Forschungsinstitut
Cyber Defence
Universität der Bundeswehr München

Bachelor Thesis

Confidential Computing via Hardware Trusted Execution Environments by an OpenStack HPC capable cloud

Second lieutenant, Representative, Valentin Pfeil

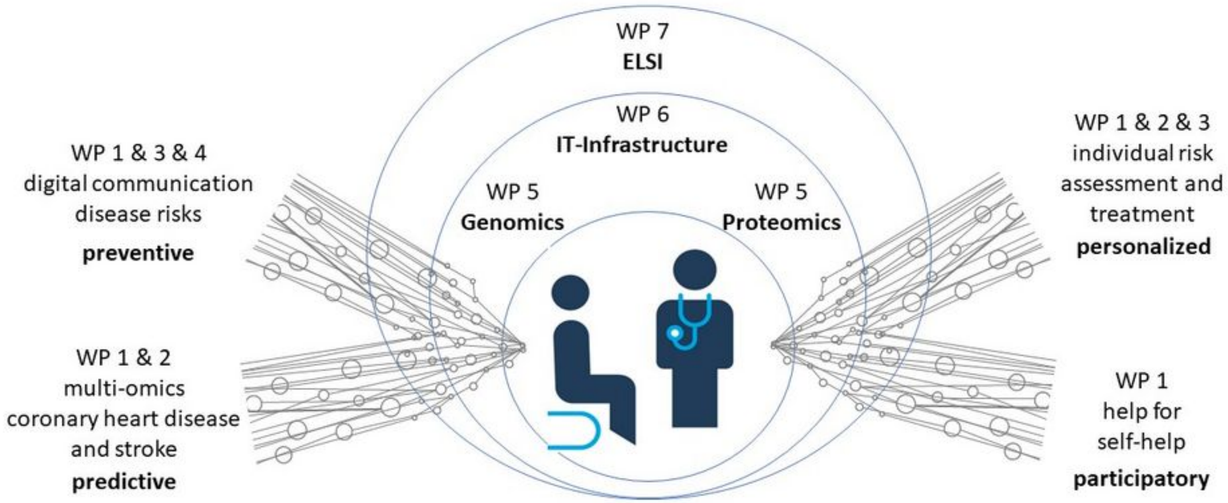
Stakeholder

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

lrz Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften

Dislocation

DigiMed Bayern



Requirements:

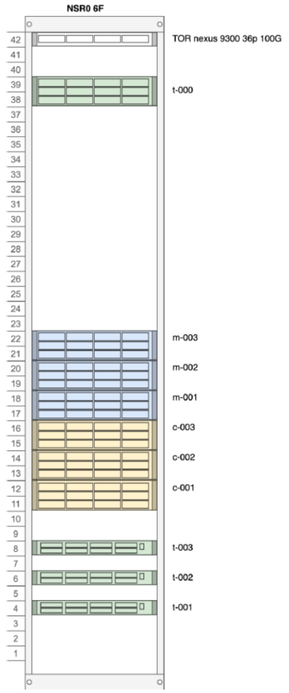
- Confidentiality (GPDR compliance)
- High Performance (Big Data and Artificial Intelligence)
- Flexibility (Private Cloud)
- **Approach:**
  - **AMD Infinity Guard (incl. AMD SEV)**
  - **HPC hardware (CPU: AMD EPYC 75F3, ...) and software components (OpenMPI/SLURM, ...)**
  - **Virtualization and cloud-services (OpenStack/QEMU-KVM)**

