Mercredi 20 Juin – Amphi H

11h30 Repas

13h30 Étienne Rivière (IRISA)

RAPPEL : *dissémination collaborative et auto-organisante de flux RSS*

L'utilisation de mécanismes de syndication Web, qui fonctionne selon le principe de publication et abonnement fondé sur le sujet, ont connu un essor important et sont au centre des nouveaux usages et technologies du Web. Des exemples de tels mécanismes sont RSS & Atom.

Ces syndication à des flux permet à un utilisateur d'être notifié de l'apparition de nouvelles entrées dans une source d'information : blog, site de journal, etc. La mise en œuvre courante consiste à utiliser un agrégateur, logiciel qui interroge périodiquement les serveurs source pour détecter les mises à jour. L'utilisateur reçoit les nouvelles informations de la même manière que pour un compte Mail ou Usenet. On observe un changement des usages de ces flux : alors qu'ils étaient au départ surtout utiliser pour suivre une source populaire de contenu (journal en ligne, site d'informations), désormais les utilisateurs adoptent un comportement actif dans la production de contenu. La « blogosphère » est un exemple typique de ces usages, où les utilisateurs se comportent à la fois comme consommateur et producteur d'informations.

Néanmoins, la mise en œuvre actuelle de la syndication engendre une utilisation élevée des ressources du serveur publiant le flux. Les clients d'un flux interrogent régulièrement le serveur pour détecter les mises à jour, générant un traffic important et rarement nécessaire si la fréquence de mise à jour est plus faible que la fréquence d'interrogation. La réponse habituelle est de limiter par des listes noires la fréquence d'interrogations en provenance d'une IP donnée. Cette solution présente le désavantage d'induire des délais de réception important pour la notification des consommateurs du flux.

Je présente la proposition et la mise en œuvre de Rappel, un réseau pair-à-pair collaboratif pour la diffusion de flux RSS. Les objectifs de Rappel sont : (1) les mises à jour des flux sont obtenues en utilisant le minimum de bande passante du nœud source, utilisant plusieurs ordre de grandeur en moins de cette ressource ; (2) les mises à jour sont rapidement (en quelques secondes) délivrées à l'ensemble des nœuds interessés ; (3) le protocole est adaptable à la topologie du réseau physique utilisé en utilisant le principe de coordonnées réseau comme une métrique principale pour la création des structures de dissémination ; (4) la proximité d'intérêt entre les différents abonnés est utilisée pour minimiser le nombre de voisins nécessaires pour assurer l'efficience de réseau logique et enfin, (5) à la différence de propositions concurrentes, Rappel ne délivre jamais ni ne fait transiter de mise à jour non désirée sur un nœud du réseau.

Rappel est mis en œuvre au sein d'un simulateur à évènements discrets utilisant un modèle de topologie transit-stub, et des mesures de délais effectués sur PlanetLab. De plus, les abonnements des usagers du système sont fondés sur des traces en provenance de la plateforme de blog populaire LiveJournal.com : nous disposons de plus de 300.000 utilisateurs et de plus de 1 million de mises à jour pour cela.

14h00 Miroslaw Korzeniowski (LaBRI)

Dating service : a tool to cope with heterogeneity in distributed systems

We propose a simple scheme to organize resource allocation in a distributed system in which nodes submit various demands and supplies for some resource. In each round our system connects demand-supply pairs (dates) in a distributed and random fashion and we prove that even distributed system (for example based on a Distributed Hash Table) produces the number of dates within a constant factor from a centralized one. We apply our scheme to two basic problems. In both nodes submit their incoming and outgoing bandwidths and we use our dating service to organize communication. In the first case we show how to do rumor spreading in a heterogenous network and in the second we show how to construct a random multigraph in polylogarithmic time.

14h30 Lionel Eyraud-Dubois (LIP)

ALNeM : Reconstruction de topologies d'un point de vue applicatif

La plupart des méthodes d'ordonnancement nécessitent une connaissance précise de la plate-forme cible, qui dans le cas de grilles n'est que rarement disponible. Nous nous intéressons donc à la reconstruction de la topologie à partir de mesures applicatives, ne nécessitant pas de privilèges particulier. Nous proposons une méthodologie pour étudier et valider des algorithmes qui effectuent cette reconstruction. Nous introduisons également des algorithmes originaux, qui permettent d'obtenir des reconstructions de bonne qualité.

