Chair of Decentralized Systems Engineering
Department of Informatics

Enabling Near Data Processing in the cloud

General information	**Advisor**	Dipl-Ing Charalampos Mainas Dr. Atsushi Koshiba
	Email	charalampos.mainas@in.tum.de, atsushi.koshiba@tum.de
	Date	17.11.2022
Type	Master / Bachelor / Guided research	
Description	In the traditional Von Neumann architecture the processing unit (e.g. CPU) is separated from the storage devices (DRAM, HDD, SSD). As a result, the data constantly moves between the processing unit and the storage devices. Although von Neumann was the dominant architecture for many years, modern workloads (ML/AI etc) and the continuously increased volume of generated data have created new challenges. The constant move of big volumes of data imposes significant performance bottlenecks and increases the energy consumption. In an effort to provide a solution for the above problem, researchers explore the idea of Near Data Processing (NDP). Smaller and more flexible computation units, such as FPGAs or embedded processors are placed close to the storage devices. Therefore instead of moving the data, the computation is taking place closer to the data storage. Some of the advantages of NDP are:	

- Lower cost and energy consumption, due to reduced data movement
- Increased performance for data-intensive applications
- Release of main computational unit, which can execute other processes

Even though NDP is not a new idea, only recently storage devices with embedded computational units have been available. Furthermore, there is limited research related to the security of NDP in a multi-tenant environment. Consequently, NDP is not mature enough to be adopted in modern data centers.

There are plenty of projects that apply the idea of NDP in well known data-intensive applications such as databases and ML training models (4,5). Hayagui (2), ZCSD (1) and blockNDP (3) target more generic use cases, yet they rely on specific hardware. Moreover, most of the research in NDP does not take into consideration multi-tenancy. In our case, we will use FPGAs as accelerators and we aim to design a framework for generic use cases, where different kinds of applications can offload part of their execution in the FPGAs. Furthermore, our approach will support virtualization technologies in order to facilitate NDP in a multi-tenant environment.

In conclusion, during the thesis we will analyze the prior work and discuss their applicability in virtualization technologies. Next, we will design and implement a system which offloads the data-intensive workloads of virtualization environments to the processing unit located closer to the data. Finally, based on a FPGA virtualization solution, created by our lab, we will enable the virtualization of NDP resources.

Keywords

FPGA, hardware acceleration, computer architecture, operating systems, NDP

Goals

Concrete outcomes

1. Analysis and description of the problem, along with a survey of relevant literature.

2. Analysis of potential solutions.
3. Design and implementation of a NDP solution for the cloud
4. Evaluation of the new solution

Bonus points

5. Preparation (with intent to publish) of a paper resulting from this work.

Prerequisites

Compulsory

- Knowledge of C and C++
- Experience in operating system or systems' programming
- Good understanding of memory & storage hierarchy

Preferred

- Knowledge of OpenCL
- Knowledge of virtualization technologies
- Experience on using FPGAs

References

1. https://arxiv.org/pdf/2112.00142.pdf
2. https://micchie.net/files/hotstorage20-hayagui.pdf
3. https://www.research.ed.ac.uk/en/publications/blockndp-block-storage-near-data-processing
4. -https://www.vldb.org/pvldb/vol7/p963-woods.pdf
5. https://www.usenix.org/conference/atc19/presentation/liang
6. http://cidrdb.org/cidr2021/papers/cidr2021_paper29.pdf

Application process

Please send an email to the advisor including the following:

- Email subject: "Thesis application (DSE)"
- CV
- A copy of your transcript(s)
- A **motivation statement**, please include samples of your work that you are proud of (e.g., major projects, open-source contributions, Github page, etc.) and/or writing samples (e.g., your technical blog, project reports, etc.)