Modelling globalized systems: challenges and examples

Cécile Germain-Renaud

Laboratoire de Recherche en Informatique

Université Paris Sud, CNRS, INRIA

Results from the GO collaboration

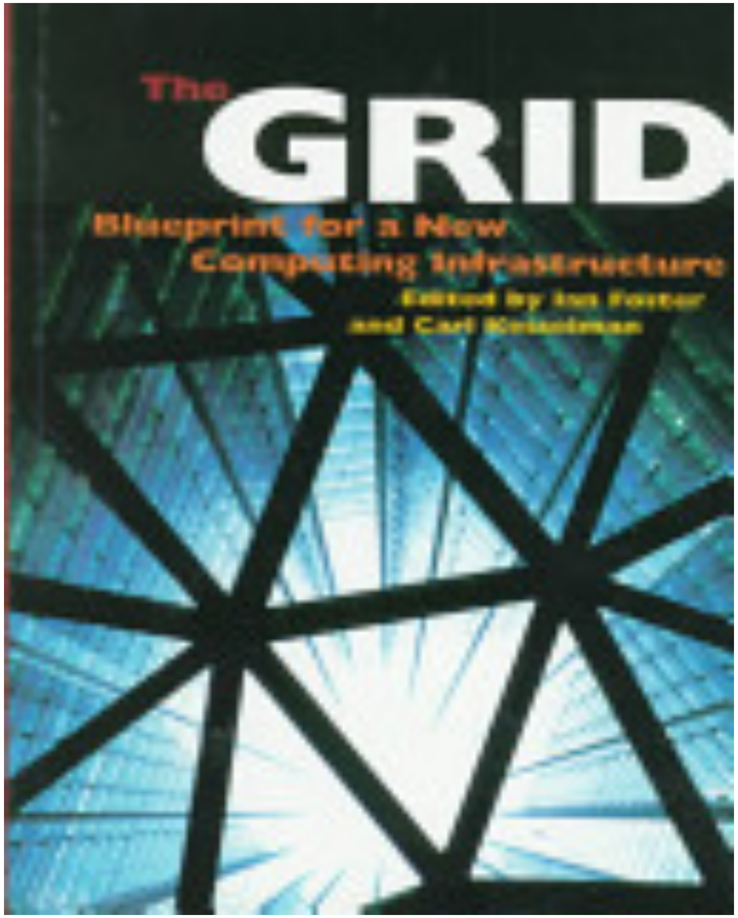
Outline

✓ Globalized systems

Remember tomorrow

A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to high-end computational capabilities

Ian Foster, 1998



The Clouds take me higher

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. NIST (US National Institute of Standards and Technology) definition of clouds

Yesterday's sorrows

How we configure our grids (EGEE 09)

Tomorrow's white lies?

Amazon's Cloud Crash Disaster Permanently Destroyed Many Customers' Data

Henry Blodget | Apr. 28, 2011, 7:10 AM | 77,816 | 76

Share | Tweet 1,297 | Like 1K | Email | A A A

71

In addition to taking down the sites of dozens of high-profile companies for hours (and, in some cases, days), Amazon's huge EC2 cloud services crash permanently destroyed some data.

The data loss was apparently small relative to the total data stored, but anyone who runs a web site can immediately understand how terrifying a prospect any data loss is.

(And a small loss on a percentage basis for Amazon, obviously, could be catastrophic for some companies).

Um...

Globalized systems

	Grid	Data Center	Cloud
Distribution	Very large	Any	Moderate
Sharing	Virtual Organisations – collective rights and control	No	Isolation – individualized access
Large data (file)	Yes	Yes	Yes
Big Data (indexed)	No	Yes	Yes
Economics	Long-term SLAs	Proprietary or usual commercial contract	Pay as you go

Outline

- ✓ Globalized systems
- ✓ Challenges

The challenges

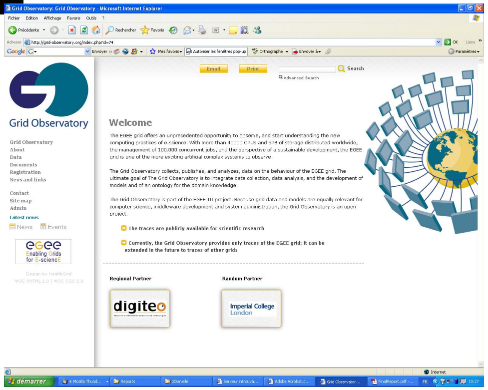
We need to show that the research has verifiable and positive impact on production systems

Demonstrating impact on complex systems

- requires experimental data
- raises serious scientific issues

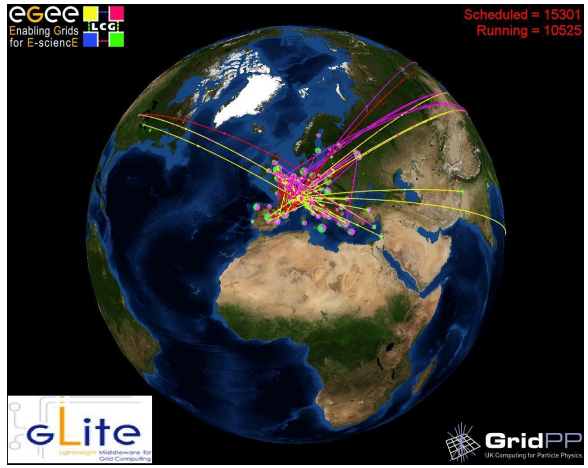
The Grid Observatory

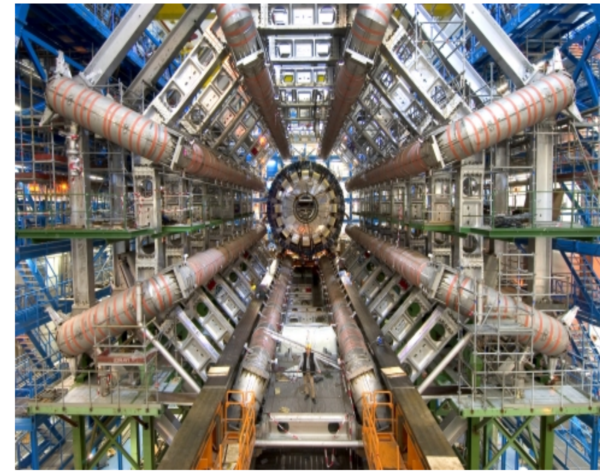
- Digital curation of the behavioural data of the EGI grid: observe and publish
- Complex systems description
- Models, optimization, Autonomics



Why the EGEE/EGI grid?

Accessible globalized production system





LHC is the

- Largest (26km),
- Fastest(14TeV)
- Coldest (1.9K)
- Emptiest (10–13 atm)

machine.

EGEE/EGI is the

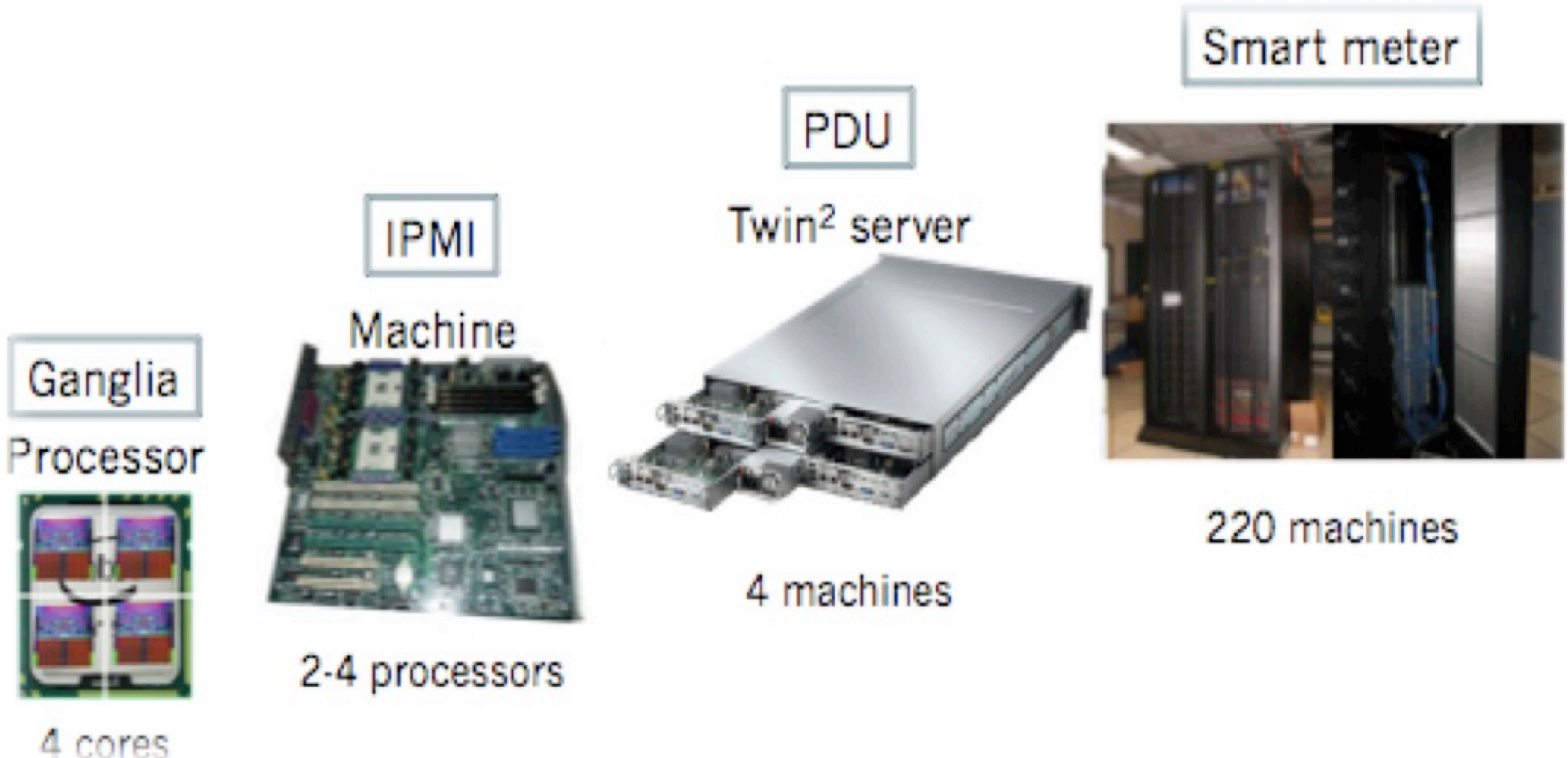
- Largest (40K CPUs),
- Most distributed (250 sites),
- Most used (300K jobs/day)

