



**D R P I**

Direction Recherche  
Partenariats Innovation

## PROPOSITION DE SUJET POUR UN CONTRAT DOCTORAL

### Laboratoire

L3i

### École doctorale EUCLIDE

### Sujet de thèse

#### ***Intitulé scientifique***

Analyse de l'efficacité énergétique d'applications basées sur des micro-services

#### ***Intitulé vulgarisé*** (explicite pour un non spécialiste)

Analyse de l'efficacité énergétique d'applications basées sur des micro-services

### Direction de la thèse Identité du·de la·des directeur·trice·s (grade, HDR) et des éventuels co-encadrant·e·s

Jean-Loup Guillaume – PR (La Rochelle Université – L3i)  
Benjamin Duthil – MCF (EIGSI – L3i)

### Priorité scientifique :

*Préciser en quoi le sujet s'inscrit dans la politique scientifique de l'établissement (création de l'Institut LUDI)*

L'objectif de cette thèse est d'offrir des outils d'analyse d'applications basées sur l'utilisation de micro-services. L'analyse sera centrée sur les communications entre micro-services et permettra d'identifier les communications coûteuses d'un point de vue énergétique, et donc à éviter, et celles qui ne posent pas de problème particulier.

La thèse s'inscrit donc dans le thème plus général du numérique responsable et se trouve donc naturellement au cœur des priorités scientifiques du LUDI.

### Descriptif du sujet

*Éléments d'explication du sujet (enjeux scientifiques, applicatifs, sociétaux...)*

En 1637, Descartes publiait anonymement le « Discours de la méthode ». Dans ce texte fondateur de la pensée occidentale, il présentait une méthode générale de résolution de problèmes. Selon lui, il faut le décomposer en éléments moins complexes, jusqu'à parvenir à ses éléments les plus simples. Mais cette approche prônant la simplification ne correspond plus désormais à la réalité d'un mode complexe.

Il y a quelques années, la chaîne de transformation de la crevette [1], depuis son lieu de pêche jusqu'à son lieu de consommation avait dévoilé un système qui séparait le processus en un ensemble de petites fonctionnalités (pêche, tri, décortiquage, emballage, vente, consommation) globalement indépendantes les unes des autres. Dès lors, si l'on ne considère pas l'impact environnemental comme un critère, il est facile d'optimiser le coût global en minimisant le coût de chacune des opérations, qu'il s'agisse de transport ou de traitement, sans jamais se demander s'il est pertinent de faire faire plusieurs milliers de kilomètres à une crevette dans un processus prenant près de 2 semaines.

De manière analogue, le développement d'applications à base de micro-services a pour objectif de concevoir une application comme un ensemble de services les plus indépendants

possible, qui communiquent entre eux par le biais d'interfaces de programmation (API). Si ces micro-services ont considérablement facilité le développement logiciel en permettant une flexibilité et une mutualisation des ressources sans précédent, ce paradigme a cependant eu un effet néfaste sur l'encombrement des réseaux et sur la consommation d'énergie associée au transport de l'information entre les micro-services. A l'heure actuelle, les critères de performance, outre sur la qualité des développements logiciels, sont essentiellement basés sur les temps de réponse des services.

L'impact écologique lié à l'utilisation massive des technologies Cloud n'est aujourd'hui plus à démontrer et notamment par rapport à la consommation nécessaire à leur fonctionnement [8,9] et les enjeux associés sont sérieusement pris en compte. Cependant, bien que l'optimisation des ressources CPU, RAM et Stockage soient relativement bien considérées, les aspects liés au transport de l'information le sont beaucoup moins alors qu'ils représentent aujourd'hui 10 à 20 % du problème énergétique du numérique, notamment pour les échanges entre services inter et intra cloud [5,6]. Pour limiter ce coût, il est nécessaire de répartir intelligemment la charge entre les différents datacenters en considérant à la fois l'endroit où se situe l'information, la disponibilité des datacenters, l'impact énergétique du datacenter et le coût énergétique du transport vers celui-ci afin de prendre les meilleures décisions. Il s'agit donc d'un problème complexe d'optimisation multicritères.

