2002: Marie-Claude Gaudel.
- ISSTA, ACM International Symposium on Software Testing and Analysis, Rome, July 2002: Marie-Claude Gaudel.
- ASW, International Conference on Applications and Services in Wireless Networks: Khaldoun Al Agha (2001).

#### Member (national events)

- DNAC, De Nouvelles Architectures pour les Communications: Khaldoun Al Agha (2000 2004).
- IHM, *Conférence Francophone sur l'Interaction Homme-Machine*: Nicolas Roussel (2002), Michel Beaudouin-Lafon (2004).
- AlgoTel, 4<sup>es</sup> et 6<sup>es</sup> Rencontres Francophones sur les aspects Algorithmiques des Télécommunications: Khaldoun Al Agha (2002), Lila Boukhatem (2004).

## Evaluation committees and invited expertise

#### International

- French Cooperation between Brasilian and French Universities (COFECUB) : Marie-Claude Gaudel is member of the board.
- National Science Foundation (NSF), USA: Marie-Claude Gaudel, referee.
- Fonds québécois de la recherche sur la nature et les technologies (Québec): Marie-Claude Gaudel, referee.
- National Sciences and Engineering Research Council (NSERC), Canada: Michel Beaudouin-Lafon, referee.
- European Science Foundation (ESF): Marie-Claude Gaudel, referee.
- Christian-Döppler-Forschungsgesellschaft, Austria: Marie-Claude Gaudel, referee.
- Engineering and Physical Sciences Research Council (EPSRC), UK: Michel Beaudouin-Lafon, referee.
- Council of Physical Sciences of the Netherlands Organization for Scientific Research (NWO): Marie-Claude Gaudel, referee.
- Research Grants Council, Hong-Kong: Marie-Claude Gaudel, referee.
- National Information Communication Technology Australia (NICTA): Marie-Claude Gaudel, referee.

• Mid-term evaluation committee of EQUATOR, a 6 years, 10M£ project funded by UK's Engineering and Physical Sciences Research Council (EPSRC): Michel Beaudouin-Lafon.

#### France

- Orientation Committees of the French Research Network on Telecommunication (RNRT) and the French Research Network on Software Technologies (RNTL): Marie-Claude Gaudel.
- Evaluation Commission of French Research Network on Software Technologies (RNTL): Michel Beaudouin-Lafon, member (since 2000) and co-chair for special area on interactive systems (2002).
- ${\ensuremath{\, }}$  Scientific board of the STIC department of CNRS, Marie-Claude Gaudel.
- Scientific expert for French Ministry of Research: Michel Beaudouin-Lafon (since 1992), Marie-Claude Gaudel, Khaldoun Al Agha (since 2002).
- Scientific Committee of IRCAM research center (Paris): Michel Beaudouin-Lafon (2004).
- Evaluation Committee of CLIPS-IMAG research laboratory (Grenoble): Michel Beaudouin-Lafon (2002).
- Prix ROBERVAL: Khaldoun Al Agha (2003), referee.

## Other evaluation activities

- Reviewer for PhD dissertation: Michel Beaudouin-Lafon (7), Khaldoun Al Agha (4), Marie-Claude Gaudel (several times).
- Reviewer for Habilitation: Michel Beaudouin-Lafon (4), Marie-Claude Gaudel (2).

## B.3.6 / Prog

## Volunteer professional service

## Management positions in scientific organisations

- Laboratoire de Recherche en Informatique (LRI): Michel Beaudouin-Lafon, Director since 2002.
- Pôle Commun de Recherche en Informatique (PCRI): Marie-Claude Gaudel, Director (2002).
- Department of Computer Science, Université Paris-Sud: Marie-Claude Gaudel, President (1998-2003).
- INRIA Scientific Board: Marie-Claude Gaudel, President (1999-2002).
- National Computer Network for Universities and Research (RENATER): Marie-Claude Gaudel, Chair of the board (2000-2002) and currently member of the board.
- CNRS Network on Models and Tools for Human-Computer Interaction, *Réseau Thématique Pluridisciplinaire Modèles et Outils pour l'Interaction Homme-Machine, RTP16*: Michel Beaudouin-Lafon, co-chair. The network gathers over 200 participants and over 25 French labs from academia and industry.
- CNRS Network on Dependability of Complex Computerized Open Systems, *Réseau Thématique Pluridisciplinaire, RTP21*: Marie-Claude Gaudel, member of the board.
- CNRS Working group on Information Visualization, *Action Spécifique Visualisation d'Information*, 2001-2002: Michel Beaudouin-Lafon, co-chair.

Research groups

Prog Evaluation of Research Volunteer professional service

## Organisation of conferences and scientific events

- MWCN, Sixth IEEE/IFIP International Conference on Mobile and Wireless Communications Networks, Paris, 2004: Khaldoun Al Agha, Conference co-chair.
- IHM, 16<sup>e</sup> Conférence Francophone sur l'Interaction Homme-Machine, Namur, Belgium, September 2004: Michel Beaudouin-Lafon, Conference vice-chair.
- OLSR Day (http://www-rocq.inria.fr/fr/actualites/seminaires/sujet/Hipercom.htm), April 5, 2004: Khaldoun Al Agha, Organizer and co-chair.
- CFIP, 10e Colloque Francophone sur l'Ingénierie des Protocoles, Paris, October 2003: Fatiha Zaïdi, member of organizing committee.
- UIST, 15th ACM Symposium on User Interface Software and Technology, Paris, October 2002: Michel Beaudouin-Lafon, Conference chair.
- ICNP, 10th IEEE International Conference on Networks Protocols, Paris, November 2002: Fatiha Zaïdi, member of organizing committee (local arrangement).

## Working groups

- Working Group 1.3 of IFIP, Specification of Systems: Marie-Claude Gaudel, member.
- CNRS Working group (*Action Spécifique*) on Plasticity of user interfaces with CLIPS, I3S, ICTT, IS3, LAMIH, LIRMM, LMP, SAP, Trigone, UCL, Xerox (2003-2004): M. Beaudouin-Lafon, R. Blanch, J. Lard, members.
- CNRS Working group (*Action Spécifique*) on Advanced Techniques for Testing Complex Systems with LAAS, LABRI, IRISA and IMAG (2002-2003): Marie-Claude Gaudel, Sandrine Gouraud, Grégory Lestiennes, members.
- CNRS Working group (*Action Spécifique*) on Test and Specification techniques for communication software components with SAMOVAR, LABRI, LAAS, LAMI, LSR/IMAG, FT R&D, CEA (2003-2004): Marie-Claude Gaudel, Grégory Lestiennes, Frédéric Voisin, Fatiha Zaidi, members.

## Other professional service

- ACM, Association for Computing Machinery, http://www.acm.org, 75000 members worldwide: Michel Beaudouin-Lafon, elected member-at-large of ACM Council (since 2000, second 4-year term), member of Publications Board (since 2002), member of Grace Murray Hopper Award committee (1999-2003, chair in 2002).
- ACM SIGCHI, ACM Special Interest Group on Computer-Human Interaction, http://www.acm.org/sigchi, 5000 members worldwide: Michel Beaudouin-Lafon, member of Publications Board (since 2001), member of awards committee (1999-2003).
- ASTI, Association des Sciences et Technologies de l'Information, http://www.asti.asso.fr: Michel Beaudouin-Lafon, member of the board (since 2002).
- AFIHM, *Association Francophone d'Interaction Homme-Machine*, http://www.afihm.org: Michel Beaudouin-Lafon (founding president in 1994, member of the board 2002-present, vice-president since 2002), Nicolas Roussel (member of the board 2001-2003).
- Steering Committee of ETAPS, *European Joint Conference on Theory and Practice of Software*: Marie-Claude Gaudel, member since its creation in 1998.
- Steering Committee of Med-Hoc-Net, *Mediterranean Ad Hoc Networking Workshop*: Khaldoun Al Agha, member since its creation in 2002.
- Steering Committee of MWCN, *IEEE/IFIP International Conference on Mobile and Wireless Communications Networks*: Khaldoun Al Agha, member since 2002.

## **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Duration (months)	Total €
IST	MC Gaudel	DSOS	UE	CEPHYTEN	06/03/00- 05/05/03	38	50000
IST	M. Beaudouin- Lafon	INTERLIVING	UE	CEPHYTEN	01/01/01- 31/12/03	36	314592
IST	M. Beaudouin- Lafon	EUD-NET	UE	UPS/SAIC	01/07/02 - 31/10/03	14	17000
RNTL	M. Beaudouin- Lafon	INDIGO	MENRT	UPS	11/12/01 - 10/12/04	36	177580
RNRT	K. Alagha	SAFARI	Min.Industry	UPS/SAIC	03/02/03 - 02/11/05	33	142645
SUBV	K. Alagha	Réseaux sans fil	MEN	UPS/SUBSE	2002		11434
ACI	K. Alagha	JE réseaux ah-hoc	MENRT	CNRS	19/07/02- 18/07/05	36	24550
ACI	L.Boukatem	Nemo (JC)	MENRT	CNRS	15/09/03- 08/09/06	36	54 347
ACI	N.Roussel	Micromégas (MDD)	MENRT	CNRS	25/07/03- 20/07/06	36	107023
SUBV	K. Alagha	PPF SUPELEC	UPS	UPS/CONTF	2004		5853
industry	B.Marre	CEA	CEA	CEPHYTEN	2000		15244
AS	M. Beaudouin- Lafon	Visualisation d'information	CNRS Stic	CNRS	23/01/01- 22/01/02	12	5945
AS	M. Beaudouin- Lafon	rtp16 interface homme machine	CNRS Stic	CNRS	16/06/03- 15/06/04	12	14000
AS	M. Beaudouin- Lafon	rtp16 plasticite des systèmes	CNRS Stic	CNRS	16/06/03- 15/06/04	12	3 0 0 0
AS	MC Gaudel	Technique avancée	CNRS Stic	CNRS	01/10/01- 30/06/03	12	9146
AS	MC Gaudel	RTP 1 Technique de spécification de test	CNRS Stic	CNRS	25/06/03- 24/06/04	12	5 000

Note: See the glossary for acronyms.

## DSoS

Partners: University of Newcastle (UK), Technische Universität Wien (Austria), LAAS-CNRS, INRIA (Rocquencourt), LRI-Université de Paris-Sud, Ulm Universität (Germany), QINETIC (UK).

The overall objective of the DSoS project was to develop significantly improved means for composing a dependable "system of systems" (SoS) from a set of largely autonomous component computer systems. Reusing existing legacy systems in a dependable fashion without the need for extensive re-engineering is a key problem currently faced by industry.

A system of systems provides new emerging services to its users, in addition to the services provided by its component systems. The goal of this project was to ensure that both types of services are provided with a level of dependability matching specified user requirements. Dependability in this context encompasses reliability, security and maintainability, though the project focused mainly on the first of these system characteristics, with some emphasis on timeliness as well as functionality issues.

project on dependable computing systems funded by the European Union. 36 months (2000-2003). Scientific director for LRI: Marie-Claude Gaudel.

Long term research

#### Research groups

Prog

Contracts and grants

#### InterLiving

Partners: KTH (Sweden, coordinator), INRIA (Rocquencourt), LRI.

The interLiving project (http://interliving.kth.se/) studied and developed, together with families, technologies that facilitate generations of family members living together or at a distance. Its main objectives were: to understand the needs of diverse families, to develop innovative artefacts that support the needs of co-located and distributed families, and to understand the

impact such technologies can have on families. The project main achievements for LRI were the concept of *technology probe* and the MirrorSpace communication technology.

#### EUD-Net

*Partners: University of Bari (Italy, coordinator) and 16 members from France, Germany, Italy, the Netherlands, Sweden and the U.K.* 

The goal of the EUD-NET Network of Excellence (http://giove.cnuce.cnr.it/eud-net.htm) was to help the European Commission prepare a research agenda in the field of end-user development and to increase contacts among highly-qualified research centres, both academic and industrial, in order to speed-up the production of innovative ideas and approaches.

European project (5th framework program, IST Disappearing Computer program). 36 months (2001-2003). Scientific director for LRI: Michel Beaudouin-

Lafon.

European Network of Excellence (5th framework program). 16 months (2002 -2003).

Scientific director for LRI: Michel Beaudouin-Lafon.

Research project funded by national network on software technology (*RNTL*).

36 months (2001-2004).

Scientific director for the project: Michel Beaudouin-Lafon.

Research project funded by the national network for telecommunications reserch (*RNRT*). 30 months (2003-2005). Scientific director for LRI: Khaldoun Al Agha.

#### INDIGO

Partners: LRI (coordinator), ILOG (Gentilly), CENA (Toulouse), W3C (Sophia-Antipolis).

The INDIGO project aims at designing and prototyping a new generation of tools to enable the development of highly interactive distributed graphical applications. The resulting protocols and sample implementations will be submitted to the World-Wide Web consortium for standardization.

#### SAFARI

*Partners: France Telecom R&D, Alcatel, INRIA, LIP6, LRI, LSIIT, LSR-IMAG, SNCF, ENST.* 

The goal of the project (https://safari-rnrt.rd.francetelecom.com/) is to create and provide an infrastructure (protocol and software) for a transparent and auto-configurable access to future multimedia networks based on IPV6. This access must take into account all constraints of existent access networks (fixed, wireless, ad-hoc) and must manage the dynamic behavior of the users.

#### Jeune équipe CNRS

Partners: LRI (coordinator).

The goal of the project is to analyze the use of small and ultramobile laptops and PDAs in ad hoc networks. Routing algorithms are proposed to take into account the capabilities of these miniature hosts.

#### NEMO

Partners: LRI (coordinator), LIP6 (Paris), ENST (Paris).

This project consists in the study of mobility of an entire network (train, bus, aircraft ...) and proposes solutions to optimize resource utilization and the attachment to the Internet. Several algorithms will be proposed in order to improve handover, multicast and resource reservation.

#### **Micromégas**

Partners: LMP (Marseille, coordinator), LRI, INRIA (Rocquencourt and Nancy), Institut Pasteur (Paris).

The goal of the project is to enrich current methods of information visualisation and navigation by exploring new approaches to deal with "familiar data", i.e. data that users know they have but have a hard time accessing. We are studying personal data, e.g., texts, photos, emails manipulated in day to day activities as well as professional data, e.g., scientific data used by biologists.

### CEA

Partners : CEA-LIST, LRI.

The goal of this research contract was to define a specialised tool for generating test data from LUSTRE descriptions. Our expertise on data selection and constraint solving was used for the development by the CEA of the GATeL tool.

Grant from CNRS to support young research group (*Jeune équipe*). 36 months (2002-2004). Scientific director for LRI: Khaldoun Al Agha.

Grant from the French Ministry of Research to support junior researchers (*ACI Jeunes chercheurs*). 36 months (2003-2006). Scientific director for LRI: Lila Boukhatem.

Research groups

Contracts and grants

Prog

Project funded by national initiative on Massive Data (ACI Masses de données). 36 months (2003-2006). Scientific director for LRI: Nicolas Roussel.

## Other contracts

The following contracts were negociated before the contract holder moved to LRI, and were therefore not managed by LRI.

#### ARCADE

Partners: France Telecom R&D, Qosmic, Thales, INRIA, LIP6.

The goal of ARCADE was to offer a generic model for the control of IP networks. For each user, an individual profile is personalized. Then, the attribution of network resources is done dynamically and depends on the user behavior. Problems treated in ARCADE are security, quality of service and mobility.

http://www-rp.lip6.fr/arcade/organisation.html

## SAMU

Partners: Motorola, Cegetel, Renault, Institut Eurécom, PRISM.

SAMU was proposed to optimize services offered by the third generation of mobile systems: UMTS. It consisted of improving the frequency allocation algorithms and mapping quality of service parameters from IP networks to the UMTS radio interface. The service optimization was tested using IP applications in a vehicular environment.

http://samu.crm-paris.com/main.htm

## 13

Partners: INRIA, LIP6, ENST.

I3 studied and implemented algorithms and protocols for the third generation Internet. This generation can adjust and allocate resources on demand. Research project funded by national network for telecommunications research (*RNRT*).

24 months (2001-2002).

Scientific director for INRIA: Khaldoun Al Agha.

Research project funded by national network for telecommunications research (*RNRT*).

30 months (2000-2002).

Scientific director for PRISM: Khaldoun Al Agha.

Project funded by INRIA (*ACI INRIA*), 24 months (2000-2001), scientific coordinator: Khaldoun Al Agha.

## B.3.8 / Prog

## Collaborations

## **Cooperation agreements**

- Collaboration France Tunisia (2004) funded by INRIA (STIC project), coordinator in France: Khaldoun Al Agha. Partners: LRI, LIP6, INRIA, CRISTAL (ENSI Tunisia), UTIC (Sup'Com Tunisia). Modelling and simulation of indoor Wi-Fi and ad-hoc networks.
- Collaboration with France Telecom R&D (September 2003 August 2006). Supervision of a doctoral student, Tinku Mohamed-Rasheed, under a CIFRE agreement. Thesis advisor: Khaldoun Al Agha.
- Collaboration with Thalès Research and Technology (Jun. 2004 May 2007). Supervision of a doctoral student, Jérôme Lard, under a CIFRE agreement. Thesis advisor: Michel Beaudouin-Lafon.

## Collaborations leading to joint publications

- Collaboration between the In Situ team and Yves Guiard, CNRS Research Director at Laboratory for Action and Perception (Marseille), on navigation in multiscale interfaces: [78, 100, 101, 102, 127, 160, 169]. Yves Guiard is now an external collaborator of the In Situ project.
- Collaboration between Hakim Badis and Jaehoon Paul Jeong, Jung-Soo Park (Researchers at ETRI/PEC in Korea), Youn-Hee Han (Researcher at Samsung Advanced Institute of Technology in Korea) and Kenichi Mase (Professor and Dr. Eng at Niigata University in Japan) on Requirements for Ad Hoc IP Address Autoconfiguration Internet-Draft. [199, 200, 201].
- Collaboration between Fatiha Zaïdi and Ana Cavalli, Institut National des Télécommunications (Evry, France), on Passive Testing [5, 10, 81, 82].

## Other collaborations

- Collaboration with the Hipercom project, INRIA-Rocquencourt: Joint Ph.D. students, and joint participation in industrial projects. Khaldoun Al Agha is external member of Hipercom.
- Collaboration between Khaldoun Al Agha, Alia Fourati and Supelec on security in ad-hoc networks.
- Collaboration between Fatiha Zaïdi and Pascale Le Gall, researcher at Université d'Evry, on integration testing.

## B.3.9 / Prog

## Dissemination and technology transfer

## Patents and licences

- Patent on "Communication terminal and system", inventors: Michel Beaudouin-Lafon (LRI), Nicolas Roussel (LRI), Jacques Martin (CSTB), Jean-Dominique Gascuel (INRIA), Georges Buchner (France Telecom), Hervé Lissek (CSTB). Filed by France Telecom in July 2000 (FR2811501), granted in France in June 2003. International extensions filed for Europe, USA, Australia and Japan. This patent is the result of a former joint project with France Telecom on media spaces.
- Patent on "Communication control device", inventors: Wendy Mackay (INRIA), Michel Beaudouin-Lafon (LRI), Nicolas Gaudron (INRIA). Patent filed by INRIA in August 2004 (PCT FR 0402135). This patent is the result of the InterLiving IST project on communication technologies for families (see above).

## Software licensing and distribution

## **CDMA simulator**

Code modulation, authors: Géraud Allard and Khaldoun Al Agha. Free software. http://www.lri.fr/asspro/groupeReseaux/site-simulateur/index.html

#### Qolyester

OLSR-IPV6 protocol for ad hoc networks with QoS, authors: Ignacy Gawedzki and Khaldoun Al Agha. Free Software. http://qolsr.lri.fr

Research groups

## Prog

Collaborations Dissemination and technology transfer

#### **OLSR Simulator**

RFC 3626 simulation model under OPNET, authors: Hakim Badis and Khaldoun Al Agha. http://qolsr.lri.fr

#### VideoSpace

A toolkit for using image streams in interactive and collaborative software, author: Nicolas Roussel. Freely available under the GNU Lesser General Public License (LGPL). http://insitu.lri.fr/~roussel/projects/videoSpace/

#### Ametista

A toolkit for exploring new window management techniques, author: Nicolas Roussel. Freely available under the GNU Lesser General Public License (LGPL).

http://insitu.lri.fr/~roussel/projects/ametista/

## Popularisation of research results

- Marie-Claude Gaudel is a member of the Scientific Board of the Club *Science et Citoyens* in Bobigny-Drancy. She participated in several meetings and the *Journées Science et Citoyens de la région Ile de france* in May 2004.
- MirrorSpace (a work that started as part of the interLiving project) appeared in the national newspaper Libération on 28 February 2003. It was exhibited at the Pompidou Center art museum, the La Villette science museum and several other Contemporary Art galleries in Paris in 2003 and 2004.
- Frédéric Voisin made a presentation about test generation ("Automatisation de la génération de jeux de test") at the 1st Forum Intertronic, Paris, March 2003.
- Marie-Claude Gaudel organised and chaired a panel on the reconfiguration of scientific disciplines for the 2003 congress of the Centre d'Alembert, in Orsay.
- Sandrine Gouraud introduced children (6/7 years) to computer science (hardware, programmation) at school within the framework of the regional project called *projets scientifiques parainnés: un enseignant, un chercheur*: a teacher, a researcher (2001-2003).
- Grégory Lestiennes introduced children (10/11 years) to computer science at school within the framework of the regional project called *projets scientifiques parainnés: un enseignant, un chercheur*: a teacher, a researcher (2001-2002).
- Nicolas Roussel gave a talk in February 2002 to students of an engineering school (EISTI) visiting LRI as members of the Club *Science et Citoyens* from Cergy Pontoise.
- Marie-Claude Gaudel organised a science café (*Bar des Sciences*) on security and safety of computer systems at the French annual science fair (*Fête de la Science*) in 2002.
- Sandrine Gouraud, Nicolas Roussel, Renaud Blanch and Stéphane Conversy presented some of the technologies developped for the interLiving project at the French annual science fair (*Fête de la Science*), ferme du Moulon, Gif-sur Yvette, October 2002.
- Sandrine Gouraud presented SiReMa, a tool used to simulate and analyse Manhattan networks, at the French annual science fair (*Fête de la Science*), ferme du Moulon, Gif-sur Yvette, October 2000.

## Summer schools, tutorials, invited seminars

## International

- Michel Beaudouin-Lafon. "Instrumental Interaction". 6 hours course, *Spring School on Multimodal and Mobile Interfaces*, Anzere, Switzerland, March 2004.
- Khaldoun Al Agha. "WLAN and WPAN : an overview on Wi-FI and bluetooth". 4 hours tutorial, *IFIP MWCN'03: International Workshop On Mobile and Wireless Communications Networks*, Singapore, October 2003.
- Nicolas Roussel. "The next five years in telepresence". Invited panelist for the workshop on Immersive Telepresence, *ACM Multimedia Conference*, December 2002.
- Fatiha Zaïdi. "Interoperability Testing. Séminaire". *ICU laboratories*, Korea, July 2001.
- ♥ Fatiha Zaïdi. "Test Methods for Communication Protocols and Services". 2 hours tutorial, International Conference on Software Engineering Applied to Networking and Parallel/Distributed Computing (SNPD'2000), May 2000, Reims, France.
- Khaldoun Al Agha. "Wireless local area networks: WiFi". 3 hours tutorial at *IFIP Med-Hoc-Net'03, Mediterranean Ad Hoc Networking Workshop,* Mahdia, Tunisia, June 2003.

#### National

- Alia Fourati. "OLSR: Attaques et conséquences".
   OLSR Day, Hypercom seminar, INRIA &LRI, April 2004.
- Davor Males. "Synthèse de la sécurité et de l'ingénierie dans les réseaux sans fil". *Winter school DNAC 2004, De Nouvelles Architectures pour les Communications,* Fortalezza, February 2004.
- Alia Fourati. "La sécurité du routage dans les réseaux ad hoc". Winter school DNAC 2004, De Nouvelles Architectures sur les Réseaux, Fortalezza, February 2004.
- Davor Males. "Sécurité dans les réseaux sans fil et ad hoc". Winter school DNAC 2004, De Nouvelles Architectures sur les Réseaux.
- Khaldoun Al Agha. 3 hours classes, Winter School, DNAC 2002 and 2004, De Nouvelles Architectures pour les Communications.
- Hakim Badis. "QOLSR, comment sélectionner les MPR pour une meilleure QoS". *OLSR Day, Hypercom Seminar*, INRIA & LRI. April 2004.
- Sandrine Gouraud, "Génération de test statistique structurel". *Lande project seminar*, IRISA Rennes, January 2003. - Hakim Badis. "QOLSR: Le protocole de routage OLSR avec QoS". *Hypercom student seminar*, INRIA Rocquencourt, April 2003.
- Nicolas Roussel. "Usages de la vidéo pour la communication médiatisée". *Invited ARTEM seminar*, Nancy, France, December 2003
- Nicolas Roussel. "Usages de la vidéo pour la communication médiatisée". *Tutorial at IHM* 2002, Conférence Francophone d'Interaction Homme-Machine, Poitiers, France, November 2002.
- Michel Beaudouin-Lafon. "L'interaction homme-machine : retour vers le futur". *Colloquium at INRIA Rocquencourt*, 20 March 2001.

Research groups Prog

Dissemination and technology transfer

## Training and education (doctoral and post-doctoral)

Defended habilitations		
Name	Date defended	Current position
AL AGHA Khaldoun	03/10/02	Professor, Université Paris XI
Defended doctorates		
Name	Date defended	Current position
ROUSSEL Nicolas	17/07/00	Assistant Professor, Université Paris XI
CASALTA Didier	22/09/00	Independent consultant
CONVERSY Stéphane	29/09/00	Assistant Professor, ENAC, Toulouse
LETONDAL Catherine	27/09/01	Research engineer, Institut Pasteur, Paris
AL-MESHHADANY Thamer	23/09/03	Lecturer (ATER), Université Paris XI
PERATO Laurent	23/09/03	Lecturer (ATER), Université Paris XI
GOURAUD Sandrine-Dominique	24/06/04	Post-doc, IRISA
BEAUDOUX Olivier	29/06/04	Assistant Professor, ESEO, Angers
BENZAID Mounir	09/07/04	Assistant Professor, ISAIP, Angers

## Graduate courses

- DEA I3 (Information, Interaction, Intelligence), Paris XI:
  - Foundations of Human-Computer Interaction: Michel Beaudouin-Lafon (1999-present).
  - Computer-Supported Cooperative Work, DEA I3: Nicolas Roussel, Michel Beaudouin-Lafon (1999-present).
- DEA ID (Informatique Distribuée) , Paris XI:
  - Mobile Networs: Khaldoun Al Agha, Lila Boukhatem, (2000-present)
  - Test of Computing Systems: Marie-Claude Gaudel and Bruno Marre.

## **Publications**

## International peer-reviewed journals

- [1] N. Achir, M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Active bandwidth sharing for multi-layered video in multi-source environments. *Networking and Information Systems Journal*, 3(1), June 2000.
- [2] K. Al Agha. Resource management in wireless networks using intelligent agents. *International journal of network management, Wiley*, 10(1), January-February 2000.
- [3] K. Al Agha, P. Jacquet, and N. Vvedenskaya. Analysis of a priority stack random access protocol in W-CDMA systems. *IEEE Transactions on Vehicular Technology*, 51(3), May 2002.
- [4] K. Al Agha, G. Pujolle, and D. Zeghlache. VCB: an efficient resource sharing scheme for cellular mobile systems. *Telecommunication Systems, Kluwer*, 19(1):101 - 110, January 2002.
- [5] E. Bayse, A. Cavalli, M. Núñez, and F. Zaïdi. A passive testing approach based on invariants: Application to the wap. *Computer Networks*, 2004. 27 pages, in press.
- [6] M. Benzaid, P. Minet, K. Al Agha, C. Adjih, and G. Allard. Integration mobile-IP and OLSR Ad-Hoc networking for futur wireless mobile systems. ACM WINET Journal on Wireless Networks, Kluwer, 10(4), July 2004.
- [7] L. Boukhatem, A.-L. Beylot, D. Gaiti, and G. Pujolle. Channel allocation with and without handover queuing in LEO satellite systems based on an earth-fixed cell coverage. *IEICE Transactions on Communications*, E86-B(7), July 2003.
- [8] L. Boukhatem, A.-L. Beylot, D. Gaiti, and G. Pujolle. TCRA: A time-based channel reservation scheme for handover requests in LEO satellite system. *International Journal of Satellite Communications and Networking, Wiley*, 21, April 2003.
- [9] L. Boukhatem, D. Gaiti, and G. Pujolle. A time-based reservation scheme for managing handovers in satellite systems. *ACM International Journal of Network Management, Wiley*, 13(2):139 145, March April 2003.
- [10] A. Cavalli, A. Mederreg, and F. Zaïdi. Application of a formal testing methodology to wireless telephony networks. *Journal of the Brazilian Computer Society*, 2004. 17 pages, in press.
- [11] H. Chaouchi and A. Munaretto. Adaptive QoS Management for IEEE 802.11 future Wireless ISPs. ACM WINET Journal, 10(4), Jul 2004, Kluwer Academic Publishers. Special Issue on Ad Hoc Networking.
- [12] M. Dias de Amorim and O. M. Duarte. Improving the quality specification of multi-layered applications. *Networking and Information Systems Journal*, 3(6), December 2000.
- [13] M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Virtual layering and efficient merging in non-cooperative multicast trees. *Journal of the Brazilian Computer Society*, 7(2), March 2001.
- [14] M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Improving user satisfaction in adaptive multicast video. *IEEE/KICS Journal of Communications and Networks*, 4(2), September 2002.
- [15] M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Distinguishing video quality through differential matrices. *ACM Multimedia Systems Journal*, 9(1), July 2003.
- [16] Y. Guiard and M. Beaudouin-Lafon. Target acquisition in multiscale electronic worlds. *International Journal of Human Computer Studies (IJHCS),* Elsevier. in press.
- [17] N. Roussel, H. Evans, and H. Hansen. Proximity as an interface for video communication. *IEEE Multimedia*, 11(3):12-16, July-September 2004.

Research groups

#### Prog Training and Education Publications

## National peer-reviewed journals

- [18] M. Beaudouin-Lafon. Ceci n'est pas un ordinateur. Perspectives sur l'Interaction Homme-Machine. *Technique et Science Informatique*, 19(1-2-3):69-74, 2000.
- [19] O. Beaudoux. Un modèle de composants (inter)actifs centré sur les documents. *Revue Information, Interaction, Intelligence*, 4(1):41-58, 2004, Cépadués éditions.
- [20] C. Cavalli, A. Mederreg, F. Zaïdi, P. Combes, W. Monin, R. Castanet, M. MacKaya, and P. Laurençot. Une plate-forme de validation multi-services et multi-protocoles - résultats d'expérimentations. *Annales des Télécommunications*, 2004. 24 pages, in press.
- [21] M.-C. Gaudel. Méthodes formelles, preuves et tests. *Technique et Science Informatique*, 19(1):239-246, 2000, Hermès Science.
- [22] M. Hascoët and M. Beaudouin-Lafon. Recherche et visualisation d'information. *Revue Information-Interaction-Intelligence*, 1(1):77-108, 2001, Cépaduès-éditions.

## Other journals

- [23] K. Al Agha. La qualité de service dans les réseaux ad hoc. *Plein Sud Spécial recherche ; université de Paris-Sud XI*, 2004.
- [24] K. Al Agha and G. Pujolle. L'alternative IP Mobile et IP Cellulaire. *Informatique Professionnelle, Gartner,* (213), April 2003.

## Books

- [25] K. Al Agha. Les Réseaux mobiles et sans fil. Hermes traité, 2004. in press.
- [26] K. Al Agha, G. Pujolle, and G. Vivier. Réseaux de mobiles et Réseaux sans fil. Eyrolles, 2001.
- [27] D. Males and G. Pujolle. Wi-Fi par la pratique, 1ère édition. Eyrolles, 2002.
- [28] D. Males and G. Pujolle. Wi-Fi par la pratique, 2è édition. Eyrolles, 2004.

## **Edited books**

- [29] M. Beaudouin-Lafon, editor. Actes 14<sup>e</sup> conférence francophone sur l'Interaction Homme-Machine (IHM 2002), ACM International Conference Proceedings Series, Poitiers, France, Nov. 2002. ACM Press. 306 pages.
- [30] Y. Guiard and M. Beaudouin-Lafon, editors. *Fitts' law 50 years later: Applications and contributions from HCI*, Special issue of International Journal of Human Computer Studies (IJHCS). Elsevier, dec 2004.
- [31] J. Jacko, A. Sears, M. Beaudouin-Lafon, and R. Jacob, editors. Proceedings of ACM Conference on Human Factors in Computing Systems, CHI 2001, volume 3(1) of CHI Letters, Seattle, USA, Apr. 2001. ACM Press. 560 pages.
## **Book chapters**

- [32] N. Achir, M. Dias de Amorim, O. M. Duarte, and G. Pujolle. *Video Data*. Hermes-Penton Science, 2002.
- [33] H. Badis and K. Al Agha. *Advances in Wireless Networks and Mobile Computing*. Kluwer Academic Publishers, 2004. in press.
- [34] M. Beaudouin-Lafon. Contexte et interaction : vers l'informatique située. In H. Prade, R. Jeansoulin, and C. Garbay, editors, *Le Temps, l'Espace et l'Evolutif en Sciences du Traitement de l'Information*. Cépaduès Editions, 2000. Ecole thématique Nouveaux défis en sciences de l'information : documents et évolution, GDR-PRC I3, Marseille, septembre 2000.
- [35] M. Beaudouin-Lafon and W. Mackay. Prototyping development and tools. In J. Jacko and A. Sears, editors, *Human Computer Interaction Handbook*, pages 1006-1031. Lawrence Erlbaum Associates, 2002.
- [36] J.-D. Fekete. Le format pdf. In Techniques de l'ingénieur. Techniques de l'ingénieur, 2004.
- [**37**] J.-D. Fekete. Le langage postscript. In *Techniques de l'ingénieur*. Techniques de l'ingénieur, 2004.
- [38] Y. Ghamri Doudane, A. Munaretto, and H. Chaouchi. Qualité de service dans les réseaux ieee 802.11. In *Traité IC2 : Internet ambiant*. Eds. Hermes, 2004. in press.
- [**39**] A. Munaretto. Qualité de service dans les réseaux ad hoc. In *Traité IC2 : Réseau de mobiles et sans fil*. Eds. Hermes, 2004. in press.
- [40] N. Roussel. Web-Based Mediaspace. In B. Furht, editor, *Handbook of Internet Computing*, chapter 9, pages 205-226. CRC Press, Boca Raton, Florida, June 2000.
- [41] B. Shneiderman, C. Plaisant, and J.-D. Fekete. *Designing the User Interface*, chapter Software Tools. Addison Wesley, 4th edition, May 2004.

## Major international peer-reviewed conferences

- [42] C. Adjih, K. Al Agha, F. Dumontet, P. Jacquet, A. López, and L. Viennot. Quality of service aspect for BRAIN architecture. In *IEEE PIMRC'00: Personal Indoors and Mobile Radio Communications*, London, UK, September 2000. IEEE Publisher.
- [43] K. Al Agha, T. Al Meshhadany, and L. Perato. Resource allocation in W-CDMA using mobile agents. In *IST'01: International Symposium on Telecommunications*, Teheran, IRAN, September 2001.
- [44] K. Al Agha and K. Boussetta. Dynamic slot allocation for multicasting in GPRS systems. In *IEEE VTC'00-Spring: Vehicular Technology Conference*, Tokyo, Japan, May 2000. IEEE Publisher.
- [45] K. Al Agha, P. Jacquet, and N. Vvedenskaya. W-CDMA random access with priority resolution. In *IEEE WCNC'00: Wireless Communications and Networking Conference*, Chicago, USA, September 2000. IEEE Publisher.
- [46] K. Al Agha, T. A. Meshhadany, and L. Perato. Resource allocation based on handover prediction in WCDMA. In *IEEE VTC'02-Fall: Vehicular Technology Conference*, Vancouver, Canada, September 2002. IEEE Publisher.
- [47] K. Al Agha and L. Viennot. Spatial reuse in wireless LAN networks. In *IFIP PWC'01: Personal Wireless Communications*, Lapeenranta, Finland, August 2001. Kluwer.
- [48] T. Al Meshhadany and K. Al Agha. A new code allocation scheme for UMTS system. In *IEEE VTC'01-Fall: Vehicular Technology Conference*, Atlantic City, USA, October 2001. IEEE Publisher.

Research groups

- [49] T. Al Meshhadany and K. Al Agha. OSSC: A new scheme for code allocation in WCDMA. In *IEEE PIMRC'01: Personal Indoors and Mobile Radio Communications*, San Diego, USA, September 2001. IEEE Publisher.
- [50] T. Al Meshhadany and K. Al Agha. Analysis of the code allocation mechanism in the WCDMA. In *IEEE MWCN'02: International Workshop On Mobile and Wireless Communications Networks*, Stockholm, Sweden, September 2002.
- [51] T. Al Meshhadany and K. Al Agha. VCB by mean of soft2hard handoff in WCDMA. In *IEEE MWCN'02: International Workshop On Mobile and Wireless Communications Networks*, Stockholm, Sweden, September 2002.
- [52] C. Appert, M. Beaudouin-Lafon, and W. Mackay. Context matters: Evaluating interaction techniques with the CIS model. In *People and Computers XVIII - Design for Life -Proceedings of HCI 2004, Leeds, UK*, pages 279-295. Springer Verlag, Sept. 2004.
- [53] M. Ayari, F. Kamoun, and D. Males. Towards a policy-based management for ad hoc networks. In *IFIP Med-Hoc-Net'04*, Bodrum, Turkey, June 2004.
- [54] H. Badis and K. Al Agha. A distributed algorithm for multiple-metric link state QoS routing problem. In *IFIP MWCN'03: International Workshop On Mobile and Wireless Communications Networks*, Singapore, October 2003.
- [55] H. Badis and K. Al Agha. An efficient mobility management in wireless overlay networks. In *IEEE PIMRC'03: Personal Indoors and Mobile Radio Communications*, Beijing, China, September 2003.
- [56] H. Badis and K. Al Agha. Scalable model for the simulation of OLSR and Fast-OLSR protocols. In *IFIP Med-Hoc-Net'03*, Mahdia, Tunisia, June 2003.
- [57] H. Badis and K. Al Agha. An efficient QOLSR extension protocol for QoS in ad hoc networks. In *IEEE VTC'04-Fall: Vehicular Technology Conference*, Los Angeles, USA, September 2004.
- [58] H. Badis and K. Al Agha. Fast and efficient vertical handoffs in wireless overlay networks. In *IEEE PIMRC'04: Personal Indoors and Mobile Radio Communications*, Barcelona, Spain, September 2004.
- [59] H. Badis and K. Al Agha. Qolsr multi-path routing for mobile ad hoc networks based on multiple metrics. In *IEEE VTC'04-Spring: Vehicular Technology Conference*, Milan, Italy, May 2004.
- [60] H. Badis and K. Al Agha. Reducing vertical handoff latency in wireless overlay networks. In ICMU'04: International Conference on Mobile Computing and Ubiquitous Networking, Yokosuka, Japan, January 2004.
- [61] H. Badis, I. Ben Hamida, L. Boukhatem, and K. Al Agha. An ad hoc networking for intermonet using mobile IPv6 and OLSR. In *IEEE VTC'04-Fall: Vehicular Technology Conference*, Los Angeles, USA, September 2004.
- [62] H. Badis, I. Gawedzki, and K. Al Agha. Qos routing in ad hoc networks using QOLSR with no need of explicit reservation. In *IEEE VTC'04-Fall: Vehicular Technology Conference*, Los Angeles, USA, September 2004.
- [63] H. Badis, A. Munaretto, K. Al Agha, and G. Pujolle. QoS for ad hoc networking based on multiple-metric: Bandwidth and delay. In *IFIP MWCN'03: International Workshop On Mobile and Wireless Communications Networks*, Singapore, October 2003.
- [64] H. Badis, A. Munaretto, K. Al Agha, and G. Pujolle. Optimal path selection on a link state QoS routing. In *IEEE VTC'04-Spring: Vehicular Technology Conference*, Milan, Italy, May 2004.
- [65] M. Beaudouin-Lafon. Instrumental interaction: an interaction model for designing post-WIMP user interfaces. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI 2000)*, volume 2(1) of *CHI Letters*, pages 446-453, The Hague, The Netherlands, 2000. ACM Press.

- [66] M. Beaudouin-Lafon. Novel interaction techniques for overlapping windows. In *Proc. ACM Symposium on User Interface Software and Technology, UIST 2001, Orlando (USA),* volume 3(2) of *CHI Letters,* pages 153-154. ACM Press, Nov. 2001.
- [67] M. Beaudouin-Lafon. Designing interfaction, not interfaces. In Proc. Conference on Advanced Visual Interfaces, AVI 2004, pages 15-22, Gallipoli (Italy), May 2004. ACM Press. Invited keynote address.
- [68] M. Beaudouin-Lafon and H. Lassen. The architecture and implementation of CPN2000, a post-WIMP graphical application. In *Proc. ACM Symposium on User Interface Software and Technology (UIST 2000)*, volume 2(2) of *CHI Letters*, pages 181-190, San Diego, USA, Nov. 2000. ACM Press.
- [69] M. Beaudouin-Lafon and W. Mackay. Reification, polymorphism and reuse: Three principles for designing visual interfaces. In *Proc. International Conference on Advanced Visual Interfaces (AVI 2000)*, pages 102-109, Palermo, Italie, May 2000. ACM Press.
- [70] M. Beaudouin-Lafon, W. Mackay, P. Andersen, P. Janecek, M. Jensen, M. Lassen, K. Lund, K. Mortensen, S. Munck, A. Ratzer, K. Ravn, S. Christensen, and K. Jensen. CPN/Tools: A post-WIMP interface for editing and simulating coloured Petri nets. In J.-M. Colom and M. Koutny, editors, *Proc. 22<sup>nd</sup> International Conference on Application and Theory of Petri Nets (ICATPN'2001)*, Lecture Notes in Computer Science, pages 71-80. Springer-Verlag, June 2001.
- [71] O. Beaudoux and M. Beaudouin-Lafon. DPI: A conceptual model based on documents and interaction instrument. In *People and Computers XV - Interaction without frontiers* (*Joint proceedings of HCI 2001 and IHM 2001, Lille, France*), pages 247-263. Springer Verlag, Sept. 2001.
- [72] I. Ben Hamida, L. Boukhatem, and K. Al Agha. A multicast-based solution for mobile-IPv6 extension in mobile networks. In *CSN'04: Communication Systems and Networks*, Marbella, Spain, September 2004. IASTED.
- [73] I. Ben Hamida, L. Boukhatem, and K. Al Agha. A mobile-IPv6 extension with multicast for nested mobile networks. In *IEEE VTC'04-Fall: Vehicular Technology Conference*, Los Angeles, USA, September 2004.
- [74] M. Benzaid, P. Minet, and K. Al Agha. A framework for integrating mobile-IP and OLSR ad-hoc networking for future wireless mobile systems. In *IFIP Med-Hoc-Net'02*, Sardinia, Italy, September 2002.
- [75] M. Benzaid, P. Minet, and K. Al Agha. Integrating fast mobility in the OLSR routing protocol. In *IEEE MWCN'02: International Workshop On Mobile and Wireless Communications Networks*, Stockholm, Sweden, September 2002.
- [76] M. Benzaid, P. Minet, and K. Al Agha. Analysis and simulation of Fast-OLSR. In *IEEE VTC'03-Fall: Vehicular Technology Conference*, Jeju, Korea, April 2003. IEEE Publisher.
- [77] M. Benzaid, P. Minet, and K. Al Agha. Performance evaluation of the implementation integrating mobile-IP and OLSR in full-ip networks. In *IEEE WCNC'04: Wireless Communications and Networking Conference*, Atlanta, USA, March 2004.
- [78] R. Blanch, Y. Guiard, and M. Beaudouin-Lafon. Semantic pointing: Improving target acquisition with control-display ratio adaptation. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI2004)*, volume 6(1) of *CHI Letters*, pages 519-526. ACM Press, Apr. 2004.
- [79] L. Boukhatem, A.-L. Beylot, D. Gaïti, and G. Pujolle. TCRA : A resource reservation scheme for handover issue in LEO satellite systems. In *WCNC 2003, IEEE Wireless Communications and Networking Conference*, New Orleans, Louisiana, USA, March 2003.
- [80] L. Boukhatem, G. Pujolle, and D. Gaïti. Handover schemes for LEO satellite systems. In IEEE WPMC'02: International Symposium on Wireless Personal Multimedia Communications, Hawaii, USA, October 2002.

Research groups

- [81] A. Cavalli, S. Maag, S. Papagiannaki, G. Vergiakis, and F. Zaïdi. A Testing Methodology for an Open Software E-Learning Platform. In 18th IFIP World Computer Congress -TC10/WG10.5 Edutech Worskshop, pages 17-32, Toulouse, France, august 2004. Kluwer Academic Publishers.
- [82] A. Cavalli, A. Mederreg, and F. Zaïdi. Application of formal testing methodology to wireless telephony Networks. In 2nd International Information and Telecommunications Technologies Symposium, I2TS'2003, Florianopolis, Brazil, october 2003.
- [83] A. Cavalli, A. Mederreg, F. Zaïdi, P. Combes, W. Monin, R. Castanet, M. MacKaya, and P. Laurençot. A multi-services and multi-protocol validation platform - experimentations results. In R. Hierons and R. Groz, editors, *The 16th IFIP International Conference on Testing of Communication Systems*, pages 165-174, Oxford, march 2004. Lectures Notes in Computer Science.
- [84] P. Costa-Cunha. The I-Book: a freely interpretive notebook for biologists. In *Proceedings* of *PDC 2004, Volume II*, pages 159-162, Toronto, Canada, July 2004. Computer Professionals for Social Responsability.
- [85] P. da S. Gonçalves, M. Dias de Amorim, J. Ferreira de Rezende, and O. M. Duarte. QoS mechanisms for multicast applications. In *International Conference in Telecommunication Systems, Modeling, and Analysis,* Monterey, CA, USA, October 2002.
- [86] A. Denise, M.-C. Gaudel, and S. Gouraud. A generic method for statistical testing. In Proceedings of 15th Int. Symposium on Software Reliability Engineering (ISSRE 2004), Saint-Malo, F, Nov. 2004. IEEE. 15 pages, in press.
- [87] M. Dias de Amorim, O. M. Duarte, and G. Pujolle. An improved mpeg behavioral analysis with autonomous parameter-extracting algorithm for strict video applications. In *IEEE International Conference on Communications*, New Orleans, Louisiana, USA, June 2000. IEEE Publisher.
- [88] M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Application-aware multicast. In *IEEE Globecom*, San Antonio, Texas, USA, November 2001. IEEE.
- [89] P. Dragicevic and J.-D. Fekete. The input configurator toolkit: Towards high input adaptability in interactive applications. In *Proc. Conference on Advanced Visual Interfaces, AVI* 2004, pages 244-247, Gallipoli (Lecce), Italy, May 2004. ACM Press.
- [90] P. Dragicevic and J.-D. Fekete. Support for input adaptability in the icon toolkit. In *Proceedings of the Sixth International Conference on Multimodal Interfaces (ICMI 2004)*, page in press, Penn State University, PA, Oct. 2004. ACM Press.
- [91] J.-D. Fekete. The infovis toolkit. In *Proceedings of the 10<sup>th</sup> IEEE Symposium on Information Visualization (InfoVis 04)*, page in press, Austin, TX, Oct. 2004. IEEE Press.
- [92] J.-D. Fekete and C. Plaisant. Interactive information visualization of a million items. In Proc. IEEE Symposium on Information Visualization 2002 (InfoVis 2002), pages 117-124, Boston, USA, Oct. 2002. IEEE Press.
- [93] M.-C. Gaudel. Testing from formal specifications, a generic approach. In *Proc. ADA-Europe Conference*, volume 2043 of *Lecture Notes in Computer Science*, pages 35-48, Louvain, Belgium, May 2001. Springer Verlag. Invited keynote address.
- [94] M.-C. Gaudel and A. Zamulin. Imperative algebraic specifications. In *Proceedings of PSI-99*, volume 1755, pages 17-39, Novossibirsk, RU, 2000. Springer verlag. Invited keynote address.
- [95] S. Ghannay, S. Mettali, F. Kamoun, and D. Males. The monitoring of ad hoc networks based on routing. In *IFIP Med-Hoc-Net'04*, Bodrum, Turkey, June 2004.
- [96] M. Ghoniem and J.-D. Fekete. A comparison of the readability of graphs using nodelink and matrix-based representations. In *Proceedings of the 10th IEEE Symposium on Information Visualization (InfoVis 04)*, page in press, Austin, TX, Oct. 2004. IEEE Press.

- [97] M. Ghoniem, N. Jussien, and J.-D. Fekete. Visualizing explanations to exhibit dynamic structure in constraint problems. In *Proc. 3th International Workshop on User Interaction in Constraint Satisfaction*, Kinsale, Ireland, Sept. 2003.
- [98] M. Ghoniem, N. Jussien, and J.-D. Fekete. Visexp: visualizing constraint solver dynamics using explanations. In *Proc. of the 17th International Conference of the Florida Artificial Intelligence Research Society (FLAIRS 2004).* AAAI Press, May 2004.
- [99] S.-D. Gouraud, A. Denise, M.-C. Gaudel, and B. Marre. A new way of automating statistical testing methods. In *Automated Software Engineering Conference (A.S.E.), IEEE*, pages 5-12, 2001.
- [100] Y. Guiard, M. Beaudouin-Lafon, J. Bastin, D. Pasveer, and S. Zhai. View size and pointing difficulty in multi-scale navigation. In *Proc. Conference on Advanced Visual Interfaces, AVI 2004*, pages 117-124, Gallipoli (Italy), May 2004. ACM Press.
- [101] Y. Guiard, R. Blanch, and M. Beaudouin-Lafon.Object pointing: a complement to bitmap pointing in GUIs. In *Proceedings of Graphics Interface 2004 (GI2004)*, pages 9-16. the Canadian Human-Computer Communications Society and A K Peters Ltd., May 2004.
- [102] Y. Guiard, F. Bourgeois, D. Mottet, and M. Beaudouin-Lafon. Beyond the 10-bit barrier: Fitts' law in multi-scale electronic worlds. In *People and Computers XV - Interaction without frontiers (Joint proceedings of HCI 2001 and IHM 2001, Lille, France),* pages 573-587. Springer Verlag, Sept. 2001. Best long paper award.
- [103] S. Huot, C. Dumas, P. Dragicevic, J.-D. Fekete, and G. Hégron. The magglite post-wimp toolkit: Draw it, connect it and run it. In *Proceedings of ACM Symposium on User Interface Software and Technology (UIST 2003)*, volume 5(2) of *CHI Letters*, page in press, Sante Fe, NM, Oct. 2004. ACM Press.
- [104] H. Hutchinson, W. Mackay, B. Westerlund, B. Bederson, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, B. Eiderbäck, S. Lindquist, and Y. Sundblad. Technology probes: Inspiring design for and with families. In *Proceedings of ACM CHI 2003 Conference on Human Factors in Computing Systems*, volume 5(1) of *CHI Letters*, pages 17-24. ACM Press, 2003.
- [105] M. A. Kaafar, L. Benazzouz, F. Kamoun, and D. Males. A kerberos-based authentication architecture for wireless LANs. In *IFIP Networking'04*, Athens, Greece, May 2004.
- [106] G. Lestiennes and M.-C. Gaudel. Testing Processes from Formal Specification with Inputs, Outputs and Data Types. In *Proceedings of 13th IEEE International Symposium on Software Reliability Engineering, Annapolis, Maryland*, pages 3-14, 2002.
- [107] C. Letondal and W. Mackay. Participatory programming and the scope of mutual responsibility: Balancing scientific, design and software commitment. In *Proc. 8th biennial Participatory Design Conference (PDC 2004)*, pages 31-41, Toronto (Canada), July 2004.
- [108] W. Mackay. The interactive thread: Exploring methods for multi-disciplinary design. In Proc. ACM Conference on Designing Interactive Systems (DIS 2004), pages 103-112, Cambridge (USA), Aug. 2004. ACM Press.
- [109] W. Mackay, C. Letondal, G. Pothier, K. Bøegh, and H. Sørensen. The missing link: Augmenting biologist's laboratory notebooks. In *Proc. ACM Symposium on User Interface Software and Technology (UIST 2002)*, volume 4(2) of *CHI Letters*, pages 41-50, Paris, France, Oct. 2002. ACM Press.
- [110] B. Marre and A. Arnould. Test sequences generation from lustre descriptions: Gatel. In Proc. Fifteenth IEEE Int. Conf. on Automated Software Engineering (ASE 2000), pages 229-237. IEEE, Sept. 2000.
- [111] D. Martin, J.-D. Fekete, and J.-C. Torres. Flattening 3D objects using silhouettes. *Computer Graphics Forum*, 21(3):239-248, Sept. 2002.
- [112] A. Munaretto, H. Badis, K. Al Agha, and G. Pujolle. A link-state QoS routing protocol for ad hoc networks. In *IEEE MWCN'02: International Workshop On Mobile and Wireless Communications Networks*, Stockholm, Sweden, September 2002.

Research groups

- [113] A. Munaretto, H. Badis, K. Al Agha, and G. Pujolle. QoS-enhanced OLSR protocol for mobile ad hoc networks. In ANWIRE 1st International Workshop, Glasgow, Scotland, April 2003.
- [114] A. Munaretto, M. Fonseca, K. Al Agha, and G. Pujolle. Fair time sharing protocol: a solution for ieee 802.11b hot spots. In *IFIP-IEEE ICT 2004*, LNCS. Springer-Verlag, August 2004. Fortaleza, Brazil.
- [115] A. Munaretto, M. Fonseca, K. Al Agha, and G. Pujolle. Virtual time synchronization for multimedia ad hoc networks. In *IEEE VTC'04-Fall: Vehicular Technology Conference*, September 2004. Los Angeles, USA.
- [116] L. Perato and K. Al Agha. Mobile agents for improving the web access in the UMTS system. In *IEEE'VTC 01-Fall: Vehicular Technology Conference*, Atlantic City, USA, October 2001. IEEE Publisher.
- [117] L. Perato and K. Al Agha. Reduction of signaling on cellular systems with mobile agents: the POP3 application. In *IEEE PIMRC'01: Personal Indoors and Mobile Radio Communications*, San Diego, USA, September 2001. IEEE Publisher.
- [118] L. Perato and K. Al Agha. Handover prediction for cellular systems in a multi-services context. In *IEEE WPMC'02: International Symposium on Wireless Personal Multimedia Communications*, Hawaii, USA, October 2002.
- [119] L. Perato and K. Al Agha. Handover prediction: User approach versus cell approach. In *IEEE MWCN'02: International Workshop On Mobile and Wireless Communications Networks*, Stockholm, Sweden, September 2002.
- [120] L. Perato and K. Al Agha. Web access for the UMTS air interface by using mobile agents. In *IEEE WCNC'02: Wireless Communications and Networking Conference*, Orlando, USA, March 2002. IEEE Publisher.
- [121] E. Petitjean. Designing timed test cases from region graphs. In *Proceedings of the Workshop* On Testing Real-Time and Embedded Systems WTRTES'03 (Pisa, Italy), Sept. 2003.
- [122] N. Roussel. Experiences in the design of the well, a group communication device for teleconviviality. In Proc. of ACM Multimedia 2002, pages 146-152. ACM Press, Dec. 2002.
- [123] N. Roussel. Ametista: a mini-toolkit for exploring new window management techniques. In *Proceedings of CLIHC 2003, Latin American Conference on Human-Computer Interaction*, pages 117-124. ACM Press, Aug. 2003.
- [124] N. Roussel, H. Evans, and H. Hansen. MirrorSpace: using proximity as an interface to video-mediated communication. In A. Ferscha and F. Mattern, editors, *Proceedings of Pervasive 2004, the second international conference on Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 345-350. Springer, Apr. 2004.
- [125] G. Vivier, C. Bonnet, and K. Al Agha. SAMU project : Tests and improvements of UMTS for users in a car. In *IEEE VTC'02-Spring: Vehicular Technology Conference*, Birmingham, Alabama, May 2002. IEEE Publisher.
- [126] B. Westerlund, S. Lindquist, W. Mackay, and Y. Sundblad. Co-designing methods for designing with and for families. In *European Academy of Design Conference, EAD'03*, Barcelona, Spain, 2003.
- [127] S. Zhai, S. Conversy, M. Beaudouin-Lafon, and Y. Guiard. Human on-line response to target expansion. In *Proc. ACM Conference on Human Factors in Computing Systems (CHI 2003)*, volume 5(1) of *CHI Letters*, pages 177-184, Fort Lauderdale (USA), Apr. 2003. ACM Press.

## National peer-reviewed conferences

- [128] N. Achir, M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Gestion de la vidéo dans un environnement multisource. In *JDIR*, Paris, France, November 2000.
- [129] G. Allard and K. Al Agha. Analysis of the OSSC mechanism in a non-synchronous transmission environment. In *AlgoTel 2003*, Banyuls-sur-mer, France, May 2003. INRIA.
- [130] H. Badis and K. Al Agha. Un modèle générique pour la simulation du protocole de routage OLSR. In *DNAC'02 : De Nouvelles Architectures pour les Communications*, Paris, France, December 2002.
- [131] O. Beaudoux. DoPIdom : Une boîte à outils pour la conception d'interfaces centrées sur les documents XML. In Actes de la conférence francophone sur l'Interaction Homme-Machine (IHM 2004). ACM, International Conference Proceedings Series, 2004. 4 pages, in press.
- [132] C. Besse, A. Cavalli, and F. Zaïdi. Génération de tests d'interopérabilité pour le protocole TCP. In 4<sup>e</sup> Colloque Francophone sur la Modélisation des Systèmes Réactifs, Metz, October 2003. Hermès. Préparé au LAMI.
- [133] R. Blanch. Programmer l'interaction avec des machines à états hiérarchiques. In Actes des 14<sup>e</sup> journées francophones sur l'Interaction Homme Machine (IHM 2002), Poitiers, ACM International Conference Proceedings Series, pages 129-136. ACM Press, Nov. 2002.
- [134] L. Boukhatem. Les réseaux NEMO. In *DNAC'04 : De Nouvelles Architectures pour les Communications*, Fortaleza, Brazil, February 2004.
- [135] S. Conversy, N. Roussel, H. Hansen, H. Evans, M. Beaudouin-Lafon, and W. Mackay. Partager les images de la vie quotidienne et familiale avec videoProbe. In *Proceedings of IHM 2003*, pages 228-231, Caen, France, Novembre 2003. ACM, International Conference Proceedings Series.
- [136] P. Costa-Cunha and W. Mackay. Papier augmenté et stylo Anoto. In *Proceedings of IHM 2003*, pages 232-235, Caen, France, Novembre 2003. ACM, International Conference Proceedings Series.
- [137] A. Crasson and J.-D. Fekete. Structuration des manuscrits : Du corpus à la région. In *Proceedings of CIFED 2004*, pages 162-168, La Rochelle, France, June 2004.
- [138] P. da S. Gonçalves, M. Dias de Amorim, J. Ferreira de Rezende, and O. M. Duarte. Agregação e supressão de feedback de grupos multiponto heterogêneos (portuguese). In *Workshop en Quality of Service (WQoSM)*, Angra dos Reis, RJ, Brazil, November 2002.
- [139] M. Dias de Amorim, O. M. Duarte, and G. Pujolle. Otimização da qualidade de vídeo em redes multicast não cooperativas. In *Brazilian Symposium on Computer Networks*, Florianópolis, SC, Brazil, May 2001.
- [140] J.-D. Fekete. La boîte à outils infovis. In *Proceedings of IHM 2004*, International Conference Proceedings Series, pages 69-76, Namur, Belgium, Sept. 2004. ACM Press.
- [141] J.-D. Fekete and C. Plaisant. Les leçons tirées des deux compétitions de visualisation d'information. In *Proceedings of IHM 2004*, International Conference Proceedings Series, pages 7-12, Namur, Belgium, Sept. 2004. ACM Press.
- [142] M.-C. Gaudel. Problems and methods for testing infinite state machines. In *VI Workshop de Métodos Formais (WMF 2003)*, volume 95 of *Electronic Notes in Theoretical Computer Science*, Campina Grande, BR, Oct. 2003. Elsevier Science Publishers.
- [143] M. Ghoniem and J.-D. Fekete. Visualisation de graphes de co-activité par matrices d'adjacence. In Actes 14<sup>e</sup> conférence francophone sur l'Interaction Homme-Machine (IHM 2002), pages 279-282, Poitiers, France, Oct. 2002. ACM Press.
- [144] M. Ghoniem and J.-D. Fekete. Visualisation matricielle des graphes et manipulation directe de hiérarchies de clusters (demonstration). In Actes 15<sup>e</sup> conférence francophone sur l'Interaction Homme-Machine (IHM 2003), International Conference Proceedings Series,

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

pages 206-207, Caen, France, Nov. 2003. ACM Press.

- [145] M. Ghoniem and J.-D. Fekete. Comparaison de la lisibilité des graphes en représentation noeuds-liens et matricielle. In *Proceedings of IHM 2004*, International Conference Proceedings Series, pages 77-84, Namur, Belgium, Sept. 2004. ACM Press.
- [146] S.-D. Gouraud. Application de la génération aléatoire de structures combinatoires au test de logiciel. In *Approches Formelles dans l'Assistance au Développement de Logiciels (AFADL)*, pages 99-112, 2001.
- [147] W. Mackay. The missing link: Integrating paper and electronic documents (keynote address). In Actes 15<sup>e</sup> conférence francophone sur l'Interaction Homme-Machine (IHM 2003), International Conference Proceedings Series, pages 1-8, Caen, France, Nov. 2003. ACM Press. Conférence invitée.
- [148] B. Marre and A. Arnould. Génération automatique de séquences de tests à partir de descriptions lustre : Gatel. In proceedings Journées Francophones sur les Langages Applicatifs 2000, Lille, France). INRIA, Jan. 2000. invited lecture.
- [149] A. Munaretto, H. Badis, K. Al Agha, and G. Pujolle. QOLSR : Routage avec QoS dans OLSR. In *AlgoTel 2003*, Banyuls-sur-mer, France, May 2003. INRIA.
- [150] E. Nars. Communiquer par groupes avec Circa. In *Proceedings of IHM 2004*, Namur, Belgique, Août 2004. ACM, International Conference Proceedings Series. 4 pages, in press.
- [151] G. Pujolle, K. Al Agha, and N. Boukhatem. Le projet I3 ou la maîtrise de la nouvelle génération Internet. In *DNAC'00 : De Nouvelles Architectures pour les Communications*, Paris, France, November 2000.
- [152] N. Roussel. Le puits : un dispositif de communication audio-vidéo pour la téléconvivialité. In Actes des 14<sup>e</sup> journées francophones sur l'Interaction Homme Machine (IHM 2002), Poitiers, pages 227-230, Novembre 2002.
- [153] N. Roussel. VideoWorkspace : une boîte à outils pour l'exploration de nouvelles techniques de gestion de fenêtres. In Actes des 14<sup>e</sup> journées francophones sur l'Interaction Homme Machine (IHM 2002), Poitiers, pages 271-274, Novembre 2002.
- [154] N. Roussel, H. Evans, and H. Hansen. Utilisation de la distance comme interface à un système de communication vidéo. In *Proceedings of IHM 2003*, pages 268-271, Caen, France, Novembre 2003. ACM, International Conference Proceedings Series.

## Other conferences and workshops

- [155] C. Appert. CIS : Evaluer les Techniques d'Interaction en Contexte. In *Annexes des Actes d'IHM 2003*, pages 23-25, Caen, France, Novembre 2003.
- [156] M. Beaudouin-Lafon and W. Mackay. Research directions in situated computing. In Extended Abstracts, ACM Conference on Human Factors in Computing Systems (CHI 2000), page 369, The Hague, The Netherlands, 2000. ACM Press. Workshop.
- [157] M. Beaudouin-Lafon, W. Mackay, P. Andersen, P. Janecek, M. Jensen, M. Lassen, K. Lund, K. Mortensen, S. Munck, A. Ratzer, K. Ravn, S. Christensen, and K. Jensen. CPN/Tools: Revisiting the desktop metaphor with post-WIMP interaction techniques. In *Extended Abstracts, ACM Human Factors in Computing Systems, CHI 2001*, pages 11-12. ACM Press, Apr. 2001. Demonstration.
- [158] M. Beaudouin-Lafon, W. Mackay, P. Andersen, P. Janecek, M. Jensen, M. Lassen, K. Lund, K. Mortensen, S. Munck, A. Ratzer, K. Ravn, S. Christensen, and K. Jensen. A tool for editing and simulating coloured Petri nets. In J.-M. Colom and M. Koutny, editors, *Tools* and Algorithms for the Construction and Analysis of Systems, European Joint Conferences on

*Theory and Practice of Software (ETAPS'2001),* volume 2031 of *Lecture Notes in Computer Science,* pages 576-579. Springer-Verlag, Apr. 2001. Tool demonstration.

- [159] O. Beaudoux. Un pivot pour la collaboration : les places publiques. In *Annexes des actes de la conférence francophone sur l'Interaction Homme-Machine (IHM 2002),* 2002.
- [160] F. Bourgeois, Y. Guiard, and M. Beaudouin-Lafon. Pan-zoom coordination in multi-scale pointing. In *Extended Abstracts, ACM Human Factors in Computing Systems, CHI 2001,* pages 157-158. ACM Press, Apr. 2001. Interactive Poster.
- [161] J.-D. Fekete. The infovis toolkit. IEEE Symposium on Information Visualization Conference Compendium (demonstration), Oct. 2003.
- [162] J.-D. Fekete. The infovis toolkit. ACM Symposium on User Interface Software and Technology (UIST 2002) Conference Compendium, Nov. 2003.
- [163] J.-D. Fekete, D. Wang, N. Dang, and C. Plaisant. Overlaying graph links on treemaps. IEEE Symposium on Information Visualization Conference Compendium (demonstration), Oct. 2003.
- [164] M.-C. Gaudel. How to Reach Software Ultra-Reliability? In *The 2002 ARO/ONR/ NSF/DARPA Monterey Workshop on Radical Innovations of Software and Systems Engineering in the Future*, pages 23-28, Venise, IT, oct 2002.
- [165] I. Gawedzki and K. Al Agha. Qolyester: A modular OLSR implementation in C++. In *The OLSR Interop & Workshop*, San Diego, CA, USA, August 2004.
- [166] W. Mackay. Educating multi-disciplinary design teams. In *Proc. of Tales of the Disappearing Computer*, pages 105-118, Santorini, Greece, June 2003.
- [167] W. Mackay, H. Evans, H. Hansen, L. Dachary, and N. Gaudron. Weaving an interactive thread: An interactive event for Tales. In *Proc. of Tales of the Disappearing Computer*, pages 409-415, Santorini, Greece, June 2003.
- [168] E. Nars. Partage de documents par identification automatique non intrusive. In *Annexes des Actes d'IHM 2003*, pages 37-39, Caen, France, Novembre 2003.

## **Dissemination of research results**

- [169] Y. Guiard, F. Bourgeois, and M. Beaudouin-Lafon. Psychologie expérimentale et nouvelles technologies de l'information - L'exemple de la loi de Fitts. *CNRS-Info*, (390):19-20, Feb. 2001.
- [170] N. Roussel. Usages de la vidéo pour la communication médiatisée. Cours (tutorial) donné à l'occasion de la conférence IHM 2002, Poitiers, France, Novembre 2002.

## **Doctoral dissertations and Habilitations**

- [171] K. Al Agha. Évoluer vers la 4<sup>e</sup> génération. Habilitation à diriger des recherches, Université de Paris-Sud XI, October 3 2002.
- [172] T. Al Meshhadany. *Optimisation du partage de ressources dans la troisième génération des systèmes mobiles*. PhD thesis, Université de Paris-Sud XI, September 2003.
- [173] O. Beaudoux. Espaces de travail interactifs et collaboratifs : vers un modèle centré sur les documents et les instruments d'interaction. Thèse de doctorat, Université Paris-Sud, Orsay (France), Juillet 2004.
- [174] M. Benzaid. *Utilisation des techniques des réseaux ad hoc pour l'optimisation de la mobilité dans IP*. PhD thesis, Université de Paris-Sud XI, July 2004.

Research groups

- [175] D. Casalta. Étude de la coordination bimanuelle dans le domaine de l'interaction hommemachine. Thèse de doctorat, Université Paris-Sud, Orsay (France), Septembre 2000.
- [176] S. Conversy. *Conception d'icônes auditives paramétrées pour les interfaces homme-machine*. Thèse de doctorat, Université Paris-Sud, Orsay (France), Septembre 2000. 132 pages.
- [177] S.-D. Gouraud. Utilisation des Structures Combinatoires pour le Test Statistique. PhD thesis, Université Paris XI, Orsay, 2004.
- [178] C. Letondal. *Interaction et programmation*. Thèse de doctorat, Université Paris-Sud, Orsay (France), Septembre 2001. 268 pages.
- [179] L. Perato. *Utilisation des agents mobiles pour l'optimisation de la qualité de service dans un réseau UMTS*. PhD thesis, Université de Paris-Sud XI, September 2003.
- [180] E. Petitjean. *Etude des méthodes de test sur les systèmes temporisés*. Thèse de doctorat, Université de Reims Champagne-Ardenne, Reims (France), Nov. 2000.
- [181] N. Roussel. *Support informatique à une communication médiatisée*. Thèse de doctorat, Université Paris-Sud, Orsay (France), Juillet 2000. 190 pages.
- [182] F. Zaïdi. Contribution a la génération de tests pour les composants de service. application aux services de réseau intelligent. Thèse de doctorat, Université d'Evry, Nov. 2001.

## Reports

- [183] H. Badis and K. Al Agha. Simulation du protocole OLSR sous OPNET. Technical report, LRI: Laboratoire de Recherche en Informatique, http://www.lri.fr, July 2002.
- [184] H. Badis and K. Al Agha. Distributed algorithms for single and multiple-metriclink state qos routing. Technical Report 1365, LRI: Laboratoire de Recherche en Informatique, http://www.lri.fr, July 2003.
- [185] H. Badis and K. Al Agha. A new technique for reducing vertical handoff latency in wireless overlay networks. Technical Report 1390, LRI: Laboratoire de Recherche en Informatique, http://www.lri.fr, August 2003.
- [186] H. Badis and K. Al Agha. Optimal path selection analysis in ad hoc networks. Technical Report 1391, LRI: Laboratoire de Recherche en Informatique, http://www.lri.fr, August 2004.
- [187] M. Beaudouin-Lafon, B. Bederson, S. Conversy, A. Druin, B. Eiderbäck, H. Evans, H. Hansen, A. Harvard, H. Hutchinson, L. Lacomme, S. Lindquist, W. Mackay, C. Plaisant, N. Roussel, Y. Sundblad, and B. Westerlund. interLiving deliverable 1.2 & 2.2, co-design and new technologies with family users. Technical report 174, CID/NADA, KTH, Suède, Sept. 2002. 121 pages.
- [188] M. Beaudouin-Lafon, B. Bederson, S. Conversy, B. Eiderbäck, and H. Hutchinson. interLiving deliverable 2.1, cooperative design with families. Technical report, CID/NADA, KTH, Suède, Jan. 2002. 47 pages.
- [189] M. Beaudouin-Lafon, S. Conversy, B. Eiderbäck, N. Gaudron, H. Evans, H. Hansen, H. Hutchinson, S. Lindquist, W. Mackay, C. Plaisant, N. Roussel, and B. Westerlund. interLiving deliverable 1.3 & 2.3, studies of co-designed prototypes in family contexts. Technical report 231, CID/NADA, KTH, Suède, Feb. 2004. 176 pages.
- [190] M. Beaudouin-Lafon, A. Druin, B. Eiderbäck, A. Harvard, S. Lindquist, W. Mackay, C. Plaisant, Y. Sundblad, and B. Westerlund. interLiving deliverable 1.1, technology probes for families. Technical report, CID/NADA, KTH, Suède, Oct. 2001. 100 pages.
- [191] R. Blanch, E. Ferley, M.-P. Cani, and J.-D. Gascuel. Non-Realistic Haptic Feedback for Virtual Sculpture. Rapport de Recherche 5090, INRIA, France, Jan. 2004.
- [192] L. Boukhatem, A.-L. Beylot, D. Gaiti, and G. Pujolle. Channel assignment with han-

dover queueing in LEO satellite systems based on an earth-fixed cell coverage. Technical report, IRIT: Institut de Recherche en Informatique de Toulouse, Martch 2001.

- [193] J.-C. Fabre, R. de Lemos, C. Gacek, M.-C. Gaudel, N. Georgantas, V. Issarny, N. Levy, A. Romanovsky, and F. Tartanoglu. Final results on architectures and dependable mechanisms for dependable soss, deliverable of the dsos ist project. Technical report, University of Newcastle, UK, Apr. 2003. 49 pages.
- [194] J.-D. Fekete. The infovis toolkit. Technical Report RR-4818, INRIA Futurs, May 2003.
- [195] M.-C. Gaudel, V. Issarny, C. Jones, H. Kopetz, E. Marsden, N. Moffat, M. Paulitsch, D. Powell, B. Randell, A. Romanovsky, R. Stroud, and F. Taiani. Final version of dsos conceptual model, deliverable of the dsos ist project. Technical report, University of Newcastle, UK, Apr. 2003. 123 pages.
- [196] B. Marre and G. Dabosville. Concepts et utilisation de l'outil de test gatel, rapport final du contrat lri/cea sav 26123. Technical report, LRI, Université de Paris-Sud and CNRS, Jan. 2001.
- [197] N. Roussel, H. Evans, and H. Hansen. MirrorSpace: using proximity as an interface to video-mediated communication. Rapport de Recherche 1363, LRI, Université Paris-Sud, France, July 2003. 7 pages.

## **Other publications**

- [198] M.-C. Gaudel and G. Lestiennes. Unfolding Symbolic Transition Systems. In *Dagstuhl seminar on Perspectives in Model Based Testing*, September 2004.
- [199] J. Jeong, J.-S. Park, K. Mase, Y.-H. Han, H. Badis, and J.-M. Orset. Dynamic requirements for ad hoc IP address autoconfiuguration. In *IETF-58 Meeting : Internet Engineering Task Force*, draft-jeong-manet-addr-autoconf-reqts-00.txt, Minneapolis, Minnesota, USA, July 2003. Draft IETF.
- [200] J. Jeong, J.-S. Park, K. Mase, Y.-H. Han, H. Badis, and J.-M. Orset. Dynamic requirements for ad hoc IP address autoconfiuguration. In *IETF-59 Meeting : Internet Engineering Task Force*, draft-jeong-manet-addr-autoconf-reqts-01.txt, Seoul, Korea, February 2004. Draft IETF.
- [201] J. Jeong, J.-S. Park, K. Mase, Y.-H. Han, H. Badis, and J.-M. Orset. Dynamic requirements for ad hoc IP address autoconfiuguration. In *IETF-60 Meeting : Internet Engineering Task Force*, draft-jeong-manet-addr-autoconf-reqts-02.txt, San Diego, USA, August 2004. Draft IETF.

Research groups Prog

Publications



B.4 / Research groups

## Databases

Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Bases de Données



' équipe Bases de Données du LRI a apporté une contribution scientifique, depuis une vingtaine d'années, à différents aspects de la modélisation des données. Les thèmes suivants ont plus particulièrement fait l'objet d'une activité de recherche intense :

- Le problème des vues et de leurs sémantique,
- les bases de données déductives,
- et les modèles de données orientés objet.

L'équipe a également eu une forte implication, au milieu des anneés 80, dans la conception et les premiers développements du premier système de bases de données orienté objet O2.

Au début des années 90, l'explosion d'Internet et du Web comme support tant de calcul que de stockage global a profondement changé la nature de la recherche en bases de données. Dans ce contexte, l'équipe a ré-orienté ses activités durant ces quatre dernières années. S'appuyant sur son expérience passée (voire même sur certains de ses résultats selon le cas) les thèmes de recherche suivants constituent désormais l'activité scientifique majeure de l'équipe :

- Intégration d'Information,
- langages et Systèmes pour données semi-structurées,
- et Bases de Données Temporelles.

Dans ce cadre, l'équipe est impliquée dans plusieurs projets nationaux ou internationaux. Notons enfin une action ponctuelle dans le domaine de la bioinformatique, sur les bases de données de gènes.

À ce jour l'équipe Bases de Données compte cinq membres permanents, six étudiants de doctorat et un post-doctorant.



## Databases

Head: Nicolas Spyratos

he database research group at LRI has contributed research in various aspects of data modelling over the last two decades, more notably in the following areas:

- Database views and their semantics,
- Deductive databases, and
- Object-oriented modelling.

It also had a strong participation in the design and early implementation phase of the O2 system, one of the first object-oriented database management systems to be released in the mid-1980's.

The advent of the Internet and the Web in the early 1990's has profoundly changed the nature of research in databases. As a consequence, the group has re-oriented its activities in the past four years, guided by previous experience in the field of databases, and in some cases by building upon its past achievements. While some effort is still devoted to the research areas mentioned earlier, the current activities of the group are centered on the following themes:

- Information Integration,
- Languages and Systems for Semi-Structured Data, and
- Temporal Databases.

The group is currently involved in several national and international projects in these areas. Note finally an action in bioinformatics on genes databases.

The group currently comprises five permanent members, six doctoral students and one postdoctoral fellow.

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups **BD**Introduction

OBJOIS

THION

## **Research Group members**

Personnel as of 01/01/2004.

Full time faculty				
Name	First Name	Position*	Institution	
BENZAKEN	Véronique	PR1	Université Paris XI	
BIDOIT	Nicole	PR1	Université Paris XI	
BOUDET	Alexandre	CR1	CNRS	
RIGAUX	Philippe	MC	Université Paris XI	
SPYRATOS	Nicolas	PR1	Université Paris XI	
WALLER	Emmanuel	MC	Université Paris XI	
Doctoral students				
BURELLE	Marwan	AM	Université Paris XI	
MELLOUKI	Mohammed	D	Université Paris XI	

AM

AM

IUT d'Orsay

Rectorat de Versailles

\*See the glossary for acronyms.