Management servers

Count	Description	CPU	RAM	Storage	Network
m-001	Lenovo ThinkSystem SR665	1x AMD EPYC 75F3 32 cores @2.95GHz up to 4.0GHz, 8 Memory Channels	4x 64 GB DDR 4	2x 3.84TB SATA SSD RAID 1	1x 100GbE 1x 40GbE 1x 1GbE
m-002	Lenovo ThinkSystem SR665	1x AMD EPYC 75F3 32 cores @2.95GHz up to 4.0GHz, 8 Memory Channels	4x 64 GB DDR 4	2x 3.84TB SATA SSD RAID 1	1x 100GbE 1x 40GbE 1x 1GbE
m-003	Lenovo ThinkSystem SR665	1x AMD EPYC 75F3 32 cores @2.95GHz up to 4.0GHz, 8 Memory Channels	4x 64 GB DDR 4	2x 3.84TB SATA SSD RAID 1	1x 100GbE 1x 40GbE 1x 1GbE

Compute servers

Count	Description	CPU	RAM	Storage	Network
c-001	Lenovo ThinkSystem SR665	2x AMD EPYC 75F3 32 cores @2.95GHz up to 4.0GHz, 8 Memory Channels	16x 64 GB DDR 4	1x 800GB NVMe SSD	1x 100GbE 1x 40GbE 1x 1GbE
c-002	Lenovo ThinkSystem SR665	2x AMD EPYC 75F3 32 cores @2.95GHz up to 4.0GHz, 8 Memory Channels	16x 64 GB DDR 4	1x 800GB NVMe SSD	1x 100GbE 1x 40GbE 1x 1GbE
c-003	Lenovo ThinkSystem SR665	2x AMD EPYC 75F3 32 cores @2.95GHz up to 4.0GHz, 8 Memory Channels	16x 64 GB DDR 4	1x 800GB NVMe SSD	1x 100GbE 1x 40GbE 1x 1GbE



- **RQ1:** How does the security attestation of TEEs work?
- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?
- **RQ3:** How is Performance affected when TEEs are implemented in a confidential HPCaaS?

- **Deployment**
  - OpenStackClient
  - Terraform
  - Ansible
  - OpenMPI
  - SLURM
  - GROMACS
- **Approach**
  - **RQ1:** One node, AMD SEV enabled, Certificate inspection
  - **RQ2:** General deployment, configuration and operation of HPC cluster, determination and evaluation of usability
  - **RQ3:** GROMACS benchmarks over SLURM/ OpenMPI
  - 1/3/10 nodes

RQ1: How does the security attestation of TEEs work?

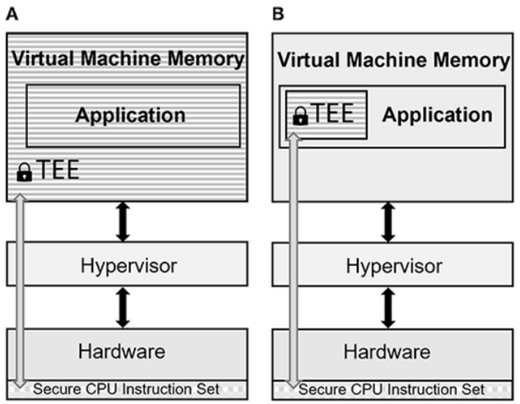
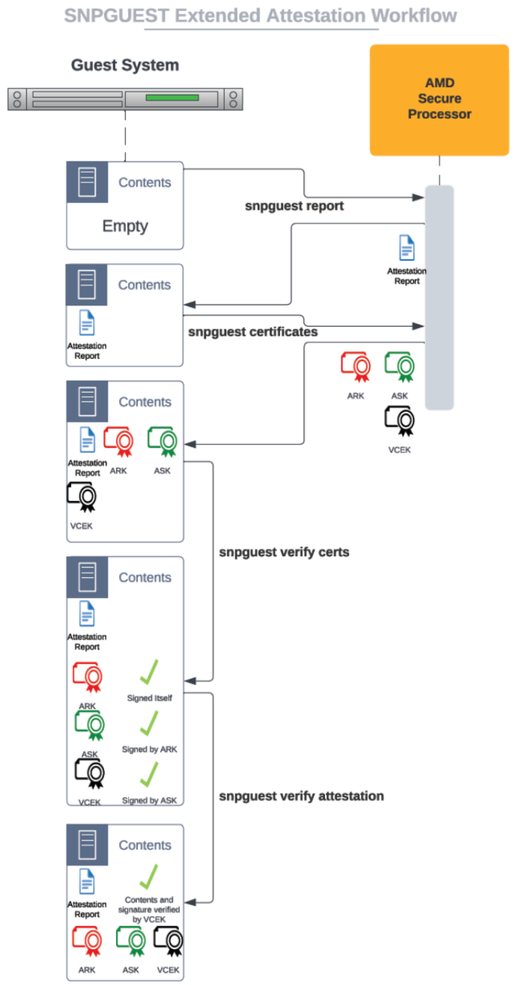


