

Research internship (Master 2)

Contribution to the adaptation of urgent applications in the IoT-to-Cloud Continuum

Hélène Coullon
IMT Atlantique, Inria, France

Daniel Balouek-Thomert
SCI Institute, University of Utah, USA

2021

Keywords : IoT-to-Cloud continuum ; urgent computing ; dynamic adaptation ; machine learning.

1 Introduction

With the advent of distributed infrastructures the Cloud computing paradigm is progressively moving towards a full continuum from IoT devices and sensors to the centralized Cloud, with Edge (edge of the network) and Fog computing (core network) in between. Simultaneously, distributed applications also evolve. Urgent computing tackles services that requires time-critical decisions that improve quality of life, monitor civil infrastructures, respond to natural disasters and extreme events, and accelerate science (e.g., autonomous cars, disaster response, precision medicine, etc.). These services are typically sensitive to latency and response time and are among the best candidate for the IoT to Cloud computing continuum [1].

In this internship, we consider a new breed of urgent intelligent services using the IoT-to-Cloud Continuum, combined with the recent advances in Artificial Intelligence and Big Data Analytics. First, these services and applications require a large computing capacity to perform well, while often being under the constraints to move data from the edge of the network to the Cloud [4]. Second, these services and applications require system support to program reactions that occurs at runtime, especially when the target infrastructure capacities and capabilities is unknown during the design [8].

It is challenging to run such services with a guaranteed performance on the continuum. From the application perspective, different types of events associated to the data to be transformed drive its configuration (i.e., what should run ?) [3, 5]. From the infrastructure perspective, objectives associated to the urgency of the results or the resources usages impacts the placement of function across the IoT-to-Cloud Continuum (i.e., where should it run ?) [2, 9].

2 Expected work

We expect the successful candidate to contribute to a comprehensive architecture that enables the composition of services from different providers across a highly heterogeneous infrastructure with geographically distributed resources. This work specifically targets end-to-end applications that are implemented as a workflow of services spanning data producers located at the edges, and data consumers requesting time-critical decisions from anywhere in the network.

The objectives of this internship are :

- to model and integrate the variability of the computing continuum ;

- to model and integrate the variability of the services involved in "urgent" applications ;
- to leverage machine learning techniques to dynamically take advantage of this variability to adapt both the application features [6, 7] and services placement on the continuum [2].

In this modeling and integration tasks, two important aspects will have to be studied :

- make a separation of concerns between domain scientists (familiar with the applications) and devops (familiar with the platforms) in the modeling ;
- consider metrics relative to latency, quality of service, and throughput during the lifecycle of an "urgent" application.

We expect the successful candidate to create repeatable processes and artifacts that will be used at scale to develop and evolve edge computing designs. Experiments and validation will occur on Grid'5000, the biggest share network dedicated to research in Computer Science.

Note that if satisfactory, the successful candidate will probably have an opportunity to start a Ph.D. thesis after the internship.

3 Skills

The following skills are expected from the successful candidate :

- a student in the last year of a Master's degree in Computer Science (or in the last year of an engineering school with a computer science option) ;
- modeling skills to be able to abstract the properties of the computing continuum and the "urgent" applications ;
- knowledge of the Python programming language, and ideally an experience in using Machine Learning libraries (SciKit-Learn, TensorFlow, Keras etc.) ;
- ideally some background in formal methods to provide formal modeling ;
- a good level of English to contribute to the writing of a research paper ;
- an ability to collaborate and communicate ;
- curiosity and an appetite for learning new things.