Matthieu

Virginie

Long term visitors						
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
AFRATI	Foto	Greek	National Technical University	1/11/99	30/06/00	Sabbatical
			of Athens			
CONSTANTOPOULOS	Panagiotis	Greek	University of Heraklion-Crete	10/10/01	9/12/01	CNRS
AKAISHI	Mina	Japanese	University ofHokkaïdo	14/11/01	29/11/01	
FURNADZHIEV	Georgi	Bulgarian	Bulgarian Academy of Science	15/03/03	15/04/03	PECO/NEI project
COLAZZO	Dario	Italian	University of Pisa	1/05/04	31/07/05	Graphduce contract

## **Group evolution**

- Nicole Bidoit joined the group as a Professor in 2001; she was formerly Professor at University of Bordeaux.
- Philippe Rigaux joined the group as Associate Professor in 2001; he was formerly Assistant Professor at CNAM (Technical University of Paris); he has recently been appointed Professor at Université Paris IX-Dauphine and will leave the group in October 2004.
- Serena Cerrito was appointed as Professor at the Université d'Evry and left the group in 2002.
- Alexandre Boudet joined the group as a CNRS Researcher in 2003.

## B.4.2 / BD

## **Research description**

The database research group conducts research on various aspects of databases. A database stores data representing the salient facts of a specific application, and the accompanying database management system provides the tools necessary for retrieving and manipulating this data in a user-friendly way.

Traditionally, a database operates from a fixed scheme, i.e., from a set of predefined data structures. The kind of structures used and the associated retrieval and manipulation languages constitute the so-called *data model*.

Several data models have been proposed since the early 70's, and some of them have led to successful commercial products in the 80's. The most prominent among them is the relational model, and the database management systems based on this model still dominate the database market today.

The database research group at LRI has contributed research in various aspects of data modelling over the last two decades, more notably in the following areas:

- Database views and their semantics,
- Deductive databases, and
- Object-oriented modelling.

The latter is connected with a major design and implementation effort by the group: the O2 System, one of the first object-oriented database management systems to be released in the mid-1980's. The O2 System originated in our group, and was designed and implemented by a consortium of several academic and industrial partners. Our group had a strong participation in the design and early implementation phase of the system.

Research groups

BD

Members Research description The advent of the Internet and the Web in the early 1990's has profoundly changed the nature of research in databases. Indeed, users today can access a multitude of information sources from the Web, which can be traditional databases, file systems, collections of Web pages and so on. These sources are developed autonomously, they are geographically dispersed, they are heterogeneous (i.e., they follow different data models) and their data no more abide by fixed schemas. As a consequence, the retrieval and manipulation of data from the Web introduce a host of new issues and challenging problems that must be solved if we are to provide a user-friendly environment.

In the light of these developments, the database research group at LRI has re-oriented its activities in the past four years, guided by previous experience in the field of databases, and in some cases building upon past achievements. While some effort is still devoted to the research areas mentioned earlier, the current activities of the group are now centred on the following themes:

- Information Integration,
- Languages and Systems for Semi-Structured Data, and
- Temporal Databases.

Our group is currently involved in several national and international projects in these areas. Note finally an action in bioinformatics on genes databases.

The database research group currently comprises five permanent members, six doctoral students and one post-doctoral fellow.

The following paragraphs briefly describe the issues involved, the results obtained and the perspectives opened in these research areas.

## Information Integration

- Participants: Nicolas Spyratos, Philippe Rigaux, Hanen Belhaj Frej.
- Collaborators: M. Akaishi (Hokkaido University, Japan), V. Christophidis (ICS-FORTH, Greece), Y. Loyer (Versailles University, France), C. Meghini (ISTI-CNR, Italy), Y. Tzitzikas (Namur University, Belgium).

In many modern applications users need to access multiple information sources such as traditional databases, file systems, collections of Web pages, etc. In such an environment, one of the main problems that a user faces is how to retrieve the desired data from a set of sources without having to deal with a multitude of different access mechanisms, and how to fuse or *integrate* the possibly heterogeneous data retrieved in order to obtain the desired results.

In principle, the answer to this is to design a single interface to the sources so that users interact only with the interface - as if it were a single information source. In other words, the users formulate their queries to the interface scheme transparently from the sources. It is up to the interface to analyze the user query, select the appropriate sources, query them and collect and integrate their answers so as to produce the answer to the original user query.

Depending on performance requirements, the interface may even store in advance data fetched from the underlying sources, in order to accelerate query processing and/or provide the necessary data volumes for statistical analyses. The so-called *mediators* and *data warehouses* fall into this general setting, and our group pursues research in both these areas, with a particular emphasis on one kind of mediators, the so-called *digital libraries*.

## **Digital Libraries**

Our work on digital libraries started in the context of a European IST Project (the SeLeNe project: Self eLearning Networks, 2002-2004), and continues in the context of a European Network of Excellence on Digital Libraries (the DELOS project, 2004-2007).

A Digital Library (DL) is a networked infrastructure supporting the creation and distribution of services over digital content. In order to realize such a vision, suitable models of the information embodied in the digital content, as well as of the interaction between a DL and its users, must be developed; these are the two objectives undertaken in the present task.

As a basic model, we assume that a DL serves a network of providers, willing to share their documents with other providers and/or consumers (collectively called users). Each document resides on the local repository of its provider, so all providers' repositories, collectively, can be seen as a database of documents spread over the network. The DL acts as a mediator, indexing all shareable documents so that users can access them transparently.

Current work assumes a central taxonomy (hierarchy of terms) according to which all shareable documents must be described if they are to be registered at the library. This taxonomy is under the control of the library. User queries are Boolean combinations of terms from the central taxonomy. From an interaction point of view, we distinguish the following basic operations:

- *Access:* a user accesses documents through the library, in order to satisfy an information need (such as learning about a specific topic), or edit them and use them in creating new documents. Accessing documents is done by querying the library taxonomy;
- *Creation:* a user creates a new document either from scratch (these are atomic documents), or by re-using existing documents (these are composite documents);
- *Storage:* a user stores every newly created document in his local repository. If he wishes to share the document then he also registers it at the library.

The DL indexes all registered documents and provides a number of services to support document access and composition. These services include:

- *maintenance*, including metadata management, registration, removal and modification of a document;
- *querying*, allowing users to query the document collection transparently, in search of documents of interest (during document composition);
- *personalization,* that is the customization of services based on user profiles; this includes notification, recommendation, ranking of query answers, context-based search, document materialization.

The design of algorithms and the implementation of a prototype supporting some or all of the above services is currently under way, in the context of a doctoral thesis (and as part of the DELOS deliverables). Part of this work is carried out jointly with the research groups of ISTI-CNR, Pise, Italy (C. Meghini) and ICS-FORTH, Crete, Greece (V. Christophidis, A. Analyti).

The basic model just described is using a central taxonomy. In the future, this model will be extended in two ways:

• *Local-to-central*. In this model, each user has a local taxonomy and shares resources through the central taxonomy. The local taxonomy is "free" while the central taxonomy is some standard taxonomy. Each local taxonomy is connected to the central taxonomy through an articulation (mapping between terms).

Research groups
BD
Research description

• *Autonomic* (or peer-to-peer). In this model, each user has its own taxonomy and connects to users of his choice. There is no central taxonomy: each user plays the role of a central taxonomy for all its connected users.

We conducted a fundamental study on the integration of rule-based sources, building upon our previous results. Roughly speaking, given two datalog-like programs, each with its semantics and hypotheses, we have investigated how their semantics and their hypotheses can be integrated into their "union". A second fundamental study is under way, in collaboration with ISTI-CNR, exploring the possibilities of applying formal concept analysis (and more generally lattice theory) to the context of digital libraries.

Our work on mediators is related to work conducted in the IASI research group at LRI, albeit their work is in different application areas. This fact has led to the introduction of a post-graduate course on information integration, taught jointly by the leaders of the two groups.

## Data warehouses

Work on data warehouses started in the context of a bilateral franco-german project (PAI Procope, 1997-2000), in collaboration with the University of Munster, Germany (G. Vossen). Although work in this area still continues in our group, the emphasis has been placed, recently, on mediators and digital libraries.

A data warehouse provides the interface between a community of data analysts and decision makers, on the one had, and the operational databases of an enterprise as well as other sources necessary in the decision making process, on the other hand. A data warehouse stores data in the form of materialized views of the underlying sources, and users interact with it in much the same way as with a conventional database. However, there are some notable differences, including the following in which our group is conducting research:

- 1. In a data warehouse users do not change the data, they simply use it. Changes to data happen only at the sources, and such changes are propagated to the warehouse. Change propagation may be implemented either by reconstructing from scratch all warehouse tables affected by the change, or by applying an incremental algorithm to update the affected tables.
- 2. Warehouse users usually ask for complex data analyses, often statistical in nature, and this requires (i) keeping very large volumes of historical data; (ii) processing very complex queries (so complex that some SQL systems refuse to evaluate them!); (iii) special kinds of schemas that facilitate the formulation and processing of such queries.

With respect to the first issue, we have proposed incremental algorithms for rendering the data warehouse self-maintainable, i.e., for updating the data warehouse based solely on knowledge of the updates in the underlying sources, but without the need to query those sources. This work is based on the notion of view complement and builds upon past work by members of the group.

With respect to the second issue, a study is under way that aims at the definition of a dimensional model for data warehouses. In our approach, a warehouse scheme is seen as a directed acyclic graph with a single root (corresponding to a key, in relational terminology). Each node has a type (domain, in relational terminology). The arrows from the root are the dimensions, and each arrow is interpreted as a function. A path algebra is defined based on well-known operations on functions (projection, restriction, composition and pairing), and aggregate queries for online analytic processing (OLAP) are defined based on function inverses.

#### **Key references**

- [6] D. Laurent, J. Lechtenberger, N. Spyratos, and G. Vossen. Monotonic complements for independent data warehouses. *VLDB Journal*, 10(4), December 2001.
- [10] Y. Loyer, N. Spyratos, and D. Stamate. Hypotheses-based semantics of logic programs in multi-valued logics. *ACM Transactions on Computational Logic*, To appear.
- [12] M. Theodorakis, A. Analyti, P. Constantopoulos, and N. Spyratos. A theory of contexts in information bases. *Information Systems Journal*, 19(4):1-54, 2001.
- [14] Y. Tzitzikas, N. Spyratos, and P. Constantopoulos. Mediators over taxonomy-based information sources. *VLDB Journal*, To appear.
- [47] C. Meghini, Y. Tzitzikas, and N. Spyratos. An abduction-based method for index relaxation in taxonomy-based sources. In *MFCS'03 (28th International Symposium on Mathematical Foundations of Computer Science)*, Bratislava, Slovak Republic, August 25-29 2003.

## Languages and Systems for Semi-Structured Data

• Participants: Véronique Benzaken, Nicole Bidoit, Marwan Burelle, Cedric Miachon, Kim NGuyen, Virginie Thion, Emmanuel Waller.

Internet has brought drastic changes and new challenges to data-management applications. More and more, data can now be accessed over the Web and is available in various formats. XML is the most promising of these formats and has become the "lingua franca" of the Web. But if the main IT actors are converging towards a standard format for the exchanged data the situation is quite the opposite on the application side.

The lack of a unique framework and of a formal foundation makes it difficult to reason on and to work with Web applications. It is difficult to query active networked data, i.e., to query data in order to obtain results and to query processes for discovering the services they provide. It is difficult to reason about Web data, to verify that services conform to their contract, that processes consume resources correctly and that they access data according to a given security policy.

Overall, it is difficult to program such applications since we do not possess the adequate primitives and abstraction mechanisms to coordinate active data, to specify their flows of control, communication, and mobility. More generally, we do not have a unified framework where different and multifarious programming, query and verification tools can be smoothly and uniformly integrated.

We aim at developing such a framework by devising a foundational core for the manipulation and processing of *Web data*, relying on XML as an exchange format, and by developing tools, test cases and proofs of concepts. The pivotal point of the project is active Web data, i.e. data that includes processes (rather than code) and their associated state, and that evolves in an environment where data can move, communicate, and coordinate over the network.

#### Query languages for web data: a case study

Note: Joint work with the GraFComm group at LRI (Evelyne Flandrin).

This work addresses the issue of web querying from the point of view of data. It explores how the structure of information impacts the semantic efficiency of query languages. Web

Research groups **BD** 

**Research description** 

data is not structured, and keyword-based querying is not efficient, returning too few or too many results. However, this information generally has an inner (usually hidden) structure. We address the issue of how some knowledge about this structure can improve web querying.

We are working on a case study about scientific papers. Our goal is to achieve a semistructured data-like labelling of scientific papers available on the Web, together with a query language for non computer scientists, e.g., for researchers wishing to query the literature more efficiently than by using only keyword-based Web querying. In order to obtain practical results, we first consider only documents of a domain of the scientific literature that is naturally ordered, namely mathematics, and more precisely graph theory, and we store theorems in a database. The purpose is to obtain reasonable answers to direct and elementary questions such as: "How many neighbours must each vertex of a 2-connex, K{1,3} graph with n vertices have, so that this graph contains cycles of all lengths?" (We are not concerned with automated theorem proving.)

Several models have been designed to represent theorems, ranging from strings to logic formulas, including for each of them an associated query language (featuring basic string pattern-matching). The expressivity of these models is studied and compared, in particular with respect to a set of (formalised) "user behaviours". A prototype is available and has been tested by researchers in graph theory. The models now have to be refined so as to capture real user behaviours and we need to better understand their expressive power.

#### Polymorphic type inference for query languages

A crucial issue in embedding a query in (Web) database application is to ensure at compile time that the generic structure of the database expected by the query is consistent; then check this generic structure against the actual databases over which the application is to be run, including their evolution over time. This is the problem of polymorphic type inference for query languages, similar to that of programming languages in the ML family. A mandatory step on the road to XML languages, e.g. XQuery, is to solve it for the mere relational algebra.

We gave a polymorphic account of the relational algebra. We introduced a formalism of "type formulas": specifically tuned for relational algebra expressions, and presented an algorithm that computes the "principal" type for a given expression. The principal type of an expression is a formula that specifies, in a clear and concise manner, all assignments of types (sets of attributes) to relation names, under which a given relational algebra expression is well-typed, as well as the output type that expression will have under each of these assignments. Topics discussed include complexity, the relationship with monadic logic, and polymorphic expressive power.

Our future work will address the general case for query languages, first for object query languages (essentially sets and records as recursive constructs, e.g. OQL), then for XML languages (essentially records allowing additional or missing attributes).

#### Logic-based query languages for XML-data

The use of the same logic to describe schemas, constraints, and queries opens many interesting possibilities in terms of optimization and automated reasoning. Problems like subtyping, constraint implication, constraint satisfiability, query correctness and query containment, become trivial to state and reveal themselves as special cases of the validity problem for this logic. The same holds for their combinations, such as query containment in the presence of constraints, or query correctness in the presence of subtyping. Queries can be manipulated by combining the condition they express on data with the schema and type information that is already known. Conversely, the implicit structural assumptions of a query can be extracted by the logical expression of the query itself.

It is well known that the semantics of the family of modal logics is defined via graphs called Kripke's models. Such logics provide a simple representation of graph properties and a rather powerful reasoning mechanism over such structures. It is thus quite natural to investigate modal logic (and its variants) in order to build the foundation for XML data much in the same way as first order logic has been investigated and has proved to be the foundation for the relational data model. Hybrid modal logic (Blackburn, 2000) is a variant of modal logics: although (multi)modal logic is a simple formalism for working with relational structures or graphs, it has no mechanism for referring to and reasoning about the individual nodes in such structures. Hybrid multimodal logic increases the effectiveness of modal logic as a representation formalism by allowing one to grasp the graph nodes via formulas. This is achieved by mixing the modal logic ingredients with first order logic ingredients and allowing one to use node names and node variables as atomic formulas. Adding the ability to name graph nodes leads to violating the tree-like model property, which will be very useful in our framework: XML data (with references) are not trees but graphs. Pioneer work can be found in the literature that addresses the issues of expressing constraints and reasoning about semistructured data, although in a rather restricted manner.

We have proposed a well-founded formalization of schema for semi-structured data allowing typed references and contributed to show that hybrid multimodal logic provides a unique formalism for expressing and reasoning about schema and constraints [83]. We also are currently working on a tableau proof technique for (finite) satisfiability logic.

#### Pattern-matching based languages for XML data: CDuce (www.cduce.org)

Note: Joint work with the Language team at Laboratoire d'Informatique de l'Ecole Normale Supérieure (LIENS), headed by Giuseppe Castagna.

XML is becoming a universal standard for the storage and exchange of structured data. One reason for its popularity is the existence of several formalisms for specifying the structure of XML documents. By using dynamic consistency checking to ensure that information being exchanged (between modules in an application, or nodes in a distributed system) has the expected structure, these schema languages significantly increase the robustness of complex XML-based information systems.

However, the use of schema languages in current XML technologies falls far short of what is possible. In particular, schemas currently play little part in the static analysis of programs that operate on XML structures: they are not used for checking code for inconsistencies at compile time, or for optimization – in short, they are not used as types in the usual programming-language sense of the term. The overall goal of our project is to take advantage of this missed opportunity, thereby improving both the robustness and efficiency of XML-based information systems. Our work has focused so far on the design of a language, CDuce, and security issues.

CDuce is a general-purpose typed functional programming language, whose design is targeted at XML applications. The closest work is, of course, XDuce (Hosoya, Pierce, 2003) from which CDuce borrows many key features, such as regular expression types, type-based pattern matching, semantic inclusion of XML types, type inference for patterns, use of recursive patterns. CDuce is not the only project that started from or was deeply influenced by XDuce. Among other projects inspired by XDuce those closest to CDuce are Xtatic

Research groups **BD**Research description

(Gapayev, Pierce, 2003) and XQuery (Boag et al., 2003, Fernández, Siméon & Wadler, 2000). The work on CDuce started from an attempt to overcome some limitations of XDuce and follows three directions: type system, language design and run-time system

*Type system.* XDuce demonstrates the adequacy of some specific features (regular expression types and type-based patterns) to XML applications, but we believe that these features could be integrated in a less specific language. Indeed, as the interface between an XDuce-like language and a mainstream language necessarily looses most of the type information, we aim at minimizing the interactions between CDuce and external languages by allowing the definition of complex applications directly in CDuce. To this end, we extended the XDuce type system by introducing less XML-specific type constructions: products, records, general Boolean connectives (union, intersection, difference), and arrow types (first-class functions), thereby extending the semantic approach that was initiated by XDuce to define subtyping. On a practical side, we implemented a type-checker that gives precise localization of error messages and exhibits samples to demonstrate type-checking failure. The theoretical foundations of CDuce's type system can be found in [39].

Language design. We added language constructions that we believe useful for XML and general-purpose processing such as overloaded functions (to allow code sharing and code reuse), iterators on sequences and trees and several extensions of the pattern algebra. We studied the precise typing of these constructions. We also made several small design decisions that turn out to be very useful. For instance, XML tags are first-class expressions in CDuce, therefore allowing computing on tags, and strings are nothing but sequences of characters. This allows the use of regular expression types and patterns on strings; also, when concatenating two sequences containing characters and XML elements, the two strings at the boundary are automatically concatenated.

*Run-time system.* We tackled the problem of executing CDuce programs efficiently. The key issue is the implementation of pattern matching. To this end, we use a new kind of deterministic tree automata that combines top-down and bottom-up automata, and we developed a compilation schema (from patterns to automata) that uses static type information to avoid unnecessary computation at run-time. This allows the programmer to use a much more declarative style for patterns without performance penalty.

Concerning language-based security, our work is deeply influenced by the work done by Francois Pottier and Vincent Simonnet in the context of ML. We currently adapt it to the context of CDuce: the challenge is that in our context we are in a typeful environment in which (the semantics of) computations are type-dependent.

Our goal is to define and characterize information flows in XML transformations in order to single out potentially insecure transformations, i.e. transformations that may leak confidential or private information. To this end we studied transformations defined in CDuce and made three main contributions:

- 1. A formal definition and study of information flows in the context of XML transformations and, more generally, in the presence of type driven computations;
- 2. A sound technique for detecting XML document transformations that cause insecure information flows, and a formal proof of its correctness;
- 3. By defining security annotations and by relating various kinds of analyses (static/dynamic, sound/complete) to different query scenarios, it proposes a general framework for checking security of middleware-located information flows.

#### **Key references**

[4]J. V. den Bussche and E. Waller. Polymorphic type inference in the standard relational algebra. *Journal of Computer and System Sciences, Special Issue PODS 99*, 64(3):694-718, 2002.

- [30] V. Benzaken, M. Burelle, and G. Castagna. Information flow security for XML transformations. In V. Saraswat, editor, *Eight Asian Computing Science Conference (ASIAN'03)*, Lecture Notes in Computer Science, Mumbai, India, 10-13 December 2003. Springer-Verlag.
- [31] V. Benzaken, G. Castagna, and A. Frisch. : an XML-centric general-purpose language. In *ACM International Conference on Functional Programming (ICFP'03)*, 2003.
- [39] A. Frisch, G. Castagna, and V. Benzaken. Semantic subtyping. In Seventeenth Annual IEEE Symposium on Logic in Computer Science (LICS'2002), pages 137-146, Copenhagen, Denmark, 2002.
- [83] N. Bidoit, S. Cerrito, and V. Thion. A first step towards modeling semistructured data in hybrid modal logic. *to appear in the Journal of Applied Non-classical Logic*, 2004.

## **Temporal databases**

• Participants: Nicole Bidoit, Matthieu Objois. Joint work with Sandra de Amo (University of Uberlandia Bresil) and Luc Segoufin (Gemo Project, INRIA).

It is well known that the choice between an implicit and explicit representation of time in a temporal data model does not affect the meaning of the information stored in a temporal database. However, one cannot state the same about temporal query languages. Following an implicit time representation approach, a temporal database is defined as a finite sequence of instances which can be interpreted as the evolution of the database state during its lifetime; on the other hand, following an explicit time representation approach, tuples in a relation are timestamped and time can represent, for instance, the valid date associated to the information in the real world (valid time) or the date this information has been entered into the database (transaction time). Although non-linear time (branching time) is potentially applicable to database problems such as version management or workflows, we concentrate on temporal domains that are linearly ordered sets.

In the context of an implicit time representation, the linear temporal logic TL, with connectives Until, Since, Next, Previous, and its extensions are the basic formalisms underlying query languages specification. When time is explicitly represented, queries are specified using the standard relational query languages with built-in linear order on the timestamps. One of these languages, the relational calculus with timestamps, i.e. the firstorder theory of linear order, denoted by FO(<=), provides a natural formalism to specify queries in the explicit perspective.

Abiteboul, Herr & Van den Bussche (1999) has shed new light on the relative expressiveness of temporal query languages: surprisingly, TL is strictly less expressive than FO(<=), in particular TL is unable to express the existence of two identical states in the temporal database, while this property is easily expressible in FO(<=). These results are of major interest and stand in contrast with the propositional case. The fact that TL is strictly less expressive than FO(<=) has clearly stimulated subsequent research work in proposing implicit and explicit query languages more expressible than TL. For instance, Abiteboul et al. (1999) propose a hierarchy of implicit and explicit temporal languages that, it is worth pointing out, are either strictly less expressive or strictly more expressive than FO(<=). We have introduced another hierarchy of implicit temporal languages (Bidoit & De Amo, 1999) with the purpose of answering the following open problem: *is there a complete implicit temporal query language?* Investigations of this problem are motivated by the simplicity and computational advantages of temporal logic which makes it especially attractive as a query language for temporal databases. Indeed, because the references to time are hidden, queries are formulated in an abstract, representation-independent way.

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups **BD** 

Research description

It is important at this point, to make some brief comments about the technique employed in Abiteboul et al (1999) in order to prove that TL is strictly less expressive than FO(<=). Their technique is based on communication protocols (Kushilevitz, Nisan, 1996). Roughly speaking, a communication protocol involves two partners exchanging information in order to prove or discover some property. The number of messages exchanged during the protocol execution measures, to some extent, the communication complexity of the property. This technique has also been used in Bidoit & De Amo (1999).

The most important contribution of our work [2] consists in (1) the syntactic characterization of the order-independent properties expressible in TL via the restricted temporal language TL(E), which by itself sheds new light on the expressiveness of implicit temporal query languages and (2) the alternative proof schema it provides, allowing the use of a more classical approach (a game à la Ehrenfeucht-Fraïssé) to show that TL is strictly less expressive than FO(<=). In order to prove (1) we introduce an implicit temporal query language TL(E) and its explicit counterpart in FO(<=), and show that they correspond exactly to the order independent properties expressible in TL. The proof of this result is quite complex and requires a strong notion of order independence involving finite and infinite instances, since a well-known counterexample given by Y. Gurevich guarantees the existence of an order-independent (with respect to finite instances) property expressible in TL which is not expressible in FO(=) therefore, not expressible in TL(E).

#### **Key reference**

[2] N. Bidoit, S. de Amo, and L. Segoufin. Order independent temporal properties. *Journal of Logic and Computation*, 14(2):277-298, 2004.

## Databases for biology

• Participant: Emmanuel Waller Work in collaboration with the BioInfo group.

Databases today are crucial for biology, since in many cases biologist cannot manage their data manually anymore, e.g., with a spreadsheet and a (computer) mouse. Databases have long been used for biological objects such as genes. However, as original research in biology develops, it may require highly specific databases, both from the point of view of data and processing. The purpose of our work is to understand the importance of databases in biological research, and more specifically in the context of a large health research institute with both fundamental and applied research (cancer hospital), and to participate in biology projects by providing original and efficient database concepts and tools.

We built a local bioinformatics group at Institut Curie, undertook there some research work including pragmatic targets and, together with the BioInfo group, started the "BIOIN-GENIERIE Plateforme Biogen 74" project. At IGM (*Institut de Génétique et de Microbiologie*, Orsay), we designed and implemented a database model and querying interface with adequate expressive power to help understand the evolution of protein modules (joint group with Christine Froidevaux and Sarah Cohen-Boulakia, BioInfo group).

#### **Key reference**

[71] S. C. Boulakia, C. Froidevaux, E. Waller, and B. Labedan. Genopage : A database of all protein modules encoded by completely sequenced genomes. In *In Actes de JOBIM2002* (short paper), Journées Ouvertes, Biologie, Informatique et Mathématiques, pages 187-193, 2002.

## **Research Perspective**

*Information integration:* Future work in digital libraries will focus mainly on metadata management in faceted taxonomies, management of user preferences for filtering information and ranking of query answers, and the use of formal concept analysis for classification purposes. These lines of research will be pursued in the context of the DELOS project. Future work in data warehouses will focus on the full development of a functional model for dimensional databases and the incorporation of user preferences in the processing of OLAP (on-line analytical processes) queries.

Languages and Systems for Semi-Structured Data: The development of proof techniques is our priority because it is the central issue for tackling semantic query optimization. Although some preliminary steps have been made, our work does not provide a query language based on hybrid multimodal logic. The satisfiability problem for hybrid multimodal logic is undecidable. Hence, another challenging issue is the characterization of decidable fragments of hybrid multimodal logic powerful enough to express interesting integrity constraints over XML data.

We will also investigate Pattern-Matching as a query language primitive. This requires investigating whether pattern-matching in the sense of CDuce is a good execution primitive for XML query processing. More precisely, to do so, we shall endow CDuce with an SQL-like query language. Optimizing XML query processing and interaction with a persistent store is another major challenge. We plan to develop a generic method for describing persistent XML storages; logical and physical algebras, capable of expressing the desired transformations, with associated equivalence laws; a cost model for our physical algebra, and finally a search strategy able to take advantage of the storage genericity to best implement the required transformation. The starting points of our work will be the XQueC persistent XML storage and querying system developed by INRIA-Gemo, the work on endowing CDuce with a query language facility and the work done in Pisa on logical and physical algebras and optimization for XQuery. Finally, while our work on security provides a sound dynamic analysis for detecting insecure transformations, we plan to provide static analysis to guarantee at compile-time that a given transformation cannot leak confidential information.

*Temporal Databases:* We are currently investigating implicit query languages that allows the extraction of temporal answers, called fully-temporal implicit query languages. In the literature, implicit temporal query languages are defined as mappings that return sets of tuples from temporal databases. This asymmetry between (temporal) input and (static) output is generally accepted based on application issues although it clearly appears that fullytemporal query languages become a central issue in the context of multimedia databases, financial databases and XML data. We will focus on setting the foundations of fully-temporal languages, specifying such languages (preferably based on modal/temporal logic) and investigating their expressiveness and complexity. Research groups
BD
Research perspective

## Highlights

Not only does the group cover most of the classical aspects of the database area both from a theoretical and practical point of view (foundation, design and implementation of query languages, logical and physical query optimization, data integration, etc.), it also has re-oriented its activities aiming at intensively exchanging with other scientific fields and groups (including programming languages, type theory and biology). This is witnessed by the group's publication list, which is not limited to the major databases journals and conferences (VLDB-Journal, TODS, PODS, SIGMOD, VLDB, EDBT, ICDE, etc.) but also includes publications in the fields mentioned above with first class journals and conferences (JLC, TCS, JCSS, ICFP, LICS, etc.). Senior members of the group are regularly serving in program committees or editorial boards.

The radar views below indicate a considerable increase in the amount of contracts, a significant increase of PC membership and a good level of publications. It also shows a significant decrease in the number of defended Ph.D.s. This is due to the fact that Veronique Benzaken was head of the IUP-Miage teaching program from 1998 to 2003 and that Nicole Bidoit was Chargée de Mission at CNRS Department STIC from 2001 to 2004. Because of this heavy involvement in administrative tasks, they decided not to take on supervising Ph.D. students over these periods.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international),
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

## B.4.4 / BD

## Honors

• Nicole Bidoit received the *Palmes académiques* in 2004.

## B.4.5 / BD

## **Evaluation of research**

## **Editorial boards**

- *Journal of Information Systems*, Pergamon Press, Elsevier Science: Nicole Bidoit, area editor since 1993.
- *Technique et Science Informatique*, Editions Hermès: Nicole Bidoit, member of the editorial board (2001-2004), Véronique Benzaken, member of the editorial board (1997-2000).
- Revue I3, Cépaduès Editions: Véronique Benzaken, member of the editorial board since1999.

## **Program committees**

#### Chair

• BDA, Bases de Données Avancées, 2005: Véronique Benzaken.

## **International events**

- FoIKS, First and 4th International Symposium on Foundations of Information and Knowledge Systems, Burg, Spreewald, Germany, February 2000; Budapest, Hungary, February, 2006: Nicolas Spyratos.
- IDEAS, 5th and 9th International Database Engineering and Applications Symposium, Grenoble, France, July 2001; Montreal, Canada, July 2005: Nicole Bidoit (2001), Nicolas Spyratos (2005).
- ISMIS, 15th International Symposium on Methodologies for Intelligent Systems, Saratoga Springs, New York, May 2005: Véronique Benzaken.
- PLAN-X, *International Workshop on Programming Languages for XML*, Long Beach, California, January 2005, joint with ACM POPL: Véronique Benzaken.
- ICDT, 8th and10th International Conference on Database Theory, London, UK, January 2001; Edinburgh, Scotland, January 2005: Nicole Bidoit.

Research groups

#### Highlights Honors Evaluation of research

- ER, 23rd International Conference on Conceptual Modelling, Shanghai, China, November 2004: Nicolas Spyratos.
- PKDD, 4th, 5th, 6th, 7th and 8th European Symposium on Principles of Data Mining and Knowledge Discovery, Lyon, France, September 00; Freiburg, Germany, September 2001; Helsinki, Finland, August 2002; Cavtat-Dubrovnik, Croatia, September 2003; Pisa, Italy, September 2004: Nicolas Spyratos.
- XIME, *International Workshop on XQuery*, Paris, France, June 2004, joint with ACM SIGMOD/PODS: Véronique Benzaken.
- Fourth SIAM International Conference on Data Mining, Orlando, Florida, USA, April 2004: Nicolas Spyratos.
- ICDM, IEEE International Conference on Data Mining, Maebashi TERRSA, Maebashi City, Japan, December 2002; Melbourne, Florida, USA, November 2003: Nicolas Spyratos.
- SIGMOD/PODS, ACM *Symposium on Principles of Database System,* Madison, Wisconsin, June 2002: Nicole Bidoit.
- ICDM, First IEEE International Conference on Data Mining, Silicon Valley, California, USA, November 2001: Nicolas Spyratos.
- Andrei Ershov Fourth International Conference "Perspectives of System Informatics", Novosibirsk, Akademgorodok, Russia, July 2001: Nicolas Spyratos.
- EDBT, 5<sup>th</sup> International Conference on Extending Database Technology, Konstanz, Germany, March 2000): Nicolas Spyratos.

#### National events

- SBBD, Simpósio Brasileiro de Banco de Dados, Brasilia, Brazil, October 2004 : Nicole Bidoit.
- BDA, Bases de Données Avancées: Nicole Bidoit (2001, 2004), Véronique Benzaken (2002, 2003).
- InFORSID 2001, Genève, Mai 2001: Nicole Bidoit.

## Evaluation committees and invited expertise

- National Committee for Scientific Research, *Comité National de la Recherche Scientifique*: Nicole Bidoit, member (2004-2008).
- CNRS, Department of Sciences and technology of information and communication (STIC): Nicole Bidoit, chargée de mission.
- Steering Committee of the national program on massive data, ACI Masses de Données: Nicole Bidoit, member (2003).
- Steering Committee (Comité de pilotage) of PCRI: Nicole Bidoit, member (2001-2004).
- Scientific Committee of LAFMI, *Laboratoire Franco-Mexicain en Informatique*: Nicole Bidoit, member (2002-2005).
- Scientific Board of LaBRI, *Laboratoire Bordelais de Recherche en Informatique*: Nicole Bidoit, member (2000- 2001).
- Best PhD award committee (SPECIF): Nicole Bidoit, member (2000-2004).
- INRIA Futurs Research Associate recruitment committee: Véronique Benzaken, member (2004).
- INRIA Rocquencourt Research Associate recruitment committee: Nicole Bidoit, member (2004).
- INRIA Rennes Research Associate recruitment committee: Véronique Benzaken, member (2003, 2004).
- INRIA Post-Doc recruitment commitee : Nicole Bidoit, member (2001-2003).

## **Other evaluation activities**

- Reviewer for Ph.-D. dissertation: Nicole Bidoit (5), Nicolas Spyratos (6).
- Reviewer for Habilitation: Nicole Bidoit (1), Nicolas Spyratos (1).

## Volunteer professional service

## Management positions in scientific organisations

- National network Information-Interaction-Intelligence, GDR I3 : Nicole Bidoit, member of the management committee (1998-2001).
- CNRS Network on databases and heterogeneous and distributed information, *Réseau Thématique Pluridisciplinaire Bases de données et d'informations hétérogènes et distribuées*, RTP 9: Véronique Benzaken, member of the steering commitee.
- CNRS Working Group on languages, types, security and integrity of semi-structured data, *Action Spécifique Langages, types, sécurite et intégrité des données semi-structurées:* Véronique Benzaken, chair (2003-2004).

## Organisation of conferences and scientific events

- International Workshop "Information Search, Integration and Personalization", Sapporo, juin 2003: Nicolas Spyratos, organizer.
- International Workshop "Peer-to-Peer Computing and Databases" (Jointly with EDBT'04 Conference), Heraklion, Greece, March 14, 2004: Nicolas Spyratos, organizer.

## Working groups

• CNRS Working Group on languages, types, security and integrity of semi-structured data, *Action Spécifique Langages, types, sécurite et intégrité des données semi-structurées:* Nicole Bidoit, member, Véronique Benzaken, chair

Research groups

BD

Evaluation of research Volunteer professionnal service

## **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Total €
IST	N. Spyratos	Selene	UE	Cephyten	31/11/02- 31/01/04	55000
NoE	N. Spyratos	Delos	UE	UPS/SAIC	01/01/04- 31/12/08	22100
RNTL	V. Benzaken	Graphduce	MENRT	UPS/SAIC	09/12/03- 08/06/05	55499
ACI	P. Rigaux	Pattern de mobilite	CNRS	CNRS	07/04/02	15000
ACI	V. Benzaken	Casc (SI)	MENRT	CNRS	29/09/03- 28/09/06	44733
ACI	V. Benzaken	Tralala (MD)	MENRT	CNRS	2004- 2007	
BQR	N. Bidoit	BQR	UPS			
AS	V. Benzaken	Langages types RTP9	CNRS (STIC)	CNRS	16/06/03- 15/06/04	9000
	N. Bidoit	Support Dept STIC	CNRS	CNRS		
	N. Spyratos	Bulgarie	CNRS (DRI)	CNRS	08/01/02	3830

Note: See the glossary for acronyms.

## Summary of each scientific project

#### Selene Self eLearning Networks

European IST project, 2002-2004.

Partners: National Institute of Education (UK); University of London, Birkbeck College (UK); University of Cyprus (Greece); FORTH Institute of Computer Science (Greece); LRI Université Paris-Sud (France).

Scientific director for LRI: Nicolas Spyratos

The main objective of the SeLeNe project is to investigate the feasibility and design of a tool to support learning communities, matching learners' needs with the educational resources potentially available on the Web. The emphasis is put on the development of personalization facilities. The proposed approach relies on semantic metadata describing educational material, and the project is developing services for the discovery, sharing, and collaborative creation of learning objects, facilitating a syndicated and personalized access to such resources. A self e-learning network, as seen in the project, consists of web-based learning objects that have been made available to the network by its users, along with metadata descriptions of these learning objects and of the network's users. The proposed personalization facilities include: querying learning object descriptions to return results tailored towards users' individual goals and preferences; the ability to define views over the learning object metadata; facilities for defining new composite learning objects; and facilities for subscribing to personalized events and change notification services.

#### Delos

The DELOS network of excellence is a 4.5 years project with a budget of about 5 million euros, bringing together over 35 European research groups. The objectives of the DELOS network European Network of Excellence, 2004-2008. Scientific director for LRI: Nicolas Spyratos are to (i) define unifying and comprehensive theories and frameworks over the life cycle of digital library information, (ii) build interoperable multimodal and multilingual services and integrated content management ranging from the personal to the global for the specialist and the general population. The Network aims at developing generic digital library technology to be incorporated into industrial-strength digital library management systems, offering advanced functionality through reliable and extensible services.

In order to achieve these ambitious goals, the DELOS Joint Programme of Activities has been organized into research clusters. Our group is leading the task on "Information and interaction models and processing schemes", within the "Information access and personalization" cluster.

#### GraphDuce

Partners: Programming Languages research group at École Normale Supérieure, the Database Research group at LRI and the BrixLogic startup. BrixLogic main activity is to provide integrated graphical environments for implementing financial data exchange protocols such as FiXML.

The main goal of GraphDuce is to explore the design and implementation of a domainspecific language targeted for processing XML documents with the following functionality and/or properties:

- declarative and/or visual user interface;
- static type checking in order to guarantee robustness and allowing for optimizations;
- well adapted for the deployment of financial data exchange protocols such as FIXML (Financial Information eXchange protocol), Visa 3D or IFX (Interactive Financial eXchange protocol).

In this context we are in charge of the declarative user interface theme, which consists in designing a query language together with graphical tools for intuitively describing queries.

## CASC

Partners: INRIA (SMIS), LIENS, LIUPPA, ENST-Bretagne.

Preserving the confidentiality and integrity of data hosted in multiple distributed sources (personal, administrative, healthcare, business or scientific data) constitutes a tremendous challenge for the database community. Unfortunately, existing access control models implemented in Data Base Management

Systems (DBMS) exhibit important weaknesses. First, existing models are unable to tackle the complexity of distributed and decentralized organizations as well as the growing diversity of channels to access the information. Second, while the semantics of access control policies is well established when applied to relational data, things become fuzzier with semi-structured and hierarchical data such as XML documents. Third, existing models suffer from a centralized access rights administration, making them more vulnerable to both internal and external attacks (according to the FBI computer crime and security report, more than 50% of database attacks are conducted by insiders). The goal of the CASC project is to address these three important issues: how to tackle complex distributed organizations, how to define accurate access control policies on XML-like data and how to secure the global architecture against attacks.

Research project funded by national network on software technology (RNTL), 2003-2005. Scientific director for LRI: Véronique Benzaken

Research groups

Contracts and grants

BD

Project funded by the national initiative on computer security (ACI Sécurité Informatique), 2003-2006. Scientific director for LRI: Véronique Benzaken.

## TRALALA

Partners: INRIA (GEMO et MOSTRARE), LIENS, LIF, LIFL. Foreign Partners: H. Hosaya (U. Tokyo), B.C. Pierce (U. Penn), T. Schwentick (U. Marburg).

This project is motivated by the increasing number of applications that produce, consume or handle large sets of data, or *datamasses*. In many cases, these are either raw data or a collecProject funded by the national initiative on masses of data (ACI Masses de Données), 2004-2007. Scientific director for

LRI: Véronique Benzaken

tion of data from various sources, both of which lack uniform descriptive criteria. Such cases require more flexibility than the classical relational model can provide, and have given rise to the so-called semi-structured data model, of which XML is one of the most prominent examples. Our project intends to study the processing, querying and handling of large datamasses whenever data is available in XML format. We pay particular attention to the programming languages and query languages problems. We aim to cover in a uniform way a wide spectrum of different areas, namely: programming languages (expressiveness, typing, new programming primitives, query underlying logics, logical optimization), data access (streamed data, compression, access to secondary memory storages, persistency engines), implementation (pattern matching compiling, physical optimization, subtyping verification, execution models for streamed data).

### **Other Projects**

Emmanuel Waller initiated and was the original leader for LRI of the "BIOINGENIERIE Plateforme, Biogen 74 project (2001-2004, 42.063 euros, funded by French national network on software technologies, RNTL"), then co-leader with Christine Froidevaux from the BioInfo group. Together with Evelyne Flandrin from the GraFComm group, he was granted a University BQR project (2001, 4.573 euros). These projects are listed in the BioInfo and GraFComm groups respectively.

## Collaborations

- Collaborations of Emmanuel Waller
  - GrafComm team at LRI (E. Flandrin).
  - Institut Curie Paris (J.-P. Thiery, F. Radvanyi).

## **Collaborations leading to joint publications**

- Collaboration between Nicole Bidoit and Sandra de Amo (University of Uberlandia, Brasil): 3-months visits every year (1997-2001).
- Collaborations of Emmanuel Waller:
  - Bioinformatique team at LRI (Ch. Froidevaux, S. Cohen-Boulakia);
  - Institut de Génétique et Microbiologie, Université Paris-Sud, Orsay (B. Labedan);
  - University of Limburg, Belgique (J. van den Bussche).

#### **Collaborations of Nicolas Spyratos**

• Meme Media Laboratory, Hokkaido University, Japan (Y. Tanaka, M.Akaishi):

- Collaboration on the theme "navigational query languages";
- Visiting professor for two months in 2001;
- Joint organization, in 2003, of the workshop "ISIP03: Information Search, Integration and Personnalization", sponsored by the french embassy in Tokyo.
- University of Crete (A. Analiti, V. Christophidis, P. Constantopoulos):
  - Collaboration on the theme "conceptual modelling and mediators";
  - Joint supervision of two doctoral theses;
  - Collaboration in the context of two european projects.
- London University, Birkbeck College (G. Loizou):
  - Collaboration on the theme "metadata management";
  - Collaboration in the context of two european projects.
- ISTI-CNR of Pise (C. Meghini):
  - Collaboration on the theme "Digital Libraries";
  - Collaboration in the context of two european projects;
  - Joint organization, in 2003, of the workshop on Peer-to-Peer Computing and Databases (P2P&DB04).

Research gro	ups
BD	

Collaborations

## Dissemination and technology transfer

## Software licensing and distribution

• CDuce is distributed under the QT Licence. The mailing list CDuce'users accounts for 50 persons.

## Popularisation of research results

• Presentation of the CDuce language by Marwan Burelle at LSM'03 (Libre Software Meeting, Metz, july 2003) in the session "High-level programming languages for developping applications".

## B.4.10 / BD

## Training and education (doctoral and post-doctoral)

Defended habilitations				
Name	Date defended	Current position		
CERRITO Séréna	31/10/00	Professor, Université d'Evry		
RIGAUX Philippe	29/11/02	Professor, Université Paris-IX Dauphine (october 2004)		
Defended doctorates				
Name	Date defended	Current position		
FAYE Abdourahmane	29/09/00			
LOYER Yann	23/05/01	Assistant Professor, Université Versailles Saint-Quentin		
LAGORCE Jean-Bernard	07/02/02			
PRAUD Sébastien	08/02/02			

#### **Graduate courses**

- Master in Computer Science, *Foundations of Database:* Véronique Benzaken, Nicole Bidoit, Nicolas Spyratos.
- Master in Computer Science, *Datawarehouses*: Nicolas Spyratos.
- Master in Computer Science, *Semantic Web*: Nicolas Spyratos.
- Master in Computer Science, Information Integration: Nicolas Spyratos.
- Master in Computer Science, Safety, Security and Integrity of Data: Véronique Benzaken.
- Master of Computer Science, Introduction to research: Emmanuel Waller.
#### **Publications**

#### International peer-reviewed journals

- [1] A. Analiti, P. Constantopoulos, and N. Spyratos. Deriving and retrieving contextual categorical information through instance inheritance. *Fundamenta Informaticae*, 44:1-31, 2000.
- [2] N. Bidoit, S. de Amo, and L. Segoufin. Order independent temporal properties. *Journal* of *Logic Computation*, 14(2):277-298, 2004.
- [3] M. Dekhtyar, A. Dikovsky, S. Dudakov, and N. Spyratos. Maximal state independent approximations to minimal real change. *Annals of Mathematics and Artificial Intelligence*, 33(2-4):157-204, 2001.
- [4] J. V. den Bussche and E. Waller. Polymorphic type inference in the standard relational algebra. *Journal of Computer and System Sciences, Special Issue PODS 99*, 64(3):694-718, 2002.
- [5] S. Grumbach, P. Rigaux, and L. Segoufin. Handling Interpolated Data. 46(6):664-679, 2003.
- [6] D. Laurent, J. Lechtenberger, N. Spyratos, and G. Vossen. Monotonic complements for independent data warehouses. *VLDB Journal*, 10(4), December 2001.
- [7] D. Laurent, V. Phan-Luong, and N. Spyratos. Querying weak instances under extension chase seamantics: A complete solution. *International Journal of Computer Mathematics* (*IJCM*), 80(5):591-613, 2003.
- [8] Y. Loyer, N. Spyratos, and D. Stamate. Interfacing decisison support systems under incomplete information. *International Journal Information Theories and Applications*], 7(1), 2000.
- [9] Y. Loyer, N. Spyratos, and D. Stamate. Parametrized semantics of logic programs. *Theoretical Computer Science*, TCS-B : Logic, Semantics and Theory of Programming(308), 2003.
- [10] Y. Loyer, N. Spyratos, and D. Stamate. Hypotheses-based semantics of logic programs in multi-valued logics. ACM Transactions on Computational Logic, To appear.
- [11] P. Rigaux, M. Scholl, L. Segoufin, and S. Grumbach. Building a constraint-based spatial database system: model, languages, and implementation. 28(6):563-595, 2003.
- [12] M. Theodorakis, A. Analyti, P. Constantopoulos, and N. Spyratos. A theory of contexts in information bases. *Information Systems Journal*, 19(4):1-54, 2001.
- [13] Y. Tzitzikas, N. Spyratos, and P. Constantopoulos. Query evaluation for mediators over web catalogs. *International Journal on Information Theories and Applications*, 9(2), 2002.
- [14] Y. Tzitzikas, N. Spyratos, and P. Constantopoulos. Mediators over taxonomy-based information sources. *VLDB Journal*, To appear.

#### National peer-reviewed journals

- [15] C. Diop, A. Giacometti, D. Laurent, and N. Spyratos. Extraction itérative de requêtes fréquentes, isi (ingénierie des systèmes d'information). To appear.
- [16] T.-Y. Jen, D. Laurent, N. Spyratos, and Y. Tanaka. Rêgles d'association significatives. *Revue des Nouvelles Technologies de l'Information (RNTI Cépaduès)*, Numéro spécial RNTI Cépaduès Mesure de qualité pour la fouille de données, To appear.

Research groups

#### BD

Dissemination and technology transfer Training and education Publications

#### **Book chapters**

- [17] N. Bidoit and S. Cluet. Introduction aux modèles objet et déductif, et aux langages. In G. J. A. Doucet, editor, *Bases de Données et Internet - Modèles, Langages, et systèmes*, pages 23-46. Hermès, 2001.
- [18] N. Bidoit and C. Collet. Contraintes d'intégrité et règles actives. In G. J. A. Doucet, editor, *Bases de Données et Internet - Modèles, Langages, et systèmes*, pages 47-74. Hermès, 2001.
- [19] S. Grumbach, M. Koubarakis, P. Rigaux, M. Scholl, and S. Skiadopoulos. *Spatio-Temporal Databases: The CHOROCHRONOS Approach*, chapter Spatio-temporal Models and Languages: An Approach Based on Constraints, pages 177-201. Springer-Verlag, 2003.
- [20] N. Spyratos and Y. Tzitzikas. *Wiley Encyclopedia of Electrical and Electronics Engineering*, chapter Searching. John Wiley & Sons, 2002.
- [21] Y. Tzitzikas, A. Analyti, N. Spyratos, and P. Constantopoulos. *Information Modelling and Knowledge Bases*, chapter An Algebra for Specifying Compound Terms in Faceted Taxonomies. IOS Press (Frontiers in AI and Applications), 2000.
- [22] Y. Tzitzikas, N. Spyratos, P. Constantopoulos, and A. Analyti. *ERCIM NEWS*, volume 51, chapter Extended Faceted Taxonomies for Web Catalogs. October 2002.
- [23] N. S. Y. Tzitzikas and P. Constantopoulos. *Information Modelling and Knowledge Bases XIII*, chapter Deriving Valid Expressions from Ontology Definitions. IOS Press, Amsterdam, 2002.

#### Major international peer-reviewed conferences

- [24] M. Akaishi, N. S. M. Ohigashi, and Y. Tanaka. Information access space framework over contextualized information bases. In 7th International Conference on Information Visualisation (IV2003), London, England, July 16-18 2003.
- [25] M. Akaishi, M. Ohigashi, N. Spyratos, and Y. Tanaka. A system for tracing implicit relationships in a web of contexts. In *The 14th European-Japanese Conference on Information Modelling and Knowledge Bases*, Skovde, Sweden, June 1-4 2004.
- [26] M. Akaishi, M. Ohigashi, N. Spyratos, Y. Tanaka, and H. Yamamoto. 3d visual construction of a context-based information access space. In *12th European-Japanese Conference on Information Modelling and Knowledge*, Krippen, Swiss Saxony, Germany, May 27-30 2002.
- [27] M. Akaishi, N. Spyratos, and Y. Tanaka. A component-based application framework for context-driven information access. In *EJC'01 (11th European-Japanese Conference on Information Modeling and Knowledge Bases)*, Maribor, Slovenia, May 29 - June 1 2001.
- [28] M. Akaishi, N. Spyratos, and Y. Tanaka. Construction of information access space based on a contextualized information base. In Proc. 7th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2003), Orlando, USA, July 27-30 2003.
- [29] M. Akaishi, N. Spyratos, and Y. Tanaka. Contextual search in large collections of information resources. In *EJC'03 : The 13th European-Japanese Conference on Information Modelling and Knowledge Bases*, Kitakyushu, Japan, June 3-6 2003.
- [30] V. Benzaken, M. Burelle, and G. Castagna. Information flow security for XML transformations. In V. Saraswat, editor, *Eight Asian Computing Science Conference (ASIAN'03)*, Lecture Notes in Computer Science, Mumbai, India, 10-13 December 2003. Springer-Verlag.
- [31] V. Benzaken, G. Castagna, and A. Frisch. : an XML-centric general-purpose language. In *ACM International Conference on Functional Programming (ICFP'03)*, 2003.

- [32] N. Bidoit, S.Maabout, and M. Ykhlef. A family of nested query languages for semi-structured data. In *Foundations of Information and Knowledge Systems, First International Symposium (FoIKS), LNCS Vol. 1762*, pages 13-30, 2000.
- [33] M. Dekhtyar, A. Dikovsky, S. Dudakov, and N. Spyratos. Maximal expansions of database updates. In *Int. Symposium on Foundations of Information and Knowledge Systems (FoIKS'00)*, Burg (Spreewald), Germany, February 2000.
- [34] C. Diop, A. Giacometti, D. Laurent, and N. Spyratos. Composition of mining contexts for efficient extraction of association rules. In *EDBT'02 (8th Int. Conference on Extending Database Technology)*, Prague, Check Republic, March 24-28 2002.
- [35] C. Diop, A. Giacometti, D. Laurent, and N. Spyratos. Mining from views : An incremental approach. In *ICT&P'02 (27th International Conference on Information and Communication Technologies and Programming)*, Primorsko, Bulgaria, June 2002.
- [36] C. du Mouza and P. Rigaux. Web architectures for scalable moving object servers. In *acmgis*, pages 17-22, Washington, USA, 2002.
- [37] C. du Mouza and P. Rigaux. Mobility Patterns. In stdbm, Toronto, Canada, 2004.
- [38] C. du Mouza and P. Rigaux. Multi-scale Classification of Moving Objects Trajectories. In *ssdbm*, Santorin, Greece, 2004.
- [39] A. Frisch, G. Castagna, and V. Benzaken. Semantic subtyping. In Seventeenth Annual IEEE Symposium on Logic in Computer Science (LICS'2002), pages 137-146, Copenhagen, Denmark, 2002.
- [40] B. Gueye, P. Rigaux, and N. Spyratos. Annotation automatique de documents XML. In Actes des journées "Extraction et gestion des connaissances" (EGC'04), Clermont-Ferrand, France, January 2004.
- [41] B. Gueye, P. Rigaux, and N. Spyratos. Taxonomy-Based Annotation of XML Documents: Application to eLearning Resources. In *Third Helenic Conference on AI (SETN'04)*, 2004.
- [42] K. Keenoy, A. Poulovassilis, V. Christophidis, P. Rigaux, G. Papamarkos, A. Magkanaraki, M. Stratakis, N. Spyratos, and P. Wood. Personalisation services for self e-learning networks. In *International Conference on Web Engineering (ICWE'04)*, Munich, Germany, July 28-30 2004.
- [43] Y. Loyer and N. Spyratos. Hypothesis-founded semantics for datalog programs with negation. In MFCS'02 (27th International Symposium on Mathematical Foundations of Computer Science), Warszawa - Otwok, Poland, August 26-30 2002.
- [44] Y. Loyer, N. Spyratos, and D. Stamate. Hypothesis-based semantics for information integration in four-valued logics. In *Workshop FICS'00 (Fixpoints In Computer Science)*, Paris, France, June 2000.
- [45] Y. Loyer, N. Spyratos, and D. Stamate. Hypothesis support for information integration in four-valued logics. In *IFIP-TCS'00 (IFIP International Conference on Theoretical Computer Science : Exploring New Frontiers in Theoretical Informatics)*, Sendai, Japan, August 17-19 2000.
- [46] Y. Loyer, N. Spyratos, and D. Stamate. Integration of information in four-valued logics under non-uniform assumtions. In *ISMVL'00 (IEEE International Symposium on Multiple-Valued Logic)*, Portland, USA, May 23-25 2000.
- [47] C. Meghini, Y. Tzitzikas, and N. Spyratos. An abduction-based method for index relaxation in taxonomy-based sources. In *MFCS'03 (28<sup>th</sup> International Symposium on Mathematical Foundations of Computer Science)*, Bratislava, Slovak Republic, August 25-29 2003.
- [48] P. Rigaux. An iterative rating method: application to web-based conference management. In *ACM Intl. Conf. on Applied Computing (ACM-SAC)*, Nicosia, Cyprus, 2004.
- [49] P. Rigaux and N. Spyratos. Metadata Management and Learning Object Composition in a Self e-Learning Network. In Proc. Intl. Workshop on Information Search, Integration and Personalisation, pages 14-29, Sapporo, Japan, 2003.

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups **BD**Publications

- [50] P. Rigaux and N. Spyratos. A metadata model for learning object management. In *Intl. Workshop on Multimedia Data and Document Engineering (MDDE-2003)*, Berlin, Germany, September 2003.
- [51] N. Spyratos. Querying a collection of intellectual assets the topica project. In *ICT&P'01* (26th International Conference on Information and Communication Technologies and *Programming*), Sofia, Bulgaria, June 26-29 2001. Invited Article.
- [52] N. Spyratos. Towards a query language for topica sheets. In *Workshop on Access Architectures for Organizing and Accessing Intellectual Assets*, Hokkaido University, Sapporo, Japan, March 8-9 2001.
- [53] N. Spyratos, Y., V. Christophides, and Y. Tzitzikas. On personalizing the catalogs of web portals. In *15<sup>th</sup> International FLAIRS Conference, Special Track on Semantic Web*, Pensacola, Florida, USA, May 16-18 2002.
- [54] M. Theodorakis and N. Spyratos. Context in artificial intelligence and information modeling. In S.-V. L. N. in Artificial Intelligence, editor, *2nd Hellenic Conference on Artificial Intelligence (SETN-02), LNAI No 2308*, Thessaloniki, Greece, April 11-12 2002.
- [55] V. Thion, S. Cerrito, and M. Cialdea-Mayer. A general theorem prover for quantified modal logics. In U. Egly and C. Fermüller, editors, *Proceedings of the International Conference on Automated Reasoning with Analytic Tableaux and Related Methods (Tableaux 2002)*, volume 2381 of *LNAI*, pages 266-280. Springer Verlag, 2002.
- [56] Y. Tzitzikas, A. Analyti, and N. Spyratos. An algebraic approach for specifying compound terms in faceted taxonomies. In *2d Hellenic Symposium on Database management*, Athens, Greece, September 4-5 2003.
- [57] Y. Tzitzikas, A. Analyti, and N. Spyratos. The semantics of the compound term composition algebra. In *ODBASE'03 (International Conference on Ontologies, Databases and Applications of Semantics)*, Catania, Sicily (Italia), November 3-7 2003.
- [58] Y. Tzitzikas, A. Analyti, N. Spyratos, and P. Constantopoulos. An algebra for specifying compound terms in faceted taxonomies. In *EJC'03 : The 13th European-Japanese Conference on Information Modelling and Knowledge Bases*, Kitakyushu, Japan, June 3-6 2003.
- [59] Y. Tzitzikas, C. Meghini, and N. Spyratos. Taxonomy-based conceptual modeling for peer-to-peer networks. In *ER2003 (22<sup>nd</sup> International Conference on Conceptual Modeling)*, Chicago, Illinois, USA, October 13-16 2003.
- [60] Y. Tzitzikas, C. Meghini, and N. Spyratos. Towards a generalized interaction scheme for information access. In *FoIKS'04 (Third Int. Symposium on Foundations of Information and Knowledge Systems)*, Vienna, Austria, February, 17-20 2004.
- [61] Y. Tzitzikas, C. Meghini, and N. Spyratos. A unifying framework for flexible information access in taxonomy-based sources. In *6th Intl Conference on Flexible Query Answering Systems (FQAS'04), Lyon France,* June 24-26 2004.
- [62] Y. Tzitzikas, N. Spyratos, and P. Constanstopoulos. Deriving valid expressions from ontology definitions. In *EJC'01 (11th European-Japanese Conference on Information Modeling and Knowledge Bases)*, Maribor, Slovenia, May 29 - June 1 2001.
- [63] Y. Tzitzikas, N. Spyratos, and P. Constanstopoulos. Mediators over ontology-based information sources. In *WISE'01 Intl Conference (The 2d International Conference on Web Information Systems Engineering)*, Kyoto, Japan, December 3-6 2001.
- [64] Y. Tzitzikas, N. Spyratos, and P. Constanstopoulos. Query evaluation for mediators over web catalogs. In *ICT&P'02 (27th International Conference on Information and Communication Technologies and Programming)*, Primorsko, Bulgaria, June 2002.
- [65] Y. Tzitzikas, N. Spyratos, and P. Constanstopoulos. Query translation for mediators over ontology-based information sources. In S.-V. L. N. in Artificial Intelligence, editor, 2nd Hellenic Conference on Artificial Intelligence (SETN-02), LNAI No 2308, Thessaloniki, Greece, April 11-12 2002.

- [66] Y. Tzitzikas, N. Spyratos, P. Constantopoulos, and A. Analyti. Extended faceted ontologies. In 14th International Conference on Advanced Information Systems Engineering, CAiSE-2002, Toronto, Canada, May 2002. short paper.
- [67] Y. Tzitzikas, N. Spyratos, P. Constantopoulos, and A. Analyti. Extended faceted taxonomies for web catalogs. In 3<sup>rd</sup> International Conference on Web Information System Engineering, WISE-02, Singapore, December 2002.
- [68] F. Wattez, S. Cluet, V. Benzaken, and C. Fiegel. Benchmarking queries over trees: Learning the hard truth the hard way. In ACM SIGMOD International Conference, Dallas, May 2000. ACM.

#### National peer-reviewed conferences

- [69] V. Benzaken, G. Castagna, and C. Miachon. CQL a pattern-based query language for XML. In *In Proceedings 20th Bases de Données Avancées (BDA'04)*, 2004.
- [70] N. Bidoit, S. Cerrito, and V. Thion. Données semistructurées et logique multi-modale hybride. In *Conférence nationale Bases de données Avancées*, pages 303-322, 2003.
- [71] S. C. Boulakia, C. Froidevaux, E. Waller, and B. Labedan. Genopage : A database of all protein modules encoded by completely sequenced genomes. In *In Actes de JOBIM2002* (*short paper*), *Journées Ouvertes, Biologie, Informatique et Mathématiques*, pages 187-193, 2002.
- [72] C. Diop, A. Giacometti, D. Laurent, and N. Spyratos. Extraction incrémentale de règles d'association par combinaison de tâches d'extraction. In *BDA' 01 (17<sup>es</sup> Journées Bases de Données Avancées)*, Agadir, Maroc, 29 octobre - 2 novembre 2001.
- [73] A. Giacometti, D. Laurent, C. Diop, and N. Spyratos. La découverte de règles d'association entre vues : vers un processus d'extraction incrémental. In *BDA'00 (16<sup>es</sup> Journées Bases de Données Avancées)*, Blois, France, Octobre 2000.
- [74] T. Jen, N. Spyratos, and Y. Tanaka. Règles d'association significatives. In BDA'02 (18es Journées Bases de Données Avancées), Every, France, October 2002.
- [75] Y. Loyer and N. Spyratos. Sémantique des programmes datalog avec négation sous hypothèses non-uniformes. In BDA'02 (18<sup>es</sup> Journées Bases de Données Avancées), Every, France, October 2002.
- [76] Y. Loyer, N. Spyratos, and D. Stamate. Test d'hypotheses pour l'integration d'informations en logique a quatre valeurs. In *Actes des 9es Journees francophones de programmation logique et avec contraintes (JFPLC'2000)*, Marseille, France, Juin 2000.

#### Other conferences and workshops

- [77] V. Benzaken, G. Castagna, and A. Frisch. Cduce: a white paper. In B. C. Pierce and P. Wadler, editors, *International workshop on Programming Languages for XML (PLANX'02,* Pittsburgh, USA, 2002.
- [78] N. Spyratos. Data warehouses : Concepts and issues. In *Taiwan-France Joint Conference on Information Sciences*, National Chiao Tung University, Hsinchu, Taiwan, November 16-17 2000. Invited Talk.
- [79] N. Spyratos. Data warehouses : Current trends and issues. In *ECD'00 (Extraction de Connaissances a partir de Donees : Data Mining, OLAP and Data Warehousing),* Tunis, Tunisia, May 8-9 2000. Invited Talk.

Research groups **BD**Publications

#### Reports

- [80] C. Pent and N. Spyratos. Utilisation des contextes en recherche d'informations. Research 1320, LRI, 2002.
- [81] P. Rigaux and N. Spyratos. Generation and syndication of learning object metadata. Research 1371, LRI, October 2003.
- [82] Y. Tzitzikas and N. Spyratos. Result fusion by mediators using voting and utility functions. Research 1332, LRI, September 2002.

#### Other publications

- [83] N. Bidoit, S. Cerrito, and V. Thion. A first step towards modeling semistructured data in hybrid modal logic. *to appear in the Journal of Applied Non-classical Logic*, 2004.
- [84] V. Thion. System description of tabfv-prover : A theorem prover for first order modal logics. In C. Areces and M. de Rijke, editors, *Proceedings of the second workshop on Methods for Modalities (M4M-2)*, 2001.



B.5 / Research groups

## Proofs and Programs

Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Démonstration et programmation

#### **Responsable: Christine Paulin**

ous étudions des techniques de démonstration assistée par ordinateur, automatique ou interactive, et leur application au développement de programmes sûrs. Nous avons une activité importante de développement d'outils. La boite à outils CiME propose des techniques avancées de preuve automatique en logique du premier ordre (en particulier pour la ter-

minaison et d'unification). L'assistant de preuve Coq est développé au sein du projet PCRI LogiCal, commun avec l'INRIA Futurs, le LIX et le LRI. Ses nouvelles fonctionnalités d'extraction de programmes et de développement modulaire, issues de travaux théoriques de l'équipe, ainsi que la nouvelle interface intégrée



CoqIDE, sont des atouts majeurs de la version 8 (janvier 2004) de Coq vis-à-vis de ses concurrents internationaux. Un livre paru en mai 2004 décrit les principes d'utilisation de Coq à destination d'un large public, il reflète son impact actuel et contribue à son rayonnement. Des résultats majeurs ont été obtenus dans le domaine de l'intégration de techniques de réécriture et de formalismes fonctionnels d'ordre supérieur, qui ont conduit au développement de prototypes de coopération entre Coq et CiME.

Une évolution de nos recherches a été d'adapter des résultats théoriques généraux à des formalismes de spécification et de programmation largement utilisés. Ainsi, nous avons proposé de nouvelles méthodes pour la preuve de programmes Java et C (outils Why, Krakatoa et Caduceus) ainsi qu'un environnement de spécification d'automates temporisés (Calife).

De 2000 à 2004, douze thèses ont été soutenues dans l'équipe et trois membres permanents ont rejoint des laboratoires prestigieux de la région parisienne. Ces départs ont amené à recentrer les activités de l'équipe autour du développement de plate-formes pour la spécification et la preuve de programmes. Ceci a conduit en 2004 à la définition d'un nouveau projet de recherche PCRI ProVal.

Forts de notre expertise en langages et techniques de preuve avancées, nous attaquons plusieurs verrous technologiques. Le premier est la collaboration intelligente de preuves automatiques et interactives, avec l'amélioration des prouveurs automatiques (en particulier pour la génération de traces). Le second consiste à fournir des langages, méthodes et outils adaptés à la spécification des programmes, phase aussi importante et complexe que la phase de preuve. Ces travaux sont menés dans le cadre de projets nationaux et en collaboration avec des partenaires industriels tels que Dassault Aviation et Axalto (cartes à puce).



## **Proofs and programs**

**Head: Christine Paulin** 

e study advanced techniques for automated or interactive computer-assisted theorem proving and their application to the certification of programs. We have a strong activity in the development of prototype tools. The CiME rewriting toolbox offers new proof techniques in first-order logic (in particular for termination and unification). The Coq proof-assistant is developed by the PCRI project LogiCal, joint between INRIA Futurs, LIX and LRI. Its new features for program extraction and modular development, resulting from theoretical work in our group, together with its new Integrated Development Environment, are major assets of the new

Coq version 8 (January 2004) over its international competitors. A new book published in May 2004 presents the principles of using Coq to a wide audience. It emphasizes its impact and contributes to its international dissemination. Major results were obtained in the integration of rewriting techniques with higher-order functional formalisms, leading to the development of prototypes for the cooperation between Coq and CiME.

The general evolution of our research has been to adapt theoretical results to widely used formalisms for specification and programming. Thus, we have proposed methods and tools for the certification of Java and C programs (Why, Krakatoa and Caduceus tools) as well as an environment for the specification of timed automata (Calife).

From 2000 to 2004, twelve Ph.D. theses were defended in our group, and three permanent members left our group to join prestigious laboratories in the Paris area. These departures led us to focus the group activities around the development of platforms for program specification and proof. This led to the proposal of a new PCRI research project called ProVal in 2004.

Thanks to our expertise in advanced proof techniques, we address several technological issues. The first one is the smart collaboration between automatic and interactive proof, with improvements to automatic provers, in particular the generation of proof traces. The second aims at providing languages, methods and tools adapted to the specification of programs, since specification is as important and complex a step as the proof itself. This research is conducted within the framework of national projects and in collaboration with industrial partners such as Dassault Aviation and Axalto (smart cards).

#### Research groups

Demons Introduction



Laboratoire de Recherche en Informatique / 2000-2004 report

### Research Group members

Personnel as of 01/01/2004.

Full time faculty				
Name	First Name	Position*	Institution	
CONCHON	Sylvain	MC	Université Paris XI	
CONTEJEAN	Evelyne	CR1	CNRS	
FILLIATRE	Jean-Christophe	CR2	CNRS	
MARCHE	Claude	MC	Université Paris XI	
PAULIN	Christine	PR2	Université Paris XI	
PUEL	Laurence	PR1	Université Paris XI	

Doctoral students				
ANDRONICK	June	D	Cifre Axalto	
CORBINEAU	Pierre	AC	Université Paris XI	
LETOUZEY	Pierre	AC	Université Paris XI	
OURY	Nicolas	AC	Université Paris XI	
SIGNOLES	Julien	AM	Université Paris XI	

Non-LRI personnel of PCRI project LogiCal				
Name	First Name	Position	Institution	
BARRAS	Bruno	CR	INRIA	
DOWEK	Gilles	PR	Ecole Polytechnique	
DUPRAT	Jean	MC	ENS Lyon	
HERBELIN	Hugo	CR	INRIA	
JOUANNAUD	Jean-Pierre	PR	Université Paris XI	
WERNER	Benjamin	CR	INRIA	
HERNEST	Dan	D		
HERMANT	Olivier	D		
KIRCHNER	Florent	D		
NARBOUX	Julien	D		
RENARD	Clément	D		
SINOT	François-Régis	D		

\*See the glossary for acronyms.

Long term vi	sitors					
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
RIOS	Alejandro	Argentina	Buenos Aires University	03/01/00	01/02/00	
CHEN	Yijia	Chinese	Shanghai University	24/03/00	23/03/01	UPS MENRT
				24/03/01	15/06/01	UPS Calife & CNET
OKADA	Mitsuhiro	Japanese	Keio University	24/03/00	23/06/00	Associate researcher CNRS
				01/04/01	30/06/01	INRIA
TERUI	Kazushige	Japanese		14/03/01	23/03/01	Post-doc
CAIRES	Luis	Portuguese	Lisbon University	02/04/01	30/04/01	invited prof UPS
FERNANDEZ FERREIRA	Maria	Spanish	Lisbon University	08/04/01	28/04/01	Invited CNRS/Portugal
THARIYAN	Ashin Paul	Indian	IIT Delhi	08/05/01	23/07/01	INRIA Egide
HUANG	Guan Shieng	Chinese	National Taïwan University	05/10/01	31/12/01	Associate researcher CNRS
				01/01/02	15/04/02	UPS-CNET
NAIT ABDALLAH	Areski	Canadian	West Ontario University, London	01/07/03	31/07/03	Invited PR UPS
Mc IVER	Annabelle	English	Macquarie University, Sydney, Australia	19/02/04	30/06/04	Invited PR UPS
CHAUDHARY	Vikrant	Indian	IIT Delhi	15/05/04	30/07/04	INRIA Egide

#### **Group evolution**

Since 2000, four permanent researchers have left the group:

- Jean-Pierre Jouannaud, Professor in the group until december 2000, is now director of the LIX laboratory at École Polytechnique;
- Delia Kesner, Assistant Professor until August 2002, is now Professor at Université Paris 7, member of the PPS laboratory;
- Ralf Treinen, Assistant Professor until August 2002, is now Assistant Professor at *École Normale Supérieure de Cachan*, member of the LSV laboratory.
- Judicaël Courant, Assistant Professor, is on leave for personal reasons since september 2003.

During this period, four post-doctoral researchers visited the group:

- Bruno Guillaume, from 1999 to 2001, now INRIA researcher at LORIA;
- Alexandre Miquel, from 2002 to 2003, now Assistant Professor at Université Paris 7, member of the PPS laboratory;
- Yijia Chen, from 2000 to 2001;
- Huang Guan-Shieng, from 2001 to 2002.

We also had twelve students who defended their Ph.D. theses, listed in section 10 below.

During the same period, two new permanent researchers were appointed:

- Jean-Christophe Filliâtre, Lecturer (ATER) from september 2000 to august 2001, then CNRS Researcher (CR) since october 2001;
- Sylvain Conchon, Assistant Professor since september 2003;

and four new students (listed above) started their Ph.D.

#### **Research description**

Our group studies advanced formal proof techniques, both automated and user-assisted, and their application to the development of certified programs. We are interested in two main logical frameworks. The first is equational logic, where a fundamental technique for computation and reasoning is rewriting. The second is type theory, a higher-order functional framework that offers both a strong logical basis for reasoning and a powerful computation mechanism, thanks to the Curry-Howard correspondence.

We conduct fundamental research, by participating in the definition of powerful formalisms and study their theoretical properties, but we also have important activities in the development of prototype tools: the CiME tool offers advanced rewriting techniques, and is a main subject of collaboration with University of Illinois at Urbana-Champaign; the Coq proof assistant is an environment for specification and proof in the Calculus of Inductive Constructions, a powerful higher-order functional language, it allows us to collaborate with numerous other research groups in Europe, the United States, Canada and Japan. It is also the basis for several actions with industrial partners.

Over the last four years, major results were obtained in the integration of rewriting techniques and higher-order functional formalisms, leading to the development of prototypes of cooperation between Coq and CiME.

We also studied applications to concrete case studies. A general evolution of our research was to adapt theoretical results, mostly obtained for abstract mathematical languages, to specific domains or to widely used specification and programming languages. We have worked on modelling problems coming from specific domains such as telecommunications, theoretical physics or CORBA objects networks.

A new direction for our research started in 2001, with new approaches for the verification of programs written in widely used programming languages: Java and C. As in our earlier work, our theoretical work as been supported by the implementation of tools: Why, Krakatoa and Caduceus. This research and these tools have been (and still are) applied to industrial case studies proposed in particular by Axalto (smart card manufacturer, formerly SchlumbergerSema), Dassault Aviation, and Trusted Logic.

#### PCRI Project LogiCal



LogiCal is a joint project between INRIA-Futurs, LRI and LIX at École Polytechnique. Its general goal is the construction of proof processing systems, i.e. systems that are able to process mathematical knowledge, with the idea that a computerized system for processing mathematical proofs allows to reach a high degree of certainty since such proofs do not contain any error. A main research theme of this project is the development of the Coq proof assistant, for which important meta-theoretic studies are required. Another theme is more applied research on the uses of Coq in particular domains, both for problems of a mathematical nature (such as geometry) and for proofs of programs (or protocols, etc.). Finally the project carries out fundamental research on the formalization of mathematics, such as the representation of proofs, the integration of a programming language in a mathematical formalism, the notion of bound variable, etc.).

Within this project, we have been heavily involved in the development of Coq. Some new features of Coq version 8, released in 2004, directly come from the theoretical work of our group, such as program extraction and modular development. We also developed the new graphical integrated development environment CoqIDE. Of course, we have also been involved in the application of Coq to program proving. In particular, we were the main team from LogiCal involved in the European IST project VerifiCard on smart card verification. In this project we developed the Krakatoa tool and we worked on real case studies provided by SchlumbergerSema.

#### **Theme 1: Rewriting and automated deduction**

- Participants: Claude Marché, Evelyne Contejean, Sylvain Conchon, Pierre Corbineau.
- Former participants: Jean-Pierre Jouannaud, Ralf Treinen, Benjamin Monate, Xavier Urbain.

Our group has a long tradition of research on automated reasoning, in particular on equational logic, rewriting, and constraint solving. One of our important contributions to equational and rewriting logic is a generalization of classical techniques to the more general setting of Horn logic with equality and membership predicates [21, 5, 4], which became the theoretical basis for the Maude system developed by J. Meseguer's group at SRI-international and at University of Illinois at Urbana-Champaign. Benjamin Monate introduced the notion of *parameterized rewriting* [73], allowing to add integer parameters to equational specifications, which has being applied to problems in theoretical physics [31].

Constraint solving differs from reasoning in general logics because formulas are interpreted into a fixed domain. Specialized techniques are usually set up for each domain, and we contributed to *features* constraints [16] (used to model natural languages and heterogeneous data bases), sets constraints [18], rewriting constraints [17] and subtyping constraints [55]. A widely used technique for constraint solving is tree automata, to which we contributed by proposing new formalisms of tree automata *with tests* [18, 56, 90]. We have edited a survey book on constraint solving [20] in which we also wrote a chapter [22].

Around 2001, several people left the group, and our research on automated reasoning has since then focused on two main subjects: termination and integration of decision procedures.

#### Termination

We continued our work on techniques for proving termination of rewriting. A major contribution is the main result of Xavier Urbain's thesis: a new criterion for checking termination *modularly* and *incrementally* [57, 75, 19], and furthermore a generalization of it to termination modulo associativity and commutativity [89, 15]. In cooperation with A.-P. Tomàs from Porto, we also proposed a new approach for solving constraints arising in search for termination orderings based on polynomial interpretations [81]. These new criteria and methods have been implemented into the CiME2 rewrite toolbox [32]. Around 2002, several new projects for developing termination tools arose around the world. We believe we pioneered this growth, and indeed we organized in 2004 the first competition of such tools.

A new research direction on termination techniques was also to apply our new approaches for rewriting to other computing formalisms, first to Prolog programs [50, 51] (see sidebar) and then to the membership equational programs [38] paradigm used in the Maude system already mentioned above.

Research groups

Demons Research description

#### Automated termination proof

ack(0,N,s(N)). ack(s(M),0,A) :- ack(M,s(0),A). ack(s(M),s(N),A) :- ack(s(M),N,A1),ack(M,A1,A).