Figure 2.4: **TEE cloud computing [15]**
A: **Virtual machine-based model**, the whole memory of the virtual machine is encrypted.
B: **Process-based model**, only the memory of the enclave is encrypted.

RQ1: How does the security attestation of TEEs work?



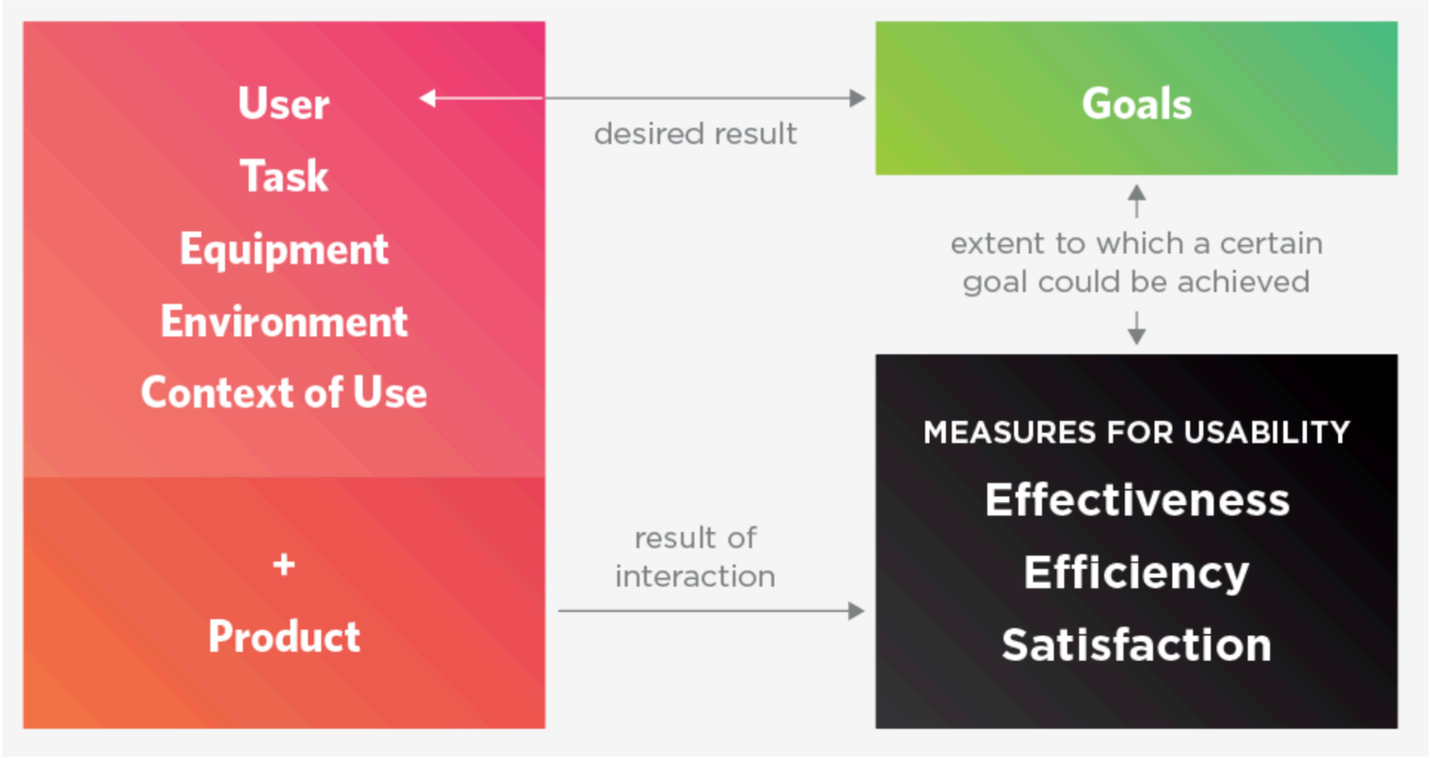

```
ubuntu@control.cloud.digimed.lrz.de (2)

ubuntu@cpu-sev-1:~$ ./sevctl/target/release/sevctl ok
[ PASS ] - AMD CPU
[ PASS ]   - Microcode support
[ FAIL ]   - Secure Memory Encryption (SME)
[ PASS ]   - Secure Encrypted Virtualization (SEV)
[ FAIL ]     - Encrypted State (SEV-ES)
[ FAIL ]     - Secure Nested Paging (SEV-SNP)
[ SKIP ]       - VM Permission Levels
[ SKIP ]         - Number of VMPLs
[ PASS ]     - Physical address bit reduction: 1
[ PASS ]     - C-bit location: 51
[ PASS ]     - Number of encrypted guests supported simultaneously: 0
[ PASS ]     - Minimum ASID value for SEV-enabled, SEV-ES disabled guest: 0
[ FAIL ]     - SEV enabled in KVM: Error - /sys/module/kvm_amd/parameters/se
v does not exist
[ FAIL ]     - SEV-ES enabled in KVM: Error - /sys/module/kvm_amd/parameters
/sev_es does not exist
[ FAIL ]     - Reading /dev/sev: /dev/sev not readable: No such file or dire
ctory (os error 2)
[ FAIL ]     - Writing /dev/sev: /dev/sev not writable: No such file or dire
ctory (os error 2)
[ PASS ]   - Page flush MSR: DISABLED
[ FAIL ] - KVM supported: Error reading /dev/kvm: (No such file or directory
 (os error 2))
[ PASS ] - Memlock resource limit: Soft: 982831104 | Hard: 982831104
Error: One or more tests in sevctl-ok reported a failure
```

- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?

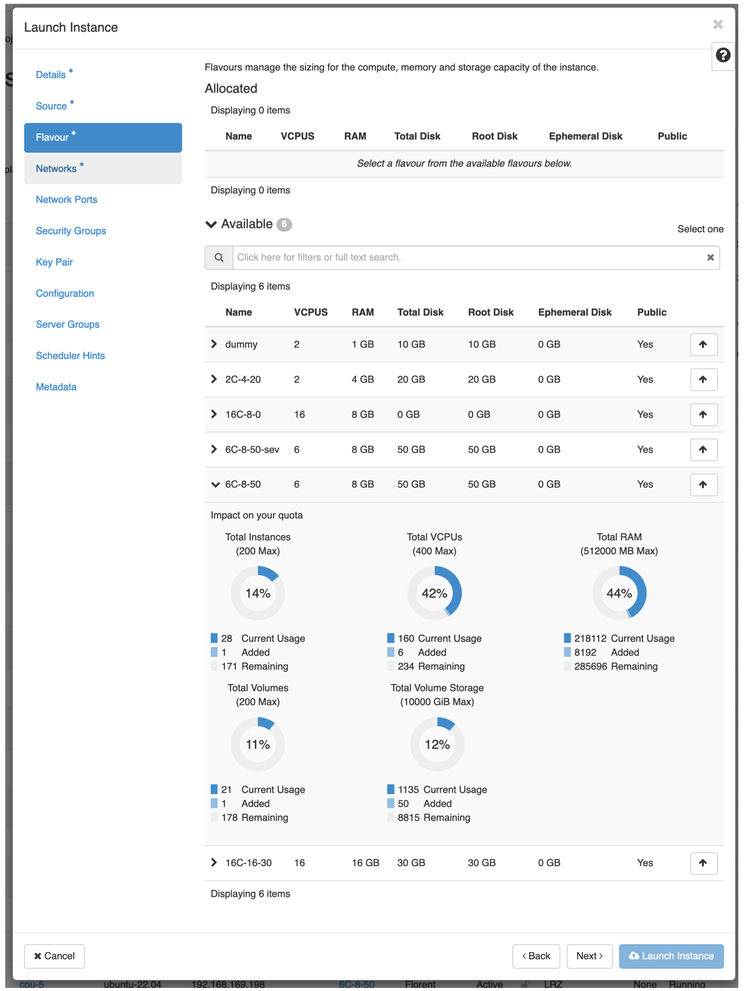
- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?