Computer system

Atlas Collaboration (one in four)

- 3000 scientists
- 38 countries
- 174 universities and labs

The Green Computing Observatory



2GBytes/day at 1 minute sampling period

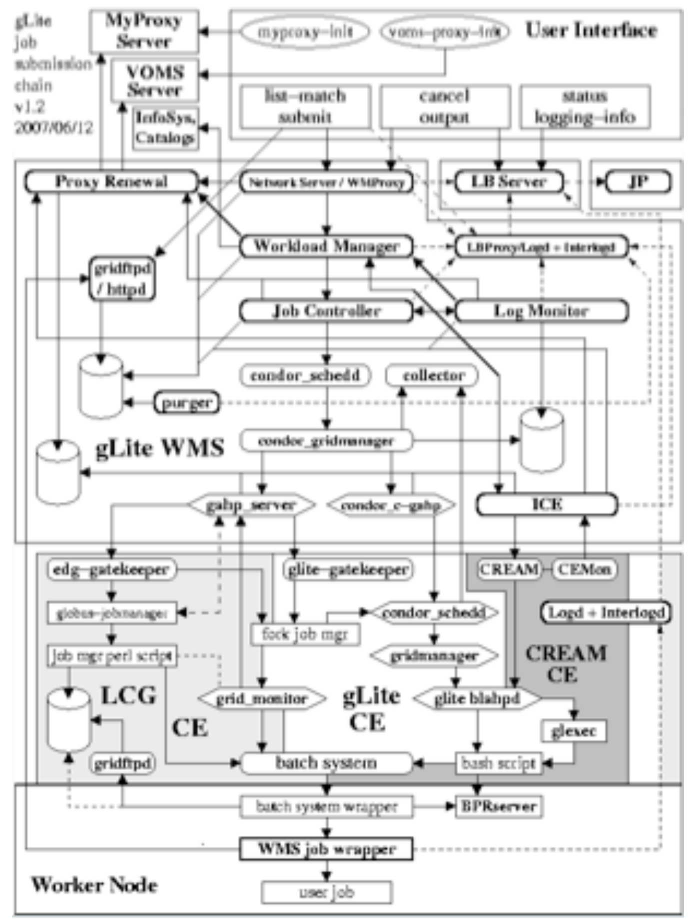
The Grid Observatory collaboration

- Born in EGEE-III, now a collaborative effort of
  - CNRS/UPS Laboratoire de Recherche en Informatique
  - CNRS/UPS Laboratoire de l'Accélérateur Linéaire
  - Imperial College London
  - France Grilles – French NGI of EGI
  - EGI-Inspire
  - Ile de France council
  - (Software and Complex Systems programme)
  - INRIA – Saclay (ADT programme)
  - CNRS (PEPS programme)
  - University Paris Sud (MRM programme)
- Scientific Collaborations
  - NSF Center for Autonomic Computing
  - European Middleware Initiative
  - Institut des Systèmes Complexes
  - Cardiff University

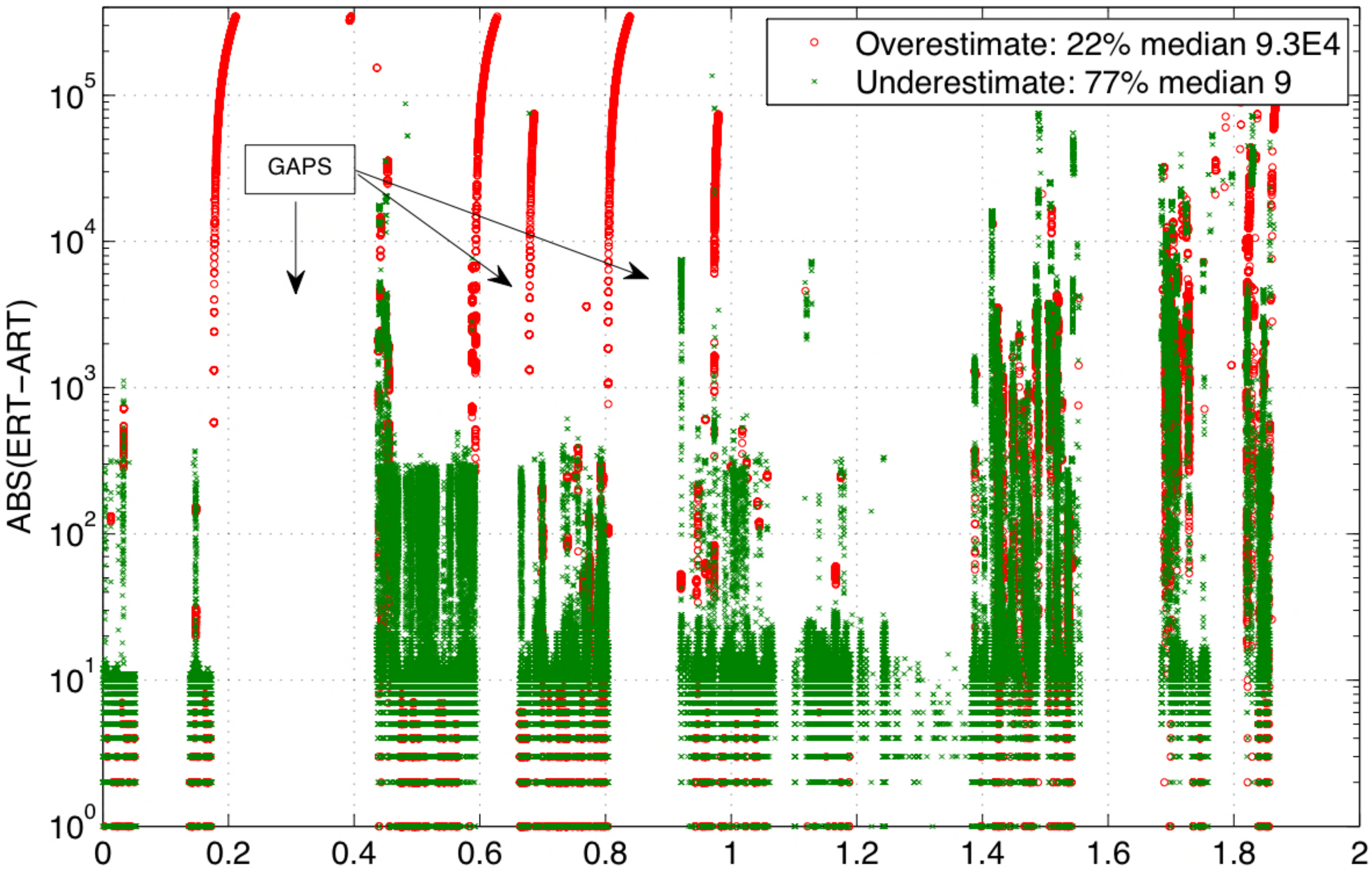
Globalized systems are complex ones

Dynamic(al) system

- Entities change behavior as an effect of unexpected feedbacks, emergent behavior
- Organized self-criticality, minority games,…



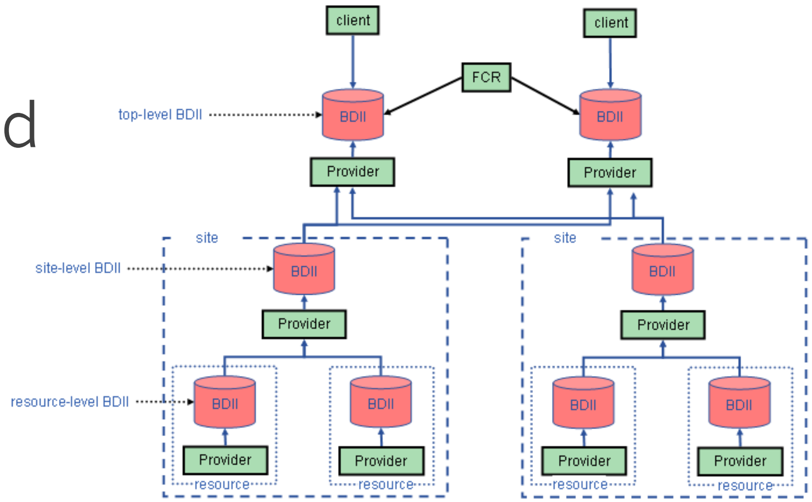
Predicting the response time



Globalized systems are complex ones

Lack of complete and common knowledge – Information uncertainty

- Monitoring is distributed too
- Resolution and calibration



Outline

- ✓ Globalized systems
- ✓ Challenges
- ✓ Towards realistic behavioural models

Issue I: Fundamentals in statistics

- “unusual” statistics: which metrics?
- Are our systems stationary?