The Prolog program above computes the famous A function of Wilhelm Ackermann. More precisely, a query ack(n,m,X) assigns to X the value of A(n,m). Proving the termination of the evaluation of such a query can be done by first transforming the program into the term rewriting system below, where extra symbols  $u_{i,j}$  are used to encode Prolog's execution stack, in the form of continuations.

$$\begin{array}{rcl} ack_{in}(0,N) & \rightarrow & ack_{out}(s(N)) \\ ack_{in}(s(M),0) & \rightarrow & u_{1,1}(ack_{in}(M,s(0))) \\ u_{1,1}(ack_{out}(A)) & \rightarrow & ack_{out}(A) \\ ack_{in}(s(M),s(N)) & \rightarrow & u_{2,1}(ack_{in}(s(M),N),M) \\ u_{2,1}(ack_{out}(A_1),M) & \rightarrow & u_{2,2}(ack_{in}(M,A_1)) \\ u_{2,2}(ack_{out}(A)) & \rightarrow & ack_{out}(A) \end{array}$$

This transformation is fully automatic and implemented by the TALP tool [51]. It guarantees that if the resulting term rewriting system terminates, then the Prolog program terminates on each query ack(n,m,X) where n and m are given. A possible technique for proving termination of this system is known as the *dependency graph* criterion. The dependency graph of the system above is given below.



The rounded nodes do not belong to a strongly connected component of the graph, hence they can be ignored. For the square nodes, one has to search for a measure that decreases when corresponding rules apply. This can be done by constraint resolution

techniques. In this example, there is no solution to the constraint system, nevertheless it is possible to find a measure that decreases for the two bottom square nodes, and does not increase for the two top square nodes: |s(x)|=|x|+1,  $|ACK_{in}(x,y)|=|x|$ ,  $|U_{2,1}(x,y)|=y+1$  and everything else 0. In such a case, one just need consider the subgraph of non-decreasing nodes, shown below.



Proving termination is then completed by looking for another measure that decreases for the remaining square node, which is easy here: |s(x)|=|x|+1,  $|ACK_{in}(x,y)|=|y|$  and everything else 0.

The process of computing the dependency graph, finding suitable decomposition into subgraphs and suitable measures is done fully automatically by the CiME rewriting toolbox [32]. The combination of TALP and CiME thus provides a fully automatic tool for proving termination of Prolog programs.

Research groups

#### Demons Research description

#### Integration of decision procedures

Our new research on automated deduction is the combination of decision procedures and their integration within user-assisted proof environments. A first result was obtained [41] in collaboration with Shankar's group at SRI-international who develops the PVS environment, showing how decision procedures for disjoint theories can be combined as soon as each of them provides a so-called *canonizer* and a *solver*. Existing combination methods in the literature are generally not very well understood, and Sylvain Conchon made a major contribution, in collaboration with Sava Krstic from OGI School of Science and Engineering (Oregon Health and Science University, USA): a uniform description of combination of decision procedures, by means of a system of inference rules, clearly distinct from their strategy of application [7, 13], allowing much clearer proofs of soundness and completeness.

We work on integrating such decision procedures into the Coq system. Pierre Corbineau proposed new inference systems for propositional and for first-order intuitionistic logic, which internally incorporate inductively defined predicates [33, 82]. These have been implemented as tactics in Coq version 8, and these tactics have proved to be very useful in practice.

#### **Key references**

[5] A. Bouhoula, J.-P. Jouannaud, and J. Meseguer. Specification and proof in membership equational logic. *Theoretical Computer Science*, 236:35-132, 2000, Elsevier Science Publishers.

- [7] S. Conchon and S. Krstic. Strategies for Combining Decision Procedures. *Theoretical Computer Science*, 2004, Elsevier Science Publishers.
- [15] C. Marché and X. Urbain. Modular and incremental proofs of AC-termination.
- [33] P. Corbineau. First-order reasoning in the Calculus of Inductive Constructions. In S. Berardi, M. Coppo, and F. Damiani, editors, 3<sup>rd</sup> International Workshop on Types for Proofs and Programs, volume 3085 of Lecture Notes in Computer Science, pages 162-177, Torino, Italy, Apr. 2004. Springer-Verlag. Journal of Symbolic Computation, 38:873-897, 2004, Academic Press.
- [55] Z. Su, A. Aiken, J. Niehren, T. Priesnitz, and R. Treinen. The First-Order Theory of Subtyping Constraints. In J. Mitchell, editor, *Conference Record of the 29th Symposium on Principles of Programming Languages*, pages 203-216, Portland, OR, USA, Jan. 2002. ACM Press.

#### **Theme 2: Higher-order formalisms**

- Participants: Christine Paulin, Nicolas Oury.
- Former participants: Frédéric Blanqui, Jacek Chrząszcz, Judicaël Courant, Pierre Courtieu, Bruno Guillaume, Jean-Pierre Jouannaud, Julien Forest, Delia Kesner, Daria Walukiewicz.

We are working on mechanizing deduction in higher-order logic. Higher-order reasoning allows manipulating functions or predicates as first-class objects, leading to more abstract and powerful deduction or computation schemes. The Coq proof assistant is based on a very general type theory: the Calculus of Inductive Constructions (CIC). Following the wellknown Curry-Howard correspondence, this theory may be used for modelling objects as functions or as inductive relations, but also for reasoning on these objects.

There are strong restrictions on the language in order to ensure consistency and decidability of type-checking (and consequently of proof verification). These include strong normalization and uniqueness of normal forms. However some extensions of the original CIC are required in order to provide a more natural representation of mathematical or programming notions. We studied several possible extensions of CIC or alternative calculi. The theoretical part requires complex proofs of properties of these languages. Some extensions have been integrated to the Coq system or are the basis of new prototypes.

Coq provides a large set of libraries, so a powerful module system was required in order for these libraries to be reusable. Such a module system was described in Judicaël Courant's thesis in 1998 [83], however it took several years before a real implementation was integrated into the Coq system. This was achieved by Jacek Chrząszcz [67], and is distributed in Coq version 8. It was successfully experimented in many applications, including our own, e.g., Java program verification [14] and case studies [39]. The integration of modules raises interesting fundamental questions like the interaction with universes. A proposal for explicit universes in Coq was made by Judicaël Courant [35].

Type theory is a powerful language: functions can be represented in an ML-like language, and computation can be implemented efficiently. However, the automation of equational reasoning on term expressions with variables is very poor. Provable equalities cannot be used in a transparent way and often lead to tedious explicit rewriting steps. It is a real challenge to get a better integration of rewriting with typed functional languages. From the theoretical point of view, a first question is how to extend the computation rules of type theory with rewrite rules while preserving strong normalization. There are different criteria. In his thesis, Frédéric Blanqui explored the idea of a general scheme [25, 65, 1] that higher-order rewrite rules should follow. His Calculus of Algebraic Constructions is an exten-

sion of the original Calculus of Constructions with new objects and rewrite rules, where inductive definitions can be encoded in a natural way and rewrite rules that do not correspond to inductive computations can be added. An alternative criterion based on an higher-order extension of the recursive path ordering (RPO) was proposed by Daria Walukiewicz [60, 76] and leads to a modular criterion that is useful in many practical cases. The basis for a module system where functions can be specified abstractly using rewrite rules has been proposed in Jacek Chrząszcz's thesis [67]. From an implementation point of view, combining rewriting and functional computations requires complex data-structures. A first prototype that uses a combination of the CiME libraries and Coq has been designed by Frédéric Blanqui. However, adding arbitrary rewrite rules can be dangerous from the semantic point of view: imagine a rule that rewrites *true* into *false*, it can be safe from the normalizing point of view but is contradictory with the usual assumption that *true*  $\neq$  *false*. In order to avoid this problem, one may prefer to work with provable equalities. Instead of including equalities internally, an alternative is to use a proof-term that reflects the equational reasoning. In his thesis [64], C. Alvarado proposed a communication scheme between Coq and the rewriting tool Elan: Elan searches for a rewriting path leading to a normal form and returns a trace that can be interpreted by Coq in a reflexive way in order to keep a small proof term justifying the equality between the original term and its normal form. This scheme was adapted by P. Guillon for an interaction between Coq and CiME.

Inductive definitions in type theory are restricted to free structures and dealing with quotient type is not as direct as it could be in first-order equational systems. Pierre Courtieu [37, 68] introduced the original idea of *normalized types*, which are quotient structures where there is a normalization function that leads to a canonical element in the equivalence class.

Unification is an important tool for automated reasoning. In higher-order languages the case of pattern unification is interesting because it shares good properties with first-order unification. We studied different extensions of pattern-unification modulo equational theories [28, 61].

Substitution plays an important role in beta-reduction, the main computation rule in functional languages. It is traditionally viewed as a meta-operation. Dealing with substitution as a first-order internal operation has led to very interesting languages, both from the theoretical point of view [8, 2, 66, 12] and for its application to efficient representation of terms in proof assistant or reduction engines for systems that combine rewriting and beta-reduction.

Another meta-operation is pattern-matching in functions defined by case analysis. Delia Kesner and Julien Forest [42, 88, 69] proposed specific languages in order to internalize this operation.

#### **Key references:**

- [2] E. Bonelli. Perpetuality in a named lambda calculus with explicit substitutions. *Mathematical Structures in Computer Science*, 11(1), 2001, Cambridge Univ. Press.
- [25] F. Blanqui. Definitions by Rewriting in the Calculus of Constructions. In *Sixteenth Annual IEEE Symposium on Logic in Computer Science*. IEEE Comp. Soc. Press, 2001.
- [28] A. Boudet and E. Contejean. Combining Pattern E-unification Algorithms. In A. Middeldorp, editor, *Rewriting Techniques and Applications*, volume 2051 of *Lecture Notes in Computer Science*, pages 63-76, Utrecht, The Netherlands, May 2001. Springer-Verlag.
- [37] P. Courtieu. Normalized Types. In L. Fribourg, editor, *Proceedings of the Annual Conference of the European Association for Computer Science Logic*, volume 2142 of *Lecture Notes in Computer Science*, pages 554-569, Paris, France, Sept. 2001. Springer-Verlag.

Research groups
Demons

#### Research description

- [67] J. Chrząszcz. *Modules in Type Theory with Generative Definitions*. PhD thesis, Warsaw University, Poland and Université de Paris-Sud, Jan. 2004.
- [76] D. Walukiewicz-Chrząszcz. Termination of Rewriting in the Calculus of Constructions. PhD thesis, Warsaw University, Poland and Université de Paris-Sud, Apr. 2003.

#### **Theme 3: Proof of programs**

- Participants: Jean-Christophe Filliâtre, Claude Marché, Christine Paulin, Julien Signoles, June Andronick.
- Former participants: Pierre Letouzey, Xavier Urbain.

Our group has been conducting research related to program verification for years. We mainly focus on the verification of behavioural specifications for programming languages such as C, Java and ML.

In 1999, Jean-Christophe Filliâtre introduced a new technique for the verification of imperative programs based on a functional translation into type theory [10]. A first implementation of this work allowed verifying imperative programs within the Coq proof assistant [9]. However it was quickly noticed that this approach was actually independent of the Coq system, and the design and implementation of a new tool, *Why*, started in 2001 [87]. This tool is a verification conditions generator: from an annotated program written in a small imperative language with Hoare logic-like specification, it generates conditions can be generated for several existing provers, including interactive proof assistants (Coq, PVS, HOL Light, Mizar) and automatic provers (Simplify, haRVey, CVC Lite). This multiprover architecture is a powerful feature of Why, as it spreads this technology well beyond the Coq community.

Since 2001, the team was involved in the European project VerifiCard [80] and started considering the verification of Java programs. We designed a method to verify Java source code annotated with the Java Modelling Language (JML) and implemented it within a tool called Krakatoa [14] (see sidebar). The main challenge was the design of a suitable model for the Java memory heap in order to tackle programs with possible aliases (even if Java programs do not manipulate pointers explicitly, objects are passed around through pointers). Indeed, the Why tool does not handle aliases by itself. The Krakatoa model is inspired by old ideas from Burstall, recently used by Bornat and others to verify programs with pointers. The key idea is to separate the heap according to objects fields, since it is statically known that two different fields cannot be aliased. Thanks to its modular architecture, the Why tool can be reused for the generation of verification conditions. The memory heap model is declared as a set of abstract types, functions and predicates and Java code is then translated into the Why input language. This original modular design has many advantages. First, the Krakatoa tool was implemented and released in less than a year, while similar projects that started before 2001 are not yet completed in 2004. Second, the Krakatoa tool benefits from the multi-provers output of the Why tool, offering a wide range of proof systems for the verification of Java programs. The Krakatoa tool was successfully used for the formal verification of a commercial smart card applet [43] proposed by SchlumbergerSema company. This case study was conducted in collaboration with B. Jacobs' group at University of Nijmegen and N. Rauch at University of Kaiserslautern. The Krakatoa tool is currently under use at Axalto (verification of other applets) and at the National Institute of Aerospace (avionics systems).

#### Verification of Java programs



The Java program on the left is a toy example of an electronic purse. It is instrumented with annotations in the Java Modelling Language (JML), written as Java comments (in red). The first annotation is a class invariant, indicating here that the field balance, the purse's current balance, must always be non-negative. The second annotation, which specifies the behavior of method withdraw for debiting the purse, is made of four clauses: a precondition requiring that the debited amount s must be non-negative ; a clause saying that the balance field is modified by the method ; a post-condition giving the behavior in case of normal termination, relating the previous (\old) value of balance and its new value; and finally another post-condition in the failure case, when an exception is thrown because the attempted debit is too large.

The Krakatoa tool generates verification conditions, which are logical formulas whose validity ensures that the body of the method correctly implements its JML specification. In this example, two verification conditions are generated and their validity can be proved using the Coq proof assistant, as shown below. The left part displays the Coq proof script, where the user has to fill in the part between Proof and Save. The right part displays the goals remaining to be proved. Apart from the syntactic differences between Java and Coq, these goals are those one would expect, which means that the Coq modelling of Java semantics does not introduce extra encoding. This is an important strength of our approach.

File Edit Navigation Try Tactics Templates Queries Compile Windows Help

<pre>((acc balance this) - si) /\ (Purse_invariant balance0 this)) (store_extends heap heap) /\ (modifiable heap balance balance0 (value_loc this)). roof. mfold Purse_invariant; intuition. ubst; krakatoa. ubst; krakatoa. ave. * Why obligation from file "Purse_withdraw.why", charact emma Purse_withdraw_body_po_2 : forall (this: value), forall (this: value), forall (balance: ((memory) Z)), forall (balance: ((memory) Z)), forall (heap: store), forall (Test: s &gt;= 0 / (neqv this Null) /\ ((Purse_invariant balance this)), forall (Test1: (acc balance this) &lt; s), forall (heap: store), forall (nesut: value), forall (resut: (acc balance this) &lt; s), forall (nesut: (acc balance this) &lt; s), forall (nesut: value), forall (nesut: (ineqv result0 Null) /\ (heap: store), forall (nest: (ineqv result0 Null) /\)</pre>	this : value s : Z balance : memory Z heap : store Test2: this # balance >= s H : s >= 0 H1 : this != Null H0 : instanceof heap this (ClassType Purse) H3 : this # balance >= 0 krak_acc : Z Post2 : krak_acc = this # balance - s balance0 : memory Z Post1 : balance0 = update balance this krak_acc (1/2) this # balance0 = this # balance - s (2/2) this # balance0 >= 0
(instanceof heap0 result0 (ClassType NoCreditE (fresh heap result0) /\	

Research groups

#### Demons Research description

The success of Krakatoa and the constant request from industrial partners encouraged us to start a similar project for C programs in 2003. The whole design of Krakatoa could be reused, apart from the specification language that had to be redesigned from scratch (but is greatly inspired by JML) and the memory model that had to be refined in order to handle pointer arithmetic [40]. The resulting tool is called Caduceus. It is currently under experimentation at Dassault Aviation and Axalto.

Beside the verification of C and Java programs and the development of the related set of tools Why-Krakatoa-Caduceus, we are exploring several other aspects of program verification and several other programming paradigms. Among these, many efforts are related to the verification of ML programs, a highly challenging goal if one considers the combination of higher-order, polymorphism, side effects and module system.

Regarding purely functional ML programs, the extraction technique has been used for years in the Coq proof assistant. It consists of extracting programs by erasing logical parts in constructive proofs of their specifications. Such programs are then correct by construction. During his Ph.D. thesis, Pierre Letouzey designed and implemented a new extraction mechanism for Coq that was much more powerful than the old version [46, 71]. With this new extraction, Jean-Christophe Filliâtre and Pierre Letouzey could verify ML finite sets libraries based on balanced trees [39]. Other ML programs have been verified using similar techniques, including the Koda-Ruskey algorithm [58, 11] and exact real arithmetic algorithms [93]. During his master thesis, Nicolas Oury investigated methods to improve the efficiency of extracted programs by a safe substitution of data structures at extraction time [99, 52].

We are also considering other ways to tackle ML programs. Julien Signoles has started a Ph.D. on the extension of ML with refinement, a methodology usually applied to imperative programming languages. The key idea is to mix types and expressions into a single syntactical entity. It leads to non-determinism, as usual with methods based on refinement, but also to dependent types. The usual notion of typing becomes a particular case of a more general notion of refinement between two programs.

Among other programming paradigms, we collaborated to the definition of a model of timed automata in Coq [53]. It is integrated in the CALIFE platform, a general tool for specification and automatic or interactive verification of protocols that is developed within the context of the CALIFE and AVERROES projects.

Verification of distributed programs was also considered in the Ph.D. of Patrick Loiseleur [72] who designed specification and proof methodologies for CORBA objects networks.

#### **Key references:**

- [10] J.-C. Filliâtre. Verification of Non-Functional Programs using Interpretations in Type Theory. *Journal of Functional Programming*, 13(4):709-745, July 2003, Cambridge Univ. Press.
- [14] C. Marché, C. Paulin-Mohring, and X. Urbain. The Krakatoa tool for certification of Java/JavaCard programs annotated in JML. *Journal of Logic and Algebraic Programming*, 58(1-2):89-106, 2004, Elsevier Science Publishers. http://krakatoa.lri.fr.
- [39] J.-C. Filliâtre and P. Letouzey. Functors for Proofs and Programs. In *Proceedings of The European Symposium on Programming*, volume 2986 of *Lecture Notes in Computer Science*, pages 370-384, Barcelona, Spain, April 2004.

- [40] J.-C. Filliâtre and C. Marché. Multi-Prover Verification of C Programs. In Sixth International Conference on Formal Engineering Methods, Lecture Notes in Computer Science, Seattle, USA, Nov. 2004. Springer-Verlag.
- [43] B. Jacobs, C. Marché, and N. Rauch. Formal Verification of a Commercial Smart Card Applet with Multiple Tools. In *Algebraic Methodology And Software Technology*, Lecture Notes in Computer Science, Stirling, UK, July 2004. Springer-Verlag.

#### **Research Perspective**

In 2004, we proposed the creation of a new PCRI project between INRIA and LRI, called *ProVal*, that emphasizes our goals for the future: we want to tackle several major scientific and technological issues in program verification in order for the methodology to scale up to industrial applications.

One important point is the specification language. We want to follow the approach that originated in the JML specification language for Java programs: with this approach, specifications are integrated as special comments and can be handled by specialized tools for different levels of verification. We are working on a similar proposal for C programs. One important point is to have appropriate primitives or high-level constructions for expressing certain classes of properties, possibly using well-established specification methods such as some classes of UML diagrams. One interesting challenge is the proof of systems developed in different languages. We believe it is accessible because of the general architecture of our approach based on a common basis, the Why tool.

Part of the specification is naturally given by the user, but many extra conditions have to be added in order to ensure basic properties such as no dereferencing of null pointers nor access outside the bounds of an array. Such conditions should be inferred by the weakest pre-condition generator, simplified and automatically integrated. This requires good interaction between program analysis techniques and automatic deduction.

The success of tools for proving programs in industry is related to their level of automation. However, because automation often fails, a good interactive prover is always required. Currently, our proof obligations can be solved by different provers but in an independent way. A major issue is to be able to combine automatic proof techniques and interactive ones in a smart way. A good automatic prover for programs needs to combine arithmetic decision procedures, theory of access and updates in arrays and first-order reasoning. Several systems explore this combination, and we are working on a modular approach to this kind of system. Our goal is to integrate these automatic procedures into a safe tool such as Coq, where checkable proof-trace are required. Such an approach is important for certification.

Even if the C and Java languages are very popular in industry, we believe other languages raise very interesting problems. One of our long-term objectives is to provide an environment for specification and proof of Ocaml programs, but the interaction of higher-order functions, polymorphism and side effects is still a major scientific challenge. Dedicated languages, in particular languages related to the manipulation of XML documents, are also interesting targets for our methods.

Our chances of success are based on our background in programming languages, automatic deduction, general proof assistant on the one hand, on our close collaboration with industrial partners such as Axalto and Dassault Aviation on the other hand and finally on the rapid and successful development of Krakatoa and Caduceus on top of the general Why tool.

#### Research groups

#### Demons Research perspective

B.5.3 / Demons

#### Highlights

Twelve Ph.D. students defended their doctoral thesis during the 2000-2004 period. Two of them received a prize for their work. As a result of several permanent members leaving the group and so many students graduating, the size of the group is significantly smaller than in 1999. With the appointment of two new permanent members in 2001 and 2002, the number of Ph.D. students is starting to go up again.

We have an important activity in developing robust prototypes that integrate our most recent research results. They are distributed under open source licences and are consequently widely available for evaluation by academic and industrial partners. CiME and Coq are long-term development projects. The distribution of CiME2 and Coq V8 mark major evolutions in the architecture of these systems. Why, Krakatoa and Caduceus are newer developments devoted to the task of proving programs. These tools evolve quickly, in close interaction with industrial partners.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international), and
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

#### Honors

#### Prizes and awards

- Jean-Pierre Jouannaud, Michel Monpetit award, "Académie des Sciences", France, 2000.
- Frédéric Blanqui, Kleene award for best student paper, "16th IEEE Symposium on Logic in Computer Science" (LICS), 2001.
- Frédéric Blanqui, best Ph.D. award, "Société des Personnels Enseignants et Chercheurs en Informatique de France" (SPECIF), 2002.
- Xavier Urbain, best Ph.D. award, "Association Française d'Informatique Théorique" (AFIT), 2003.

#### **Keynote addresses**

• Christine Paulin, invited keynote speaker, Conference on Theoretical Aspects of Computer Software (TACS 2001), Sendai, Japan, October 29-31, 2001 [53].

#### B.5.5 / Demons

#### **Evaluation of research**

#### **Program committees**

#### Chair

 JFLA, 14<sup>es</sup> Journées Francophones des Langages Applicatifs, Chamrousse, France, January 27-28, 2003: Jean-Christophe Filliâtre.

#### **International events**

- MFCS, 25th Conference on Mathematical Foundations of Computer Science, Bratislava, Slovaquia, August 28-September 1st, 2000: Claude Marché.
- RTA, *12th and 14th conference on Rewriting Techniques and Applications*, Utrecht, Netherland, May 22-24 2001; Valencia, Spain, June 9-11, 2003: Christine Paulin.
- CAV, 13th Conference on Computer Aided Verification, Paris, France, July 18-23, 2001: Christine Paulin.
- THPOLs, *Theorem Proving in Higher Order Logics*, Edinburgh, Scotland, September 3-6, 2001; Hampton, Virginia, U.S.A, August 20-23, 2002; Rome, Italy, September 8-12, 2003; Park City, Utah, USA, September 14-17, 2004: Christine Paulin (2001, 2002, 2003, 2004), Jean-Christophe Filliâtre (2003, 2004).
- CSL, *15th International Workshop Computer Science Logic*, Paris, France, September 10-13, 2001: Christine Paulin.
- ACM SIGPLAN Workshop on Types in Language Design and Implementation (TLDI'03), New Orleans, Louisiana, January 18, 2003: Christine Paulin.
- 7<sup>th</sup> International Workshop on Termination (WST 2004), Aachen, Germany, June 1-2, 2004: Claude Marché.

#### Research groups

#### Demons

Highlights Honors Evaluation of research

#### National events:

• JFLA, 14es Journées Francophones des Langages Applicatifs, Chamrousse, France, January 27-28, 2003: Claude Marché.

#### Evaluation committees and invited expertise

- Expert for the French ministry of Research and Education: Christine Paulin.
- Expert for the 6<sup>e</sup> chambre de la Cour des Comptes: Laurence Puel.
- President of 2004 audition committee for hiring junior researchers at INRIA Futurs: Christine Paulin.
- Members of University hiring committee, *commission de spécialistes*, section 27, Université Paris-Sud: Evelyne Contejean (2000-2003), Claude Marché (2004), Christine Paulin (2000-2004).
- Members of University hiring committee, *commission de spécialistes*, section 27, École Normale Supérieure de Cachan: Evelyne Contejean (2004), Claude Marché (2004).

#### B.5.6 / Demons

#### Volunteer professional service

#### Management positions in scientific organisations

- CNRS network on complex or constrained embedded systems, Réseau Thématique Pluridisciplinaire, Systèmes Embarqués Complexes ou Contraints (RTP 19): Christine Paulin, member of the steering committee.
- CNRS network on mathematical methods for computer science, *Réseau Thématique Pluridisciplinaire Méthodes Mathématiques de l'Informatique* (RTP 23): Christine Paulin, member of the steering committee.
- EACSL, *European Association for Computer Science Logic* http://www.dimi.uniud.it/~eacsl : Christine Paulin, member of the executive committee.

#### Organisation of conferences and scientific events

- JFLA, 14es Journées Francophones des Langages Applicatifs (French-speaking workshop on applicative languages), Chamrousse (France), January 27-28, 2003, http://pauillac.inria.fr/jfla/2003 : Jean-Christophe Filliâtre, organizer.
- TYPES Conference, Jouy en Josas (France), December 15-18, 2004, http://types2004.lri.fr: Christine Paulin, Jean-Christophe Filliâtre, organizers.

#### Working groups

- IFIP Working Group on Programming Methodology (WG 2.3) : C. Paulin, invited member (2001).
- IFIP Working Group on Program Generation (WG 2.11) : Christine Paulin, member since 2004.

#### Other professional service

 Organization of the 1st Termination Competition, held in conjunction with the 7th International Workshop on Termination (WST 2004), Aachen, Germany, June 1-2, 2004, Claude Marché. http://www.lri.fr/~marche/wst2004-competition/

#### B.5.7 / Demons

#### **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Total €
IST	C. Paulin	TYPES	UE	CEPHYTEN	01/10/00- 31/09/03	20400
RNTL	C. Paulin	AVERROES	MENRT	UPS	30/09/02- 29/09/05	75000
RNRT	C. Paulin	CALIFE	MENRT	UPS	12/07/99- 11/07/02	58729
ACI	C. Paulin	GECCOO (SI)	MENRT	CNRS	21/07/03- 20/07/06	29264
ACI	E. Contejean	CIME (JC)	CNRS	CNRS	22/10/01- 21/10/03	25958
Industry	JP. Jouannaud	COQ	France Telecom	UPS	01/12/99- 30/11/02	79582
SUBV	E. Contejean	BQR AREPHYT 2001	UPS	UPS		4573
SUBV	D. Kesner	ICCII	CNRS	CNRS		1 296
SUBV	D. Kesner	Portugal	CNRS	CNRS		2020

Note: See the glossary for acronyms.

#### **TYPES**

#### http://www.dur.ac.uk/TYPES/

Partners: 32 groups from academia (Helsinki, Chambéry, Paris, Lyon, Rocquencourt, Sophia Antipolis, Orsay, Darmstadt, Freiburg, München, Birmingham, Cambridge, Durham, Edinburgh, Manchester, London, Sheffield, Padova, Torino, Udine, Nijmegen, Utrecht, Bialystok, Warsaw, Minho, Chalmers) and from industry (Prover Technology, France Telecom, Nokia, Dassault Aviation, Trusted Logic, Xerox).

The subject of this project is Computer-Assisted Formal Reasoning, an area which is of deep interest to industry. The goal is to develop the technology of formal reasoning based on Type Theory by improving the languages and tools of reasoning, and apply the technology to several domains such as programming languages, certified software, and formalization of mathematics.

The project was a working group of the IST program in the fifth framework, it ended in

Working group of the Information Society Technologies (IST) European program, 5th framework, 2000-2003. Scientific director for LRI: Christine Paulin. Demons

Evaluation of research Volunteer professional service Contracts and grants 2003. A new coordinated action under the sixth framework started in september 2004. It is coordinated by Chalmers University.

#### **AVERROES**

http://www-verimag.imag.fr/AVERROES/

Partners: France Télécom R&D (coordinator); CRIL Technology Systèmes Avancés ; LaBRI, Bordeaux; LORIA, Nancy; PCRI (LRI, LIX, INRIA Futurs), Saclay ; LSV, Cachan.

The goal of the project is the development of formal methods for verifying, in a secure way, various properties (fonctional, non-fonctional, quantitative) of interest in an industrial context.

#### CALIFE

http://www.loria.fr/projets/calife/

Partners: France Télécom R&D (coordinator); CRIL Technology Systèmes Avancés; LaBRI, Bordeaux; LORIA, Nancy; LRI, Orsay; INRIA, Rocquencourt; LSV, Cachan.

The goal of the project was the development of an environment for formal specification, proofs and tests of algorithms used in telecommunications. During the project, a platform, CALIFE, was developed which supports specification using parameterized timed automata and the automatic generation of proofs in different automatic or interactive provers.

#### **GECCOO**

http://geccoo.lri.fr/

Partners: LogiCal PCRI project, LRI, Orsay (coordinator); TFC group, LIFC, Besançon; CASSIS project, LORIA, Nancy; Everest project, INRIA Sophia-Antipolis ; VASCO group, LSR, Grenoble.

Project funded by French national research program on computer security, ACI Sécurité et Informatique, 2003-2006.

Research project funded

by national network on software technology (RNTL), 2002-2005.

Scientific director for

LRI: Christine Paulin.

Research project funded

by national network on telecommunication (RNRT), 1999-2002.

Scientific director for LRI: Christine Paulin.

Coordinator and scientific director for LRI: Christine Paulin.

The objective of this project is to propose new methods and tools for the development of object-oriented programs, with strong guarantees of security. The project specifically focuses on

programs embedded in smart cards or terminals which are written in subsets of Java, such as JavaCard.

#### Other contracts

Other contracts of PCRI project LogiCal managed by INRIA, involving LRI personnel.

#### VerifiCard

http://www.verificard.com/

IST European project.

Partners: INRIA, University of Nijmegen, University of Munich, Unversity of Hagen, Swedish Institute of Computer Science, and Schlumberger-Sema.

The VerifiCard project aims at providing the European smart card manufacturers with the latest technology (models and tools) for verification of the JavaCard smart card platform and smart card applications, so that they can comply with the highest quality



evaluation requirements, as laid down in the so-called Common Criteria, the ISO standard for certifying the security of ICT products. It proposes to boost the field of Java program specification and verification, so that it can develop from an academic discipline into an industrially relevant field.

#### B.5.8 / Demons

#### Collaborations

#### **Cooperation agreements**

- Bilateral CNRS-STIC/University of Illinois at Urbana-Champaign research project *Rewriting calculi, logic and behavior,* managed by the LIX laboratory at Ecole Polytechnique, 2003-2005.
- Cooperation with Axalto for funding the Ph.D. thesis of June Andronick under a CIFRE agreement, 2002-2004.

#### Collaborations leading to joint publications

- Bart Jacobs's group LOOP, University of Nijmegen: [43].
- José Meseguer's group, University of Illinois at Urbana-Champaign: [38].
- Ana-Paula Tomas, University of Porto: [81], to appear in Journal of Automated Reasoning.

#### Informal collaborations with industrial partners

- Dassault Aviation, Suresnes: Proof of C programs for avionics.
- Trusted Logic: Javacard applets, formal modeling with Coq.
- Axalto: Proof of Java and C programs, Javacard applets and operating systems.
- France Télécom, Lannion: CALIFE/AVERROES project, communication protocols.
- Jacques Raguideau's research group, LIST laboratory, CEA, Saclay: proof of C programs, CAVEAT project.
- César Muñoz, Natiuonal Institute of Aerospace (NIA), Hampton, USA: Proof of Java programs for avionics. 9. Dissemination and technology transfer.

#### B.5.9 / Demons

#### Dissemination and technology transfer

#### Software licensing and distribution

• CiME, a rewriting toolbox. Distributed since 1996 as open source, under the LGPL licence, http://cime.lri.fr. In addition to a few dozen users, CiME is used as a back-end for other tools such as the TALP tool http://bibiserv.techfak.uni-bielefeld.de/talp/ for termination of logic programs, the Mu-Term tool http://www.dsic.upv.es/~slucas/csr/termination/muterm/ for termination of context-sensitive rewriting, and the Cariboo tool http://www.loria.fr/equipes/protheo/SOFTWARES/CARI-BOO/ for termination of rewriting under strategies.

Research groups

#### Demons

Contracts and grants Collaborations Dissemination and technology transfer

- Coq, a proof assistant based on Type Theory. Coq is the result of 20 years of development supported by INRIA in collaboration with ENS Lyon and LRI. It is distributed as open source since 1984, Its current version 8 was released on January 2004, under the LGPL licence, http://coq.inria.fr/. Coq is used in hundreds of sites. We have demanding users in industry (France Telecom R & D, Dassault-Aviation, Trusted Logic, Gemplus, Axalto) in the academic world in Europe (Scotland, Netherlands, Spain, Italy, Portugal, Poland) and in France (Bordeaux, Lyon, Marseille, Nancy, Nantes, Nice, Paris, Strasbourg). Coq is also used in the USA, Canada and Japan. Coq'Art, The first book dedicated to the Coq proof assistant, written by Yves Bertot and Pierre Casteran, has been published by Springer-Verlag in May 2004.
- Why, a verification condition generator with multi-prover output. Distributed as open source since July 2002, under the GPL licence, http://why.lri.fr. Why has been used by external researchers in published verifications of non-trivial algorithms (Efficient square root used in GMP, Knuth's algorithm for prime numbers). It is also an essential tool for Krakatoa and Caduceus.
- Krakatoa, a verification tool for Java programs, using Why as a back-end. Distributed as open source since March 2003, http://krakatoa.lri.fr. It is currently under experimentation at Axalto company (France), at NIA and DoCoMo labs (USA), and at University of Toronto (Canada).
- **Caduceus**, a verification tool for C programs, using Why as a back-end. Distributed since 2004 as open source, under the GPL licence, **http://why.lri.fr/caduceus**. It is currently under experimentation at Axalto and Dassault Aviation.
- Bibtex2html, a generator of HTML pages of bibliographic references. Distributed as open source, under the GPL licence, http://www.lri.fr/~filliatr/bibtex2html/. A Google search of the appropriate banner string reveals that, as of August 31, 2004, 4580 web pages have been generated using Bibtex2html.

#### Summer schools, tutorials, invited seminars

- Claude Marché, *Preuve Automatique de Terminaison*, invited seminar at École Normale Supérieure de Lyon (May 2003).
- Jean-Christophe Filliâtre, *Des programmes certifiés corrects*, invited seminar at École Normale Supérieure de Lyon (October 2001).
- Jean-Christophe Filliâtre, *Introduction à Coq*, invited seminar at the QSL meeting (Nancy, France, May 2002).
- Jean-Christophe Filliâtre, *Why : un outil de vérification générique*, invited seminars at CEA (February 2004), École Normale Supérieure (March 2004), and École Normale Supérieure de Lyon (June 2004).
- Jean-Christophe Filliâtre, Verifying C and Java programs, invited seminar at National Institute of Aerospace (Hampton, USA, March 2004).
- Christine Paulin, *Inductive Types* Course (8h) at the Summer school on Foundations of Security (Oregon, USA, June 2003).
- Christine Paulin, *Randomized Algorithms in Type Theory*, Seminar at University of Nijmegen (May 2004).

### Training and education (doctoral and post-doctoral)

Defended habilitation	s	
Name	Date defended	Current position
KESNER Delia	19/11/2001	Professor at Université Paris 7
Defended doctorates		
Name	Date defended	Current position
ABDICHE Mina	05/07/00	
URBAIN Xavier	01/10/01	assistant professor at IIE, Evry
BONELLI Eduardo	08/11/01	post-doc research assistant,
		Stevens Institute of Technology, USA
COURTIEU Pierre	14/12/01	assistant professor at Université d'Orléans
MONATE Benjamin	07/01/02	research engineer at CEA, Saclay
ROTARU Mina	12/07/02	engineer in private company
LOISELEUR Patrick	27/09/02	engineer in private company
WALUKIEWICZ Daria	09/04/03	teaching assistant at Warsaw University, Poland
FOREST Julien	26/09/03	post-doc at LORIA, Nancy
CHRZASZCZ Jacek	12/01/04	teaching assistant at Warsaw University, Poland
LETOUZEY Pierre	09/07/04	post-doc at University of Munich, Germany
BLANQUI Frédéric	28/09/01	researcher (CR) at INRIA Lorraine

#### **Graduate courses**

• DEA Programmation: Sémantique, Preuves et Langages:

- Termes en logique du premier ordre: Jean-Pierre Jouannaud (2000), Claude Marché (2000).
- *Calcul des constructions inductives:* Christine Paulin (2000-2004), Claude Marché and Jean-Christophe Filliâtre (2004).
- Terminaison: Claude Marché (2003-2004).

#### Other research-related teaching activities

• Christine Paulin organized 3-days tutorials on the Coq proof assistant (April 10-12, 2001; February 25-27, 2002), with 10-15 participants from academic or industrial research centers each time.

Demons

Dissemination and technology transfer Training and education

#### Publications

#### International peer-reviewed journals

- [1] F. Blanqui, J.-P. Jouannaud, and M. Okada. Inductive Data Type Systems. *Theoretical Computer Science*, 272(1-2):41-68, 2002, Elsevier Science Publishers.
- [2] E. Bonelli. Perpetuality in a named lambda calculus with explicit substitutions. *Mathematical Structures in Computer Science*, 11(1), 2001, Cambridge Univ. Press.
- [3] A. Boudet. Unification of higher-order patterns modulo simple syntactic equational theories. *Discrete Mathematics and Theoretical Computer Science*, 4(1):11-30, 2000, French association of DMTCS. Electronic journal.
- [4] A. Bouhoula and J.-P. Jouannaud. Automata-driven automated induction. *Information and Computation*, 169(1), 2001, Elsevier Science Publishers.
- [5] A. Bouhoula, J.-P. Jouannaud, and J. Meseguer. Specification and proof in membership equational logic. *Theoretical Computer Science*, 236:35-132, 2000, Elsevier Science Publishers.
- [6] G. Ciobanu and M. Rotaru. A pi-calculus machine. *electronic Journal of Universal Computer Science*, 6(1), 2000, Springer-Verlag.
- [7] S. Conchon and S. Krstic. Strategies for Combining Decision Procedures. *Theoretical Computer Science*, 2004, Elsevier Science Publishers.
- [8] R. David and B. Guillaume. A λ-calculus with explicit weakening and explicit substitution. *Mathematical Structures in Computer Science*, 11(1), 2001, Cambridge Univ. Press.
- [9] J.-C. Filliâtre. Formal Proof of a Program: Find. *Science of Computer Programming*, 2001, Elsevier Science Publishers. To appear.
- [10] J.-C. Filliâtre. Verification of Non-Functional Programs using Interpretations in Type Theory. *Journal of Functional Programming*, 13(4):709-745, July 2003, Cambridge Univ. Press.
- [11] J.-C. Filliâtre and F. Pottier. Producing All Ideals of a Forest, Functionally. *Journal of Functional Programming*, 13(5):945-956, September 2003, Cambridge Univ. Press.
- [12] D. Kesner. Confluence of extensional and non-extensional λ-calculi with explicit substitutions. *Theoretical Computer Science*, 238(1-2):183-220, 2000, Elsevier Science Publishers.
- [13] S. Krstic and S. Conchon. Canonization for Disjoint Unions of Theories. *Information and Computation*, 2004, Elsevier Science Publishers. Special Issue of Information and Computation dedicated to a refereed selection of papers presented at CADE-19.
- [14] C. Marché, C. Paulin-Mohring, and X. Urbain. The Krakatoa tool for certification of Java/JavaCard programs annotated in JML. *Journal of Logic and Algebraic Programming*, 58(1-2):89-106, 2004, Elsevier Science Publishers. http://krakatoa.lri.fr.
- [15] C. Marché and X. Urbain. Modular and incremental proofs of AC-termination. *Journal of Symbolic Computation*, 38:873-897, 2004, Academic Press.
- [16] M. Müller, J. Niehren, and R. Treinen. The first-order theory of ordering constraints over feature trees. *Discrete Mathematics and Theoretical Computer Science*, 4(2):193-234, Sept. 2001, French association of DMTCS. Electronic journal.
- [17] J. Niehren, S. Tison, and R. Treinen. On Rewrite Constraints and Context Unification. *Information Processing Letters*, 74(1-2):35-40, 2000, Elsevier Science Publishers.

- [18] F. Seynhaeve, S. Tison, M. Tommasi, and R. Treinen. Grid structures and undecidable constraint theories. *Theoretical Computer Science*, 258(1-2):453-490, May 2001, Elsevier Science Publishers.
- [19] X. Urbain. Modular and Incremental Automated Termination Proofs. *Journal of Automated Reasoning*, 2004, Kluwer Academic Publishers.

#### **Edited books**

[20] H. Comon, C. Marché, and R. Treinen, editors. Constraints in Computational Logics, volume 2002 of Lecture Notes in Computer Science. Springer-Verlag, 2001.

#### **Book chapters**

- [21] J. A. Goguen, T. Winkler, J. Meseguer, K. Futatsugi, and J.-P. Jouannaud. *Software Engineering with OBJ: Algebraic Specification in Action*, chapter Introducing OBJ\*. Kluwer Academic Publishers, 2000.
- [22] J.-P. Jouannaud and R. Treinen. *Constraints and Constraint Solving: An Introduction*, chapter 1, pages 1-46. Volume 2002 of Comon et al. [20], 2001.
- [23] L. Puel. *La sécurité sociale*, chapter Rapport de la 6<sup>e</sup> chambre de la Cour des Comptes: Le système d'information de la caisse nationale d'Assurance Vieillesse. édition des journaux officiels, Sept. 2003.

#### Major international peer-reviewed conferences

- [24] F. Blanqui. Termination and confluence of higher-order rewrite systems. In L. Bachmair, editor, 11<sup>th</sup> International Conference on Rewriting Techniques and Applications, volume 1833 of Lecture Notes in Computer Science, Norwich, UK, July 2000. Springer-Verlag.
- [25] F. Blanqui. Definitions by Rewriting in the Calculus of Constructions. In *Sixteenth Annual IEEE Symposium on Logic in Computer Science*. IEEE Comp. Soc. Press, 2001.
- [26] E. Bonelli, D. Kesner, and A. Ríos. A de bruijn notation for higher-order rewriting. In L. Bachmair, editor, 11th International Conference on Rewriting Techniques and Applications, volume 1833 of Lecture Notes in Computer Science, pages 62-79, Norwich, UK, July 2000. Springer-Verlag.
- [27] E. Bonelli, D. Kesner, and A. Ríos. From Higher-Order to First-Order Rewriting (Extended Abstract). In A. Middeldorp, editor, *Rewriting Techniques and Applications*, volume 2051 of *Lecture Notes in Computer Science*, Utrecht, The Netherlands, May 2001. Springer-Verlag.
- [28] A. Boudet and E. Contejean. Combining Pattern E-unification Algorithms. In A. Middeldorp, editor, *Rewriting Techniques and Applications*, volume 2051 of *Lecture Notes in Computer Science*, pages 63-76, Utrecht, The Netherlands, May 2001. Springer-Verlag.
- [29] G. Ciobanu and M. Rotaru. JC-Nets. In *Machines, Computations and Universality* (*MCU2001*), volume 2055 of *Lecture Notes in Computer Science*, pages 190-201, Chisinau, Moldova, May 2001. Springer-Verlag.
- [30] E. Contejean. A certified AC matching algorithm. In V. van Oostrom, editor, *15th International Conference on Rewriting Techniques and Applications*, volume 3091 of *Lecture Notes in Computer Science*, pages 70-84, Aachen, Germany, June 2004. Springer-Verlag.

Laboratoire de Recherche en Informatique / 2000-2004 report

Demons Publications

- [31] E. Contejean, A. Coste, and B. Monate. Rewriting techniques in theoretical physics. In L. Bachmair, editor, 11<sup>th</sup> International Conference on Rewriting Techniques and Applications, volume 1833 of Lecture Notes in Computer Science, pages 80-94, Norwich, UK, July 2000. Springer-Verlag.
- [32] E. Contejean, C. Marché, B. Monate, and X. Urbain. Proving termination of rewriting with cime. In A. Rubio, editor, *Extended Abstracts of the 6th International Workshop on Termination, WST'03*, pages 71-73, June 2003. http://cime.lri.fr.
- [33] P. Corbineau. First-order reasoning in the calculus of inductive constructions. In S. Berardi, M. Coppo, and F. Damiani, editors, 3<sup>rd</sup> International Workshop on Types for Proofs and Programs, volume 3085 of Lecture Notes in Computer Science, pages 162-177, Torino, Italy, Apr. 2004. Springer-Verlag.
- [34] R. D. Cosmo, D. Kesner, and E. Polonovski. Proof Nets and Explicit Substitutions. In J. Tiuryn, editor, *Foundations of Software Science and Computation Structures*, volume 1784 of *Lecture Notes in Computer Science*, Berlin, Germany, Mar. 2000. Springer-Verlag.
- [35] J. Courant. Explicit Universes for the Calculus of Constructions. In V. A. Carreño, C. A. Muñoz, and S. Tahar, editors, *Theorem Proving in Higher Order Logics: 15th International Conference, TPHOLs 2002*, volume 2410 of *Lecture Notes in Computer Science*, pages 115-130, Hampton, VA, USA, Aug. 2002. Springer-Verlag.
- [36] J. Courant. Strong Normalization with Singleton Types. In S. V. Bakel, editor, Second Workshop on Intersection Types and Related Systems, volume 70 of Electronic Notes in Theoretical Computer Science, Copenhaguen, Danemark, July 2002. Elsevier Science Publishers.
- [37] P. Courtieu. Normalized Types. In L. Fribourg, editor, *Proceedings of the Annual Conference of the European Association for Computer Science Logic*, volume 2142 of *Lecture Notes in Computer Science*, pages 554-569, Paris, France, Sept. 2001. Springer-Verlag.
- [38] F. Durán, S. Lucas, J. Meseguer, C. Marché, and X. Urbain. Proving termination of membership equational programs. In *ACM SIGPLAN 2004 Symposium on Partial Evaluation and Program Manipulation*, Verona, Italy, Aug. 2004. ACM Press.
- [**39**] J.-C. Filliâtre and P. Letouzey. Functors for Proofs and Programs. In *Proceedings of The European Symposium on Programming*, volume 2986 of *Lecture Notes in Computer Science*, pages 370-384, Barcelona, Spain, April 2004.
- [40] J.-C. Filliâtre and C. Marché. Multi-Prover Verification of C Programs. In *Sixth International Conference on Formal Engineering Methods*, Lecture Notes in Computer Science, Seattle, USA, Nov. 2004. Springer-Verlag.
- [41] J.-C. Filliâtre, S. Owre, H. Rueß, and N. Shankar. ICS: Integrated Canonization and Solving (Tool presentation). In G. Berry, H. Comon, and A. Finkel, editors, *Proceedings* of CAV'2001, volume 2102 of Lecture Notes in Computer Science, pages 246-249. Springer-Verlag, 2001.
- [42] J. Forest and D. Kesner. Expression Reduction Systems with Patterns. In R. Nieuwenhuis, editor, 14th International Conference on Rewriting Techniques and Applica tions, volume 2706 of Lecture Notes in Computer Science, pages 107-122, Valencia, Spain, June 2003. Springer-Verlag.
- [43] B. Jacobs, C. Marché, and N. Rauch. Formal Verification of a Commercial Smart Card Applet with Multiple Tools. In *Algebraic Methodology And Software Technology*, Lecture Notes in Computer Science, Stirling, UK, July 2004. Springer-Verlag.
- [44] J.-P. Jouannaud. Model checking versus theorem proving for verifying protocols. In *In Proc. Distributed Systems Verification and Validation*, Apr. 2000.
- [45] A. Koller, J. Niehren, and R. Treinen. Dominance Constraints: Algorithms and Complexity. In M. Moortgat, editor, *Logical Aspects of Computational Linguistics 1998*, volume 2014 of *Lecture Notes in Artificial Intelligence*, pages 106-125, Grenoble, France, 2001. Springer-Verlag.

- [46] P. Letouzey. A new extraction for Coq. In H. Geuvers and F. Wiedijk, editors, *TYPES 2002*, volume 2646 of *Lecture Notes in Computer Science*. Springer, 2003.
- [47] P. Letouzey and L. Théry. Formalizing Stålmarck's algorithm in Coq. In J. Harrison and M. Aagaard, editors, *Theorem Proving in Higher Order Logics: 13th International Conference, TPHOLs 2000*, volume 1869 of *Lecture Notes in Computer Science*, pages 387-404. Springer-Verlag, 2000.
- [48] A. Miquel. A strongly normalising Curry-Howard correspondence for IZF set theory. In *Computer Science Logic, CSL'03*, Lecture Notes in Computer Science. Springer-Verlag, 2003.
- [49] A. Miquel and B. Werner. The not so simple proof-irrelevant model of CC. In H. Geuvers and F. Wiedijk, editors, *TYPES 2002*, volume 2646 of *Lecture Notes in Computer Science*, pages 240-258. Springer, 2003.
- [50] E. Ohlebusch, C. Claves, and C. Marché. TALP: A Tool for the Termination Analysis of Logic Programs. In L. Bachmair, editor, 11<sup>th</sup> International Conference on Rewriting Techniques and Applications, volume 1833 of Lecture Notes in Computer Science, pages 270-273, Norwich, UK, July 2000. Springer-Verlag. Available at http://bibiserv.techfak.uni-bielefeld.de/talp/.
- [51] E. Ohlebusch, C. Claves, and C. Marché. The talp tool for termination analysis of logic programs. In A. Rubio, editor, *Extended Abstracts of the 6th International Workshop on Termination, WST'03*, June 2003. http://bibiserv.techfak.uni-bielefeld.de/talp/.
- [52] N. Oury. Observational equivalence and program extraction in the Coq Proof Assistant. In M. Hofmann, editor, *TLCA*, volume 2701 of *Lecture Notes in Computer Science*, pages 271-285. Springer, 2003.
- [53] C. Paulin-Mohring. Modelisation of timed automata in Coq. In N. Kobayashi and B. Pierce, editors, *Theoretical Aspects of Computer Software (TACS'2001)*, volume 2215 of *Lecture Notes in Computer Science*, pages 298-315. Springer-Verlag, 2001.
- [54] F. Prost. A Static Calculus of Dependencies for the λ-cube. In M. Abadi, editor, *Fiveteenth Annual IEEE Symposium on Logic in Computer Science*, Santa Barbara, California, June 2000. IEEE Comp. Soc. Press.
- [55] Z. Su, A. Aiken, J. Niehren, T. Priesnitz, and R. Treinen. The First-Order Theory of Subtyping Constraints. In J. Mitchell, editor, *Conference Record of the 29th Symposium on Principles of Programming Languages*, pages 203-216, Portland, OR, USA, Jan. 2002. ACM Press.
- [56] R. Treinen. Predicate Logic and Tree Automata with Tests. In J. Tiuryn, editor, *Foundations of Software Science and Computation Structures*, volume 1784 of *Lecture Notes in Computer Science*, pages 329-343, Berlin, Germany, Mar. 2000. Springer-Verlag.
- [57] X. Urbain. Automated Incremental Termination Proofs for Hierarchically defined Term Rewriting Systems. In R. Goré, A. Leitsch, and T. Nipkow, editors, *International Joint Conference on Automated Reasoning*, volume 2083 of *Lecture Notes in Artificial Intelligence*, pages 485-498, Siena, Italy, June 2001. Springer-Verlag.

#### National peer-reviewed conferences

- [58] J.-C. Filliâtre. La supériorité de l'ordre supérieur. In *Journées Francophones des Langages Applicatifs*, pages 15-26, Anglet, France, Jan. 2002.
- [59] J. Signoles. Calcul statique des applications de modules paramétrés. In *Journées Francophones des Langages Applicatifs*, 2003.
- [60] D. Walukiewicz-Chrząszcz. Termination of rewriting in the Calculus of Constructions. In *Proceedings of the Workshop on Logical Frameworks and Meta-languages, Santa Barbara, California,* 2000. Part of the LICS'2000.

Research groups

#### Demons Publications

#### Other conferences and workshops

- [61] E. Contejean and A. P. Tomas. On Symmetries in Systems Coming from AC-Unification of Higher-Order Patterns. In P. Flener and J. Pearson, editors, *SymCon'01, Symmetry in Constraints,* Paphos, Cyprus, December 2001.
- [62] V. Cremet, K. Hasebe, J.-P. Jouannaud, A. Kremer, and M. Okada. Fatalis : Real time processes as linear logic specifications. In *International Workshop on Automated Verification* of *Infinite-State Systems, Varsaw*, 2003.

#### **Doctoral dissertations and Habilitations**

- [63] M. Abdiche. π-calcul et sous-typage: inférence de types et codages du λ-calcul dans le π-calcul. PhD thesis, Université Paris-Sud, July 2000.
- [64] C. Alvarado. *Réflexion pour la réécriture dans le calcul des constructions inductives*. PhD thesis, Université Paris-Sud, Dec. 2002.
- [65] F. Blanqui. Théorie des Types et Récriture. PhD thesis, Université Paris-Sud, 2001.
- [66] E. Bonelli. *Substitutions explicites et réécriture de termes*. PhD thesis, Université Paris-Sud, Orsay, France, Nov. 2001.
- [67] J. Chrząszcz. *Modules in Type Theory with Generative Definitions*. PhD thesis, Warsaw University, Poland and Université de Paris-Sud, Jan. 2004.
- [68] P. Courtieu. Représentation d'algèbres non libres en théorie des types. PhD thesis, Université Paris-Sud, 2001.
- [69] J. Forest. Réécriture d'ordre supérieur avec motifs. Thèse de doctorat, Université Paris-Sud, Orsay, France, Sept. 2003. http://www.pps.jussieu.fr/~forest/these/TheseForest.ps.gz.
- [70] D. Kesner. Un parcours de la réécriture d'ordre supérieur à la réécriture du premier ordre. Thèse d'habilitation à diriger des recherches, Université Paris-Sud, 2001.
- [71] P. Letouzey. *Programmation fonctionnelle certifiée: l'extraction de programmes dans l'assistant Coq.* PhD thesis, Université Paris-Sud, July 2004.
- [72] P. Loiseleur. *Spécifications et preuves de programmes distribués, une approche basée sur les réseaux d'objets CORBA*. PhD thesis, Université Paris-Sud, Sept. 2002.
- [73] B. Monate. Propriétés uniformes de familles de systèmes de réécriture de mots paramétrées par des entiers. PhD thesis, Université Paris-Sud, Orsay, France, Jan. 2002.
- [74] M. Rotaru. *Sur une théorie unificatrice des modèles de concurrence*. PhD thesis, Cluj University, Romania and Université de Paris-Sud, Jan. 2004.
- [75] X. Urbain. Approche incrémentale des preuves automatiques de terminaison. Thèse de doctorat, Université Paris-Sud, Orsay, France, Oct. 2001. http://www.lri.fr/~urbain/textes/ these.ps.gz.
- [76] D. Walukiewicz-Chrząszcz. Termination of Rewriting in the Calculus of Constructions. PhD thesis, Warsaw University, Poland and Université de Paris-Sud, Apr. 2003.

#### Reports

[77] F. Baader, V. Diekert, C. Tinelli, and R. Treinen. Unif 2001 - 15<sup>th</sup> international workshop on unification. Technical Report DII 09/01, Università degli Studi di Siena, Dipartimento di Ingegneria dell'Informazione, Siena, Italy, June 2001.

- [78] B. Bérard, P. Castéran, E. Fleury, L. Fribourg, J.-F. Monin, C. Paulin, A. Petit, and D. Rouillard. Automates temporisés CALIFE. Fourniture F1.1, Calife, 2000.
- [79] P. Castéran, E. Freund, C. Paulin, and D. Rouillard. Bibliothèques coq et isabelle-hol pour les systèmes de transitions et les p-automates. Fourniture F5.4, Calife, 2000.
- [80] N. Cataño, M. Gawlowski, M. Huisman, B. Jacobs, C. Marché, C. Paulin, E. Poll, N. Rauch, and X. Urbain. Logical Techniques for Applet Verification. Deliverable 5.2, IST VerifiCard project, 2003. http://www.cs.kun.nl/VerifiCard/files/deliverables/deliverable\_5\_2.pdf.
- [81] E. Contejean, C. Marché, A.-P. Tomás, and X. Urbain. Mechanically proving termination using polynomial interpretations. Research Report 1382, LRI, 2004.
- [82] P. Corbineau. First-order reasoning in the calculus of inductive constructions. Research Report 1380, LRI, Dec. 2003.
- [83] J. Courant. MC<sub>2</sub>: A Module Calculus for Pure Type Systems. Research Report 1292, LRI, Sept. 2001.
- [84] P. Courtieu. Normalized Types. Research Report 1277, LRI, 2001.
- [85] J.-C. Filliâtre. Design of a proof assistant: Coq version 7. Research Report 1369, LRI, Oct. 2000.
- [86] J.-C. Filliâtre. Hash consing in an ML framework. Research Report 1368, LRI, September 2000.
- [87] J.-C. Filliâtre. Why: a multi-language multi-prover verification tool. Research Report 1366, LRI, Mar. 2003. http://www.lri.fr/~filliatr/ftp/publis/why-tool.ps.gz.
- [88] J. Forest. A Weak Calculus with Explicit Operators for Pattern Matching and Substitution. Research Report 1313, LRI, May 2002.
- [89] K. Kusakari, C. Marché, and X. Urbain. Termination of Associative-Commutative Rewriting using Dependency Pairs Criteria. Research Report 1304, LRI, 2002. http://www.lri.fr/~urbain/textes/rr1304.ps.gz.
- [90] R. Treinen. Undecidability of the emptiness problem of reduction automata with component-wise tests. June 2001.
- [91] X. Urbain. Modular & incremental automated termination proofs. Research Report 1326, LRI, 2002.
- [92] X. Urbain. Modular & incremental proofs of AC-termination. Research Report 1317, LRI, 2002.

#### Other publications

- [93] J. Creci. Certification d'algorithmes d'arithmétique réelle exacte dans le système Coq. Rapport de DEA, Université Paris-Sud, Orsay, France, Sept. 2002.
- [94] Coq development team. *The Coq Proof Assistant Reference Manual Version V7.4*, Feb. 2003. http://coq.inria.fr/doc/main.html.
- [95] G. Huet, G. Kahn, and C. Paulin-Mohring. *The Coq Proof Assistant A tutorial Version* 7.1, Oct. 2001.
- [96] G. Huet, G. Kahn, and C. Paulin-Mohring. *The Coq Proof Assistant A tutorial Version* 8.0, Apr. 2004.
- [97] A. Kremer. Spécification et vérification de systèmes temporisés réactifs en logique linéaire. Rapport de DEA, Université Paris-Sud, 2002.
- [98] P. Letouzey. Exécution de termes de preuves: une nouvelle méthode d'extraction pour

Research groups

Demons Publications le Calcul des Constructions Inductives. Rapport de DEA, 2000.

- [99] N. Oury. Équivalence vis-à-vis de l'observation et extraction de programmes. Rapport de DEA, Université Paris-Sud, Orsay, France, Sept. 2002.
- [100] J. Signoles. Calcul statique des applications de foncteurs en présence d'effets de bord. Rapport de DEA, Université Paris-Sud, Sept. 2002. In French.
- [101] The Coq Development Team. The Coq Proof Assistant Reference Manual Version V7.3, May 2002. http://coq.inria.fr.
- [102] The Coq development team. The Coq Proof Assistant Reference Manual V7.2. Technical Report 255, INRIA, France, Mar. 2002.
- [103] The Coq Development Team. *The Coq Proof Assistant Reference Manual Version V8.0,* Apr. 2004. http://coq.inria.fr.


Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Architectures parallèles

**Responsable : Olivier Temam** 

e groupe Architectures Parallèles s'est transformé en deux équipes de recherche distinctes, l'une focalisée sur l'architecture et la programmation des processeurs, et l'autre sur les grilles de calcul et les clusters. Cette évolution correspond à l'évolution du domaine de recherche, où les efforts, initialement consacrés aux super-calculateurs conçus à partir de processeurs spécialisés (requèrant donc une expertise en architecture parallèle comme en architecture de processeur), ont été progressivement remplacés par des machines allant de stations de travail monoprocesseurs à des clusters de stations/PCs au niveau d'un réseau local voire d'Internet. Tous ces clusters utilisent maintenant des processeurs généralistes très rapides et complexes, et les accroissements de performance sont obtenus à la fois grâce à l'évolution de la performance des processeurs individuels et à l'augmentation de la taille des clusters. En conséquence, la recherche s'est également scindée en deux parties: l'une consacrée à l'organisation de ces clusters, et l'autre à l'évolution des processeurs eux-mêmes. Bien que les deux sujets apparaissaient ici au sein de la même équipe de recherche, le groupe Clusters et Grille est actuellement en train de rejoindre l'équipe Parallélisme.

> Chacun des deux groupes est impliqué dans un projet commun du PCRI. Le groupe Architecture est impliqué dans le projet PCRI Alchemy, décrit dans cette section, tandis que le groupe



Clusters et Grilles est impliqué dans le projet PCRI Grand Large, décrit dans la section sur l'équipe Parallélisme.



## **Parallel Architectures**

Head: Olivier Temam

The Parallel Architecture research group has evolved from a single research team into two distinct research teams, one focusing on processor architecture and programming, the other on grid and cluster computing. This evolution matches the evolution of the research domain, where efforts which initially focused on supercomputers built with custom processors (requiring expertise in both parallel machines and processors) have been replaced by machines ranging from single workstations to LAN-based and even Web-based clusters. All these clusters now rely on very fast and complex general-purpose processors, and performance improvements are achieved both through the fast evolution of processors and the increasing cluster size. As a result, research has been split into two domains: one focusing on the organization of these clusters and the other focusing on the processors themselves. While the two topics are presented in this report as two research teams within the same research group, the Cluster and Grid team is in the process of joining the Parall group.

Each of the two groups is involved in a PCRI project. The Micro-Architecture team is involved in the Alchemy PCRI project, described in this section. The Cluster and Grid group is involved in the Grand Large PCRI project, described with the Parall group.



Research groups Archi Introduction

## Research group members

Personnel as of 01/01/2004

Full time faculty					
Name	First Name	Position*	Institution		
Processor research tean	n				
DRACH	Nathalie	MC	Université Paris XI		
ETIEMBLE	Daniel	PRHC	Université Paris XI		
GRUAU	Frédéric	MC	Université Paris XI		
TEMAM	Olivier	PR2	Université Paris XI		
Cluster and Grid research te	eam				
GERMAIN-RENAUD	Cécile	МС	Université Paris XI		
Doctoral students					
Processor research team	n				
COVELIERS	Alexandre	AM	IUT d'Orsay		
DUPRE	Michaël	AM	Rectorat de Versailles		
FAVRE	Sébastien	D	НР		
GIRBAL	Sylvain	D	CEA		
GRACIA PEREZ	Daniel	D	Université Paris XI		
LHUILLIER	Yves	AC	Université Paris XI		
MOUCHARD	Gilles	ATER	Université Paris XI		
PARELLO	David	D	COMPAQ France		
YEHIA	Sami	ATER	Université Paris XI		
Cluster and Gria resear	cri leam				
BOSILCA	George	AM	IUT d'Orsay		
DJILALI-SAIAH	Abderrahmane	AM	Université Paris XI		
KRAWEZIK	Géraud	D	EADS		
TEXIER	Romain	AM	IUT d'Orsay		
BOUTEILLIER	Aurélien	AM	IUT d'Orsay		

Non-LRI personnel of PCRI project Alchemy			
COHEN	Albert	CR	INRIA
EISENBEIS	Christine	DR	INRIA
AMIRANOFF	Pierre	D	
BASTOUL	Cédric	D	
CARRIBAULT	Patrick	D	
GONZALEZ SIGLER	Marc	Engineer	INRIA

\*See Glossary for acronyms.

See Parall group for non-LRI personnel of PCRI project Grand Large.

Long term visitor						
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
Cluster and	Grid researc	h team				
SELIKOV	Anton	Russian	Mathematics Institute Novosibirsk	1/10/01	31/03/03	UPS Ministère de la Rech.

## **Group evolution**

Since 2000, three permanent researchers have left the group, while four new assistant professors have been recruited :

Processor research team:

#### **Faculty Arrivals:**

- 2001: Frédéric Gruau, Assistant Professor;
- 2004: Julien Cohen, ATER, finishing his Ph.D. from Université d'Evry.

#### **Faculty Departures:**

- 2000: Jean-Louis Béchennec, CR CNRS; left for IRCCYN, Nantes;
- 2004: Nathalie Drach, Assistant Professor; became Professor at LIP6, Paris 6, in Alain Greiner's group;
- Daniel Etiemble, Professor, was on leave at University of Toronto, Canada, in 2000-2002.
- Olivier Temam, Professor, was appointed as research director (DR) at INRIA in 2004, where he continues to lead the ALCHEMY PCRI project.
- Four Ph.D. students defended their thesis, while eleven new Ph.D. students joined the group (see section 10 for details).

• Franck Cappello, CR CNRS, was appointed research director (DR) at INRIA in 2003, where he started the Grand Large PCRI project described in the Parall group.

Cluster and Grid research team:

- Cecile Germain spent two years at *Laboratoire de l'Accélérateur Linéaire* (LAL, Université Paris-Sud), on a temporary CNRS position in 2002-2003.
- George Bosilca is currently at ICL laboratory Knoxville.
- Two Ph.D. students defended their thesis, while five new Ph.D. students joined the group (see section 10 below for details).

Res	earch g	group	s	
Ar	chi			
Me	mbers			

B.6.2 / Archi

## Research description: micro-architecture team



## ALCHEMY AND MICRO-ARCHITECTURE TEAM APPROACH

Figure 1: Alchemy and Micro-architecture team approach.

The general research topics of the micro-architecture group are architectures, languages and compilers for high-performance embedded and general-purpose processors (Figure 1). The micro-architecture group investigates alternative solutions to incremental architecture and compiler optimizations for high-performance general-purpose and embedded processors. The increasing complexity of high-performance processor architectures has two main consequences. In the short term, the inability to embed a sufficiently accurate architecture model in compilers makes it increasingly difficult to generate efficient program optimizations, and thus to achieve high sustained performance. In the long term, the architecture complexity makes it increasingly difficult to scale processor architectures, more exactly to translate technology improvements into higher sustained performance. We are developing two approaches that correspond respectively to the short-term and long-term issues outlined above, and we are also investigating associated technology and methodology issues.

## PCRI Project ALCHEMY



ALCHEMY is a PCRI project created in the fall of 2003 as a result of the merger between the former INRIA group A3 on compilation, and the LRI research team on processor architecture. ALCHEMY stands for Architectures, Languages and Compilers to Harness the End of Moore Years. It addresses both the complex but traditional processor architectures implemented using the current photolithographic processes, and novel architecture/language paradigms compatible with future and alternative technologies.

## **Scientific Context**

For the past 30 years, processor performance has apparently regularly and smoothly increased in parallel with the technology improvements. In fact these improvements both bred and resulted from a considerable increase in architecture complexity. For the industry, this complexity means that architecture design is an increasingly sophisticated process; not surprisingly American companies increasingly rely on Ph.D.-level engineers to design such processors, while European companies are only starting to edge in that direction.

Scientifically, this complexity has two serious consequences. A short-term consequence is that it is increasingly difficult to generate an efficient program for such complex architectures. As a result, current architectures only exploit a fraction of their potential/peak performance; this is especially true for the more sophisticated general-purpose processors, but it is increasingly true for recent embedded processors as well. Current optimization techniques rely on static optimization (program analysis at compile-time only), which requires embedding a detailed architecture model in the compiler in order to generate that best matches the underlying architecture; naturally, as the architecture complexity increases, this is increasingly difficult to achieve and the resulting optimization efficiency does not increase as fast.

A long-term consequence of this complexity is that it is increasingly difficult to scale up such architectures. Each new component added to the architecture improves the overall architecture performance, but it also creates potential bottlenecks that will need to be addressed by updated or new mechanisms, and so on. The current approach essentially relies on speeding up a Von Neumann-like centralized architecture (a single/centralized processing unit, and a single memory), and to devote most of the available on-chip space to the new components required to achieve the targeted speed. Besides the increasing architecture hardships, technology may ultimately limit this speed-oriented approach. For both reasons, some researchers are investigating alternative approaches to exploit the available on-chip space and translate it into performance.

## **General Approach**

The goal of our group is to tackle both of the above short-term and long-term issues that respectively threaten the efficient exploitation of current architectures and the evolution of future architectures. The first approach (short-term research on program optimizations) aims at coming up with a software environment that can be used by both research and industry for simplifying the task of optimizing programs on complex processor architectures. The role of the second approach is to investigate possibly radical modifications to the current architecture and programming models in order to come up with architectures that can scale up more easily and are compatible with upcoming (new) technologies. We are less concerned with medium-term architecture research for practical reasons: while incremental architecture modifications can certainly improve performance, we believe they will not be sufficient to let architectures scale up smoothly and regularly again; on the practical side, processor manufacturers already have very skilled architecture research and development groups for coming up with medium-term innovations for their next-generation products; on the other hand, processor manufacturers cannot always afford to investigate very long-term and risky alternatives, and for them to accept such radical changes they have to be anticipated long in advance. We believe that is a core role for academic researchers in this domain.

Research groups

Archi

Research description Micro architecture team

## Theme 1: A practical approach to program optimizations for complex architectures

• Participants: Alexandre Coveliers, Nathalie Drach, Michaël Dupré, Sylvain Girbal, David Parello, Olivier Temam, Nicolas Vasilache, Sami Yehia.



Figure 2: The increasing gap between maximal performance (Max) and observed performance (FP, INT).

The principle of our approach is to heavily rely on dynamic (run-time) analysis as a way to overcome the architecture complexity bottleneck (Figure 2). Using an iterative compilation approach, we want to learn over executions the appropriate method for optimizing a program based on detailed low-level information on the behaviour of the program on the architecture. In recent years, iterative optimization has received increased attention thanks to researchers like Mike O'Boyle at University of Edinburgh, UK, or Keith Cooper at Rice University. In this research, iterative optimization is used primarily to fine-tune the parameters of program optimizations applied to restricted code constructs. Moreover, most approaches attempt to show that they work by exhaustively searching the parameter space, but this is not a practical approach for iterative optimization. We want to address the issues faced by practical applications of iterative optimization and extend it to whole-program optimization parameters but program optimizations themselves and the location of their application in the program. We want to show that we can both achieve significant performance improvements and significantly reduce the program optimization effort by making it more systematic.

Toward practical applications of iterative optimization. This research work is divided into several steps and projects. First, instead of the top-down approach where compilers are progressively augmented with information from the architecture, such as current profile-based, iterative or dynamic compilation techniques, we have adopted a bottom-up approach to the architecture complexity issue: assuming we know everything about the behaviour of the program on the architecture (using cycle-level processor simulators), what can we do to improve its performance ? Based on extensive analysis of programs behaviours on a complex processor architecture, we have designed a systematic and iterative optimization process [54, 53]. While it is not yet implemented as a fully automatic iterative environment, it is systematic, and it has already been (and is still being) used successfully at HP France for quickly optimizing programs on complex processors for prospective customers (on the Alpha for now, extension to x86 is planned).

The second big issue is how to let iterative optimizations control the application of program transformations themselves. If program optimizations have to be applied/selected automatically, the search space now includes not only optimization parameters but optimizations themselves, and especially compositions of optimizations. While compilers include rigid sequences of optimizations, an iterative process can seek the best sequence for each code section. However, for that purpose, we must be able to compose long sequences of program transformations, and the current syntactic implementation of program transformations strongly limits this ability. Therefore, we are currently developing and implementing a framework based on the polyhedral representation of programs and transformations to easily compose very long sequences of program transformations.

The third issue is the software environment for scanning the search space, apply program transformations, collect feedback information and deduce the most appropriate next transformations to be tested. Moreover, for a practical application of iterative optimization, performance information deduced from one program execution, i.e., one data set, must be somehow exploited for other/next executions, i.e., other data sets). We have gathered preliminary results on iterative low-level optimizations for embedded processors that show that this approach is feasible [31], but we still have to confirm this with larger-scale experiments. We also have a prototype version of our software environment and are now investigating search space strategies.

Increasingly, we are now moving toward automating the whole process. A first step in that direction is to show that, for simple low-level optimizations, it is possible to directly rely on the detailed architecture description implicitly embedded in a cycle-level processor simulator to replace the static analysis of compilers. We have shown that a modified simulator of an embedded VLIW processor can automatically schedule assembly instructions, much like complex and costly Out-of-Order superscalar processors, but without any additional hardware cost [31]. We are also investigating applications of this approach to the specialization/idiomization of embedded processors [60]. Next steps include fully automated runs of the above software environment with the prospect of finding complex compositions of high-level program transformations.

*Related activities:* Esprit LTR MHAOTEU and OCEANS, Digital/Compaq/HP France grants, ACI grant, RNTL COP; in the past four years, four Ph.D. students have been or are still working on these topics.

## Theme 2: Revisiting the processor architecture/programming approach

• Participants: Frédéric Gruau, Yves Lhuillier, Pierre Palatin, Olivier Temam.

In the long term, we consider that both excessive processor architecture complexity and low sustained performance are rooted in the current architecture/programming model itself. More precisely, the current model fails in two ways: passing enough program semantics to the compiler and the architecture, and efficiently managing the increasing chip space brought by technology.

*Spatial architectures and programming.* With a very large chip space, it is likely that the centralized control used in current processor architectures may no longer be acceptable. One of the main challenges then becomes the management of one or several programs on a very large space. We are investigating very regular/homogeneous architectures, such as large arrays of processors. Note that tiled architectures have received increased attention in the past few years as a result of the increased difficulty of scaling up current superscalar processors. However, while some researchers agree that regular and space-oriented architectures are much easier to scale, most of them still attempt to program them with conventional programming styles and approaches. We are exploring different programming styles that may be more compatible with these space-oriented architectures. We are particularly interested in programming approaches that break down a program into a set of coordinated "local" actions (local to a node and data), and that relieve the programmer from thinking through the global management of his program on the architecture [87, 43] (Figure 3).

Research groups

Archi Research description Micro architecture team The architecture is augmented with a support that is aware of this local program breakdown and that actually manages the program execution, instead of the compiler or the programmer. We argue that, provided that we strike the right balance between the architecture, compiler and user effort, it is possible to unveil relatively coarse-grain parallelism (coarser than ILP) and to take advantage of it without excessively complex architectures or compilers.



We call this combination of local programming and hardware support *self-organized architectures*. Such self-organized systems are in fact widespread in nature, especially in biological systems, and part of our longer-term work consists in understanding/extracting the simple rules used in complex natural systems (in cooperation with researchers in biology) that can serve to implement desired properties in computing systems. For instance, pressure and elasticity can respectively express the need for more space and to the need to get closer during communications. This research direction may seem futuristic, and part of our objective is indeed to investigate alternative computing models that may be compatible with future technologies such as molecular electronics or biologically assembled chips. Another shorter-term goal however is to apply such local rules to current architectures, e.g. Simultaneous Multithreaded architectures (SMTs) and Chip Multiprocessors (CMPs).

Related activities: two Ph.D. students are working on these topics.

## Theme 3: A transversal research direction: methodology and simulation

#### • Participants: Nathalie Drach, Daniel Gracia-Perez, Gilles Mouchard, Olivier Temam.

Simulators are needed for both architecture and program optimization research. For architecture research, they serve to implement detailed cycle-level models to evaluate new ideas; for program optimization research, they serve to get a better understanding of the detailed behaviour of a program on processor architectures. As a result, they are key tools for our research group. However, as processor architectures (Figure 4) and program complexity increase so does the development and execution time of these simulators. Therefore,

we have progressively investigated methodology issues for the sake of efficiency of our other research activities. The main principles of our approach are the development of modular simulators, and the increased sharing and reuse of researchers' expertise and efforts through a common library of simulator components [42].



Figure 4: Processor model.

Simulators are used in most processor architecture research. While most research papers include performance measurements such as IPC and more specific metrics, these numbers tend to be distrusted because the simulator used is rarely publicly available, or at least not in a standard and reusable form. As a result, it is not easy to check for design and implementation hypotheses, potential simplifications or errors. Since the goal of most processor architecture research is to improve performance, it is frustrating not to be able to clearly quantify the benefit of a new architecture mechanism with respect to previously ones. Many researchers wonder what is the best mechanism, at least for a given processor architecture and benchmark suite, or even for a single benchmark. But many consider that it is too timeconsuming to implement a significant set of previous mechanisms based on the research articles only. We argue that, provided that a few groups start populating a common library of modular simulator components, a broad and systematic quantitative comparison of architecture ideas may become realistic, at least for some research topics. We are in the process of developing such a library, called MicroLib. Because the main flaw of modular simulators is their poor performance, we are also working on techniques to speed up simulation, using parallelization and sampling.

Several simulator models, components and tools that were initially used only internally are now being disseminated through a general web site (www.microlib.org). Some tools were used heavily (2000+ downloads for the PowerPC simulator on June 2004, several hundred downloads for other tools, several articles using/referencing these tools). Besides program optimization and architecture research, this work is also being applied to the software test of large embedded systems in collaboration with CEA.

*Related activities:* RNTL ATLAS, ACI grant. Two Ph.D. students have been or are working on these topics.

#### Research groups Archi

Research description Micro architecture team

## Theme 4. Beyond performance

• Participants: Sébastien Favre, Olivier Temam.