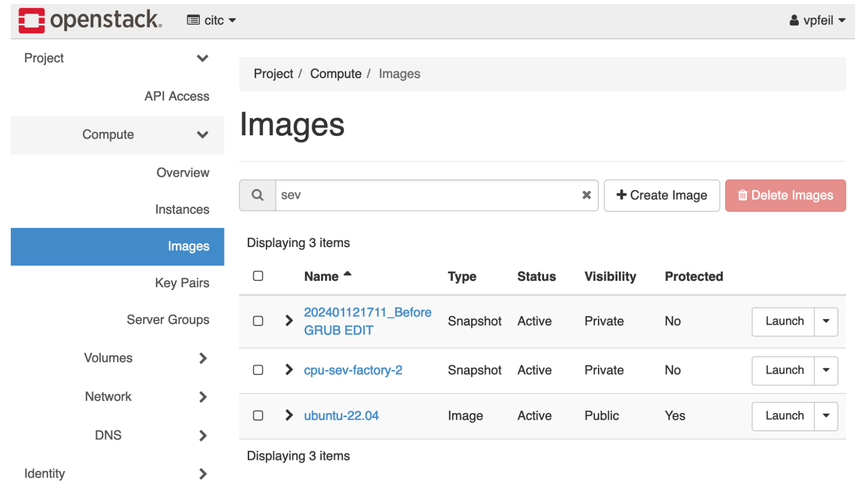
EN ISO 9241-11:2018



- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?

Frontend





- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?

Backend

- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?

Backend

Prerequisites:

- QEMU-KVM with **libvirt.virt_type** (driver)
- At least one of the **Nova compute nodes** must be capable of **supporting SEV**
- **Flavor/image requirements:**
  - Flavor property **hw:mem_encryption=true**
  - **In any case,** SEV instances have to have their boot images with **hw_firmware_type property** set to **uefi**
  - Images property have to have **hw_machine_type=q35** or per compute node via **libvirt.hw_machine_type** set to **x86_64=q35**