La question se pose désormais de savoir si le protocole de transport est adapté ou encore si la quantité de données nécessaires au transport n'est pas démesurée par rapport à la donnée "utile" (encapsulation). On peut également remettre en question la nécessité même des échanges entre micro-services, voire la fusion de certaines briques logicielles qui communiquent uniquement entre elles, d'autant plus lorsque ces micro-services communiquent entre eux alors qu'ils sont déployés dans des datacenters différents. Tout comme la crevette qui parcourrait des milliers de kilomètres pour revenir à son point de départ, combien de micro-services impliquent du transport d'informations d'un bout à l'autre de la planète ? Si seuls la simplicité de développement ou le temps de réponse sont des critères, cela ne pose pas de problème particulier mais d'un point de vue numérique responsable, on ne peut se limiter à ces seuls critères.

Serait-il temps de passer de la simplicité à la simplexité ? [7]

L'objectif de cette thèse n'est pas de proposer de nouveaux protocoles et nous ne cherchons pas ici à remettre en question l'utilité et encore moins le fonctionnement du protocole TCP-IP, considéré aujourd'hui comme un standard. Il s'agit, au contraire, de s'attaquer au format utilisé pour l'échange des données afin d'éviter les consommations énergétiques liées notamment aux transformations de l'information (parsing) ainsi qu'à la localisation des services. Ainsi, même si on ne décide pas de la manière dont les données sont transportées, on peut contrôler tout le reste : qu'est-ce qui est envoyé, depuis où et vers où, à quelle fréquence, etc. L'objectif est donc plutôt d'établir un comparatif énergétique entre les différents standards (SOAP, REST...) en analysant les flux de données et, à terme, d'être capable d'automatiser l'analyse d'un ensemble de micro-services d'un point de vue énergétique.

Le problème formel revient à considérer un ensemble de datacenters hébergeant des micro-services qui communiquent les uns avec les autres et à déterminer la route la plus pertinente d'un point de vue environnemental (consommation énergétique), la pertinence d'une route (son impact) étant déterminée par les critères de distance, de protocole, de quantité de données pour un niveau d'acceptabilité. Il s'agit alors de trouver un compromis entre rapidité du transport et performance énergétique des datacenters cibles.

Une modélisation naturelle du problème consiste à considérer un graphe de relations entre datacenters (avec éventuellement une granularité plus fine) avec des pondérations sur les arêtes (le coût du transport) et un graphe de communications entre micro-services dans le cadre d'une application. Les micro-services étant hébergés dans des datacenters, on a donc affaire à un graphe multi-couches [4] avec un couplage entre les services et les datacenters qui permet d'analyser le problème d'un point de vue communications entre micro-services, communications entre datacenter ou couplée entre les deux niveaux. Via cette modélisation,

un certain nombre de problèmes peuvent se modéliser et se résoudre avec de l'algorithmique de graphes, notamment pour :

- le placement de micro-services dans des datacenters en fonction de la disponibilité et des coûts de transport ;
- l'analyse des communications entre micro-services et les coûts associés ;
- l'analyse des trajectoires des données entre micro-services ou entre datacenters.

Plusieurs thèses récentes ou en cours au sein du L3i tournent autour de l'analyse de trajectoires, qu'il s'agisse de trajectoires au sein d'une ville ou de trajectoires sur des graphes. La présente thèse pourra donc offrir un terrain d'application aux méthodes développées dans les autres thèses : clustering de trajectoires, détection de communautés dans les graphes multi-couches [3, 4], etc.

## Références

[1] [https://www.youtube.com/watch?v=f\\_KGyt4XBYw](https://www.youtube.com/watch?v=f_KGyt4XBYw)

[2] A. K. Bhowmick, K. Meneni, M. Danisch, J.-L. Guillaume, B. Mitra. LouvainNE: Hierarchical Louvain Method for High Quality and Scalable Network Embedding. WSDM2020.

[3] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre. Fast unfolding of communities in large networks. J. Stat. Mech. P10008, 2008.

[4] S. Pramanik, R. Tackx, A. Navelkar, J.-L. Guillaume, B. Mitra. Discovering Community Structure in Multilayer Networks. ACM/IEEE International Conference on Data Science and Advanced Analytics (DSAA 2017).

[5] M. Dayarathna, Y. Wen and R. Fan, "Data Center Energy Consumption Modeling: A Survey," in IEEE Communications Surveys & Tutorials, vol. 18, no. 1, pp. 732-794, Firstquarter 2016, doi: 10.1109/COMST.2015.2481183.

[6] Huigui Rong, Haomin Zhang, Sheng Xiao, Canbing Li, Chunhua Hu, Optimizing energy consumption for data centers, Renewable and Sustainable Energy Reviews, Volume 58, 2016, Pages 674-691, ISSN 1364-0321, <https://doi.org/10.1016/j.rser.2015.12.283>.