We have recently started a new research project that is both different from and complementary to our other projects. All of the above projects are focused on the issue of performance, especially the performance of processors and systems. However, to a large extent, we are not taking enough advantage of our knowledge of processor and system architecture. Since we know several years in advance the capabilities of future processors, we can anticipate as well the possible structures of future systems and their applications. For instance, architects knew several years in advance when cheap processors or circuits would be capable of real-time MPEG-2 encoding, and thus when hard-drive based VCRs could become a reality. We want to open up more to system architecture in the general sense and especially its applications, and propose software or hardware prototypes when appropriate. While this type of work may be considered borderline research and harder to publish, we also believe it is a healthy (and fun) exercise for a research group to think about the potential applications of its mainstream research projects. When possible, we want these to be federating projects within the research group. We have started a first project on the evolution of the home PC towards more operating system-level simplicity and the geographic distribution of peripherals and computing elements.

Related activities: HP France grant. One Ph.D. student is working on this topic.

#### B.6.2 / Archi

## Research description: cluster and grid team

• Participants: Franck Cappello, Cécile Germain, George Bosilca, Abderrahmane Djilali-Saiah, Gilles Fedak, Géraud Krawezik, Romain Texier.

The Cluster and Grid research team is investigating the issues raised by high performance computing on Clusters and Large Scale Distributed Systems (LSDS). Our research focuses on middleware and low-level programming environments design, proof and experiments. Fundamentally, we address the impact of large scale on distributed systems with several methodological tools: theoretical models, simulators, emulators and real-size systems.

The goals of our work are:

- to study experimentally, and formally, the fundamental mechanisms of LSDS for high performance computing;
- to design, implement, validate and test real software, middleware and platforms;
- to define, evaluate and experiment approaches for programming applications on these platforms.

Compared to other European and French projects, we gather skills in large scale systems (large scale scheduling, volatility tolerance, heterogeneity, inter administration domain security, etc.) acquired with the XtremWeb project, formal design and validation of algorithms and protocols for distributed systems and programming, evaluation, analysis and definition of programming languages and environments for parallel architectures and distributed systems.

We pursue short and long-term research aiming to have a scientific and industrial impact in the areas of LSDS, grid computing and large clusters. Research topics include:

- the design of middleware to expand the application domain of Desktop Grid;
- resource discovery engines on large-scale systems with volatile participants;
- large-scale storage on volatile nodes;
- simulation of large-scale scheduling;
- fault-tolerant MPI for large-scale systems;
- algorithms for large-scale fault tolerance;
- protocol verification;
- algorithms, programming and evaluation of scientific applications on desktop Grids;
- tools and languages for large-scale computing.

In our longer-term research, we investigate the conditions for the convergence of Global Computing, Peer-to-Peer (P2P) and Grid systems (how Grid Services can be used in Desktop Grid) and experimental tools for improving the methodology associated with research in LSDS. For example we are in charge of the Grid eXplorer project funded by the French ministry of research as well as the Grid5000 national project (Franck Cappello chair of the steering committee).

The Cluster and Grid research team has now joined the Parall group. Please refer to the sections describing the PCRI Grand Large project (B.7.2) and the Global Computing research issues (B.7.2) for an overview of our on-going and future research activities.

Over the past years, our research activities were focused on 3 projects, OVM, QUID, XtremWeb and MPICH-V that investigated models for high-performance execution environments on clusters and desktop grid systems. Our scientific goals were to design and develop innovating software for experimental platforms and to conduct experiments and perform optimizations for the platform.

## Theme 1: Cluster Computing

In the 1990s, the dominant parallel architectures were based on message passing, SMP, with moderate number of processors (8-16) and NUMA machines. At the end of the 1990s, parallel architectures evolved towards cluster of multiprocessors. Today, large shared-memory nodes (32-64-128 processors) are becoming available (IBM, Compaq, Fujitsu). Such systems may be used as parts of a cluster of multiprocessors. Two major issues have emerged or are emerging for programming high-performance applications:

- how to efficiently program large shared memory machines/nodes?
- how to efficiently program clusters of multiprocessors?

The QUID project studies programming models for complex parallel architectures. The constraint is to only rely on standard programming models, i.e. OpenMP and MPI (Message Passing Interface). We study the interactions between message passing and shared memory by considering a single program/multiple data (SPMD) execution style. We try to understand in what conditions the hybrid programming model (MPI+OpenMP) provides superior performances compared to the unified memory model. The QUID project has also addressed the issue of efficient programming of clusters of clusters. Géraud Krawesik has recently defended his Ph.D. thesis on this topic. These new results compare MPI to three OpenMP programming styles using a subset of the NASA Advanced Supercomputing (NAS) benchmark and two shared memory multiprocessors (IBM SP3 Night Hawk II, SGI Origin 3800). We presented the first SPMD OpenMP version of the NAS benchmark and compared it with other OpenMP versions from independent sources (PBN, SDSC and RWCP). Experimental results demonstrate that OpenMP provides competitive performance compared to MPI for a large set of experimental conditions.

Research groups

Archi Research description Cluster and grid team set adaptation and inter-thread communications. MPI still provides the best performance under some conditions.

The OVM project also considers parallel architectures, clusters of multiprocessors and clusters. Its goal is to design and develop an execution environment with dynamic control, based on a high performance client/server architecture. It targets a large range of applications: regular applications, irregular applications and applications with real-time constraints. The main topics of study encompass the optimization of the dataflow engine, the remote procedure call mechanism and global communication, which are the core of OVM. George Bosilca defended his Ph.D. thesis on OVM in October 2004. This work showed on a large set of experimental results that OVM could compare with MPI on regular applications while providing an efficient and convenient way to program irregular applications that require load-balancing facilities.

#### Theme 2: Large Scale Distributed and Parallel Computing.

Two projects represent the research of the Cluster and Grid research team around Grid Computing: XtremWeb and MPICH-V. These projects are now part of the Grand Large PCRI project.

XtremWeb is an open source middleware that generalizes traditional global computing platforms to a multi-user and multi-parallel programming context. XtremWeb relies on the notion of services to deploy a Desktop Grid based on a three-tier architecture. This architecture gathers three main services: Clients, Coordinators and Workers. Clients submit requests to the coordinator, which uses the worker resources to execute the corresponding tasks. Currently, tasks concern computation but we are also considering the integration of storage and communication capabilities. Coordinator sub-services provide resource discovery, service construction, service instantiation and data repository for parameters and results. A major concern is fault tolerance. XtremWeb relies on passive replication and message logging to tolerate the mobility of Clients, transient and definitive crashes of the Coordinator and the volatility or Workers. The Client service provides a Java API that unifies the interactions between the applications and the Coordinator. Three client applications are available: the Java API that can be used in any Java application, a command line (shell-like) interface and a Web interface allowing users to easily submit requests, consult the status of their tasks and retrieve the results. A second major issue is security. The origins of the threats are the applications, the infrastructure, the data (parameters and results) and the participating nodes. Currently XtremWeb provides user authentication, application sandboxing and communication encryption. We have developed deployment tools for harnessing home PCs, PCs in University or industry laboratories and PCs in clusters. XtremWeb provides a RPC interface for bag of tasks, parameter sweep, master worker and workflow applications. Associated with MPICH-V, XtremWeb allows the execution of unmodified MPI applications on Desktop Grids.

XtremWeb has been tested extensively by harnessing a thousand Workers and computing a million tasks. XtremWeb is deployed at several sites: Lille, University of Geneva, University of Tsukuba and University of Paris-Sud. At this latter site, XtremWeb is the Grid engine of the University Desktop Grid, which gathers about 500 PCs. Two multi-parametric applications are used in production since early 2004: the first one in physics for understanding the origin of cosmic rays (Aires - Air showers extended simulation - for the HEP Auger project); the second one in biology for a protein conformation predictor using a molecular dynamic simulator. MPICH-V is a research effort combining theoretical studies, experimental evaluations and pragmatic implementations in order to provide a MPI implementation based on MPICH featuring multiple fault tolerant protocols.

There is a long history of research in fault tolerance for distributed systems. We can distinguish the automatic/transparent approach from the manual/user-controlled approach. The first approach relies either on coordinated checkpointing (global snapshot) or uncoordinated checkpointing associated with message logging. A well-known algorithm for the first approach has been proposed by Chandy and Lamport. This algorithm requires restarting all processes even if only one process crashes, so it is believed not to scale well. Several strategies have been proposed for message logging: optimistic, pessimistic, causal. Several optimizations have been studied for the three strategies. The general context of our study is high-performance computing on large platforms. One of the most widely used programming environments for such platforms is MPI.

Within the MPICH-V project, we have developed and published 3 original fault-tolerant protocols for MPI: MPICH-V1, MPICH-V2 and MPICH-V/CL. The first two protocols rely on uncoordinated checkpointing associated with either remote pessimistic message logging or sender-based pessimistic message logging. We have demonstrated that MPICH-V2 outperforms MPICH-V1. MPICH-V/CL implements a coordinated checkpoint strategy (Chandy-Lamport) removing the need for message logging. MPICH-V2 and V/CL are concurrent protocols for large clusters. We have compared them by defining a new parameter for evaluating the merits of fault-tolerant protocols: the impact of fault frequency on performance. We have demonstrated that the stress of the checkpoint server is the fundamental source of performance differences between the two techniques. Under the considered experimental conditions, message logging becomes more efficient than coordinated checkpoint when the fault frequency reaches 1 fault every 4 hours, for a cluster of 100 nodes sharing a single checkpoint server, with a data set of 1 GB on each node and a 100 Mb/s network.

The next step in our research is to investigate a protocol dedicated to Hierarchical Desktop Grid (HDG) that would also apply to other Grids. In such context, several MPI executions take place on different clusters, possibly using heterogeneous networks. An automatic fault-tolerant MPI for HDG or Grids should tolerate faults inside clusters and the crash or disconnection of a full cluster. We are currently considering a hierarchical fault-tolerant protocol combined with a specific run-time module allowing the migration of full MPI executions on clusters independently of their high performance network hardware.

The performance and volatility tolerance of MPICH-V make it attractive for large clusters, clusters made from collection of nodes in a LAN environment (Desktop Grid), Grid deployments harnessing several clusters and campus/industry-wide desktop Grids with volatile nodes, i.e., when all infrastructures feature synchronous networks or controllable area networks. Research groups

Archi Research description Cluster and grid team

## B.6.3 / Archi

## Highlights

Processor research team:

- 2002: SPECIF award for Alexandre Farcy.
- Steering committee member of the HiPEAC european network of excellence.
- Several contracts with Hewlett-Packard.
- Broad diffusion of MicroLib simulators through our Web site.
- Publications in top conferences: ISCA, MICRO, Sigmetrics.
- Creation of the Alchemy PCRI project.

Cluster and Grid research team:

- Steering committee of GridExplorer and Grid'5000 national projects.
- Publications in top conferences: SC, Cluster, SPAA, HPCA.
- Creation of the Grand Large PCRI project.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- $\bullet$  Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international), and,
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

## Honors

## Prizes and awards

Processor research team:

- David Parello, nominated with 4 other articles for the best student paper award at *Supercomputing 2004* (SC'04) [53].
- Alexandre Farcy, best Ph.D. thesis award, SPECIF, 2002.
- Gregory Watts, selected among the 10 best papers at *Compilers for Parallel Computers* (CPC 2001) [83].

Cluster and Grid research team:

• George Bosilca, Best Paper Award, *IEEE International Symposium on Cluster Computing and the Grid* (CCGRID'2001) [21].

### **Keynote addresses**

Cluster and Grid research team:

- Franck Cappello, "OVM: High Performance Computing with RPC programming style", *Score Users Group Meeting*, September 2000, Oxford, UK.
- Franck Cappello, "Understanding Performance of SMP Clusters for the NAS Benchmark", *Workshop on Grid and Cluster Computing*, March 2000, Tsukuba, Japan.
- Franck Cappello, "Comparing Performance of MPI and MPI+OpenMP for NAS benchmark on IBM SP3", *IBM Watson ACTC European Workshop*, May 2000, Paris, France.
- Franck Cappello, "Desktop Grid With XtremWeb: User experiences and feedback", *SC2002 SuperComputing Conference panel on Desktop Grids*: 10,000 fold parallelism for the masses.
- Franck Cappello, "XtremWeb: toward High Performance Computing on P2P systems", *High Performance Computing Symposium, HPC2002,* Cetraro, Italy, July, 2002.
- Franck Cappello, "Ontology of Desktop Grids: a Pragmatic View from the XtremWeb Experience", *DGRID workshop, associated with GRID 2003 and SC 2003*, Phoenix, USA, November, 2003.
- Franck Cappello, "High Performance Computing on P2P Platforms: Recent Innovations", *Terena Conference*, Zagreb, Croatia, Mai, 2003.
- Franck Cappello, "The MPICH-V Project", ENS/NSF Workshop, Lyon, September 2004.
- Franck Cappello, "Hybrid Preemptive Scheduling of MPI Applications on the Grids", *Scheduling Workshop*, Modane, August 2004.

## Other honors

Cluster and Grid research team:

• Franck Cappello, member of *IEEE/ACM Cluster Computing and the Grid (CCGRID) Steering Committee*.

Research groups

Archi Highlights Honors

## **Evaluation of research**

## **Editorial boards**

Processor research team:

• IEEE Transactions on Computers: Daniel Etiemble, associate editor (1999-2003).

Cluster and Grid research team:

- International Journal of Grid Computing, Kluwer Academic Publishers: Franck Cappello, member of the editorial board.
- IJGUC, International Journal of Grid and Utility Computing, Inderscience: Franck Cappello, associate editor (2002).
- TSI, Technique et Science Informatique, Hermès: Franck Cappello, member of the editorial board.
- Edited book, "Parallel I/O for Cluster Computing", Hermès, 2002: Franck Cappello, member of the editorial board.

#### **Program committees**

Processor research team:

- CGO, ACM/IEEE International Symposium on Code Generation and Optimization: Olivier Temam, 2005.
- ACM/IEEE International Conference on Parallel Architectures and Compilation Techniques: Olivier Temam, 2004 & 2005.
- HiPC, ACM/IEEE International Conference on High Performance Computing: Olivier Temam, 2004.
- ASPLOS, ACM Conference on Architectural Support for Programming Languages and Operating Systems: Olivier Temam, 2004.
- ISPASS, IEEE International Symposium on Performance Analysis of Systems and Software: Olivier Temam, 2004.
- ISHPC, Fourth International Symposium on High Performance Computing: Olivier Temam, 2002.
- MEDEA Workshop, ACM/IEEE International Conference on Parallel Architectures and Compilation Techniques (PACT): Olivier Temam, 2002.
- SympA, Symposium Francophone en Architecture de Machines: Olivier Temam, 2000-2003.

Cluster and Grid research team:

- HCW, IEEE Heterogeneous Computing Workshop:Franck Cappello, 2005.
- EGC, European Grid Conference:Franck Cappello, 2005.
- PaCT, Parallel Computing Technologies: Franck Cappello, 2005.
- HPDC 13, IEEE International Symposium on High Performance Distributed Computing: Franck Cappello, 2004.
- SC 04, ACM/IEEE SuperComputing 2004, Conference on High Performance Computing, Networking and Storage: Franck Cappello, 2004.
- HiPC, ACM/IEEE International Conference on High Performance Computing: Franck Cappello, 2002 & 2004.
- DiDaMIC, workshop on Distributed Databases and processing in Medical Image Computing: Cécile Germain, 2004.
- ISGC, International Symposium on Grid Computing: Franck Cappello, 2003.
- EuroPar, International Conference on Parallel and Distributed Computing: Franck Cappello, 2003.
- IMAGE, workshop on Imaging, Medical Analysis and Grid Environments: Cécile Germain, 2003.
- CSCW, Conference on Computer Supported Cooperative Work: Franck Cappello, 2002.

- CCGrid, IEEE/ACM International Symposium on Cluster Computing and the Grid: Franck Cappello, 2001.
- Renpar 14, 15 & 16, Rencontres du Parallélisme: Franck Cappello, 2002, 2003, 2005.

#### **Evaluation committees and invited expertise**

Processor research team:

- Member of the SPECIF award jury: Olivier Temam, 2000 & 2001.
- Member of INRIA post-doc scholarships jury: Olivier Temam, 2000-2002.

Cluster and Grid research team:

• Evaluation Board member of Grid'5000 proposals: Franck Cappello, 2003.

## Other evaluation activities

• Reviewer for PhD dissertation: Olivier Temam (3), Franck Cappello (6).

#### B.6.6 / Archi

## Volunteer professional service

#### Management positions in scientific organisations

Processor research team:

- IEEE Computer Society representative for IFIP Technical Committee 10: Daniel Etiemble since 1999.
- IFIP Technical Committee 10: Daniel Etiemble, vice-chair since 2002.
- CNRS network on architecture and compilers, *Réseau Thématique Pluridisciplinaire Architecture-Compilation*: Nathalie Drach & Olivier Temam, members of the steering committee since 2003.
- PCRI project ALCHEMY (INRIA/CNRS/Université Paris-Sud): Olivier Temam, project leader since its creation in 2003.
- HiPEAC (HIgh-PErformance Architectures and Compilers) European Network of Excellence: Nathalie Drach & Olivier Temam, members of the steering committee since 2002.
- CNRS *Action Spécifique*, working group on New technologies and new architecture paradigms, 2002-2003: Nathalie Drach, chair.

#### Cluster and Grid research team:

- CNRS network on GRID computing, *Réseau Thématique Pluridisciplinaire*: Franck Cappello, member of the steering committee since 2003.
- Grid'5000 national project: Franck Cappello, chair of the steering committee.
- CNRS "Action Spécifique", working group on enabling Grid'5000: Franck Cappello, Chair
- CNRS-Num@tec working group "Large Scale Architectures, Simulation, modelisation", 2003-2004: Cécile Germain, chair.

Research groups

Archi Evaluation of research Volunteer professional service

## Organisation of conferences and scientific events

*Cluster and Grid research team:* 

- Workshop GP2PC (Global and Peer-to-Peer Computing), in conjunction with CCGRID: Franck Cappello, co-organizer with Spyros Lalis, 2000-2005.
- GRID2002 Winter School: Franck Cappello, co-organizer with Emmanuel Jeannot and Fréderic Desprez.
- GRID tutorial, Renpar 2002 Summer School: Franck Cappello, co-organizer with Jean-Louis Pazat.
- Symposium SYMPA7, Paris, 2001: Franck Cappello, organizer.
- Winter School iHperf2000, *Programmation parallèle et hybride*, Aussois, Décembre 2000: Franck Cappello, organizer.
- IEEE HPCA6 conference, Toulouse, France, 2000: Franck Cappello, organizer.

## Working groups

Processor research team:

- CNRS Action Spécifique, SocLib working group (simulation of Systems-On-Chip), 2002: Olivier Temam, member.
- CNRS Action Spécifique, working group Architecture-OS Adequation, 2003-2004: Nathalie Drach, member.

Cluster and Grid research team:

- CNRS Action Spécifique, Programmation Grid'5000 working group, 2003: Franck Cappello, member.
- "SuperComputers" mission organized by the french ministry of foreign affairs, 2001: Franck Cappello, Experts Board member.

## **Contracts and grants**

	C :		<b>F</b> <i>V</i>			D ('	
Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Duration (months)	Total €
RNTL	O. Temam	СОР	UE	UPS/SAIC	11/12/03- 10/12/05	13	148 796
RNTL	N. Drach	ATLAS		UPS	01/06/02- 01/06/04	24	106 993
RNTL	F. Cappello	CASPer	MENRT	UPS/SAIC	01/10/02 - 01/10/04	24	119 317
CIFRE	O. Temam	COMPAQ COMPUTER		Cephyten	15/02/01- 14/01/04	36	22 867
Industry	F. Cappello	EADS Etude des perfor- mances machines parallèles		Cephyten	20/07/01- 19/07/04	36	18 293
ACI	F. Cappello	Datagraal	MENRT	CNRS	23/10/02- 22/10/04	24	10 744
ACI	F. Cappello	CGP2P (grid)	CNRS fns	CNRS	01/02/02- 30/01/05	36	109 621
ACI	F. Cappello	GRID2 (grid)	CNRS fns	CNRS	01/02/02- 30/01/05	36	21 414
ACI	F. Cappello	data grid explorer (MDD)	MENRT	CNRS	10/10/03- 09/10/06	36	689 023*
ACI	F. Cappello	Grid'5000 (grid)	MENRT	CNRS	10/10/04- 09/10/07	36	155 000
ACI	C. Germain	AGIR (MDD)	CNRS fns	CNRS	21/07/04- 20/07/07	36	29 933
AS	F. Cappello	Coop urbana	CNRS dri	CNRS	06/05/03- 05/05/04	12	7 000
AS	F. Cappello	RTP 8 Etude préparatoire	CNRS Stic	CNRS	16/06/03- 15/06/04	12	3 750
AS	N.Drach	Nouveaux paradigmes d'architecture	CNRS Stic	CNRS	19/07/02- 18/07/03	12	15 240
SUBV	O. Temam	Action Blanche	MENRT	UPS/SUBSF			58 040
SUBV	C. Germain	iHPerf98	GDR ARP	UPS	1/01/00- 31/12/00	12	1 829
SUBV	C.Germain	Accueil Jeunes chercheurs séjour postdoctoral	MENRT	UPS/SUBSF	01/10/01- 31/03/03	36	35 930
	O. Temam	MAHOTEU (Solde ESPRIT)	UVQS	Cephyten	1/10/00- 31/12/00	3	28 797

Note : See the glossary for acronyms

\* Most of the budget (630435  $\in$ ) is devoted to equipment shared by the partners.

## Summary of main projects:

#### **Processor architecture team:**

СОР

Centre d'Optimisation de Programmes

Partners: LRI, Hewlett-Packard, STMicroelectronics, INRIA, CEA-LETI, Université de Toulouse.

The goal of the project is to set up a center for program tuning. The center will target general-purpose and embedded processors. Techniques for rapidly optimizing programs are being developed, based on automatic or manual iterative optimization techniques [53].

Project funded by national network on software technology (*RNTL*), 2004-2006. Scientific director for LRI: Olivier Temam. Research groups Archi Volunteer professional

service Contracts and grants

#### ATLAS

*Analyse par Tests de Logiciels embarqués par Application de la Simulation Généralisée* 

Partners: LRI, CEA-LETI, EADS, TNI, Université de Toulouse.

The goal is to develop software test techniques for embedded systems using embedded processors simulation. As part of this project, a library for simulator components will be set up. Several processor embedded models will be added to the library [42].

#### **COMPAQ / HP France**

This contract includes two projects. The first one is the development of a manual methodology for rapidly optimizing scientific applications. This project was part of the CIFRE Ph.D. thesis of David Parello (see Ph.D. students and [54]).

The second project is related to the software and hardware development of new computing systems in-between traditional computer systems and consumer electronics products. Essentially, we want to understand how computer systems, especially home computer systems, will evolve.

#### ACI Blanche

Partners: LRI, Université de Versailles, Université de Toulouse.

This project was used as seed funding for two of the current main group projects: the development of a modular simulation environment and methods for optimizing programs on complex processor architectures [83].

#### ΜΑΗΟΤΕυ

Memory Hierarchy Analysis and Optimization Tools for the End-User.

*Partners: Université de Versailles, Universitat Politecnica de Barcelona, University of Edinburgh, ONERA, Edinburgh Portable Compilers (now Analog Devices), Digital France (now HP France).* 

This project targeted the development of tools and techniques for analyzing and optimizing programs on processor memory hierarchy. These tools were designed for end-users who are currently left unaided in the task of optimizing programs on the memory hierarchy, since compilers mostly fail to properly address this issue [32].

Project funded by national research program (ACI Blanche), 2000-2002.

Scientific director: Olivier Temam.

Extension of European ESPRIT Long Term Research contract, 2000. Scientific director for LRI: Olivier Temam.

Collaboration with Hewlett-Packard, 2000-2004, Scientific director: Olivier Temam.

Project funded by national network on software technology (RNTL), 2002-2004.

Scientific director for LRI: Nathalie Drach.

#### Cluster and Grid research team.

#### CASPer

Community-based Application Service Provider

*Partners: LRI, IRISA Rennes, EADS (coordinator), ALCATEL, IDEAMECH, CADOE, FFT.* 

Development of an Application Service Provider for numerical simulation applications. CASPer is a joint work between academic and industrial partners. It aims at providing a framework for enterprises to access collections of numerical simulations, to use computing resources and to share information and data among the participants.

#### CGP2P

Calcul Global Pair à Pair

Partners: LRI (coordinator), LIFL Lille, ID-IMAG Grenoble, LIP Lyon, LARIA Amiens

This project, chaired by Franck Cappello, investigates the convergence between Global Computing and P2P systems with the objective of developping software for high performance computing on P2P systems.

#### **Data Grid Explorer**

Coordinating partners: LRI, ID-Imag Grenoble, LIP6 Paris Other parnters: CEA, INRIA Sophia Antipolis, IBCP Lyon, IRISA Rennes, LAAS Toulouse, LABRI Bordeaux, LARIA Amiens, LIFL Lille, LIP Lyon, LORIA Nancy Project funded by national research program on massive data (*ACI Masses de données*), 2003-2006. Scientific director: Franck Cappello.

The GRID eXplorer project is headed by Franck Cappello and

aims at building a large scale emulator for Grids and P2P systems based on a large cluster equiped with network emulators.

#### eGrid5000

The goal of this working group is to prepare the deployment and useage of the Grid'5000 experimental platform. The group brings together most of the members of the French Grid community. The Grid'5000 project, partially funded by the *ACI* 

*GRID* national program, seeks to establish an experimental Grid platform for the community of Grid reserachers. Franck Cappello chairs the streering committee of Grid5000.

By the end of 2005, this platform will gather seven sites distributed accross France, each equiped with a 256 CPUs cluster. All the clusters will be connected via the RENATER national network infrastructure with 1 Gigabits/s dedicated links. This platform will be used for experiments in the FRAGILE project.

Project funded by national network on software technology (*RNTL*), 2002-2004. Scientific director for LRI: Franck Cappello.

Project funded by national research program on grid comput-

2005

ing (ACI GRID), 2002-

Scientific director:

Franck Cappello.



2003-2004. Scientific director: Franck Cappello. Research groups

Contracts and grants

Archi

#### AGIR

Analyse Globalisée des Images Radiologiques

Partners: CNRS-STIC, CNRS-IN2P3, INRIA, INSERM and physicians.

This multi-disciplinary project aims to define, validate and deploy new grid services for individual and collective analysis of medical images.

#### B.6.8 / Archi

## Collaborations

## **Cooperation agreements**

Cluster and Grid research team:

- 2004-2005: Collaboration with the ICMMG lab. in Novosibirsk funded by a CNRS PAI Franco-Russe follow-up of *Action Jeunes Chercheurs* (C.Germain).
- 2003-2004: Collaboration with University of Illinois at Urbana-Champaign (Marc Snir). This project studies langages for parallel programation and their application to real-world parallel programs. More specifically this study investigates two tools: GA (Global Arrays) and UPC (Unified Parallel C). This grant has already permitted several working visits (F. Cappello and G. Krazewick to UIUC and M. Snir to Paris).
- 2001-2004: CIFRE Grant in collaboration with EADS to support a Ph.D. student. Investigates the performances of parallel machines and hybrid programming models (shared memory and message passing).

## Collaborations leading to joint publications

Processor research team:

- since 2003: Collaboration with HPLabs in Palo Alto on memory systems architecture (Olivier Temam).
- since 2000: Collaboration with Computer Science department at University of Edinburgh on performance analysis of memory systems (Olivier Temam).
- since 1999: Collaboration with CEA LETI (Grenoble) on processor simulation and optimization; possible extension within the CNRS/CEA Num@tec project (Olivier Temam, Nathalie Drach).
- 1996-2000: Several collaborations with Universitat Politecnica de Catalunya (UPC) on processor architecture research (Olivier Temam).
- 1996-2000: Collaboration with ONERA (Toulouse) on optimization of numerical programs for memory systems (Olivier Temam).
- since 1996: Collaboration with HP France on program optimization (Olivier Temam).

Project funded by national research program on massive data (*ACI Masses de données*), 2004-2007. Scientific director for LRI: Cécile Germain.

## Informal collaborations with industrial partners

Processor research team:

• since 2003: Collaboration with STMicroelectronics on iterative compilation; this is the Advanced System research group(AST) located in Switzerland, distinct from the ST group in Grenoble participating in the RNTL COP project (Olivier Temam).

## Other collaborations

Processor research team:

• 2004: Recent collaboration with the Liberty Group at Princeton University on processor simulation (Olivier Temam).

Cluster and Grid research team:

- 2004: Franck Cappello is a member of the FP6 CoreGrid Network of Excellence.
- 2004: Cécile Germain is a member of the European EGEE project (Enabling Grids for E-science in Europe), Activity NA4 "Application Identification and support".

#### B.6.9 / Archi

## Dissemination and technology transfer

## Software licensing and distribution

Processor research team:

#### MicroLib

Microlib is a processor simulation software based on SystemC, distributed under GPL at http://www.microlib.org and developed by Gilles Mouchard, Daniel Gracia-Perez, Olivier Temam.

Several components have been downloaded several hundred times, and the PowerPC simulator has been downloaded more than 2000 times. The software features:

- a retargetable processor emulator
- a library of processor simulator components (published at Micro 2004)
- a generic superscalar processor simulator
- a PowerPC~750 simulator

• DiST, a distributed simulation environment for speeding up processor simulation (published at Sigmetrics 2003)

• a new fast SystemC engine (published at DATE 2004)

#### Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

#### Archi

Collaborations Dissemination and technology transfer

#### Cluster and Grid research team:

Our group develops and distributes several free softwares for cluster and grid computing under the GPL Licence. These softwares are delivered to end users in both academic and industry and to researchers in computer science. Our users include EADS, Alcatel, IFP, Université Paris-Sud, University of Tsukuba, University of San Diego.

- Xtrem Web: Desktop Grid Middleware.
- **MPICH-V**: An automatic fault tolerant MPI implementation for volatile nodes.
- **OVM**: Client broker server system for parallel applications on clusters.
- QUID: Programming and performance evaluation of Clusters of Multiprocessors.

#### Summer schools, tutorials, invited seminars

#### Processor research team:

Olivier Temam was invited to give seminars at the following places :

- From Sequences of Dependent Instructions to Functions: An Approach for Improving Performance without ILP or Speculation, TU Delft, The Netherlands, 2004.
- *MicroLib: A case for the quantitative comparison of micro-architecture mechanisms*, Ghent University, Belgium, 2004.
- Agent programming and self-organizing architectures, IBM Thomas Watson reseach center, USA, 2004.
- Spatial-oriented architectures, University of Texas, USA, 2004.
- Iterative optimization environment, STMicroelectronics, Lugano, Switzerland, 2004.
- Bridging the Gap between Programs and Architectures, HPLabs, Palo Alto, California, 2003.
- Program optimization for complex processors, CEA LIST, 2002.
- Mhaoteu: Tools for memory hierarchy management, IMACS Conference, 2000.

#### Cluster and Grid research team:

- *Calcul dans les systèmes distribués à grande échelle,* GRID2002 winter school, Aussois, December 2002: Franck Cappello.
- Calcul sur Grilles et Calcul Pair à Pair, Ecole thématique, Avril 2002, Hammamet, Tunisie, in conjunction with RenPar2002: Franck Cappello.
- *Systèmes distribués de calcul global et pair à pair,* Tutorial at RenPar 2001, April 2001, Paris: Franck Cappello, Gille Fedak, Olivier Richard.
- *Programmation parallèle et hybride*, iHperf2000 Winter school, Aussois, December 2000: Franck Cappello, Fabien Coelho, Frédéric Desprez.

Franck Cappello was invited to give the following seminars outside France:

- Destkop Grid: fault tolerance and programming models, Center for Computational Physics, Université de Tsukuba, May 2003, Japan.
- Various aspects of cluster computing, Center for Computational Physics, Tsukuba, October 2001, Japan.
- Cluster & GRID research at LRI, AIST, Tsukuba, September 2001, Japan.
- *XtremWeb : a Peer-to-Peer Global Computing experimental platform,* Matsuoka Research Lab, TiTech, Tokyo, September 2001, Japan.
- *XtremWeb : a Peer-to-Peer Global Computing experimental platform,* Fujistsu Research Lab, Kawazaki, September 2001, Japan
- *Peer-to-Peer Global Computing*, with Mark Baker, Cécile Germain et Gille Fedak, Tutorial at Peer-to-Peer Europe, Décember 2001, Paris (accepted but the conference was cancelled).
- XtremWeb : une plate-forme expérimentale pour le calcul Global et Pair à Pair, March 2001, Genève, Suisse.

Franck Cappello was invited to give the following seminars in France:

- *Systèmes distribué à grande échelle : Associer l'observation, la théorie et la réalisation, s*éminaire Graphes, Reseaux et Modelisation, Jussieu, October 2003.
- *The MPICH-V project: towards a fault tolerant MPI for clusters and Grids,* journée Calcul Scientifique et le GRID, Versailles, September 2003.
- Grid eXplorer, réunion du comité de direction du RTP CNRS réseau, Marseille, September 2003.
- Des architrectures haute performance et des Grid en particulier, journée calcul Sanofi-Synthélabo, March 2003, Montpellier.
- L'ACI CGP2P, journée GRID au CINES, March 2003, Montpellier.
- Calcul distribué à grande échelle pour la biophysique moléculaire à partir de la plate-forme logicielle XtremWeb, Grid pour la génomique, Institut de Biologie et Chimie des Proteines, IBCP CNRS, January 2003, Lyon.
- Calcul à grande échelle : XtremWeb et MPICH-V, CEA, January 2003, Bordeaux.
- ACI GRID CGP2P : problématique de la fusion P2P et GRID pour le calcul à grande échelle, GRID@Inria, Juy 2002, Sophia Antipolis.
- *Etudes autour d'une plate-forme pour le calcul distribué à grande échelle,* groupe ORDO du GDR ARP, LIP6, June 2002, Paris.
- *Calcul Global Pair à Pair : classification, problèmes techniques et scientifiques,* projet Inria SOR, May 2002, Rocquencourt.
- Calcul dans les système Peer-to-Peer, Séminaire X-Aristote, école polytechnique, April 2002, Palaiseau.
- *Calcul numérique Pair à Pair à grande échelle : classifications, exemples et problèmes scientifiques,* projet INRIA OASIS, March 2002, Sophia Antipolis.
- *Techniques et problèmes scientifiques du calcul numérique à grande échelle par l'approche Pair à Pair,* laboratoire ID, February 2002, Grenoble.
- Calcul numérique à très grande échelle, Journée IrisaTech, Rennes, February 2002.
- Vers une nouvelle génération de systèmes P2P (pair à pair), Club Sciences et Citoyens, February 2002, Orsay.
- *P2P et XtremWeb*, Journées Tarot, November 2001, Paris.
- Cluster de PCs pour les applications hautes performances, Rencontre sur les Cluster, April 2001, Orsay.
- *XtremWeb : une plate-forme expérimentale pour le calcul Global et Pair à Pair,* LRI, séminaire de l'équipe Parallélisme, February 2001, Orsay.
- Programmation hybride MPI+OpenMP sur l'IBM SP3, séminaire du LAMIA, Valenciennes, Décember 2000.
- *XtremWeb : une plate-forme expérimentale pour le calcul Global et Pair à Pair,* sémimaire LARIA, November 2000, Amiens.
- *XtremWeb : une approche académique et pluridisciplinaire d'expérimentation du Global Computing*, Journées Tarot, November 2000, Paris.
- XtremWeb : une plate-forme expérimentale pour le Calcul Global, Réunion Data Grid, July 2000, Lyon.

#### B.6.10 / Archi

## Training and education (doctoral and post-doctoral)

Defended habilitations					
Name		Date defende	d Current position		
Processor research team					
DRACH-TEMAM	Nathalie	8/11/02	Professor at Université Paris VI		
Cluster and Grid research group					
CAPPELLO	Franck	23/10/01	DR INRIA Futurs (Grand Large PCRI Project)		

Research groups

#### Archi

Dissemination and technology transfer Training and education

Defended doctorates					
17/01/00					
4/05/00					
12/10/01	start-up company on video graphics processing				
19/09/02	researcher at Intel Portland (Pentium 4 processors)				
27/09/02	engineer at the IRISA/Texas Instruments joint lab				
8/11/02	researcher at Intel Israël (Centrino processors)				
Cluster and Grid research group					
11//06/03	CR INRIA Futurs (Grand Large PCRI project)				
30//09/04	Assistant Professor (University of Tennesser at Knoxville)				
	17/01/00 4/05/00 12/10/01 19/09/02 27/09/02 8/11/02 oup 11//06/03 30//09/04				

Eleven new Ph.D. students have joined the Processor team since 2000: Gilles Mouchard, David Parello and Sami Yehia in 2000, Sylvain Girbal and Michaël Dupré in 2001, Daniel Gracia Perez and Yves Lhuillier in 2002, Alexandre Coveliers and Sébastien Favre in 2003, Nicolas Vasilache and Pierre Palatin in 2004.

Five new Ph.D. students have joined the Cluster and Grid team since 2000: Abderrahmane Djilali, Geraud Krawezik and Oleg Lodigensky in 2001, Romain Texier, Aurelien Bouteiller in 2002, Benjamin Quetier in 2004.

## Graduate courses

Processor research team:

• DEA ID, now Master Informatique Paris-Sud, *Futurs Processeurs*: Nathatlie Drach, Frédéric Gruau, Olivier Temam.

Cluster and Grid reasearch team:

• Master Informatique Paris-Sud, *Calcul Haute Performance et Metacomputing*: Franck Cappello, Cécile Germain, Jean-François Méhaut (ENS Lyon), Raymond Namyst.

## **Publications**

### International peer-reviewed journals

- [1] G. Bosilca, G. Fedak, and F. Cappello. Ovm: Out-of-order execution parallel virtual machine. *FGCS Future Generation Computer Science*, 2002.
- [2] A. Bouteiller, P. Lemarinier, G. Krawezik, and F. Cappello. Coordinated checkpoint versus message log for fault tolerant MPI. *Int. Journal of High Performance Computing and Networking (IJHPCN)*, (3), September 2004.
- [3] T. Brandes and C. Germain. A Schedule Cache for Data-Parallel Unstructured Computations. *Parallel Computing*, 26(13):1807-1823, October 2000.
- [4] F. Cappello, S. Djilali, G. Fedak, T. Herault, F. Magniette, V. Néri, and O. Lodygensky. Computing on large scale distributed systems: Xtremweb architecture, programming models, security, tests and convergence with grid. *FGCS Future Generation Computer Science*, 2004.
- [5] F. Cappello, O. Richard, and D. Etiemble. Understanding performance of smp clusters running mpi programs. *Futur Generation Computer Systems*, 2001.
- [6] S. Djilali, T. Herault, O. Lodygensky, T. Morlier, and F. Cappello. Rpc-v: Toward fault-tolerant rpc for grids with volatile nodes. In *IEEE/ACM SC 2004*, Pittsburg, USA, 2004.
- [7] N. Drach, J.-L. Béchennec, and O. Temam. Increasing hardware data prefetching performance using the second-level cache. *Journal of Systems Architecture*, 48(4-5):137-149, Nov 2002, Elsevier Science.
- [8] M. Dupré and N. Drach. Algorithme d'ordonnancement dynamique pour exécution statique. *Revue Technique et Science Informatiques*, 22:713-735, Nov 2003, Lavoisier et Hermes Science.
- [9] G. Fursin, M. O'Boyle, O. Temam, and G. Watts. A fast and accurate method for evaluating the upper-bound of memory performance. *Concurrency : Practice and Experience*, 16(2-3):271-292, Jan 2004, John Wiley & Sons.
- [10] F. Gruau and J. Tromp. Cellular gravity. *Parallel Processing Letters*, 10(4), Dec 2000, World Scientific.
- [11] G. Krawezik and F. Cappello. Performance comparison of mpi and openmp on shared memory multiprocessors. *Journal of Concurrency and Computation: Practice and Experience*, page 46, 2003.
- [12] J. Sébot and N. Drach. Simd isa extensions: Power efficiency on multimedia. *IEICE Transactions on Electronics*, E85-C(2):297-303, Feb 2002, Oxford University Press.
- [13] A. Vartanian, C. Limousin, J. Sébot, and N. Drach. Architecture optimization for multimedia application exploiting data and thread-level parallel, journal of systems architecture. *Journal of Systems Architecture*, A paraître, Dec 2004, Elsevier Science.

## National peer-reviewed journals

[14] F. Cappello and D. Etiemble. Mpi ou mpi+openmp pour les grappes de multiprocesseurs. *TSI (Technique et Science Informatiques)*, 2001.

Research groups

Archi Training and education Publications

- [15] F. Cappello, D. Litaize, J.-F. Mehaut, C. Morin, S. Petiton, and D. Trystram. Metacomputing : vers une nouvelle dimension pour le calcul haute performance. *TSI* (*Technique et Science Informatiques*), Juin 2000.
- [16] A. Osorio, S. Merran, C. Germain, J. Atif, X. Ripoche, A. Tarault, and R. Texier. Segmentation 3D d'images radiologiques : application à la reconstruction et mesure du volume d'organes et de lésions. *Journal de Radiologie*, 84(10), 2003. Journées Françaises de Radiologie.

#### Other journals

[17] F. Cappello. Calcul global pair a pair : extension des systemes pair a pair au calcul. *Lettre de l'IDRIS*, 2002.

#### **Books**

- [18] F. Cappello, A. Djilali, G. Fedak, C. Germain, O. Lodygensky, and V. Néri. *Calcul réparti à grande échelle Metacomputing*, chapter XtremWeb, une plate-forme de recherche sur le Calcul Global et Pair à Pair. Hermes Science, Lavoisier, 2002.
- [19] K. McKinley and O. Temam. *How to quantify loop nest locality*. MIT Press, Rudi Egeinmann, Jan 2000.

#### Major international peer-reviewed conferences

- [20] G. Bosilca, A. Bouteillier, F. Cappello, S. Djilali, G. Fedak, C. Germain, T. Herault, P. Lemarinier, O. Lodygensky, F. Magniette, V. Neri, and A. Selhikov. Mpich-v: Toward a scalable fault tolerant mpi for volatile nodes. In *Proceedings of ACM/IEEE International Conference on Supercomputing SCO2*, Baltimore, USA, November 2002. IEEE/ACM, IEEE Press.
- [21] G. Bosilca, G. Fedak, and F. Cappello. Ovm: Out-of-order execution parallel virtual machine. In *Proceedings of CCGRID'2001*, Brisbane, Australia, May 2001. IEEE/ACM, IEEE Press.
- [22] G. Bosilca, G. Fedak, O. Richard, and F. Cappello. High performance computing with rpc programming style. In *Proceedings of the First Myrinet User Group Conference MUG2000*, September 2000.
- [23] A. Bouteiller, H.-L. Bouziane, P. Lemarinier, T. Herault, and F. Cappello. Hybrid preemptive scheduling of mpi applications on the grids. In *5th IEEE/ACM International Workshop on Grid Computing (Grid 2004)*, Pittsburgh USA, November 2004. IEEE CS Press/ACM.
- [24] A. Bouteiller, F. Cappello, T. Hérault, G. Krawezik, P. Lemarinier, and F. Magniette. MPICH-V2: a fault tolerant MPI for volatile nodes based on pessimistic sender based message logging. In *High Performance Networking and Computing (SC 2003)*, Phoenix USA, November 2003. IEEE/ACM.
- [25] A. Bouteiller, P. Lemarinier, and F. Cappello. MPICH-V3 preview: A hierarchical fault tolerant MPI for multi-cluster grids. In *IEEE/ACM High Performance Networking and Computing (SC 2003), poster session,* Phoenix USA, November 2003.

- [26] A. Bouteiller, P. Lemarinier, G. Krawezik, and F. Cappello. Coordinated checkpoint versus message log for fault tolerant MPI. In *IEEE International Conference on Cluster Computing (Cluster 2003)*. IEEE CS Press, December 2003.
- [27] F. Cappello and D. Etiemble. Mpi versus mpi+openmp on ibm sp for the nas benchmarks. In *Supercomputing 2000*, Dallas, November 2000.
- [28] F. Cappello, P. Fraigniaud, B. Mans, and A. Rosenberg. Hihcohp: Toward a realistic communication model for hierarchical hyperclusters of heterogeneous processors. In *IEEE/ACM IPDPS'2001*, 2001.
- [29] F. Cappello, O. Richard, and D. Etiemble. Investigating the performance of two programming models for clusters of smp pcs. In *IEEE HPCA6*, 2000.
- [30] A. Cohen, S. Girbal, and O. Temam. A polyhedral approach to ease the composition of program transformations. In *Euro-Par*. ACM/IFIP/IEEE, Aug 2004.
- [31] M. Dupré, N. Drach, and O. Temam. Quickly building an optimizer for complex embedded architectures. In *International Symposium on Code Generation and Optimization*. ACM/IEEE, Mar 2004.
- [32] C. Eisenbeis, A. Gonzalez, J. Llosa, M. O'Boyle, O. Temam, and G. Watts. Mhaoteu: Tools for memory hierarchy management. In *Conference on Applications of Computer Algebra*. IMACS, Aug 2000.
- [33] D. Etiemble and L. Lacassagne. 16-bit fp sub-word parallelism to facilitate compiler vectorization and improve performance of image and media processing. In *ICPP2004*.
- [34] D. Etiemble and L. Lacassagne. Simd 16-bit fp instructions for image and media processing on general purpose microprocessor. In *ODES-2004*.
- [35] G. Fedak, C. Germain, V. Néri, and F. Cappello. Xtremweb: A generic global computing platform. In *Proceedings of CCGRID'2001 Special Session Global Computing on Personal Devices*, Brisbane, Australia, May 2001. IEEE/ACM, IEEE Press.
- [36] C. Germain, G. Fedak, V. Neri, and F. Cappello. Global computing systems. In S. LNCS, editor, *3rd Int. Conf. on Scale Scientific Computations SciCom01*, volume 2179, pages 218-227, 2001.
- [37] C. Germain, V. Néri, G. Fedak, and F. Cappello. Xtremweb: building an experimental platform for global computing. In *Grid'2000 First Internationnal Workshop on Grid Computing*, volume 1971, Bangalore, India, December 2000. IEEE/ACM, Springler LNCS.
- [38] C. Germain, A. Osorio, and R. Texier. A Case Study in Medical Imaging and the Grid. In S. Norager, editor, *Procs. 1st European HealthGrid Conference*, pages 110-118. EC-IST, 2003.
- [39] C. Germain and N. Playez. Result-Checking in Global Computing Systems. In *Procs.17th ACM Int.Conf. on Supercomputing*, pages 226-233, San Francisco, June 2003. ACM Press.
- [40] C. Germain, R. Texier, and A. Osorio. Interactive Exploration of Medical Images on the Grid. In *Procs. 2nd european HealthGrid Conference*, Clermont-Ferrand, Jan. 2004. Extended version to appear in Methods of Information in Medecine.
- [41] S. Girbal, G. Mouchard, A. Cohen, and O. Temam. Dist: A simple, reliable and scalable method to significantly reduce architecture simulation time. In *International Conference on Measurement and Modeling of Computer Systems*. ACM, May 2003.
- [42] D. Gracia-Perez, G. Mouchard, and O. Temam. Microlib: A case for the quantitative comparison of micro-architecture mechanisms. In *International Symposium on Microarchitecture*. ACM, Dec 2004.
- [43] F. Gruau, Y. Lhuillier, P. Reitz, and O. Temam. Blob computing. In *Computing Frontiers*. ACM, Apr 2004.
- [44] F. Gruau and P. Malbos. The blob: A basic topological concept for hardware-free distributed computation. In *Unconventional Models of Computation*, pages 151-163. Lecture Notes in Computer Science, Springer Verlag, Oct 2002.

Research groups Archi Publications

- [45] G. Krawezik, G. Alleon, and F. Cappello. Spmd openmp versus mpi on a ibm smp for 3 kernels of the nas benchmarks. In WOMPEI 2002, ISHPC-IV, LNCS, 2002.
- [46] G. Krawezik and F. Cappello. Performance comparison of mpi and three openmp programming styles on shared memory multiprocessors. In *ACM Symposium on Parallel Algorithms and Architectures*, San Diego, USA, June 2003.
- [47] P. Lemarinier, A. Bouteiller, T. Herault, G. Krawezik, and F. Cappello. Improved message logging versus improved coordinated checkpointing for fault tolerant MPI. In *IEEE International Conference on Cluster Computing (Cluster 2004)*, 2004.
- [48] C. Limousin, J. Sébot, A. Vartanian, and N. Drach. Improving 3d geometry transformations on a simultaneous multithreaded simd. In *International Conference on Supercomputing*, pages 236-245. ACM, Jun 2001.
- [49] G. Lindenmaier, K. McKinley, and O. Temam. Load scheduling using hardware counters. In *Euro-Par*. ACM/IFIP/IEEE, Aug 2000.
- [50] O. Lodygensky, G. Fedak, F. Cappello, V. Neri, and A. Cordier. Augernome & xtremweb: Monte carlos computation on a global computing platform. In *CHEP(Conference on High Energy Physics)*, San Diego, California, USA, 2003.
- [51] O. Lodygensky, G. Fedak, F. Cappello, V. Neri, M. Livny, and D. Thain. Xtremweb & condor : sharing resources between internet connected condor pools. In *Third Workshop on Global and Peer to Peer Computing, IEEE/ACM CCGRID2003*, Tokyo, JAPAN, 2003.
- [52] G. Mouchard, D. Gracia-Perez, and O. Temam. A fast systemc engine. In *Design, Automation and Test in Europe.* EDA/IEEE, Feb 2004.
- [53] D. Parello, O. Temam, A. Cohen, and J.-M. Verdun. Toward a systematic, pragmatic and architecture-aware program optimization process for complex processors. In *Supercomputing*. IEEE, Nov 2004.
- [54] D. Parello, O. Temam, and J.-M. Verdun. On increasing architecture awareness in program optimizations to bridge the gap between peak and sustained processor performance : Matrix-multiply revisited. In *Supercomputing*. IEEE, Nov 2002.
- [55] J. Sébot and N. Drach. Memory bandwidth: the true bottleneck of simd multimedia performance on a superscalar processor. In *Euro-Par*, pages 439-447. ACM/IFIP/IEEE, Aug 2001.
- [56] J. Sébot and N. Drach. Reducing power consumption on a superscalar processor for multimedia applications. In *International Symposium on Low-Power and High-Speed Chips*. ACM-IEEE, Apr 2001.
- [57] A. Selhikov, C. Germain, G. Bosilca, F.Cappello, and G. Fedak. mpich-cm: a p2p mpi implementation. In 9th EuroPVM/MPI, Johannes Kepler University, Linz, Austria, September 2002.
- [58] A. Selikhov and C. Germain. CMDE: a Channel Memory based Dynamic Environment for Message Passing based on MPICH-V Architecture. In *Procs. Parallel Computing Technologies Conf.(PaCT)*, LNCS 2763, pages 528-537, 2003.
- [59] A. Vartanian, J.-L. Béchennec, and N. Drach. The best distribution for a parallel opengl 3d engine with texture caches. In *International Symposium on High Performance Computer Architecture*. IEEE, Jan 2000.
- [60] S. Yehia and O. Temam. From sequences of dependent instructions to functions: An approach for improving performance without ilp or speculation. In *International Symposium on Computer Architecture*, May 2004.

#### National peer-reviewed conferences

- [61] G. Bosilca, G. Fedak, T. Herault, and F. Magniette. Evaluation de performance de différentes techniques de confinement d'exécution pour le calcul pair-à-pair. In *Rencontre du parallélisme Renpar'13*, Hammamet, Tunisie, Avril 2002.
- [62] F. Cappello and D. Etiemble. Mpi ou mpi+openmp sur un cluster de multiprocesseurs : l'exemple des nas avec la sp3 du cines. In 6<sup>e</sup> Symposiun en Architectures Nouvelles de Machines, 2000.
- [63] G. Fedak. Exécution délocalisée et répartition de charge : une étude expérimentale. In *RenPar'2000,* 2000.

#### Other conferences and workshops

- [64] C. Bastoul, A. Cohen, S. Girbal, S. Sharma, and O. Temam. Putting polyhedral loop transformations to work. In *International Workshop on Languages and Compilers for Parallel Computing*, Oct 2003.
- [65] F. Cappello. Comparing performance of mpi and mpi+openmp for nas benchmark on ibm sp3. In *IBM Watson ACTC European Workshop*, Paris, France, 2000.
- [66] F. Cappello. Ovm: High performance computing with rpc programming style. In *Score Users Group Meeting*, Oxford, UK, 2000.
- [67] F. Cappello. Understanding performance of smp clusters for the nas benchmark. In *Workshop on Grid and Cluster Computing*, Tsukuba, Japan, 2000.
- [68] F. Cappello. Desktop grid with xtremweb: User experiences and feedback. In *SC2002* panel on Desktop Grids: 10,000 fold parallelism for the masses, Baltimore, USA, 2002.
- [69] F. Cappello. Xtremweb: toward high performance computing on p2p systems. In *HPC2002*, Cetraro, Italia, 2002.
- [70] F. Cappello. High performance computing on p2p platforms: Recent innovations. In *Terena Conference*, Zagreb, Croatia, 2003.
- [71] F. Cappello. Ontology of desktop grids: a pragmatic view from the xtremweb experience. In *DGRID associated with GRID 2003 and SC 2003*, Phoenix, USA, 2003.
- [72] F. Cappello. Hybrid preemptive scheduling of mpi applications on the grids. In *Scheduling Workshop*, Modane, 2004.
- [73] F. Cappello. The mpich-v project. In ENS/NSF Workshop, Lyon, 2004.
- [74] F. Cappello, S. Djilali, G. Fedak, T. Morlier, and O. Lodygensky. Calcul global, desktop grids et xtremweb. In *GridUse-2004 Ecole thématique sur la Globalisation des Ressources Informatiques et des Données : Utilisation et Services*, Metz, France, 2004.
- [75] F. Cappello and G. Fedak. Desktop grids. In *France-Korea Joint Workshop on Grid Computing*, IRISA Rennes, 2004.
- [76] F. Cappello, G. Fedak, and O. Richard. Systèmes distribués de calcul global et pair à pair. In *Tutoriel à Renpar'2001*, Paris, April 2001.
- [77] A. Cohen and O. Temam. Digital lc-2: From bits and gates to a little computer. In *International Workshop on Computer Architecture Education*. ISCA, May 2002.
- [78] M. Dupré and N. Drach. Algorithme d'ordonnancement dynamique pour exécution statique. In *Symposium en Architecture et Adéquation Algorithme Architecture*, Apr 2002.
- [79] D. Etiemble. Des flottants 16 bits sur microprocesseurs généraliste pour le traitement d'images. In sympAA'2003, pages 253-260.

g

Research groups

Publications

- [80] A. Farcy and O. Temam. Prediction de branchement: les limites de la correlation. In *Symposium en Architecture et Adéquation Algorithme Architecture*, Jun 2000.
- [81] G. Fedak. Xtremweb: an experimental platform for global and peer-to-peer computing. In *Présentation invitée, HEPiX 2001, UNIX users in the High Energy Physics,* LAL-Orsay, France, April 2001.
- [82] G. Fedak. Xtremweb: A peer-to-peer global computing experimental platform. In *Free Software and Open Source Developers Meeting FOSDEM*, Brussel, Belgium, February 2002.
- [83] G. Fursin, M. O'Boyle, O. Temam, and G. Watts. A fast and accurate method for evaluating the upper-bound of memory performance. In *International Workshop on Compilers for Parallel Computers*, Jun 2001.
- [84] D. Gracia-Perez, G. Mouchard, and O. Temam. Microlib: A case for the quantitative comparison of micro-architecture mechanisms. In *International Workshop on Duplicating*, *Deconstructing, and Debunking*. ISCA, May 2004.
- [85] F. Gruau, Y. Lhuillier, and O. Temam. Blob computing. In *Symposium en Architecture et Adéquation Algorithme Architecture*, Oct 2003.
- [86] F. Gruau and G. Moszkowski. The blob division. In *Biologically Inspired Approaches to Advanced Information Technology*. Lecture Notes in Computer Science, Springer Verlag, Jan 2004.
- [87] Y. Lhuillier and O. Temam. Ap+somt: Agentprogramming selforganized. In International Workshop on Complexity-Effective Design. ISCA, May 2004.
- [88] C. Limousin, N. Drach, and J.-L. Béchennec. Politiques d'émission pour un processeur smt. In *Symposium en Architecture et Adéquation Algorithme Architecture*, Jun 2000.
- [89] C. Limousin, J. Sébot, A. Vartanian, and N. Drach. Amélioration des performances des transformations géométriques 3d temps réel sur un processeur smt simd. In *Symposium en Architecture et Adéquation Algorithme Architecture*, Apr 2001.
- [90] J. Sébot and N. Drach. Tradeoff between power consumption and performance on a superscalar processor. In *International Kool Chips Workshop*. ACM-IEEE, Dec 2000.
- [91] J. Sébot and N. Drach. Extensions simd et superscalaire dans l'ordre : une solution pour réduire la consommation des applications multimédia. In *Symposium en Architecture et Adéquation Algorithme Architecture*, Apr 2001.
- [92] S. Yehia and O. Temam. From sequences of dependent instructions to functions: a complexity-effective approach for improving performance without ilp or speculation. In *International Workshop on Complexity-Effective Design*. ISCA, Jun 2003.

#### Reports

[93] D. Berry, C. Germain-Renaud, D. Hill, S. Pieper, and J. Saltz. Report on workshop IMAGE 03 : Images, Medical Analysis and Grid Environments. TR UKeS-2004-02, UK National e-Science Centre, http://www.nesc.ac.uk/technical\_papers/UKeS-2004-02.pdf, Feb. 2004.



B.7 / Research groups
Parallelism

Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Parallélisme

## Responsable: Brigitte Rozoy

'équipe parallélisme est composée actuellement de treize personnes : deux professeurs, quatre maîtres de conférences et sept doctorants ou postdoctorants. Au cours des quatre dernières années, huit étudiants y ont préparé et soutenu leur thèse, quatre d'entre eux devenant maître de conférences, les autres rejoignant le privé. Enfin une maître de conférences, promue profes-

seure, a quitté le laboratoire.

L'activité de l'équipe est structurée autour d'un séminaire bi-hebdomadaire, commun maintenant avec le projet Grand Large du PCRI, et des réunions informelles régulières. Outre son activité propre, l'équipe est ou a été impliquée dans diverses actions, en particulier dans le cadre du PCRI avec l'INRIA et le CNRS,



cinq ACI du CNRS (Grid CGP2P, Mobicoop, SR2I, Fragile, Star), trois Actions Spécifiques (Algorithmique répartie, Algorithmique distribuée, Réseaux et Algorithmes), un projet BQR (AirCast) avec des physiciens et un Plan Pluri-Formations, avec l'Ecole Supérieure d'Electricité.

Les permanents du groupe sont fortement impliqués dans l'animation scientifique nationale et internationale, telles les évaluations des organismes de recherche (ministère, CNRS, INRIA, CNU,...), les comités de programme, l'organisation de journées et de conférences, etc... Ils publient régulièrement dans les revues et les conférences nationales et internationales; un outil et une plate-forme de vérification sont issus de l'activité du groupe, ainsi que l'implémentation, avec le groupe "Grand Large" d'algorithmes dans une plate-forme de grille de calcul. Enfin plusieurs des membres sont fortement impliqués dans l'animation et l'administration de l'université ainsi que dans les cursus d'enseignement.

L'objectif scientifique général de l'équipe est la description, la compréhension et la maîtrise des phénomènes exprimables et exécutables dans les contextes de communications que sont les systèmes parallèles et répartis, en particulier asynchrones, éventuellement de très grande taille.

Outre la modélisation théorique, l'équipe s'est attaché plus particulièrement aux problèmes liés à la vérification des systèmes et à la résistance aux pannes via l'auto-stabilisation. Il s'agit de pouvoir faire face à une large gamme de défaillances, qui vont des pannes provoquant un arrêt de fonctionnement total de certains composants aux défaillances byzantines, conséquences de logiciels imparfaits ou même d'actions malveillantes en passant par les défaillances transitoires de type perte ou duplication de messages : ces types de problèmes ne sont pas nouveaux, mais ce qui est nouveau et qui rend les solutions anciennes finalement peu utilisables, c'est le passage à l'échelle. En particulier, nous nous concentrons sur :

• la conception d'algorithmes auto-stabilisants, éventuellement probabilistes;

- l'observation de l'auto stabilisation;
- l'étude de la fiabilité des systèmes à grande échelle, et en particulier les systèmes de calcul global;
  - la vérification efficace des systèmes répartis synchrones et asynchrones.




# Parallelism

Head: Brigitte Rozoy

The research group consists of thirteen people: two professors, four assistant professors and seven Ph.D. students or post-docs. During the last four years, eight students have defended their Ph.D. thesis, four of them became permanent university members while the others went to industry. Finally, one assistant became professor and has left the group.

The group holds a bi-weekly seminar with the Grand Large PCRI project, and meets frequently for teamwork. The group members are also involved in various collaborations with universities and research organizations: Grand Large with INRIA and CNRS, five national research pro-

jects of the ACI program (Grid CGP2P, Mobicoop, SR2I, Fragile, Star), three CNRS national working groups on distributed algorithms, one university BQR project with physicists, and one PPF project from the ministry of research with the SupElec school.

Permanent members are involved in national and international animation and evaluation activities, including the ministry of research, CNRS, INRIA, program committees, organization of conferences, summer schools, etc. They regularly publish papers in journal and conferences. Two verification tools are developed in the group and we also participate in the implementation of a grid platform. Finally, several group members are heavily involved in the university management as well as in teaching activities and responsibilities.

The general scientific goal of the research group is the description, understanding and control of the phenomena that occur in distributed environments, whether synchronous or asynchronous and including very large systems. In addition to the theoretical modelling that provides our basic framework, we deal more particularly with issues related to transient failure resilience, such as self-stabilization, and distributed system checking. Since non-faulty systems are in practice impossible to realize and control, we explicitly deal with faults, e.g. by designing and verifying algorithms that automatically correct incorrect behaviours. We particularly focus on the following problems:

• design of self-stabilizing algorithms (we mainly concentrate on scalable self-stabilization, stabilization with mobile agents, observing self-stabilization),

study of fault-tolerant protocols in the context of global computing,

- development of accurate and relevant emulation tools to emulate global computing,
- verification of synchronous and asynchronous distributed systems.

Research groups

commun de recherche

Introduction

## Research group members

#### Personnel as of 01/01/2004

Full time faculty			
Name	First Name	Position*	Institution
BEAUQUIER	Joffroy	PRCE	Université Paris XI
DELAET	Sylvie	MC	IUT d'Orsay
JOHNEN	Colette	MC	IUT d'Orsay
ROSAZ	Laurent	MC	Université Paris XI
ROZOY	Brigitte	PR1	Université Paris XI
TIXEUIL	Sébastien	MC	IUT d'Orsay
Doctoral students			
ABED	Nazha	D	Université Paris XI
HADDAD	Sammy	А	Université Paris XI
LEMARINIER	Pierre	AM	Université Paris XI
PILARD	Laurence	AM	Université Paris XI
QUESNEAU	Denys	AM	Université Paris XI
Temporary personnel			
CANTARELL	Sébastien	ATER	Université Paris XI
HERAULT	Thomas	ATER	Université Paris XI
Non-LRI personnel of	PCRI project Gran	nd Large	
CAPPELLO	Franck	DR	INRIA
GRISVARD	Gina	AJT	INRIA
PETITON	Serge	PR	EUDIL
REDON	Xavier	MCF	EUDIL
AOUAD	Lamine	D	
BOUTEILLER	Aurélien	D	
GAURON	Philippe	D	
LODYGENSKY	Oleg	D	

Engineer

INRIA

\*See glossary for acronyms

MORLIER

## Long term visitors

Long term the						
Name	First Name	Nationality	Institution	Arrival	Departure	Funding
TCHEEKO	Lot	Camerounese	ENSP de Yaoundé	17/02/00	30/06/00	
BEJAOUI	Tarek	Tunisian	Sup' Com Tunis	1/07/00	31/07/00	CNRS-DGRST Cooperation
SCHILLER	Elad	Israeli	Ben-Gurion University of the Negev Israel	12/10/00	20/10/00	
DATTA	Ajoy	American	University of Nevada	1/06/02	30/06/02	Invited Professor UPS
HERMAN	Ted	American	University of Iowa	1/04/03	31/05/03	Invited Professor UPS
HIGHAM	Lisa	Canadian	University of Calgary	30/06/03	28/07/03	Invited Professor UPS
ISRAELI	Amos	Israeli	Netanya Academic College	7/10/03	17/10/03	
MASUZAWA	Toshimitsu	Japanese	University of Osaka	16/08/04	16/09/04	Invited Professor UPS

Tangui

## **Group evolution**

Since 2000, one faculty member has left our research group :

• Véronique Vèque, assistant professor until september 2000, is now a professor at IEF (Unversité Paris-Sud).

Two new faculty members have joined the group:

- Sébastien Tixeuil as assistant professor in september 2000.
- Thomas Hérault as assistant professor in september 2004.

During this period, eight Ph.D. students have defended their thesis; four of them became permanent faculty members, the other four left the University for industry. Finally, with the creation of the Grand Large PCRI projects, members of the Cluster and Grid team from the Archi group are in the process of joining the Parall group.

## B.7.2 / Parall

## **Research description**

The general scientific goal of the research group is the understanding, design and control of the phenomena that occur in a distributed environment, whether synchronous or asynchronous one and including very large systems. In addition to the theoretical modelling that provides our basic framework, we deal more particularly with issues related to transient failure resilience, such as self-stabilization, and distributed system checking. Thus we pay a special attention to problems linked with faulty processes or faulty networks.

Since non-faulty systems are in practice impossible to realize and control, we explicitly deal with faults, e.g. by designing and verifying algorithms that automatically correct incorrect behaviours.

We particularly focus on the following problems:

- design of self-stabilizing algorithms (mainly we concentrate on scalable self stabilization, stabilization with mobile agents, observing self stabilization),
- study of fault-tolerant protocols in the global computing context,
- development of accurate and relevant emulation tool to emulate global computing,
- verification of synchronous and asynchronous distributed systems.

## PCRI Project Grand Large



• Participants: Joffroy Beauquier, Thomas Hérault, Pierre Lemarinier, Laurent Rosaz, Brigitte Rozoy, Sébastien Tixeuil.

*Grand Large* is a Grid research project investigating the issues raised by computing on Large Scale Distributed Systems (LSDS), where participants execute different applications on shared resources belonging to other participants, possibly geographically and administratively independent. More specifically, we consider large-scale parallel and distributed computing on Peer-to-peer (P2P), Global Computing and Desktop Grid systems. Our research focuses on middleware and low-level programming environments design, proof and experiments. Fundamentally, we address the impact of LSDS using several methodological tools: theoretical models, simulators, emulators and real-size systems.

The goals of the project are:

- to study both formally and experimentally the fundamental mechanisms of LSDS for high-performance computing;
- to design, implement, validate and test real software, middleware and platforms;
- to define, evaluate and experiment approaches for programming applications on these platforms.

Compared to other European and French projects, we gather skills in large-scale systems (large-scale scheduling, volatility tolerance, heterogeneity, inter-administration domain security, etc.) acquired with the XtremWeb project (LRI, Cluster and Grid team), formal design and validation of algorithms and protocols for distributed systems (LRI, Parall team) and programming, evaluation, analysis and definition of programming languages and environments for parallel architectures and distributed systems (LIFL-Lille, methodologies and parallel algorithms).

Grand Large pursues short and long-term research aiming to have a scientific and industrial impact in the areas of LSDS, Grid and large clusters. Research topics include:

- the design of a middleware to expand the application domain of Desktop Grid;
- resource discovery engines on large-scale system with volatile participants;
- large-scale storage on volatile nodes;
- simulation of large-scale scheduling;
- fault-tolerant MPI for large-scale systems;
- algorithms for large-scale fault tolerance;
- protocol verification;
- algorithms, programming and evaluation of scientific applications on desktop Grids;
- tools and languages for large-scale computing.

In our longer-term research, we investigate the conditions for the convergence of Global Computing, P2P and Grid systems (how Grid Services can be used in Desktop Grid) and experimental tools for improving the methodology associated with research in LSDS. For example we are in charge of the Grid eXplorer project founded by the French ministry of research and we are deeply involved in the Grid5000 national project.

#### Building a Large Scale Distributed System (LSDS) for Computing

The main application domain for the Large Scale Distributed Systems developed in Grand Large is high-performance computing. The two main programming models associated with our platform, remote procedure call (RPC) and message passing interface (MPI), allow to program a large variety of distributed/parallel algorithms with computational paradigms such as bag of tasks, parameter sweep, workflow, dataflow, master worker, recursive exploration with RPC, and single program/multiple data (SPMD) with MPI. The RPC programming model can be used to concurrently execute different application codes or the same application code with different parameters and library function codes. In all these cases, there is no need to change the code. The code must only be compiled for the target execution environment. LSDSs are particularly useful for users with large computational needs. They can typically be used in Research and Development departments of Pharmacology, Aerospace, Automotive, Electronics, Petroleum, Energy and Meteorology industries. LSDSs can also be used for purposes other than CPU intensive applications. Other resources of the connected PCs can be used, e.g., memory, disc space and networking capacities. A Large-Scale Distributed System such as XtremWeb can typically be used to harness and coordinate the usage of these resources. In that case XtremWeb deploys Workers services dedicated to provision and management of disc space and network connection. The storage service can be used for large-scale distributed fault-tolerant storage and distributed storage of very large files. The networking service can be used for server tests in real-world conditions, e.g. workers deployed on the Internet are coordinated to stress a Web server, and for networking infrastructure tests in real-world conditions, e.g. workers with known characteristics are coordinated to stress the network infrastructure that connects them.

#### Security and reliability of network control protocols

The main application domain for self-stabilizing and secure algorithms is LSDSs where correct behaviours must be recovered within finite time. Typically, in LSDSs such as high-performance computing systems, a protocol is used to control the system, submit requests, retrieve results and ensure that the computation is carried out accordingly to its specification. Yet, since the scale of the system is large, it is likely that nodes will fail while the application is executing. While nodes that actually perform the computation can fail unpredictably, a self-stabilizing and secure control protocol ensures that a user submitting a request will obtain the corresponding result within a (presumably small) finite time. Examples of LSDSs that use self-stabilizing and secure algorithms include global computing platforms and peer-to-peer file-sharing systems.

Another application domain is routing protocols, which are used to carry information between nodes that are not directly connected. Routing should be understood here in its most general sense, e.g. at the network level (Internet routing) or at the application level (on virtual topologies that are built on top of regular topologies in peer-to-peer systems). Since the topology (actual or virtual) evolves quickly through time, self-stabilization ensures that the routing protocol eventually provides accurate information. However, for the protocol to be useful, it is necessary that it provides extra guarantees either on the stabilization time (to recover quickly from failures) or on the routing time of messages sent when many faults occur.

Finally, additional applications can be found in distributed systems that are composed of many autonomous agents that are able to communicate only with a limited set of nodes, e.g. due to geographical or power consumption constraints, and whose environment is evolving rapidly. Examples of such systems are wireless sensor networks, while are typically large (10000+ nodes), mobile autonomous robots, etc. It is completely unrealistic to use centralized control on such networks because they are intrinsically distributed; yet strong coordination is required to provide efficient use of resources such as bandwidth or battery life.

Research groups

**Research** description

#### **End-User Tools for Computational Science and Engineering**

Another Grand Large application domain is Large-Scale Programming for Computational Science and Engineering. Two main approaches are proposed. First, we want to experiment and evaluate such programming. Second, we want to develop tools for end-users.

In addition to classical supercomputing and GRID computing based on virtual organizations, the large scale P2P approach proposes new computing facilities for computational scientists and engineers. On the one hand, many applications already exist, some of them classical, e.g. Computational Fluid Dynamic or Quantum Physic, others new and of strategic importance, e.g. Nanotechnology, which will have to use a lot of computing power for long periods of time in the near future. On the other hand, a new large-scale programming paradigm emerges for existing computers that can be accessible by scientists and engineers for classical application domains as well as new ones, such as some Non-Governmental Organisations. During a first period, many applications would be based on large simulations rather than classical implicit numerical methods, which are more difficult to adapt for such large problems and new programming paradigm. Nevertheless, we expect that more complex implicit methods will be adapted to such programming . The potential number of peers and the planed evolution of network communications, especially multicast ones, will allow to solve some of the grand challenges in scientific applications.

Simulations and large implicit methods will always have to compute linear algebra routines, which will be our first targeted numerical methods (we also note that the most powerful worldwide computing facilities are still rated using a linear algebra benchmark http://www.top500.org). We will especially focus on divide-and-conquer and block-based matrix methods to solve dense problems and on iterative hybrid methods to solve sparse matrix problems. As these applications are used by many applications, it is possible to extrapolate the results to different scientific domains.

Many advanced tools must be developed in order to help end-users program such environments. At present, the main goal is to experiment on large platforms so as to evaluate and extrapolate performance with respect to many parameters and under some specific hypothesis concerning scheduling strategies and multicast speed. We must also propose tools for end-users, keeping in mind that the end-user is at the center of scientific programming. Finally, we have to propose a framework to program P2P architectures that completely virtualizes the P2P middleware and heterogeneous hardware. Our approach is based, on the one hand, on component-based programming and coordination languages, on the other hand, on the development of an Application Service Provider (ASP), which may be dedicated to a given scientific domain. The result would be a P2P scientific programming methodology based on experimentation and evaluation on an actual P2P development environment.

## Modelling and Verification

• Participants: Brigitte Rozoy, Dominique Ambroise, Patrick Augé, Laurence Pilard, Laurent Rosaz

This research theme deals with algorithms for distributed systems and more precisely with the verification by simulation of distributed algorithms running on synchronous or asynchronous networks. Indeed, such systems can usually exhibit an extremely large number of different behaviours. This is due to the combinatorial explosion resulting from all the possible interactions between the different concurrent components of the system and the many race conditions that may arise among them. Verification provides the means to ensure the correctness of the design of such concurrent systems. It means checking that a system description conforms to its expected properties. These properties can range from several forms of consistency to complex correctness requirements; verification is thus a tool that enables the designer to be confident that the formal description of the system does satisfy the problem requirements.

State-space exploration is one of the most successful strategies for analyzing and verifying such concurrent reactive systems: it consists of exploring a global state graph representing the combined behaviour of all concurrent components, which represents the set of all possible executions. It is well known that the main difficulty is the huge number of nodes in the graph. This is inherently connected to concurrency and non-determinism due to asynchronous communications.

The originality of our approach is to try to take advantage of automata, partial orders and traces at the same time: we give the semantic of these systems in terms of prime event structures with binary conflict.

This event-based model is related to other approaches that may be grouped under the generic heading of "Partial Order Semantics". These approaches essentially rely on the view that the description of a run of the system must take into account events, which are actions occurring more or less independently at various locations of a network. Here the Event Structure model is based on events, seen as occurrences of actions; it emphasizes relationships between these actions and describes them by two relations, an order relation and a conflict relation.

Word-based models are another classical description of distributed systems, commonly known as "the interleaving point of view". Given an alphabet, letters can be viewed as actions and words may be used to represent possible firing sequences. Thus languages can describe the behaviour of the system in terms of the possible sequences of actions that it exhibits: trace languages use an independence relation defined on the alphabet of actions and captures the causal concurrency of a pair of actions by their adjacent occurrence in a word. Within this framework, tools and techniques from formal language theory can be used to study distributed systems. After studying some properties of these languages [19, 5] we have tried to use them for verification [24, 25, 26, 70, 69, 18].

In practice, for the asynchronous case, when starting with programs written in an eventdriven language, the process consists of exhibiting a single and compact object that represents all its possible executions, namely an event structure. Using this structure, the state graph can be constructed and studied in a very efficient way. For the synchronous case, we use a state-based model using a product of automata. We propose two strategies whose efficiency results from the underlying model. The first algorithm covers the graph by a tree; it is efficient in the sense that any state is constructed exactly once. The second algorithm does not cover the graph but constructs a number of nodes that is sufficient for the detection of the desired properties; it works by executing at once as many concurrent events as possible. We showed that this could be applied to a large class of distributed algorithms. Research groups

Research description



The theoretical part of our work resulted in a concrete tool, the Marrella simulator, which was developed in order to observe our strategies and test their performance. We have tested our technique with numerous examples of classical asynchronous algorithms with a large set of states and tried to compare it with partial order reductions. For valid comparisons with other techniques, we needed a similar environment for other techniques. We used an open validation environment, called IF, with the idea of having heterogeneous tools cooperate through translators. IF includes a translator from SDL to IF. The VIRES project provides a translator from IF to PROMELA. At LRI, we have developed a translator from IF to an ESTELLE-like language, which is the input language for Marrella. The translators ensure that the semantic interpretation of an IF program is the same in SDL, PROMELA and ESTELLE-like programs. Each algorithm was written in the IF language. Thus the reduction techniques were applied to identical state graphs. This work showed that Marrella compared favourably with other tools using known strategies.



*Figure 2: Marella in practice* 

The reduction ratios between the complete and reduced graphs are large. In most cases, the reduction obtained by Spin Ulg and Marrella are similar: they reduce the number of states from o(np) to o(n'), where *p* is the number of processors and *n* the average number of actions that each processor performs. Second, since Marrella does not construct intermediate states, it may gain an additional factor. This factor results from the degree of parallelism, which is less than the number of processors. The division factor obtained with Marrella is most often about *p*/2, *p*/3 or *p*/4, where *p*'' is the number of processors. For the bounded retransmission protocol however, the performance drops and Marrella may even run out of space. An explanation may be that the considered protocol is mostly sequential and cannot be reduced efficient-

ly with this kind of methods. Note that our technique does not cache states: intermediary states are neither stored nor constructed. This is clearly important since state-space caching involves a run-time penalty. As a consequence, even though Marrella is only a prototype written in a Lisp-like language, its performance is also good in terms of construction time. Surprisingly, Marrella's construction times are often similar to those of Spin and Spin Ulg..

Another set of experiments were performed in order to compare Marrella with the PeP tool, based on Petri Nets and written in C: Marrella is better for states but worse for execution time. Finally, in collaboration with the electrical engineering department, we have used the same tools and ideas in order to design and verify a multi-agent embedded system and showed a performance gain too.