15h00 William Hoarau (LRI)

Versatile Fault-injection with FAIL-FCI

A long standing trend in high performance distributed systems is the increase of the number of nodes. As a consequence, the probability of failures in supercomputers and distributed systems also increases. So fault tolerance becomes a key property of parallel applications. Designing and implementing fault-tolerant software is a complex task. Fault-tolerance is a strong property which implies a theoretical proof of the underlying protocols. The protocol implementation should then be checked with respect to the specification.

In order to validate this implementation, a rigorous testing approach can be used. Automatic failure injection is a general technique suitable for evaluating the effectiveness and robustness of distributed applications against various and complex failure scenarios.

After having validated the fault-tolerant implementation, it is necessary to evaluate and tune its performance, in order to determine the best protocol suitable for an actual distributed system. Fault-Tolerant distributed applications are classically evaluated without failures. However, performance under a failure-prone environment is also a important information for evaluating and tuning a fault-tolerant distributed system. Automatic failure injection is desirable to evaluate fairly different heuristics or parameters under the same failure conditions.

In this talk we explore the versatility of FAIL-FCI, our tool for fault injection in distributed applications. In particular, we show that not only we are able to fault-load existing distributed applications (as used in most current papers that address fault-tolerance issues), we are also able to inject qualitative faults, e.g. inject specific faults at very specific moments in the program code of the application under test. Finally, and although this was not the primary purpose of the tool, we are also able to inject specific patterns of workload, in order to stress test the application under test. Interestingly enough, the whole process is driven by a simple unified description language, that is totally independent from the language of the application, so that no code changes or recompilation are needed on the application side.

As a case study, we strain XtremWeb and the MPICH-Vcl non-blocking implementation of the Chandy-Lamport protocol. XtremWeb is a general purpose platform that can be used for high performance distributed computation. A list of tasks (or jobs) is described by the user and then distributed over the different available nodes of the system. The basic operating mode of XtremWeb is based on a participant community, e.g. it allows a High School, a University or a Company to setup and run a Global Computing or Peer to Peer distributed system for either a dedicated application or a whole range of applications. MPICH-Vcl is a high performance fault-tolerant MPI library which provides a generic framework to add transparently fault-tolerance to any MPI application.

15h30 Pause

16h00 François Bonnet (IRISA)

Petits mondes : Y a-t-il désaccord entre pratique et théorie ?

In small-world networks each peer is connected to its closest neighbors, in the network topology, as well as to additional long-range contact(s), also called shortcut(s). In 2000, Kleinberg showed that greedy routing in a n peer small world network, performs in $O(n^{\frac{1}{3}})$ steps when the distance to shortcuts is chosen uniformly at random, and in $O(\log^2 n)$ when the distance to shortcuts is chosen according to a harmonic distribution in a d-dimensional mesh. Yet, we observe through experimental results that peer to peer gossip-based protocols achieving small-world topologies where shortcuts are randomly chosen, perform well in practice.

The motivation of this paper is to explore this mismatch and attempts to reconcile theory and practice in the context of small-world overlay networks. More precisely, based on the observation that, despite the fact that the routing complexity of gossip-based small-world overlay networks is not polylogarithmic (as proved by Kleinberg), this type of networks ultimately provide reasonable results in practice. This leads us to think that the asymptotic big $O()$ complexity alone might not always be sufficient to assess the practicality of a system. The paper consequently proposes a refined routing complexity measure for small-world networks. Simulation results confirm that random selection of shortcuts can achieve "practical" systems. Then, given that Kleinberg proved that the distribution of shortcuts has a strong impact on the routing complexity, arises the question of leveraging this result to improve upon current gossip-based protocols. We show that it is possible to design gossip-based protocols providing a good approximation of Kleinberg-like small-world topologies. Along, are presented simulation results that demonstrate the relevance of the proposed approach.

