

# Rapport sur le manuscrit de thèse de doctorat

par Monsieur **Loic Guegan**

**pour l'obtention de la thèse de doctorat  
de l'École Normale Supérieure de Rennes**

Rapporteur: Stéphane GENAUD,  
Professeur des Universités,  
Université de Strasbourg.

Titre du document évalué :

## ***Scalable end-to-end models for the time and energy performance of Fog infrastructures***

Le document *Scalable end-to-end models for the time and energy performance of Fog infrastructures* présente le travail de thèse de Loic Guegan, mené à l'Ecole Normale Supérieure de Rennes sous la direction de Anne-Cécile Orgerie et Martin Quinson. L'objectif global de cette recherche est de proposer des modèles permettant la simulation du comportement des applications sur les infrastructures distribuées appelées communément *Fog*. Ce terme désigne les infrastructures constituées à la fois des ressources présentes dans les *data-centers* et celles qui appartiennent à l'*Internet of Things* (IoT). Ce type d'infrastructure, reposant sur des communications filaires et non-filaires, fait apparaître des défis nouveaux et considérables quand on cherche à prédire la performance et la consommation énergétique engendrées par son utilisation. Des travaux de recherche ont déjà produit des résultats reconnus pour la simulation du comportement de telles infrastructures, mais ils s'appuient sur des méthodes de simulation qui modélisent à un niveau très fin le comportement du réseau (on parle de *Packet Level Simulation* (PLS)) et pour cette raison nécessitent des temps extrêmement longs de simulation, incompatibles avec la taille des infrastructures que l'on souhaite étudier. La thèse de Loic Guegan est qu'il est possible de construire des modèles pour une simulation faisant abstraction de certains détails (on parle de *Flow Level Simulation* (FLS)) produisant des prédictions de performance et de consommation énergétique assez proche de ceux prédits par une PLS avec des temps de restitution qui sont plus courts de un à deux ordres de grandeur, permettant ainsi l'étude de ces environnements.

## **1 Analyse du document**

Le document, rédigé en anglais, est organisé en quatre chapitres, précédés d'une introduction, qui établit la question centrale traitée dans ce travail et terminés par une conclusion.

### **1.1 Chapitre 2**

Le premier chapitre (chapitre 2) décrit le contexte technologique actuel concernant les infrastructures distribuées. C'est un chapitre très complet qui brosse un panorama très bien illustré à la fois du matériel utilisé et des applications qui utilisent aujourd'hui ce type d'infrastructure, ainsi que des enjeux en termes énergétiques de leurs développements récents. Ce chapitre introduit également le vocabulaire et les notions nécessaires pour la suite. Une grande partie du chapitre décrit l'état de

l'art concernant les modèles de simulation de ces plateformes. Une attention particulière est portée à la description du fonctionnement du WiFi et des phénomènes pathologiques que l'on peut rencontrer (utile pour comprendre le dernier chapitre). La dernière partie du chapitre est consacrée à l'état de l'art des travaux scientifiques qui visent à simuler le comportement des réseaux sous l'angle de la performance et de l'énergie, en distinguant les propositions à base de PLS et de FLS. Le chapitre conclut en soulignant l'importance d'être capable de mener des expérimentations scientifiques à l'échelle réelle de ces infrastructures, et que pour ce faire la proposition de la thèse est d'étendre le simulateur FLS éprouvé SimGrid.

À ce stade du document, je suis bien convaincu que la simulation de l'ensemble d'une telle plateforme, dont l'ordre de grandeur du nombre d'équipements connectés atteint vite  $10^5$  ou  $10^6$  ne peut être praticable qu'à un niveau plus abstrait. Les temps rapportés pour les simulations NS-3 tout au long du document montrent la limite de l'approche PLS. Par conséquent l'approche FLS pour simuler le réseau apparaît comme un choix pertinent.