*Figure 3: Multi-agent system for collision-risk avoidance.* 

#### Key references:

- [18] F. Magniette, L. Pilard, and B. Rozoy. A method for the verification of distributed and synchronized algorithm. *International Journal of Production Research*, page to appear, 2004.
- [19] L. Rosaz. The word problem for 1lc-congruences is np-hard. *Theoretical Computer Science*, 306:245-268, 2003.
- [24] D. Ambroise, P. Augé, K. Bouchefra, and B. Rozoy. Marrella and the verification of an embedded system. In M. M. Furio Honsell, editor, *Foundations of Software Science and Computation Structures, 4th International Conference, FOSSACS 2001 Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2001 Genova, Italy,* number 2030 in Lecture Notes in Computer Science, pages 409-412. Springer-Verlag, April 2001.
- [25] D. Ambroise, B. Rozoy, and J. Saquet. Deadlock detextion in distributed systems. In *Proceedings of the 18th International Conference on Computers And Their Applications (ISCA'2003)*, pages 210-213, Honolulu, Hawaii, USA, March.
- [58] L. Rosaz. Self-stabilizing token circulation on asynchronous uniform unidirectional rings. In Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing (PODC'2000), pages 249-258, Portland, Oregon, USA, July 2000. ACM.

Research groups

## Parall

**Research description** 

## **Distributed Algorithms**

• Participants: Joffroy Beauquier, Sylvie Delaët, Sammy Haddad, Thomas Hérault, Christophe Genolini, Maria Gradinariu, Colette Johnen, Frédéric Magniette, Laurence Pilard, Denis Quesneau, Laurent Rosaz, Brigitte Rozoy, Sébastien Tixeuil

Note: Some members of this theme (Joffroy Beauquier, Thomas Hérault, Laurent Rosaz, Brigitte Rozoy and Sébastien Tixeuil) also participate in the Grand Large PCRI Project.



Starting from an arbitrary initial state, a self-stabilizing system eventually exhibits correct behavior.

Dijkstra introduced the notion of self-stabilization in 1974 by defining a system as selfstabilizing when "regardless of its initial state, it is guaranteed to arrive at a legitimate state in a finite number of steps". Such a property is very desirable for any distributed system: after any unexpected perturbation modifying the memory state, the system eventually recovers and returns to a legitimate state without any outside intervention.

Dijkstra's notion of self-stabilization, which originally had a very narrow scope of application, has proved to encompass a formal and unified approach to fault-tolerance under a model of transient failures for distributed systems. Compared to traditional protocols, selfstabilizing protocols are more general, because they do not need an initialization phase. Moreover they have a larger application domain, due to their intrinsic resistance to failures.

Self-stabilization has received a lot of attention over the last few years, with papers regularly presented in the best distributed computing conferences and special issues of journals. Our work on self-stabilization is concentrated on several precise issues.

#### **Self-stabilizing Routing**

In the context of computer networks, resuming correct behaviour after a fault occurs can be very costly: the whole network may have to be shut down and globally reset to a proper initial state. While this approach is feasible for small networks, it is far from practical in large networks such as the Internet. Self-stabilization provides a way to recover from faults without the cost and inconvenience of a generalized human intervention: after a fault is diagnosed, one simply has to remove, repair, or reinitialize the faulty components, and the system, by itself, will return to a good global state within a relatively short amount of time.

In 2000, Varghese and Jarayam showed that crash and restart failures could lead distributed systems into arbitrary configurations, highlighting the need for algorithms that are resilient to those failures, such as self-stabilizing algorithms. Fault-tolerance is an important issue in the design of network routing protocols since the topology changes due to the link/node failures and repairs. As a matter of fact, many existing routing protocols are selfstabilizing. For example Nancy Lynch has proven that the Routing Information Protocol (RIP) and the Open Shortest Path First (OSPF) algorithms are self-stabilizing.

Our contribution to self-stabilizing routing (Figure 4) is threefold. First, we studied precisely to which extent various routing metrics are self-stabilizing in the context of unidirectional networks, where not every link is bidirectional (wireless networks and satellite communication fall into this category). It turns out that restricted metrics enable stabilization even over unreliable communication media [15, 14], while broader metrics permit stabilization in reliable networks [16]. Second, we added self-stabilizing properties to existing communication networks [12]: in the context of Wormhole routing networks, [39, 8] study the case of the standard inter-domain routing protocol in the Internet (BGP). Finally, we proposed in [55] a routing algorithm that is not only self-stabilizing but also route preserving. This last property means that, a tree being constructed, any message sent to the root is received in a bounded amount of time even in the presence of continuous edge cost changes.

#### Scalable self-stabilization

Recovering from memory or communication channel corruptions has been the main goal of the theory of self-stabilization. A self-stabilizing system, regardless the initial states of the processors and initial messages in the communication channels, is guaranteed to converge to the intended behaviour in finite time. But most of the self-stabilizing algorithms are not scalable. This is not surprising because they were intended to recover even after the corruption of all processors and all channels. In most cases, even if the effective number of corruptions is small, the general mechanism is such that the cost of recovery (in terms of time) is still proportional to the size of the network.

To circumvent this problem, some authors, such as Kutten, Peleg and Patt-Shamir developed the idea that global fault handling methods are inefficient and that there is a growing need for scalable methods, i.e. methods whose cost depends on the effective extent of the faults rather than the size of the network. These papers argued for the development of a systematic methodology for designing what they called fault-local error handling algorithms.

On the negative side, we proved [47] that there exists a lower bound of D atomic steps (where D is the network diameter) before stabilization after a single fault hits an asynchronous dynamic system. This implies that for non-trivial problems, scalable stabilization is only meaningful using the round complexity model where each processor executes at least one step in a round.

Research groups

Research description





Eventually, Each Node Knows All The Network



Figure 4: Self-stabilizing Routing.

On the positive side, we proposed a technique [29] for designing static algorithms that have the following three properties:

- (i) They are state self-stabilizing, i.e. they reach a legitimate state after any number of processors have been corrupted and stay in legitimate states as long that there are no new corruptions,
- (ii) Given a bound k lower than half the total number of processors, if affected by f<k corruptions of processors after stabilization, they output-stabilize in a time proportional to f , and
- (iii) They use a memory size that depends on k.
- We also developed a dynamic algorithm for mutual exclusion [86] that is also scalable and self-stabilizing.

On the practical side, scalable stabilization is well suited so sensor networks (Figure 5), where hundreds of thousands of sensor nodes collaborate to perform a monitoring task (alerting in case of fire, intrusion detection, etc.). In this context, we proposed [52] a general schema to enable reusing self-stabilizing algorithms in the communication model of sensor networks, where communications can collide in a hidden manner. Our approach is self-stabilizing and scalable, because its stabilization time is bound by a constant independent of the number of actual failures.



Research groups

Parall Research description

Figure 5: Sensor Networks.

#### Stabilization with mobile agents

As stated above, self-stabilization is a framework for dealing with transient failures corrupting the RAM of processors and the communication messages. Usually the code of processors is assumed to be stored in ROM and henceforth to be incorruptible. We have studied how to overcome this limitation design self-stabilizing algorithms even if some code is corrupted.

We proposed a technique based on mobile agents [30]. A mobile agent can be consid-

ered (in an abstract way, independently of its implementation) as a message of a special type, which is created by an upper layer (the outside). It may carry two types of data: code that can be installed at some nodes and information (referred to as contained in the briefcase) that can be read or modified by the processors. Code and briefcase are specific to the application controlled by the agent. Once created, an agent circulates in the network according to some agent circulation rules. A particular rule is presented in [59].

Because it is supposed to travel, an agent (code and briefcase) can be corrupted. In our agent model, there are two crucial properties: non-interference and stabilization. The non-interference property states that if no failure (corruption) occurred after the creation of the previous agent, then after the destruction of this agent the controlled application must not behave differently whether or not a new agent is present in the network. This property implies that the application cannot use the agents for solving its problem and that the agents cannot perform resets in a predetermined global configuration. The stabilization property states that if a failure occurred before the creation of an agent and if no failure occurs during the lifetime of this agent, then when the agent disappears the application behaves according to its specification. We gave several examples on how to use agents for resisting corruption failures.



Figure 6: Merging Mobile Agents

Of course, unicity of the agent is a crucial property of the system, so we also studied the well-known problem of self-stabilizing token passing. Starting from an arbitrary initial state where any number of tokens can be present in the system, the algorithm has to guarantee that after a finite time a single token remains and circulates fairly in the system. We concentrated on optimizing the three main complexity measures for this problem [38, 58, 53, 13, 42]: (i) stabilization time, the time before a single token remain, (ii) service time, the time between two arrivals of the token at a given node when stabilized, and (iii) the memory space needed for the algorithm. Our solutions perform in uniform and anonymous networks and are thus randomized. The overall principle is simple (Figure 6): tokens are given different speeds so that they eventually meet at the same node and merge and only one token remains at the end.

#### **Observation of self-stabilization**

Traditionally, self-stabilization is considered as suffering two main drawbacks. The first is that a self-stabilizing system recovers only after some time, meaning that during that time the behaviour is not correct. The second is that a process can never know whether or not the system is stabilized. There is little we can do about the first problem because it is inherently bound to the very definition of self-stabilization. We simply have to pay a much smaller overhead than for robust algorithms approach, e.g., consensus, and the payload is the temporary loss of safety.

Surprisingly enough, we are not aware of any study dealing with the second issue, apart from the important paper by Liu and Simon in 1992. If it is obvious that no detection of stabilization from the inside is possible since any local variable used for this could be corrupted, nevertheless it is perfectly feasible to detect stabilization from the outside. For instance and although this is just a theoretical remark, when a bound on the number of steps before stabilization is known, one can simply count the steps. "From the outside" can be replaced by "from the inside but using stable memory" (memory not subject to failures). By focusing on ring networks, Liu and Simon propose "a new model, in which it is meaningful to say that a processor knows that the ring is stable". This model introduces the notion of a distributed observer located at each node of the network and responsible for detecting stabilization and does not influence the self-stabilizing protocol. Liu and Simon did not address the implementation of their observer, but it is straightforward to see that implementing the observer involves, at each node, the presence of a stable memory. If such an assumption can be valid on small local area networks, in which strong security and reliability can be ensured, it is unrealistic for world-wide, heterogeneous networks in which implementing a stable memory on each machine would be, if ever possible, extremely costly.

We tried to solve the observation problem by using much less safe memory. In particular, we raised the question of whether or not stabilization detection is feasible if only one node has some stable memory. Our main result [4] can be stated as follows: if there exists a self-stabilizing distributed solution for some problem in some network, then there exists a self-stabilizing distributed solution (not necessarily the same) for the same problem in the same network that can be observed by an observer located at a particular node. Moreover the observable self-stabilizing solution can be effectively built from the simply self-stabilizing solution in a canonical way, and therefore it can be built into a compiler.

#### **Key references:**

- [4] J. Beauquier, L. Pilard, and B. Rozoy. Observing locally self-stabilization. *Journal of High Speed networks*, page to appear, 2004.
- [14] S. Delaët and S. Tixeuil. Tolerating transient and intermittent failures. *Journal of Parallel and Distributed Computing*, 62(5):961-981, May 2002.
- [42] P. Duchon, N. Hanusse, and S. Tixeuil. Optimal randomized self-stabilizing mutual exclusion in synchronous rings. In *Proceedings of the 18th Symposium on Distributed Computing* (*DISC(2004)*, page to appear. Springer Verlag, October 2004.
- [52] T. Herman and S. Tixeuil. A distributed tdma slot assignment algorithm for wireless sensor networks. In *Proceedings of the First Workshop on Algorithmic Aspects of Wireless Sensor Networks (AlgoSensors'2004)*, number 3121 in Lecture Notes in Computer Science, pages 45-58, Turku, Finland, July 2004. Springer-Verlag.
- [55] C. Johnen and S. Tixeuil. Route preserving stabilization. In *Proceedings of the Sixth Symposium on Self-stabilizing Systems (SSS'03)*, Lecture Notes in Computer Science, San Francisco, USA, june 2003. Springer Verlag. Also in the Proceedings of DSN'03 as a one page abstract.

Research groups

Research description

## Global Computing

• Participants: Sébastien Tixeuil, Joffroy Beauquier, Thomas Hérault, William Hoarau, Pierre Lemarinier

Note: Members of this theme also participate in the Grand Large PCRI Project.

The global computing theme of the research group has two main topics. The first domain of research is the study of fault tolerant protocols in the context of global computing. The second domain of research consists in providing accurate and relevant emulation tools. Such tools allow emulating the execution of global computing systems in smaller (and more manageable) distributed systems.

#### Fault-tolerance in global computing

Global computing implies reconsidering the basic hypothesis of distributed computing. The network size is much larger, involving different network types and administration domains. The research effort in this context has two main purposes. First it consists in adapting and validating classical fault-tolerant protocols. Second, it implies elaborating new fault-tolerant protocols fitting different architectures.

Classical fault-tolerant protocols, such as Chandy and Lamport's algorithm (1985) make all the components of the distributed system simultaneously take frequent global snapshots. In order to increase the asynchrony of the technique, three main techniques of uncoordinated checkpoints and message logging have been proposed: pessimistic, causal and optimistic. During the last four years, we adapted these fault-tolerant protocols to fit global computing environments [31, 33, 57].

Based on our previous experience, a new fault-tolerant protocol targeting cluster GRIDs is currently studied. GRIDs provide a natural hierarchy among the set of clusters that are part of the GRID. However, there is no known protocol that can take advantage of this hierarchy and classical protocols performance is tightly bound to the number of components. Thus, new protocols that use the hierarchy should be designed for GRIDs. These protocols can be a composition of the classical protocols at different levels, but their composition is difficult: indeed, the performance of a protocol depends also on the amount of non-determinism inside a single node, and the non-determinism of a set of machines working together is exponentially higher than the non-determinism of a single node. Thus, the composition of protocols has to follow precise rules in order to ensure good performance and preserve correctness.

#### **Global computing emulation**

Global computing is based on the use of a large number of computing nodes. Testing and validating new applications dedicated to global computing is very difficult since it requires handling a large number of computers and it is impossible in practice to reproduce an experiment. One way to address this problem is to run applications on an emulated global computing system. This approach is accurate only if the emulated system behaves in the same way as an effective global computing system. This requires a significant effort to model the system.

Many aspects of the distributed system must be emulated in order to provide accurate results for the tested application. One of the aspects which are often overlooked is the pres-

ence of failures inside the system. For global computing systems, computer failure are not a rare event but a characteristic of the system, which behaves according to a fault scheme. Moreover, failures have dramatic impact on the performance of many fault-tolerant distributed applications [34]. We are designing a language and a set of tools to precisely describe and submit emulated applications to reproducible fault scenarios. Another important research topic is the extraction of fault models from collections of traces of real global computing systems [95].

#### **Key references**

- [31] G. Bosilca, A. Bouteiller, F. Cappello, S. Djilali, G. Fedak, C. Germain, T. Herault, P. Lemarinier, O. Lodygensky, F. Magniette, V. Neri, and A. Selikhov. MPICH-V: Toward a scalable fault tolerant MPI for volatile nodes. In *Proceedings of the 15th High Performance Networking and Computing conference (SC'02)*, page electronic, Baltimore, USA, November 2002. IEEE/ACM.
- [33] A. Bouteiller, F. Cappello, T. Herault, G. Krawezik, P. Lemarinier, and F. Magniette. Mpichv2: a fault tolerant mpi for volatile nodes based on pessimistic sender based message logging. In *Proceedings of the 16th High Performance Networking and Computing conference* (SC'03), page electronic, Phoenix, USA, November 2003. IEEE/ACM.
- [34] A. Bouteiller, P. Lemarinier, G. Krawezik, and F. Cappello. Coordinated checkpoint versus message log for fault tolerant MPI. In *IEEE International Conference on Cluster Computing (Cluster 2003)*, pages 242-250. IEEE Press, december 2003.
- [57] P. Lemarinier, A. Bouteiller, T. Herault, G. Krawezik, and F. Cappello. Improved message logging versus improved coordinated checkpointing for fault tolerant MPI. In *Proceedings* of the 6th international conference on Cluster Computing (CLUSTER'04), San Diego, USA, September 2004. IEEE. to appear.

## **Research Perspective**

For two years the team has been involved in the Grand Large PCRI project. The development of this project is our main perspective.

On the one hand, the need to solve problems such as security, volatility and fault tolerance in a global computing framework will lead to a new algorithmic approach in which the important issue is scalability. We have already worked on scalability, e.g., self-stabilizing time adaptivity, fault containment, mobile agent techniques, and have gathered some experience in this area. We intend to use this expertise to develop fault-tolerant algorithms for global systems under scalability constraints. Such algorithms are required for the software development involved in "Grand Large". Note that these issues are also faced by sensor networks, which have the supplementary requirement of electrical power saving. We have started to address such networks in the context of the FRAGILE ACI project.

On the other hand, the activity of several members of the group is now more focused on software development. This orientation also appears in the topics of several Ph.D.s that were proposed recently. The effort to make MPICH-V fault-tolerant will be continued, and new solutions will be tested. We have also initiated the design and development of a fault injection system. This system will consist of a language allowing to specify complex fault patterns. Then a compiler will be implemented, allowing to emulate complex distributed applications with reproducible faults. A possible target for this system is the Grid Explorer platform. Research groups

Parall

Research perspective

#### B.7.3 / Parall

## Highlights

The permanent members are involved in national and international animation and evaluation activities, including the ministry of research, CNRS, INRIA, program committees, organization of conferences, summer schools, etc. They regularly publish papers in journal and conferences. Two verification tools are developed in the group and we also participate in the implementation of a grid platform. Finally, several group members are heavily involved in the university management as well as in teaching activities and responsibilities.

Since 1999, the research group has maintained the quantity of referred publications and Ph.D. defenses. The number of program committee participations has slightly increased, while the amount of contracts raised significantly over the past four years.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international), and,
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

## Honors

## Prizes and awards

• Sébastien Tixeuil, Best paper award, *IEEE International Conference on Parallel and Distributed Systems*.

#### **Keynote addresses**

- Joffroy Beauquier, International conference on Principles of Distributed Systems (OPODIS'2003): "Future trends in self-stabilization".
- Joffroy Beauquier, *ACM Conference on Principles of Distibuted Computing* (PODC'2004): "Self-stabilization : past, present, future".

#### B.7.5 / Parall

## **Evaluation of research**

## **Editorial boards**

• PPL, Parallel Processing Letters: Joffroy Beauquier, member of the editorial board.

## **Program committees**

#### **International events**

- WODES, IFAC Workshop on Discrete Event Systems : Brigitte Rozoy (2004).
- OPODIS, *International Conference on Principles of Distributed Systems*: Joffroy Beauquier (2000-2004), Colette Johnen (2003-2004).
- MOVEP, MOdelling and VErification of Parallel Processes: Brigitte Rozoy (2000, 2002, 2004).
- SSS, Fifth Symposium on Self-stabilizing Systems: Joffory Beauquier (2003).
- DISC, International Symposium on Distributed Computing: Colette Johnen (2003).
- ICPADS, IEEE International Conference on Parallel and Distributed Systems: Sébastien Tixeuil (2002).
- WSS, Fourth Workshop on Self-stabilizing Systems: Joffroy Beauquier, Sébastien Tixeuil (2001).

#### **National events**

- MSR, Modélisation des Systèmes Réactifs: Brigitte Rozoy (2000, 2001, 2003).
- RenPar, Rencontres sur le Parallélisme: Brigitte Rozoy (2000).

#### **Other evaluation activities**

- Reviewer for the Israel Science Foundation (ISF): Sébastien Tixeuil.
- Reviewer for CNRS and the Ministry of Research: Brigitte Rozoy.
- Reviewer for PhD dissertation: Joffroy Beauquier (3).

#### Research groups

#### Parall

Highlights Honors Evaluation of research

## Volunteer professional service

#### Management positions in scientific organisations

- Scientific coordinator, French Ministry of Research (Mission Scientifique Universitaire, MSU and Mission Scientifique, Techniquet et Pédagogique, MSTP): Brigitte Rozoy.
- University hiring committee, *Commission de Spécialistes*, section 27, Université Paris-Sud: Joffroy Beauquier, Chair.
- Computer Science Department, Université Paris-Sud: Brigitte Rozoy, Chair.
- Université Paris-Sud: vice president for information and communication technology : Brigitte Rozoy.

#### Organisation of conferences and scientific events

- *Journées Internationales de l'Auto-stabilisation*, Luminy, October 2002: Joffroy Beauquier, Colette Johnen, Chairs.
- Journée Réseaux et Algorithmes répartis, Orsay, June 2002: Sylvie Delaët, Thomas Hérault, Colette Johnen, Sébastien Tixeuil, Chairs.
- MOVEP, MOdelling and VErification of Parallel processes: Brigitte Rozoy, Organization Commitee member (2000, 2002).
- Algotel, *Rencontres Francophones sur l'Algorithmique des Télécommunications*, 2001: S. Tixeuil, Organization Commitee member.
- WSS, Fourth Workshop on Self-stabilizing Systems, 2001: S. Tixeuil, Publicity Chair.

## Working groups

- CNRS Working group on dynamic networks, *Action Spécifique Dynamo* (led by Pierre Fraigniaud, LRI): Colette Johnen, Sébastien Tixeuil.
- CNRS Working group and distributed algorithms and their applications, *Action Spécifique Algorithmique Distribuée et Applications* (led by Carole Delporte and Hugues Fauconnier, LIAFA): Joffroy Beauquier, Sylvie Delaët, Colette Johnen, Sébastien Tixeuil.
- CNRS Working group communication networks and , *Action Spécifique Réseaux de Communication et Automatique*: Sébastien Tixeuil.
- Research Action *TAROT* of the national network on architecture, networks, systems and parallelism, *GdR ARP* : Sylvie Delaët, Colette Johnen, Sébastien Tixeuil.

#### **Other professional service**

• Reviewer for National University Concil, Conseil National des Universités (CNU): Joffroy Beauquier.

## **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Total €
ACI	S. Tixeuil	Mobicoop (JC)	MENRT	CNRS	2001-2004	12 958
ACI	C. Johnen	Star (JC)	MENRT	CNRS	2001-2004	25 958
ACI	J. Beauquier	CGP2P	MENRT	CNRS	2002-2005	34 033
ACI	S.Tixeuil	SR2I	MENRT	CNRS	2004-2007	25084
ACI	S.Tixeuil	Fragile	MENRT	CNRS	2004-2007	91656
AS	C. Johnen	Réseaux et algorithmes / GDR 725	CNRS vdl	CNRS	2002	1 000
AS	J. Beauquier	RTP5 Algorithmes distribués et Applications	CNRS Stic	CNRS	2003-2004	3 000
BQR	C. Johnen	BQR (physique Projet AIRCAST)	UPS	UPS		3 049
BQR	C. Johnen	BQR	UPS	UPS		15 245
SUBV	L. Rosaz	Algorithmique répartie	UPS/SUBSF	MEN		7 584
SUBV	V. Vèque	Tunisie	CNRS	CNRS		1 220
SUBV	J. Beauquier	PPF SUPELEC	UPS	UPS/CONTF		9 057

Note: see glossary for acronyms

## Summary of scientific projects and main results:

#### Mobicoop

Partners: LRI, LaBRI (Université Bordeaux I)

This project focused on studying the influence of mobility and topology for problems involving cooperating mobile agents.

STAR

STAbilisation des Réseaux d'ordinateurs fondés sur la technologie Internet.

The goal of the STAR project was to apply the principles of self-stabilization to the management of the Internet. We mainly studied routing protocols. We designed a self-stabilizing routing protocol that guarantes service even during the stabiliza-

tion phase (phase where more efficient routes are built according to the link costs). A first self-stabilizing Border Gateway protocol (BGP) was also designed.

Project funded by the national research program for junior researchers (*ACI Jeunes Chercheurs*). Scientific director for LRI: Sébastien Tixeuil.

Project funded by the national research program for junior researchers (*ACI Jeunes Chercheurs*). Scientific director: Colette Johnen.

#### Research groups

Parall

Volunteer professional service Contracts and grants

#### Calcul Global Pair à Pair.

CGP2P

Partners: LRI(Parall group and Cluster & Grid team from Archi group), IMAG ID (Université de Grenoble), LaRIA (Université de Picardie), LIP (ENS Lyon), LIFL (Université de Lille).

We participated in sub-project 5 "Verification of distributed

protocols". We investigated the different approaches for fault-tolerance in large-scale systems: checkpointing, replication and consensus, self-stabilization. Two consequences of this study were the adoption of checkpointing for making the MPICH-V distributed library fault-tolerant and the initiation of a new project for realizing a reproducible fault injector.

#### SR2I

Sécurité du Routage Interdomaine dans Internet.

Partners: LRI, PriSM (Université de Versailles), LIRMM (Université de Montpellier)

Project funded by the national research program on computer security (ACI Sécurité Informatique). Scientific director for LRI: Sébastien Tixeuil.

Project funded by the

gram on computer security (ACI Sécurité Informatique).

national research pro-

Scientific director for

LRI: Sébastien Tixeuil.

Project funded by the

national research program for grid computing (*ACI Grid*).

Scientific director for

the group: Ioffrov

Beauquier.

This new project is aimed at better understanding the reliability and security issues in the Internet inter-domain routing

protocol (BGP). This includes studying the self-stabilization property, active and passive tomography to recover all information about the network, statistical models to detect variations in traffic, the construction of a more realistic image of the network topology in order to limit the exchange of messages, to speed up the convergence phase and to improve interdomain routing robustness.

#### FRAGILE

*Failure Resilience and Application Guaranteed Integrity in Largescale Environments.* 

Partners: LRI (Parall and GraFComm groups), INRIA ARES (Lyon), LIAFA (Paris), LPD - EPFL (Switzerland), NEST (University of Iowa).

This new project focuses on software solutions to handle failures in large-scale environments. The archetypes of large-scale

environments we consider are wireless sensor networks, grid networks and peer-to-peer networks.

#### AIRCAST

Plate-forme de réseau sans fil pour l'étude de la gestion des ressources volatiles.

Project funded by the university (*BQR*). Scientific director for LRI: Colette Johnen.

Partners: LRI, Institut d'Electronique Fondamentale (IEF, Université Paris-Sud).

In this project we studied the problems of instability and volatility in a wireless network. We installed an IEEE 802.11 wireless network where users and terminals can be disconnected or move from one access point to the next at any time, therefore creating some instability in the system and the volatility of resources.

#### Algorithmique Répartie

This project addressed two aspects of distributed computing: the modelling of parallel system and its use for verification, and self-stabilizing algorithms for distributed computing with a focus on algorithms with a short stabilizing phase and small loss with respect to their non-stabilizing versions, in fault-free settings.

## B.7.8 / Parall

## Collaborations

## Collaborations leading to joint publications

#### **Foreign countries**

#### France

- University of Nevada Las Vegas, USA (Prof. Ajoy K. Datta).
- University of Iowa, USA (Prof. Ted Herman).
- University of Osaka, Japan (Prof. Toshimitsu Masuzawa).
- University of the Negev, Israel (Prof. Shlomi Dolev).
- Technion, Israel (Prof. Shay Kutten).
- University of Toronto, Canada (Prof. Faith Fich).
- University of Calgary, Canada (Prof. Lisa Higham).

- Université Bordeaux I (Prof. Philippe Duchon and Dr. Nicolas Hanusse).
- Université de Picardie (Prof. Franck Petit and Prof. Vincent Villain).
- Université de Technologie de Compiègne (Prof. Bertrand Ducourthial).
- INSA de Lyon (Prof. Eric Fleury, Dr. Isabelle Guérin-Lassous).
- Université de Reims (Duy-So Nguyen).

## Other collaborations

• Université Denis Diderot (Prof. Carole Delporte, Prof. Hugues Fauconnier).

#### B.7.9 / Parall

## Dissemination and technology transfer

## Software licensing and distribution

**APMC** is a tool dedicated to the approximate verification of probabilistic systems. It uses a Monte-Carlo algorithm to approximate the probability that a temporal formula is true, by using a randomized sampling method.

http://www.lri.fr/~syp/APMC/

#### Funded by ministry of research. Scientific director for LRI: Laurent Rosaz.

Research groups

#### Parall

Contracts and grants Collaborations Dissemination and technology tra fer **MPICH-V** (Message Passing Interface for Volatile resources) is a research effort with theoretical studies, experimental evaluations and pragmatic implementations aiming to provide an MPI implementation based on MPICH, featuring multiple fault tolerant protocols. http://www.lri.fr/~gk/MPICH-V

**Marella** is a model checker for asynchronous reactive systems that uses event structures to derive properties of graph reduction.

http://www.lri.fr/~ambroise/marrella\_site/version\_hdef/home\_1.html

**Verification Platform** compares the performance of several model-checkers using translators.

http://www.lri.fr/~auge

## Popularisation of research results

- National Science fair, *Fête de la science*, 2003: Sylvie Delaët, "Sorting and recursivity", scientific popularisation for schools and the general public.
- Articles in the french scientific magazine La Recherche: Christophe Genolini, Le rêve d'un réseau antibugs (november 2001), Ces problèmes sans solutions (january 2002).

## Summer schools, tutorials, invited seminars

- Spring school *Algorithmique Distribuée* (Porquerolles, May 2003): Joffroy Beauquier, tutorial; Sébastien Cantarell, Sylvie Delaët, Pierre Lemarinier, Denis Quesneau, Laurence Pilard, participants.
- Journées Internationales de l'Auto-stabilisation (Luminy, October 2002) : Joffroy Beauquier, Sylvie Delaët, Colette Johnen, Sébastien Tixeuil, invited seminars.
- Journées de recherche sur l'Automatique et les Télécoms (CNAM, Paris, 2002) : Sébastien Tixeuil, invited tutorial.
- Journée Thématique Outils pour les Systèmes Grande Echelle Dynamiques: Sébastien Tixeuil, invited talk.
- Seminar Informatique Fondamentale (LaBRI, Bordeaux, 2003) : Sébastien Tixeuil, invited seminar.
- Seminar Algorithmes Distribués et Protocoles (IRISA, Rennes, 2000) : Sébastien Tixeuil, invited seminar.
- Self-stabilization Seminar (Dagsthul, Germany, 2000): Joffroy Beauquier, Sylvie Delaët, Maria Gradinariu, Thomas Hérault, Colette Johnen, Sébastien Tixeuil, participants.

## Training and education (doctoral and post-doctoral)

Defended doctorates					
Name	Date defended	Current position			
TIXEUIL Sébastien	14/01/00	Assistant Professor (Paris-Sud)			
GRADINARIU Maria	20/12/00	Assistant Professor (Rennes)			
GENOLINI Christophe	22/12/00	Assistant Professor (Nanterre)			
AMBROISE Dominique	26/10/01	High-school teacher			
AUGE Patrick	23/11/01	industry			
MAGNIETTE Frédéric	19/06/02	system administrator			
HERAULT Thomas	28/05/03	Assistant Professor (Paris-Sud)			
CANTARELL Sébastien	8/12/03	industry			

Six students started a Ph.D. since 2000:

- Denis Quesneau in september 2001.
- Pierre Lemarinier and Laurence Pilard in september 2002.
- Sammy Haddad in september 2003.
- William Hoarau and Xiaolong Liu in september 2004.

#### **Graduate courses**

• DEA *Informatique Distribuée*: Joffroy Beauquier (Distributed Algorithms) and Brigitte Rozoy (Model and Verification of Asynchronous Systems).

Research groups

## Parall

Collaborations Training and education

## Publications

## International peer-reviewed journals

- J. Beauquier, B. Bérard, L. Fribourg, and F. Magniette. Proving convergence of self-stabilizing systems using first-order rewriting and regular languages. *Distributed Computing*, 14(2):83-95, 2001.
- [2] J. Beauquier, S. Cantarell, A. Datta, and F. Petit. Group mutual exclusion in tree networks. *Journal of Information Science and Engineering*, 19(3):415-432, 2003.
- [3] J. Beauquier, M. Gradinariu, C. Johnen, and J. Durand-Lose. Token-based self-stabilization uniform algorithms. *Journal of Parallel and Distributed Computing*, 62(5):899-921, May 2002.
- [4] J. Beauquier, L. Pilard, and B. Rozoy. Observing locally self-stabilization. *Journal of High Speed networks*, page to appear, 2004.
- [5] I. Biermann, L. Rosaz, and B. Rozoy. Occurence preserving congruences and associated graphs. *Rairo Informatique théorique et applications*, page to appear, 2004.
- [6] A. Bouteiller, P. Lemarinier, G. Krawezik, and F. Cappello. Coordinated checkpoint versus message log for fault tolerant MPI. *Int. Journal of High Performance Computing and Networking (IJHPCN)*, (3), September 2004.
- [7] F. Cappello, S. Djilali, G. Fedak, T. Herault, O. Lodygenski, F. Magniette, and V. Neri. Computing on large scale distributed systems: experience of the xtremweb project. Special issue on Peer-to-Peer Computing and Interaction with Grids of Future Generation Computer Systems, 2004. to appear.
- [8] Y. Chen, A. K. Datta, and S. Tixeuil. Stabilizing inter-domain routing in the internet. *Journal of High Speed Networks*, page to appear, 2004.
- [9] A. Datta, J. Derby, J. Lawrence, and S. Tixeuil. Stabilizing hierarchical routing. *Journal* of *Interconnexion Networks*, 1(4):283-302, 2000.
- [10] A. Datta, C. Johnen, F. Petit, and V. Villain. Self-stabilizing depth-first token circulation in arbitrary rooted networks. *Distributed Computing*, 13(4):207-218, 2000.
- [11] A. Datta and S. Tixeuil. Self-stabilizing sorting on tree networks. *Parallel Algorithms and Applications*, 16(1):1-15, January 2001.
- [12] A. K. Datta, M. Gradinariu, A. B. Kenitzky, and S. Tixeuil. Self-stabilizing wormhole routing on ring networks. *Journal of Information Science and Engineering*, 19:401-414, 2003. Extended abstract in IEEE ICPADS 2002, best paper award.
- [13] A. K. Datta, M. Gradinariu, and S. Tixeuil. Self-stabilizing mutual exclusion with arbitrary scheduler. *The Computer Journal*, 47(3):289-298, 2004.
- [14] S. Delaët and S. Tixeuil. Tolerating transient and intermittent failures. *Journal of Parallel and Distributed Computing*, 62(5):961-981, May 2002.
- [15] B. Ducourthial and S. Tixeuil. Self-stabilization with r-operators. *Distributed Computing*, 14(3):147-162, July 2001.
- [16] B. Ducourthial and S. Tixeuil. Self-stabilization with path algebra. *Theoretical Computer Science*, 293(1):219-236, 2003. Extended abstract in Sirrocco 2000.
- [17] C. Johnen, L. Alima, A. K. Datta, and S. Tixeuil. Optimal snap-stabilizing neighborhood synchronizer in tree networks. *Parallel Processing Letters*, 12(3-4):327-340, 2002.
- [18] F. Magniette, L. Pilard, and B. Rozoy. A method for the verification of distributed and synchronized algorithm. *International Journal of Production Research*, page to appear, 2004.

[19] L. Rosaz. The word problem for 1lc-congruences is np-hard. *Theoretical Computer Science*, 306:245-268, 2003.

## National peer-reviewed journals

- [20] C. Johnen, F. Petit, and S. Tixeuil. Auto-stabilisation et protocoles réseaux. *Technique et Science Informatiques*, page to appear, 2004.
- [21] I. Lavallée, C. Lavault, and C. Johnen. Reliable self-stabilizing communication for quasi rendezvous. *Studia Informatica*, 1(1):59-76, 2002, Suger.

## **Edited books**

- [22] F. Cassez, C. Jard, B. Rozoy, and M. Ryan, editors. *MOdelling and VErification of Parallel processes*, Lecture Notes in Computer Science, Nantes, France, June 2002. Springer-Verlag.
- [23] S. Delaët, T. Hérault, C. Johnen, and S. Tixeuil, editors. *Actes de la première journée Réseaux et Algorithmes répartis*, volume 1325, Orsay, France, June 2002. LRI, Université Paris Sud.

#### Major international peer-reviewed conferences

- [24] D. Ambroise, P. Augé, K. Bouchefra, and B. Rozoy. Marrella and the verification of an embedded system. In M. M. Furio Honsell, editor, *Foundations of Software Science and Computation Structures, 4th International Conference, FOSSACS 2001 Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2001 Genova, Italy*, number 2030 in Lecture Notes in Computer Science, pages 409-412. Springer-Verlag, April 2001.
- [25] D. Ambroise, B. Rozoy, and J. Saquet. Deadlock detextion in distributed systems. In *Proceedings of the 18th International Conference on Computers And Their Applications (ISCA'2003)*, pages 210-213, Honolulu, Hawaii, USA, March.
- [26] P. Augé, K. Bouchafra, T. Maurin, R. Reynaud, and B. Rozoy. Simulation and verification of an embedded system. In *Proceedings of the ISCA 16th International Conference on Computers And Their Applications*, Seattle, March 2001.
- [27] J. Beauquier, A. Datta, M. Gradinariu, and F. Magniette. Self-stabilizing local mutual exclusion and daemon refinement. In *Proceedings of DISC'2000*, Lecture Notes in Computer Science, pages 223-237. Springer-Verlag, 2000.
- [28] J. Beauquier, M. Gradinariu, and C.Johnen. Cross-over composition enforcement of fairness under unfair adversary. In *Proceedings of Self-Stabilizing Systems, 5th International Workshop (WSS'2001)*, number 2194 in Lecture Notes in Computer Science, pages 19-34, Lisbon, Portugal, October 2001. Springer.
- [29] J. Beauquier and T. Herault. Fault local stabilization : the shortest path tree. In *Proceedings of the 21st Symposium on Reliable Distributed Systems (SRDS'02)*, pages 62-69, Osaka, Japan, October 2002.
- [30] J. Beauquier, T. Herault, and E. Schiller. Easy self-stabilization with an agent. In *Proceedings of the 5th Internation Workshop on Self-Stabilizing systems (WSS'01)*, volume 2194 of *LNCS*, pages 25-50, Lisbon, Portugal, October 2001.

Research groups

Publications

- [31] G. Bosilca, A. Bouteiller, F. Cappello, S. Djilali, G. Fedak, C. Germain, T. Herault, P. Lemarinier, O. Lodygensky, F. Magniette, V. Neri, and A. Selikhov. MPICH-V: Toward a scalable fault tolerant MPI for volatile nodes. In *Proceedings of the 15th High Performance Networking and Computing conference (SC'02)*, page electronic, Baltimore, USA, November 2002. IEEE/ACM.
- [32] A. Bouteiller, H.-L. Bouziane, T. Herault, P. Lemarinier, and F. Cappello. Hybrid preemptive scheduling of MPI applications on the grids. In *Proceedings of the 5th International Workshop on Grid Computing (Grid'04)*, Pittsburgh, USA, November 2004. IEEE/ACM. to appear.
- [33] A. Bouteiller, F. Cappello, T. Herault, G. Krawezik, P. Lemarinier, and F. Magniette. Mpichv2: a fault tolerant mpi for volatile nodes based on pessimistic sender based message logging. In *Proceedings of the 16th High Performance Networking and Computing conference (SC'03)*, page electronic, Phoenix, USA, November 2003. IEEE/ACM.
- [34] A. Bouteiller, P. Lemarinier, G. Krawezik, and F. Cappello. Coordinated checkpoint versus message log for fault tolerant MPI. In *IEEE International Conference on Cluster Computing (Cluster 2003)*, pages 242-250. IEEE Press, december 2003.
- [35] S. Cantarell, A. Datta, and F. Petit. Self-stabilizing atomicity refinement allowing neighborhood concurrency. In *Proceedings of the Sixth Symposium on Self-Stabilizing Systems* (SSS'03).
- [36] Y. Chen, A. Datta, and S. Tixeuil. Self-stabilizing inter-domain routing in the internet. In S. Verlag, editor, *Euro-Par'2002*, pages 749-752, Paderborn, Germany, August 2002.
- [37] A. Datta, M. Gradinariu, A. Kenitzki, and S. Tixeuil. Self-stabilizing wormhole routing in ring networks. In *Proceedings of the IEEE International Conference on Parallel and Distributed Systems (ICPADS'2002)*, Taiwan, ROC, December 2002. Best paper award.
- [38] A. Datta, M. Gradinariu, and S. Tixeuil. Self-stabilizing mutual exclusion using unfair distributed scheduler. In *IEEE International Parallel and Distributed Processing SYmposium* (*IPDPS'2000*), pages 465-470, Cancun, Mexico, May 2000. IEEE Press.
- [**39**] S. Delaët, D.-S. Nguyen, and S. Tixeuil. Stabilité et auto-stabilisation du routage interdomaine dans internet. In *Proceedings of RIVF 2003*, pages 139-144, Hanoi, Vietnam, february 2003. Studia Informatica Universalis.
- [40] S. Delaët and S. Tixeuil. Tolerating transient and intermittent failures. In *International Conference on Principles of Distributed Systems (OPODIS'2000)*, pages 17-36, Paris, France, December 2000.
- [41] S. Djilali, T. Herault, O. Lodygensky, T. Morlier, G. Fedak, and F. Cappello. RPC-V: Toward fault-tolerant rpc for internet connected desktop grids with volatile nodes. In *Proceedings* of the 17th High Performance Networking and Computing conference (SC'04), Pittsburgh, USA, November 2004. IEEE/ACM. to appear.
- [42] P. Duchon, N. Hanusse, and S. Tixeuil. Optimal randomized self-stabilizing mutual exclusion in synchronous rings. In *Proceedings of the 18th Symposium on Distributed Computing* (*DISC(2004)*, page to appear. Springer Verlag, October 2004.
- [43] B. Ducourthial and S. Tixeuil. Self-stabilization with path algebra. In *International Colloquium on Structural Information and Communication Complexity(Sirroco'2000)*, pages 95-110, L'Aquila, Italy, June 2000. Carleton University Press.
- [44] M. Duflot, L. Fribourg, T. Herault, R. Lassaigne, F. Magniette, S. Messika, S. Peyronnet, and C. Picarrony. Verification of the CSMA/CD protocol using PRISM and APMC. In *Proceedings of the 4th Internation Workshop on Automated Verification of Critical Systems* (AVoCS'04), Electronic Notes in Theoretical Computer Science, London, UK, September 2004. Elsevier Science Publishers. to appear.
- [45] F. E. Fich and C. Johnen. A space optimal, deterministic, self-stabilizing, leader election algorithm for unidirectional rings. In *Distributed Computing 15th International Symposium, Springer LNCS:2180*, pages 224-239, 2001.

- [46] C. Genolini. Optimal k-stabilization: the case of synchronous mutual exclusion. In Proceedings of Parallel and Distributed Computing Systems (PDCS'2000), pages 371-376, November 2000.
- [47] C. Genolini and S. Tixeuil. A lower bound on k-stabilization in asynchronous systems. In Proceedings of IEEE 21st Symposium on Reliable Distributed Systems (SRDS'2002), Osaka, Japan, October 2002.
- [48] M. Gradinariu and C. Johnen. Self-stabilizing neighborhood unique naming under unfair scheduler. In *Euro-Par'01 Parallel Processing, Proceedings LNCS:2150*, pages 458-465, 2001.
- [49] M. Gradinariu and S. Tixeuil. Self-stabilizing vertex coloring of arbitrary graphs. In *International Conference on Principles of Distributed Systems (OPODIS'2000),* pages 55-70, Paris, France, December 2000.
- [50] M. Gradinariu and S. Tixeuil. Tight space uniform self-stabilizing l-mutual exclusion. In *IEEE International Conference on Distributed Computing Systems (ICDCS'01)*, pages 83-90, Phoenix, Arizona, May 2001. IEEE Press.
- [51] T. Herault, R. Lassaigne, F. Magniette, and S. Peyronnet. Apmc: Approximate probabilistic model checker. In *Proceedings of the 5th Verification Model Checking and Abstract Interpretation (VMCAI'04) conference*, Venice, Italy, January 2004.
- [52] T. Herman and S. Tixeuil. A distributed tdma slot assignment algorithm for wireless sensor networks. In *Proceedings of the First Workshop on Algorithmic Aspects of Wireless Sensor Networks (AlgoSensors'2004)*, number 3121 in Lecture Notes in Computer Science, pages 45-58, Turku, Finland, July 2004. Springer-Verlag.
- [53] C. Johnen. Service time optimal self-stabilizing token circulation protocol on anonymous unidrectional. In 21st Symposium on Reliable Distributed Systems, pages 80-89. IEEE Computer Society Press, 2002.
- [54] C. Johnen. Bounded service time and memory space optimal self-stabilizing token circulation protocol on unidirectional rings. In I. C. S. Press, editor, *IEEE 18th International Parallel and Distributed Processing Symposium*, 2004.
- [55] C. Johnen and S. Tixeuil. Route preserving stabilization. In *Proceedings of the Sixth Symposium on Self-stabilizing Systems (SSS'03)*, Lecture Notes in Computer Science, San Francisco, USA, june 2003. Springer Verlag. Also in the Proceedings of DSN'03 as a one page abstract.
- [56] I. Lavallée, C. Lavault, and C. Johnen. Fair and reliable self-stabilizing communication. In 4<sup>th</sup> International Conference On Principles Of DIstributed Systems, pages 163-176, 2000.
- [57] P. Lemarinier, A. Bouteiller, T. Herault, G. Krawezik, and F. Cappello. Improved message logging versus improved coordinated checkpointing for fault tolerant MPI. In *Proceedings* of the 6<sup>th</sup> international conference on Cluster Computing (CLUSTER'04), San Diego, USA, September 2004. IEEE. to appear.
- [58] L. Rosaz. Self-stabilizing token circulation on asynchronous uniform unidirectional rings. In Proceedings of the Nineteenth Annual ACM Symposium on Principles of Distributed Computing (PODC'2000), pages 249-258, Portland, Oregon, USA, July 2000. ACM.
- [59] S. Tixeuil. On a space-optimal distributed traversal algorithm. In S. Verlag, editor, *Fifth Workshop on Self-stabilizing Systems (WSS'2001)*, volume LNCS 2194, pages 216-228, Lisbon, Portugal, October 2001.

## National peer-reviewed conferences

[60] D. Barth, C. Johnen, and V.Vèque. Routage et reroutage dans un réseau mobile. In 2<sup>es</sup> *Rencontres Francophones sur les aspects Algorithmiques des Télécommunications*, 2000.

Research groups

Publications

- [61] G. Bosilca, G. Fédak, T. Herault, and F. Magniette. Évaluation de performances de différentes techniques de confinement d'exécutions pour le calcul pair-à-pair. In *Actes des Rencontres Francophone du parallélisme (RenPar'02)*, Hammamet, Tunisia, 2002.
- [62] S. Delaët, D.-S. Nguyen, and S. Tixeuil. Stabilité et auto-stabilisation de bgp. In *Proceedings* of *Algotel 2003*, Banyuls, France, may 2003. INRIA.
- [63] P. Duchon, N. Hanusse, and S. Tixeuil. Protocoles auto-stabilisants synchrones d'exclusion mutuelle pour les anneaux anonymes. In *Proceedings of Rencontres Francophones sur l'Algorithmique des Communications (Algotel'2004)*, Batz sur Mer, France, 2004. ENST Bretagne.
- [64] B. Ducourthial and S. Tixeuil. Self-stabilizing construction of optimal broadcast forests. In *Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel'2000)*, pages 149-154, La Rochelle, France, May 2000. in French.
- [65] B. Ducourthial and S. Tixeuil. Adaptive multi-sourced multicast. In *Rencontres Francophones sur les aspects Algorithmiques des Télécommunications (AlgoTel'2001)*, pages 135-142, St-Jean de Luz, France, May 2001. in French.
- [66] T. Herault and P. Lemarinier. A rollback-recovery protocol on peer-to-peer systems. In Proceedings of the MOVEP'2002 international summer school, pages 313-319, Nantes, France, June 2002.
- [67] T. Herman and S. Tixeuil. Un algorithme tdma réparti pour les réseaux de capteurs. In *Proceedings of Rencontres Francophones sur l'Algorithmique des Communications (Algotel'2004)*, Batz sur Mer, France, 2004. ENST Bretagne.
- [68] C. Johnen. Service time of self-stabilizing token circulation protocol on anonymous unidirectional rings - extended abstract. In *Proceedings of the Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (Algotel'2003).* INRIA, 2003.
- [69] F. Magniette, L. Pilard, and B. Rozoy. Model-checking et produit synchronisé. In *Proceedings of Modélisation des Systèmes Réactifs (MSR'2003)*, 2003.
- [70] L. Pilard and F. Magniette. A method for the verification of distributed and synchronized algorithms. pages 384-390. MOVEP, May 2002.

## **Doctoral dissertations and Habilitations**

- [71] S. Cantarell. *Group Mutual Exclusion and Self-Stabilization*. PhD thesis, University of Paris Sud XI, December 2003.
- [72] C. Genolini. *Raffinements de l'Auto-stabilisation*. PhD thesis, University of Paris Sud XI, December 2000.
- [73] M. Gradinariu. *Modélisation, vérification et Raffinement des Algorithmes Auto-stabilisants.* PhD thesis, University of Paris Sud XI, December 2000.
- [74] T. Hérault. *Mending of transient failure in self-stabilizing systems*. PhD thesis, University of Paris Sud XI, June 2003.
- [75] S. Tixeuil. Auto-stabilisation Efficace. PhD thesis, University of Paris Sud XI, January 2000.

#### Reports

[76] J. Beauquier and C. Johnen. Analyze of randomized self-stabilizing algorithms under nondeterministic scheduler classes. Technical Report 1327, University of Paris Sud, July 2002.

- [77] J. Beauquier, L. Pilard, and B. Rozoy. Observing locally self-stabilization. Technical Report 1387, University of Paris Sud XI, May 2004.
- [78] Y. Chen, A. K. Datta, and S. Tixeuil. Stabilizing inter-domain routing in the internet. Technical Report 1302, Laboratoire de Recherche en Informatique, Université Paris Sud XI, 2002.
- [79] A. K. Datta, J. L. Derby, J. E. Lawrence, and S. Tixeuil. Stabilizing hierarchical routing. Technical Report 1244, University of Paris Sud XI, Laboratoire de Recherche en Informatique, February 2000.
- [80] A. K. Datta, M. Gradinariu, A. B. Kenitzki, and S. Tixeuil. Self-stabilizing wormhole routing on ring networks. Technical Report 1324, Laboratoire de Recherche en Informatique, Université Paris Sud XI, 2002.
- [81] A. K. Datta, M. Gradinariu, and S. Tixeuil. Self-stabilizing mutual exclusion using unfair distributed scheduler. Technical Report 1227, University of Paris Sud XI, Laboratoire de Recherche en Informatique, February 2000.
- [82] S. Delaët, B. Ducourthial, and S. Tixeuil. Self-stabilization with r-operators in unreliable directed networks. Technical Report 1361, Laboratoire de Recherche en Informatique, 2003.
- [83] S. Delaët and S. Tixeuil. Tolerating transient and intermittent failures. Technical Report 1246, University of Paris Sud XI, Laboratoire de Recherche en Informatique, March 2000.
- [84] P. Duchon, N. Hanusse, and S. Tixeuil. Optimal self-stabilizing mutual exclusion on synchronous rings. Technical Report 1389, University of Paris Sud XI, Laboratoire de Recherche en Informatique, June 2004.
- [85] B. Ducourthial and S. Tixeuil. Self-stabilization with r-operators. Technical Report 2001/358, Université de Technologie de Compiègne, HEUDIASYC, 2001.
- [86] C. Genolini and S. Tixeuil. Reactive k-stabilization and time adaptivity: possibility and impossibility results. Technical Report 1276, Laboratoire de Recherche en Informatique, University of Paris Sud XI, 2001.
- [87] M. Gradinariu and S. Tixeuil. Self-stabilizing vertex coloring of arbitrary graphs. Technical Report 1260, Laboratoire de Recherche en Informatique, University of Paris Sud XI, September 2000.
- [88] M. Gradinariu and S. Tixeuil. Tight space uniform self-stabilizing l-mutual exclusion. Technical Report 1249, University of Paris Sud XI, Laboratoire de Recherche en Informatique, March 2000.
- [89] T. Herman and S. Tixeuil. A distributed tdma slot assignment algorithm for wireless sensor networks. Technical Report 1370, Laboratoire de Recherche en Informatique, 2003.
- [90] C. Johnen, F. Petit, and S. Tixeuil. Auto-stabilisation et protocoles réseau. Technical Report 1357, Laboratoire de Recherche en Informatique, 2003.
- [91] C. Johnen and S. Tixeuil. Route preserving stabilization. Technical Report 1353, Laboratoire de Recherche en Informatique, 2003.
- [92] N. Mitton, E. Fleury, I. Guérin-Lassous, and S. Tixeuil. Density-driven clustering in wireless multi-hop networks. Technical report, INRIA, 2004.

## **Other publications**

[93] A. Bouteiller, P. Lemarinier, and F. Cappello. MPICH-V3 preview: A hierarchical fault tolerant MPI for multi-cluster grids. IEEE/ACM High Performance Networking and Computing (SC 2003), poster session, November 2003. Phoenix USA. Research groups

Publications

- [94] S. Haddad. Augmentation du champ d'utilisation de l'auto-stabilisation. Master's thesis, University of Paris Sud XI, 2003.
- [95] W. Hoarau. Fail fault injection language. Master's thesis, University of Paris Sud XI, 2004. http://www.lri.fr/hoarau/fail.html.
- [96] P. Lemarinier. Défaillance et calcul global x-trem web. Master's thesis, University of Paris Sud, June 2002.
- [97] L. Pilard. Une méthode pour la vérification d'algorithmes répartis synchronisés. Master's thesis, University of Paris Sud XI, July 2002.



Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Intelligence Artificielle et Systèmes d'Inférence

#### **Responsable : Marie-Christine Rousset**

'équipe IASI a travaillé de nombreuses années sur la conception et la construction de systèmes à base de connaissances. Les problèmes qui ont été plus particulièrement étudiés portent sur la représentation déclarative et logique de connaissances complexes et sur la conception d'algorithmes d'inférences permettant de mécaniser des raisonnements sur ces connaissances.

Au cours des 4 dernières années, l'équipe a focalisé ses travaux sur le problème de l'intégration d'informations distribuées et hétérogènes, devenu un enjeu crucial pour l'exploitation de l'immense gisement d'informations rendu



accessible par le Web. Cette nouvelle orientation s'est traduite par la création du projet Gemo commun entre le LRI et l'UR Futurs de l'INRIA. Gemo est un des trois projets fondateurs du PCRI et est un des rares groupes de recherche en France et en Europe réunissant des chercheurs en Intelligence Artificielle et en Bases de Données.

L'équipe IASI a développé une collaboration étroite et suivie avec France Telecom R&D à travers la série de projets PICSEL qui porte sur la mise en oeuvre d'un environnement générique et déclaratif pour la construction de portails sémantiques. IASI participe activement à l'effort collectif de recherche internationale vers un Web Sémantique pour une meilleure exploitation du Web grâce à un marquage sémantique des documents, des données et des ressources à l'aide d'ontologies.



# Artificial Intelligence and Inference Systems

Head: Marie-Christine Rousset

he IASI group has a long tradition of research on inference algorithms and knowledge representation as the formal basis for the construction of knowledge-based systems.

In the last 4 years, the group has focused its work on the new and hot topic of information integration. This new orientation has resulted in the creation of the Gemo PCRI project, which was born from the merging of IASI and the INRIA-



Rocquencourt Verso project, leading to one of the first artificial intelligence and database groups in France.

IASI has developed a very tight and productive collaboration with France Telecom R&D through the PICSEL series of projects, which aim at building a generic and declarative environment for the construction of semantic portals. IASI participates actively in the world wide collective initiative towards a Semantic Web in which data and resources will be better retrieved thanks to a semantic mark-up based on ontologies. Research groups

## Research Group Members

Personnel as of 01/01/2004.

Full time faculty			
Name	First Name	Position*	Institution
CHATALIC	Philippe	MC	Université Paris XI
GAGLIARDI	Hélène	MC	Université Paris XI
GOASDOUE	François	MC	Université Paris XI
HAEMMERLE	Ollivier	MC	INAPG
PERNELLE-MANSCOUR	Nathalie	MC	IUT Sceaux
REYNAUD	Chantal	PR2	Université Paris X-Nanterre
ROUSSET	Marie-Christine	PR1	Université Paris XI
SAFAR	Brigitte	MC	Université Paris XI
SCHLIENGER	Daniel	MCHC	IUT de Sceaux
SIMON	Laurent	MC	Université Paris XI
VENTOS	Véronique	MC	Université Paris XI
Doctoral students			
ADJIMAN	Philippe	AM	IUT Orsay
GIRALDO	Gloria	D	Université Paris XI
KEFI	Hassen	D	Université Paris XI
MEZAOUR	Amar	А	Université Paris XI
SUAREZ BARON	Marco Javier	D	Bourse ALPHA CORDIAL
TERMIER	Alexandre	ATER	Université Paris XI
Temporary personnel			
AYAT	Nedjem Eddine	Post-doc	CNRS
Non-LRI personnel of	PCRI project Gem	10	
ABITEBOUL	Serge	DR	INRIA
MANOLESCU	loana	CR	INRIA
SEGOUFIN	Luc	CR	INRIA
BENJELLOUN	Omar	D	
NGUYEN	Benjamin	D	
RUBERG	Gabriela	D	
SAMUELIDES	Mathias	D	
AMANN	Bernd	on contract	INRIA
BAUMGARTEN	Jérôme	Engineer	INRIA
RUBERG	Nicolaas	Engineer	

\* See the glossary for acronyms.

Long term visitor								
Name	First Name	Nationality	Institution	Arrival	Departure	Funding		
BARFOROUSH	Ahmad	Iranian	Teheran Univ.	16/02/01	31/12/01			
#### Group evolution

Since 2000, the following faculty members got a permanent position in our group:

- Nathalie Pernelle, post-doc in our group from 2000 to 2001, Assistant Professor since September 2001.
- Laurent Simon, Lecturer (ATER) from September 2001 to August 2002, Assistant Professor since September 2002.
- François Goasdoué, Lecturer (ATER) from September 2002 to August 2003, Assistant Professor since September 2003.
- Olivier Haemmerlé (Associate Professor at INA-PG) is on a 2-year visiting position since May 2003.

The following people left our group in September 2002 to join the new BioInfo group:

- Christine Froidevaux (Professor)
- Jean-Paul Forest (Ph.D. student)

#### B.8.2 / IASI

## **Research Description**

#### **Overview**

The IASI group has a long tradition of work on inference algorithms and knowledge representation as the formal basis for the construction of knowledge-based systems. The knowledge representation formalisms that are studied from a computational and algorithmic point of view are different fragments of propositional or first-order logic that are useful for modelling real-world applications. In the last 4 years, while pursuing our previous work on inference algorithms for logical formalisms, we have focused our work on knowledge representation for information integration.

Information integration is a hot topic both from a practical and theoretical viewpoint. The explosion of the number of information sources that are now available through the Web raises new needs and new problems for building information integration systems over distributed data sources. An information integration system provides a uniform interface for querying collections of pre-existing data sources that were created independently. It is based on a single mediated schema in terms of which users' queries are issued, and the data sources to integrate are described. The descriptions specify semantic relationships between the contents of data sources and the mediated schema. They are exploited by the query processor of the information integration system which must reformulate the original query into queries against the source schemas. Information integration systems must deal with large and constantly changing collections of data sources. This requires powerful languages and flexible mechanisms for describing and handling data sources which may have overlapping or contradictory contents, semantic mismatches, limited capabilities, etc. Knowledge representation techniques are good candidates for building mediated systems because they provide declarative knowledge representation formalisms enabling a precise and formal modelling of the semantic relationships between source schemas through a mediated schema, together with generic inference algorithms for query reformulation.

Research groups

IASI

In our group, we are studying the problem of information integration at different scales, the key challenge being then to identify formalisms offering a reasonable trade-off between expressive power and good computational properties for the corresponding reformulation algorithm. In particular, in the context of the PICSEL series of projects funded by France Telecom R&D, we have developed a generic and declarative environment for building semantic portals, based on a logical formalism combining the expressive power of rules and classes. We have recently started working on the Semantic Web for which we are investigating the problem of semantic mediation in a peer-to-peer setting.

#### PCRI Project Gemo



Gemo's main theme is the integration of distributed and heterogeneous information, where information is seen as a general concept covering data, knowledge and services. More precisely, the problems addressed in Gemo deal with discovering meaningful data sources or services, understanding their content or goal, integrating them, and finally monitoring their evolution over time. Our aim is to offer environments that are both powerful and flexible in order to simplify the deployment of applications that give fast access to meaningful data distributed over the network. The creation of data warehouses and mediators offering a wide access to multiple heterogeneous sources provides a good means of achieving these goals.

Addressing these new problems requires combining Artificial Intelligence techniques, such as reasoning on knowledge representation formalisms, classification or data mining, and Database techniques such as indexing or query processing optimizations.

Gemo was born from the merging of the INRIA-Rocquencourt Verso project with members of the IASI group working on knowledge representation for information integration. It became a joint project of LRI and INRIA Futurs within the PCRI in 2002.

## Theme 1: Knowledge Representation and Algorithmic

• Participants: Philippe Chatalic (leader), Philippe Adjiman, François Goasdoué, Marie-Christine Rousset, Laurent Simon.

Automated reasoning requires the choice of a given knowledge representation formalism and the design of associated inference algorithms, the complexity of the latter being strictly related to the expressive power of the former. Studying formal properties is essential, in order to circumscribe the class of problems that can be addressed with a given formalism/technique. Studying practical aspects of algorithms is also important and a strong accent is put on experimental evaluation of the designed algorithms.

#### Inference algorithms for propositional logic

The poor expressive power of propositional logic is counterbalanced by the efficiency of their associated inference algorithms. Many industrial problems are addressed today with such techniques, thus making this framework particularly challenging for designing new methods of reasoning.

#### Compressed formulas for propositional reasoning procedures

While most SAT solvers are based on model elimination techniques (DPLL-like algorithms), variable elimination (DP-like) can also prove to be useful. However, propositional encoding of problems often leads to large sets of clauses. A key point is therefore the ability to represent and handle clause sets efficiently, as pointed out by the first real experimental study of the DP algorithm that we have conducted [80, 101]. We have introduced the concept of compressed sets of clauses, using binary decision diagram like data structures, and defined the multi-resolution inference rule [43, 6]. Our implemented solver ZRes [44] allowed handling exponentially large formulae efficiently. Particularly, it was able to handle the pigeon-holes and Urquhart's problems, two problems known to be untractable for classical resolution, improving the size of solved instances by an order of magnitude.

#### Knowledge base compilation

We have also investigated the use of multi-resolution and compressed sets of clauses in the context of knowledge base compilation, e.g. prime implicates. New target languages have been introduced, a number of specialized algorithms have been proposed and studied, both from a theoretical and an extensive experimental point of view [68, 101]. This work has clearly shown the first and significant scaling-up, in terms of the size of formulae, for prime implicates computation since the introduction of these resolution-based algorithms as well as other compilation techniques.

#### Distributed reasoning in a peer-to-peer setting

Peer-to-peer systems are increasingly used to design scalable and flexible distributed applications over the Internet, without any centralized control. We have investigated peer-to-peer inference systems in which each peer can answer queries by reasoning from its local theory but also by asking queries to some other peers with which it is semantically related by sharing part of its vocabulary. We have proposed the first distributed consequence-find-ing algorithm. It is incremental, anytime, and is guaranteed to be complete if we wait until termination. Its experimentation on random generated theories, over several hundreds of peers looks very promising [34, 77].

#### Experimental comparison of SAT solvers

Initially motivated by the need of testing and comparing our own solvers with state-ofthe-art SAT solvers in a satisfactory way, our work has evolved to stress the need and importance of fair and extensive experimental studies of algorithms in artificial intelligence. We have identified good criteria for such comparisons and designed a new framework, called SATEX [67, 81, 66], that publishes the experimental comparison of SAT algorithms on a web site (Figure 1). This tool rapidly became the worldwide reference on this topic. As a result, we have been selected for organizing the SAT contest in 2002, 2003 and 2004, whose results are similarly published online [17, 12, 36, 37].



*Figure 1: The SAT-Ex site* (*http://www.lri.fr/~simon/satex*).

Research groups

**Research description** 

#### Inference algorithms for the CARIN fragment of first order logic

CARIN combines in an expressive and uniform logical setting two decidable fragments of first order logic: function-free Horn rules (logical foundations of Datalog) and descriptions logics (logical foundations of the forthcoming Semantic Web with the Ontology Web Language recommended by the W3C). CARIN was initiated by joint work between Marie-Christine Rousset and Alon Levy when he was at AT&T Bell Laboratories (USA).

#### Compilation and approximation

In our previous work on CARIN, we have studied the complementarity of Horn rules (without function) and description logics. By studying the compilation and approximation of Horn rules by concept descriptions, we have investigated the logical overlapping between those two fragments of first order logic. Our work [48] is the first one on the compilation and approximation of Horn rules by concept descriptions. Our results enable transferring complexity results and algorithms between description logics and Horn rules or relational database theory. This work has been the key for proving the completeness of the query-rewriting algorithm presented below.

#### Query rewriting

Query rewriting using views is the core algorithmic problem in information integration where users queries are formulated in terms of a virtual mediated schema while the data resides in distant sources described as views over the mediated schema. Queries, views and the mediated schema, are in fact logical formulas and the problem of query rewriting using views given a certain mediated schema amounts to computing (prime) implicants of a logical theory in the corresponding fragment of first order logic. We have studied this problem in the setting of CARIN, which enabled us to offer a uniform logical framework encompassing and relating previous work on query answering and rewriting. In particular, in [10, 100], we were the first to exhibit a complete algorithm for rewriting conjunctive queries using views defined with respect to an ontology expressed in description logics. This theoretical work has been applied in the setting of the PICSEL project. The Java implementation of our query-rewriting algorithm led to the OntoQuery software, which has been deposited by France Telecom.

#### **Key references**

- [6] P. Chatalic and L. Simon. Multiresolution for SAT checking. *Journal of Artificial Intelligence Tools*, 10(4), December 2001.
- [10] F. Goasdoué and M.-C. Rousset. Answering queries using views: a krdb perspective for the semantic web. In ACM Journal - Transactions on Internet Technology (TOIT), volume 4, pages 255-288, 2004.
- [48] F. Goasdoué and M.-C. Rousset. Compilation and approximation of conjunctive queries by concept descriptions. In *European Conference on Artificial Intelligence (ECAI)*, pages 267-271, Lyon, 2002.
- [67] L. Simon and P. Chatalic. SATEX: a Web-based framework for SAT experimentation. In H. Kautz and B. Selman, editors, *Proceedings of the Workshop on Theory and Applications of Satisfiability Testing SAT2001*, June 14th-15th 2001.
- [68] L. Simon and A. del Val. Efficient consequence finding. In 17th International Joint Conference on Artificial Intelligence IJCAI'01, pages 359-365, Seattle, Washington, USA, 2001.
- [77] P. Adjiman, P. Chatalic, F. Goasdoué, M.-C. Rousset, and L. Simon. Raisonnement distribué dans un environnement de type pair-à-pair. In Actes des Dixièmes Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets (JNPC), 2004.

## Theme 2: Information extraction from Semi-structured Data

• Participants: Marie-Christine Rousset (leader), Alexandre Termier, Gloria Giraldo, Nathalie Pernelle, Chantal Reynaud, Véronique Ventos

More and more available data is semi-structured, meaning that it does not conform to a rigid pre-established schema. We have investigated different information extraction tasks in the setting of different projects (Xyleme, GAEL, PICSEL). The common denominator of our work is to target the extraction from (semi-structured, irregular, numerous) data of an abstract view better suited to interrogation by end-users.

#### Extraction of frequent trees from an heterogeneous XML collection

Note: joint work with Michele Sebag, I&A group.

With the rapid growth of XML data available on-line, discovering frequent tree structures in huge collections of tree data becomes a crucial issue for information extraction. We have proposed two novel algorithms for discovering frequent trees. Their common distinguishing feature is to handle a type of tree inclusion which is more general than those considered so far in the tree mining literature, thus leading to the discovery of frequent trees even in the presence of variations in the nesting of the labels of XML documents. These two algorithms differ in their way of coping with complexity issues. The solution experimented in the TreeFinder algorithm [71] is to approximate the tree representation by a flat boolean representation of the trees, coding the presence or absence of all possible label pairs (ancestor relations) in each tree. The counterpart of this simplification is that TreeFinder is not guaranteed to be complete. The solution experimented in the Dryade algorithm [72, 106] is to compute closed frequent trees, which has the advantage of providing a compact representation of frequent trees without loss of information. The performance of TreeFinder and Dryade have been experimentally tested on artificial medium-size data and also on real data in the setting of the Xyleme project [1].

#### Flexible clustering with Galois lattices

In many applications, it becomes crucial to help users to interactively access huge amounts of data. One way consists in clustering data in classes organized in a hierarchy and described at an appropriate level of abstraction. We have defined a framework based on Galois lattices for dealing with this problem. Our goal was to bypass the main drawback of the large size of Galois lattices when applied to real-world data. We have introduced new Galois lattices, which we have called alpha Galois lattices, allowing the control of the number of nodes of the lattice and thus the adaptability to users and data. The basic idea of our approach is to first build a coarse (but complete) Galois lattice covering the whole set of instances and then to make successive extensional and intentional zooms on the part of Galois lattices delimited by two nodes chosen by the user. This approach has been implemented in the ZooM system [55, 13, 20, 56, 96] devoted to the construction of Galois lattices at different levels of abstraction, and applied to real data from an electronic catalog of computer products in the setting of the RNRT GAEL project. The theoretical foundations of alpha Galois lattices have been presented in [57, 92, 76].

#### Semi-automated construction of an ontology

In the setting of the PICSEL2 project, we are investigating how to semi-automatically construct a class-based ontology from a set of Document Type Definitions (DTDs) related to a given application domain. An ontology composed of classes, properties and relations is needed in order to provide a mediated schema over several XML sources. DTDs can be

Research groups
IASI
Research description

represented as labelled trees, but due to the lack of formal semantics of DTDs, there is no clear correspondence between the tags labelling the different nodes of a DTD and the classes in an ontology structured according the subsumption relation. Our solution to deal with this semantic mismatch is to extract classes, properties and relations of the ontology from the DTDs using heuristics. The extraction phase is a structure-based process done on a sample of DTDs covering the whole application domain. The set of discovered elements is then structured: classes are linked; properties are associated with the classes that they characterize. The extraction phase is entirely automatic but is followed by a manual validation phase. During the validation step, the designer may eliminate non-relevant terms or rewrite abbreviations. The extraction process is then iterated until the results are satisfactory. A first prototype (OntoMedia) has been implemented for extracting components of the ontology from DTDs [83]. An application has been designed from DTDs provided by the Open Travel Alliance (OTA), an organization working to define standardized e-business messages for exchanges in the travel industry [58, 87].

#### **Key references**

- [13] N. Pernelle, M.-C. Rousset, H. Soldano, and V. Ventos. Zoom : a nested Galois latticesbased system for conceptual clustering. *Journal of Experimental and Theoritical Artificial Intelligence (JETAI) on concept lattices for KDD*, 1(14):157-187, 2002.
- [58] C. Reynaud and G. Giraldo. An application of the mediator approach to services over the Web. In *Concurrent Engineering*, volume 1, pages 26-30, 2003.
- [71] A. Termier, M.-C. Rousset, and M. Sebag. TreeFinder: a first step towards XML data mining. In *International Conference on Data Mining (ICDM02)*, 2002.
- [72] A. Termier, M.-C. Rousset, and M. Sebag. DRYADE :a new approach for discovering closed frequent trees in heterogeneous tree databases. In *International Conference on Data Mining* (*ICDM*), 2004.
- [76] V. Ventos, H. Soldano, and T. Lamadon. Alpha Galois lattices. In *International Conference on Data Mining, ICDM 2004*, 2004.

## Theme 3: Mediation for the Semantic Web

• Participants: Chantal Reynaud (leader), Philippe Adjiman, Alain Bidault, Christine Froidevaux, François Goasdoué, Hassen Kefi, Marie-Christine Rousset, Brigitte Safar.

The Semantic Web aims at building a Web with a semantic representation of the contents of the sources, this representation being both understandable by human beings and accessible to automatic reasoning. A first step towards the Semantic Web is the construction of semantic portals that provide users with a uniform interface for accessing collections of pre-existing sources related to a same domain. Semantic portals correspond to a centralized mediation between users and distributed data sources. We are also investigating a distributed approach to mediation in a peer-to-peer setting. Finally, we are working on cooperative answering systems serving as mediating tools between end-users and information integration systems.

#### Semantic portals

A semantic portal (Figure 2) is a centralized mediation system that provides a uniform access to heterogeneous data sources, based on an ontology describing the semantic of the target application domain. The data sources are connected to the portal by views (predefined queries) that describe their contents in terms of the domain ontology. In such portals, the key problem for query answering is query rewriting: the user's queries must be

reformulated in terms of the available views, leading to query plans that can then be evaluated on the data sources. In the PICSEL project, in collaboration with France Telecom R&D, we have developed the first declarative environment for building semantic portals based on an ontology language enriching the OWL-like description logics with rules. This was made possible by the theoretical work described in Theme 1 above on query rewriting in CARIN [9]. The resulting software systems, OntoClass and OntoQuery, have been patented by France Telecom R&D and have been successfully re-used for e-commerce in the European IST MKBEEM project. They are the basis for the development and deployment by France Telecom of an information integration platform based on PICSEL.



Figure 2: Architecture of a semantic portal.

#### Peer-to-peer data management systems

Semantic peer-to-peer data management systems (PDMS) have been proposed recently to generalize the centralized approach of information integration systems based on single semantic portals connected to heterogeneous data sources views. Such a centralized vision of mediation is appropriate for integrating several data sources but cannot handle some Semantic Web applications for which PDMSs based on distributed mediation are more adapted. In a PDMS, there is no central portal (Figure 3): each peer has its own schema and data and can mediate with some other peers to answer queries. We have designed the Somewhere PDMS whose data model complies with the W3C recommendations since it is captured by the propositional fragment of the OWL ontology language. Our work on distributed reasoning in propositional logic provides the query-answering algorithm of Somewhere since we have shown that in our data model query answering amounts to a proper prime implicant calculus [34, 77]. This simple data model already scales up to more than a thousand peers in our first benchmarks. Moreover, it guarantees that the queryanswering problem is decidable. Somewhere is the basis for the forthcoming MediaD project in collaboration with France Telecom R&D on efficient distributed mediation. The problem of distributed storage of XML documents in a peer-to-peer architecture is also addressed by Gemo [28].

Research groups
IASI
Research description



Figure 3: Architecture of a peer-to-peer data management system.

#### **Mediation with end-users**

In information integration systems, the actual content of the different data sources is hidden to the users who formulate their queries in terms of a domain ontology serving as a global schema for the application. The possible limitations on the data stored in the available data sources are expressed by constraints attached to the description of the sources content according to the concepts of the domain ontology. For example, in an information integration system related to tourism, users could ask for 2 stars hotels in Paris while the only available sources provide either Hotels located in London or four stars hotels. Therefore, queries that are valid with respect to the domain ontology may have no answer because of the constraints stated on the available data sources. Conversely, queries that are too general may lead to a huge number of answers, which is impossible to display to the user. In both cases, query refinement tools are needed to help the user to reformulate his query. We have designed an algorithm in order to identify when queries have no answers [102]. It tests the satisfiability of the queries given the set of constraints. Another algorithm enables then either to repair the query by a generalisation process [39], either to choose a close query among a set of predefined ones specified from a description of the available sources [40]. In order to address large numbers of answers, we have designed an algorithm for query refinement based on the construction of a Galois lattice for clustering the answers. Our approach has two distinguishing features. It integrates in a single framework standard Galois lattices with taxonomies of concepts. It avoids the cost of the construction of the whole Galois lattice by the iterative and dynamic construction of its most general concepts at each step of the interaction with the user [64].

#### **Key references**

- [9] F. Goasdoué, V. Lattes, and M. C. Rousset. The use of carin language and algorithms for information integration. *International Journal of Cooperative Information Systems*, (9):383-401, 2000.
- [34] P. Adjiman, P. Chatalic, F. Goasdoué, M.-C. Rousset, and L. Simon. Distributed reasoning in a peer-to-peer setting. In *European Conference on Artificial Intelligence (ECAI)*, 2004.
- [**39**] A. Bidault, C. Froidevaux, and B. Safar. Repairing queries in a mediator approach. In *European Conference on Artificial Intelligence (ECAI)*, pages 406-410, August 2000.
- [40] A. Bidault, C. Froidevaux, and B. Safar. Similarity between queries in a mediator. In *European Conference on Artificial Intelligence (ECAI)*, pages 235-239, Lyon, 2002.

[64] B. Safar, H. Kefi, and C. Reynaud. OntoRefiner, a user query refinement interface usable for semantic web portals. In *Application of Semantic Web Technologies to Web Communities Workshop, European Conference on Artificial Intelligence (ECAI)*, 2004.

## **Theme 4: Thematic warehouses**

 Participants: Brigitte Safar (leader), Hélène Gagliardi, Olivier Haemmerlé, Hassen Kefi, Amar Mezaour, Nathalie Pernelle, Marie-Christine Rousset, Chantal Reynaud, Marco Suarez, Véronique Ventos

This theme is recent in our group. It has been initiated by the RNTL e.dot project. Thematic warehouses integrate data collected from multiple sources and relative to a same application domain described by an ontology. In contrast with mediators, which manage virtual data, warehouses store data extracted from information sources and then integrated with the data already present in the warehouse. In all our work on thematic warehouses, a domain ontology is used. It guides data acquisition, extraction and integration. Moreover, as XML is nowadays the new standard for the exchange of data, it has been chosen as the data model of the warehouse. Our current work is developed in the context of the e.dot RNTL project which aims at enriching an existing data warehouse (in the domain of food risk assessment) with data from the Web (Figure 4). In e.dot, the domain ontology describes the contents of the warehouse in two parts: a relational database and data represented in the conceptual graph formalism. The ontology plays the role of the schema of the warehouse. It is a set of concepts organized in a hierarchy with their synonyms in French and their translation in English. The organisation and the materialisation of the warehouse is specified in a high-level model specific to thematic warehouses and with all operations invoked through Web services (ACWARE Active Content Warehouse). These specifications are executed by Active XML, a framework developed by Gemo to deal with embedded Web services calls inside Web documents [31, 45].



