

Sujet de TER

Détection de ruptures, application à la tolérance aux défaillances

Sujet

Le fonctionnement des grands systèmes de calcul et de stockage répartis n'est pas stationnaire. Il présente en particulier des pics de charge dus à des externalités. Ces pics peuvent surcharger des composants logiciels ou matériels du système, et entraîner des défaillances. L'objectif du TER est de prédire ces surcharges avec assez d'avance pour les éviter. La méthode sera un algorithme classique de détection de saut de moyenne (test de Page-Hinkley) ; celui-ci sera expérimenté sur des données réelles. Une extension possible consiste à faire évoluer le test vers un comportement autonome. Ce TER se situe dans le contexte du projet Grid Observatory¹ et de la grille européenne EGI².

Travaux prévus

L'étude portera sur deux situations :

- Accès en écriture à des systèmes de stockage : le but est d'anticiper les dépassements de capacité.
- Accès à un système d'information : le but est d'anticiper la surcharge du serveur.

Dans les deux cas, des traces (données) sont disponibles. Elles contiennent les retours de sondes logicielles, et une caractérisation des situations fautives. Dans le premier cas, un échantillonnage à fréquence fixe et élevée est soutenable. Dans le second cas, un schéma de type "exponential backoff" à la TCP devra probablement être envisagé pour obtenir des résultats précis sans créer d'effet d'entraînement supplémentaire.

Le test de Page-Hinkley inclut plusieurs paramètres, qui sont définis par essai et erreur lorsque la distribution des données n'est pas assez régulière. L'extension vers un système autonome consiste à définir automatiquement ces paramètres, à partir des données, en fonction de la précision et de la sensibilité souhaitées par l'utilisateur final³.

Résultats attendus

- Conception, réalisation et documentation d'un environnement d'expériences
 - Code du test de Page-Hinley, en C++ ou en Matlab, suivant les compétences de l'étudiant.
 - Interface permettant la réutilisation, ainsi que la sélection manuelle ou automatique des paramètres.
 - Base de données de résultats associée.
- Expérimentation sur les données d'accès en écriture, présentation et interprétation des résultats.
- Expérimentations sur l'accès à un système d'information ; éventuellement amélioration du logiciel d'acquisition des données.
- Etude bibliographique sur l'autonomisation (méthodes de bandits, apprentissage par renforcement, autres stratégies d'optimisation).

Suivant les intérêts de l'étudiant, deux prolongements du TER sont possibles : intégration de l'environnement dans un outil de monitoring d'EGI, ou approfondissement de la stratégie d'autonomisation.

Compétences requises

Bon niveau en programmation (C++ ou Matlab, et Perl ou Python), éléments de génie logiciel et de bases de données.

Organisation

Contact : Cécile Germain-Renaud Laboratoire de Recherche en Informatique, cecile.germain@lri.fr

Lieu de travail : un bon déroulement du TER suppose que l'étudiant puisse être présent au Laboratoire de Recherche en Informatique une demi journée par semaine.

¹ www.grid-observatory.org

² www.egi.eu

³ Xiangliang Zhang, Cécile Germain-Renaud and Michèle Sebag, "Adaptively Detecting Changes in Autonomic Grid Computing", Workshop on Autonomic Computational Science (ACS) in conjunction with 11th ACM/IEEE International Conference on Grid Computing (Grid'2010), Brussels, Belgium, October 25-29, 2010.

Prolongements

Le TER peut éventuellement se prolonger par un stage d'été.