

# Sujet de thèse - Bourse école doctorale

## Simuler l'Internet des objets

### 1 Contexte

Les objets communicants envahissent progressivement notre quotidien avec des champs d'application toujours plus étendus : équipements personnels de santé, bâtiments intelligents, smart grids, véhicules connectés, etc. Une étude récente estime en effet qu'il y aura 50 milliards d'objets connectés en 2020 avec des retombées économiques conséquentes dans les secteurs de la santé, de l'énergie, de l'automobile et du bâtiment.

Tous ces objets, reliés à des réseaux de télécommunication (Internet le plus généralement), peuvent interagir avec d'autres objets connectés ou avec des infrastructures de calcul distribuées, telles que des Clouds par exemple, pour stocker des informations ou effectuer des calculs. La croissance du nombre d'objets connectés et d'infrastructures de support pose des défis scientifiques concernant notamment la gestion du passage à l'échelle, de l'hétérogénéité des réseaux de communications employés (Ethernet, WiFi, 3G, etc.), de la migration des calculs entre les objets et les infrastructures de support, et de leur consommation électrique inquiétante.

Ces objets communicants et les infrastructures de support exécutent des applications distribuées telles que des systèmes communicants, des jeux, des calculs scientifiques ou des applications web. Par nature, ces environnements matériels et logiciels sont extrêmement complexes et dynamiques, car ils agrègent des multitudes d'éléments hétérogènes, dynamiques et potentiellement mobiles. Cette complexité rend ces systèmes très difficiles à observer, tester et évaluer. Les travaux purement théoriques sont limités par des suppositions au réalisme discutable, ce qui explique que la plupart des études utilisent des résultats expérimentaux, le plus souvent obtenus sur des infrastructures de test dédiées, à petite échelle. Mais la taille et la complexité que ces systèmes atteignent de nos jours compliquent grandement les tests, et rendent les expériences difficiles à reproduire de manière scientifique et à grande échelle. La simulation constitue alors une alternative intéressante pour pouvoir étudier les problèmes auxquels se confrontent ces systèmes : passage à l'échelle, gestion de l'hétérogénéité ou optimisation énergétique globale par exemple.

SimGrid (développé par un consortium national piloté par M. Quinson) est un outil open-source permettant de simuler des applications distribuées dans des environnements à large échelle. Au fil des années, SimGrid a émergé comme l'un des instruments scientifiques majeurs de sa communauté. Il a été utilisé pour évaluer les expériences d'une dizaine de thèses de doctorat et plus de 150 articles scientifiques. Ses principaux avantages sont : des modèles de performances fiables (capables de prédictions satisfaisantes même dans des cas non-triviaux), son efficacité (temps de simulation) et son extensibilité (modularité logicielle), ainsi que sa capacité à étudier formellement la correction d'algorithmes distribués par exploration exhaustive des exécutions possibles.

### 2 Problématique

L'objectif de cette thèse est d'enrichir le simulateur open-source SimGrid pour pouvoir étudier à grande échelle l'Internet des objets, en tant qu'infrastructure distribuée, dynamique et fortement hétérogène, ainsi que les applications qui se déploient sur ces objets. Les trois problèmes principaux que nous avons identifiés pour arriver à ce résultats sont :

1. l'intégration dans SimGrid de modèles de réseau de communication hétérogènes (filaire et sans fil) pour simuler les moyens de communication employés par les objets connectés. Actuellement, le

modèle disponible correspond à un réseau Ethernet et SimGrid ne permet pas d'utiliser plusieurs modèles réseau simultanément dans une même simulation.

2. l'intégration dans SimGrid de modèles de Clouds distribués pour simuler les infrastructures de support des objets connectés. Il est actuellement possible de simuler des machines virtuelles dans SimGrid, mais les modèles de migration de machine virtuelle et ceux de stockage restent très limités.
3. la validation du simulateur comprenant tous les modèles proposés sur des cas d'études réalistes. Ceci nécessitera d'effectuer des comparaisons simulation vs. vraie plate-forme pour évaluer la précision du simulateur. On s'attachera également à obtenir des temps de simulation raisonnables pour offrir un vrai outil de simulation de ce type de plates-formes qui passe à l'échelle. L'implémentation et la complexité des modèles proposés joueront un rôle primordial pour cela.

### 3 Approche méthodologique

Cette thèse appliquera aux systèmes distribués modernes des outils méthodologiques emprunté au domaine de la modélisation des systèmes complexes, en visant à la fois des modèles prédictifs des performances de ces systèmes et une architecture logicielle adaptée. Ce travail regroupe donc une forte composante de génie logiciel, mais sans bien sûr se limiter à cette dimension. L'objectif est d'établir une preuve de concept suffisamment solide pour permettre des mesures objectives et reproductibles sur les qualités de l'approche. Ces travaux seront utilisables en pratique après consolidation par des ingénieurs de recherche financés actuellement sur le projet SimGrid. Le plan de travail envisagé est le suivant : La première année sera dévolue à l'étude bibliographique des modèles existants de réseau sans fil ainsi que de migration de ressources virtuelles. Afin de mieux cerner leurs limitations dans notre contexte, l'étudiant.e sélectionnera plusieurs scénarios d'usage et mettra en place un cadre de travail technique permettant de comparer les modèles entre eux. Evaluer différentes méthodes d'étude pour des applications distribuées peut être difficile en soi, et l'étudiant.e devra renforcer ses pratiques expérimentales pour rendre nos travaux pleinement reproductibles par d'autres. Une campagne de tests évaluant la puissance prédictive des modèles de la littérature grâce SimGrid permettra de valider le cadre de travail établi. Pendant la deuxième année, l'étudiant.e proposera de nouveaux modèles allant au-delà de l'état de l'art. Des interfaces de programmation très similaires à celles utilisées pour écrire les applications réelles seront également implémentées au dessus du noyau de SimGrid. Durant la troisième année, l'étudiant.e évaluera l'applicabilité de ses propositions sur des cas d'études aussi réalistes que possible, et écrira le manuscrit de sa thèse.