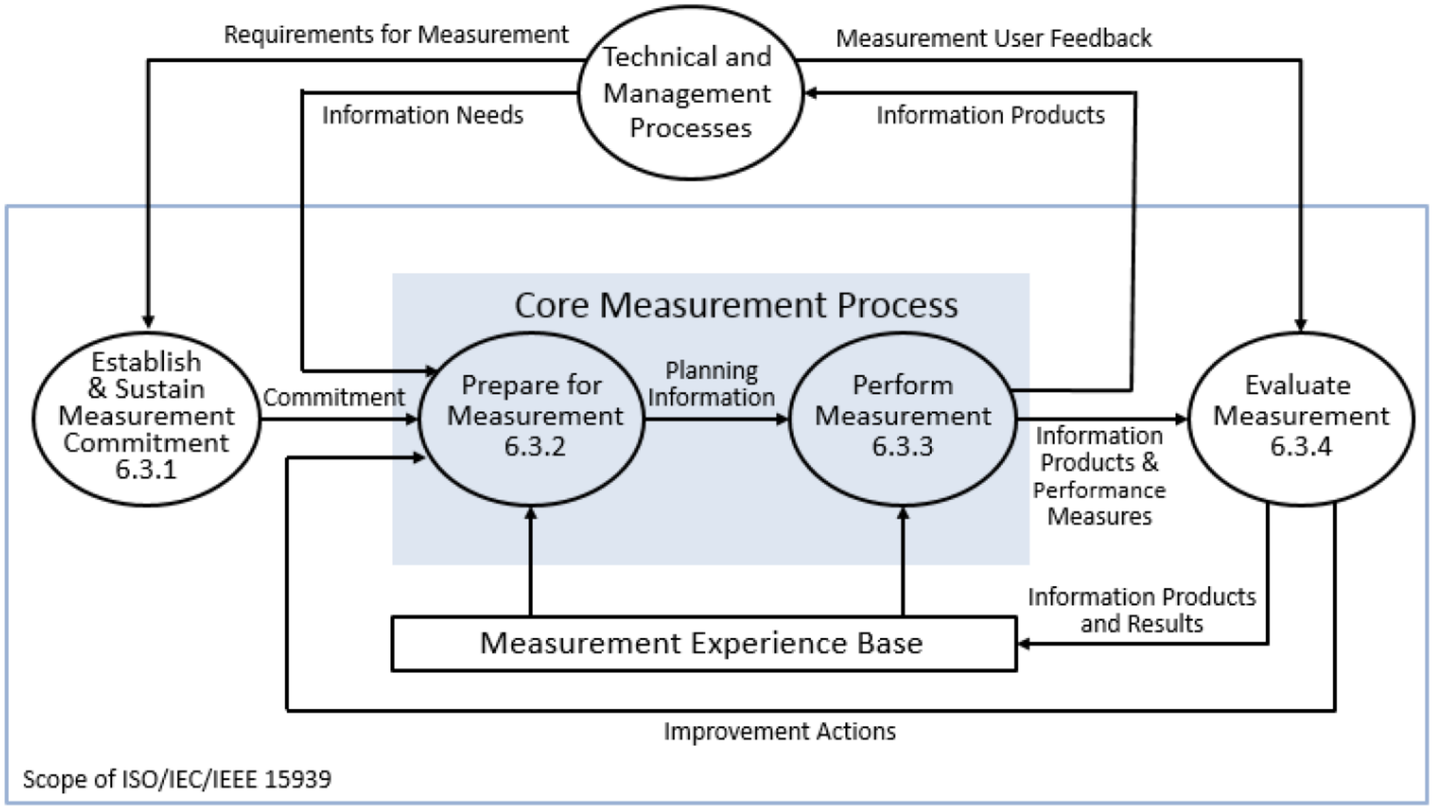
Limits:

- **Permanent:**
  - On the first generation of EPYC machines, the number of guests is **limited to 15**
  - **OS** needs to **support SEV**
- **Impermanent:**
  - **Live migration** and **suspension** of VMs
  - **PCI passthrough** to VMs
  - Boot disk limited to **virtio**

- **RQ3:** How is Performance affected when TEEs are implemented in a confidential HPCaaS?

- **RQ3:** How is Performance affected when TEEs are implemented in a confidential HPCaaS?

ISO/IEC/IEEE 15939:2017



- **RQ3:** How is Performance affected when TEEs are implemented in a confidential HPCaaS?

