Keynote Talk: Deployment and Management of Edge Microservices

Keynote Talk: Deployment and Management of Edge Microservices

Carla Fabiana Chiasserini
Politecnico di Torino
Torino, Italy

KEYWORDS

Edge computing, Service virtualization, Service interference, Migration, Container retention

ACM Reference Format:
Carla Fabiana Chiasserini. 2023. Keynote Talk: Deployment and Management of Edge Microservices. In *Proceedings of the 24th International Conference On Distributed Computing And Networking (ICDCN '23).* ACM, New York, NY, USA, 1 page. https://doi.org/XXXXXXX.XXXXXXX

ABSTRACT

Edge computing is an emerging technology for present and next generation mobile networks, which, unlike the cloud, can meet the low latency or bandwidth consumption requirements of time- and mission-critical services. One essential component of edge computing is service virtualization, with each service being often defined as a set of virtual functions implemented through containers.

This talk introduces the main scientific challenges in microservice deployment and management at the edge, including microservice interference avoidance, migration, and retention. Specifically, interference among microservices arises whenever the associated containers run on the same server and, hence, compete for memory resources, even if they are allocated dedicated cores. Such interference can lead to severe throughput degradation, thus harming the microservices performance [1]. Microservice migration is instead pivotal to continuously meeting low-latency requirements, as users or devices move from one access point to another [2]. Characterizing container migration is therefore critical for guaranteeing that the expected QoE is ensured, while minimizing the migration cost for the system. Finally, microservice retention becomes relevant in the presence of serverless computing platforms that can launch multiple isolated containers to fulfill service requests by mobile users. The creation of a new container implementing a microservice may require fetching the corresponding image from the remote repository and fetching and loading essential libraries and dependencies before executing the actual function. This long delay involved in the initialization setup is known as cold-start latency, which represents one of the main performance issues faced by the serverless computing platforms [3]. Reducing the cold-start latency is a hard task due to the infrequent function invocations and their unpredictable patterns.

For each of these aspects, we present possible approaches and solutions, providing interesting insights obtained through experimental measurements, as well as highlighting how such aspects can be conveniently modelled.

ICDCN '23, January 04–07, 2023, Kharagpur, India

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

ACKNOWLEDGMENTS

This work was supported by the EU Commission through the H2020 MSCA-ITN SEMANTIC Grant No. 861165 and by the Qatar National Research Fund (a member of Qatar Foundation) under the NPRP-S 13th Cycle Grant No. NPRP13S-0205-200265.

REFERENCES

[1] Madhura Adeppady, Carla Fabiana Chiasserini, Holger Karl, and Paolo Giaccone. 2022. iPlace: An Interference-aware Clustering Algorithm for Microservice Placement. In *IEEE ICC*.
[2] TianZhang He, Adel N. Toosi, and Rajkumar Buyya. 2021. SLA-aware multiple migration planning and scheduling in SDN-NFV-enabled clouds. *Journal of Systems and Software* 176 (June 2021), 1069–1078.
[3] Li Pan, Lin Wang, Shutong Chen, and Fangming Liu. 2022. Retention-Aware Container Caching for Serverless Edge Computing. In *IEEE INFOCOM 2022 - IEEE Conference on Computer Communications*. 1069–1078. https://doi.org/10.1109/INFOCOM48880.2022.9796705

Received 27 September 2023