### 4 Profil du candidat

En plus des compétences méthodologiques que l'on attend d'un étudiant en début de thèse, le/la candidat.e devra avoir un profil orienté réseau et/ou une expérience dans le domaine des systèmes distribués. Un niveau raisonnable de programmation en C/C++ est demandé, tandis qu'une expérience en simulation et modélisation de systèmes complexes sera un atout apprécié, sans être indispensable.

### 5 Positionnement de l'équipe d'accueil

Cette thèse sera effectuée à l'IRISA dans l'équipe-projet Myriads (<http://www.irisa.fr/myriads>) qui s'intéresse à la conception et mise en œuvre de systèmes distribués. Elle sera encadrée par :

- Martin Quinson, Professeur ENS Rennes, [martin.quinson@ens-rennes.fr](mailto:martin.quinson@ens-rennes.fr)
- Anne-Cécile Orgerie, Chargée de recherche CNRS, [anne-cecile.orgerie@irisa.fr](mailto:anne-cecile.orgerie@irisa.fr)

Martin Quison (<http://people.irisa.fr/Martin.Quinson>) est professeur à l'École Normale Supérieure de Rennes depuis septembre 2015. Entre 2005 et 2015, il était maître de conférences à l'université de Lorraine, où il a défendu son habilitation à diriger les recherches intitulée "*Computational Science of Computer Systems*" en 2013. Ses recherches portent sur l'étude de systèmes répartis à large échelle, tant par simulation pour leur performance que par vérification formelle pour leur correction. Il co-développe depuis 2003 l'environnement SimGrid, dans lequel la plupart de ses recherches

sont mises en pratique au bénéfice de la communauté de recherche et développement des systèmes distribués. Il est l'auteur d'un chapitre de livre, 6 articles de journaux, 28 articles de conférences et colloques avec comités de lecture, ainsi que 22 tutoriaux principalement sur l'étude par simulation de systèmes distribués à très large échelle.

Anne-Cécile Orgerie (<http://people.irisa.fr/Anne-Cecile.Orgerie>) est chargée de recherche au CNRS dans l'équipe-projet Inria Myriads à l'IRISA depuis octobre 2012. Elle a obtenu son doctorat en informatique de l'École Normale Supérieure de Lyon en septembre 2011. Sa thèse intitulée "*An Energy-Efficient Reservation Framework for Large-Scale Distributed Systems*" a reçu le prix de thèse de la recherche en système de l'ACM SIGOPS France (pour les thèses soutenues entre 2008 et 2012) et s'inscrit parfaitement dans les thématiques de recherche qui sont développées dans cette proposition de thèse. Elle a notamment été fortement impliquée dans la conception et la mise en œuvre du système de collecte de mesures énergétiques du site lyonnais de Grid'5000. D'octobre 2011 à septembre 2012, elle a travaillé en tant que post-doctorante à l'Université de Melbourne (Australie) sur le projet Franco-Japonais PetaFlow où elle a acquis des compétences en instrumentation réseau et synchronisation d'horloge. Ses intérêts de recherche se concentrent sur l'efficacité énergétique, le Cloud computing et les systèmes distribués tant au point de vue théorique qu'appliqué. Depuis décembre 2008, elle est l'auteur de 10 articles de journaux internationaux avec comité de lecture, 5 chapitres de livres et 34 articles de conférences avec comité de lecture essentiellement sur des problématiques d'économie d'énergie dans les systèmes distribués.

## 6 Références

- [BDG<sup>+</sup>13] P. Bedaride, A. Degomme, S. Genaud, A. Legrand, G. Markomanolis, M. Quinson, M. L. Stillwell, F. Suter, and B. Videau, "Toward Better Simulation of MPI Applications on Ethernet/TCP Networks," in *PMBS13 - 4th International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems*, Denver, United States, Nov. 2013.
- [CGL<sup>+</sup>14] H. Casanova, A. Giersch, A. Legrand, M. Quinson, and F. Suter, "Versatile, Scalable, and Accurate Simulation of Distributed Applications and Platforms," *Journal of Parallel and Distributed Computing*, vol. 74, no. 10, pp. 2899–2917, Jun. 2014.
- [GNQ13] M. Geier, L. Nussbaum, and M. Quinson, "On the Convergence of Experimental Methodologies for Distributed Systems : Where do we stand ?" in *WATERS - 4th International Workshop on Analysis Tools and Methodologies for Embedded and Real-time Systems*, Paris, France, Jul. 2013.
- [OLGLLP11] A.-C. Orgerie, L. Lefèvre, I. Guérin-Lassous, and D. Lopez Pacheco, "ECOFEN : an End-to-end energy Cost mOdel and simulator For Evaluating power consumption in large-scale Networks," in *SustaInet : International Workshop on Sustainable Internet and Internet for Sustainability, in conjunction with WoWMoM (IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks)*, Jun. 2011.
- [QRT12] M. Quinson, C. Rosa, and C. Thiery, "Parallel Simulation of Peer-to-Peer Systems," in *CCGrid 2012 – The 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing*, ser. CCGRID '12 Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing. Ottawa, Canada : IEEE, May 2012, pp. 668–675.
- [ROQ16] I. Rais, A.-C. Orgerie, and M. Quinson, "Impact of Shutdown Techniques for Energy-Efficient Cloud Data Centers," in *International Conference on Algorithms and Architectures for Parallel Processing (ICA3PP)*, Granada, Spain, Dec. 2016, pp. 203–210.