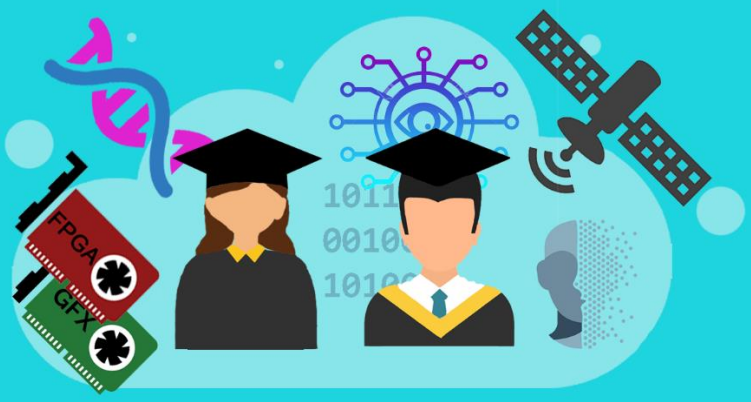
Diploma Thesis

Microproccessors and Digital Systems Laboratory



Interference-aware container orchestration on Kubernetes based on Synthetic to Real-world Workload Pressure Mapping

Over the last decade, the adoption of cloud computing has seen explosive growth, both at consumer and enterprise levels and will continue to rise in the future [1]. The evolution and endorsement of container-based virtualization technology, as well as the advantages that the Cloud computing offers both to users and operators, have acted as enablers towards this direction. Users can launch different kind of applications and services, only paying for the resources used at a time, while economies of scale are enabled for operators, who are sharing their resources across several users. This increment in the amount of workloads uploaded and executed on the cloud, have forced data-center (DC) operators and cloud providers, such as Google Cloud Platform [2] and Amazon EC2 [3], to embrace workload co-location and multi-tenancy as first class system design concern.

However, this sharing of resources between separate users does not come “on the house”. Workloads placed on the same physical machines continuously contest for shared resources, such as cache and memory occupancy, memory/network band- width and others, causing interference to each other, which, in turn, induces huge negative impact on their performance. This situation becomes even more consecutive, as cloud providers currently provide users with elasticity and resizability of their computing capacity, leading to a dynamic provisioning of resources.

iBench[4] is a suite that provides contentious micro-benchmarks that apply pressure on different resources of a system. However, in contrast to real workloads this pressure is stable. In this thesis, we need to map different phases of real-world workloads to combinations of various artificial (or synthetic) workload pressures in order to enhance the robustness of a model that identifies interference on a specific machine.

Keywords:

Cloud computing, Kubernetes, Interference-aware

Related Work:

[1] C. V. networking Index, “Forecast and methodology, 2016-2021, whitepaper,” San Jose, CA, USA, vol. 1, 2016.

[2] Google Cloud Platform, https://cloud.google.com

[3] Amazon Elastic Compute Cloud, https://aws.amazon.com/ec2

[4] Delimitrou, Christina, and Christos Kozyrakis. "ibench: Quantifying interference for datacenter applications." 2013 IEEE international symposium on workload characterization (IISWC). IEEE, 2013.

Prerequisites:

- Linux, Bash/Shell scripting, eager to learn new things
- Experience on other programming languages, e.g. C, Golang, Python, would be a plus

Knowledge & Experience the student will acquire:

- A broader understanding of cloud computing architectures
- Work on technologies for automation and deployment
- Research, and become familiar with Kubernetes container orchestrator and its internals

Duration: At least 9 months

Contact:

Achilleas Tzenetopoulos Ph.D. student: (atzenetopoulos@microlab.ntua.gr)

Dimosthenis Masouros Ph.D. student: (dmasouros@microlab.ntua.gr)

Sotirios Xydis Ass. Prof.: (sxydis@microlab.ntua.gr)

Dimitrios Soudris Prof.: (dsoudris@microlab.ntua.gr)