



Offer #2025-08736

PhD Position F/M Latency-Driven container network optimization in edge industrial IoT

Contract type : Fixed-term contract

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

The Inria Rennes - Bretagne Atlantique Centre is one of Inria's eight centres and has more than thirty research teams. The Inria Center is a major and recognized player in the field of digital sciences. It is at the heart of a rich R&D and innovation ecosystem: highly innovative PMEs, large industrial groups, competitiveness clusters, research and higher education players, laboratories of excellence, technological research institute, etc.

Assignment

Cloud computing and its three facets (IaaS, PaaS and SaaS) have become essential in today Internet Applications, offering many advantages such as scalability, elasticity or flexibility. With its different service models, the cloud still faces many issues prone to impact either the end-user (QoS), the provider (Cost) and the environment (Sustainability).

The Edge computing is a paradigm that address such issues by provisioning resources outside the cloud and closer to the end-device, at the edge of the network.

This allows to reduce the latency and minimize the traffic between end-user and the cloud platform [4]. Several studies have shown that edge systems can indeed reduce the latency compared to cloud systems, but this reduction is not guaranteed and will highly depend on the components placement, leading sometimes to worse performance [11]. It has been also demonstrated that less traffic is sent to the cloud when using edge systems. However, a lack of a proper monitoring or reconfiguration mechanism in the edge exists, especially when the application is related to the IoT [9], where the cloud infrastructure is known not to be a viable solution. Critical industrial IoT use case scenario, such as real-time manufacturing robot application or monitoring of critical equipment, leverage the adequacy of the edge paradigm that could mitigate the latency in such specific scenario.

However, evaluating realistic large-scale edge infrastructure constitute a complex task given the cost of deployment and the absence of a realistic view of the real-world deployments. In an IoT context, geo-distributed edge infrastructures mostly rely on SDN approaches[6] that contribute to conceal the networking aspects such as the topology or the routing decisions[10]. In consequence, it appears that the impact of the elasticity of a edge solution is mainly evaluated on the data plane side[5].

Moreover, standard container orchestration platforms, such as Kubernetes, rely on the Container Network Interface (CNI) to define the network requirements and manage the connectivity of containerized workloads. Consequently, the performance of the network is heavily influenced by the specific implementation of the CNI [3]. In latency-critical applications, dynamic changes in the network can significantly impact the responsiveness of the CNI, posing challenges for routing technologies. Segment Routing (SR) is one among multiple technologies that offers potential solutions to these challenges due to its adaptability to network changes. While some studies have explored the integration of SR with Kubernetes [8], they have predominantly focused on throughput metrics. There remains a significant research gap in addressing SR's implications for latency, which warrants further investigation.

In the context of IIoT applications (i.e. critical response time environments such as Manufacturing or Monitoring of critical equipment) latency is at the center of a tremendous number of studies to optimize the placement of resources in distributed architectures. To ensure that the quality of service is guaranteed, several solutions exist to reconfigure the components placement (migration) and can reduce the overall latency by changing the components and routes. However, knowing precisely which component is the source of the problematical latency remains scarcely addressed. When taking a decision for a reconfiguration or a migration, which can be triggered due to latency issue, it can be beneficial to check if the source of the latency can be solved before instantiating a migration or a full reconfiguration. Some studies exist where a comparison of response time is done between the major cloud actors depending on the load [7]. Proper measurement protocols exist but always refer to specific case studies [1, 2] and would not allow to be integrated in edge systems.

Objectives:

The objective of this thesis is to study the optimization of the container network in Edge-based IIoT systems based on latency measurement, by evaluating the control plane cost of a change in the architecture. It will particularly address the problem of how to identify the origin of a latency issue, and based on this finding, propose a routing optimization that take into account the cost and elasticity of the control plane.

