Research internship (Master 2)
Contribution to decision-making models for the Edge-Cloud Continuum

Daniel Balouek-Thomert
Inria, IMT Atlantique, France

2024

Keywords : Edge-to-Cloud continuum ; urgent computing ; Cloud Computing ; Event-driven systems

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 [1]. Simultaneously, distributed applications also evolve. Urgent computing tackles services that require 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 [2].

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 [3]. Second, these services and applications require system support to program reactions that occur at runtime, especially when the target infrastructure capacities and capabilities are unknown during the design [4].

This internship focuses on decision support for the management on resources distributed across the Edge-Cloud Computing continuum. The location (which physical server) and the execution parameters (which software configuration) have significant effect on the performance of applications. Particularly, emerging data-driven applications are built as compositions of individual functions deployed between the edge of the network (where data is produced from physical/virtual sensors) and the cloud (where final data-processing steps are usually performed). Each function consists in stand-by code with input parameters depending on the previous step and the overall performance expected by the developers. Managing the application as a whole requires decision mechanisms to perform the placement of individual functions through uncertainty and constraints by identify runtime events, and trigger appropriate policies.

2 Expected work

In this context, the successful candidate will be in charge of proposing and evaluating a decision model. The main challenge consists in modeling the different outcomes of the system, and deducing management policies based on their impact.

The objectives of this internship are :

- a state-of-the-art to assess the concepts associated to the computing continuum and data-driven analytics
- proposing a mathematical model integrating the status, events and performance metrics of a system deployed across edge and cloud resources
- evaluating the model on a real-platform using an urgent application

In this modeling and evaluation tasks, consideration of metrics relative to latency, quality of service, and throughput during the lifecycle of an urgent application is of particular importance.

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
- Basics in probability and statistic
- 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

- Daniel Balouek-Thomert, Inria, IMT Atlantique daniel.balouek@inria.fr

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, Eduard Gibert Renart, Ali Reza Zamani, Anthony Simonet, and Manish Parashar. Towards a computing continuum : Enabling edge-to-cloud integration for data-driven workflows. *The International Journal of High Performance Computing Applications*, 33(6) :1159–1174, 2019.

[2] 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`.

[3] Kevin Fauvel, Daniel Balouek-Thomert, Diego Melgar, Pedro Silva, Anthony Simonet, Gabriel Antoniu, Alexandru Costan, Véronique 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`.

[4] 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.