Metrics

Root Mean Squared Error is inadequate

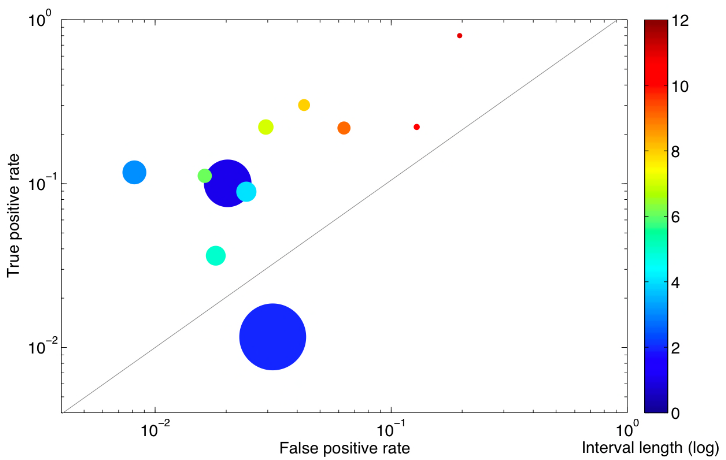
	Atlas		Biomed	
	ART	ERT	ART	ERT
Mean	1.33E3	2.74E4	3.01E2	2.66E2
Median	11	1	11	1
Std	1.09E4	7.41E4	4.33E3	5.99E3
RMSE	7.94E4		7.21E3	
$q_{90\%}$	1.35E2	1.16E5	25	4
Over. fraction	22%		3%	
Over. median	9.34E4		228	
Under. fraction	77%		96%	
Under. median	9.01E0		9.00E0	

Metrics

Should make sense for the end user

	Atlas		Biomed	
	ART	ERT	ART	ERT
Mean	1.33E3	2.74E4	3.01E2	2.66E2
Median	11	1	11	1
Std	1.09E4	7.41E4	4.33E3	5.99E3
RMSE	7.94E4		7.21E3	
$q_{90\%}$	1.35E2	1.16E5	25	4
Over. fraction	22%		3%	
Over. median	9.34E4		228	
Under. fraction	77%		96%	
Under. median	9.01E0		9.00E0	

The ROC metrics: à la BQP



- Evaluation of binary predictors: False positives vs true positive curve
- Intervals of the response time define as many binary predictors
- Intervals of increasing size
- gLite prediction is definitely better than random

[C. Germain-Renaud et al. The Grid Observatory. CCGRID 2011]

Statistical significance

Extreme values may dominate the statistics

	Atlas		Biomed	
	ART	ERT	ART	ERT
Mean	1.33E3	2.74E4	3.01E2	2.66E2
Median	11	1	11	1
Std	1.09E4	7.41E4	4.33E3	5.99E3
RMSE	7.94E4		7.21E3	
$q_{90\%}$	1.35E2	1.16E5	25	4
Over. fraction	22%		3%	
Over. median	9.34E4		228	
Under. fraction	77%		96%	
Under. median	9.01E0		9.00E0	

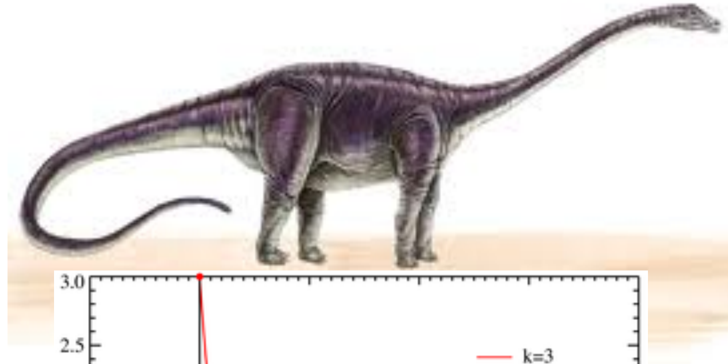
More on statistical significance

Can we predict anything?
Maybe as difficult as earthquakes and markets

	Atlas		Biomed	
	ART	ERT	ART	ERT
Mean	1.33E3	2.74E4	3.01E2	2.66E2
Median	11	1	11	1
Std	1.09E4	7.41E4	4.33E3	5.99E3
RMSE	7.94E4		7.21E3	
$q_{90\%}$	1.35E2	1.16E5	25	4
Over. fraction	22%		3%	
Over. median	9.34E4		228	
Under. fraction	77%		96%	
Under. median	9.01E0		9.00E0	

A few keywords

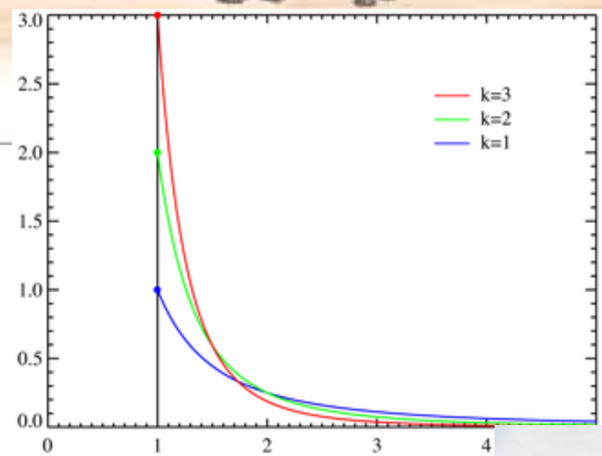
Heavy tail

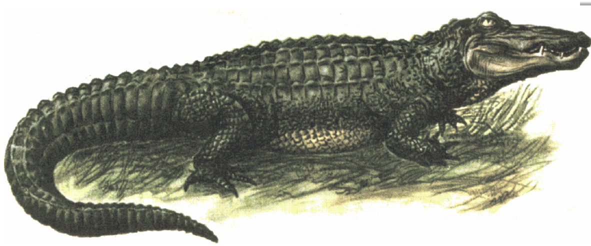


Self-similarity

Long range dependence

Heteroskedasticity





Harold Edwin Hurst
1880-1978

Stationarity

Joint probability distribution of the time series does not change when shifted

Do naïve statistics make sense?

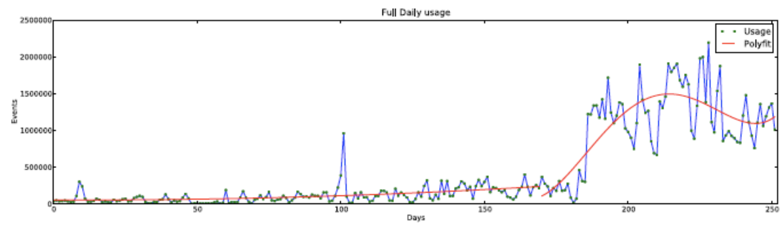
Non-stationarity and long-range dependence can easily be confused

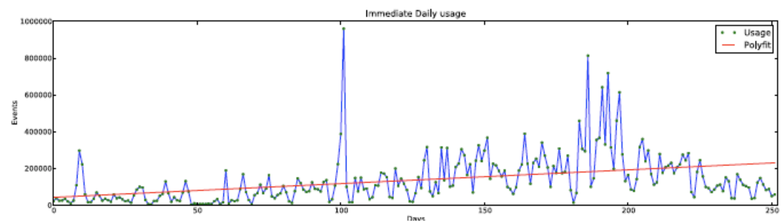
- The Hurst effect under trends. J. Appl. Probab., 20(3), 1983.
- Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. J. Empirical Finance, 11(3), 2004.
- Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. J. Time Ser. Anal., 18(3), 1997.
- Long memory and regime switching. J. Econometrics, 105(1), 2001.

Do naïve statistics make sense?

The "physical" process is not stationary

- Trends: Rogers's curve of adoption
- Technology innovations
- Real-world events
  - Experimental discoveries
  - Slashdotted accesses



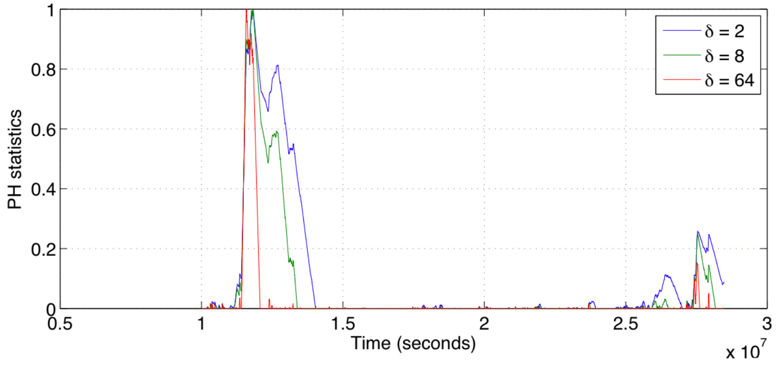


NON-STATIONARITY IS A REASONABLE ALTERNATIVE

Dealing with non-stationarity

1. Statistical testing

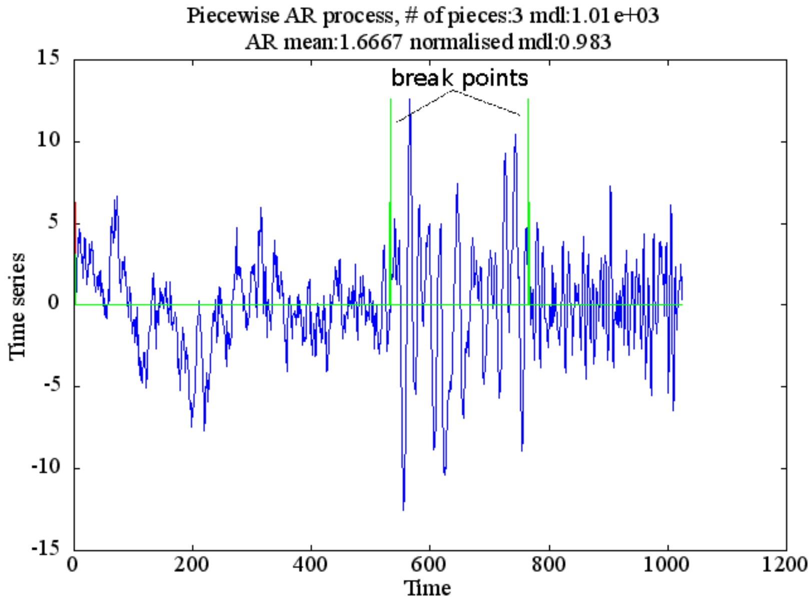
- Sequential jump detection
- Theoretical guarantees for known distributions
- Predictive, not generative
- Example: blackhole detection
- Calibration and Validation: by the Expert