Research groups
IASI
Research description

Figure 4: Enrichment of data in e.dot

Laboratoire de Recherche en Informatique / 2000-2004 report

#### Acquisition of data from the Web

Our work combines a search engine such as Google or a web crawler (such as that of Xyleme) with a filtering tool that can distinguish, among the possible thousands of web pages returned by Google or the Xyleme crawler, those that really contain useful data for the warehouse. In the first e.dot experiments, it was shown that guiding the search through the Web by keywords extracted from the domain ontology was not precise enough to guarantee that the returned Web pages were relevant to the topic of the warehouse. Our approach for designing a filtering tool is generic and declarative. We have defined and implemented a query language, called WebQL [51], which enables the combination of different criteria for specifying the Web pages of interest. Those criteria allow for combining content and structure of searched documents. For example, using WebQL, it is possible for the warehouse administrator to specify that he is interested in HTML pages containing tables with titles including a given keyword. The evaluation of a WebQL query does not provide a "yes or no" answer but a matching coefficient, which is the basis for ranking the pages that are filtered. We plan to integrate WebQL within a focused crawling tool.

Another work focuses on characterizing Web documents. A prototype system, called Thesus, extracts keywords from the incoming links of a page and converts them into semantics by mapping them to a domain ontology [11, 54]. In the e.dot experiments, this system has been used to annotate previously extracted Web pages and to help eliminate irrelevant ones.

#### Extraction and integration of web data driven by an ontology

We want to populate the thematic data warehouse with data found in tables of HTML or PDF documents because tables often contain relevant and synthetic data. We explicitly separate tasks that are specific to the format of a web source, e.g., HTML, from the tasks that are independent of any source but specific to the domain. Thus, we first automatically transform the various tables into a generic XML representation called XTab. Because of the semantic heterogeneity among sources, extracting data from Web pages is insufficient to support integration. Data has to be organized in a different way with a different vocabulary, i.e. we have to find an XML representation where most of the values and tags belong to the ontology. In our approach, we want this transformation to be as automatic and flexible as possible, only driven by the ontology and the way the data has been structured in the original table. Thus, we have defined a Document Type Definition named SML (Semantic Markup Language) that can automatically be generated using the ontology and can deal with additional or incomplete information in a semantic relation, ambiguities or possible interpretation errors. This transformation has been partly implemented and experimented on real data from the e.dot project. The extracted data will be exploited by an interrogation engine named MIEL++, which is developed by the INAPG team. The complete architecture is described in figure 4.

#### Mappings between ontologies

Our aim is to allow a user to interrogate with a single query several data sources related to the same topic but annotated with terms from multiple ontologies. In order to be able to query the other sources, mappings between terms of the different ontologies are needed. We propose to compute these mappings once and for all and to store them in the warehouse in order to be used by the query engine. The identification of the mappings combines both syntactic and semantic comparison techniques, i.e. name-based matches and structure-level matches. This research work is being applied in the e.dot project with two ontologies: Sym'Previus and Com'Base. A first tool supporting automatic syntactic mapping techniques has been developed.

#### **Key references**

- [31] S. Abiteboul, O. Benjelloun, and T. Milo. Positive Active XML. In ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, June 2004.
- [45] G. Cobena and B. Nguyen. Model, design and construction of a service-oriented Webwarehouse. In *European Conference on Digital Libraries*, 2003.
- [51] A.-D. Mezaour. Focused search on the web using wequel. In *10th International Workshop* on Knowledge Representation meets Databases (KRDB), pages 63-74, 2003.

## **Research Perspective**

We plan to consolidate the most recent theme on thematic warehouses, and to pursue our work on the Semantic Web in two main directions: (1) query answering in semantic peer-to-peer systems, (2) semi-automatic construction and exploitation of ontologies for building semantic portals.

Research groups

Research perspective

## B.8.3 / IASI

## Highlights

The IASI group has reached a good scientific visibility in the new and hot topic of information integration at the confluence of the two fields: artificial intelligence and databases. It has developed a very tight and productive collaboration with France Telecom R&D through several projects. The creation of the PCRI project Gemo as a joint group between IASI and the former INRIA Verso project has lead to one of the few Artificial Intelligence and Database groups in France and in Europe.



*Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

- Size of the group at the end of the period (01/01/1999, 01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international), and
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

## Honors

## Prizes and awards

• Marie-Christine Rousset has been a member of *Institut Universitaire de France* (IUF) until August 2002.

### **Keynote addresses**

- Chantal Reynaud, invited keynote speaker, *Topical Session Workshop, Semantic Integration of Heterogeneous Data, IFIP World Computer Congress,* Toulouse, France, August 2004: "Building Scalable Mediator Systems" [97].
- Marie-Christine Rousset, invited keynote speaker, *International Conference on the Semantic Web*, Hiroshima, Japan, Nov. 2004: "Small can be Beautiful in the Semantic Web" [61].
- Marie-Christine Rousset, invited keynote speaker, *Journée Grilles Bio-Médicales*, Orsay, France, April 2003: "Integrating Heterogeneous Data Sources: a Centralized versus a Peerto-Peer Approach".
- Marie-Christine Rousset, invited keynote speaker, *International Symposium on Methodologies* for Intelligent Systems (ISMIS), Lyon, France, June 2002: *Knowledge Representation for* Information Integration.
- Marie-Christine Rousset, invited keynote speaker, *Journées de l'IUF*, Montpellier, France, March 2001: "Représentation de connaissances et modélisation du raisonnement en Intelligence Artificielle".
- Marie-Christine Rousset, invited keynote speaker, *Journée scientifique de l'UFR d'Orsay*, Orsay, France, October 2000: "Les faces cachées du Web : problèmes et modèles".

#### B.8.5 / IASI

## **Evaluation of research**

## **Editorial boards**

#### **International journals**

- TOIT, ACM Transactions on Internet Technology: Marie-Christine Rousset.
- AICOM, *AI Communications*: Marie-Christine Rousset.
- ETAI, Electronic Transactions on Artificial Intelligence for the Concept-based Knowledge Representation and Semantic Web domains: Marie-Christine Rousset.
- SAT-Journal: Laurent Simon.

#### **National journals**

- JEDAI, Journal Electronique d'Intelligence Artificielle: Chantal Reynaud since 2001.
- RIA, Revue d'Intelligence Artificielle: Chantal Reynaud, since 2001.
- Interstices (corpora of "on-line knowledge" hosted by INRIA): Marie-Christine Rousset.

#### Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

#### IASI

Highlights Honors Evaluation of research

- RI3, Revue Information Interaction Intelligence: Marie-Christine Rousset.
- Revue *Psychologie de l'Interaction*: Brigitte Safar, member of editorial board of a special issue, n°9-10, Feb. 2000.

#### **Program committees**

#### Chair

- KRDB, 9th International Workshop on Knowledge Representation meets Databases, 2002: Marie-Christine Rousset, Chair of Program Committee and of Organization Committee.
- RFIA, *Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle*, 2000: Marie-Christine Rousset, Co-chair of the Program Committee.

#### International events

- ICSS, International Conference on Conceptual Structures: Olivier Haemmerlé, (2003, 2004).
- EKAW, International Conference on Knowledge Engineering and Management: Chantal Reynaud, (2002, 2004).
- ISWC, 3rd International Semantic Web Conference: Chantal Reynaud, Marie-Christine Rousset (2004).
- KR, International Conference on Principles of Knowledge Representation and Reasoning: Marie-Christine Rousset (2004).
- SIGMOD, ACM International Conference on Management of Data: Marie-Christine Rousset (2004).
- FQAS, International Conference on Flexible Query Answering Systems: Brigitte Safar (2004).
- SAT, International Conference on Theory and Applications of Satisfiability Testing: Laurent Simon (2003, 2004).
- EKAW workshop, Workshop on Knowledge management and the Semantic Web, 14th International Conference on Knowledge Engineering and Management: Chantal Reynaud (2004).
- IJCAI, International Joint Conference on Artificial Intelligence: Marie-Christine Rousset (2003).
- PODS, International Conference on Principles of Database Systems: Marie-Christine Rousset (2003).
- DEXA workshop, *Workshop on Web semantics, International Conference on Database and Expert Systems Applications:* Chantal Reynaud (2003).
- K-CAP workshop, Workshop on Knowledge Management and the Semantic Web, Second International Conference on Knowledge Capture: Chantal Reynaud (2003).
- ECAI, European Conference on Artificial Intelligence: Marie-Christine Rousset (2000, 2002).
- ISMIS, *International Symposium on Methodologies for Intelligent Systems*: Marie-Christine Rousset (2002).
- ECAI workshop, Workshop on Knowledge Transformation for the Semantic Web, European Conference on Artificial Intelligence: Chantal Reynaud (2002).
- ECAI workshop, *Workshop on Machine Learning and Natural language Processing for Ontology Engineering, European Conference on Artificial Intelligence,* Lyon, France: Chantal Reynaud (2002).
- EKAW workshop, Workshop on Knowledge Management through Corporate Semantic Webs, International Conference on Knowledge Engineering and Management: Chantal Reynaud (2002).
- WebDB, International Workshop on the Web and Databases: Marie-Christine Rousset (2002).
- FMII, Foundations of Models for Information Integration workshop: Marie-Christine Rousset (2001).

- IJCAI workshop, *Workshop on Ontology Learning, International Joint Conference on Artificial Intelligence:* Marie-Christine Rousset (2001).
- WWW workshop, *Workshop on Semantic Web, World-Wide Web Conference:* Marie-Christine Rousset (2001).
- UIDIS, Second International Workshop on User Interfaces to Data Intensive Systems: Marie-Christine Rousset (2001).
- EKAW, 12<sup>th</sup> International Conference Knowledge Engineering and Knowledge Management: Chantal Reynaud (2000).
- ECAI workshop, Workshop on Applications of Ontologies and Problem-Solving Methods, European Conference on Artificial Intelligence: Chantal Reynaud (2000).
- KRDB, 7th International *Workshop on Knowledge Representation meets Databases:* Marie-Christine Rousset (2000).

#### **National events**

- Journées Ingénierie des Connaissances: Chantal Reynaud since its creation in 1997.
- Journée Web sémantique médical: Chantal Reynaud (2004).
- JNPC, *Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets:* Laurent Simon (2003, 2004).
- CAP, Conférence d'Apprentissage: Véronique Ventos (2004).
- Journée Web sémantique médical: Chantal Reynaud (2003).
- JFT, Premières Journées Francophones de la Toile, Tours'Polytech: Chantal Reynaud (2003).
- BDA, Conférence sur les Bases de Données Avancées: Marie-Christine Rousset (2003).
- EGC, *Journées Francophones d'Extraction et de Gestion des Connaissances:* Marie-Christine Rousset (2002, 2003).
- Journées Semantic Web, Action Spécifique Web sémantique du département STIC du CNRS: Chantal Reynaud (2002).
- RFIA, *Congrès Francophone de Reconnaissances des Formes et Intelligence Artificielle:* Marie-Christine Rousset, Member of the Editorial Board (2002).
- Congrès INFORSID: Chantal Reynaud (2001).

## Evaluation committees and invited expertise

- Scientific Committee of LIFO, *Laboratoire d'Informatique Fondamentale d'Orléans:* Marie-Christine Rousset, member.
- Scientific Committee of national program on telemedicine and health technologies, *ACI Télémédecine et Technologies pour la Santé:* Marie-Christine Rousset, member.
- Scientific Committee of national program for junior researchers, *ACI Jeunes Chercheurs:* Marie-Christine Rousset, member in 2000.
- Scientific Committee of national program on masses of data, *ACI Masses de Données:* Marie-Christine Rousset, member.
- Strategic Committee of French network on health technologies, *Réseau National des Technologies de Santé (RNTS):* Marie-Christine Rousset, member.
- Expertise for French Research Ministry, *Mission Scientifique Universitaire:* Marie-Christine Rousset (evaluation of research labs, masters and graduate schools, jury for the attribution of research bonus "prime d'encadrement doctoral et de recherche").
- Expertise for research projects of the *Contrat de Plan Etat-Région de l'Ile de France:* Marie-Christine Rousset, *chargée de mission pour les Mathématiques et l'Informatique*, appointed

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

Evaluation of research

by the Research Ministry near the Délégué Régional à la Recherche et Technologie d'Ile de France.

• Steering Committee of PCRI: Marie-Christine Rousset, member.

#### **Other evaluation activities**

- Reviewer for Ph.D. dissertation: Marie-Christine Rousset (4), Chantal Reynaud (1).
- Reviewer for "Habilitation": Marie-Christine Rousset (2).

#### B.8.6 / IASI

## Volunteer professional service

#### Management positions in scientific organisations

- CNRS network *(Réseau Thématique Pluridisciplinaire)* "Information and Intelligence: Reasoning and Deciding": Marie-Christine Rousset, member of the steering committee.
- National network on Information, Interaction and Intelligence (PRC-GDR I3), Acquisition and Knowledge Engineering task force: Chantal Reynaud, Co-chair.
- CNRS Working group (*Action Spécifique*) "Algorithms and Experimental Approaches for Evaluating Quantified Formulas": Philippe Chatalic and Laurent Simon, co-chairs (2002-2003).
- CNRS Working group (*Action Spécifique*) "Semantic Web": Chantal Reynaud, co-chair (Dec. 2001 to Sept. 2003).
- CNRS Working group (*Action Spécifique*) "Web Services and Information Integration": Chantal Reynaud, co-chair (2004).

## Organisation of conferences and scientific events

#### International

- FQAS, International Conference on Flexible Query Answering Systems, 2004: Marie-Christine Rousset, General Co-chair.
- ECAI, *European Conference on Artificial Intelligence*, 2004: Marie-Christine Rousset, Tutorial Chair.
- SAT Competition: Laurent Simon, Chair in 2002, 2003 and 2004.

#### National

- *Journée Nationale sur les formules booléennes quantifiées,* Nov. 21, 2003, Paris, France: Philippe Chatalic and Laurent Simon, co-organizers.
- Journée du LRI, June 2003 and June 2004: Nathalie Pernelle, organizer.
- CNRS working group on the Semantic Web: *Journées Scientifiques de l'Action Spécifique Web sémantique*, October 10-11, 2002, Paris, France, *Journée Web sémantique Médical*, March 17, 2003, Rennes, France and *Journée Web sémantique et Sciences Humaines et Sociales*, May 7, 2003, Paris, France: Chantal Reynaud, Co-organizer.

## Working groups

- European Network OntoWeb: Chantal Reynaud, member (June 2001 to December 2002).
- CNRS Working Group (*Action Spécifique*) "Semantic Web", sub-group "Integration of Heterogeneous Data Sources": Chantal Reynaud, member (2002-2003).
- CNRS Working group (*Action Spécifique*) "Interactive Solving and Compiling of Combinatorial Problems": Laurent Simon, member (2003-2004).
- University project (*Bonus Qualité Recherche*, Univ. Paris XI) "Highlighting Latent Structure in XML Documents", in collaboration with members of the LIMSI laboratory: Nathalie Pernelle, Marie-Christine Rousset, Alexandre Termier, members (2002) [94].

#### B.8.7 / IASI

## **Contracts and grants**

Project Type	Scientific director	Project Name	Funding institution	Managing institution	Dates	Total €
Industry	MCh. Rousset	PICSEL1	CNET	UPS/REVAF		17997
Industry	MCh. Rousset	PICSEL2	CNET	UPS	2001-2004	104882
RNTL	MCh. Rousset	E.DOT	MENRT	UPS/SAIC	2003-2005	75681
ACI MD	MC. Rousset	MDD	MENRT	CNRS	2003-2006	67726
AS	Ph. Chatalic	Algorithms and Experimental Approach for evaluating quantified boolean formulas	CNRS STIC	CNRS	2002-2003	10290
AS	C. Reynaud	Semantic Web	CNRS STIC	CNRS	2001-2002	9147
AS	MCh. Rousset	Médiation d'information via les meta-données	CNRS STIC	CNRS	2002-2003	5 600
AS	L. Simon	Interactive Solving and Compiling of combinatorial Problems	CNRS STIC	CNRS	2003-2004	
PPF	MC. Rousset	Plan Pluri-Formations Supelec				9057
SUBV	MC. Rousset	Dotation Institut Universitaire de France	IUF	UPS/SUBSF		20076
SUBV	MC. Rousset	Dotation Institut Universitaire de France	IUF	UPS/SUBSE		15244

Note: See the glossary for acronyms.

## Summary of each scientific project:

#### PICSEL1

*Partners : France Telecom Research and Development (Rennes) and IASI/LRI (Paris XI).* 

Industrial contract, 1998-2001 Scientific director: Marie-Christine Rousset

This project was focused on designing a declarative environment for building ontology-based mediators. Two Java tools

have been developed and patented by France Telecom R&D: OntoClass, a tool for automatic classification of concepts defined using a description logic, and OntoQuery, a tool for rewriting conjunctive queries expressed in terms of the mediated ontology, into query plans that use only views describing source contents. The mediator system PICSEL was the kernel Research groups

service

Volunteer professional

Contracts and grants

for constitution of the European project "MKBEEM", a project to develop intelligent knowledge-based multilingual key components for application in multilingual electronic commerce platform. http://mkbeem.elibel.tm.fr/

#### PICSEL2

Partners: France Telecom Research and Development (Rennes) and IASI/LRI (Paris XI).

This project is the continuation of PICSEL1, which focused on designing a declarative environment for building ontology-based

Industrial contract, 2001-2004 Scientific director: Marie-Christine Rousset.

mediators. It aims at scaling up to the Web the mediator approach that has been implemented in PICSEL1. The goal is to facilitate the automatic construction of a mediated schema over several XML sources described by DTDs and related to a same domain. A prototype (OntoMedia) has been developed, which extracts ontology components automatically from a set of DTDs. In PICSEL2, we also develop methods initiated in PICSEL1 for cooperative query answering.

#### E.DOT

Partners: Verso/INRIA, IASI/LRI, Group BIA/INRA, the Xyleme Company.

The goal of the project is to propose generic solutions for enriching existing data warehouses with data from the web. The target application domain is food risk assessment. The technical aspects Research project funded by national network on software technology (RNTL), 2003-2005 Scientific director for LRI: M.-Ch. Rousset.

involve Web crawling, classification and integration of heterogeneous information and mass storage in the Xyleme repository. One of the challenges is the integration of XML data with legacy data (typically relational). It is thus an excellent test bed for Gemo's technology.

#### ACI MDD

*Partners: Mostrare/INRIA, IASI/LRI, Statistical Machine Learning Group at LIP6.* 

The goal of this project is to study fundamental problems raised by modern information retrieval and to determine novel solutions to solve these problems. In particular, we want to build tools for retrieving and extracting information, which fully and Project funded by national initiative on Massive Data (ACI Masses de données), 2003-2006

Scientific director for LRI: M.-Ch. Rousset.

jointly exploit the structure and contents of the XML documents. The distinguishing feature of our approach is to use machine-learning techniques for building flexible and robust tools applicable to large corpora of structured documents, which are possibly heterogeneous, varied and dynamic.

#### **PPF Supelec**

#### Partners: IASI/LRI, Supelec.

This project deals with the structuring and classification of complex data. It is composed of three themes: classification of

data with Galois Lattice structures, construction of ontologies over complex data (hypermedia and spatio-temporal data), query refinement directed by a taxonomy of terms.

University grant (Plan Pluri-Formations) Scientific director for LRI: M.-Ch. Rousset.



## Collaborations

## **Cooperation agreements**

• Collaboration project, Programme Pluri-Formation (PPF) Structuration et classification de données complexes between Supelec (Gif-sur-Yvette, France) and LRI: Hassen Kefi, Nathalie Pernelle, Chantal Reynaud, Marie-Christine Rousset, Brigitte Safar, Véronique Ventos, members.

#### Collaborations leading to joint publications

- Xyleme research action initiated by the Verso project at INRIA-Rocquencourt that also involved members of our group, of the Vertigo group in Paris and of the university of Mannheim in Germany: Chantal Reynaud, Marie-Christine Rousset. This collaboration resulted in two joint international publications [8, 60].
- Université Paris-Nord: Nathalie Pernelle, Véronique Ventos, collaboration with the machine learning research group [13, 20, 75, 96, 57, 86, 92, 76].

## Other collaborations

#### International

- Databases Research Group, University of California at San Diego (UCSD): Marie-Christine Rousset, collaboration with Victor Vianu.
- Vrije Universiteit, Amsterdam: Laurent Simon, collaboration with Frank van Harmelen and Perry Groot, invitation for two weeks (February 2002).
- University of Michigan, USA: Laurent Simon, collaboration with Karem A. Sakallah, invitation for two weeks (June 2002)

#### National

- Engineering of a peer-to-peer semantic integration information platform with Michel Plu from France Telecom R&D: François Goasdoué.
- Engineering of a centralized semantic integration information platform with François Paulus from France Telecom R&D: François Goasdoué.
- Collaboration with Alain Léger, Director of Scientific Program "Knowledge Processing" at France Telecom R&D: François Goasdoué.
- Collaboration with the natural language processing research team at France Telecom R&D: François Goasdoué.
- Collaboration with the human interaction research team at France Telecom R&D: François Goasdoué.
- Co-training of Ph.D. students with Supelec, Gifsur-Yvette, France (with Yolaine Bourda): Chantal Reynaud.
- Co-training of Ph.D. students with the University of Jerusalem (with Shuki Sagiv): Marie-Christine Rousset.

Research groups

IASI Contracts and grants Collaborations

## Dissemination and technology transfer

## **Patents and licences**

• Deposit of 2 softwares (OntoClass and OntoQuery) by France Telecom R&D, authors: François Goasdoué, Marie-Elisabeth Goncalves and Marie-Christine Rousset. Agence pour la Protection des Programmes IDDN.FR.001.260017.00.S.P.2000.000.10500.

#### Popularisation of research results

- Magazine article: François Goasdoué and Marie-Christine Rousset, "Intégration d'Information par Médiation", *Plein-Sud Spécial Recherche 2004*, Journal de l'Université Paris Sud XI, 2004.
- Magazine article: François Goasdoué and Marie-Christine Rousset, "Querying Distributed Data through Distributed Ontologies: a Simple but Scalable Approach", *IEEE Intelligent Systems magazine*, 2003.
- Presentation: Chantal Reynaud, "Ontologies et Recherche d'information", two research seminars, one organized by the *Bibliothèque Nationale de France* (BNF), Paris, France, September 11, 2000 and the second organized by École Polytechnique, Palaiseau, France, October 19, 2000.
- Book chapter: Chantal Reynaud, "Ingénierie des Connaissances pour les systèmes d'information", in C. Cauvet and C. Sabroux-Rosenthal (eds), *Ingénierie des Systèmes d'Information*, Traité IC2 (Informatique Commande Communication), Hermès, 2001.
- Presentation: Chantal Reynaud, "Ontologies et Web Sémantique : Problématique, méthodes, outils, retour d'expérience et enjeux", MEMENTO Seminar, *Club Ecrin Ingénierie du traitement de l'information*, INRIA Sophia-Antipolis, France, November 15, 2001.
- Presentation: Chantal Reynaud, "Sur quelques aspects du Web sémantique", Assises du GDR-I3, Nancy, France, December 4-6, 2002.
- Seminar: *Le Web sémantique : de nouveaux enjeux documentaires ?*, organized by ADBS and the University Institute of Technology of Paris V, Sources d'Europe, Socle de la Grande Arche, Paris La Défense, October 14, 2003: Chantal Reynaud, member of the program committee.
- Presentation: Chantal Reynaud, *Le Web Sémantique : une infrastructure d'intégration de sources de données*, research seminar organized by ADBS and University Institute of Technology of Paris V, *Le Web sémantique : de nouveaux enjeux documentaires ?*, Sources d'Europe, Socle de la Grande Arche, Paris La Défense, October 14, 2003.
- Science fair: Marie-Christine Rousset, managing the theme *L'Informatique, demain*, Fête de la science, Orsay, October 2004.

#### Summer schools, tutorials, invited seminars

- Seminar I3, GREYC, Caen, May 2004, *Towards an automatic structuration of textual data in XML documents:* Nathalie Pernelle.
- Seminar of the ExeCo project, Sophia-Antipolis, France, July 2004, *Médiation sémantique dans un réseau pair à pair : la clé pour le Web sémantique ?* Marie-Christine Rousset.
- Summer School, *DistRibUtIon de Données à grande Echelle* (DRUIDE), Le Croisic, France, May 2004, *Médiation sémantique dans un réseau Pair-à-Pair*: Marie-Christine Rousset.
- Invited Seminar, Journées de l'ACI PairApair, Arcachon, France, May 2004, *Coupling DHT with sophisticated query capabilities*: Marie-Christine Rousset.

- Seminar of LISI/ENSMA, Poitiers, France, Nov. 2003, *Intégration d'informations Le projet PICSEL*: Chantal Reynaud.
- Invited Seminar, Journée Loriatech, Nancy, France, Oct. 2001, Gestion et interrogation de données dans Xyleme: Marie-Christine Rousset.
- RFIA Seminar, LORIA, Nancy, France, January 2001, *Représentation de connaissances et intégration d'informations*: Marie-Christine Rousset.
- Invited Seminar, Journées Exploration de données issues d'Internet, March 2001, LIPN, University Paris 13, *Construction et affinement de hiérarchies à partir de données semi-structurées*: Véronique Ventos, Nathalie Pernelle.

#### B.8.10 / IASI

## Training and education (doctoral and post-doctoral)

Defended doctorates					
Name	Date defended	Current position			
GOASDOUE François	23/11/01	Assistant Professor in the IASI group			
SIMON Laurent	17/12/01	Assistant Professor in the IASI group			
BIDAULT Alain	8/07/02	post-doc in the BioInfo group			
TERMIER Alexandre	28/04/04	post-doc at University of Osaka, Japan			

#### **Graduate courses**

- DEA I3, Fondements de la Représentation des Connaissances: Marie-Christine Rousset.
- DEA I3, Formation à la Recherche: Marie-Christine Rousset.
- DEA I3, *Ontologie et raisonnement sur des sources d'information*: Chantal Reynaud, Christine Froidevaux and Brigitte Safar.
- DEA I3, Intégration de sources d'informations multiples et hétérogènes: Marie-Christine Rousset and Brigitte Safar.

## Other research-related teaching activities

- Marie-Christine Rousset is the director of the graduate school École Doctorale d'Informatique de Paris-Sud.
- Christine Froidevaux has been co-responsible of the masters program DEA I3 Information, Interaction, Intelligence until 2001.

Research groups

IASI

Dissemination and technology transfer

## Publications

## International peer-reviewed journals

- [1] S. Abiteboul, S. Cluet, G. Ferran, and M.-C. Rousset. The Xyleme project. *Computer Networks*, 39:225-238, 2002.
- [2] M. Benedikt, M. Grohe, L. Libkin, and L. Segoufin. Reachability and connectivity queries in constraint databases. *JCSS*, 66(1):169-206, 2003.
- [3] M. Benedikt, L. Libkin, T. Schwentick, and L. Segoufin. Definable relations and firstorder query languages over strings. *Journal ACM*, 50:694-751, 2003.
- [4] M. Brambilla, S. Ceri, S. Comai, P. Fraternali, and I. Manolescu. Specification and design of workflow-driven hypertexts. *Journal of Web Engineering*, 2003.
- [5] P. Buche, C. Dervin, R. Thomopoulos, and O. Haemmerlé. Fuzzy querying on incomplete, imprecise and heterogeneously structured data in the relational model using ontologies and rules (to appear). *IEE Computer Society*, 2004.
- [6] P. Chatalic and L. Simon. Multiresolution for SAT checking. *Journal of Artificial Intelligence Tools*, 10(4), December 2001.
- [7] S. de Amo, N. Bidoit, and L. Segoufin. Order independent temporal properties. *Journal of Logic and Computation*, 14(2):277-298, 2004.
- [8] C. Delobel, C. Reynaud, M.-C. Rousset, J.-P. Sirot, and D. Vodislav. Semantic integration in Xyleme: a uniform tree-based approach. *Data and Knowledge Engineering Review*, 44(3):267-298, March 2003.
- [9] F. Goasdoué, V. Lattes, and M. C. Rousset. The use of carin language and algorithms for information integration. *International Journal of Cooperative Information Systems*, (9):383-401, 2000.
- [10] F. Goasdoué and M.-C. Rousset. Answering queries using views: a krdb perspective for the semantic web. In ACM Journal - Transactions on Internet Technology (TOIT), volume 4, pages 255-288, 2004.
- [11] M. Halkidi, B. Nguyen, I. Varlamis, and M. Vazirgiannis. THESUS: Organizing Web document collections based on semantics and clustering. *The International Journal on Very Large Databases*, 12(4):320-332, November 2003.
- [12] D.L. Berre, P. Burdom, and L. Simon. A phylogenetic tree for the SAT 2002 contest. Accepted for publication in Annals of Mathematics and Artificial Intelligence (AMAI), 43, 2005.
- [13] N. Pernelle, M.-C. Rousset, H. Soldano, and V. Ventos. Zoom : a nested Galois latticesbased system for conceptual clustering. *Journal of Experimental and Theoritical Artificial Intelligence (JETAI) on concept lattices for KDD*, 1(14):157-187, 2002.
- [14] E. Pitoura, S. Abiteboul, D. Pfoser, G. Samaras, and M. Vazirgiannis. DBGlobe: a service-oriented P2P system for global computing. *SIGMOD Record 32(3): 77-82, 2003.*
- [15] M.-C. Rousset. The Semantic Web needs languages for representing (complex) mappings between (simple) ontologies. *IEEE Intelligent Systems*, 17(2), 2002.
- [16] M.-C. Rousset and C. Reynaud. Knowledge representation for information integration. *Information Systems International Journal*, 29(1):3-22, 2003.
- [17] L. Simon, D. L. Berre, and E. Hirsch. The SAT2002 competition. Accepted for publication in Annals of Mathematics and Artificial Intelligence (AMAI), 43:343-378, 2005.
- [18] R. Thomopoulos, P. Buche, and O. Haemmerlé. Representation of weakly structured imprecise data for fuzzy querying. *Fuzzy Sets and Systems*, 140(1):111-128, October 2003.

#### National peer-reviewed journals

- [19] M. Baker, M. Joab, B. Safar, and D. Schlienger. Analyses et modélisations d'explications sur un corpus de dialogues finalisés. *Revue Psychologie de l'Interaction*, 10(9):325-356, 2000.
- [20] N. Pernelle, M.-C. Rousset, and V. Ventos. Construction et affinement de hiérarchies à partir de données semi-structurées. *numéro spécial de la Revue d'extraction des connais*sances et apprentissage, 1(1-2):81-92, 2001.
- [21] M.-C. Rousset, A. Bidault, C. Froidevaux, H. Gagliardi, F. Goasdoué, C. Reynaud, and B. Safar. Construction de médiateurs pour intégrer des sources d'information : le projet PICSEL. *revue 13*, 2002.
- [22] B. Safar and D. Schlienger. Elaboration de stratégies d'explication coopératives pour des dialogues conflictuels. *Revue Psychologie de l'Interaction*, 10(9):325-356, 2000.

#### **Book chapters**

- [23] J. Charlet, C. Reynaud, and R. Teulier. *Ingénierie des connaissances pour les systèmes d'information*. Hermès, 2003.
- [24] C. Reynaud, B. Safar, and H. Gagliardi. Une expérience de représentation d'une ontologie dans le médiateur PICSEL (to appear). Eyrolles, 2004.
- [25] C. Reynaud and F. Tort. Diriger la réutilisation de composants à l'aide d'ontologies. In Ingénierie des connaissances, évolutions récentes et nouveaux défis, pages 59-74. Eyrolles, 2001.
- [26] M.-C. Rousset and C. Reynaud. Picsel and Xyleme: two illustrative information integration agents. In *Intelligent Information Agents Research and Development in Europe: An agentLink Perspective*. Springer-Verlag, 2003.

## Major international peer-reviewed conferences

- [27] S. Abiteboul. Managing an XML warehouse in a P2P context. In Advanced Information Systems Engineering, 15th International Conference, CAiSE, volume 2681, pages 4-13. Springer-Verlag, 2003.
- [28] S. Abiteboul. Distributed information management with XML and Web services. In L. N. in Computer Science, editor, *European Joint Conferences on Theory and Practice of Software (ETAPS), in proc. FASE,* 2004.
- [29] S. Abiteboul, B. Amann, J. Baumgarten, O. Benjelloun, F. D. Ngoc, and T. Milo. Schemadriven customization of Web services (demonstration). In *International Conference on Very Large Data Bases (VLDB)*, pages 1093-1096, September 2003.
- [30] S. Abiteboul, O. Benjelloun, B. Cautis, I. Manolescu, T. Milo, and N. Preda. Lazy query evaluation for Active XML. In *ACM SIGMOD Conference on Management of Data*, June 2004.
- [31] S. Abiteboul, O. Benjelloun, and T. Milo. Positive Active XML. In ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, June 2004.
- [32] S. Abiteboul, A. Bonifati, G. Cobena, I. Manolescu, and T. Milo. Dynamic XML documents with distribution and replication. In *International ACM Special Interest Group for the Management of Data (SIGMOD) conference*, 2003.
- [33] S. Abiteboul, M. Preda, and G. Cobena. Adaptive on-line page importance computation. In *World Wide Web Conference*, May 2003.

Research groups
IASI
Publications

- [34] P. Adjiman, P. Chatalic, F. Goasdoué, M.-C. Rousset, and L. Simon. Distributed reasoning in a peer-to-peer setting. In *European Conference on Artificial Intelligence (ECAI)*, 2004.
- [35] A. Arion, A. Bonifati, G. Costa, I. Manolescu, and A. Pugliese. Efficient query evaluation over compressed XML data. In *International Conference on Extending Database Technologies (EDBT)*, March 2004.
- [36] D. L. Berre and L. Simon. The essentials of the SAT'03 competition. In A. T. E. Giunchiglia, editor, *Proceedings of the Sixth International Conference on Theory and Applications of Satisfiability Testing (SAT2003)*, volume Lecture Notes in Computer Science 2919, pages 452-467. June 2003.
- [37] D. L. Berre, L. Simon, and A. Tachella. Challenges in the QBF arena: the SAT'03 evaluation of QBF solvers. In A. T. E. Giunchiglia, editor, *Proceedings of the Sixth International Conference on Theory and Applications of Satisfiability Testing (SAT2003)*, volume Lecture Notes in Computer Science 2919, pages 468-485. June 2003.
- [38] A. Bidault, C. Froidevaux, and B. Safar. Finding successful queries in a mediator context. In *International Conference on Flexible Query Answering Systems (FQAS)*, october 2000.
- [39] A. Bidault, C. Froidevaux, and B. Safar. Repairing queries in a mediator approach. In *European Conference on Artificial Intelligence (ECAI)*, pages 406-410, August 2000.
- [40] A. Bidault, C. Froidevaux, and B. Safar. Similarity between queries in a mediator. In *European Conference on Artificial Intelligence (ECAI)*, pages 235-239, Lyon, 2002.
- [41] P. Buche, J. Dibie-Barthélemy, O. Haemmerlé, and M. Houhou. Towards flexible querying of XML imprecise data in a dataware house opened on the web. In *Flexible Query Answering Systems (FQAS)*, pages 28-40. Springer-Verlag, June 2004.
- [42] P. Buche, O. Haemmerlé, and R. Thomopoulos. Integration of heterogeneous, imprecise and incomplete data: an application to the microbiological. In 14th International Symposium on Methodologies for Intelligent Systems (ISMIS), volume 2871 of Lecture Notes in Artificial Intelligence, pages 98-107. Springer-Verlag, October 2003.
- [43] P. Chatalic and L. Simon. Multi-resolution on compressed sets of clauses. In Twelth International Conference on Tools with Artificial Intelligence (ICTAI'00), pages 2-10, 2000.
- [44] P. Chatalic and L. Simon. Zres: The old Davis-Putnam procedure meets ZBDDs. In D. McAllester, editor, 17th International Conference on Automated Deduction (CADE'17), number 1831 in Lecture Notes in Artificial Intelligence (LNAI), pages 449-454, June 2000.
- [45] G. Cobena and B. Nguyen. Model, design and construction of a service-oriented Webwarehouse. In *European Conference on Digital Libraries*, 2003.
- [46] C. Delobel and M.-C. Rousset. A uniform approach for querying large tree-structured data through a mediated schema. In *Foundations of Models For Information Integration Workshop (FMII-2001)*, 2001.
- [47] F. Gayral, N. Pernelle, and P. Saint-Dizier. On verb selectional restrictions : Advantages and limitations. In *Natural Language Processing (NLP)*, pages 57-68. Springer-Verlag, 2000.
- [48] F. Goasdoué and M.-C. Rousset. Compilation and approximation of conjunctive queries by concept descriptions. In *European Conference on Artificial Intelligence (ECAI)*, pages 267-271, Lyon, 2002.
- [49] F. Goasdoué and M.-C. Rousset. Querying distributed data through distributed ontologies : a simple but scalable approach. In *Information Integration on the Web*, 2003.
- [50] F. Goasdoue and M. Rousset. Rewriting conjunctive queries using views in description logics with existential restrictions. In *Description Logics conference (DL)*, pages 113-122, 2000.
- [51] A.-D. Mezaour. Focused search on the web using wequel. In 10th International Workshop on Knowledge Representation meets Databases (KRDB), pages 63-74, 2003.
- [52] T. Milo, S. Abiteboul, B. Amann, O. Benjelloun, and F. D. Ngoc. Exchanging intensional XML data. In ACM Special Interest Group for the Management of Data conference (SIG-MOD), pages 289-300, June 2003.

- [53] A. Muscholl, T. Schwentick, and L. Segoufin. Active context-free games. In STACS, 2004.
- [54] B. Nguyen, G. Cobena, and M. Vazirgiannis. Organization of Web document collections based on link semantics. In *European Conference on Digital Libraries*, 2003.
- [55] N. Pernelle, M.-C. Rousset, and V. Ventos. Automatic construction and refinement of a class hierarchy over multi-valued data. In *Principles and Practice of Knowledge Discovery in Database (PKDD)*, pages 386-398. Springer-Verlag, 2001.
- [56] N. Pernelle, M.-C. Rousset, and V. Ventos. Automatic construction and refinement of a class hierarchy over semi-structured data. In *Ontology Learning Workshop associated with the International joint Conference Conference on Artificial Intelligence (IJCAI)*, 2001.
- [57] N. Pernelle, V. Ventos, and H. Soldano. Alpha Galois lattices. In *First International Conference on Formal Concept Analysis: The State of the Art (ICFCA03)*, 2003.
- [58] C. Reynaud and G. Giraldo. An application of the mediator approach to services over the Web. In *Concurrent Engineering*, volume 1, pages 26-30, 2003.
- [59] C. Reynaud and B. Safar. Representation of ontologies for information integration. In *International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, 2002.
- [60] C. Reynaud, J.-P. Sirot, and D. Vodislav. Semantic integration of XML heterogeneous data sources. In *International Database Engineering and Applications Symposium (IDEAS)*, 2001.
- [61] M.-C. Rousset. Small can be beautiful in the semantic web. In *3rd International Semantic Web Conference (ISWC)*, 2004.
- [62] C. Rouveirol and V. Ventos. Towards learning in Carin-ALN. In *International conference on Inductive Logic Programming (ILP)*, pages 191-208. Springer-Verlag, 2000.
- [63] B. Safar and H. Kefi. Domain ontology and Galois lattice structure for query refinement. In *International Conference on Tools with Artificial Inteligence (ICTAI)*, pages 597-601, November 2003.
- [64] B. Safar, H. Kefi, and C. Reynaud. OntoRefiner, a user query refinement interface usable for semantic web portals. In *Application of Semantic Web Technologies to Web Communities Workshop,European Conference on Artificial Intelligence (ECAI)*, 2004.
- [65] L. Segoufin. Typing and querying XML documents: some complexity bounds. In *PODS*, 2003.
- [66] L. Simon. SatEx: towards and exhaustive and up-to-date SAT experimentation. IJCAI'01 Workshop on Empirical Methods in Artificial Intelligence (Organized by H. Hoos and T. Stützle), August 4th 2001.
- [67] L. Simon and P. Chatalic. SATEX: a Web-based framework for SAT experimentation. In H. Kautz and B. Selman, editors, *Proceedings of the Workshop on Theory and Applications of Satisfiability Testing SAT2001*, June 14th-15th 2001.
- [68] L. Simon and A. del Val. Efficient consequence finding. In *17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, pages 359-365, Seattle, Washington, USA, 2001.
- [69] A. Termier, M.-C. Rousset, and M. Sebag. Combining statistics and semantics for word and document clustering. In Ontology Learning Workshop associated with the International joint Conference Conference on Artificial Intelligence (IJCAI), 2001.
- [70] A. Termier, M.-C. Rousset, and M. Sebag. Mining XML data with frequent trees. In DBFusion workshop, 2002.
- [71] A. Termier, M.-C. Rousset, and M. Sebag. TreeFinder: a first step towards XML data mining. In *International Conference on Data Mining (ICDM02)*, 2002.
- [72] A. Termier, M.-C. Rousset, and M. Sebag. DRYADE :a new approach for discovering closed frequent trees in heterogeneous tree databases. In *International Conference on Data Mining (ICDM)*, 2004.

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

- [73] R. Thomopoulos, P. Bosc, P. Buche, and O. Haemmerlé. Logical interpretation of fuzzy conceptual graphs. In *International Conference NAFIPS*, pages 173-178, July 2003.
- [74] R. Thomopoulos, P. Buche, and O. Haemmerlé. Different kinds of comparisons between fuzzy conceptual graphs. In 11th International Conference on Conceptual Structures (ICCS), volume 2746 of Lecture Notes in Artificial Intelligence, pages 54-68. Springer-Verlag, July 2003.
- [75] V. Ventos, P. Brezellec, and H. Soldano. Explicitly using default knowledge in concept learning: an extended dl plus strict and default rules. In *International Conference on Logic Programming and Non-monotonic Reasoning (LPNMR)*, 2001.
- [76] V. Ventos, H. Soldano, and T. Lamadon. Alpha Galois lattices. In *International Conference* on *Data Mining, ICDM 2004*, 2004.

#### National peer-reviewed conferences

- [77] P. Adjiman, P. Chatalic, F. Goasdoué, M.-C. Rousset, and L. Simon. Raisonnement distribué dans un environnement de type pair-à-pair. In Actes des Dixièmes Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets (JNPC), 2004.
- [78] A. Bidault, C. Froidevaux, and B. Safar. Proximité entre requêtes dans un contexte médiateur. In *Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, volume 2, pages 653-662, Angers, 2002.
- [79] A. Bidault and B. Safar. Réparation de requêtes posées à un médiateur. In *Congrès de Reconnaissance des formes et Intelligence Artificielle (RFIA)*, pages 225-235, february 2000.
- [80] P. Chatalic and L. Simon. Résolution sur des ensembles de clauses compressés : comment doper DP. In Actes des Cinquièmes Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets (JNPC'00), pages 107-118, Marseille, 2000.
- [81] P. Chatalic and L. Simon. SatEx : un cadre de travail dynamique pour l'expérimentation SAT. In Actes des Septièmes Journées Nationales sur la Résolution Pratique de Problèmes NP-Complets (JNPC'01), pages 259-270, Toulouse, 27-28-29 Juin 2001.
- [82] J. Dibie-Barthélemy, O. Haemmerlé, and E. Salvat. Validation de graphes conceptuels. In *Extraction et Gestion de Connaissances (EGC)*, pages 135-146. Cépaduès, Janvier 2004.
- [83] G. Giraldo and C. Reynaud. Construction semi-automatique d'ontologies à partir de DTDs relatives à un même domaine. In *13<sup>es</sup> journées francophones d'Ingénierie des Connaissances (IC)*, 2002.
- [84] A. D. Mezaour. Recherche ciblée de documents sur le web. Revue des Nouvelles Technologies de l'Information (RNTI-E2), numéro spécial EGC'2004, 2:491-502, January 2004.
- [85] B. Nguyen, I. Varlamis, M. Halkidi, and M. Vazirgiannis. Construction de classes de documents Web. In *Journées Francophones de la Toile (JFT)*, 2003.
- [86] N. Pernelle, V. Ventos, and H. Soldano. Treillis de Galois alpha. In 1ère Journée IA et Treillis de Galois, Plateforme AFIA, 2003.
- [87] C. Reynaud and G. Giraldo. Médiation de services sur le Web. In *Journées Francophones de la Toile*, pages 59-68, 2003.
- [88] C. Reynaud and B. Safar. Aide à la formulation de requêtes dans un médiateur. In *Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, volume 1, pages 201-210, Angers, 2002.
- [89] C. Reynaud, B. Safar, and H. Gagliardi. La représentation de l'ontologie dans le médiateur PICSEL. In journees francophones d'Ingénierie des connaissances (IC), pages 329-348, 2001.
- [90] B. Safar and H. Kefi. Apport d'une ontologie du domaine pour affiner une requête à l'aide d'un treillis de Galois. In *Journées Francophones de la Toile (JFT)*, 2003.
- [91] R. Thomopoulos, P. Buche, and O. Haemmerlé. Sous-ensembles flous définis sur une

ontologie. In *Extraction et Gestion des Connaissances (EGC)*, pages 147-158. Cépaduès, Janvier 2004.

[92] T. Lamadou, V. Ventos, H. Soldano. Treillis de galois alpha. In *Conférence d'Apprentissage CAP*, 2004.

### Other conferences and workshops

- [93] P. Buche, C. Dervin, O. Haemmerlé, E. Surleau, and R. Thomopoulos. Carrying out the microbial risk in food products using the MIEL software: a new tool to query incomplete, imprecise and heterogeneously structured experimental data stored in a relational database. In *International Conference on Predictive Modelling in Food, ICPMF 03*, pages 58-60, June 2003.
- [94] H. Folch, B. Habert, M. Jardino, N. Pernelle, M. Rousset, and A. Termier. Highlighting latent structure in documents. In *International Conference on Language Resources and Evaluation (LREC)*, 2004.
- [95] G. Giraldo and C. Reynaud. Vers l'automatisation de la construction de systèmes de médiation pour le commerce électronique. In *Journées Web Sémantique, Action Spécifique du département STIC du CNRS*, 2002.
- [96] N. Pernelle, V. Ventos, and H. Soldano. ZooM: Alpha Galois lattices for conceptual clustering. In *Managing Specialization/Generalization Hierarchies*, 2003.
- [97] C. Reynaud. Building scalable mediator systems. In *Topical Day in Semantic Integration* of Heterogeneous Data, IFIP World Computer Congress, 2004.
- [98] B. Safar, A. Bidault, C. Froidevaux, H. Gagliardi, and C. Reynaud. Production de réponses coopératives dans un contexte médiateur. In *Journées Web Sémantique, Action Spécifique du département STIC du CNRS*, 2002.
- [99] R. Thomopoulos, P. Buche, and O. Haemmerle. Carrying out the microbial risk in food products using the Conceptual Graph model: a flexible complement to relational databases to represent and query heterogeneously structured experimental data. In *International Conference on Predictive Modelling in Food, ICPMF 03*, pages 61-63, June 2003.

## **Doctoral dissertations and Habilitations**

- [100] F. Goasdoué. *Réécriture de requêtes en terme de vues*. Ph.D. thesis, Laboratoire de Recherche en Informatique, Université Paris-Sud, 2001.
- [101] L. Simon. Multirésolution pour le test de consistance et la déduction en logique propositionnelle. Ph.D. thesis, Laboratoire de Recherche en Informatique, Université Paris-Sud, December 2001.
- [102] A. Bidault. *Aide à l'affinement de requêtes posées à un serveur d'informations*. Ph.D. thesis, Laboratoire de Recherche en Informatique, Université Paris-Sud, July 2002.
- [103] B. Nguyen. *Construction et maintenance d'un entrepôt de données du Web*. Ph.D. thesis, Laboratoire de Recherche en Informatique, Université Paris-Sud, 2003.
- [104] O. Benjelloun. *A data-centric perspective on Web services.* Ph.D. thesis, Laboratoire de Recherche en Informatique, Université Paris-Sud, 2004.
- [105] A. Termier. *Extraction d'arbres fréquents dans un corpus hétérogène de données semi-structurées*. PhD thesis, Laboratoire de Recherche en Informatique, Université Paris-Sud, April 2004.

Research groups
IASI
Publications

## Other publications

- [106] M. Brambilla, S. Ceri, S. Comai, P. Fraternali, and I. Manolescu. Specification and design of workflow-driven hypertexts. In *World Wide Web Conference, Poster*, 2003.
- [107] P. Chatalic and L. Simon. Are pigeon hole and urquhart problems really exponential for DP? In *Third Workshop on the Satisfiability Problem (SAT2000)*, May 2000. Slides of the presentation only.
- [108] F. Goasdoue and M.-C. Rousset. Intégration d'information par médiation. *Plein Sud spécial recherche*, 2004.
- [109] A. Lerner and I. Manolescu. The nuts and bolts of DBMS construction: building your own prototype. In *Simposio Brasileiro de Banco de Dados (Brazilian national database conference), tutorial,* 2003.
- [110] I. Manolescu, S. Ceri, M. Brambilla, P. Fratemali, and S. Comai. Exploring the combined potential of Web sites and Web services. In *World Wide Web Conference, Poster*, 2003.



Laboratoire de Recherche en Informatique / 2000-2004 report

# équipe Inférence et Apprentissage

Responsable : Michèle Sebag

léquipe Inférence et Apprentissage a été sous la direction de Yves Kodratoff une des équipes pionnières de l'apprentissage artificiel au niveau mondial. Les thèmes de recherche ont évolué pour répondre aux défis posés par la croissance exponentielle de l'information disponible, et permettre l'extraction de connaissances stratégiques à partir de ces données.

Les points majeurs des quatre dernières années concernent :

• La jonction entre la programmation logique inductive (Céline Rouveirol et Michèle Sebag ont présidé la 11ème conférence internationale du domaine) et le domaine de la satisfaction de contraintes, démontrant l'existence d'une région aveugle en relation avec la transition de phase (collaboration avec l'Université du Piémont).

• La jonction entre la fouille de texte et la linguistique systémique fonctionnelle, formalisant les ressources requises (ontologies, schémas linguistiques) comme les solutions de problèmes d'apprentissage.

• L'extension des recherches en optimisation stochastique et algorithmes d'évolution, renforcées par l'arrivée de Michèle Sebag. Le couplage de la programmation génétique et des grammaires probabilistes hors contexte permet l'identification non paramétrique dans des espaces syntaxiquement contraints (par exemple la découverte de lois dimensionnellement cohérentes).

• La création du groupe Robotique, renforcée par l'arrivée de Nicolas Bredèche et centrée sur le projet "Action, Anticipation and Adaptation" du programme national Robea. Ce projet unit I&A, le projet Fractal de l'INRIA et le groupe Perception Située du LIMSI.

• Le développement d'approches de fouille de données visuelles, motivées par la fouille de données médicales (collaboration avec l'équipe INSERM ERM0202). Cette approche, permettant l'identification de facteurs de risque, a conduit à des avancées théoriques en sélection d'attributs (collaboration avec l'Université d'Amsterdam).

Les liens entre l'équipe I&A et le projet Fractal de l'INRIA ont conduit à la création en 2003 du projet PCRI TAO (Thèmes Apprentissage et Optimisation), centré sur la fertilisation croi-



sée de l'apprentissage et de la fouille de données d'une part, et de l'optimisation et des algorithmes d'évolution d'autre part. TAO est l'une des seules équipes européennes disposant d'une expertise internationale dans ces deux domaines. L'enjeu du projet est de proposer une perspective unifiée qui couvre les aspects adaptation, apprentissage et optimisation, indissociables pour la modélisation et le contrôle de systèmes complexes.



## Inference and Learning

Head: Michèle Sebag

The Inference & Learning research group, which used to be headed by Yves Kodratoff, has been among the world pioneer teams in machine learning since the 80's. It later broadened its scope to knowledge discovery and data mining, motivated by the ever growing amount of available data and the strategic importance of their exploitation. The I&A group is a member of the European Network of Excellence PASCAL (FP6).

The main highlights of the past four years can be summarized as follows: • The research team on Inductive Logic Programming (ILP) and learning from structured data (Céline Rouveirol and Michèle Sebag co-chaired the 11th ILP Conference, 2001) focused on the hybridation of ILP and constraint satisfaction. A major negative result showed the existence of a failure region in connection with the phase transition (2000, collaboration with Università del Piemonte).

• The research team on Text Mining summarized the formalization of the linguistic resources (ontologies, linguistic schemata) needed as solutions to learning problems. Recently, the junction was made between this line of research and the Systemic Functional Linguistics framework.

• The Evolutionary Computation team was reinforced by the arrival of Michèle Sebag. Evolutionary computation, for its flexibility and ability to incorporate domain knowledge, offers unrivalled abilities for ill-posed optimisation problems, such as arise in machine learning and data mining. The coupling of probabilistic context-free grammars with genetic programming was instrumental in searching for optimal hypotheses in nonparametric spaces, e.g., dimensionally consistent equation discovery.

• The Robotics team was created in 2003, fostered by the arrival of Nicolas Bredèche. The national Robea project "Action, Anticipation and Adaptation", tackles the challenging application of machine learning and evolutionary computation to autonomous robot control (collaboration with LIMSI and the INRIA project Fractal).

• Along applications in medical data mining (collaboration with INSERM ERM0202), an evolutionary computation-based approach to visual data mining was developed for risk factor identification and sensitivity analysis. This work resulted in a fundamental research on feature selection (collaboration with University of Amsterdam).

The links between I&A and the Fractal INRIA project led to the creation of the PCRI Project TAO in 2003. TAO, one of the few European research groups with international expertise in machine learning and evolutionary



computation, focuses on the cross-fertilisation of learning, data mining and evolutionary computation. an unified framework, accounting for the tightly linked facets of adaptation, learning and optimisation, is needed to model complex systems and scale up to real-world applications.

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups

1 & A

Introduction

## **Research Group Members**

#### Personnel as of 01/01/2004

Full time faculty			
Name	First Name	Position*	Institution
BOURNAUD	Isabelle	MC	IUT de Sceaux
BREDECHE	Nicolas	MC	Université Paris XI
CORNUEJOLS	Antoine	MC	IIE EVRY
FRANOVA	Marta	CR1	CNRS
KODRATOFF	Yves	DR2	CNRS
SEBAG	Michèle	DR2	CNRS
Doctoral students			
AMRANI	Ahmed	D	Sté ESIEA Recherche
BASKIOTIS	Nicolas	D	ENS
FELKIN	Mary	D	INRIA
GODZIK	Nicolas	ATER	Université Paris XI
MALOBERTI	Jérôme	D	University Yokohama
			& Université Paris XI
MARY	Jérémie	AM	Université Paris XI
OULD ABDEL VETAH	Mohamed	ATER	Université Paris XI
ROCHE	Mathieu	AM	Université Paris XI
ISHIDA	Celso	D	Université fédérale du Parana (Brésil)
			& Université Paris XI
Temporary personnel			
AZE	Jérôme	ATER	Université Paris XI
MATTE-TAILLEZ	Oriane		Université Paris XI

Non-LRI personnel of PCRI project TAO					
SCHOENAUER	Marc	DR2	INRIA		
ABBOUD	Kamal	D	École Polytechnique		
AUGER	Anne	D	École Polytechnique		
ROUDENKO	Olga	D	École Polytechnique		
SINGH	Vijay Pratap	D	CIFRE IFP		
PIERRES	Matthieu	D	CIFRE Airbus		

\*See the glossary for acronyms.

Long term visitors							
Name	First Name	Nationality	Institution	Arrival	Departure	Funding	
COSTA	Ernesto	Portugese	University of Coïmbra (Portugal)	1/05/01	31/05/01	Visiting professor,	
						Université Paris XI	
MATWIN	Stan	Canadian	Université d'Ottawa (Canada)	1/09/01	31/08/02	Ministère de la Recherche	
						CNRS + UPS	
POZO	Ramirez	Brazilian	Federal University of Parana (Brazil)	1/11/03	31/08/04	Capes grant -Brazil	

## **Group evolution**

- Claire Nédellec, Assistant Professor, left in 2001, now researcher (CR) at INRA
- Céline Rouveirol, Assistant Professor, joined the Bioinformatics group at LRI in 2002
- Arrivals:
- Isabelle Bournaud, Assistant Professor, 2000 (Ph.D. Paris-6)
- Michèle Sebag, CNRS Researcher, 2002 (formerly at Laboratoire de Mécanique des Solides, Ecole Polytechnique)
- Nicolas Bredèche, Assistant Professor, 2003 (Ph.D. Paris-11).

The research group had two departures and three arrivals over the 2000-2004 period. The departures of Claire Nédellec and Céline Rouveirol were motivated by the increasing importance of machine learning in the Bio Computing field. The arrivals of Michèle Sebag, Isabelle Bournaud and Nicolas Bredèche resulted in the development of the Evolutionary Computation team and in the creation of a new research team on Robotics, a challenging application for both machine learning and evolutionary computation.

Research groups

I & A Research group members

## **Research Description**

The I&A group, historically centered on Machine Learning (ML) and Inductive Logic Programming (ILP), has evolved in the last 4 years to address the challenges of Knowledge Discovery from Databases (KDD), Data Mining (DM) and Evolutionary Computation (EC).

Machine learning, data mining and knowledge discovery can all be characterized as search problems, finding "interesting" hypotheses from (massive) amounts of data; the difference lies in the role devoted to the expert user, almost negligible in machine learning, important in data mining, and central in knowledge discovery.

Two main approaches of this search problem have been investigated in the I&A group: Constraint Satisfaction (ILP Theme) and Optimisation (Fundamental ML/KDD/DM and EC Themes). With respect to Optimisation, the key issues are the choice of the search space, or hypothesis language, and the choice of the learning criteria. Both issues (Model Selection, Learning Risk Bounds) are at the core of the PASCAL Network of Excellence (FP6), involving LRI and the Mathematics Lab at Paris-Sud together with major European Research groups in the field.

The I&A group also features a long expertise in Evolutionary Computation, dating back from 1994. EC, including genetic algorithms, genetic programming and evolution strategies, is a population-based stochastic optimisation framework, known for its efficiency with respect to ill-posed optimisation problems and explained from its ability to incorporate domain knowledge.

## PCRI Project TAO: Thème Apprentissage & Optimisation



The cross-fertilization of machine learning, knowledge discovery and evolutionary computation is the main motivation for the PCRI TAO project, created in 2003 and including members of the IA group and the INRIA Fractal group. On the one hand, evolutionary computation offers unrivalled opportunities for machine learning and knowledge discovery, specifically for investigating non-convex learning criteria and controlling the trade-off between the computational resources needed and the hypothesis quality (any-time algorithms). On the other hand, evolutionary computation will highly benefit from a principled, machine learning-based, exploitation of the wealth of data generated by an EC application. This data is at the moment poorly exploited, if at all.

The I&A - TAO group involves six highly coupled research teams / themes. The first two are concerned with the fundamentals of machine learning, knowledge discovery and evolutionary computation. A third one is concerned with inductive logic programming. The last three teams are concerned with applications calling for specific advances in ML/KDD and EC: Text Mining, Inverse Problems (Numerical Engineering), and Robotics.

## Fundamentals of Machine Learning, Knowledge Extraction and Data Mining

• Participants: Antoine Cornuéjols, Michèle Sebag, Yves Kodratoff, Céline Rouveirol, Nicolas Bredèche, Jérôme Azé, Jérémie Mary, Nicolas Baskiotis.

This theme focuses on machine learning, knowledge discovery and data mining (ML/KDD/DM) considered as optimisation problems, and particularly on the key issues of the search space/hypothesis language, and the learning criteria.

#### **AUC-based Learning**

While the main stream of Statistical Learning focuses on well posed, quadratic, optimisation criteria, e.g., kernel methods, our expertise in evolutionary computation allows us to consider non convex optimisation criteria such as the Wilcoxon statistics, or area under the Receiver Operating Characteristics (ROC) curve. The ROC curve describes the trade-off between the two types of error of a hypothesis: the false positive and the false negative error rates. Among its main advantages, the ROC approach is adapted to imbalanced example distributions (one positive for 100 negative training examples), and to cost-sensitive learning (the cost of mistaking an ill person for a healthy one is much higher than vice versa).

The evolutionary optimisation of the area under the ROC curve (AUC) showed very good learning performances compared to prominent approaches such as Support Vector Machines [81], illustrated by applications in medical data mining [80] and text mining [130].

#### **Feature Selection**

Most available databases were not constructed with data mining in mind, and they usually involve a number of features irrelevant to the learning task at hand. The irrelevant features not only result in a significant increase of the computational and memory resources needed; they might also mislead the learning search, ultimately resulting in a poor predictive accuracy.

Feature Selection (FS) is thus recognised as a central task for ML/KDD/DM applications, and particularly so in bio-informatics. In collaboration with the bio-informatics team of LRI and INSERM, a novel algorithm inferring the relevance of attributes from the structure and parameters of hypotheses was proposed [12].

Along the same lines, the variability of solutions provided by stochastic optimisation can be exploited for sensitivity studies. In collaboration with the Vrije Universiteit of Amsterdam, we showed that AUC-based evolutionary optimisation provides the means for Ensemble Feature Ranking [64, 63]. This theoretical result extends the state of the art (see Special Issue on Feature and Variable Selection, Journal of Machine Learning Research, 2003) to non-linear target concepts, while only linearly increasing the size of the search space.

#### **Human Computer Interaction and Visual Data Mining**

Two main requirements are put on knowledge discovery and data mining systems: the experts want the hypotheses produced to be intelligible and accurate. To address these antagonistic requirements, a new graphical approach for comparing sub-populations, evaluating the expert's conjectures, etc., was proposed in [81] (see sidebar Encart1). The expert can interactively exploit precise and accurate hypotheses, asking for more details, and/or

Research groups

 Research description

B.9.2 / 313

checking his/her own conjectures; the answers are expressed in terms of pictures, and human eyes are incomparably efficient at decoding the rich information conveyed by these pictures.



An estimation of the risk for cardio-vascular diseases (CVD) (Challenge Principles and Practice of Knowledge Discovery from Databases, 2001) is used to visually compare the impact of tobacco and alcohol ingestion.

The lower curve shows the risk for the 100 non-smokers in the database. It is always lower than the median risk (horizontal line). The upper curve depicts the risk associated with the 100 heaviest smokers in the database. Although 40% of them have a CVD risk lower than the median risk, 10% of them present a huge risk.



Likewise, the above picture compares the risk respectively associated with the 100 heaviest and lightest drinkers in the database. These curves, unexpectedly showing little or no impact of alcohol ingestion, are explained by the fact that the studied population was found to involve no light drinkers...

#### Methodological aspects, Meta-learning and Competence Maps

The Meta-Learning problem (finding a priori the algorithm most suited to a given dataset) is a bottleneck issue for machine learning algorithms, hampering their widespread use outside research labs, in industry and services.

For this purpose, a methodology inspired by the Constraint Satisfaction and Phase Transition paradigm (see Inductive Logic Programming Theme below) was developed, based on the definition of order parameters problem. In this paradigm, assessing an algorithm amounts to drawing its "Competence Map", picturing its average performance in the landscape defined by the order parameters.
The Competence Map of the best-known (propositional) learning algorithm, C4.5 (decision tree) was built, and effectively led to characterise the regions of problems where this prominent algorithm is inefficient on average. Further, this map gives an unified and precise perspective on long known empirical observations, e.g., regarding the small disjunctive problem [53].

It must be emphasised that this approach significantly differs from an analytical algorithmic study; instead, it postulates that many heuristics are packed into really efficient algorithms, the interaction of which is hardly amenable to analytical modelling. Therefore, an empirical framework originating from natural and physical sciences is relevant to observe the failures of algorithms.

#### **Key references**

- [12] G. Mercier, N. Berthault, J. Mary, A. Antoniadis, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. Biological detection of low radiation by combining results of two analysis methods. *Nucleic Acids Research (NAR)*, 32(1):1-8, 2004.
- [25] A. Cornuéjols and L. Miclet. Apprentissage artificiel. Concepts et algorithmes. Eyrolles, 2002.
- [53] N. Baskiotis and M. Sebag. C4.5 competence map: a phase transition-inspired approach. In R. Greiner and D. Schuurmans, editors, 21<sup>st</sup> International Conference on Machine Learning, ICML 2004, pages 73-80. Morgan Kaufmann, 2004.
- [64] K. Jong, J. Mary, A. Cornuéjols, E. Marchiori, and M. Sebag. Ensemble feature ranking. In 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-04), pages 267-278, Pisa, Italy, 2004. Springer-Verlag.
- [81] M. Sebag, J. Azé, and N. Lucas. Roc-based evolutionary learning: Application to medical data mining. In *Artificial Evolution'03*, Lecture Notes in Computer Science, pages 384-396. Springer-Verlag, 2004.

## Inductive Logic Programming, Constraint Satisfaction and Phase Transition

• Participants: Céline Rouveirol, Michèle Sebag, Antoine Cornuéjols, Isabelle Bournaud, Erick Alphonse, Jérôme Maloberti, Nicolas Baskiotis.

Inductive Logic Programming (ILP) is about learning from structured examples, such as chemical molecules (graphs), XML data (trees), and/or learning structured hypotheses such as toxicological patterns (graphs, sequences) or dimensional differential equations (mechanical models).

The I&A team developed an international competence in ILP and Relational Learning; Céline Rouveirol and Michèle Sebag co-chaired the 11<sup>th</sup> International Conference in Inductive Logic Programming in 2001.

The major difficulties of ILP can be illustrated on an application such as Predictive Toxicology Evaluation (IJCAI Challenge, looking for patterns discriminating carcinogenic molecules in nitro-aromatic compounds). The cornerstone of Machine Learning is the covering test, checking whether a given hypothesis (a chemical pattern) appears in the training examples (molecules). In first-order logic, this problem is NP-complete. Research groups

**Research description** 

ILP thus faces an additional difficulty compared to propositional machine learning, namely the fact that the hypothesis assessment is intractable in terms of worst-case complexity analysis. This drawback is even more serious for relational data mining as it both calls upon relational approaches and challenges their scalability.

#### **Bounded Relational Reasoning**

A first approach addressing the relational covering test, or theta-subsumption, is based on stochastic reasoning, sampling the subsumption lattice. This work pioneered the field of any-time algorithms for relational induction and deduction [14] and was successfully applied to predictive toxicology evaluation. Furthermore, it provides hints into a principled junction between first order logic and probabilistic frameworks, needed to achieve efficient knowledge discovery in highly relational and noisy domains, e.g., Web mining or ontology construction.

#### **Phase Transitions**

The formalisation of the subsumption lattice led us to investigate more deeply the connections between ILP and Constraint Satisfaction Problems (CSP). Independently, L. Saitta and A. Giordana (Università del Piemonte) formalised this connection, and imported into ILP the phase transition complexity paradigm developed for CSP. In this paradigm, the performance of the algorithm (predictive accuracy, running time) is viewed as random variables depending on the order parameters of the problem instances, e.g., the density and tightness of constraints.

The practical implications on ILP of the phase transition paradigm were investigated in collaboration [3], showing the existence of a "failure region" where the up-to-date algorithms behave no better than random guessing. This negative result is acknowledged as significant in the literature, and the explanations offered for this failure region shed unexpected light on the actual limitations of ILP.

#### Junction with Constraint Satisfaction

In parallel, the constraint satisfaction perspective on relational learning was investigated by Erick Alphonse (Ph.D. 2002, advisor Céline Rouveirol), who proposed a constraintbased exploration of the subsumption lattice [49, 48].

Independently, a new theta-subsumption algorithm was devised by Jérôme Maloberti, based on an ad hoc rewriting of theta-subsumption problems into CSPs, and calling upon efficient CSP procedures [8]. This algorithm was shown to outperform the state of the art by several orders of magnitude (see sidebar Encart2).

Lastly, a new relational learning mechanism was proposed to address the failure region mentioned earlier [46]. Formally, this bottom-up approach exploits the constraints that originated from training examples, confronted to that of the order parameters, e.g., total number of variables, of the sought hypothesis (to be guessed by the expert). A stochastic complexity analysis, confirmed by empirical validation, shows that the approach actually addresses part of the failure region observed for greedy (top-down) algorithms.



The standard covering test in ILP (theta-subsumption) shows a phase transition: the percentage of clauses covering an example abruptly drops from almost 1 to almost 0 as the specificity of the clause (number n of variables, number m of predicates) and of the example (number L of constants, number N of litterals built on each predicate symbol) increases. The figure below depicts the percentage of successful theta-subsumption test for n=10 and N=100, in plane (m,L), averaged over 100 pairs (clause, example) uniformly drawn for each (m,L) point.



The complexity of the former best known theta-subsumption algorithm (Scheffer, 1994) is depicted above.



Research groups

Research description

The complexity of the Django algorithm [8] reported below shows an improvement of several orders of magnitude. Note in particular that the complexity is negligible in the region to the right, contrasting with Scheffer (1994).

Laboratoire de Recherche en Informatique / 2000-2004 report

#### Software

The Django algorithm [8] is available; it is used at University of Bari, Tufts University and Yokohama University.

#### **Key references**

- [2] H. Blockeel and M. Sebag. Scalability and efficiency in multi-relational data mining. ACM SIGKDD, Special Issue on Multi-Relational Data Mining, 5(1):17-30, 2003.
- [3] M. Botta, A. Giordana, L. Saitta, and M. Sebag. Relational learning as search in a critical region. *Journal of Machine Learning Research*, 4:431-463, 2003.
- [8] J. Maloberti and M. Sebag. Fast theta-subsumption with constraint satisfaction algorithms. *Machine Learning Journal*, 55:137-174, 2004, Kluwer Academic Publishers.
- [14] M. Sebag and C. Rouveirol. Resource-bounded relational reasoning: Induction and deduction through stochastic matching. *Machine Learning Journal*, 38:41-62, 2000, Kluwer Academic Publishers.
- [48] E. Alphonse and S. Matwin. Feature subset selection and inductive logic programming. In C. Sammut and A. Hoffmann, editors, 19th International Conference on Machine Learning (ICML'02), pages 11-18, Sydney Australia, 2002. Morgan Kaufmann.

## Fundamentals of Evolutionary Computation

• Participants: Michèle Sebag, Antoine Cornuéjols, Jérémie Mary.

Evolutionary Computation (EC) is a unifying framework for population-based optimisation algorithms, including genetic algorithms, evolution strategies, evolutionary programming and genetic programming. These approaches differ by the search space, respectively boolean (genetic algorithms), real-valued (evolution strategies) and structured (genetic programming).

EC is now widely acknowledged as a powerful optimisation framework dedicated to illposed optimisation problems. The main reason for this efficiency comes from the possibility for EC to incorporate background knowledge about the application domain into the representation and the variation operators.

#### Self-adaptive operators in continuous search spaces

A first research direction considers the self-adaptive operators embedded in evolution strategies, which are the most effective EC algorithms for optimisation in  $\mathbb{R}^{n}$ .

In continuous search spaces, genetic individuals are mutated by adding a Gaussian perturbation. It is clear that the most critical issue is the tuning of the Gaussian law, most generally its covariance matrix. A flexible mechanism was proposed, using a quadratic approximation of the target function and constructing the covariance matrix from the Hessian of this estimate in "sufficiently" regular regions, and switching to the standard Covariance Matrix Adaptation otherwise [51].

#### **Estimation of Distribution Algorithms**

Estimation of Distribution Algorithms (EDAs) proceed by alternatively sampling and updating a distribution on the search space. The sampled individuals are evaluated, i.e. their fitness is computed, and the distribution is updated and biased towards the best individuals in the current sample. Extensions of this framework to continuous optimisation, initialised by Ducoulombier & Sebag (1998), showed failures in specific cases where the solutions lie on the edge of the search space. Regularisation heuristics, which calibrate the eigenvalues of this distribution, were shown to successfully overcome such failures.

On-going work (Ph.D. Celso Ishida, co-advised with A. Pozo, Universidad Federale do Parana, Brazil) is concerned with using mixtures of distributions, borrowing to the MIX-MOD EM-like approaches developed in the SELECT project at INRIA, to extend EDAs to multi-modal optimisation.

#### **Genetic Programming and Context-Free Grammars**

As mentioned earlier, EC should take advantage of any available knowledge about the application domain to enable an efficient exploration of huge search spaces and non-regular landscapes. A rigorous framework to express constraints on the search/solutions is by way of grammars. We investigated the coupling of context-free grammars with genetic programming (CFG-GP), first proposed by Gruau (1996), and discovered why this elegant approach was not intensively used in the literature: the constraints in non-toy grammars severely harm the initialization step, and EC cannot recover from a bad initialisation.

This limitation was addressed by using a specific constraint solver within the initialization step [13]. The application motivating this theoretical development was the identification of dimensionally consistent equations for Law Identification in Mechanics (see the Inverse Problems theme below).

Finally, the EDA approach was hybridised with CFG-GP; the underlying distribution was expressed in terms of stochastic grammars, and the approach was shown to solve the "bloat" problem (uncontrolled growth of individuals in genetic programming), which is the major limitation of genetic programming [73]. Applications concern inverse problems, particularly identification of behavioral laws for numerical engineering.

#### **Key references**

- [13] A. Ratle and M. Sebag. Grammar-guided genetic programming and dimensional consistency: Application to non-parametric identification in mechanics. *Applied Soft Computing*, 8:1-14, 2001, Elsevier Science Publishers.
- [44] M. Sebag. *Principles of non-parametric learning*, chapter Genetic Programming for Model Identification, pages 287-354. Springer Wien New-York, 2002.
- [51] A. Auger, N. Vanhaecke, and M. Schoenauer. LS-CMA-ES: Better Covariance Matrix Adaptation using Least Square Approximation. In 8th International Conference on Parallel Problems Solving from Nature (PPSN 2004), LNCS, pages 182-191. Springer-Verlag, 2004.
- [73] A. Ratle and M. Sebag. Avoiding the bloat with probabilistic grammar-guided genetic programming. In P. Collet et al., editor, *Artificial Evolution*, VI, pages 255-266. Springer-Verlag, 2001.

Research groups I & A Research description

## **Text Mining**

• Participants: Yves Kodratoff, Marta Franova, Jérôme Azé, Mathieu Roche, Ahmed Amrani, Michèle Sebag.

Text Mining (TM) is concerned with exploiting/transforming documents to achieve particular tasks. The difficulty lies in the delicate balance to keep between texts, transformations and tasks. Problem resolution implies the existence of cognitive entities, called concepts of specialty, necessary to the resolution of the current tasks.

Our basic hypothesis calls on a theorization of the language, minimal but sufficient to both characterize concepts of specialty and spot their occurrences in the documents. For example, in corpora related to Human Resources, the verb "to give up", followed by the noun "influence" evokes the presence of a concept that the psychologist recognises as describing the nature of relations in the company and that he named "relational". In the same way, the verb "to give up" followed by the noun "task" is said to evoke a kind of behaviour called "implication" (in this case, it is evidently the lack of implication). Along the same lines, the noun "action", followed by the adjective "together" evokes the concept "relational", while the adjective "adequate" followed by the noun "action", evokes the concept "activity".

Our study focuses on the induction of concepts from texts and examples provided by a domain expert, and the development of efficient procedures for text mining as defined above. Our approach relies on two postulates:

1. The importance of the field expert. The expert's help is needed to understand texts in his/her specialty. It follows that user-friendly tools must enable the expert to transfer the expertise to the program, e.g, in terms of ontologies; however, since the expert's time is expensive, such tools must incorporate inductive steps to speed up the transfer and make it feasible. The difficulty is that induction has to take place from data that is both incomplete and very noisy, a well-known cause of failure for most inductive approaches. Therefore specific inductive algorithms have to be developed.

In particular, the central step for understanding is to spot the occurrences of (interesting) specialty concepts in the documents. The field expert is required to provide a (good) few examples of what he or she identifies as a concept in a text, that is the linguistic instances of the concept's presence. Induction will thus provide a (generalised) characterisation for the occurrences of such concepts.

2. Text mining is a chain of linguistic processes involving, e.g., lemmatisation, tagging, terminology, concept recognition, and information extraction. It must be emphasised that there is no point in independently optimising any step of the above chain (as is currently the main stream in the literature): the chain must be considered and optimised as a whole.

Our approach is rooted in systemic functional linguistics (Halliday 1978, 1994) (see sidebar).



Systemic Functional Linguistics proposes an unified framework involving the manifold aspects of communication, specifically categorising the various social functions of language.

The three top-level categories are:

- 1. interpersonal metafunctions, accounting for interaction between speaker and addressee(s);
- 2. Ideational metafunctions, representing our experience in the world;
- 3. textual metafunctions, the communication means embedding the first two metafunctions.

We are presently working with five technical corpora.

- A "biology" corpus obtained by querying the National Library of Medicine (NLM) Medline database with the keywords "DNA-binding", "proteins" and "yeast", thus obtaining a 10Mb corpus of 6119 abstracts. We are presently in the process of gathering (some of) the corresponding complete papers. It illustrates the problem of dealing with many different texts written in a highly technical language.
- A "Data Mining" corpus (369 Kb), obtained from the introduction of papers published at KDD conferences. The "English speakers" sub-corpus contains 100 texts authored by people working in English-speaking countries, the "French speakers" sub-corpus contains 31 texts in English authored by native French speakers known personally by the creator of the corpus. It illustrates the problem of dealing with few texts written in a semi-technical language, and, as a side effect, it gives indications on the problems met by French speakers when they write in English.
- A "Human Resources" corpus, made of 3784 Kb of a text belonging to the company PerformanSe. It is written in French by a psychologist who serves as expert, and it illustrates the problem of mining well-balanced sentences in a single style.
- A "Resume" corpus (2470 Kb) that contains 1144 Resumes written in French, belonging to the company VediorBis. It illustrates the problem of mining texts written in a semi-telegraphic style with many spelling mistakes.
- A "newspapers" corpus of 40 Mb, which is part of the training data provided by the TREC competition.

