

Diploma Thesis

Microprocessors and
Digital Systems
Laboratory



Explore Resource Configurations for Serverless Functions

Serverless computing has emerged as a transformative paradigm, enabling Cloud Service Providers (CSPs) to provision, deploy, and scale client resources seamlessly. By offloading infrastructure management to CSPs, serverless computing facilitates efficient multiplexing of data center resources, significantly reducing operational overhead and delivering cost savings to users.

However, serverless platforms still require users to provision the resources for their functions, i.e., Memory, and vCPUs. As shown in prior art [1], current resource provisioning does not always provide Pareto Optimal solutions. Other works [2] have started including Core frequency scaling in the resource configuration mix, further increasing the solution space.

This thesis aims to investigate a broader range of configuration parameters that influence function latency. By analyzing key performance metrics—such as latency, resource utilization, and power/energy consumption—the student will develop a methodology to “right-size” serverless functions based on different constraints, e.g., user-specified Service Level Objectives (SLOs), and current Virtual Machine resources availability. The proposed solution will be integrated into widely adopted serverless frameworks, such as Knative [3].

Prerequisites: Familiarity with Linux, bash scripting, Python, distributed computing, and basic concepts of AI/ML.

Contact:

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

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

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

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

[1] Bilal, Muhammad, et al. "With great freedom comes great opportunity: Rethinking resource allocation for serverless functions." Proceedings of the Eighteenth European Conference on Computer Systems. 2023.

[2] A. Tzenetopoulos, D. Masouros, D. Soudris and S. Xydis, "DVFaas: Leveraging DVFS for FaaS Workflows," in IEEE Computer Architecture Letters, vol. 22, no. 2, pp. 85-88, July-Dec. 2023, doi: 10.1109/LCA.2023.3288089. keywords: {Quality of service;Frequency control;Frequency measurement;Time-frequency analysis;Power demand;Monitoring;Timing;Quality of Services;power management;distributed systems;emerging technologies},

[3] <https://knative.dev/docs/>