Dealing with non-stationarity

2. Segmentation

- Fit a piecewise time-series: infer the parameters of the local models and the breakpoints
- Model selection: AIC, MDL,… – based
- a priori hypotheses on the segment models: AR, ARMA, FARMA,…

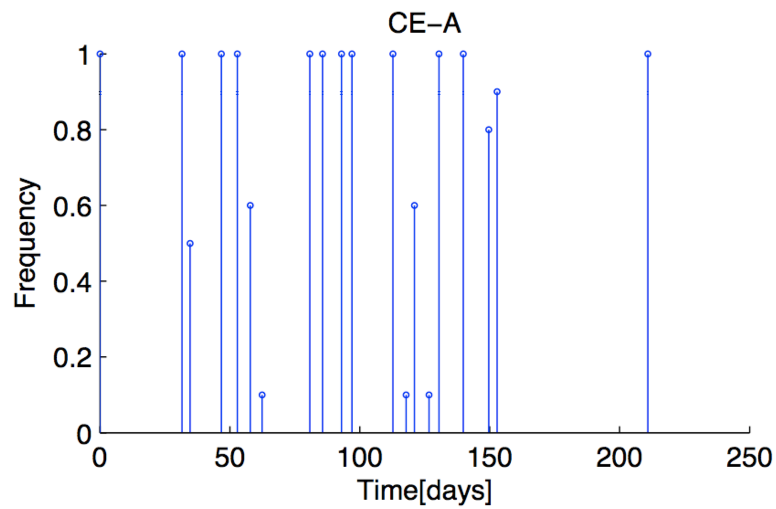


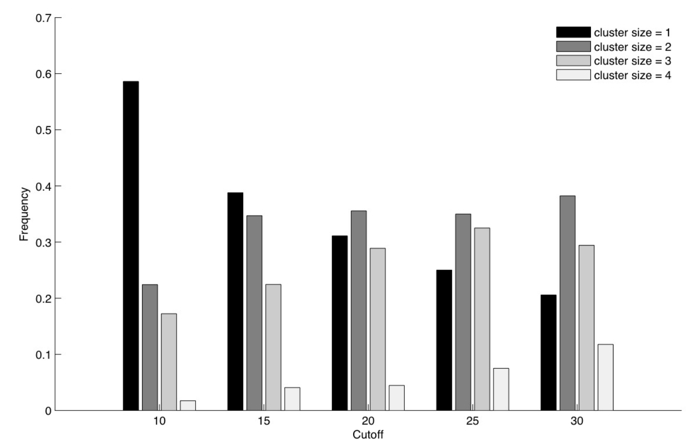
[Towards non stationary Grid Models, JoGC Dec. 2011]

Dealing with non-stationarity

2. Segmentation

- Mostly off-line and computationally expensive: generative, explanatory models
- Validation is not trivial
  - Fit quality
  - Stability: bootstrapping
  - Randomized optimization: clustering the results
- Hints at global behavior

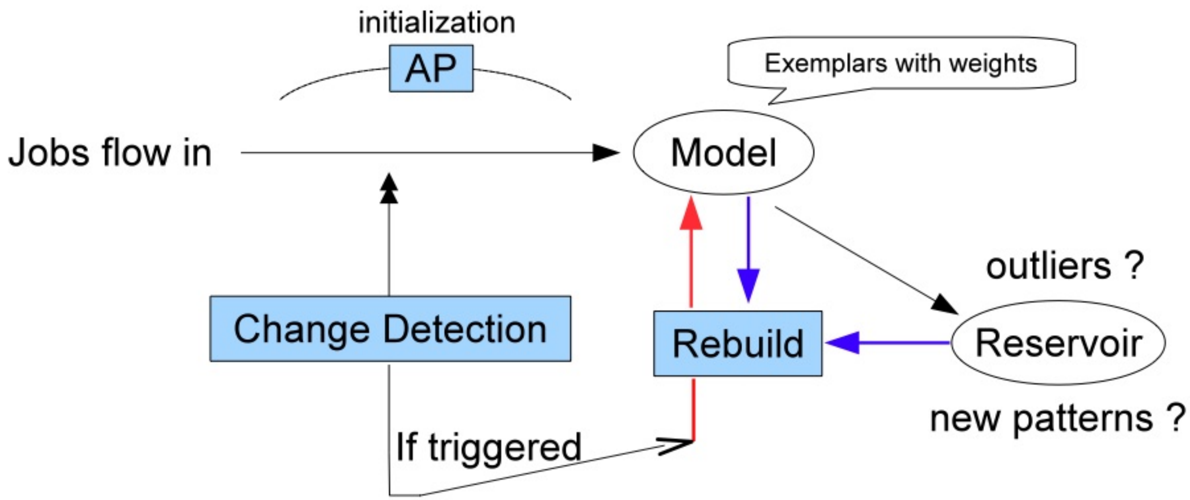




Dealing with non-stationarity

3. **Adaptive clustering**:

- Adaptive: on-line rupture detection
- Back to statistical testing, but on the model, not on the data

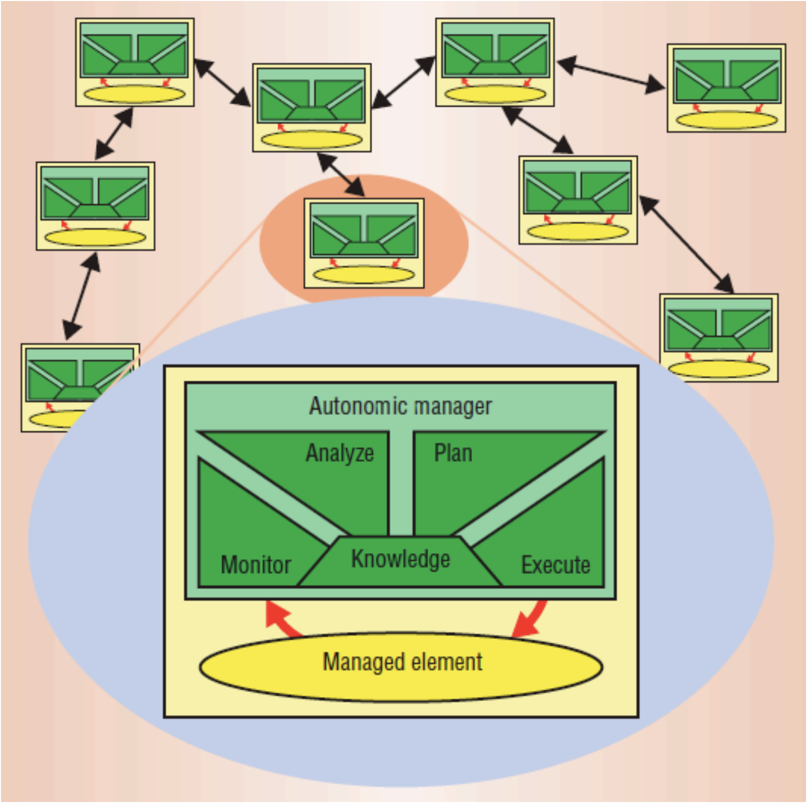


[Toward Autonomic Grids: Analyzing the Job Flow with Affinity Streaming". SIGKDD'2009]

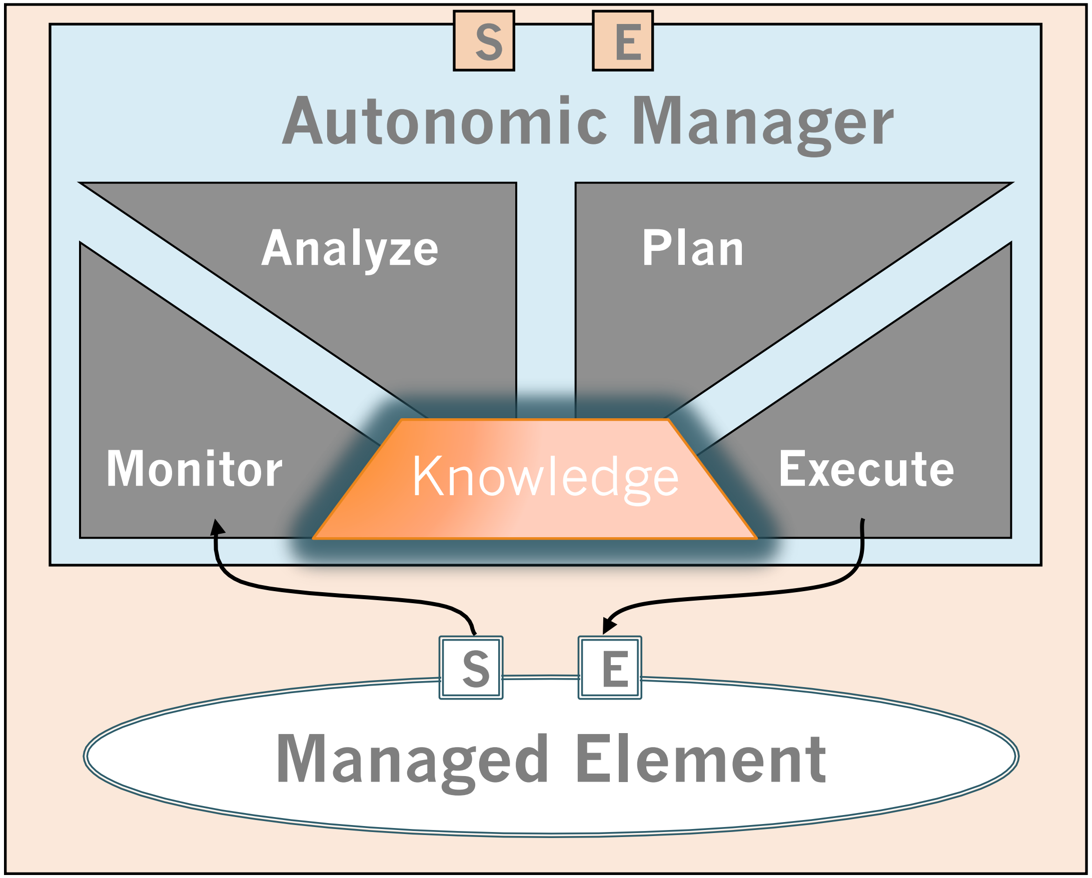
Remember tomorrow (5 years later)