Research groups
I & A
Research description

#### **Key references**

- [19] L. Fontaine and Y. Kodratoff. La notion de concept dans les textes spécialisés: une étude comparative entre la progression thématique et la texture des concepts. In *Revue Anglais Spécialité (ASp) 37-38, 2002.*
- [21] Y. Kodratoff. Induction extensionnelle: définition et application à l'acquisition de concepts à partir de textes. In *Revue des Nouvelles Technologies de l'Informatique*, volume 2, 2004.
- [37] Y. Kodratoff. *Machine Learning and its applications*, volume 2049 of *Lecture Notes in Computer Science*, chapter Comparing Machine Learning and Knowledge Discovery in DataBases :An Application to Knowledge Discovery in Texts, pages 1-21. Springer-Verlag, 2001.
- [50] A. Amrani, Y. Kodratoff, and O. Matte-Tailliez. A semi-automatic system for tagging specialized corpora. In H. Dai, R. Srikant, and C. Zhang, editors, 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004), volume 3056 of Lecture Notes in Artificial Intelligence, pages 670-681. Springer-Verlag, 2004.
- [75] M. Roche, J. Azé, O. Matte-Tailliez, and Y. Kodratoff. Mining texts by association rules discovery in a technical corpus. In *Conference on Intelligent Information Processing and Web Mining (IIPWM 2004)*, Advances in Soft Computing, pages 89-98. Springer-Verlag, 2004.

## **Inverse Problems**

• Participants: Michèle Sebag, Antoine Cornuéjols, Jérémie Mary, Nicolas Baskiotis, Jérôme Azé.

Inverse Problems (IP) aim to determine unknown causes based on the observation of their effects. In contrast, direct problems are concerned with computing the effects of (exhaustively described) causes. Inverse problems are mathematically ill-posed in the sense that the existence, uniqueness and stability of solutions cannot be assured.

IPs are present in many areas of science and engineering, such as mechanical engineering, meteorology, heat transfer, electromagnetism, material science, etc. The I&A group has focused on the problems of scientific discovery, modelling physical (mechanical, chemical, biological, etc.) phenomena from available observations and current theories.

Each discovery task can be best characterised by the search space explored to accomplish that task; depending on this search space, diverse Machine Learning (ML) methods have been applied and among them Genetic Programming (GP) has shown promising results. However, genetic programming suffers from a major limitation with respect to machine learning: although the knowledge-based issues of evolutionary computation have been widely acknowledged, canonical genetic programming offers no way of exploiting domain knowledge.

Significant advances have been achieved in the I&A group (see Evolutionary Computation theme above) to combine Genetic Programming and Context Free Grammars (CFG). The motivations for this work stemmed from Law Identification in Solid Mechanics. More precisely, the strain and stress observed along indentation experiments are recorded; the behavioural law characterizing the material at hand is sought for in terms of differential equations. Indeed, the available observations are measured along specific units, e.g., meters Newton, and the expert is willing to find dimensionally consistent laws, e.g., not adding meters with Newtons. Such constraints were implemented in terms of large CFG, with hundreds of non-terminal symbols and thousands of derivation rules. Interestingly, dimensional consistency allows for massive contractions of the GP search space.

Another kind of Inverse Problem is known as Feature Construction, mapping the problem at hand onto a space more amenable to the resolution of the problem. The Feature Construction problem, subsuming the Feature Selection one, is central to artificial intelligence in general, and machine learning and data mining in particular.

The problem of Feature Construction is studied on an application, concerned with characterising good meshes for numerical engineering, particularly the design of 3D meshes in the aerospace industry (Ph.D. Matthieu Pierres, Airbus CIFRE, co-advised by Marc Schoenauer and Michèle Sebag). This challenging real-world application involves relational issues (a mesh is but a set of finite elements and their relations) and probabilistic issues (as usual for realworld applications, the solution is to be sought as a trade-off between conflicting logical rules).

#### **Key references**

- [6] H. Hamda, F. Jouve, E. Lutton, M. Schoenauer, and M. Sebag. Compact unstructured representations in evolutionary topological optimum design. *International Journal of Applied Intelligence*, 16:139-155, 2001.
- [13] A. Ratle and M. Sebag. Grammar-guided genetic programming and dimensional consistency: Application to non-parametric identification in mechanics. *Applied Soft Computing*, 8:1-14, 2001, Elsevier Science Publishers.
- [20] H. Hamda, F. Jouve, E. Lutton, M. Schoenauer, and M. Sebag. Représentations non structurées en optimisation topologique de formes par algorithmes évolutionnaires. *ESAIM:Proc*, Actes du 32e Congrés national d'analyse numérique, 2000.
- [42] M. Schoenauer and M. Sebag. Evolution artificielle et problèmes spatiaux. In H. Prade, R. Jeansoulin, and C. Garbay, editors, *Le temps, l'espace et l'évolutif en sciences du traitement de l'information*, pages 377-396. Cépaduès, 2000.
- [43] M. Schoenauer and M. Sebag. Using Domain Knowledge in Evolutionary System Identification. In K. Giannakoglou, editor, *Evolutionary Algorithms in Engineering and Computer Science*. John Wiley, 2002.

#### **Robotics**

• Participants: Antoine Cornuéjols, Michèle Sebag, Nicolas Bredèche, Jérémie Mary, Mary Felkin, Nicolas Godzik.

Autonomous Robotics is a challenging application for both machine learning and evolutionary computation. From a machine learning perspective, robotics asks for extending machine learning algorithms beyond the classical assumptions of independent and identically distributed examples (indeed the robot exploration results in non independent and identically distributed data). Furthermore, an autonomous robot is immersed in a noisy and dynamic environment where its actions are bound to modify its very perception of the environment (the perception-action loop). Last, robotics perfectly illustrates the learnability issues: robotic sensors with high definition result in a very high dimensional instance space, hindering the learning task; but if the sensor definition is low, this results in "perceptual aliasing": the robot mistakes different places, e.g. corners in a maze, as being the same.

As the rewards for the robot actions are delayed, sparse and noisy, e.g., due to perceptual aliasing, and the state space is huge, an alternative to reinforcement learning is offered by Evolutionary Robotics, pioneered by Nolfi and Floreano (2000). Along this line, the robot controller is sought for by evolutionary computation, as the goal at hand, e.g., exploring the arena, getting to a specific location in a maze, combining exploration and search for a source of energy, is formalised as optimising a manually designed fitness function.

A central issue for Evolutionary Robotics is to decide whether the search space is discrete (as for classifier systems) or continuous (using, e.g., neural nets as the search space). A hybrid approach called "Symbolic Controllers" was proposed in [61]. It involves continuous inputs and discrete outputs. Compared to the now standard Action Selection framework, this approach allows to gradually construct and exploit libraries of hierarchical actions, enforcing the scalability of the approach towards more complex target behaviours.

Another issue is the generality of the robotic controller and the adaptation to environments that are different from those encountered during the (limited) training period. In collaboration with LIMSI (Robea project), we proposed an architecture combining three functionalities (see sidebar): action selection, anticipation (predicting the next robot sensations based on the selected action) and adaptation (reacting to the difference between the predicted sensations and those actually encountered by the robot). This architecture, inspired from cognitive science and sensory-motor contingency, contains a model of the world (the anticipation module) which can be costlessly confronted with the world and provides hints into the needed adaptation [62]. Interestingly, this approach was shown to be much more robust in the long run than the competitive auto-teaching approach proposed by Nolfi and Parisi (1997).

The team has also taken part in software licensing and distribution, developing and making available the "World in a Bottle" simulator (C++/OpenGL), which is interconnected to EO (Evolving Object), a library for Evolutionary Computation developed in the EvoNet framework.

#### **Key references:**

- [4] N. Bredeche, Z. Shi, and J.-D. Zucker. Perceptual learning and abstraction in machine learning : An application to autonomous robotics. *accepted in IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews.*
- [61] N. Godzik, M. Schoenauer, and M. Sebag. Evolving symbolic controllers. In G. Raidl et al., editor, *Applications of Evolutionary Computing*, volume 2611 of *Lecture Notes in Computer Science*, pages 638-650. Springer-Verlag, 2003.
- [62] N. Godzik, M. Schoenauer, and M. Sebag. Robustness in the long run: Auto-teaching vs anticipation in evolutionary robotics. In X. Yao et al., editor, 8th International Conference on Parallel Problem Solving from Nature (PPSN 2004), pages 932-941. Springer-Verlag, 2004.
- [65] C. Kavka and M. Schoenauer. Evolution of voronoi-based fuzzy controllers. In 8<sup>th</sup> International Conference on Parallel Problem Solving From Nature (PPSN 2004), pages 541-550. Springer-Verlag, 2004.



The Action-Anticipation-Adaptation architecture (A3) is an autonomous robot controller composed of three interacting modules, currently implemented as Elman neural nets, the weights of which are optimized using Evolutionary Computation.

The decision making module (agent network on the left) takes as input the robot sensations (provided by the robot sensors) and produces the actions (motor actuators).

The anticipation module (model network on the upper right) predicts the sensations at time t+1 from the sensations and actions at time t.

*The difference between the actual sensations and the anticipated ones provides the robot with a (very basic) self-awareness: if something unexpected happens, the adaptation mechanism is triggered.* 

## **Research Perspective**

Our main research perspective relies on the hybridization of data mining, machine learning and evolutionary computation.

The current challenges of data mining and text mining are related to the "expert in the loop", and the modelling of his/her preferences. It now clearly appears that no general measure of interest can efficiently capture the preferences of the user, which are both subjective and changing as the mining process goes along. Evolutionary computation offers several approaches for handling such preferences.

The first approach is related to multi-objective optimisation, and the identification of the Pareto front. Multi-objective optimisation is in our opinion a relevant and expressive framework for formalising the conflicting criteria of the expert, e.g., simultaneously looking for general and accurate hypotheses.

Another expert priority is to restrict the search to relevant hypotheses. Since distinguish-

I & A

Research description Research perspective ing a relevant regularity from a trivial one in the data can be brought back to the hard AI problem of common sense, two possibilities remain. The first possibility is based on ad hoc heuristics and consists of encoding sufficient domain knowledge. Another possibility is inspired from On-line Learning and Interactive Evolutionary Optimisation (which was proposed in the mid 90s to optimise non computable fitness functions, e.g., related to taste). In this framework, the hypotheses presented to the expert at each step are sampled from his/her interest function and estimated from his/her reactions to the previous hypotheses. The main differences when compared with on-line learning are the distribution of examples, non-independent and identically distributed, and the exploitation/exploration trade-off (dynamic learning). Incidentally, such preference models would be invaluable ways of coding and exchanging knowledge.

Last, interactive data mining can be combined with the visual data mining approach first investigated in [81, 80]that allows the expert to compare sub-populations, evaluate conjectures, etc., through graphics. This combination would allow the expert to obtain implicit hypotheses fitting his/her expectations, which indeed corresponds to an Inverse Problem.

The specific problems mentioned above (multi-objective and interactive optimisation) have been tackled with success through specific evolutionary computation heuristics, which might benefit to learning settings too.

Furthermore, the connection between ensemble learning and multi-modal optimisation will be investigated in more depth. Ensemble learning, a most successful yet not fully understood advance of statistical learning since the early 1990s, is based on the search of a population of solutions (hypotheses); the vote of these diverse hypotheses results in a lower variance error. Conversely, multi-modal optimisation (simultaneously finding the multiple global/local optima of the target optimisation problem) is concerned with maintaining a population of diversified solutions. Ideally, such a study would formalise the sought population of learning solutions as the Pareto front corresponding to a set of antagonistic criteria, e.g., generality vs. precision or false positive vs. false negative error rates.

Symmetrically, a key issue is the exploitation of the wealth of data generated along EC runs. These data are currently not exploited, except for the best individual generated along each run. However, their principled exploitation, through mining the archives of evolution could significantly alleviate the computational burden of artificial evolution. Among the innovative uses of such an archive are:

- Adjusting operator rates (determining the circumstances where a given operator is useful, during the early or late stages of evolution, in combination with other operators, etc);
- Immigration (restoring the population diversity by including non-random migrants, e.g., selected in the evolution archive);
- Policy restart (estimating whether a run can provide worthwhile solutions, and stop early the runs that are stuck in known or bad local optima).
- Characterisation of the Pareto front in multi-objective optimisation, characterisation of the regions that are insufficiently covered, and eventually delegating their coverage to mono-objective optimisation.

## Highlights

The I&A Group, with international expertise in Machine Learning, Data Mining and Knowledge Discovery since the inception of these fields, has been extended with the creation of the TAO PCRI Project in 2003. This project inherits the competence of the Artificial Evolution and Learning Group at Ecole Polytechnique (Applied Maths Center and Solids Mechanics Lab., 1986-2002), and brings together a critical mass of researchers in learning and optimisation.

From a fundamental research perspective, the main milestones of this period are i) Importation of the Phase Transition framework to Inductive Logic Programming; ii) Feature Selection. A reference book on Machine Learning, co-authored by Antoine Cornuéjols, was published in 2002 (Cépadués, 632 pages).

Regarding the applications, although Medical Data Mining (in collaboration with INSERM) and Text Mining are still at the core of our research, some changes have been observed. On the one hand, Céline Rouveirol and Claire Nédellec now actively contribute to the development of two new Bio-Informatics groups (at INRA and LRI), acknowledging the central role of data mining and machine learning in this emerging field. On the other hand, a new, exciting and very challenging field of applications appeared: Robotics (in collaboration with LIMSI). This new theme, at the intersection of learning and optimisation, nicely fits in the TAO Project.

Independently, some links between the I&A group and other LRI groups have been established: with IASI (Alexandre Termier, Ph.D. student co-advised by Marie-Chirstine Rousset and Michèle Sebag, taking advantage of the common background between structured data mining and XML) and with GraFComm (Abdel Lisser, using evolutionary computation for robust dynamic routing in telecom networks).



Research groups **I & A** Research perspective *Radar views display a synthetic view of the activity of the group over the 1995-1999 and 2000-2004 periods, by means of five statistics:* 

•Size of the group at the end of the period (01/01/1999, 01/01/2004),

- •Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- •Annualised amount in Euro of new contracts and grants,
- •Number of participations in program committees (national and international), and
- •Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

B.9.4 / I & A

## Honors

## Prizes and awards

• Yves Kodratoff, corresponding member, Romanian Academy of Science, June 2001.

#### Keynote addresses

- Antoine Cornuéjols, invited keynote speaker, *Journées Nationales de la Recherche en Robotique* (JNRR-03). Murol, France, 7-9 oct. 2003.
- Yves Kodratoff, invited keynote speaker, *Congrès Extraction et Gestion des Connaissances*, Montpellier, France, Janvier 2002.
- Yves Kodratoff, invited keynote speaker, *Premières Journées d'Etudes sur les Systèmes d'Information pour l'Aide à la Décision en Ingénierie Système*, JESIADIS, Brest, November 2002
- Michèle Sebag, invited keynote speaker, *Workshop on Inductive Logic Programming*, Hintergarten 2002.
- Michèle Sebag, invited keynote speaker, Conférence d'Apprentissage, CAP 2001.
- Michèle Sebag, invited keynote speaker, *Journée Romande d'Hygiène Hospitalière* (SWISS-NOSO), Genève 2001.
- Michèle Sebag, invited keynote speaker, *Workshop on Learning and Induction, Genetic and Evolutionary Computation Conference*, GECCO 2001.
- Michèle Sebag, invited keynote speaker, *Workshop on Inductive Logic Programming, International Conference on Machine Learning*, ICML 2000.

## Other honors

• Yves Kodratoff, invited plenary session, Romanian Academy of Science, June 2001.

## **Evaluation of research**

## **Editorial boards**

- *Knowledge and Information Systems:* Michèle Sebag, member of the editorial board since 2003.
- *Machine Learning Journal:* Michèle Sebag, member of the editorial board since 2001 ; Céline Rouveirol & Michèle Sebag, guest editors, 2001.
- *Genetic Programming and Evolvable Hardware:* Michèle Sebag, member of the editorial board since 2000.
- *International Journal of Expert Systems:* Yves Kodratoff, member of the editorial board since 1998.
- Applied Artificial Intelligence: Yves Kodratoff, member of the editorial board since 1998.
- IEEE Transactions on Evolutionary Computation: Michèle Sebag, associate editor, 1999-2003.
- Artificial Intelligence Review: Yves Kodratoff, member of the editorial board since 1996.
- *Revue des Nouvelles Technologies de l'Information:* Y. Kodratoff, 2004, M. Sebag, 2004, guest editors.
- *Revue d'Intelligence Artificielle:* Michèle Sebag, member of the editorial board since 2002, guest editor, 2002.

## Program committees

#### Chair

- RFIA, 14<sup>e</sup> Congrès Reconnaissance des Formes et Intelligence Artificielle, 2004: Michèle Sebag, Editor.
- ICDM, 3rd IEEE International Conference on Data Mining, 2003: Michèle Sebag, Vice-Chair.
- EGC, Extraction et Gestion des Connaissances, 2003: Yves Kodratoff, Co-Chair.
- ICML, 18th International Conference on Machine Learning, 2001: Céline Rouveirol, area chair.
- ILP, 11th Inductive Logic Programming, 2001: Céline Rouveirol and Michèle Sebag, Co-chairs.
- CAP, Conférence d'apprentissage, 1999: Michèle Sebag, Chair.

#### **International events**

- ICIIP, International Conference on Intelligent Information Processing, 2004: Nicolas Bredeche.
- ECAI, 16th European Conference on Artificial Intelligence, 2004: Michèle Sebag, Yves Kodratoff.
- ICDM, *IEEE International Conference on Data Mining:* Michèle Sebag since 2001, Yves Kodratoff since 2002.
- DaWaK, 4<sup>th</sup> International Conference on Data Warehousing and Knowledge Discovery, 2002: Yves Kodratoff.
- ISMIS, International Symposium on Methodologies for Intelligent Systems: Yves Kodratoff since 1998.
- ICML, International Conference on Machine Learning: Michèle Sebag since 1998.
- ILP, Inductive Logic Programming: Céline Rouveirol since 1994, Michèle Sebag, since 1996.
- PPSN, Parallel Problem Solving from Nature: Michèle Sebag since 1996.
- EuroGP, European Conference on Genetic Programming: Michèle Sebag since 2000.
- GECCO, Genetic and Evolutionary Computation Conference: Michèle Sebag since 1999.

Research groups

I & A

Highlights Honors

- ECML-PKDD, European Conference on Machine Learning, Conference on Principle and Practice of Knowledge Discovery from Databases: Michèle Sebag since 1996;Yves Kodratoff, Céline Rouveirol since 1997.
- ICEC then CEC, *IEEE International Conference on Evolutionary Computation:* Michèle Sebag since1994.

#### National events

- RFIA, Reconnaissance des Formes et Intelligence Artificielle: Michèle Sebag, member since 1996.
- EGC, *Extraction et Gestion des Connaissances:* Yves Kodratoff since 2002, M. Sebag since 2002.
- CAP, *Conférence d'apprentissage:* Michèle Sebag since 1999; Yves Kodratoff, Antoine Cornuéjols, Céline Rouveirol since 2001.
- EA, Evolution Artificielle: Michèle Sebag since 1994.

## Evaluation committees and invited expertise

- European commission, 6<sup>th</sup> framework program, IST: Michèle Sebag, expert reviewer, 10-19 november 2003, Brussels.
- National research program ACI Nouvelles Interfaces des Mathématiques: Michèle Sebag, reviewer, 2004.
- Expert for teaching programs LMD (Licence-Master-Doctorat), 2004.
- Journal of Artificial Intelligence; Knowledge Discovery from Databases: Michèle Sebag, reviewer, 2002, 2004.

#### **Other evaluation activities**

- Reviewer for PhD dissertation: Michèle Sebag (13, including Royal Institute of Technology Stockholm, University of Leuven) ; Yves Kodratoff (4) ; Antoine Cornuéjols (4).
- Reviewer for Habilitation: Michèle Sebag (1).

## Volunteer professional service

## Management positions in scientific organisations

- National research program on knowledge management, machine learning and new technologies ACI TCAN *Traitement des Connaissances, Apprentissage, Nouvelles Technologies:* Michèle Sebag, member of the steering committee, 2002-2004; Antoine Cornuéjols, member of the steering committee since 2004.
- CNRS Network "Discovering and Summarizing", *Réseau Thématique Pluridisciplinaire* Découvrir et résumer, RTP 12: Michèle Sebag, member of the steering committee since 2002.
- CNRS Working group on Text Mining, *Action Spécifique Fouille de Textes*, 2003-2004: Yves Kodratoff, co-chair.
- CNRS Working group on Data Mining, *Action Spécifique Extraction et Fouille*, 2000-2002: Michèle Sebag, co-chair.

## Organisation of conferences and scientific events

- ISMIS, 13<sup>th</sup> International Symposium on Methodologies for Intelligent Systems, 2002: Yves Kodratoff, General Chair.
- ECML/PKDD, 12th European Conference on Machine Learning, 5th Principles and Practice of Knowledghe Discovery from Databases, 2001: Michèle Sebag, tutorial chair.
- CAFE-IN, Workshop, Paris-V, 2002: Michèle Sebag, organizer.
- CAFE-BAYES, Workshop, ESIEA, 2001: Michèle Sebag, organizer.
- CAFE-BD, Workshop, Paris-V, 2001: Michèle Sebag, organizer.
- JET, Journées Évolutionnaires Trimestrielles: organization since 1998.

## Working groups

- PASCAL, Network of Excellence, 6<sup>th</sup> Framework Program: Michèle Sebag, corresponding member for Université Paris-Sud since 2003.
- Evonet, European Network on Evolutionary Computation: Michèle Sebag, member (1999-2003).
- KDNet, European Network of Knowledge Discovery: Michèle Sebag, member (1996-2003).

## Other professional service

- AFIA, *Association Française d'Intelligence Artificielle:* Michèle Sebag, member of steering committee since 2000; Jérémie Mary, treasurer since 2004.
- Institut de la complexité: Michèle Sebag, member of the Controversy Seminar since 2003.

Research groups

#### I & A

Evaluation of research Honors Volunteer professional service

## **Contracts and grants**

	Scientific		Funding insti-	Managing	5.7	Duration Budget
Туре	director	Project Name	tution	institution	Dates	(months) (euro)
NoE	M. Sebag	PASCAL	UE	UPS/SAIC	01/12/03 - 30/11/07	26932
Inter-EPST	C. Rouveirol	Caderige		CNRS	04/02/03 - 03/02/04	5183
ACI	A. Cornuéjols	Accamba	MENRT	CNRS	01/07/04 - 01/07/07	9900
ACI-NIM	M. Sebag	Neurodyne	MENRT	CNRS	16/09/03 - 15/09/06	5017
ACI-NIM	M. Sebag	Traffic	MENRT	CNRS	01/07/04 - 01/07/07	17500
PI	A. Cornuéjols	Robea		CNRS	06/12/01 - 05/12/02	1 5 2 4
PI	M. Sebag	Robea		CNRS	30/06/03 - 30/12/04	11 500
PI-TCAN	N. Bredèche	Herobot	PIRVSEV	CNRS	05/07/04 - 05/07/07	14000
SUBV	A. Cornuéjols	Indana/F	MENRT	UPS	13/09/02 - 12/02/05	10402
SUBV	A. Cornuéjols	Indana	MENRT	UPS	13/09/02 - 12/02/05	2 287
SUBV	M. Sebag	Telemedecine	MENRT	CNRS	13/09/02 - 12/02/05	12747
SUBV	Y. Kodratoff		Fondation de France FDF	UPS		9406
SUBV	A. Cornuéjols	BQR/FISICA	UPS	UPS	01/01/03 - 31/12/04	16722
SUBV	A. Cornuéjols	GIS/Cognition	CNRS	CNRS		13720
SUBV	Y. Kodratoff	Mining CVs	Vediorbis	UPS	01/10/01 - 01/10/04	15000
Industry	C. Nédellec	Astuxe	Spartcom	UPS	01/11/99 - 30/04/02	152995
AS	Y. Kodratoff	Fouille texte	STIC	CNRS	09/07/03 - 08/07/04	3 000
AS	M. Sebag	Extraction fouille	STIC	CNRS	01/10/01 - 01/10/02	17379
AS	M. Sebag C. Froidevaux	Bioinfo Apprent.	STIC	CNRS	09/07/03 - 08/07/04	4 3 7 5
Joint scient	ific projects hand	lled by a third party				
BQR	P. Tarroux	Sparse coding and Support Vector Machines for Vision of Natural Scenes	UPS	UPS	01/01/01 - 31/12/02	9146
BQR	A. Lisser	Routage robuste	UPS	UPS	01/01/04 - 01/01/05	5 000

Note: see glossary for acronyms.

## Summary of main scientific projects

The European network of excellence **PASCAL**, (6th Framework Program, 2003-2007) includes most major European Research Centers in Machine Learning, Vision, Speech and Text processing. At University Paris Sud, it involves LRI and the Mathematics Laboratory (coordinator for Paris-Sud: M. Sebag).

The Accamba and Caderige projects are related to the application of Machine Learnig techniques developed in our group to Bio-Informatics [64]. The partners are *Laboratoire Leibniz*-IMAG (UMR 5522, Grenoble)(leader), *Centre de criblage de molécules bioactives*, INSERM 366 (Grenoble), *Service de Marquage Moléculaire et de Chimie Biorganique* (CEA-Saclay),

*Service de développement pour la Bio-Informatique* (CEA-Grenoble) and laboratoire RéSO (UMR 6516 Marseille).

The **Robea** and **Herobot** projects are related to our work in Robotics [102, 61]. The partners are LIMSI-CNRS and the research team *Cognition et Usages*, Université Paris 8.

The two ACI-NIM projects, **Neurodyne** (2003-2006) and **Traffic** (2004-2007) are related to mining spatio-temporal data, respectively in Brain imagery and in Road traffic. The partners are *Laboratoire de Mathématiques* (Paris-Sud), INRETS, *Laboratoire Imagerie cérébrale et Neuro-sciences*, LENA (UPR 640).

The **INDANA** project (2001-2005) is concerned with medical data mining, and the identification of risk factors for cardio-vascular diseases. This application led to a new learning approach for sensitivity analysis [80, 81] and feature ranking [63]. The partners are INSERM, Université Paris 6, Université de Lille 3 and Universitié Technologique de Compiègne.

The **VediOrbis** and **Astuxe** projects are related to Text Mining and the development of learning-based approaches for developing textual resources and using them in Document Extraction and Text Retrieval [107, 19, 21, 96, 50, 40].

## Other contracts where I&A members are involved

The following contracts are managed by INRIA (Marc Schoenauer, leader). Two contracts are funded by the national research program "ACI-NIM" on new interfaces with Mathematics: Chimie quantique ( $8k \in$  for 3 years) and Chromalgena ( $14k \in$  for 3 years). Two contracts correspond to the funding by industry of a Ph.D. student, under 3-year CIFRE agreements: Airbus ( $15 \ k \in$ /year) and IFP ( $6k \in$ /year). Three contracts are industrial contracts with Thalès ( $50k \in$ ), EADS ( $5k \in$ ) and SNCF ( $50k \in$ ).

#### B.9.8 / I & A

## Collaborations

## **Cooperation agreements**

- Laboratoire de Mathématiques, Université Paris-Sud (ACI-NIM, Road traffic project, since 2004).
- Laboratoire LENA, UPR CNRS 640 (ACI-NIM, Brain imagery project, since 2003).
- Laboratoire d'Informatique Médicale, Hopital Broussais (ACI, Indana project, since 2000).
- Laboratoire IMAG, Grenoble (ACI-Bio, Caderidge, Astuxe, Accamba projects, since 2000).

### Collaborations with joint publications

- Università del Piemonte Orientale, Italie [60, 3].
- Vrije Universiteit Amsterdam, The Netherlands [63, 64].
- University of Leuven, Belgium [2].
- IASI group, LRI [85, 86].
- Situated Perception group (Perception Située), LIMSI-CNRS [102].

Research groups

I & A

Contracts and grants Collaborations

## **Other collaborations**

- Centre de Mathématiques Appliquées, Ecole Polytechnique.
- ERIC, Université Lyon 2.
- Grappa, Université Lille 3-INRIA Futurs.
- Laboratoire d'Acoustique Musicale, Paris-VI.
- University of Aberystwyth, UK.
- Albert-Ludwig University Friburg, Germany.
- University of Yokohama, Japan, (co-advised Ph.D. student Jérôme Maloberti).
- University of Bari, Italie.
- University Hokkaido, Japan.
- Universitade Federale do Parana, Brazil (co-advised Ph.D. student Celso. Ishida).

## B.9.9 / I & A

## Dissemination and technology transfer

## Software licensing and distribution

#### WOB

World in a Bottle. Robot simulator (C++ / OpenGL) developed by Jérémie Mary, interfaced with EO.

http://www.lri.fr/~mary/WoB/

#### **BioRelief**

Software for feature selection in the context of bioinformatics (C++ / OpenGL) developed by Jérémie Mary.

http://www.lri.fr/chris/bioinfo/BioRelief

#### Django

Fast theta-subsumption operator (C++), developed by Jérôme Maloberti. http://www.lri.fr/~malobert/

### Popularisation of research results

- Interview, Michèle Sebag. Automates Intelligents http://www.automatesintelligents.com/interviews/2002/oct/sebag.html.
- Interview, Nicolas Bredèche. *Sciences et Avenir,* special issue *Des robots et des hommes,* no.655, september 2001.

## Summer schools, tutorials, invited seminars

- Summer School Non parametric learning, CISM, Udine Italy, 2002, Michèle Sebag.
- Summer School I3, Information, Interaction, Intelligence, 2001, Michèle Sebag.
- Tutorial, Extraction et Gestion de Connaissances, EGC 2003, Michèle Sebag.
- Invited seminars, Michèle Sebag: Université Lille 2001, 2003, 2004; University of Leuven,

2002; Royal Stockholm Institute, 2003; University of Hokkaido, 2003; Laboratoire de Physique Théorique et Modèles Statistiques, Université Paris-Sud, 2004.

• Invited seminar, Nicolas Bredèche: Dipartimento Di Informatica, Università di Torino (Italy), 2002.

#### B.9.10 / I & A

## Training and education (doctoral and post-doctoral)

Defended habilitations					
Name	Date defended	Current position			
ROUVEIROL Céline	22/12/00	BioInfo research group, LRI			

Defended doctorates						
Name	Date defended	Current position				
TORRE Fabien	28/01/00	Assistant Professor, Université Lille 3				
ALES BIANCHETTI Jacques	17/07/00	Assistant Professor, Université Lille 3				
AUGIER Sébastien	19/12/00	Assistant Professor, ESIEA				
FAURE David	20/12/00	engeneer, Thalès				
ALPHONSE Erick	11/07/03	post-doc, INRA				
AZE Jérôme	16/12/03	ATER, Université Paris-Sud				

## Graduate courses

- DEA I3, Data mining and machine learning : Michèle Sebag, Antoine Cornuéjols, Céline Rouveirol.
- DEA I3, Text mining: Yves Kodratoff, Claire Nédellec.
- DEA ECD, Foundations of data mining: Yves Kodratoff, director.
- DEA ECD, Optimisation for data mining: M. Sebag.

## Other teaching-related activities

- Summer School, CISM Udine Italy: Michèle Sebag, 2001.
- EvoNet Summer School, Thessaloniki, Greece: Michèle Sebag, 2001.
- EvoNet Summer School, Szeged, Hungary: Michèle Sebag, 2002.
- Stages d'option, Ecole Polytechnique: Michèle Sebag, 2000, 2001, 2003, 2004.
- Summer School, Fast Modelling in Numerical Engineering, Ecole Polytechnique: Michèle Sebag, 2000.
- Collège de l'École Polytechnique, Algorithmes Evolutionnaires: Michèle Sebag, 2000.

Research groups

1 & A

Dissemination and technilogy transfer Training and education

## Publications

## International peer-reviewed journals

- [1] J. Ales Bianchetti. a machine learning approach to analogical reasonning. *Applied Artificial Intelligence*, 4:16-24, 2001. Expert Update, ECCAI, issn 1465-4091.
- [2] H. Blockeel and M. Sebag. Scalability and efficiency in multi-relational data mining. ACM SIGKDD, Special Issue on Multi-Relational Data Mining, 5(1):17-30, 2003.
- [3] M. Botta, A. Giordana, L. Saitta, and M. Sebag. Relational learning as search in a critical region. *Journal of Machine Learning Research*, 4:431-463, 2003.
- [4] N. Bredeche, Z. Shi, and J.-D. Zucker. Perceptual learning and abstraction in machine learning : An application to autonomous robotics. *accepted in IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews.*
- [5] A. Cornuéjols, A. Tiberghien, and G. Collet. A new mechanism for transfer between conceptual domains in scientific discovery and education. *Foundations of Science*, 5(2):129-155, 2000.
- [6] H. Hamda, F. Jouve, E. Lutton, M. Schoenauer, and M. Sebag. Compact unstructured representations in evolutionary topological optimum design. *International Journal of Applied Intelligence*, 16:139-155, 2001.
- [7] E. Lepage, E. Marguet, C. Geslin, O. Matte-Tailliez, W. Zillig, P. Forterre, and P. Tailliez. Molecular diversity of new thermococcale isolates from hydrothermal deep-sea vents as revealed by RAPD, 16S rDNA and 16S-23S rDNA intergenic spacer regions analysis. *Applied and Environmental Microbiology*, 70:1277-1286, 2004.
- [8] J. Maloberti and M. Sebag. Fast theta-subsumption with constraint satisfaction algorithms. *Machine Learning Journal*, 55:137-174, 2004, Kluwer Academic Publishers.
- [9] O. Matte-Tailliez, C. Brochier, P. Forterre, and H. Philippe. Archaeal phylogeny based on ribosomal proteins. *Molecular Biology and Evolution*, 19:631-639, 2002.
- [10] O. Matte-Tailliez, E. Lepage, M. Tenenhaus, and P. Tailliez. Use of predictive modeling for bacterial strain classification. *Systematic and Applied Microbiology*, 25:386-395, 2002.
- [11] O. Matte-Tailliez, P. Quénée, J. V. Opstal, F. Dessevre, O. Firmesse, and P. Tailliez. Detection and identification of lactic acid bacteria in milk and industrial starter culture with fluorescently labeled rDNA-targeted peptide nucleic acid probes. *Lait*, 81:237-248, 2001.
- [12] G. Mercier, N. Berthault, J. Mary, A. Antoniadis, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. Biological detection of low radiation by combining results of two analysis methods. *Nucleic Acids Research (NAR)*, 32(1):1-8, 2004.
- [13] A. Ratle and M. Sebag. Grammar-guided genetic programming and dimensional consistency: Application to non-parametric identification in mechanics. *Applied Soft Computing*, 8:1-14, 2001, Elsevier Science Publishers.
- [14] M. Sebag and C. Rouveirol. Resource-bounded relational reasoning: Induction and deduction through stochastic matching. *Machine Learning Journal*, 38:41-62, 2000, Kluwer Academic Publishers.

## National peer-reviewed journals

- [15] J. Azé. Une nouvelle mesure de qualité pour l'extraction de pépites de connaissances. *Revue d'Intelligence Artificielle*, 17(1-2-3):171-182, 2003.
- [16] J. Azé, S. Guillaume, and P. Castagliola. évaluation de la Résistance au Bruit de quelques Mesures Quantitative. *Revue des Nouvelles Technologies de l'Information, numéro spécial Entreposage et Fouille de données*, 1:159-170, 2003.
- [17] J. Azé and Y. Kodratoff. Evaluation de la résistance au bruit de quelques mesures d'extraction de règles d'association. *Revue d'Intelligence Artificielle, numéro spécial Extraction des connaissances et apprentissage (EGC-02)*, 1(4):143-154, 2002.
- [18] J. Azé and Y. Kodratoff. Extraction de pépites de connaissance dans les données : une nouvelle approche et une étude de la sensibilité au bruit. *Revue des Nouvelles Technologies de l'Information, numéro spécial Mesures de qualité pour la fouille de données*, pages 247-270, 2004.
- [19] L. Fontaine and Y. Kodratoff. La notion de concept dans les textes spécialisés: une étude comparative entre la progression thématique et la texture des concepts. In *Revue Anglais Spécialité (ASp)* 37-38, 2002.
- [20] H. Hamda, F. Jouve, E. Lutton, M. Schoenauer, and M. Sebag. Représentations non structurées en optimisation topologique de formes par algorithmes évolutionnaires. *ESAIM:Proc*, Actes du 32e Congrés national d'analyse numérique, 2000.
- [21] Y. Kodratoff. Induction extensionnelle: définition et application à l'acquisition de concepts à partir de textes. In *Revue des Nouvelles Technologies de l'Informatique*, volume 2, 2004.
- [22] Y. Kodratoff, J. Azé, M. Roche, and O. Matte-Tailliez. Des textes aux associations entre les concepts qu'ils contiennent. *Revue des Nouvelles Technologies de l'Information, numéro* spécial Entreposage et Fouille de données, 1:171-182, 2003.
- [23] I. C. Lerman and J. Azé. Une mesure probabiliste contextuelle discriminante de qualité des règles d'association. *Revue d'Intelligence Artificielle*, 17(1-2-3):247-262, 2003.
- [24] I. C. Lerman and J. Azé. Indice probabiliste discriminant (de vraisemblance du lien) d'une règle d'association en cas de très grosses données. *Revue des Nouvelles Technologies de l'Information (RNTI-2), numéro spécial Mesures de qualité pour la fouille de données*, 1:69-94, 2004.

## **Books**

[25] A. Cornuéjols and L. Miclet. Apprentissage artificiel. Concepts et algorithmes. Eyrolles, 2002.

## **Edited books**

- [26] H. Briand, M. Sebag, R. Gras, and F. G. (eds). *Numéro spécial Mesures de qualité pour la fouille de données*, volume 1. 2004.
- [27] C. Cardie, W. Daelemans, C. Nédellec, and E. Tjong Kim Sang. *Proceedings of the Fourth Conference on Computational Natural Language Learning and of the Second Learning Language in Logic Workshop.* Omni Press, septembre 2000.
- [28] E. Diday, Y. Kodratoff, P. Brito, and M. Moulet. *Induction symbolique-numérique à partir de données*. Editions Cépaduès, 2000.
- [29] M.-S. Hacid, Z. W. Ras, D. A. Zighed, and Y. Kodratoff, editors. *Foundations of Intelligent Systems*, volume 2366 of *Lecture Notes in Computer Science*. Springer-Verlag, 2002.

- [30] C. Rouveirol and M. Sebag. Proceedings of Eleventh International Conference on Inductive Logic Programming, volume 2157 of Lecture Notes in Artificial Intelligence. Springer-Verlag, 2001.
- [31] C. Rouveirol and M. Sebag, editors. *Machine Learning Journal, Special Issue on Inductive Logic Programming and Relational Learning*, volume 55. Kluwer Academic Publishers, 2004.
- [32] S. Stabb, A. Maedche, C. Nédellec, and E. Hovy. IJCAI Workshop Notes of the Ontology Learning, workshop of the 17<sup>th</sup> International Joint Conference on Artificial Intelligence (IJCAI). 2001.
- [33] S. Stabb, A. Maedche, C. Nédellec, and P. Wiemer-Hastings. *ECAI Workshop Notes of the Ontology Learning, workshop of the European Conference on Artificial Intelligence (ECAI)*. août 2000.

## **Book chapters**

- [34] A. Cornuéjols. Machine learning: The necessity of order. In E. Lehtinen and F. Ritten, editors, *In Order to Learn: How ordering processes and sequencing effects in machines illuminate human learning and vice-versa*. Cambridge University press, 2005.
- [35] D. Kayser, S. Vosniadou, C. Nédellec, A. Tiberghien, and J.-D. Zucker. *Introduction*, chapter Knowledge Representation Changes in Humans and Machines. Elsevier Science Publishers, 2000.
- [36] Y. Kodratoff. Induction symbolique-numérique à partir de données : des outils pour le data mining, chapter L'induction symbolique et numérique en ECD. Editions Cépaduès, 2000.
- [37] Y. Kodratoff. *Machine Learning and its applications*, volume 2049 of *Lecture Notes in Computer Science*, chapter Comparing Machine Learning and Knowledge Discovery in DataBases :An Application to Knowledge Discovery in Texts, pages 1-21. Springer-Verlag, 2001.
- [38] J. Mary, G. Mercier, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. An attribute estimation technique for the analysis of microarray data. In P. Amar, F. Képès, V. Norris, and P. Tracqui, editors, *Proceedings of the Dieppe School on Modelling and Simulation of Biological processes in the Context of Genomics*, pages 69-77. Publisher Frontier Group, 2003.
- [39] J. Mary, G. Mercier, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. Utilisation d'une méthode d'estimation d'attributs pour l'analyse du transcriptome de cellules de levures exposées à de faibles doses de radiation. In J.-F. Boulicaut and O. Gandrillon, editors, *Informatique pour l'analyse du transcriptome*, pages 189-205. Hermès, 2004.
- [40] C. Nédellec. *Corpus-based learning of semantic relations by the ILP system, Asium,* chapter Learning Language in Logic. Springer-Verlag, october 2000.
- [41] C. Nédellec and A. Tiberghien. *Teaching the energy concept with the Machine Learning,* chapter Knowledge Representation Changes in Humans and Machines. Elsevier Science Publishers, 2000.
- [42] M. Schoenauer and M. Sebag. Evolution artificielle et problèmes spatiaux. In H. Prade, R. Jeansoulin, and C. Garbay, editors, *Le temps, l'espace et l'évolutif en sciences du traitement de l'information*, pages 377-396. Cépaduès, 2000.
- [43] M. Schoenauer and M. Sebag. Using Domain Knowledge in Evolutionary System Identification. In K. Giannakoglou, editor, *Evolutionary Algorithms in Engineering and Computer Science*. John Wiley, 2002.
- [44] M. Sebag. *Principles of non-parametric learning*, chapter Genetic Programming for Model Identification, pages 287-354. Springer Wien New-York, 2002.
- [45] M. Sebag and P. Gallinari. Apprentissage artificiel : Acquis, limites, enjeux. In J. L. Maitre, editor, *Assises 2002 : Information Interaction Intelligence*. Cépadués, 2002.

#### Major international peer-reviewed conferences

- [46] J. Ales-Bianchetti, C. Rouveirol, and M. Sebag. Constraint-based learning of long relational concepts. In C. Sammut, editor, 19th International Conference on Machine Learning (ICML 2002), pages 35-42. Morgan Kaufmann, 2002.
- [47] E. Alphonse and S. Matwin. A dynamic approach to dimensionality reduction in relational learning. In 13th International Symposium on Methodologies for Intelligent Systems (ISMIS 2002), pages 255-264. Springer-Verlag, 2002.
- [48] E. Alphonse and S. Matwin. Feature subset selection and inductive logic programming. In C. Sammut and A. Hoffmann, editors, 19th International Conference on Machine Learning (ICML'02), pages 11-18, Sydney Australia, 2002. Morgan Kaufmann.
- [49] E. Alphonse and C. Rouveirol. Lazy propositionalisation for relational learning. In W. Horn, editor, 14th European Conference on Artificial Intelligence (ECAI 2000), pages 256-260. IOS Press, 2000.
- [50] A. Amrani, Y. Kodratoff, and O. Matte-Tailliez. A semi-automatic system for tagging specialized corpora. In H. Dai, R. Srikant, and C. Zhang, editors, 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2004), volume 3056 of Lecture Notes in Artificial Intelligence, pages 670-681. Springer-Verlag, 2004.
- [51] A. Auger, N. Vanhaecke, and M. Schoenauer. LS-CMA-ES: Better Covariance Matrix Adaptation using Least Square Approximation. In 8th International Conference on Parallel Problems Solving from Nature (PPSN 2004), LNCS, pages 182-191. Springer-Verlag, 2004.
- [52] J. Azé and Y. Kodratoff. A study of the Effect of Noisy Data in Rule Extraction Systems. In 16th European Meeting on Cybernetics and Systems Research (EMCSR 2002), volume 2, pages 781-786, 2002.
- [53] N. Baskiotis and M. Sebag. C4.5 competence map: a phase transition-inspired approach. In R. Greiner and D. Schuurmans, editors, *21st International Conference on Machine Learning, ICML 2004*, pages 73-80. Morgan Kaufmann, 2004.
- [54] M. Botta, A. Giordana, L. Saitta, and M. Sebag. Relational learning: Hard problems and phase transition. In *Selected papers from AIIA'99*, volume 1792 of *Lecture Notes in Artificial Intelligence*, pages 178-189. Springer-Verlag, 2000.
- [55] I. Bournaud, M. Courtine, and J. Zucker. Abstractions for knowledge organization of relational descriptions. In *Symposium on Abstraction, Reformulation and Approximation, SARA2000,* volume 1864 of *Lecture Notes in Artificial Intelligence*, pages 87-106. Springer-Verlag, 2000.
- [56] I. Bournaud, M. Courtine, and J. Zucker. Propositionalization for clustering symbolic relational descriptions. In 12th International Conference on Inductive Logic Programming (ILP 2002), Lecture Notes in Artificial Intelligence. Springer-Verlag, 2002.
- [57] M. Courtine and I. Bournaud. Building a pruned inheritance lattice structure for relational description. In *Workshop on Concept Lattice-based theory, methods and tools for Knowledge Discovery in Databases, International Conference on Conceptual Structures, ICCS2001,* 2001.
- [58] M. Courtine, I. Bournaud, and J. Zucker. Kids: an iterative algorithm to organize relational knowledge. In O.Corby and R. Ding, editors, 12<sup>th</sup> International Conference on Knowledge Engineering and Knowledge Management (EKAW 2000), volume 1937 of Lecture Notes in Artificial Intelligence, pages 217-232. Springer-Verlag, 2000.
- [59] M. Franova and X. Popelinsky. Synthesis of formal specifications of predicates: Why and how? In R. Trappl, editor, *15th European Meeting on Cybernetics and Systems Research*, volume II, pages 739-744. Austrian Society for Cybernetic Studies, 2000.
- [60] A. Giordana, L. Saitta, M. Sebag, and M. Botta. Analyzing relational learning in the phase transition framework. In P. Langley, editor, 17<sup>th</sup> International Conference on Machine Learning (ICML 2000), pages 311-318. Morgan Kaufmann, 2000.

Laboratoire de Recherche en Informatique / 2000-2004 report

Publications

Research groups

1 & A

- [61] N. Godzik, M. Schoenauer, and M. Sebag. Evolving symbolic controllers. In G. Raidl et al., editor, *Applications of Evolutionary Computing*, volume 2611 of *Lecture Notes in Computer Science*, pages 638-650. Springer-Verlag, 2003.
- [62] N. Godzik, M. Schoenauer, and M. Sebag. Robustness in the long run: Auto-teaching vs anticipation in evolutionary robotics. In X. Yao et al., editor, *8th International Conference on Parallel Problem Solving from Nature (PPSN 2004)*, pages 932-941. Springer-Verlag, 2004.
- [63] K. Jong, E. Marchiori, and M. Sebag. Ensemble Learning with Evolutionary Computation: Application to Feature Ranking. In X. Yao et al., editor, 8<sup>th</sup> International Conference on Parallel Problem Solving from Nature (PPSN 2004), Lecture Notes in Computer Science, pages 1133-1142. Springer-Verlag, 2004.
- [64] K. Jong, J. Mary, A. Cornuéjols, E. Marchiori, and M. Sebag. Ensemble feature ranking. In 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD-04), pages 267-278, Pisa, Italy, 2004. Springer-Verlag.
- [65] C. Kavka and M. Schoenauer. Evolution of voronoi-based fuzzy controllers. In 8<sup>th</sup> International Conference on Parallel Problem Solving From Nature (PPSN 2004), pages 541-550. Springer-Verlag, 2004.
- [66] J. Maloberti and M. Sebag. Theta-subsumption in a constraint satisfaction perspective. In 11th International Conference on Inductive Logic Programming (ILP 2001), volume 2157 of Lecture Notes in Artificial Intelligence, pages 164-178. Springer-Verlag, 2001.
- [67] J. Maloberti and E. Suzuki. Improving efficiency of frequent query discovery by eliminating non-relevant candidates. In 6<sup>th</sup> International Conference on Discovery Science (DS 2003), volume 2843 of Lecture Notes in Artificial Intelligence, pages 219-231. Springer-Verlag, 2003.
- [68] J. Maloberti and E. Suzuki. An efficient algorithm for reducing clauses based on constraint satisfaction techniques. In 14<sup>th</sup> International Conference on Inductive Logic Programming (ILP 2004), Lecture Notes in Artificial Intelligence. Springer-Verlag, 2004.
- [69] C. Nédellec. Knowledge extraction from text, a machine learning approach. In K. Zreik, editor, *3e Colloque International sur l'Apprentissage Personne Système (CAPS'3), Paris*. Europia, 2000.
- [70] C. Nédellec, M. Ould Abdel Vetah, and P. et Bessières. Sentence filtering for information extraction in genomics, a classification problem. In 5<sup>th</sup> European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2001), volume 2168 of Lecture Notes in Artificial Intelligence, pages 326-338. Springer-Verlag, 2001.
- [71] B. Paechter, T. Back, M. Schoenauer, M. Sebag, A. Eiben, J. Merelo, and T. Fogarty. A Distributed Resource Evolutionary Algorithm Machine (DREAM). In *Proceedings of the* 2000 Congress on Evolutionary Computation, pages 951-958. IEEE press, 2000.
- [72] A. Ratle and M. Sebag. Genetic programming and domain knowledge: Beyond the limitations of grammar-guided machine discovery. In M. Schoenauer et al., editor, 6<sup>th</sup> International Conference on Parallel Problem Soving from Nature (PPSN 2000), volume 1917 of Lecture Notes in Computer Science, pages 211-220. Springer-Verlag, 2000.
- [73] A. Ratle and M. Sebag. Avoiding the bloat with probabilistic grammar-guided genetic programming. In P. Collet et al., editor, *Artificial Evolution*, VI, pages 255-266. Springer-Verlag, 2001.
- [74] A. Ratle and M. Sebag. A novel approach to machine discovery: Genetic programming and stochastic grammars. In S. Matwin and C. Sammut, editors, 12th International Conference on Inductive Logic Programming (ILP 2002), volume 2583 of Lecture Notes in Computer Science, pages 207-222. Springer-Verlag, 2003.
- [75] M. Roche, J. Azé, O. Matte-Tailliez, and Y. Kodratoff. Mining texts by association rules discovery in a technical corpus. In *Conference on Intelligent Information Processing and Web Mining (IIPWM 2004)*, Advances in Soft Computing, pages 89-98. Springer-Verlag, 2004.

- [76] M. Roche, T. Heitz, O. Matte-Tailliez, and Y. Kodratoff. Exit: Un système itératif pour l'extraction de la terminologie du domaine à partir de corpus spécialisés. In *International Conference on Statistical Analysis of Textual Data (JADT 2004)*, volume 2, pages 946-956, 2004.
- [77] O. Roudenko and M. Schoenauer. Dominance based crossover operator for evolutionary multi-objective algorithms. In 8th International Conference on Parallel Problem Solving from Nature (PPSN 2004), pages 812-821. Springer-Verlag, 2004.
- [78] O. Roudenko and M. Schoenauer. A steady performance stopping criterion for paretobased evolutionary algorithms. In 6<sup>th</sup> International Multi-Objective Programming and Goal Programming Conference, 2004.
- [79] C. Rouveirol and V. Ventos. Towards learning in CARIN-ALN. In J. Cussens and A. Frisch, editors, 10th International Conference on Inductive Logic Programming (ILP 2000), volume 1866 of Lecture Notes in Artificial Intelligence, pages 191-208. Springer-Verlag, 2000.
- [80] M. Sebag, J. Azé, and N. Lucas. Impact studies and sensitivity analysis in medical data mining with roc-based genetic learning. In *IEEE International Conference on Data Mining*, *ICDM03*, pages 637-640, 2003.
- [81] M. Sebag, J. Azé, and N. Lucas. Roc-based evolutionary learning: Application to medical data mining. In *Artificial Evolution'03*, Lecture Notes in Computer Science, pages 384-396. Springer-Verlag, 2004.
- [82] Y. Semet, U.-M. O'Reilly, and F. Durand. An interactive artificial ant approach to nonphotorealistic rendering. In Kalyanmoy Deb et al., editor, *Proceedings of the Genetic and Evolutionary COmputation Conference (GECCO'04)*, volume 1 of *LNCS 3102*, pages 188-200, Seattle, WA, USA, June 26-30 2004. Springer-Verlag.
- [83] Y. Semet, S.Gelly, M. Schoenauer, and M. Sebag. Artificial agents and speculative bubbles. In Wessex Institute of Technology, editor, *Computational Finance and its Applications*, pages 35-44, Southampton, UK, 2004. WIT Press.
- [84] P. Tailliez and O. Matte-Tailliez. Characteristics of DNA sequences used as targets in whole-cell hybridization determined and predicted by PLS models. In *International Conference on Partial Least Square (PLS'03), Lisbon, Portugal, Lisbon, Portugal, 15-17* September 2003.
- [85] A. Termier, M.-C. Rousset, and M. Sebag. Treefinder: a first step towards xml data mining. In *IEEE International Conference on Data Mining, ICDM02*, pages 450-457, 2002.
- [86] A. Termier, M.-C. Rousset, and M. Sebag. Dryade: a new approach for discovering closed frequent trees in heterogeneous tree databases. In *IEEE International Conference on Data Mining, ICDM04*, 2004.

## National peer-reviewed conferences

- [87] J. Ales-Bianchetti, C. Rouveirol, and M. Sebag. Apprentissage contraint de longs concepts relationnels. In *Actes de la Conférence d'Apprentissage (CAp2002)*, pages 1-12, 2002.
- [88] E. Alphonse. Sélection d'attributs pour les systèmes de pli ascendants. In G. Bisson, editor, *Actes de la Conférence d'Apprentissage (CAP'01)*, pages 11-20. Hermès, 2001.
- [89] A. Amrani, Y. Kodratoff, and O. Matte-Tailliez. Etiq, un étiqueteur inductif convivial pour les corpus de spécialité. In *Revue des Nouvelles Technologies de l'Information, numéro spécial Entreposage et Fouille de Données (EGC-2004)*, volume 2, page 477. Cépaduès, janvier 2004.
- [90] A. Amrani, Y. Kodratoff, and O. Matte-Tailliez. Un système dédié à l'étiquetage morphosyntaxique des corpus de spécialité. In *Colloque International sur la Fouille de Texte (CIFT 2004)*, 22-24 juin 2004.

Research groups

Publications

- [91] J. Azé, N. Lucas, and M. Sebag. Fouille de données visuelle et analyse de facteurs de risque médical. *Revue d'Intelligence Artificielle, numéro spécial Extraction et Gestion des Connaissances (EGC-03)*, 17:183-188, 2003.
- [92] J. Azé, H. Richy, and G. Lorette. Interface-stylo de correction en ligne de documents électroniques: Application aux pages web. In *Colloque International sur l'Ecrit et le Document (CIFED'2000)*, pages 261-269, 2000.
- [93] J. Azé and M. Roche. Une application de la fouille de textes : l'extraction des règles d'association à partir d'un corpus spécialisé. *Revue d'Intelligence Artificielle, numéro spécial Extraction et Gestion des Connaissances (EGC-03)*, 17:283-294, 2003.
- [94] N. Baskiotis and M. Sebag. Carte de competence de C4.5 : une approche inspiree de la transition de phase. In *Actes de la Conference d'Apprentissage, CAp04,* 2004.
- [95] G. Bisson and C. Nédellec. Aide à la conception de méthode de classification pour la construction d'ontologies : l'atelier mo'k. In H. Briand, editor, *Conférence Extraction et Gestion des Connaissances (EGC'2001)*, pages 213-226. Hermès, janvier 2001.
- [96] G. Bisson, C. Nédellec, and D. Cañamero. Designing clustering methods for ontology building - the mo'k workbench. In S. Stabb, A. Maedche, and C. Nédellec, editors, ECAI Workshop Notes of the Ontology Learning, workshop of the 14<sup>th</sup> European Conference on Artificial Intelligence (ECAI 2000), Berlin, Germany, pages 13-19, août 2000.
- [97] I. Bournaud and M. Courtine. Kids : un algorithme itératif pour l'organisation de données relationnelles. In *Actes de la Conférence d'Apprentissage, CAP2000,* pages 47-56, 2000.
- [98] I. Bournaud and M. Courtine. Un espace de généralisations pour l'extraction de règles d'association. *Extraction des Connaissances et Apprentissage, numéro spécial Extraction et Gestion des Connaissances (EGC 2001)*, 1(1-2):123-135, 2001, Hermès.
- [99] A. Cornuéjols and L. Miclet. Introduction à l'apprentissage artificiel. In P. Martinet, editor, *Journées Nationales de Recherche en Robotique (JNNR-03)*, pages 121-126, Murol, France, 2003.
- [100] D. Faure and T. Poibeau. Extraction d'information utilisant intex et des connaissances sémantiques apprises par Asium, premières expérimentations. In F. Schmitt and I. Bloch, editors, 12ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA 2000), Paris, pages 91-100, February 2000.
- [101] A. Giordana, L. Saitta, M. Sebag, and M. Botta. La programmation logique inductive à la lumière de la transition de phase. In C. de la Higuera, editor, *Actes de la Conférence d'Apprentissage, CAP2000*, pages 157-172. Hermès, 2000.
- [102] S. Jouteau, A. Cornuéjols, M. Sebag, P. Tarroux, and J.-S. Liénard. Nouveaux résultats en classification à l'aide d'un codage par motifs fréquents. *Revue d'Intelligence Artificielle, numéro spécial Conférence Extraction et Gestion des Connaissances, EGC-2003*, 17(1-3):521-532, 2003.
- [103] Y. Kodratoff. Applications de l'apprentissage automatique et de la fouille de données. In *Conférence Extraction et Gestion des Connaissances (EGC'2001)*, pages 57-68, janvier 2001.
- [104] C. Nédellec and A. Nazarenko. Application de l'apprentissage à la recherche et à l'extraction d'information - un exemple, le projet caderige : identification d'interactions géniques. In Actes de la Journée thématique Exploration de données issues d'Internet, LIPN-CNRS, Institut Galilée, Université Paris 13, Villetaneuse, 2001.
- [105] M. Roche. Extraction paramétrée de la terminologie du domaine. *Revue d'Intelligence Artificielle, numéro spécial Extraction et Gestion des Connaissances (EGC-03),* 17:295-306, 2003.
- [106] M. Roche and Y. Kodratoff. Utilisation de LSA comme première étape pour la classification des termes d'un corpus spécialisé. In Actes (CD-ROM) de la conférence MAJECSTIC'03 (MAnifestation des JEunes Chercheurs dans le domaine STIC), 2003.

- [107] M. Roche, O. Matte-Tailliez, J. Azé, and Y. Kodratoff. Extraction de la Terminologie du Domaine : étude de Mesures sur un Corpus Spécialisé Issu du Web. In Actes des Journées Francophones de la Toile (JFT 2003)), pages 279-288, 2003.
- [108] M. Roche, O. Matte-Tailliez, and Y. Kodratoff. Etude de Mesures de Qualité pour Classer les Termes Extraits de Corpus Spécialisés. In *Actes de INFORSID'04*, pages 371-386, 2004.
- [109] M. Sebag, J. Azé, and N. Lucas. Apprendre et optimiser la courbe roc. In *Actes de la Conférence d'Apprentissage (CAp 2003)*, pages 315-330, 2003.

## Other conferences and workshops

- [110] E. Alphonse and S. Matwin. Feature selection for bottom-up inductive logic programming learners. In *International Conference on Inductive Logic Programming (ILP 2001), Workin-progress track*, pages 1-12, 2001.
- [111] E. Alphonse and N. Stroppa. Filtering multi-instance problems for feature selection in ilp. In *International Conference on Inductive Logic Programming (ILP 2002), Work-in-progress track,* 2002.
- [112] A. Amrani, Y. Kodratoff, and O. Matte-Tailliez. Etiquetage morpho-syntaxique de corpus de spécialité. In *Colloque Jeunes Chercheurs (Modyco)*, 29-30 avril 2004.
- [113] J. Azé, N. Lucas, and M. Sebag. A new medical test for atherosclerosis detection : Geno. In *Discovery Challenge PKDD 2003*, september 2003.
- [114] J. Azé, N. Lucas, and M. Sebag. A genetic roc-based classifier. In *Physiological Data Modeling Contest, ICML 2004, Canada*, July 2004.
- [115] A. Cornuéjols, M. Sebag, and J. Mary. Classification d'images à l'aide d'un codage par motifs fréquents. In *Workshop sur la fouille d'images (RFIA-04)*, Toulouse (France), 2004.
- [116] M. Courtine and I. Bournaud. Propositionalization for clustering symbolic relational descriptions. In *Advances in Formal Concept Analysis for Knowledge Discovery in Databases, FCAKDD2002, Workshop at the 15th European Conference on Artificial Intelligence, ECAI2002,* 2002.
- [117] M. Dutreix, J.-P. Comet, A. Cornuéjols, and C. Froidevaux. Determination of cellular drug targets: searching for functional information in the jungle of microarrays data. In *Current Trends in Drug Discovery Research (CTDDR-04)*, India, 2004.
- [118] D. Faure and T. Poibeau. First experiments of using semantic knowledge learned by asium for information extraction task using intex. In S. Staab, A. Maedche, and C. Nédellec, editors, *Proceedings of ECAI-2000 Ontology Learning Workshop*, pages 7-12. AAAI Press, August 2000.
- [119] Y. Kodratoff, O. Matte-Tailliez, M. Roche, J. Azé, and A. Amrani. Extraction de règles d'association de concepts à partir de différents corpus spécialisés. Session Poster de la plateforme AFIA 2003, Juillet 2003.
- [120] Y. Kodratoff, O. Matte-Tailliez, M. Roche, J. Azé, and A. Amrani. Une chaine complète pour le traitement de quantités de textes : des textes bruts a l'extraction de l'information. In *Actes de l'atelier Acquisition, apprentissage et exploitation de connaissances sémantiques pour l'accès au contenu textuel à la plateforme AFIA 2003, 4 juillet 2003, Laval, 2003.*
- [121] N. Lucas, J. Azé, and M. Sebag. Atherosclerosis risk identification and visual analysis. In Discovery Challenge PKDD 2002, august 2002.
- [122] O. Matte-Tailliez. The Bio-ontologies: an analysis of ontologies effectively used in biology. In *NATO Workshop on Information Exchange for Coalition Operations*. NATO Research and Technology Organization, 24-25 November 2003.

Research groups

- [123] O. Matte-Tailliez. Création d'ontologies en biologie et leur utilisation en extraction d'information. In *Journées Ontologies et Méta-données*, 11 juin 2003.
- [124] O. Matte-Tailliez and A. Amrani. Processus de fouille de textes: normalisation et étiquetage, action et rétroaction. In *Actes de l'atelier Fouille de Textes à la conférence EGC'04*, 2004.
- [125] O. Matte-Tailliez, M. Roche, J. Azé, and Y. Kodratoff. Extraction automatique de la terminologie en biologie moléculaire: une étape fondamentale pour l'extraction d'information. In Atelier Fouille de textes en génomique à EGC'2003, Lyon, Janvier 2003.
- [126] O. Matte-Tailliez, M. Roche, Y. Kodratoff, and M. Termier. Semi-automatic construction of a taxonomy for a biological field. In *IEEE Computer Society Bioinformatics Conference*, 14-16 August 2002.
- [127] N. Nazarenko and C. Nédellec. Extraction de connaissances à partir de documents textuels. In *Journée thématique Exploration de données issues d'Internet*, Mars 2001.
- [128] N. Nazarenko and C. Nédellec. *Journée A3CTE, Applications, Apprentissage et Acquisition de Connaissances a partir de Textes Electroniques.* 2001.
- [129] M. Roche. Utilisation de mesures pour extraire la terminologie de corpus spécialisés de domaines différents. In *Actes de l'atelier Fouille de Textes à la conférence EGC'04*, 2004.
- [130] M. Roche, J. Azé, Y. Kodratoff, and M. Sebag. Learning Interestingness Measures in Terminology Extraction. A ROC-based approach. In Workshop ROC Analysis in AI, 16<sup>th</sup> European Conference on Artificial Intelligence (ECAI 2004), pages 81-88, 21 august 2004 2004.

## **Dissemination of research results**

- [131] A. Cornuéjols. Une introduction aux SVM. Bulletin de l'AFIA (Association Française d'Intelligence Artificielle), 51(Juin, 2002):14-23, 2002.
- [132] P. Forterre, Y. Zivanovic, O. Matte-Tailliez, F. Matsunaga, and H. Myllykallio. Ce qui est vrai pour l'archaea est (presque) vrai pour l'éléphant! *Médecine/Sciences*, 17:939-945, 2001.
- [133] O. Matte-Tailliez. Biologistes, à vos textes. *Plein Sud, le Magazine d'Information de l'Université Paris-Sud*, 57:27-28, Avril-Mai 2004.
- [134] O. Matte-Tailliez. Vers une nouvelle forme de recherche en biologie. In *Forum des Jeunes Mathématiciennes: Informatique, Mathématiques et Sciences du Vivant*, Paris, France, 31 mai 2004.

## **Doctoral dissertations and Habilitations**

- [135] J. Ales-Bianchetti. Le raisonnement par analogie. Une unification des modèles cognitifs et des théories de l'induction pour l'étude du raisonnement pas analogie. Thèse de doctorat, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, july 2000.
- [136] E. Alphonse. *Macro-opérateurs et Sélection Relationnelle en Programmation Logique Inductive: théorie et algorithmes*. Thèse de doctorat, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, juillet 2003.
- [137] S. Augier. *Apprentissage Supervisé Relationnel par Algorithmes d'Évolution*. Thèse de doctorat, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, 2000.
- [138] J. Azé. Extraction de Connaissances dans des Données Numériques et Textuelles. Thèse de doctorat, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, decembre 2003.

- [139] D. Faure. Conception de méthode d'apprentissage symbolique et automatique pour l'acquisition de cadres de sous-catégorisation de verbes et de connaissances sémantiques à partir de textes : le système ASIUM. Thèse de doctorat, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, décembre 2000.
- [140] C. Rouveirol. *Expressivité / Efficacité : le Dilemme de la Programmation Logique Inductive*. Mémoire d'habilitation, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, décembre 2000.
- [141] O. Rudenko. *Application des algorithmes évolutionnaires aux problèmes d'optimisation multiobjectif avec contraintes*. Thèse de doctorat, École Polytechnique, 2004.
- [142] F. Torre. Intégration des biais de langage à l'algorithme générer-tester. Contribution à l'apprentissage disjonctif. Thèse de doctorat, Laboratoire de Recherche en Informatique, Université Paris-Sud, France, 2000.

### Reports

- [143] O. Matte-Tailliez, M. Roche, and Y. Kodratoff. A precise automatic extraction of terminology in genomics, 2002.
- [144] M. Roche. Astuxe : Analyse des causes d'erreur de l'analyse syntaxique dans les documentations logicielles. Research Report Doc.Ast.LRI.4.2b, Projet RNRT Astuxe, 2001.
- [145] M. Roche. Astuxe : Etude de l'influence de différents paramètres sur l'apprentisssage de connaissances du domaine. Research Report Doc.Ast.LRI.4.3f, Projet RNRT Astuxe, 2001.
- [146] M. Roche. Astuxe : Evaluation de la qualité de l'apprentissage sur le corpus de mails Microsofts. Research Report Doc.Ast.LRI.4.5b, Projet RNRT Astuxe, 2001.
- [147] M. Roche. Astuxe : Evaluation de la qualité de l'apprentissage sur le corpus de mails Prologue. Research Report Doc.Ast.LRI.4.5a, Projet RNRT Astuxe, 2001.
- [148] Y. Semet and M. Schoenauer. Optimisation évolutionnaire pour la régulation des circulations. Technical report, Collaboration SNCF/INRIA, Juin 2004.
- [149] Y. Semet, S.Gelly, M. Schoenauer, and M. Sebag. Agent-based computational economics and finance: early research and design issues. Economic Modeling report, DREAM project, 2003.

## Other publications

[150] C. Buty and A. Cornuéjols. Evolution des connaissances chez l'apprenant. In A. Tiberghien, editor, *Des connaissances naives aux savoirs scientifiques*, pages 41-66. Rapport au Ministère de l'Education Nationale, 2002.

Research groups

Publications



Laboratoire de Recherche en Informatique / 2000-2004 report



# équipe Bioinformatique

Responsable : Christine Froidevaux



L'équipe réunit en une fructueuse synergie des compétences en Algorithmique et Combinatoire, en Intégration de sources de données, en Représentation des Connaissances, en Apprentissage et Extraction de Connaissances dans les Bases de Données ainsi qu'en Interaction-Homme-Machine, qui lui permettent de développer un travail de recherche pluridisciplinaire à large spectre. L'équipe BioInfo tire sa richesse et son originalité de la diversité des champs informatiques couverts, diversité qui lui permet d'appréhender les problèmes de la génomique dans leur globalité et leur complexité.

Les travaux de l'équipe, menés en étroite collaboration avec des biologistes/bioinformaticiens (notamment à l'IGM Orsay et à l'Institut Curie), portent principalement sur l'étude des phénomènes de recodage, l'étude de la structure des molécules d'ARN, l'analyse de données de puces (transcriptome, CGH), la comparaison de réseaux d'interactions protéiques et l'annotation de données génomiques.

L'équipe est déjà bien reconnue au niveau national, comme en témoigne sa participation active à l'animation de la recherche française en bioinformatique. Au niveau international, elle est impliquée dans un projet R&D (5<sup>e</sup> PCRD) visant à la conception et au développement d'une plate-forme biomédicale intégrative.



## **Bioinformatics**

Head: Christine Froidevaux

The Biolnfo group is very recent: it was created in 2002 as a followup to the cross-group research team of bioinformatics created in 2001. Today it consists of five full-time permanent researchers, five Ph.D. students and one Post-Doc. The group was created because decoding genomes and thoroughly analysing their components and relationships raises new bioinformatics problems. Solving these problems requires a close cooperation between computer scientists and biologists, together with varied and complementary skills in computer science are required to solve these problems.

The group combines, in a fruitful synergy, skills in Algorithms, Combinatorics, Data sources Integration, Knowledge Representation, Machine Learning, Knowledge Discovery in Databases and Human-Computer Interaction, enabling it to carry out a wide range of multidisciplinary research. The originality and richness of the Biolnfo group stem from the diversity of the computer science fields it covers. The group is therefore able to tackle intricate genomics problems in a comprehensive way.

The group has worked in close collaboration with biologists, especially from *IGM* in Orsay and from the *Institut Curie*. Our work primarily addresses the prediction of recoding events, the study of RNA structure, the analysis of microarray and CGH data, the comparison of protein-protein interaction networks and the annotation of genomic data.

The Biolnfo group is well recognised at the national level, e.g., it actively participates in the direction of several French working groups on bioinformatics. At the international level, it is involved in a European Fifth Framework R&D project on the design and development of an integrative platform for biomedical data.

# Research groups Bioinfo Introduction

## **Research group members**

#### Personnel as of 01/01/2004

Full time faculty							
Name	First Name	Position*	Institution				
AMAR	Patrick	MC	Université Paris XI				
DENISE	Alain	PR2	Université Paris XI				
FROIDEVAUX	Christine	PR1	Université Paris XI				
ROUVEIROL	Céline	MC HDR	Université Paris XI				
VIALETTE	Stéphane	MC	Université Paris XI				
Doctoral students							
COHEN-BOULAKIA	Sarah	AM	Université Paris XI				
FOREST	Jean-Paul	ATER	Université Paris XI				
HERRBACH	Claire	D	Université Paris XI				
PONTY	Yann	AM	IUT d'Orsay				
RIVIERE	Romain	AM	Université Paris XI				
Temporary personnel							
BIDAULT	Alain	Post-doc	Université Paris XI				

\* See glossary for acronyms

## Group evolution: arrivals and departures

The Bioinformatics group is very recent. In April 2001, a cross-group research team on Bioinformatics was created at LRI with members from four groups:

- Four full-time faculty: Alain Denise (Algo group), Christine Froidevaux (IASI group), Claire Nédellec (I&A group) and Emmanuel Waller (BD group).
- Two doctoral students : Jean-Paul Forest and Mohamed Ould Vetah.

A year later, in September 2002, Alain Denise (MC) was appointed as Professor and the Bioinformatics group was officially started. Some researchers of the original team did not join the research group: Emmanuel Waller stayed with the BD group, Claire Nédellec and Mohamed Ould Vetah left LRI and went to INRA. However, two other members joined the group :

- Patrick Amar (from the Prog Group)
- Céline Rouveirol (from the I&A Group).

In 2003, a new assistant professor joined the group:

• Stéphane Vialette.

Céline Rouveirol is currently on leave (*délégation*) at the *Institut Curie* (Paris) from September 2003 to August 2005.
# **Research description**

The members of the BioInfo group have diverse backgrounds and this mix is a great asset of the group. It leads to original research in which each scientist benefits from the expertise of the other members of the group and it facilitates broad coverage of the topics concerned. This cross-fertilisation extends to biologists through close collaborations. Our primary collaborators in biology are at the *Institut de Génétique et de Microbiologie* (IGM, Orsay) and at the *Institut Curie* (Orsay and Paris).

Since its creation, the BioInfo group has participated in numerous national working groups on bioinformatics (IMPG, AS STIC CNRS, ACI IMPBio). In particular, the group has actively contributed to the co-ordination of the French bioinformatics community in the last three years, as its members are either co-leaders or members of the steering committees of these working groups. The BioInfo group is acquiring an international reputation through its involvement in the HKIS R&D project (5th Framework). It also has informal scientific relations with groups in Switzerland (GeneBio, Geneva), Canada (University of Montreal) and the USA (University of Pennsylvania).

The work carried out in the BioInfo group is of two types: pure computer science studies, motivated by biological problems, and bioinformatics studies more tightly focused on the analysis of biological objects themselves. We use a *multifield* approach in our bioinformatics studies, *i.e.* the biological objects are studied using complementary concepts, models, algorithms or tools from the various fields of computer science. The originality of our group lies in our efforts to study biological objects in all their aspects. For example our study of eukaryotic -1 frameshifting combines knowledge representation, data mining, algorithm design and visualisation techniques (see Theme 1). We feel that this *multifield* approach is essential to tackle intricate genomic problems in a comprehensive way.

The main biological problems addressed in the group so far are the prediction of recoding events, the study of RNA structure, the analysis of microarray and CGH array data, the comparison of protein-protein interaction networks and the annotation of genomic data. Our biologist colleagues often face these problems and new concepts and algorithms are required to tackle them. Each study is initiated in close collaboration with our biologist colleagues, who then evaluate our results (models/concepts/algorithms). This validation phase may result in *in vivo* experimentation that may lead to revision of the model or of the algorithm designed.

We will present our research work in three sections as follows. The first section deals with recoding events and illustrates the *multifield* approach introduced above. This approach also underlies the work presented in the other two sections. The second section concerns the analysis of biological sequences and structures. Several themes, theoretical to various degrees, are presented in this section: combinatorial and statistical models for sequence analysis, modelling, visualisation and algorithms for RNA structures, comparison of protein interaction networks across species and modelling and simulating biological processes. The third section concerns the analysis and storage of massive biological data. It deals with our work on the integration of heterogeneous biomedical data sources and on the extraction of knowledge from numerous biological data. Indeed, with ever-increasing amounts of disparate biomedical data, there is a clear need for interoperability between sources in bioinformatics. Biologists are now faced with the double problem of integrating information from multiple heterogeneous public sources (e.g. Internet web sites) with their own experimental data (*e.g.* results concerning gene expression) contained in individual sources and extracting relevant information from this huge amount of data.

Research groups

Bioinfo

Members Research description

# Theme 1: A multifield approach for the prediction of a recoding event

#### • Participants: Christine Froidevaux, Patrick Amar, Alain Denise, Jean-Paul Forest

Genetic information is passed on from DNA to RNA (transcription) and then to protein (translation). The genetic information contained in the RNA molecule is transformed into the corresponding protein by the ribosome. Translation begins at an initiation codon followed by the decoding of a sequence of nucleotide triplets, and ends with a termination mechanism on a stop codon. One messenger RNA (mRNA) gives one polypeptide.

Several mechanisms make it possible for the translational machinery to decode the genetic information in non-canonical ways. These programmed events are called *recoding* and, in most cases, they give rise to extended proteins carrying new functional domains without being translational error. Recoding makes it possible to synthesise two polypeptides from the same mRNA, obtained by canonical decoding or recoding. The frequency of recoding is usually low, at around two to twenty percent.

In collaboration with the IGM, we focused on eukaryotic -1 frameshifting, in which translating ribosomes are induced to slip one nucleotide backwards at a specific recoding site on the mRNA. Most examples of -1 frameshifting sites have been found in virus genomes (e.g. IBV, HIV). Enhancing or reducing the effectiveness of the mechanism can dramatically influence virus viability. Our aim was to model this event and to establish an algorithm for the identification of new genes that use -1 frameshifting in their expression. We combined two independent strategies to look for frameshifting sites in silico: a model-based approach and a similarity-based approach. We tried to improve the previously established computational model of eukaryotic -1 frameshifting sites by designing a module based on a combination of bioinformatics methods, computer science concepts (knowledge representation, data mining, algorithm design and visualisation) and biological experimentation. We first created an attribute-value version of the existing model, taking into account the consensus organisation (a typical site contains a slippery heptamer, followed by a stimulatory structure - stemloop or pseudoknot - downstream, [Brierley et al. 1989]) and properties demonstrated or thought to be relevant in the frameshifting process. We then applied a machine learning supervised algorithm to refine the current model, providing new constraints to make the model more precise. For this, we assumed that frameshifting was a binary event (occurring/not occurring). Positive and negative examples provided to the machine learning system were built by setting a threshold on the efficiency level of the frameshifting event.

We thus identified new properties that may be involved in frameshifting and specified limits for the ranges of some parameters [4, 34]. We then tried to find a prediction algorithm for potential frameshifting sites. The algorithm designed, ORFEO, finds candidates according to the current model. More precisely, it searches for genomic regions in which a pseudoknot immediately follows a slippery sequence. Data mining techniques (voting techniques) were used to sort the set of candidates obtained, making it possible to limit the candidates selected to the most promising. Finally, we added the visualisation of pseudoknots to our module to improve assessment of the relevance of the candidates found. This complements the other two approaches in that it makes it possible to formulate hypotheses concerning the structure of these candidates.