[7] [Alain Berthoz](#), *La simplicité*, Ed. Odile Jacob, 2009

[8] <https://www.clarke-energy.com/applications/data-centre-chp-trigeneration/>

[9] <https://www.energystream-wavestone.com/2017/02/data-center-grands-pollueurs-numerique/>

## Contexte partenarial

*S'il est prévu que la thèse se déroule en partenariat (cotutelle internationale, EU-Conexus, partenariat avec un autre laboratoire, une entreprise...) merci de le préciser ici et d'en décrire les principaux éléments*

-

## « Plus value » Etablissement :

*Préciser l'ambition du sujet et ses retombées : impacts mesurables sur la dynamique du/des chercheurs, du laboratoire, de l'établissement*

L'objectif de cette thèse est de produire des outils de modélisation et d'analyse de logiciels basés sur les micro-services. La thèse entre parfaitement dans le cadre général du numérique responsable et émerge donc de fait sur deux des quatre axes du LUDI avec des aspects liés d'une part, à la durabilité et, d'autre part, au numérique.

Les méthodes et algorithmes feront également l'objet de publications, les outils d'évaluation seront mis à disposition, de préférence en accès ouvert.

## Travail demandé au doctorant

*Préciser les tâches qui seront confiées au doctorant (programme de travail)*

Trois axes de développement seront étudiés. Le premier consistera à récupérer des traces d'exécution et à les modéliser sous forme de graphes multi-couches (plusieurs traces sont déjà disponibles). Le second consistera à proposer des mesures pour analyser le coût global

d'une exécution et le troisième à proposer, de manière dynamique, du placement de micro-services sur des datacenters pour optimiser le coût des échanges.

L'implémentation efficace de tous les algorithmes et métriques ainsi que leur mise à disposition feront l'objet d'une attention particulière.

### **Accompagnement du doctorant / Fonctionnement de la thèse**

*Qu'est-ce qui est prévu pour accompagner le doctorant (accompagnement humain, matériel, financier), en particulier pour la prise en charge du fonctionnement de la thèse et des dépenses associées*

Accompagnement standard du L3i et de l'école doctorale : locaux + matériel + 2k€ minimum pour conférences (rang B ou mieux).

Les travaux en lien avec l'analyse de trajectoires sont en lien avec le projet ANR MITIK (<https://project.inria.fr/mitik/>). Les travaux plus fondamentaux sur l'analyse de réseaux multi-couches pourront entrer dans le cadre du projet ANR JCJC Limass (porté par Sorbonne Université - <https://sites.google.com/view/limass>). Enfin, des liens avec le projet CHIST-ERA SCORING qui étudie également des problématiques de placement de micro-services (sans prise en compte spécifique de l'aspect énergétique cependant) seront également possibles. Le doctorant, s'il effectue des travaux qui rentrent clairement dans le cadre de ces projets, pourra également émarger dessus.

### **Compétences du candidat recherché**

**Master en informatique** ou diplôme conférant le grade de Master ou diplôme étranger jugé équivalent par le conseil de l'école doctorale.

Compétences en développement logiciel, notamment micro-services, et en théorie des graphes.

Intérêt marqué pour le numérique responsable.

### **Candidater**

Envoyer toutes les pièces demandées dans le dossier de candidature au plus tard le 31 mai 2021 :

- Lettre de candidature et de motivation (**précisant notamment les objectifs professionnels à l'issue de la thèse**)
- Un curriculum vitae détaillé avec liste des publications (si existantes)
- Descriptif de la **formation suivie** les 2 dernières années et **résultats obtenus**
- Un exemplaire des publications ou de tout mémoire réalisé par le candidat, même écrit dans une langue autre que le français
- Une ou deux lettres de recommandation (**dont une du directeur de mémoire**) envoyées directement par la personne qui recommande (sous pli confidentiel ou par mail)

Les documents doivent être envoyés par courrier postal ou par e-mail :

[jean-loup.guillaume@univ-lr.fr](mailto:jean-loup.guillaume@univ-lr.fr) :

Jean-Loup Guillaume  
L3i - Université de la Rochelle  
Avenue Michel Crépeau  
17042 La Rochelle Cedex 1 - France

Suite à l'étude des dossiers, une audition sera effectuée courant juin pour les candidats retenus.