Systems manage themselves according to an administrator's goals. New components integrate as effortlessly as a new cell establishes itself in the human body

J. Kephart and David M. Chess, the Autonomic Computing Manifesto, 2003



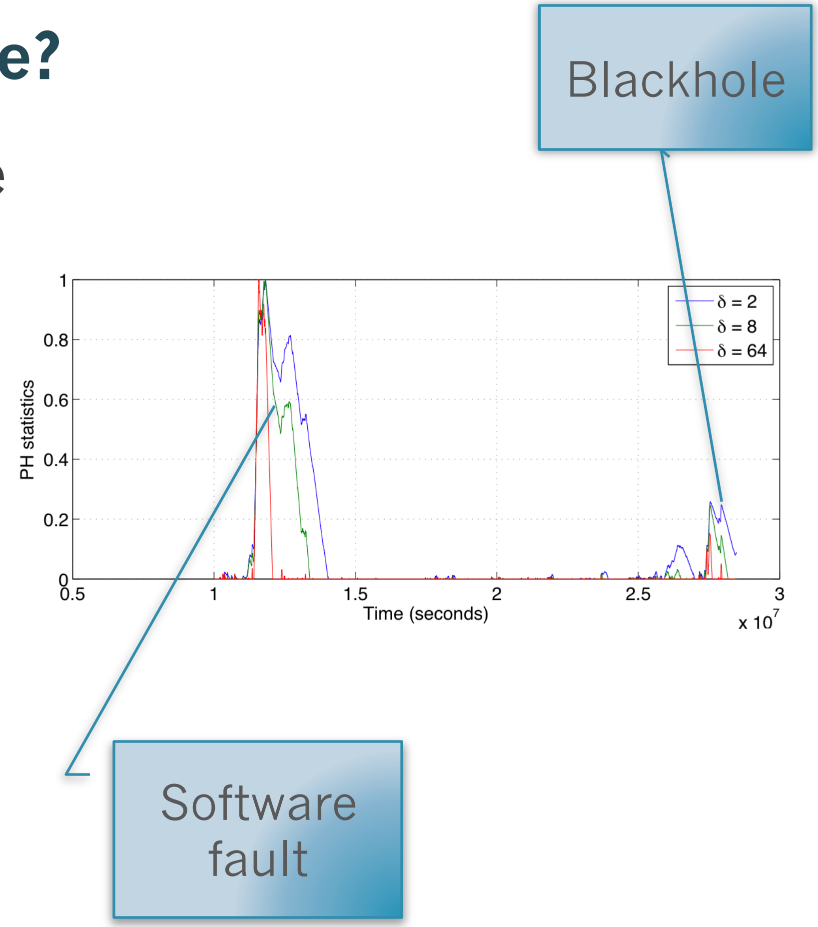
Issue II: Intelligibility



Issue II: Intelligibility

How to build the knowledge?

- No Gold Standard, too rare experts



Issue II: Intelligibility

How to build the knowledge?

- No Gold Standard, too rare experts
- Let's build it on-line! Model-free policies eg Reinforcement Learning!
- Unfortunately, tabula rasa policies and vanilla ML methods are too often defeated (Rish & Tesauro ICML 2006, Tesauro)

Exploration/exploitation tradeoff

… Issue II: Intelligibility

- Transaction traces are text files, thus we can infer causes from data as latent topics, in the spirit of text mining.

  [Characterizing E-Science File Access Behavior via Latent Dirichlet Allocation, UCC 2011]

- The internals of a globalized system might be so complex that it might be more effective to consider it as a black box, but the causes of failures or performance can be elucidated from external observation.

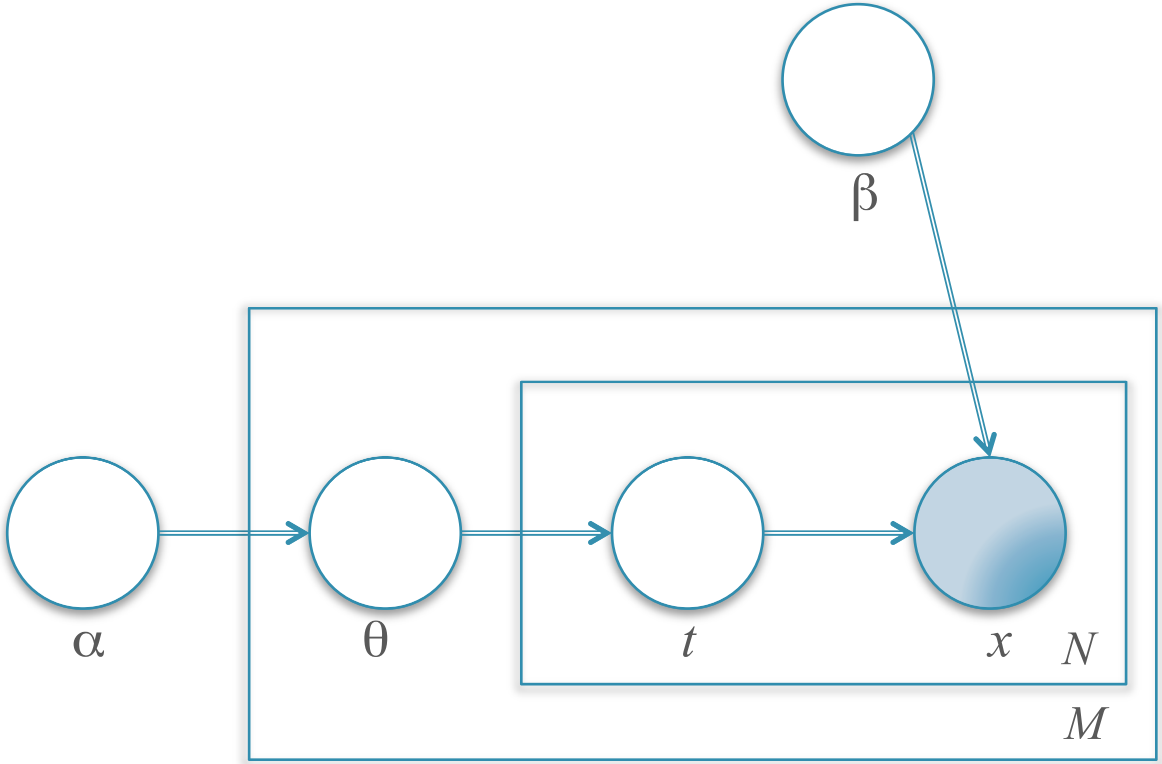
  [Distributed Monitoring with Collaborative Prediction. CCGrid 2012]

Latent Dirichlet Allocation...

- A corpus is a set of documents, each built over a dictionary (set of words)
  - A document is characterized by a mixture distribution over *topics*. Best example of topics: scientific keywords
  - A topic is characterized by a distribution over words.
  - The only observables are words.
  - Bag of words - interchangeability
- LDA is a generative model
  - For each document, choose the topic distribution.
  - For each topic, choose a word distribution.
  - For each word, choose:
    - the topic along the selected topic distribution
    - the word along the selected word distribution for this topic

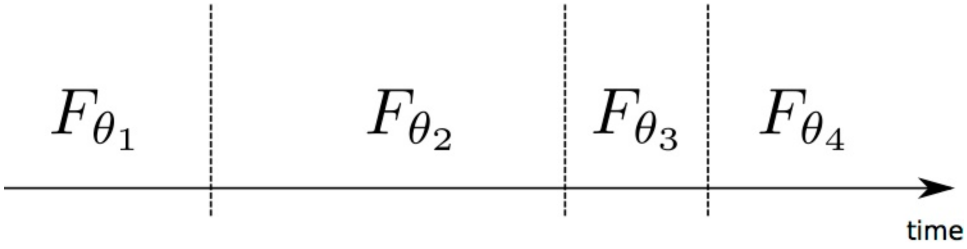
...Latent Dirichlet Allocation

M is the number of documents, N the size of a document



Analogy

- An analogy between text corpora and transaction traces
  - Corpus ~ Complete trace
  - Document ~ Segment of a trace (phase)
  - Topic ~ Activity
  - Word ~ Filename



Grid Observatory

And differences

- Unlike text corpora, trace files have...
  - No natural segmentation.
  - No well established, predefined set of activities equivalent to a set of topics.
- This work makes crude assumptions to avoid dealing with these issues.
  - 1 week phase
  - Arbitrarily fix the number of activities