4 Additional information

Advisors

- [Hélène Coullon](mailto:helene.coullon@imt-atlantique.fr), IMT Atlantique & Inria équipe Stack, helene.coullon@imt-atlantique.fr
- [Daniel Balouek-Thomert](mailto:daniel.balouek@utah.edu), SCI Institute, University of Utah, daniel.balouek@utah.edu

Duration 6 months

Salary legal amount of 3,90€ / hour, full time

Location IMT Atlantique, équipe Inria Stack, laboratoire LS2N à Nantes

Références

- [1] Daniel Balouek-Thomert, Ivan Rodero, and Manish Parashar. Harnessing the computing continuum for urgent science. *SIGMETRICS Perform. Eval. Rev.*, 48(2) :41–46, November 2020. ISSN 0163-5999. doi:10.1145/3439602.3439618. URL <https://doi.org/10.1145/3439602.3439618>.
- [2] Emile Cadorel, H el ene Coullon, and Jean-Marc Menaud. Online Multi-User Workflow Scheduling Algorithm for Fairness and Energy Optimization. In *CCGrid2020 : 20th International Symposium on Cluster, Cloud and Internet Computing*, Melbourne, Australia, November 2020. doi:10.1109/CCGrid49817.2020.00-36. URL <https://hal.archives-ouvertes.fr/hal-02551733>.
- [3] Maverick Chardet, H el ene Coullon, and Simon Robillard. Toward Safe and Efficient Re-configuration with Concerto. *Science of Computer Programming*, 203 :1–31, March 2021. doi:10.1016/j.scico.2020.102582. URL <https://hal.inria.fr/hal-03103714>.
- [4] Kevin Fauvel, Daniel Balouek-Thomert, Diego Melgar, Pedro Silva, Anthony Simonet, Gabriel Antoniu, Alexandru Costan, V eronique Masson, Manish Parashar, Ivan Rodero, and Alexandre Termier. A distributed multi-sensor machine learning approach to earthquake early warning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(01) :403–411, Apr. 2020. doi:10.1609/aaai.v34i01.5376. URL <https://ojs.aaai.org/index.php/AAAI/article/view/5376>.
- [5] Luc Lesoil, Mathieu Acher, Arnaud Blouin, and Jean-Marc J ez equel. Deep software variability : Towards handling cross-layer configuration. In Paul Gr unbacher, Christoph Seidl, Deepak Dhungana, and Helena Lovasz-Bukvova, editors, *VaMoS'21 : 15th International Working Conference on Variability Modelling of Software-Intensive Systems, Virtual Event / Krems, Austria, February 9-11, 2021*, pages 10 :1–10 :8. ACM, 2021. doi:10.1145/3442391.3442402. URL <https://doi.org/10.1145/3442391.3442402>.
- [6] Hugo Martin, Mathieu Acher, Juliana Alves Pereira, and Jean-Marc J ez equel. A comparison of performance specialization learning for configurable systems. In Mohammad Mousavi and Pierre-Yves Schobbens, editors, *SPLC '21 : 25th ACM International Systems and Software Product Line Conference, Leicester, United Kingdom, September 6-11, 2021, Volume A*, pages 46–57. ACM, 2021. doi:10.1145/3461001.3471155. URL <https://doi.org/10.1145/3461001.3471155>.
- [7] Juliana Alves Pereira, Hugo Martin, Paul Temple, and Mathieu Acher. Machine learning and configurable systems : a gentle introduction. In Roberto Erick Lopez-Herrejon, editor, *SPLC '20 : 24th ACM International Systems and Software Product Line Conference, Montreal, Quebec, Canada, October 19-23, 2020, Volume A*, page 40 :1. ACM, 2020. doi:10.1145/3382025.3414976. URL <https://doi.org/10.1145/3382025.3414976>.
- [8] Eduard Gibert Renart, Daniel Balouek-Thomert, and Manish Parashar. An edge-based framework for enabling data-driven pipelines for iot systems. In *2019 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*, pages 885–894, 2019. doi:10.1109/IPDPSW.2019.00146.
- [9] Eduard Gibert Renart, Alexandre Da Silva Veith, Daniel Balouek-Thomert, Marcos Dias De Assun c ao, Laurent Lefevre, and Manish Parashar. Distributed operator placement for iot data analytics across edge and cloud resources. In *2019 19th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*, pages 459–468. IEEE, 2019.