16h30 Philippe Duchon (LaBRI)

Résultats de non-navigabilité dans certains modèles de graphes

sans échelle

17h00 Christophe Dürr (LIX)

Équilibres de Nash pour des jeux de Voronoï sur des graphes

Les jeux de Voronoï ont été introduits pour étudier la séparation d'un espace par des agents visant à maximiser la surface attribuée. Nous étudions la version discrète qui se joue sur un graphe où la longueur du plus court chemin définit une métrique. Ceci peut être vu comme le jeu associé au problème de facility location. Il y a des graphes où le jeu converge vers un équilibre et des graphes où il n'y a pas de convergence. Nous montrons que distinguer ces deux cas est NP-complet. Pour finir nous étudions la différence du coût social entre les différents équilibres.

Jeudi 21 Juin – Amphi H

8h30 Café

9h00 Anne Benoît (LIP)

Complexity results for throughput and latency optimization of replicated and data-parallel workflows

Mapping applications onto parallel platforms is a challenging problem, even for simple application patterns such as pipeline or fork graphs. Several antagonist criteria should be optimized for workflow applications, such as throughput and latency (or a combination). In this talk, we consider a simplified model with no communication cost, and we provide an exhaustive list of complexity results for different problem instances. Pipeline or fork stages can be replicated in order to increase the throughput of the workflow, by sending consecutive data sets onto different processors. In some cases, stages can also be data-parallelized, i.e. the computation of one single data set is shared between several processors. This leads to a decrease of the latency and an increase of the throughput. Some instances of this simple model are shown to be NP-hard, thereby exposing the inherent complexity of the mapping problem. We provide polynomial algorithms for other problem instances. Altogether, we provide solid theoretical foundations for the study of mono-criterion or bi-criteria mapping optimization problems.

9h30 Ralf Kalsing (LaBRI)

Searching for black-hole faults in a network using multiple agents

We consider a fixed communication network where (software) agents can move freely from node to node along the edges. A "black hole" is a faulty or malicious node in the network such that if an agent enters this node, then it immediately "dies". We are interested in designing an efficient communication algorithm for the agents to identify all black holes. We assume that we have k agents starting from the same node s and knowing the topology of the whole network. The agents move through the network in synchronous steps and can communicate only when they meet in a node. At the end of the exploration of the network,

at least one agent must survive and must know the exact locations of the black holes. If the network has n nodes and b black holes, then any exploration algorithm needs $\Omega(n/k + D_b)$ steps in the worst-case, where D_b is the worst case diameter of the network with at most b nodes deleted. We give a general algorithm which completes exploration in $O((n/k) \log n / \log \log n + bD_b)$ steps for arbitrary networks, if b <= k/2. In the case when $b \leq k/2$, $bD_b = O(\sqrt{n})$ and $k = O(\sqrt{n})$, we give a refined algorithm which completes exploration in asymptotically optimal O(n/k) steps.

10h00 Pause

10h30 Matthieu Gallet (LIP)

Ordonnancement de tâches divisibles

Cet exposé aura pour objectif la présentation générale du modèle des tâches divisibles. Un premier exemple sur une plate-forme maître-esclave montrera l'intérêt de ce modèle par rapport à l'approche classique, plus difficile à résoudre. Nous verrons également les avantages et les inconvénients de ce modèle, et si on peut l'améliorer en découpant les envois en plusieurs tournées ou en introduisant des latences dans le modèle de communication. Enfin, nous verrons un exemple pratique sur un réseau linéaire de processeurs.

11h00 Raphaël Bolze (LIP)

Multi-workflow scheduling

Au cours de cet exposé, je commencerai par poser le problème auquel nous nous intéressons : l'ordonnancement d'un ensemble de graphes de tâches dans un environnement grille. La motivation de ce travail sera illustrée par quatre applications. Je présenterai un bref état de l'art du sujet et vous détaillerai un ensemble d'heuristiques qui découlent de l'étude de ce problème. Dans ce cadre je vous présenterai un outil que j'ai développé permettant de simuler les heuristiques précédemment évoquées. Enfin, je présenterai l'intégration de ces heuristiques dans l'intergiciel DIET.

11h30 Repas de midi