Inference and parameter estimation

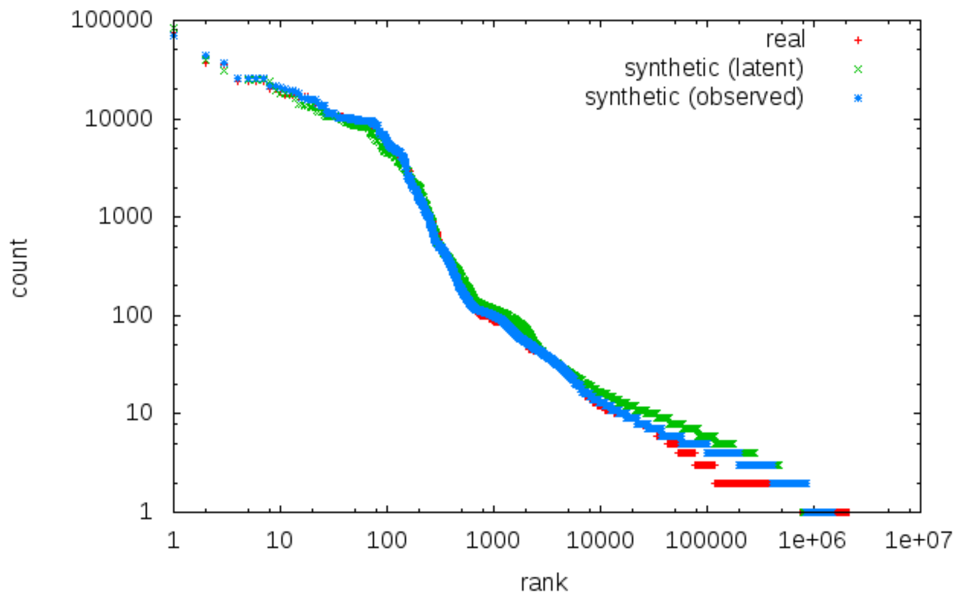
- Exact algorithms are intractable due to coupling between θ and β
- Alternating variational EM for the MLE estimates of α and β [Blei,Ng,Jordan, JMLR 2003]
- Gibbs sampling for estimating θ and Φ [Griffiths&Steyvers, Procs Nat Academy Science 2004]
- …

A tentative simpler model

- User data is included in each transaction thus is observable
- Assume each activity is associated with a unique user.
- Estimation and inference is much easier than standard LDA
- Goal: check the validity of this assumption

Experimental results

- Synthetic trace generated using the estimated parameters of the 2 models.
- 2M different files. 63 activities (standard LDA, number of clustered users), 262 activities (observed).
- File popularity: χ-square test gives p-value of 1

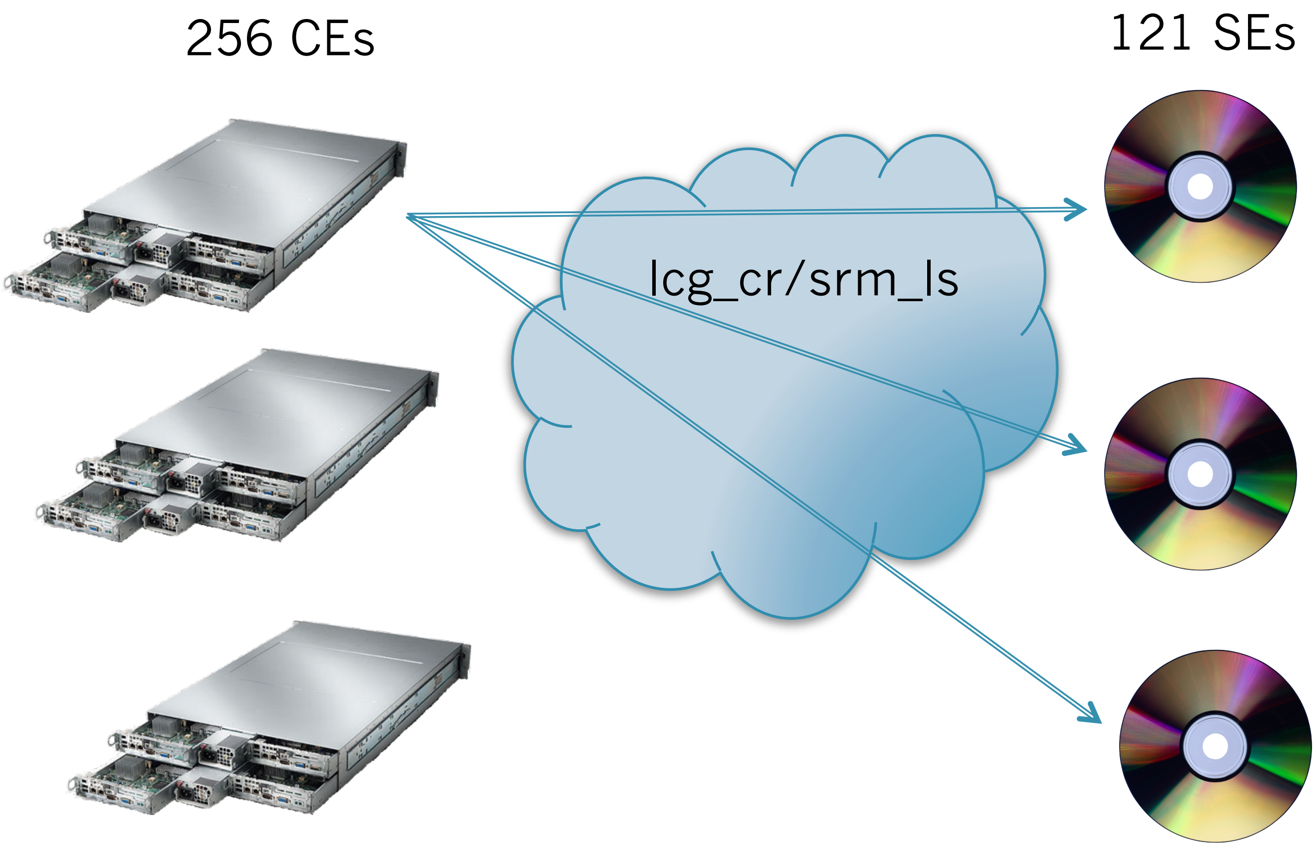


Ongoing work

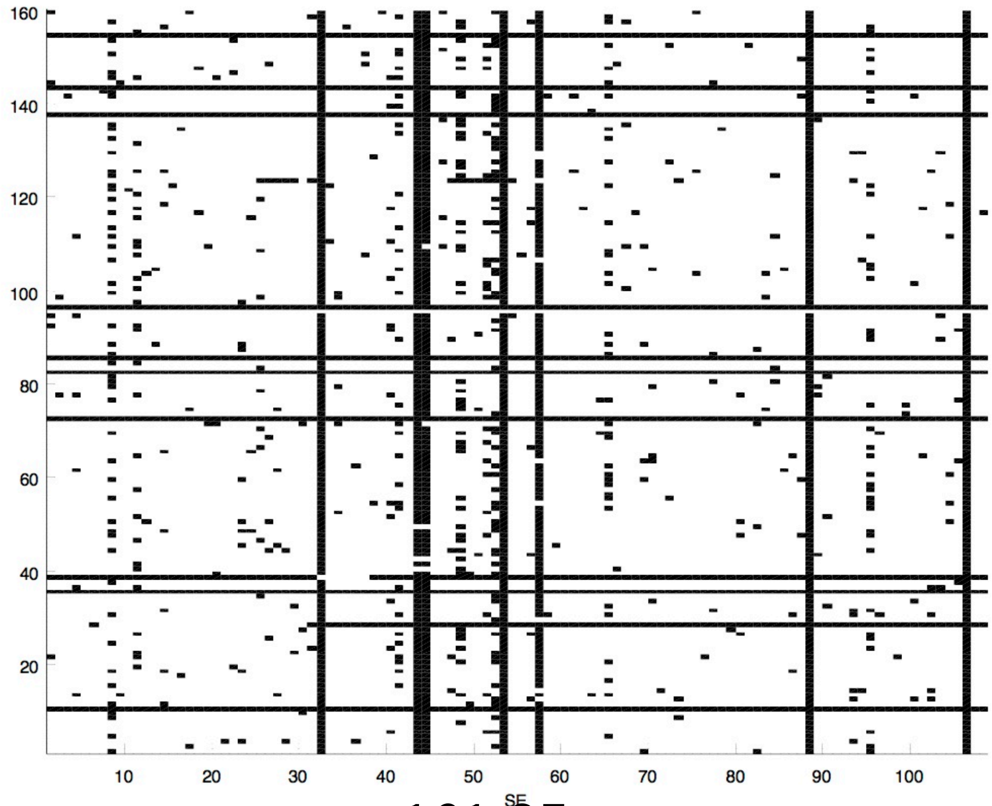
- Inferred segmentation: Probabilistic Context-Free Grammar