## 1.2 Chapitre 3

Le chapitre 3 décrit la problématique consistant à modéliser la consommation énergétique de *bout-en-bout* d'une infrastructure impliquant des appareils de type IoT communiquant en WiFi et des serveurs de calcul accessibles par un réseau filaire. La démarche pour aboutir au modèle général de consommation est convaincante dans ce chapitre : un cas d'utilisation impliquant un nombre variable de capteurs envoyant des données dans une base de données gérée par une machine virtuelle sur un serveur permet d'introduire les différents paramètres influant sur la consommation, lesquels sont pré-éminents, et dans quelles situations. Partir d'un cas d'usage minimal me semble être une idée saine pour établir les paramètres de base du modèle. Je regrette néanmoins que la partie IoT ait dû être simulée par NS-3, alors que la consommation énergétique des serveurs a pu être mesurée réellement. Bien que j'imagine la difficulté à mettre en place les conditions expérimentales concernant le WiFi et à mesurer également la consommation des capteurs, je pense que ce micro-benchmark eût été encore plus satisfaisant.

## 1.3 Chapitre 4

Le chapitre 4 décrit une première contribution de cette thèse qui est un modèle de consommation énergétique pour les communications sur réseaux filaires, implanté dans SimGrid et évalué par comparaison avec NS-3. L'exposé est la suite logique du chapitre précédent et explique comment le modèle présenté peut être intégré dans la mécanique interne de SimGrid. Le reste du chapitre est consacré à une évaluation conséquente de cette proposition. Le cas test choisi me semble pertinent. Il consiste en une description d'une infrastructure de type data-center, où des messages extérieurs arrivent sur 8 serveurs frontaux qui les distribuent à des serveurs de calcul. Cette expérience montre que dans ce cas, le modèle proposé et implanté dans SimGrid ne diverge que de quelques pourcents avec la prédiction de consommation énergétique de NS-3. Cette expérience met également en évidence qu'une simplification du modèle (*Optimized Homogeneous*) donne des résultats non-significativement différents. Cette contribution est donc d'importance car elle apporte la prédiction de consommation à un simulateur FLS, permettant de travailler sur des cas d'infrastructure à taille réelle. J'aurais été curieux de voir une comparaison avec un autre FLS existant (comme CloudSim ou CDartSim+ qui peuvent probablement décrire et simuler le cas test choisi).

## 1.4 Chapitre 5

Le dernier chapitre présente la deuxième contribution qui est un modèle pour simuler les communications non-filaires WiFi. C'est une réalisation difficile car il faut intégrer le modèle de tranches de temps de communication dans le solveur interne de SimGrid. Cette tentative est ensuite évaluée de manière approfondie sur des micro-benchmarks dont les prédictions sont comparées à celles de NS-3 qui fait référence. De façon très méthodique, les différences sont analysées et il ressort que les différences croissantes observées au fur et à mesure que le nombre de stations augmentent proviennent essentiellement de la prise en compte des interférences dans NS-3, non-prises en compte dans le modèle proposé pour SimGrid. En négligeant les interférences, les prédictions ne diffèrent que de quelques pourcents. C'est donc un résultat encourageant, la limitation étant bien identifiée, et ouvrant de nouvelles perspectives pour tenter d'ajouter une modélisation des interférences. Sur ce point, j'aurais été curieux de savoir si des idées prospectives avaient été envisagées, mais le document n'en mentionne aucune.

## 2 Conclusion

Les travaux de thèse de M. Loic Guegan ouvrent des perspectives nouvelles pour la simulation de bout-en-bout des infrastructures à large-échelle, dites *Fog*, pour lesquelles les simulateurs à grain fin de type Packet Level atteignent très vite leurs limites en terme de scalabilité, et sont donc inutilisables dans la pratique.

Si des recherches comparables existent dans le domaine des simulations de type Flow Level, la démarche scientifique de validation des résultats produits qui est proposée dans ce travail la rend plus convaincante que bien d'autres. Il faut souligner l'énorme quantité de travail que nécessite de telles expérimentations pour valider la modélisation proposée pour des infrastructures sans cesse plus complexes. C'est le cas du travail présenté ici. Il faut également noter que ces expériences ont été menées avec un objectif de reproductibilité, les paramètres d'expérimentation étant publiés et toute la base logicielle de Simgrid disponible en open source.

C'est au final un travail qui apporte des contributions scientifiquement bien construites concernant les modèles de simulation de la consommation énergétique et de la performance. De plus ces contributions ont été intégrées à SimGrid et bénéficient donc immédiatement à la communauté des utilisateurs de ce logiciel.

Pour toutes les raisons évoquées ci-dessus, je suis favorable à la soutenance de cette thèse de doctorat de l'École Normale Supérieure de Rennes.

Stéphane GENAUD,  
fait à Strasbourg, le 14 décembre 2020.