Our methodology for the model-based approach can therefore be summarised as a cycle consisting of the following steps: 1) take a set of sequences that induces frameshifting and represent them; 2) refine the model by means of supervised machine learning techniques; 3) design a prediction algorithm according to the refined model; 4) sort the candidates

obtained according to their similarity to known examples; 5) test the most promising candidates *in vivo*. This last step can also be used to evaluate the model, which must be modified if the candidates tested are not frameshifting sites. In such cases, they can be considered counterexamples of the event and added to the set of initial data. The whole process can then be performed again with this extended set of data and the cycle repeated several times. When the model is considered sufficiently reliable, it will be used to construct an effective prediction tool.



RNAedit is a program developed at LRI for the three-dimensional visualisation and editing of RNA secondary structures. Two types of information are input into the program: the primary sequence of the RNA, and folding information indicating which bases are paired. This information is obtained from programs that predict secondary RNA structures, the results of which are not always realistic, especially if particular structural features, such as pseudoknots, are present. RNAedit computes a three-dimensional view of the secondary structure in order to display a clear representation of that structure, including pseudoknots. The paired bases are displayed with a colour indicating the strength of binding. It is possible to navigate through the structure in all directions to focus on interesting regions. Bases can be selected and modified to test the effect of a mutation on secondary structure. Bonds can be added or broken to test structural constraints. Using this software, biologists can input their own knowledge into the system to obtain a structure more realistic than that predicted by folding software. The figure shows the RNAedit display of a pseudoknotted structure in the genome of the avian infectious bronchitis virus (IBV). This structure, which has been proposed by the ORFEO system developed at LRI, is the structural part of a good candidate for a -1 frameshifting site.

Research groups Bioinfo Research description For the similarity approach, we searched for genomic regions in which two domains, each carrying a protein pattern, can be associated within a polypeptide by a single -1 frameshift. This approach does not rely on any particular model of a frameshifting site. We then crossed the two methods and evaluated common hits *in vivo* [17]. Four common candidates were identified for the *S. cerevisiae* genome and tested *in vivo*. Frameshifting efficiency was highly significant (13%), with respect to the background rate (0.1%) and to the programmed L-A virus frameshifting (10%) for one of these candidates.

The combination of two simple approaches (model-based and similarity-based approaches) led to the identification of several candidate genes potentially controlled by a -1 frameshift mechanism. This approach is very promising and we plan to extend it directly to other eukaryotic and prokaryotic organisms. We also plan to improve the visualisation module by increasing the knowledge included on the structure of pseudoknots, making it more precise. It could then be used as a filter for selecting promising candidates and would be included in our cyclic scheme (see methodology above).

This research was supported by a CNES grant on the Preparatory Program Mars Sample Analysis "Identification et réactivité de molécules biologiques" (Identification and reactivity of biological molecules) and by a joint CNRS-INRA-INRIA-INSERM Bioinformatics action "Phénomène de recodage traductionnel : étude bioinformatique du décalage de phase en -1" (Phenomenon of translational recoding: bioinformatics study of -1 frameshifting).

This work was the starting point of a close collaboration between LRI and IGM. It gave rise to the definition of a multidisciplinary program "Programme PluriFormation" (PPF, University Paris-Sud) on Bioinformatics and Genomics co-directed by Jean-Pierre Rousset (IGM) and Christine Froidevaux (LRI). This PPF brings together the bioinformatics groups from three biological laboratories (IGM, LEBS and IBP), two computer science laboratories (LIMSI, LRI) and from the laboratory of mathematics at Orsay campus. This PPF aims to develop increasing numbers of research projects in the same spirit as the frameshift project, involving a multi-disciplinary and *multifield* approach.

#### **Key references:**

- [4] M. Bekaert, L. Bidou, A. Denise, G. Duchateau-Nguyen, J.-P. Forest, C. Froidevaux, I. Hatin, J.-P. Rousset, and M. Termier. Towards a computational model for eukaryotic -1 frameshifting sites. *Bioinformatics*, 19:327-335, 2003.
- [17] M. Bekaert, J.-P. Forest, L. Bidou, A. Denise, G. Duchateau-Nguyen, C. Fabret, C. Froidevaux, I. Hatin, J.-P. Rousset, and M. Termier. Searching the saccharomyces cerevisiae genome for -1 frameshifting sites. In *Proceedings of ECCB 2003 (European Conference on Computational Biology)*, pages 4-6, September 2003.
- [34] A. Denise, M. Bekaert, L. Bidou, G. Duchateau-Nguyen, J.-P. Forest, C. Froidevaux, I. Hatin, J.-P. Rousset, and M. Termier. Vers une modélisation des gènes à décalage de cadre en -1 dans les génomes eucaryotes. In *Proc. of JOBIM 2001, Journées Ouvertes, Biologie, Informatique et Mathématiques*, pages 229-237, 2001.

## Theme 2: Analysis of Biological Sequences and Structures

• Participants: Alain Denise, Patrick Amar, Claire Herrbach, Yann Ponty, Romain Riviere and Stéphane Vialette.

This theme is closer in spirit to theoretical computer science. We have three main efforts in this field. First, we are interested in developing algorithms for specific biological problems (sequence analysis, RNA secondary and tertiary structures, and protein-protein interaction networks) that are both biologically and mathematically sound. This work includes the design and analysis of efficient algorithms solving these problems in practice and strategies for validating the results. We are also interested in developing software for visualising RNA secondary structures. Finally, we also address the problem of modelling and simulating biological processes.



Molecular phylogeny involves constructing the "tree of life" by comparing homologous genes of several different species. It is difficult to compare genes encoding RNA because homologous RNA genes may have very different sequences, even in closely related species. As RNA structure is of particular importance, comparisons of RNA molecules must take into account both the structure and the sequence of the RNAs. The figure shows a comparison of the secondary structures of two transfer RNA (tRNA) molecules, Homo sapiens isoleucine-tRNA (left) and Bacillus subtilis arginine-tRNA (right) based on an algorithm developed at LRI in collaboration with LaBRI. Red and purple areas indicate identities and substitutions with no effect on structure, respectively, whereas blue areas indicate substitutions that modify structure (pairings and unpairings). This figure was produced with Tulip, a graphics program developed at LaBRI.

#### Combinatorial and statistical models for sequence analysis

As the amount of data from sequenced genomes is rapidly increasing, there is a crucial need for efficient computer-based means of extracting new biological information from sequences. For this purpose, a widely used method involves comparing biological sequences with random ones, representing the *background noise* from which relevant biological information must stand out. This powerful paradigm is implemented in several fields in sequence

Laboratoire de Recherche en Informatique / 2000-2004 report

Research groups
Bioinfo
Research description

analysis. The search for exceptional motifs in biological sequences constitutes a key example of this paradigm. In this context, an exceptional motif is a pattern that is over- or underrepresented in a biological sequence, by comparison with the expected number of occurrences in random sequences. The underlying hypothesis is that overrepresented or underrepresented motifs may be indicative of important biological functions. For the observations to be relevant, random sequences must obey a model that takes into account wellchosen characteristics of biological sequences. One of the most used models of random sequences follows a Markovian process, based on the number of occurrences of all factors of a given fixed length in one or several biological sequences taken as a reference. Properties of random sequences are studied either analytically (when possible) or experimentally, by random generation.

We have developed, in collaboration with the INRIA Algorithms project, an analytical method for determining the distribution of motif occurrences in random Markovian sequences ([24] and [11]). In particular, we have developed the first closed formula for calculating the conditional probability distribution of one motif given that another known motif is overrepresented, and have applied it to biological data. This work constitutes a first step towards models of random sequences taking into account parameters other than the classical statistical ones. This is important because adding more biological information to sequence models may make it possible to identify the properties of biological sequences more accurately. We have therefore designed an algorithm for the random generation of sequences according to the so-called *shuffling* model, with additional motif constraints [27]. A number of structural features of genomic sequences (such as secondary RNA structures) may be represented in terms of formal context-free grammars. We design models of sequences taking into account these structural properties as well as classical statistical properties, and develop algorithms for the random generation of sequences according to these models [23]. This work has resulted in the GenRGenS software (http://www.lri.fr/~denise/GenRGenS) [37], which can be used for the random generation of biological sequences and structures according to several models. In particular, we have randomly generated structured sequences to estimate the sensitivity of certain alignment algorithms, in collaboration with LORIA [26]. We plan to develop models of random sequences that take into account more complex structural features. We are particularly interested in refining models of RNA structures by taking pseudoknots into account.

This work was partly supported by the IMPG working group "Algorithmique, combinatoire et statistique des séquences génomiques" and "Action spécifique STIC-CNRS Modélisation et algorithmique des structures d'ARN".

#### Visualisation and algorithm design for RNA structures

The structure and function of RNA molecules has become a topic of considerable interest in the last few years. Indeed, biologists discovered very recently that the role of RNA in cells is far more important than previously thought, particularly as concerns the regulation of gene expression. The biological function of an RNA molecule is strongly related to its three-dimensional conformation (tertiary structure). This conformation is generally represented by means of a graph on which the nodes are the nucleotides and the edges are the physicochemical bonds between nucleotides.

We search for structural patterns in RNA tertiary structures. This is important for comparing RNA structures and for elucidating their functions. In collaboration with TAGC (INSERM, Marseille) and the Bioinformatics group of the University of Montreal, we are designing and applying algorithms for the counting and classification of structural patterns of a given RNA structure, and for identifying patterns shared by different RNAs from the same functional family. We are also studying, in collaboration with LINA (Nantes), the computational complexity of finding potential structured patterns in RNA sequences. We are using an approach involving the geometric description of the graph by means of a natural generalisation of intervals, namely 2-intervals ([21] and [31]).

In collaboration with LaBRI (Bordeaux) and IGM (Orsay), we are developing algorithms for the pair-wise comparison of RNA molecules. This problem is much more difficult for RNA than for other biological molecules, such as DNA or proteins. Indeed, very different RNA sequences may give rise to very similar structures, with similar biological functions. Thus, sequence comparison algorithms cannot be used to compare RNA molecules. Our approach involves representing the secondary structure of an RNA molecule as an enriched tree. We are investigating comparison algorithms that take into account a number of constraints due to the biological nature of the structures concerned.

Finally, we are developing interactive software for the visualisation of RNA structures. We aim to offer the user the possibility of acting directly on the structure, by breaking bonds, for example, or by computing an alternative structure according to given constraints. We will do this by associating our software with programs for the prediction of RNA structure. This will make it possible, not only to visualize, but also to explore interactively the potential structures of any given RNA molecule.

Our research benefited from stimulating discussions in several national working groups: "Navgraphe" (ACI Masses de Données), Actions spécifiques STIC-CNRS "Modélisation et algorithmique des structures d'ARN" et "Nouveaux modèles et algorithmes de graphes pour la Biologie", and the new national "AReNa" working group (ACI-IMPBIO).

#### **Comparing protein interaction networks across species**

High-throughput analysis makes possible the study of protein-protein interactions at a genome-wide scale. The ever-increasing amount of information available for protein networks for different organisms lends itself to comparative analysis, in which the aim is to determine the extent to which protein networks are conserved between species. Indeed, there is increasing evidence to suggest that proteins that function together in a pathway or structural complex are likely to evolve in a correlated fashion, and during evolution, all such functionally linked proteins tend to be either preserved or eliminated in a new species.

Protein interactions identified on a genome-wide scale are commonly visualised as protein interaction graphs, in which the proteins are vertices and the interactions are edges. Experimentally derived interaction networks may be extremely complex, and it is a challenging problem to extract biological functions or pathways from them. In collaboration with the Laboratoire de Génétique Moléculaire (*Ecole Normale Supérieure*, Paris) and the *Institut Gaspard Monge* (University of Marne-la-Vallée), we studied the problem of automatically identifying occurrences of a given protein complex in the protein-protein interaction graph of another species [25]. We demonstrated that the problem is polynomial-time solvable provided that each protein has at most two orthologues in the other species, but is computationally hard if there are at least three orthologs. We also developed practical ways of coping with hardness by translating the problem into the well-known satisfiability framework, and by giving a fixed-parameter algorithm for a restricted common case.

We intend to structure our research in the following ways. First, we wish to improve our model for dealing with false negatives and false positives (the more complete the model, the more accurate the model). Secondly, we plan to design algorithms for comparing whole protein-protein interaction graphs comparing protein-protein interaction graphs without previous knowledge of the particular protein complex we are looking for. We are therefore focusing our efforts on adapting satisfiability solvers for "fast" exponential-time algorithms.

Research groups Bioinfo Research description

#### Modelling and simulation of biological processes

In collaboration with the Bioinfo group at LaMI (University of Evry), the Epigenomic project of Genopole (Evry) and LIPC (University of Rouen) we have designed models and simulation tools to study the formation and behaviour of large assemblies of proteins.

These studies have addressed two different biological objects: studies of *functioning-dependent structures* (FDS), which are a cascade of enzymes involved in a metabolic pathway [1], and studies of the interactions between the cell cytoskeleton and external mechanical constraints.

FDS result from reactions catalysed by certain enzymes, in which the product of the reaction catalysed by enzyme N is the substrate of enzyme N+1. We study the assemblies formed as a result of an increase in affinity between two consecutive enzymes in the chain. These macromolecular assemblies persist as long as the initial substrate is being metabolised. They disassemble when no longer needed, *i.e.* when all the initial substrate has been consumed.

All metabolic pathways in eukaryote and prokaryote cells involve biochemical reactions catalysed by proteins (enzymatic function). In *in vitro* experiments, the environment is homogeneous and the volume considerably larger than in a real cell, so large quantities of reagents are required to achieve the concentrations required for initiation of the chemical reactions. Conversely, in a real cell, the quantity of protein present is so small that, even in very small volumes, the concentrations are too low to produce a sufficient quantity of metabolites for the cell to survive ... The basic idea is that large enough local concentrations must be present in the cell to maintain the metabolic pathways required at an efficiency high enough to feed the cell.

Modelling the behaviour of such metabolic pathways *in vivo* with continuous systems, such as partial derivative equations, is difficult as it is not straightforward to fit models to the very small numbers of enzymes involved in some pathways (tens to hundreds of copies). Can we really use the macroscopic properties of these models when they are applied to so few proteins? In such conditions, does the word 'concentration' actually mean anything?

To take into account these particular conditions, we have chosen to model these biochemical reactions at the macromolecular level, using a discrete approach [2]. The cell is represented by a 3D volume bound by a 2D membrane containing a realistic number of copies of each kind of protein involved in the studied pathway. When available, these numbers are taken from published studies. For computational reasons, the simulated cell is a bacterium such as *E. coli*, which is much smaller than a eukaryote cell. The simulation programme is a variant of a multi-agent system in which each macromolecule is an agent, driven by local rules, depending on its type. The simulator is fed with a description of the molecules, their diffusion speed, the reaction equations and the initial number of copies of each molecule species.

Many simulations have been made using a model of the phosphotransferase system (PTS) and glucose glycolysis in *E. coli*. These simulations have shown that the efficiency of this metabolic pathway is considerably enhanced if we allow the assembly of enzymes in the PTS branch.

The cytoskeleton of a cell consists mostly of actin and tubulin strands. These filaments are permanently in the process of polymerising and depolymerising, thereby forming a dynamic structure. The simulation *in silico* of these filaments has shown the reorganisation of their orientation in direct response to a mechanical constraint applied to the cell membrane.

Future improvements to the simulator include the modelling of the membrane as a collection of lipids, making it possible to predict more accurately local modifications of the membrane. We also plan to modify the programme to make it run on a distributed environment such as a cluster of PCs.

#### **Key references:**

- P. Amar, P. Ballet, G. Barlovatz-Meimon, A. Benecke, G. Bernot, Y. Bouligand, P. Bourguine, F. Delaplace, J.-M. Delosme, M. Demarty, I. Fishov, J. Fourmentin-Guilbert, J. Fralick, J.-L. Giavitto, B. Gleyse, C. Godin, R. Incitti, F. Képès, C. Lange, L. L. Sceller, C. Loutellier, O. Michel, F. Molina, C. Monnier, R. Natowicz, V. Norris, N. Orange, H. Pollard, D. Raine, C. Ripoll, J. Rouviere-Yaniv, M. S. jnr., P. Soler, P. Tambourin, M. Thellier, P. Tracqui, D. Ussery, J.-C. Vincent, J.-P. Vannier, P. Wiggins, and A. Zemirline. Hyperstructures, genome analysis and i-cell. *Acta Biotheoretica*, pages 357-373, 2002, Kluwer Academic Publishers.
- [2] P. Amar, G. Bernot, and V. Norris. Hsim: a simulation programme to study large assemblies of proteins. *Journal of Biological Physics and Chemistry*, 4:79-84, 2004, Collegium Basilea & AMSI.
- [11] M. Régnier and A. Denise. Rare events and conditional events on random strings. *Discrete Mathematics and Theoretical Computer Science*, 6(2):191-214, 2004.
- [25] I. Fagnot, G. Lelandais, and S. Vialette. Bounded list injective homomorphism for comparative analysis of protein-protein interaction graphs. In *Proc. of 1st Algorithms and Computational Methods for Biochemical and Evolutionary Networks (CompBioNets 2004)*. KCL publications, King's College, London, 2004.
- [26] G. Kucherov, L. Noe, and Y. Ponty. Estimating seed sensibility on homogenous alignments. In IEEE, editor, *Proceedings of Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE04)*, pages 387-394, 2004.
- [27] R. Rivière, D. Barth, J. Cohen, and A. Denise. Shuffling biological sequences with motifs constraints. In *Proc 1st Algorithms and Computational Methods for Biochemical and Evolutionary Networks (CompBioNets 2004)*. KCL publications, King's College, London, 2004.

## Theme 3: Analysis and Storage of Massive Biological Data

• Participants: Céline Rouveirol, Alain Bidault, Sarah Cohen Boulakia and Christine Froidevaux

Most biologists now store their experimental data in a structured computer format, if not in a relational database. These private data may describe several points of view for a set of objects of study e.g. patients, proteins. For instance, a patient may be described by his clinical file, measurements of his genes expression at a specific moment, reports of chromosomal alterations detected in his tumour, etc. Moreover, when interpreting their own experimental data, biologists are now faced with the problem of integrating information from multiple heterogeneous public sources with their own experimental data. We try to support them in the increasingly complex process of interpreting their data by following two complementary lines of research: the integration of biological data, with the aim of providing biologists with transparent access to distributed heterogeneous data, and data mining, providing biologists with potentially relevant hypotheses extracted from these complex data.

Research groups Bioinfo Research description

#### Integration of biomedical data sources

#### HKIS: an integrative platform

Most of the work in this area has been performed in the context of the European HKIS project, which aims to develop an Integrative Software Platform for Biological and Biomedical Data Processing in Cancerology. The goal of the HKIS project is to provide the user with a workspace in which he can use analysis tools to treat his data. The global approach of an HKIS user is based on a set of scenarios describing different analysis methodologies and reflecting the expertise of the biologists and health professional partners



The CGH technique can be used to assess chromosome alterations (gain or loss) in the genome of a tumour cell. For array-CGH, chromosome fragments about 100,000 base pairs in length are extracted from the human genome and inserted into BACs (Bacterial Artificial Chromosomes) clones, which are then spotted onto glass slides. Tumour DNA (test DNA) and normal DNA (reference DNA) are then labelled separately with two different fluorophores (red Cy5 for test DNA and green Cy3 for reference DNA). If a genomic region is amplified, *i.e.*, has more than two copies in the tumour cell genome, then more tumour DNA than normal DNA will hybridise to the chromosome fragments corresponding to this region, and the spots will therefore appear red. In contrast, if a genomic region is lost in a tumour, *i.e.*, zero or only one copy in the cell genome, more normal DNA than tumour DNA corresponding to this region will hybridise to chromosome fragments, and the spot will appear green. In collaboration with the Institut Curie (Paris), we analyse CGH arrays (A.M. Snijders et al., Nat. Genet., 29:263-4, 2001) containing 3000 chromosome fragments of about 100,000 base pairs each. As the haploid human genome consists of about 3 billion base pairs and chromosome fragments are uniformly distributed throughout the genome, such fragments occur every 1 million base pairs and about 10% of the genome is represented on the array.

involved in the project. Our aim is to provide a formal framework for the uniform representation of all the scenarios in the HKIS project. This framework will enable us to compare the methods used for data analysis and to identify the need for new tools or methods.

At a given stage in a scenario, a user may wish to improve his understanding of the results of his analysis by asking questions necessitating the consultation of various data sources available through the Web. In the context of the HKIS project, we have selected 30 sources frequently consulted by health professionals in studies of cancer. We have thoroughly studied the selected sources and have designed a unified framework describing each source at meta-data level. This unified framework specifies, for each source, the list of biomedical entities included, the central entity, the confidence that certain users place in its data and so on. We have used this framework to design the unifying data model used by the HKIS platform.

The selection of the data sources to be considered at a given step of the scenario is critically important. We have thus developed a module to assist the user in selecting the data sources to be consulted during the querying process ([5, 32]). We have designed the Data Sources Selection algorithm (DSS), which takes into account both the query and the user preferences and exploits information provided by the unifying data model (described above). DSS was designed to reflect the way in which HKIS biologists search for information in public data sources. DSS yields a list of all the possible ways of obtaining information about the entities of the query. Thanks to DSS, the user does not need to know *a priori* which sources to query because they are selected automatically. The various ways of obtaining information provided by DSS can be used to exploit the complementary aspects of the data sources. Moreover, the user preferences are taken into account, making it possible to filter and to sort the results. The usefulness of DSS has been illustrated through an example, namely the BAC (Bacterial Artificial Chromosome) augmentation scenario [5].

We plan to generalise our work by designing a more generic data source selection algorithm. This algorithm would take into account other users' ways of querying the sources and would permit the automatic selection of tools useful for managing data.

#### User preferences and mediator

We have extended the work done in a mediator context [20], to address the problem of querying distributed biomedical databases. We propose exploiting the metadata of the sources to take into account user preferences [19]. A mediator approach tightly integrates multiple and heterogeneous data sources within a uniform query interface. The mediator system presented in [12] is designed within a tractable logical framework. Users ask queries in terms of a set of predicates. Thus, answering a query means translating a user query into query plans that refer directly to the relevant sources. It enables users to focus on specifying their demand, may combine data from multiple sources to obtain answers and is suitable for frequent updates of the biomedical databases. This work makes it possible to keep track of the origins of the instances provided as answers. The metadata included in the mediator system concern focus, reliability and cross-references.

We also plan to address the problem of synonymy, which arises when queries cannot be answered because the arguments of the predicates that refer to the same concept do not have the same name.

We have studied the problem of the interoperability of genomic data sources as part of various working groups (IMPG, AS STIC, ACI IMPBio).

Research groups Bioinfo Research description

#### **Data Analysis**

#### Analysis of array\_CGH data

As cancer is a disease with a genetic component, obtaining knowledge about recurrent genomic alterations observed in specific cancers is one of the many ways in which researchers increase their understanding of the mechanisms underlying cancer initiation and progression. The CGH (Comparative Genomic Hybridisation) array technique, is a more recent development than DNA microarray analysis and can be used to measure chromosomic alterations, with a resolution of about 1 or 2 Mb.

Array\_CGH data measure the number of occurrences of small fragments of chromosomes in the genome of a cell. In a normal situation, each chromosome has two copies, therefore each fragment of chromosome occurs twice in the genome. Measured alterations are fragment loss (a given chromosome fragment occurs 0 or 1 times in the tumoural cell) or fragment gain (a given chromosome fragment occurs more than twice). We have proposed [39] an algorithm that calculates minimal regions of recurrent gain or loss from a set of discretised CGH profiles. These regions of recurrent gain or loss potentially contain genes of interest, such as oncogenes, which are likely to belong to gained regions of the tumoural genome, whereas tumour suppressor genes are likely to be located in lost regions.

The technique we have proposed is inspired by frequent itemsets computation algorithms. The most original feature of our work is that we are looking for sequences rather than itemsets and that we have studied and used biologically founded constraints to reduce the number of such recurrent regions. The first application of the system to Array\_CGH data for bladder cancer patients, and to public cancer-related CGH data, has given very promising results: regions containing well-known oncogenes and tumour suppressor genes have been found in the data, and other regions have been isolated that require further biological validation.

The most promising aspect of this work is that these minimal regions can be used to search efficiently for patterns of genomic alterations predicting anatomo-clinical parameters (stage and grade of cancer, good/poor prognosis etc.). We are also considering applying similar techniques to search for co-expressed regions in transcriptome data. However, defining a region of co-expressed genes is much more complex than defining regions of gain and loss.

This work was performed in collaboration with the *Institut Curie*, Paris (UMR 144, the Molecular Oncology Laboratory) and was supported by the FP5 R&D European project HKIS.

#### Analysis of microarray data: BioRelief

The starting point for this study was the assumption that exposure to genotoxic substances induces various transcriptional changes that can be estimated by comparing the messenger RNA (mRNA) levels between two populations growing in different environments. We used DNA microarray technology to measure levels of all the mRNAs of the yeast *S. cerevisiae* to quantify the biological effects of low doses of ionising radiation, particularly when delivered at low dose rates, a situation relevant to most environmental exposures. As massive amounts of data are generated from microarrays, data analysis methods are required to determine whether changes in gene expression are significant. We tried to find a method for analysis of the transcriptome of these yeasts, to identify the genes involved in the response to environmental changes. The use of microarrays to discriminate between various environmental conditions and to determine the biological processes involved poses a great challenge. One reason for this is that very few measurements are generally available, with respect to the large number of genes. Another is that the usual statistical assumptions regarding both the distribution of the gene expression levels and the independence of genes are unwarranted, if not downright erroneous (groups of genes are known to be correlated), given our current level of knowledge. We have therefore suggested the use of an original attribute estimation method overcoming these drawbacks that is also robust to noise.

The BioRelief algorithm we have designed [44, 36, 16] is based on Relief, an attribute estimation technique. It looks for the genes that seem to be the most significantly correlated to the class to be predicted. In addition to its low computational cost (after code optimisation), BioRelief has the major advantage that, unlike many other statistically oriented techniques, its does not assume that genes are independent, and does not require information on the data distribution (*e.g.* are the data Gaussian). Moreover, it remains valid if there are missing data (lack of expression of some genes on some microarrays) and the possibility of choosing the number of neighbours makes it possible to control the trade-off between sensitivity and robustness to noise. This method prevents unreasonable sensitivity to aberrant values and is more precise than methods taking into account only mean values. Using BioRelief, we have been able to demonstrate the transcriptional response induced at low doses of radiation [8]. We ranked the genes according to their ability to differentiate samples exposed to radioactive environment or not. Based on this ranking, we identified a set of genes involved in different processes: three of these processes are clearly induced by low radiation exposure, and all participate in the same oxidative phosphorylation cascade.

This opens up very promising perspectives. For instance, we plan to use dedicated microarray data (based on selected genes) for diagnostic purposes. We also plan to use the BioRelief tool with other kinds of data, *e.g.*, human mRNA. It is also being used (free software) at the Centre for Pharmacology and Health Biotechnology (Montpellier), with other kinds of data. This work was performed in collaboration with the *Institut Curie* (UMR 2027, Orsay) and the LaMI (Evry). It was supported by the national project Biogen n°74.

Works on Data Analysis (Array\_CGH data and microarray data) benefited from stimulating discussions in several working groups: "Consensus" (Genopole Evry), "Gafopuces" and "Gafoum" (Action Spécifique STIC GafoDonnées), "Apprentissage et Biopuces" (Action Spécifique STIC ASAB - Apprentissage et Biologie).

#### **Key references:**

- [5] S. Cohen-Boulakia, S. Lair, N. Stransky, S. Graziani, F. Radvanyi, E. Barillot, and C. Froidevaux. Selecting biomedical data sources according to user preferences. In *Proc. of ISMB/ECCB 2004, Bioinformatics Special Number*, volume Vol 20 Suppl. 1, pages i86-i93, 2004.
- [8] G. Mercier, N. Berthault, J. Mary, J. Peyre, A. Antoniadis, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. Biological detection of low radiation by combining results of two microarray analysis methods. *Nucleic Acids Research, 2004, Vol.32 (1), e12, 2004.*
- [16] J. Mary, G. Mercier, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. *Informatique pour l'analyse du transcriptome*, chapter Utilisation d'une méthode de sélection d'attributs pour l'analyse du transcriptome de cellules de levure exposées à de faibles doses de radiation, pages 189-205. Informatique et systèmes d'information. Traité IC2 Hermes-Lavoisier, 2004.
- [19] A. Bidault, S. Cohen-Boulakia, and C. Froidevaux. Preferences for queries in a mediator approach. In Proc. of ECAI2004 (short paper), European Conference on Artificial Intelligence, 2004.
- [39] C. Rouveirol, N. Stransky, and F. Radvanyi. Vers l'identification de nouveaux oncogènes et gènes suppresseurs de tumeurs. analyse automatique des données de puces CGH. In *Proc. of JOBIM'04, Journées Ouvertes, Biologie, Informatique et Mathématiques*, 2004.

Research groups Bioinfo Research description

## **Research Perspective**

We will carry on our work on the themes presented above by addressing pure computer science issues in relation to biological problems, as well as bioinformatics issues more tightly focused on analysis of the biological objects themselves. We will work on a new theme, the annotation of genomic data, in the RAFALE project. We intend to maintain a *multifield* group with a good balance between the various computer science skills and plan to reinforce this group in the coming years by recruitment.

Collaborations with computer science groups will be continued, especially with the other groups at LRI and with the bioinformatics groups at LaBRI (Bordeaux) and LaMI (Evry). Multi-disciplinary studies with biologists will be intensified, particularly with biologists at the University Paris-Sud, through the PPF project and the IFR *Génome*, and with biologists at the *Institut Curie*, where Céline Rouveirol is currently on leave. In particular, we plan to collaborate more closely with IGM (Orsay). The BioInfo group, although only recently formed, is well-recognised within the French scientific community. We now aim to develop our reputation in the international community.

# Highlights

Since 2002, members of the BioInfo group have led or co-led seven national working groups (STIC AS, IMPG, ACI IMPBio, Génopole Evry), one multi-disciplinary programme at University Paris-Sud (PPF) and two national projects (EPST joint grant and ACI IMPBio), and have been part of the steering committees of four working groups.

At the international level, the group is involved in the HKIS R&D project (5th Framework Programme) led by the ISoft company.

The group has been highly involved in the creation of the Masters degree in Bioinformatics and Biostatistics of University Paris-Sud, aiming to train both researchers and engineers. All of the permanent group members are involved in teaching for this Masters course. The originality and richness of the BioInfo group stem from the diversity of computer science fields used to tackle intricate genomics problems in a comprehensive way. The group publishes in journals and conferences in Computer Science as well as in Bioinformatics and in Biology. This reflects our desire to carry out fundamental research relevant to Biology.

#### **Radar view**

The statistics shown in the radar view below are standardised with respect to the size of the group, but not with respect to the duration of its activity. The group was formed in 2002 (initially with 7 members), and this radar view therefore takes into account only the publications of its members since that date. This accounts for the small number of publications for the period as a whole (2000-2004) and the lack of PhDs obtained (only one habilitation). However, in the next year, three students from our group will defend their Ph.D.s.

This radar view also shows that despite the short duration of its activity the amount of contract funding obtained and the level of programme committee participation are as would be expected for a group of this current size.



Research groups

Bioinfo

Research perspectivee Highlights The radar view displays a synthetic view of the activity of the group over the 2000-2004 period, by means of five statistics:

- Size of the group at the end of the period (01/01/2004),
- Number of publications (peer-reviewed journals, books, book chapters, major international conferences),
- Annualised amount in Euro of new contracts and grants,
- Number of participations in program committees (national and international), and,
- Number of defended Ph.D.s and habilitations.

These statistics are displayed as a percentage with respect to the overall laboratory, together with a reference line corresponding to the size of the group. Data points outside the reference line show above-average performance whereas those inside the line show below-average performance, relative to the laboratory.

As with any statistics, these data must be interpreted carefully. In particular, they are not a direct measure of relative quality of the groups within the laboratory, because different research areas place different values and offer different opportunities with respect to these five measures.

B.10.5 / Bioinfo

# **Evaluation of research**

## **Program committees**

#### International events:

- ECML, European Conference on Machine Learning: Céline Rouveirol (2002, 2003, 2004).
- IDCM, International Conference on Data Mining: Céline Rouveirol (2004).
- MRDM, Workshop on Multi-Relational Data Mining: Céline Rouveirol (2002, 2003, 2004).
- MSBP, Modelling and Simulation of Biological Processes in the context of genomics: Patrick Amar (2002, 2003, 2004).
- ICML, International Conference on Machine Learning: Céline Rouveirol (2002).
- ILP, Inductive Logic Programming: Céline Rouveirol (since 1994).

#### **National events:**

- EGC, Journées francophones d'Extraction et de Gestion des Connaissances: Christine Froidevaux (2003, 2004).
- JOBIM, Journées Ouvertes Biologie, Informatique, Mathématiques: Alain Denise (2002, 2004).
- CAp, Conférence Francophone d'Apprentissage: Céline Rouveirol (2002, 2003, 2004).
- Workshop Analyse de données d'expression (in conjunction with Journées Post Génomiques de la Doua, 2003): Céline Rouveirol.

## Evaluation committees and invited expertise

- Scientific Committee of national research program for junior researchers, *ACI Jeunes Chercheurs*, 2002: Christine Froidevaux.
- Scientific Committee of national research program on computer science, mathematics and physics in molecular biology, *ACI IMPBio*, 2004: Alain Denise.
- Scientific Committee of national research center on agronomics, *INRA*, Jouy-en-Josas, 2003-2006: Alain Denise.
- External expert committee of Ouest-Génopole: Alain Denise, member.

- Scientific Expertise for French ministry of research, *Mission Scientifique Universitaire MSU*, 2003-2004: Christine Froidevaux (evaluation of ERT research groups and ACI projects).
- Scientific Expertise for French ministry of research, *Mission Scientifique, Technique et Pédagogique MSTP*, 2003: Alain Denise.
- CNRS National Committee, *Comité National du CNRS*, interdisciplinary commission CID 44, 2003-2004: Christine Froidevaux, member.
- Scientific advisor of the GEMO team (PCRI INRIA): Christine Froidevaux.
- Research Associate recruitment committee, INRA, "Mathématiques et Informatique", 2002 & 2004: Alain Denise.
- Evaluation Committee of MIG research team (*Mathématique, Informatique et Génome,* INRA, Jouy-en-Josas), 2004: Alain Denise.

#### **Other evaluation activities**

- Scientific Committee (*Comité Scientifique de l'UFR des Sciences*), Université de Versailles-St-Quentin-en-Yvelines, 2003-present: Alain Denise.
- Scientific expert for France Telecom R&D, 2004: Christine Froidevaux.
- External referee for European project IST-1999-11993 Mining Mart (2000-2003): Céline Rouveirol.
- Specialized Scientific Committee on Mathematics, Bioinformatics and Artificial Intelligence, National research center on agronomics (INRA), 2003-2004: Alain Denise.
- Reviewer for PhD dissertation: Alain Denise (4), Céline Rouveirol (2).
- Reviewer for Habilitation: Christine Froidevaux (1).

## B.10.6 / Bioinfo

# Volunteer professional service

## Management positions in scientific organisations

- CNRS network on Bioinformatics, *Réseau Thématique Pluridisciplinaire Bioinformatique : de la séquence génomique à la fonction biologique, RTP 41,* 2001-2004: Christine Froidevaux and Alain Denise, members of the Steering Committee.
- Multidisciplinary program on Bioinformatics and Genomics, *Plan Pluri-Formations*, Université Paris Sud, 2002-2005: Christine Froidevaux, co-chair; Alain Denise, member of the Steering Committee.
- CNRS Working group on transcriptome and data mining, *Action Spécifique Gafodonnées, sous-groupe Gafopuces*, 2001: Christine Froidevaux, co-chair.
- CNRS Working group on integration and interoperability of genomic data sources, *Action Spécifique* Intégration et interopérabilité de sources de données génomiques, 2003-2004: Christine Froidevaux, co-chair.
- CNRS Working group on models and algorithms for RNA, *Action Spécifique Modélisation et Algorithmique des Structures d'ARN*, 2002-2003: Alain Denise, co-chair.
- CNRS Working group on machine learning and biology, *Action Spécifique Apprentissage et Biologie*: Céline Rouveirol, member of the Steering committee.
- National research program on computer science, mathematics and physics for molecular biology, *ACI IMPBIO*:
  - Project on semi-automatic functional annotation rules for yeast, *RAFALE*, *Règles pour l'annotation Fonctionnelle semi-Automatisée pour la LEvure*, 2004-2007: Christine Froidevaux, chair.
  - Working group on integration of information systems in biology, *ISIBio, Intégration des Systèmes d'Information en BIOlogie*, 2004-2006: Christine Froidevaux, member of the Scientific Committee.

Research groups

#### Bioinfo

Evaluation of research Volunteer professional service

- Working group on epigenomics, *Epigénomique*, 2004-2007: Patrick Amar, member of the Scientific Committee.
- Working group on structure and function of RNA, *AReNa* : *Structure et fonction des ARN*, 2004-2007: Alain Denise, chair.
- National research program on computer science, mathematics and physics for genomics, *IMPG*, working group on ontologies and metadata, *Ontologies et Métadonnées*, 2003: Christine Froidevaux, member of the scientific committee.
- Joint CNRS-INRA-INRIA-INSERM project on multi-criteria approaches for genome analysis, AMASIG: Approches multicritères pour l'analyse des génomes, 2000-2001: Alain Denise, chair.
- Working group on biological objects, Ateliers Epigenèse de Génopole, 2003-2004: Patrick Amar, chair.
- Epigenomics program of Génopole, 2004-2007: Patrick Amar, member of the Steering Committee.

## Organisation of conferences and scientific events

- ALEA'02 meeting of the national network *GDR ALP* (2002, Marseille): Alain Denise, co-organiser with D. Gouyou-Beauchamps.
- Spring Schools MSBP'02, MSBP'03 and MSBP'04, *Modelling and Simulation of Biological Processes in the context of Genomics*: Patrick Amar, co-organiser.
- Working group *Data mining and microarrays*, EGC'02 conference, 2002: Christine Froidevaux, coorganiser.
- Meeting of the European project HKIS, LRI, 2003: Christine Froidevaux, organiser.
- CNRS Working group on integration and interoperability of genomic data sources, *Action Spécifique Intégration et interopérabilité de sources de données génomiques*, LRI, 2003: Christine Froidevaux, organiser.
- One-day meeting Transcriptome & Proteome of multidisciplinary program *Plan PluriFormation Bioinformatique et Génomique*, University Paris Sud, 2004: Christine Froidevaux, organiser.

## Working groups

- CNRS Working groups ("Actions Spécifiques"):
  - GafoDonnées, Gafoum sub-group on frequent pattern usage, 2001-2002: Céline Rouveirol.
  - Apprentissage et Biologie (ASAB): Christine Froidevaux.
  - Nouveaux modèles et algorithmes de graphes pour la biologie, 2003-2004: Stéphane Vialette.
  - *Algorithmique pour la Bioinformatique*, 2002-: Alain Denise, Christine Froidevaux.
- National research program *ACI IMPBio*, working group *AReNa : Structure et fonction des ARN*", 2004-2007: Patrick Amar, Christine Froidevaux, Claire Herrbach, Yann Ponty, Romain Rivière, Stéphane Vialette.
- National research program on masses of data, *ACI Masses de données*, working group on navigation in large graphs, "Navgraphe", 2003-2006: Alain Denise.
- National research program on new interfaces with mathematics, *ACI Nouvelles Interfaces des Mathématiques*, "working group on mathematical and algorithmical aspects of biochemical and evolutionary networks", 2004-2007: Alain Denise, Stéphane Vialette.
- National research program IMPG (*Informatique, Mathématique et Physique pour le Génome*) working groups:
  - Bases de données en Bioinformatique et Génomique, 2001-2002: Christine Froidevaux, Céline Rouveirol.
  - Algorithmique et Statistiques des Séquences Génomiques, 2000-2003: Alain Denise.
- Génopole d'Evry working groups:
  - Consensus, 2002-: Christine Froidevaux, Céline Rouveirol.
  - Epigenèse, 2001-: Patrick Amar.
- *ALEA* working group, 2000-: Alain Denise.

# **Contracts and grants**

Туре	Scientific Director	Project Name	Funding institution	Managing institution	Dates	Duration (months)	Total €
IST	Ch. Froidevaux	HKIS - Integrated Software System	UE	Cephyten	07/02- 12/03		70000
RNTL	Ch. Froidevaux & E. Waller	BIOINGENIERIE Plateforme	MEN	UPS	11/01- 05/04		42063
Inter EPST	A. Denise	Amasig	CNRS	CNRS	11/00- 11/03		15245
Inter EPST	A. Denise	Phénomène de recodage (bio-info)	CNRS	CNRS	11/00- 11/03		19056
IMPG	A. Denise	Statistique des Séquences Génomiques	CNRS	CNRS	12/02- 12/05		10000
ACI	A. Denise	Navgraphe	MENRT	CNRS	12/00- 12/03		6773
ACI	A. Denise	Mathematical and Algorithmical Aspects of Biomedical and Evolutionnary Networks	MENRT	CNRS	2004-2007		6000
ACI	Ch. Froidevaux	Rafale	MENRT	CNRS	2004-2007		27000
ACI	A. Denise	Arena	MENRT	CNRS	2004-2007		30000
PPF	Ch. Froidevaux & J-P. Rousset	Plan Pluri Formation	UPS	CONTF/ CONTE	01/02- 12/05		66944
AS	A. Denise	Modélisation et Algorithmiques	CNRS Stic	CNRS	07/02- 07/03		6100
AS	Ch. Froidevaux	RTP 41 Intégration des sources de données	CNRS Stic	CNRS	06/03- 06/04		5000

Note: see glossary for acronyms.

## Summary of main scientific projects:

### HKIS

Integrated Software Platform For Biological and Biomedical Data processing in Cancerology.

Partners : ISoft company, Institut Curie (France), Faculty of Medecine of University of Ulm (Germany) and European Institute of Oncology (Milano) European project, IST-5th framework program, 2002-2004 (30 months). Scientific director for LRI:

Christine Froidevaux.

The goal of this project is to design an integrated and evolutionary environment able to address in real-time heterogeneous and large volume data sources, useful for cancer study purpose. The global approach of a HKIS user is based on a set of analysis scenarios that we have defined in close collaboration with biologists. They reflect their expertise, and are represented by workflows. We have also built a unified data model for guiding the query process. Moreover, we have proposed two algorithms. The first one learns minimal regions of chromosomal gain and loss regions from CGH array data. The second one (DSS) for selecting data sources according to user preferences. DSS is available at http://www.lri.fr/~cohen/dss/dss.html (free software).

Research groups

Bioinfo

Volunteer professional service Contracts and grants

#### **BIOINGENIERIE Plateforme "Biogen 74"**

Partners : Institut Curie (Paris and Orsay), Laboratoire de Physique statistique (ENS Ulm, Paris), Hospital Henri Mondor (GETU, Créteil), University Paris-Sud (LRI and Kremlin-Bicêtre) and the ISoft company.

The goal of this project is to support health professionals working in cancerology with an informatics platform providing them with intelligent retrieval and user-friendly navigation

through vital medical information, therefore helping the development of diagnostic tools and the identification of new therapeutic targets. In this context, we have elaborated a data mining algorithm, BioRelief, for analysing gene expressions yielded by micro-arrays. BioRelief is freely available at http://www.lri.fr/~chris/bioinfo/BioRelief.html.

#### Phénomène de recodage

*Partners : Institut de Génétique et Microbiologie and LRI (Paris XI).* 

This project developed a methodology for automatically identifying genes subject to translational -1 frameshifting in eukaryotic genomes. The results were validated using biological experiments. In particular, we showed how a machine learning

approach could be used to refine the previously known model. An algorithm and a prototype software were developed, which allowed to find new putative frameshifting genes in the Yeast genome.

#### RAFALE

Règles pour l'Annotation Fonctionnelle semi-Automatisée de la LEvure.

Partners : IBBMC (Orsay), MIG (INRA), LRI.

Determining the function of a sequenced protein is a crucial and complex task that is usually performed manually and for which annotators have to cross different kinds of information.

The RAFALE project aims at helping biologists in this difficult annotation task by providing them with a semi-automatic module for protein annotation. Our goal is to learn rules from the known cases, which will be used to infer the functional classes of proteins of unknown function. We will select for validating our system yeast genomic data.

#### **Other projects**

**Programme préparatoire à l'analyse des échantillons martiens** project, CNES grant on the PREMIER program, *"Identification et réactivité de molécules biologiques"*, with Institut Jacques Monod (Paris 6) and IGM Orsay, 2001.

INSERM action on Bioinformatics, 2001-2003. Scientific director for LRI: Alain Denise.

Project of the joint CNRS-INRA-INRIA-

Project funded by the *GenHomme* national network, 2001-2004

Scientific director

Christine Froidevaux.

(30 months).

for LRI:

Project of the national research program *ACI IMPBIO*, 2004-2007. Scientific director for LRI: Christine Froidevaux

**Caderige** national project (see I&A group). Members are research groups active in Natural Language Processing (LIPN), Information Extraction (Thales), Learning (LRI and Leibniz) and Biology (MIG, INRA).

ARN theme of CNRS Working group ("Action Spécifique") ALBIO (ALgorithmique pour la BIOinformatique), 20.000 F for the whole theme, 2001-2002.

**ASAB**, CNRS Working group ("Action Spécifique") organized in order to advertise bioinformatical problems to a machine learning audience.

**ISIBio** working group, Intégration des Systèmes d'Information en BIOlogie, supported by the national program ACI IMPBio (2004-2007), managed at LORIA (30.000 Euros for the whole project).

#### B.10.8 / Bioinfo

# **Collaborations**

## **Cooperation agreements**

- Consortium agreement with european project partners (HKIS): ISoft company, University of Ulm (Germany) and European Institute of Oncology (Italy): Christine Froidevaux, Céline Rouveirol, Sarah Cohen Boulakia.
- GeneBio, Genève: Sarah Cohen-Boulakia, Christine Froidevaux.

## Collaborations leading to joint publications

- Institut Curie, Paris: Sarah Cohen Boulakia, Christine Froidevaux, Céline Rouveirol.
- Institut Curie, Orsay: Christine Froidevaux.
- PRISM, University of Versailles: Alain Denise, Romain Rivière.
- LaBRI, University of Bordeaux: Alain Denise.
- Programme Epigénomique, Genopole d'Evry: Patrick Amar.
- LPIC, University of Rouen: Patrick Amar.
- LAMI, University of Evry: Patrick Amar, LORIA, INRIA: Alain Denise, Yann Ponty, Christine Froidevaux.
- INRIA Rocquencourt: Alain Denise.

- Institut de Génétique et Microbiologie (IGM), University of Paris-Sud Orsay: Sarah Cohen-Boulakia, Alain Denise, Jean-Paul Forest, Christine Froidevaux, Yann Ponty.
- LINA, University of Nantes: Stéphane Vialette.
- LGM, Ecole Normale Supérieure Paris: Stéphane Vialette.
- Institut Gaspard Monge (IGM), University of Marne-la-Vallée: Stéphane Vialette
- Romain Rivière.

## Other collaborations

- Stat. et Génome, Evry: Alain Denise, Christine Froidevaux.
- TAGC, INSERM ERM 206, Luminy: Alain Denise, Romain Rivière.
- DIRO, University of Montréal: Alain Denise, Romain Rivière.
- LaBRI, University of Bordeaux 1, Thèse de Claire Herrbach en co-tutelle: Alain Denise.
- Institut Jacques Monod, University Paris 6: Alain Denise, Yann Ponty, Stéphane Vialette.

Laboratoire de Recherche en Informatique / 2000-2004 report

#### Research groups Bioinfo

Contracts and grants Collaborations

- Susan Davidson, Center for Bioinformatics, University of Pennsylvania: Christine Froidevaux, Sarah Cohen-Boulakia.
- IBBMC, Orsay: Christine Froidevaux.
- MIG, INRA: Christine Froidevaux, Céline Rouveirol, Sarah Cohen-Boulakia.
- LIRMM, University of Montpellier: Christine Froidevaux, Sarah Cohen Boulakia.
- LORIA, INRIA: Christine Froidevaux, Sarah Cohen Boulakia.
- LIF, University of Marseille: Christine Froidevaux, Sarah Cohen Boulakia.
- LIPN, University Paris 13, Céline Rouveirol.
- GRAPPA research group, University Lille 3: Céline Rouveirol.

B.10.9 / Bioinfo

# Dissemination and technology transfer

## Software licensing and distribution

**GenRGenS** (*Generation of Random Genomic Sequences and Structures*) is a software for random generation of genomic sequences and structures. It uses several models and techniques in sequence analysis: classical ones (such as Markov chains) as well as new ones (weighted context free grammars) and combinations of different models. GenRGenS is written in Java, and hence can be run on most systems. GenRGenS is free software under the GNU license. Authors: Alain Denise, Yann Ponty: http://www.lri.fr/~denise/GenRGenS

**HSIM** Protein-Protein Interaction simulation Software, free Software under the GNU licence. Author: Patrick Amar.

**RNAedit** Visualisation of RNA Secondary and Tertiary Structures, free Software under the GNU licence. Author: Patrick Amar.

**DSS** Data Sources Selection, is an algorithm that selects data sources according to user preferences. Free software. Authors: Sarah Cohen Boulakia and Christine Froidevaux. http://www.lri.fr/~cohen/dss/dss.html

**BioRelief** is an analysis tool for microarrays. Free software. Authors: Antoine Cornuéjols, Christine Froidevaux, Jérémy Mary. http://www.lri.fr/~chris/bioinfo/BioRelief

## Popularisation of research results

- Conference at the Mathematics Olympiads (Académie de Versailles), December 2001: Alain Denise.
- Computer Science Booth at the National Science Fair (*Fête de la Science*), University Paris-Sud 11, 2002-2004: Alain Denise.

## Summer schools, tutorials, invited seminars

- Seminar GeneBio (Genève): Christine Froidevaux.
- Tutorial on random generation of combinatorial structures, journées ALEA 2002 of the national network GDR ALP: Alain Denise.

- Seminar, DIRO, university of Montréal, Canada, 2003: Alain Denise.
- Seminar, Dagstuhl Extracting local patterns workshop, co-organised by J.F. Boulicaut, K. Morik and A. Siebes, "Pattern Discovery in CGH array Data", 2004: Céline Rouveirol.
- Seminar, IRIT, "Programmation Logique Inductive : état de l'art et perspectives", 2002: Céline Rouveirol.
- Seminar Ecole de printemps de Dieppe, 2003: Christine Froidevaux, Patrick Amar.
- Seminar Ecole de printemps d'Evry, 2004: Patrick Amar.
- Seminar Rencontres Franco-taiwanaises, Ecole Polytechnique, 2004: Alain Denise.
- Seminar Algorithmique et Biologie, INRIA Rhône-Alpes, 2003: Stéphane Vialette.
- Seminars at LaBRI (Bordeaux), LPIC (Rouen) and LaMI (Evry): Patrick Amar.
- Invited Seminar, UCSD, Laboratory of Biology, San Diego, 2003: Patrick Amar.

## Other dissemination and technology transfer

- Seminar Langage du Génome, University Paris-Sud: Alain Denise, co-chair.
- Seminar, ECRIN (*Club Recherche et Entreprise*): Christine Froidevaux.

## B.10.10 / Bioinfo

# Training and education (doctoral and post-doctoral)

# **Defended habilitations**

Name	Date defended	Current position
DENISE Alain	10/12/01	Professor, BioInfo group

## Graduate courses

- Chair of Master program DEA I3 *Information, Interaction, Intelligence,* Université Paris-Sud, 2000-2003: Céline Rouveirol.
- Co-chair of the Master program *Bioinformatique et Biostatistiques*, Université Paris-Sud, 2004-: Alain Denise.
- DEA I3, Traitement de l'Information génomique, (created in September 2001): Christine Froidevaux.
- DEA I3, Fondements de la représentation des connaissances: Christine Froidevaux and Céline Rouveirol.
- DEA I3, *Fouille de données et apprentissage*, Céline Rouveirol.
- DEA Algorithmique, *Combinatoire énumérative et génération aléatoire* 2000-2002: Alain Denise.
- DEA Algorithmique, *Modèles combinatoires en bioinformatique*, 2002-2004: Alain Denise.

- DEA AMIB (Application des Mathématiques et de l'Informatique à la Biologie), *Algorithmes combinatoires pour la biologie*, Evry, 2001-2004: Alain Denise.
- Master "Recherche Bioinformatique et Biostatistiques", Combinatoire et algorithmique des structures moléculaires, 2004: Alain Denise & Stéphane Vialette.
- Master Recherche "Bioinformatique et Biostatistiques", Fouille de données dans de grandes sources de données hétérogènes, 2004: Céline Rouveirol & Christine Froidevaux.
- Master Recherche "Bioinformatique et Biostatistiques", Simulation de processus biologiques, 2004: Patrick Amar.
- Master Recherche "Informatique", *Formation à la recherche*, 2004: Alain Denise.

Research groups

## Bioinfo

Dissemination and technology transfer Training and education

# Other research-related teaching activities

- Mastère de Bioinformatique, IIE, 2004: Christine Froidevaux.
- Organisation of a one day training meeting on Swiss-Prot, University Paris Sud, Amos Bairoch, (Swiss Institute of Bioinformatics, Genève), September 2002: Christine Froidevaux.
- Member of the "*Comité de cours pour l'enseignement de l'informatique en biologie*", Institut Pasteur, since 2002, Christine Froidevaux.
- Post-graduate course, Ecole Doctorale de Mathématiques et Informatique Fondamentale, "Algorithmes en Bioinformatique", ENS Lyon, 2004: Alain Denise.
- Post-graduate course, Ecole Doctorale de Mathématiques et Informatique, "Algorithmique pour la bioinformatique", Université Bordeaux I, 2004: Alain Denise.

# **Publications**

## International peer-reviewed journals

- P. Amar, P. Ballet, G. Barlovatz-Meimon, A. Benecke, G. Bernot, Y. Bouligand, P. Bourguine, F. Delaplace, J.-M. Delosme, M. Demarty, I. Fishov, J. Fourmentin-Guilbert, J. Fralick, J.-L. Giavitto, B. Gleyse, C. Godin, R. Incitti, F. Képès, C. Lange, L. L. Sceller, C. Loutellier, O. Michel, F. Molina, C. Monnier, R. Natowicz, V. Norris, N. Orange, H. Pollard, D. Raine, C. Ripoll, J. Rouviere-Yaniv, M. S. jnr., P. Soler, P. Tambourin, M. Thellier, P. Tracqui, D. Ussery, J.-C. Vincent, J.-P. Vannier, P. Wiggins, and A. Zemirline. Hyperstructures, genome analysis and i-cell. *Acta Biotheoretica*, pages 357-373, 2002, Kluwer Academic Publishers.
- [2] P. Amar, G. Bernot, and V. Norris. Hsim: a simulation programme to study large assemblies of proteins. *Journal of Biological Physics and Chemistry*, 4:79-84, 2004, Collegium Basilea & AMSI.
- [3] C. Banderier, M. Bousquet-Mélou, A. Denise, P. Flajolet, D. Gardy, and D. Gouyou-Beauchamps. Generating functions for generating trees. *Discrete Mathematics*, 246:29-55, 2002.
- [4] M. Bekaert, L. Bidou, A. Denise, G. Duchateau-Nguyen, J.-P. Forest, C. Froidevaux, I. Hatin, J.-P. Rousset, and M. Termier. Towards a computational model for eukaryotic -1 frameshifting sites. *Bioinformatics*, 19:327-335, 2003.
- [5] S. Cohen-Boulakia, S. Lair, N. Stransky, S. Graziani, F. Radvanyi, E. Barillot, and C. Froidevaux. Selecting biomedical data sources according to user preferences. In *Proc.* of ISMB/ECCB 2004, Bioinformatics Special Number, Vol 20 Suppl. 1, pages i86-i93, 2004.
- [6] S. Corteel, D. Gardy, M. Valencia-Pabon, D. Barth, and A. Denise. The permutation-path coloring problem on trees. *Theoretical Computer Science*, 297:119-143, 2003.
- [7] F. Képès, P. Amar, G. Barlovatz, G. Bernot, C. Froidevaux, J.-L. Giavitto, J. Guespin, F. Molina, V. Norris, and V. Schächter. Modelling and simulating biological processes in the genomic era: An account of a multidisciplinary thematic school held in Évry (France) in april 2004. *Journal of Biological Physics and Chemistry*, 4(2), 2004.
- [8] G. Mercier, N. Berthault, J. Mary, J. Peyre, A. Antoniadis, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. Biological detection of low radiation by combining results of two microarray analysis methods. *Nucleic Acids Research*, 2004, Vol.32 (1), e12, 2004.
- [9] V. Norris, P. Amar, G. Bernot, A. Delaune, C. Derue, A. Cabin-Flaman, M. Demarty, Y. Grondin, G. Legent, C. Monnier, H. Pollard, and D. Raine. Questions for cell cyclists. *Journal of Biological Physics and Chemistry*, 4:124-130, 2004, Collegium Basilea & AMSI.
- [10] V. Norris, P. Amar, G. Bernot, J.-L. Giavitto, C. Godin, J. Guespin, H. Pollard, P. Tracqui, and F. Képès. Modelling and simulation of biological processes in the context of genomics: Conference synthesis. *Journal of Biological Physics and Chemistry*, 3:106-110, 2003, Collegium Basilea & AMSI.
- [11] M. Régnier and A. Denise. Rare events and conditional events on random strings. *Discrete Mathematics and Theoretical Computer Science*, 6(2):191-214, 2004.

Research groups Bioinfo

Publications

## National peer-reviewed journals

[12] M.-C. Rousset, A. Bidault, C. Froidevaux, H. Gagliardi, F. Goasdoué, C. Reynaud, and B. Safar. Constructions de médiateurs pour intégrer des sources d'informations multiples et hétérogènes : le projet PICSEL. *Revue I3 (Information - Interaction - Intelligence)*, 2(1):9-59, 2002.

## Edited books

- [13] P. Amar, J.-P. Comet, F. Képès, and V. Norris, editors. Proceedings of the Évry spring school on Modelling and simulation of biological processes in the context of genomics. Platypus Press, 2004.
- [14] P. Amar, F. Képès, V. Norris, and P. Tracqui, editors. Proceedings of the DIEPPE spring school on Modelling and simulation of biological processes in the context of genomics. Platypus Press, 2003.
- [15] C. Rouveirol and M. Sebag, editors. *Machine Learning Journal, Special Issue on Inductive Logic Programming and Relational Learning*, volume 55. Kluwer Academic Press, 2004.

## **Book chapters**

[16] J. Mary, G. Mercier, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. *Informatique pour l'analyse du transcriptome*, chapter Utilisation d'une méthode de sélection d'attributs pour l'analyse du transcriptome de cellules de levure exposées à de faibles doses de radiation, pages 189-205. Informatique et systèmes d'information. Traité IC2 Hermes-Lavoisier, 2004.

## Major international peer-reviewed conferences

- [17] M. Bekaert, J.-P. Forest, L. Bidou, A. Denise, G. Duchateau-Nguyen, C. Fabret, C. Froidevaux, I. Hatin, J.-P. Rousset, and M. Termier. Searching the saccharomyces cerevisiae genome for -1 frameshifting sites. In *Proceedings of ECCB 2003 (European Conference on Computational Biology)*, pages 4-6, September 2003.
- [18] J. A. Bianchetti, C. Rouveirol, and M. Sebag. Constraint-based learning of long relational concepts. In C. Sammut, editor, *Proceedings of International Conference on Machine Learning, ICML02*, pages 35-42, 2002.
- [19] A. Bidault, S. Cohen-Boulakia, and C. Froidevaux. Preferences for queries in a mediator approach. In Proc. of ECAI-2004 (short paper), European Conference on Artificial Intelligence, 2004.
- [20] A. Bidault, C. Froidevaux, and B. Safar. Similarity between queries in a mediator. In Proc. of 15th European Conference on Artificial Intelligence, ECAI-2002, pages 235-239, 2002.
- [21] G. Blin, G. Fertin, and S. Vialette. New results for the 2-interval pattern problem. In Proc. of 15th Annual Symposium on Combinatorial Pattern Matching (CPM 2004), volume 3109 of Lecture Notes in Computer Science, pages 311-322. Springer, 2004.
- [22] A. Denise, M.-C. Gaudel, and S.-D. Gouraud. A generic method for statistical testing. In Proc. of 15th IEEE International Symposium on Software Reliability Engineering (ISSRE 2004), 2004.

- [23] A. Denise, Y. Ponty, and M. Termier. Random generation of structured genomic sequences. In *Proceedings of RECOMB 03 (Berlin)*, 2003. Poster.
- [24] A. Denise, M. Régnier, and M. Vandenbogaert. Assessing statistical significance of overrepresented oligonucleotides. In *Proceedings of WABI01 (Workshop on Algorithms in Bioinformatics)*, pages 85-97. Elsevier, August 2001. Lecture Notes in Computer Science 2149.
- [25] I. Fagnot, G. Lelandais, and S. Vialette. Bounded list injective homomorphism for comparative analysis of protein-protein interaction graphs. In *Proc. of 1st Algorithms and Computational Methods for Biochemical and Evolutionary Networks (CompBioNets 2004)*. KCL publications, King's College, London, 2004.
- [26] G. Kucherov, L. Noe, and Y. Ponty. Estimating seed sensibility on homogenous alignments. In IEEE, editor, *Proceedings of Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE04)*, pages 387-394, 2004.
- [27] R. Rivière, D. Barth, J. Cohen, and A. Denise. Shuffling biological sequences with motifs constraints. In Proc 1st Algorithms and Computational Methods for Biochemical and Evolutionary Networks (CompBioNets 2004). KCL publications, King's College, London, 2004.

## National peer-reviewed conferences

- [28] D. Abergel, S. Cohen-Boulakia, F. Lemoine, C. Froidevaux, and M. Termier. WInGS: A reliability controlled data warehouse for yeast. In *Proc. of JOBIM'04 (short paper), Journées Ouvertes, Biologie, Informatique et Mathématiques*, page 76, 2004.
- [29] J. Ales-Bianchetti, C. Rouveirol, and M. Sebag. Apprentissage contraint de longs concepts relationnels. In PUG, editor, *Proc. of Conférence d'Apprentissage (CAP 2002)*, pages 1-12, 2002.
- [30] A. Bidault, C. Froidevaux, and B. Safar. Proximité entre requêtes dans un contexte médiateur. In *Proc. of 13<sup>e</sup> Congrès AFRIF-AFIA, RFIA-02 (Reconnaissances des Formes et Intelligence Artificielle)*, pages 653-662, 2002.
- [31] G. Blin, G. Fertin, R. Rizzi, and S. Vialette. Pattern matching in arc-annotated sequences: new results for the APS problem. In *Proc. of JOBIM'04 (short paper), Journées Ouvertes, Biologie, Informatique et Mathématiques*, page 46, 2004.
- [32] S. Cohen-Boulakia, C. Froidevaux, and S. Lair. Interrogation de sources biomédicales : gestion des préférences de l'utilisateur. In *Proc. of EGC'04, Extraction et Gestion des Connaissances*, pages 53-64, 2004.
- [33] S. Cohen-Boulakia, C. Froidevaux, E. Waller, and B. Labedan. Genopage : A database of all protein modules encoded by completely sequenced genomes. In *Proc. of JOBIM'02* (*short paper*), *Journées Ouvertes, Biologie, Informatique et Mathématiques*, pages 187-193, 2002.
- [34] A. Denise, M. Bekaert, L. Bidou, G. Duchateau-Nguyen, J.-P. Forest, C. Froidevaux, I. Hatin, J.-P. Rousset, and M. Termier. Vers une modélisation des gènes à décalage de cadre en -1 dans les génomes eucaryotes. In *Proc. of JOBIM 2001, Journées Ouvertes, Biologie, Informatique et Mathématiques*, pages 229-237, 2001.
- [35] C. Froidevaux and S. Cohen-Boulakia. Intégration de sources de données génomiques du web. In *Journées scientifiques du Web Sémantique (electronic proc.)*, 2002. http://www.lalic.paris4.sorbonne.fr/stic/octobre/programme0209.html.
- [36] J. Mary, G. Mercier, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. Utilisation d'une méthode de sélection d'attributs pour l'analyse du transcriptome de cellules de levure exposées à de faibles doses de radiation. In *Journées Post-Génomique de la Doua* (*JPGD-03*), Lyon, France, Mai 14-15, 2003.

Research groups Bioinfo Publications

- [37] Y. Ponty. GenRGenS : Generation of random genomic sequences. In Proc. of JOBIM'02, Journées Ouvertes Biologie Informatique et Mathématiques(JOBIM 02), 2002. poster.
- [38] R. Rivière. Génération aléatoire uniforme de shuffling de séquences avec contraintes de motifs. In *Proc. of Conférence JOBIM'04, Journées Ouvertes, Biologie, Informatique et Mathématiques*, 2004. Poster.
- [39] C. Rouveirol, N. Stransky, and F. Radvanyi. Vers l'identification de nouveaux oncogènes et gènes suppresseurs de tumeurs. analyse automatique des données de puces CGH. In *Proc. of JOBIM'04, Journées Ouvertes, Biologie, Informatique et Mathématiques,* 2004.
- [40] B. Safar, A. Bidault, C. Froidevaux, H. Gagliardi, and C. Reynaud. Production de réponses coopératives dans un contexte médiateur. In *Journées scientifiques du Web Sémantique (electronic proc.)*, 2002.

#### Other conferences and workshops

- [41] P. Amar, G. Bernot, and V. Norris. Modelling and simulation of large assemblies of proteins. In *Proceedings of the DIEPPE spring school on Modelling and simulation of biological processes in the context of genomics*, pages 41-48. Platypus Press, 2003.
- [42] G. Bernot, J. Guespin-Michel, J.-P. Comet, P. Amar, A. Zemirline, F. Delaplace, P. Ballet, and A. Richard. Modelling, observability and experiment: a case study. In *Proceedings of the DIEPPE spring school on Modelling and simulation of biological processes in the context of genomics*, pages 49-56. Platypus Press, 2003.
- [43] F. Képès, P. Amar, G. Barlovatz, G. Bernot, C. Froidevaux, J.-L. Giavitto, J. Guespin, F. Molina, V. Norris, and V. Schachter. Modelling and simulating biological processes in the genomic era: An account of a multidisciplinary thematic school held in Évry. In *Proceedings of the Évry spring school on Modelling and simulation of biological processes in the context of genomics*, pages 9-16. Platypus Press, 2004.
- [44] J. Mary, G. Mercier, J.-P. Comet, A. Cornuéjols, C. Froidevaux, and M. Dutreix. An attribute estimation technique for the analysis of microarray data. In *Proc. of Dieppe Spring school on Modelling and simulation of biological processes in the context of genomics* (eds. P. Amar, F. Képès, V. Norris and P. Tracqui) Publisher Frontier group, pages 69-77, 2003.
- [45] V. Norris, P. Amar, P. Ballet, G. Bernot, F. Delaplace, M. Demarty, J.-L. Giavitto, C. Ripoll, M. Thellier, and A. Zemirline. Hyperstructures. In *Proc. of colloque d'Autrans : Modelling and simulation of biological processes in the context of genomics*, pages 169-192. Génopole d'Évry, 2002.
- [46] A. Zemirline, P. Ballet, L. Marc, P. Amar, G. Bernot, F. Delaplace, J.-L. Giavitto, O. Michel, J.-M. Delosme, R. Incitti, P. Bourgine, C. Godin, F. Képès, P. Tracqui, V. Norris, J. Guespin, M. Demarty, and C. Ripoll. Cellular automata, reaction-diffusion and multiagents systems for artificial cell modelling. In *Proc. of Colloque d'Autrans : Modelling and simulation of biological processes in the context of genomics*, pages 257-280. Génopole d'Évry, 2002.
- [47] A. Zemirline, P. Ballet, L. Marc, P. Amar, G. Bernot, F. Delaplace, J.-L. Giavitto, O. Michel, J.-M. Delosme, R. Incitti, P. Bourgine, C. Godin, F. Képès, P. Tracqui, V. Norris, J. Guespin, M. Demarty, and C. Ripoll. Neural networks. In *Proc. of colloque d'Autrans : Modelling and simulation of biological processes in the context of genomics*, pages 281-293. Génopole d'Évry, 2002.

# **Doctoral dissertations and Habilitations**

[48] A. Denise. Structures aléatoires, modèles et analyse des génomes. Technical Report 1301, LRI Orsay, December 2001. Habilitation à Diriger des Recherches.

## Reports

[49] G. Kucherov, L. Noe, and Y. Ponty. Estimating seed sensibility on homogenous alignments. Technical report, LORIA, 2003.

Research groups

Bioinfo

Publications

# Glossary

А	Allocataire.	CIFRE	Convention Industrielle de Formation
	Doctoral Student with government grant.		par la Recherche.
AC	Allocataire Couplé.		Funding scheme for joint academia-industry
	Doctoral Student with joint grant between the		training of doctoral students.
	government and another organisation.		http://www.anrt.asso.fr
ACI	Action Concertée Incitative.	CNAM	Conservatoire National des Arts & Métiers.
	National Research Program funding academ-		National Centre for adult training, techno-
	ic research projects. Some recent programs:		logical research and innovation, and spread-
	- ACI GRID: Glabolisation de Ressources		ing of scientific and technical knowledge.
	Informatiques Distribuées - grid comput-		http://www.cnam.fr
	ing.	CNET	Centre National d'Études des Telecom-
	- ACI IMPBio: Informatique, Mathéma-		munications.
	tiques, Physique en Biologie moléculaire		National Research Centre for Telecom-
	- computer science, mathematics and physics		munications. Became FT R&D in 2000.
	for molecular biology.		http://www.cnet.fr
	- ACI MD: Masses de Données - masses of	CNRS	Centre National de la Recherche Scien-
	data.		tifique.
	- ACI NIM: Nouvelles Interfaces des		National Centre for Scientific Research.
	Mathématiques - new interfaces with Math-		http://www.cnrs.fr
	ematics.	CNRS-STI	IC
	- ACI SI ACI Sécurité Informatique - com-		Département Sciences et Technologies de
	puter security.		l'Information et de la Communication du
AI JC	Action Incitative Jeunes Chercheurs.		CNRS.
	National research program for junior res-		CNRS department in Science and Technolo-
	earchers.		gy of Information and Communication.
AM	Allocataire Moniteur.		http://www.cnrs.fr/STIC
	Doctoral Student with government grant and	CR	Chargé de Recherche.
	part-time lecturer position.		Research Scientist.
AS	Action Spécifique CNRS-STIC.		– CR2 Research Scientist second class.
	National working group of CNRS STIC dep-		– CR1 Research Scientist first class.
	artment.	D	Doctorant. Doctoral Student.
ATER	Attaché Temporaire d'Enseignement et de	DEA	Diplôme d'Études Approfondies.
	Recherche. Lecturer.		French Master program, now called Master.
BQR	Bonus Qualité-Recherche.	DR	Directeur de Recherche.
	University research grant.		Research Director.
CEA	Commissariat à l'Energie Atomique /		– DR2 Research Director second class.
	Centre d'Études Atomiques.		– DR1 Research Director first class.
	Public organisation for atomic and nuclear	EGIDE	Non-profit organisation that manages inter-
	research. http://www.cea.fr		national cooperation initiatives.
CEPHYTEN			http://www.egide.asso.fr
	Centre d'Études de Physique Théorique	ENS	École Normale Supérieure.
	Et Nucléaire.		Élite public institution for research and high-
	Non-profit organization that manages		er education.
	research contracts on behalf of Universtité		http://www.ens.fr
	Paris-Sud.	EUDIL	École Universitaire d'Ingénieurs de Lille.
			University Engineering School, Lille,

http://www.eudil.fr

FP5	European Fifth Framework Program (1998-2002).		INRIA Futurs are the three member labora- tories of the PCRI.
	http://www.cordis.lu/fp5		http://www-futurs.inria.fr
FP6	European Sixth Framework Program	INSERM	Institut National de la Santé et de la
	(2003-2007). http://fp6.cordis.lu		Recherche Médicale.
FT R&D	France Telecom Recherche et Développe- ment.		National Institute of Health and Medical Research.
	Research centre of France Telecom. formerly		http://www.inserm.fr
	known as CNET.	IRISA	Institut de Recherche en Informatique et
	http://www.rd.francetelecom.fr		Systèmes Aléatoires.
GdR	Groupement de Recherches.		Research Centre in Computer Science joint
	National Research Network funded by CNRS.		between INRIA, CNRS, Université de Rennes
	– GdR ALP: Algorithms, Languages & Pro-		and INSA Rennes.
	gramming.		http://www.irisa.fr
	– GdR ARP: Architecture, Network and Sys-	IST	Information Society Technologies.
	tems, Parallelism.		The major European research program for
	- GdR I3: Information, Interaction, Intelli-		computer science.
	gence.		http://www.cordis.lu/ist
HDR	Habilitation à Diriger des Recherches.	IUF	Institut Universitaire de France.
	Required degree for applying to a professor		University Institute of France, an institution
	position. Usually defended six to ten years		whose members are University faculty recog-
	after one's Ph.D.		nized for their excellence.
IGM	Institut de Génomique et de Microbiolo-		http://www.cpu.fr/luf
	gie.	IUT	Institut Universitaire de Technologie.
	Institute of Genetics and Microbiology, Uni-		University Institute of Technology.
	versité Paris-Sud & CNRS.	LIMSI	Laboratoire d'Informatique et de Méca-
	http://www.igmors.u-psud.fr		nique pour les Sciences de l'Ingénieur.
IIE	Institut d'Informatique d'Entreprise.		Laboratory for Computer Science and Mechan-
	Information Technology Institute, Evry.		ical Engineering. LRI and LIMSI are the two
	http://www.iie.cnam.fr		research laboratories of the Computer Science
IMPG	Action Incitative Informatique, Mathé-		Department at Université Paris-Sud.
	matiques, et Physique pour la Géno-		http://www.limsi.fr
	mique.	LIX	Laboratoire d'Informatique de l'École
	National research program on computer sci-		Polytechnique.
	ence, mathematics and physics for genomics.		Laboratory for Computer Science at École
	Followed up by ACI IMPBio.		Polytechnique. LRI, LIX and INRIA Futurs
INAPG	Institut National Agronomique Paris-		are the three member laboratories of the
	Grignon.		PCRI.
	National Agronomics Institute, Paris.		http://www.lix.polytechnique.fr
	http://www.inapg.inra.fr	LORIA	Laboratoire Lorrain de Recherche en
INRA	Institut National de Recherche Agrono-		Informatique et ses Applications.
	mique.		Research Centre in Computer Science joint
	National Institute for Agronomics Research.		between INRIA, CNRS, Universités Nancy 1
	http://www.inra.fr		& 2 and Institut National Polytechnique de
INRIA	Institut National de Recherche en Infor-		Nancy.
	matique et Automatique.		http://www.loria.fr
	National Institute for Research in Computer	LRI	Laboratoire de Recherche en Informatique.
	Science and Automation.		Laboratory for Computer Science, Universite
	http://www.inria.tr		Paris-sua & CNRS.
iinkia fu	turs	MC	nttp://www.iri.tr
	The request DIDIA Descende Contra to the	MC	Maine de Conferences.
	i ne newest INKIA Kesearch Centre, partial-		ASSISTANT FTOJESSOF.
	iy iocatea on the Plateau de Saciay, near the		- MCHC Mattre de Conferences Hors
	Orsay University Campus. LRI, LIX and		Classe.

MENRT	Ministère de l'Éducation Nationale, de la Recherche et de la Technologie. <i>Ministry in charge of Education, Research</i> <i>and Technology.</i>	SOCRATE	European Program for Training and Educa- tion. http://europa.eu.int/comm/education/ programmes/socrates/socrates en.html
	http://www.recherche.gouv.fr	UE	Union Européenne.
NoE	Network of Excellence.		European Union.
	One of the instruments of the European	RNRT	Réseau National de Recherche en Télé-
	Framework Programme.		communications.
ONERA	Office National d'Études et de Recherch-		National Network for Research in Telecom-
	es Aérospatiales.		munications. Provides public funding for
	National aerospace research institution.		joint academia-industry project.
	http://www.onera.fr		http://www.telecom.gouv.fr/rnrt
Paris XI	Université Paris-Sud XI, also known as UPS.	RNTL	Réseau National des Technologies Logi-
PCRD	Programme-Cadre Recherche et Dévelop-		cielles.
	pement.		National Network on Software Technology.
	European Framework Program (FP) for		Provides public funding for joint academia-
	research and development.		industry project.
PCRI	Pôle Commun de Recherche en Informa-		http://www.telecom.gouv.fr/rntl
	tique du Plateau de Saclay.	SAIC	Service d'Activités Industrielles et Com-
	Federation of three research laboratories: LRI,		merciales.
	LIX and INRIA Futurs.		University department in charge of technol-
	http://www.pcri.info		ogy transfer.
PECO/NE	EI	SPECIF	Société des Personnels Enseignants et
	Pays d'Europe Centrale et Orientale -		Chercheurs en Informatique de France.
	Nouveaux Etat Indépendants.		Professional association of French computer
	Grants for cooperation with central and east-		science teachers and researchers.
	european countries and newly independant	CLIDIA	http://specif.org
DI	states.	SUBV	Subvention de recherche.
PI	CNRS multidisciplingra programs	LIDC	Research grant.
	DI TCAN Knowledge management	UPS	Universite Paris-Sua A1, also known as
	- PI-ICAN Knowledge manugement,		Pulls AI.
	information and communication		http://www.u-psuu.n
PICS	Programmes Internationaux de Coopéra-		
1105	tion Scientifique du CNRS.		
	CNRS programs for international scientific		
	cooperation.		
PPF	Programme Pluri-Formation.		
	Multi-disciplinary research grant funded by		
	the Ministry of research as part of the four-		
	year contract between each University and		
	the government.		
PR	Professeur des Universités.		
	University Professor.		
	– PR2 Professeur deuxième classe.		
	Professor second class.		
	– PR1 Professeur première classe.		
	Professor first class.		
	– PRCE Professeur classe exceptionnelle.		
	Professor emeritus.		
RTP	Réseau Thématique Pluridisciplinaire		
	CNRS-STIC.		
	National research network of CNRS STIC		
	department.		

383