Prof. Ece Güran Schmidt / TUBITAK

ACCLOUD (ACcelerated CLOUD): A Novel, FPGA Accelerated Cloud Architecture

Principle Investigator: Prof. Ece Güran Schmidt [http://eee.metu.edu.tr/tr/personel/senan-ece-guran-schmidt]

Project Manager: Prof. Ece Güran Schmidt [http://eee.metu.edu.tr/tr/personel/senan-ece-guran-schmidt] (METU), Yük. Müh. Alper Yazar [https://www.alperyazar.com] (ASELSAN [http://www.aselsan.com.tr])

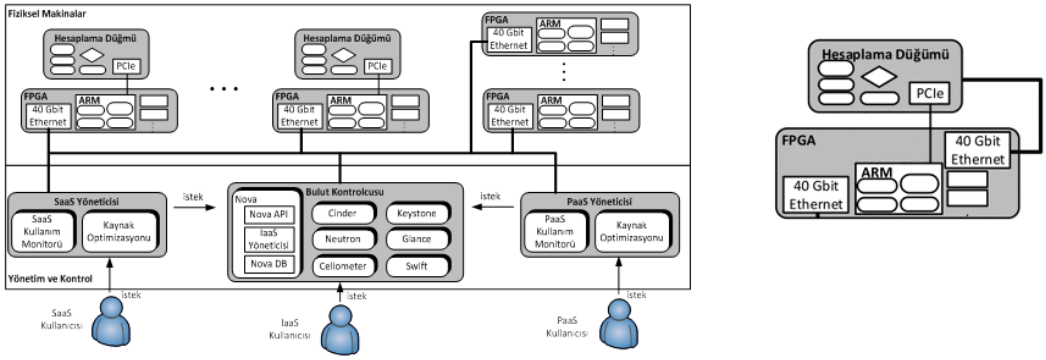
Project Type: TÜBİTAK/ARDEB 1003 – Primary Subjects R&D Funding Program

Project Duration: 36 Months

Project Start Date: April 2018

Project Budget: 1.117.059,00 TL

Funded Personnel: 1 PhD (full-time, 36 months), 2 MSc (full-time, 18 months) student(s)



Cloud computing has become a popular computing model in the recent years as it exploits the economies of scale for efficient use of computing resources. The data centers of today are mostly cloud based with virtualized servers to provide on-demand scalability and flexibility of the available resources such as CPU, memory, data storage and network bandwidth. A cloud data center provider may offer Infrastructure as a Service (IaaS), where the user gets a virtual machine (VM) with processing, memory, storage and networking resources, which can be installed with any desired operating system and software. Differently, Platform as a Service (PaaS) commonly provides a ready environment with operating system, programming language execution environment, database and web server for developers to test and deploy their programs and applications. Finally in Software as a Service (SaaS), the user only accesses the provided application for example via a web browser without any control of the underlying infrastructure.

In this respect, it is necessary to maintain a proper mapping of the virtual resources to the underlying physical hardware. In addition, the available physical resources must be efficiently allocated according to user requests and the respective service such as IaaS, PaaS or SaaS. This is a particularly difficult task, considering the cloud resource heterogeneity, the unpredictable nature of workload, and the diversified objectives of cloud users. Finally, it is desirable to always select the most appropriate resource type depending on the user requests.

The slowdown of Moore's law and the increased data and problem sizes together with the development of the high performance programmable hardware platforms such as FPGAs increase the popularity of hardware accelerators which can provide better performance and less energy consumption depending on the problem properties and size. To this end, the very recent research focuses on employing hardware resources in cloud based data centers. Integrating hardware resources in the cloud based data center should be seamless, together with virtualization and dynamic resource allocation capabilities.

This project proposes a novel architecture for cloud-based data centers that we call ACCLOUD (ACcelerated CLOUD). The proposed architecture features FPGA hardware resources that can be offered to users in the scope of IaaS, PaaS and SaaS. To this end, we propose augmenting the cloud servers with FPGA as well as to employ standalone FPGA units. All FPGA units are directly connected to the data center fabric to avoid slowdown of the processing of high speed network traffic. The FPGA resources are virtualized using a number of run-time partially reconfigurable regions. We propose to use the OpenStack software framework to allocate these partially reconfigurable regions to the users together with other virtualized computing resources. **To this end, the first major contribution of this project is offering hardware resources as IaaS, PaaS and SaaS in the ACCLOUD architecture different from all previous work.** In the scope of IaaS, a user can get partially reconfigurable regions to program his own bit streams to realize any desired functionality. A user can get a ready hardware accelerator as PaaS to process his own data. Furthermore ACCLOUD enables SaaS on FPGA by providing both the hardware and software realizations of the same software application, while running the users data on the appropriate realization based on its size and properties. **The second major contribution is the entirely novel resource management approach ACCLOUD-MAN that incorporates the hardware resources into the existing CPU, memory, bandwidth and disk resources and coordinates the IaaS, PaaS and SaaS management.** To this end, ACCLOUD-MAN achieves near globally optimum resource allocation together with increased performance by employing hardware resources for SaaS whenever appropriate. In the scope of the project, we will develop the FPGA components together with reconfigurable regions and high-speed switching structures between different sources of data. We will extend the widely used OpenStack Cloud software framework to include the hardware resources in the databases and messaging APIs. This enables allocating them together with the existing computing

resources. We will formulate a bin-packing based optimization problem for the coordinated optimal allocation of resources, which utilizes resource monitoring and prediction to achieve most efficient resource allocation that meets the user requirements in ACCLOUD. **The third major contribution of the project is the extensive optimization evaluation, simulation and real test bed implementation and evaluation of ACCLOUD and ACCLOUD-MAN.** To this end, we will develop hardware accelerators for high speed network processing applications such as deep packet inspection and encryption. These accelerators and their software implementations will be used for performance testing together with request workload generation modules that will be developed in the scope of the project.

The project is proposed in the form of one main project (METU) with one sub-project (ASELSAN) together with competent project teams. The METU project team has experience in research and development of high speed hardware architectures for computer networking and network processing. The ASELSAN project team has experience in the development and implementation of advanced FPGA-based architectures. The project outcomes will be papers in high quality journals, PhD and MSc theses as well as possible patent applications for the SaaS on hardware approach.