Fault management

Operational motivation: all (CExSE) pairs tests



All (CExSE) pairs tests



256 CEs

121 SEs

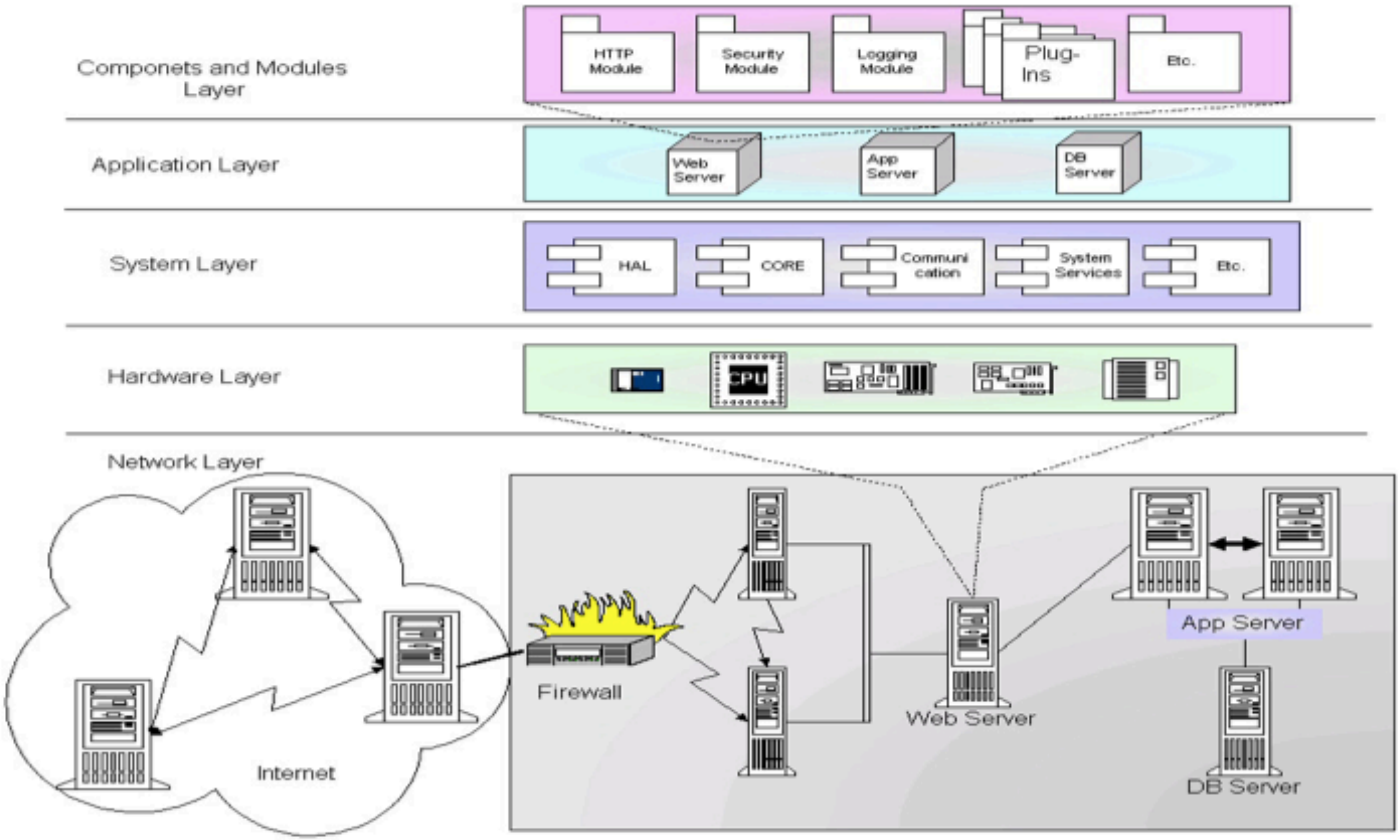
Goal: Detection/Diagnosis, or Prediction?

Detection/diagnosis: define a minimal set of probes that discovers all / any faulty component

Equivalent to the minimum cover set problem

Assumes that we know the internal dependencies

Assumes that we know the internal dependencies



Goal

Prediction! More precisely

- (Minimal) probe selection: choose which subset of the (CE,SE) pairs will actually be tested
- Prediction: predict the availability of all (CE,SE) pairs from a small number of them.

Less probes

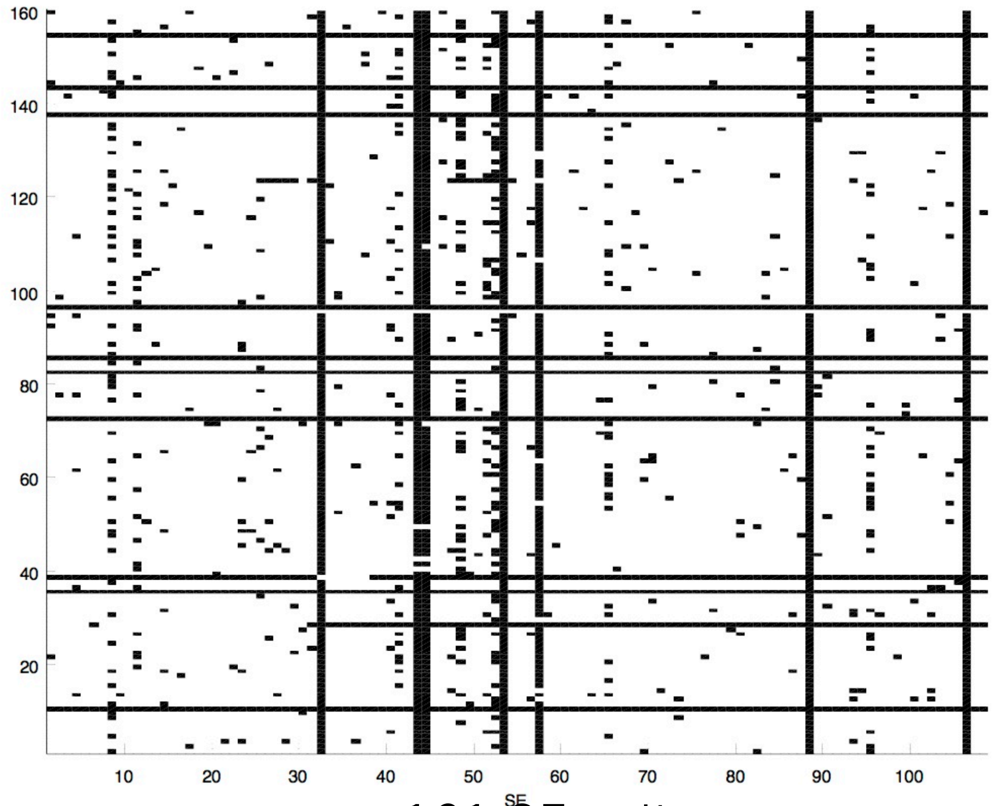
	-1		1	1
-1			1	
		1	1	
		1		-1
	1	1		1

	X	-1	X	1	1
	-1	X	X	1	X
Predicted	X	X	1	1	X
	X	X	1	X	-1
	X	1	1	X	1

All (CExSE) pairs tests

A case for collaborative filtering

256 CEs
~ users



121 SEs ~items

Grid Observatory

Collaborative filtering

- Major aplication: recommendation systems eg netflix challenge
- Neigborhood approach
- Latent factor models approach
  - Transform items AND users into the same latent factors space
  - Factors are *inferred* from data
  - Better if interpretable eg comedy, drama, action, scenery, music,... but this is another task

Latent topics: LDA

Implicit mapping to high-dimensional space: SVM

Maximum Margin Matrix factorization

(Srebro, Rennie, Jaakkola, NIPS 2005)

- Linear factor model

X the observed $n\text{x}m$ sparse matrix

$$X = UV \qquad\qquad U \text{ is } n\text{x}k,\ V \text{ is } k\text{x}m$$

each line i of U is a feature vector (« tastes » of user i)
each column j of V is a linear predictor for movie j

- Low-rank CP: regularizing by the rank k

Trace (or Frobenius) norm as a convex surrogate for rank

- With uniform sample selection, theoretical bounds on misclassification error: learning both U and V is within log factors of learning only one

Maximum Margin Matrix factorization

(Srebro, Rennie, Jaakkola, NIPS 2005)

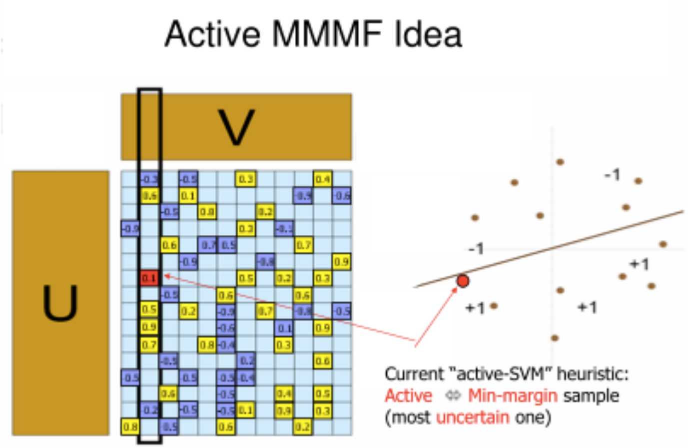
- Linear factor model
- Low-rank CP: regularizing by the rank k

Trace (or Frobenius) norm as a convex surrogate for rank

$$\|X\|_{\Sigma} + C \sum_{ij \in S} \max(0, 1 - X_{ij} Y_{ij})$$

- With uniform sample selection, theoretical bounds on misclassification error: learning both U and V is within log factors of learning only one

MMMF with Active Learning



- Rish & Tesauro, 2007
- Min margin selection: get the label for the most uncertain entry
- More applicable to system problems than to movies ratings
- In our case, just launch a probe

Evaluation

For once, we have ground truth: 51 days (March-April 2011) of all (CExSE) probes outcomes

Original

1	-1	1	1	1
-1	-1	1	1	1
1	1	1	1	-1
-1	-1	1	1	-1
-1	1	1	1	1

Selected

	-1		1	1
-1			1	
		1	1	
		1		-1
	1	1		1

Less probes

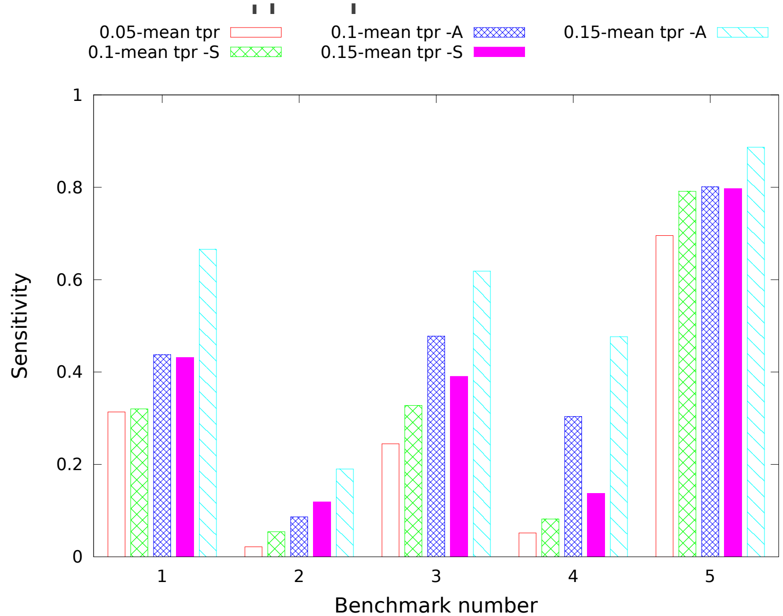
$\neq$?