References

- [1] Da?niel Ge?hberger, Da?vid Balla, Markosz Maliosz, and Csaba Simon. Performance evaluation of low latency communication alternatives in a containerized cloud environ- ment. In 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), pages 9–16, 2018.
- [2] Devasena Inupakutika, Gerson Rodriguez, David Akopian, Palden Lama, Patricia Chalela, and Amelie G. Ramirez. On the performance of cloud-based mhealth ap- plications: A methodology on measuring service response time and a case study. *IEEE Access*, 10:53208–53224, 2022.
- [3] Zhuangwei Kang, Kyoungho An, Aniruddha Gokhale, and Paul Pazandak. A compre- hensive performance evaluation of different kubernetes cni plugins for edge-based and containerized publish/subscribe applications. 2021 IEEE International Conference on Cloud Engineering (IC2E), pages 31–42, 2021.
- [4] Zheng Li and Francisco Millar-Bilbao. Characterizing the cloud’s outbound network latency: An experimental and modeling study. In 2020 IEEE Cloud Summit, pages 172–173, 2020.
- [5] Carla Mouradian, Diala Naboulsi, Sami Yangui, Roch H. Glitho, Monique J. Morrow, and Paul A. Polakos. A comprehensive survey on fog computing: State-of-the-art and research challenges. *IEEE Communications Surveys and Tutorials*, 20(1):416–464, 2018.
- [6] Feyza Yildirim Okay and Suat Ozdemir. Routing in fog-enabled iot platforms: A survey and an sdn-based solution. *IEEE Internet of Things Journal*, 5(6):4871–4889, 2018.
- [7] Istva?n Pelle, Ja?nos Czentye, Ja?nos Do?ka, and Bala?zs Sonkoly. Towards latency sen- sitive cloud native applications: A performance study on aws. In 2019 IEEE 12th International Conference on Cloud Computing (CLOUD), pages 272–280, 2019.
- [8] Jose? Santos, Jeroen van der Hooft, Maria Torres Vega, Tim Wauters, Bruno Volck- aert, and Filip De Turck. Srfog: A flexible architecture for virtual reality content delivery through fog computing and segment routing. 2021 IFIP/IEEE International Symposium on Integrated Network Management (IM), pages 1038–1043, 2021.
- [9] U. Tomer and P. Gandhi. An enhanced software framework for improving qos in iot. *Engineering, Technology and Applied Science Research*, 12(5):9172–9177, Oct. 2022.

[10] Benjamin Warnke, Yuri Cotrado Sehgelmeble, Johann Mantler, Sven Groppe, and Stefan Fischer. Simora: Simulating open routing protocols for application interoperability on edge devices. 2022 IEEE 6th International Conference on Fog and Edge Computing (ICFEC), pages 42–49, 2022.

[11] Sami Yangui, Pradeep Ravindran, Ons Bibani, Roch H. Glitho, Nejib Ben Hadj-Alouane, Monique J. Morrow, and Paul A. Polakos. A platform as-a-service for hybrid cloud/fog environments. In 2016 IEEE International Symposium on Local and Metropolitan Area Networks (LANMAN), pages 1–7, 2016.

Main activities

- Explore the State-of-the-Art of the IoT/Edge Emulation/Simulation platforms
- Integrate an IIoT solution in a Edge architecture platform with a latency measurement
- Propose a profile and a classification of latency issues
- Through the use of Segment Routing, propose an innovative way to optimize CNI taking into account the latency metrics and the control plane capabilities

Skills

- A master degree in distributed systems and/or Cloud computing/Networking
- Good knowledge of distributed systems
- Good programming skills (e.g., C++ and Python)
- Basic knowledge of simulation
- Excellent communication and writing skills in English (Note that knowledge of French is appreciated but not required for this position)
- Knowledge of the following technologies is not mandatory but will be considered as a plus:
 - Cloud resource scheduling
 - Routing, Software Defined networks
 - Revision control systems: git, svn
 - Linux distribution: Debian, Ubuntu

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking (after 6 months of employment) and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

Remuneration

monthly gross salary amounting to 2200 euros

General Information

- **Theme/Domain** : Distributed Systems and middleware System & Networks (BAP E)
- **Town/city** : Rennes
- **Inria Center** : [Centre Inria de l'Université de Rennes](#)
- **Starting date** : 2025-09-01
- **Duration of contract** : 3 years
- **Deadline to apply** : 2025-05-14

Contacts

- **Inria Team** : [MAGELLAN](#)
- **PhD Supervisor** :
Lemerrier Francois / francois.lemerrier@irisa.fr

About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

The keys to success

Doing a PhD is a job unlike any other. Please read this document carefully to decide whether a PhD is the right career move for you:

<https://medium.com/great-research/do-you-need-a-ph-d-f78d2fb0f286>

Warning : you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

Instruction to apply

Defence Security :

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

Recruitment Policy :

As part of its diversity policy, all Inria positions are accessible to people with disabilities.