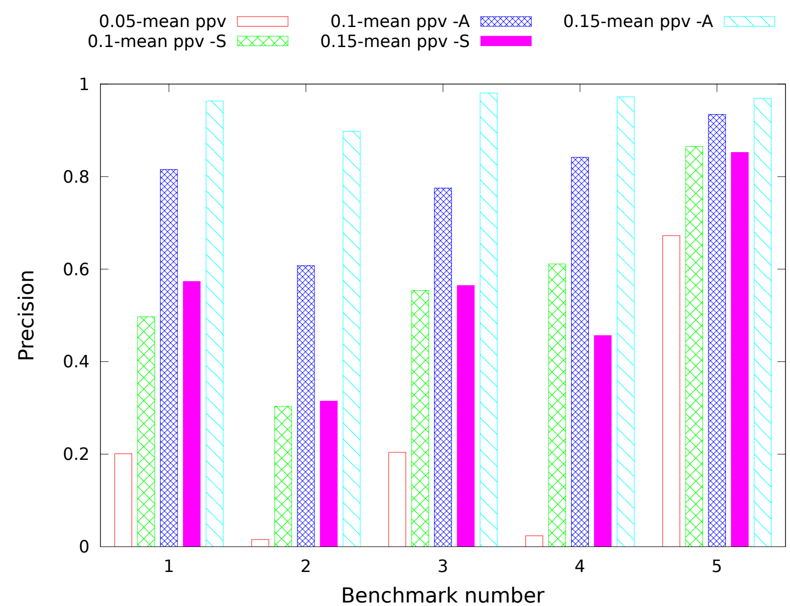
Predicted

X	-1	X	1	1
-1	X	X	1	X
X	X	1	1	X
X	X	1	X	-1
X	1	1	X	1

Evaluation

- Systematic failures: excellent results, too easy problem
- Without systematic failures: accuracy is excellent, but not a significant performance indicator
- MMMF-based Active probing
  - provides good results
  - outperforms M3F, a combined low-rank/latent topic



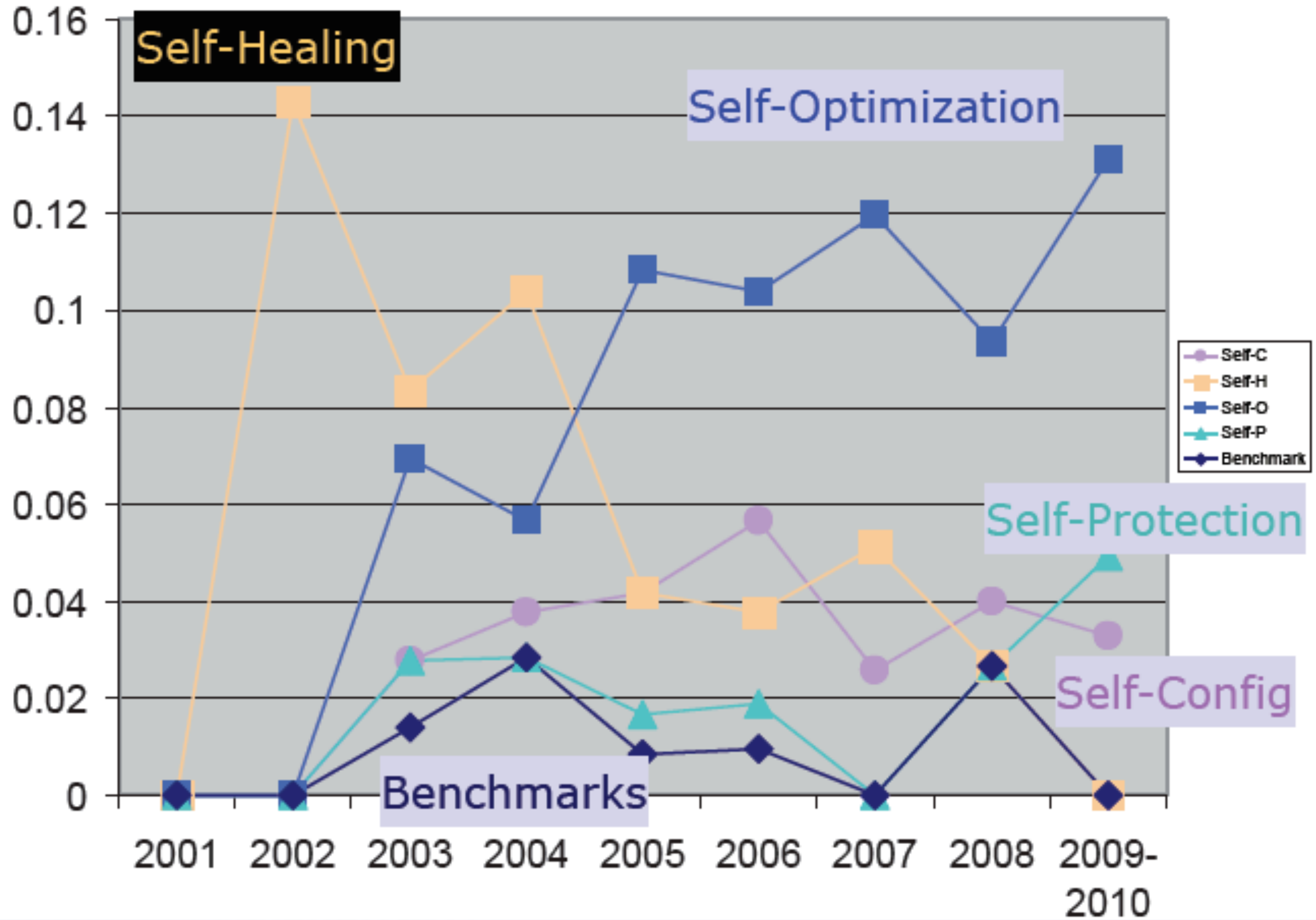


Outline

- ✓ Globalized systems
- ✓ Scientific challenges
- ✓ Towards realistic behavioural models
- ✓ Conclusion and questions

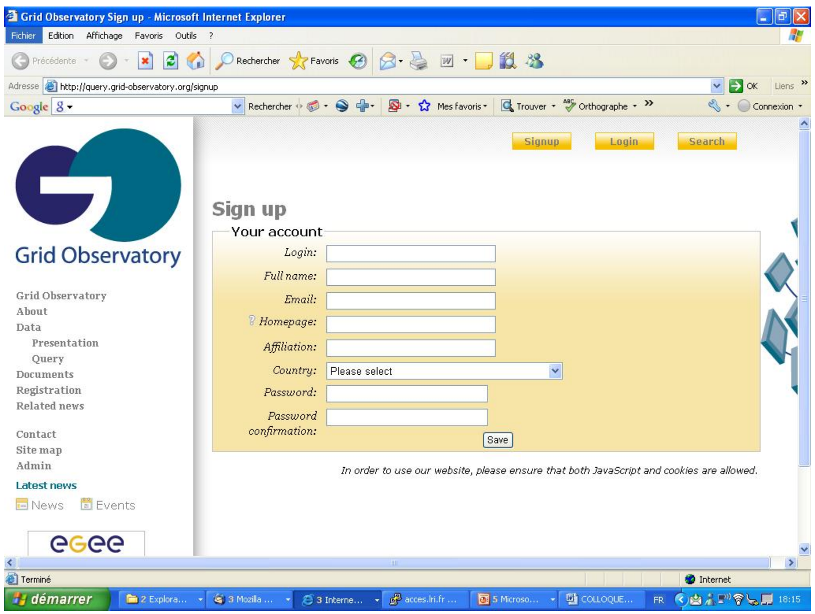
AC Paper Trends 2001-2010: Self-*, Benchmarks

- David Patterson warned us that we needed benchmarks for self-{C,H,P} in order to drive work in the field
- It appears that he was right
- **We need to revive the benchmark work**
- **We need more work on self-{C,H,P}**

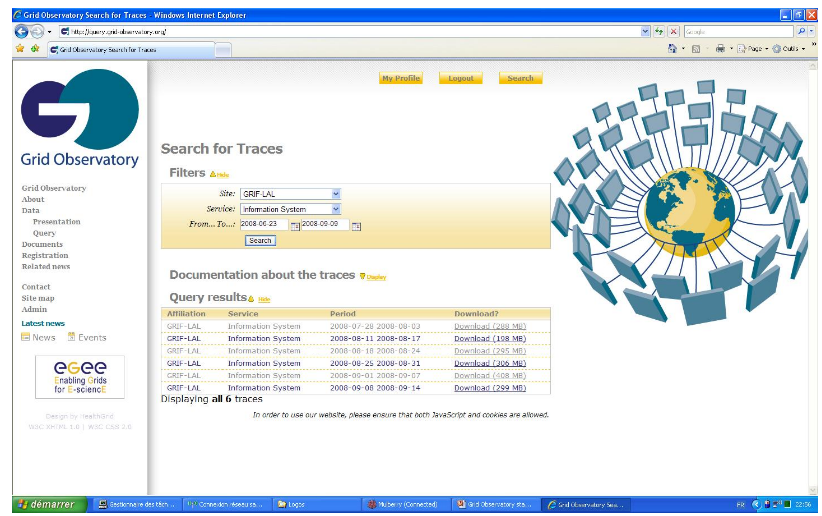


How to

- Get an account



- Download files



www.grid-observatory.org

Questions ?