Benchmarks - MFLOPS Accounting

cpu			
	benchRIB	**cmet_eq**	**benchBFC**
M-Flops	925205908.230	14527316278.663	34223340.327

Table 5.1: **Partition cpu - Single Node - Mega-Flops Accounting**

cpu-sev			
	benchRIB	**cmet_eq**	**benchBFC**
M-Flops	925063918.719	14542858037.511	34287275.083

Table 5.2: **Partition cpu-sev - Single Node - Mega-Flops Accounting**

cpu			
	benchRIB	**cmet_eq**	**benchBFC**
M-Flops	938244867.570	15230958917.053	35912739.227

Table 5.5: **Partition cpu - Three-Node Cluster - Mega-Flops Accounting**

cpu-sev			
	benchRIB	**cmet_eq**	**benchBFC**
M-Flops	938197070.312	15228015556.021	35903490.594

Table 5.6: **Partition cpu-sev - Three-Node Cluster - Mega-Flops Accounting**

cpu			
	benchRIB	**cmet_eq**	**benchBFC**
M-Flops	1071630614.665	15299855815.791	58838419.978

Table 5.9: **Partition cpu - Small Cluster - Mega-Flops Accounting**

cpu-sev			
	benchRIB	**cmet_eq**	**benchBFC**
M-Flops	1104463557.104	19590361804.160	47120266.026

Table 5.10: **Partition cpu-sev - Small Cluster - Mega-Flops Accounting**

- **RQ3:** How is Performance affected when TEEs are implemented in a confidential HPCaaS?

Benchmarks - Time Accounting

cpu			
	benchRIB	**cmet_eq**	**benchBFC**
Wall t (s)	3208.421	53174.533	214.008
Core t (s)	19250.521	319047.194	1284.048
Effective t (mm:ss)	53:28	14h46:14	3:40

Table 5.3: **Partition cpu - Single Node - Time Accounting**

cpu-sev			
	benchRIB	**cmet_eq**	**benchBFC**
Wall t (s)	2697.893	60848.689	151.757
Core t (s)	16187.353	365092.133	910.543
Effective t (mm:ss)	44:57	16h54:08	2:31

Table 5.4: **Partition cpu-sev - Single Node - Time Accounting**

cpu			
	benchRIB	**cmet_eq**	**benchBFC**
Wall t (s)	2278.896	37748.322	106.490
Core t (s)	41020.113	679469.783	1916.798
Effective t (mm:ss)	37:58	10h29:08	1:46

Table 5.7: **Partition cpu - Three-Node Cluster - Time Accounting**

cpu-sev			
	benchRIB	**cmet_eq**	**benchBFC**
Wall t (s)	2300.233	37748.322	107.563
Core t (s)	41403.986	679469.783	1936.119
Effective t (mm:ss)	38:20	11h19:57	1:48

Table 5.8: **Partition cpu-sev - Three-Node Cluster - Time Accounting**

cpu			
	benchRIB	**cmet_eq**	**benchBFC**
Wall t (s)	1153.569	21768.311	86.232
Core t (s)	69212.669	1306098.487	5173.522
Effective t (mm:ss)	19:16	6h02:48	1:26

Table 5.11: **Partition cpu - Small Cluster - Time Accounting**

cpu-sev			
	benchRIB	**cmet_eq**	**benchBFC**
Wall t (s)	665.509	22587.554	87.012
Core t (s)	39919.275	1355252.592	5219.928
Effective t (mm:ss)	11:05	6h16:27	1:27

Table 5.12: **Partition cpu-sev - Small Cluster - Time Accounting**

- **Course of project**
  - **01.09.2023 - 15.01.2024**
- **Phase I**: Information gathering, familiarisation with the LRZ Compute Cloud (CC)

01.09.2023 - 01.10.2023

- **Phase II**: Writing the thesis

02.10.2023 - 08.12.2023

- **Phase III**: Writing the thesis, Working on DigiMed prototype system

09.12.2023 - 15.01.2024

- **Challenges**
  - **Phase I**: Information gathering, familiarisation with the LRZ Compute Cloud (CC)
    - Complexity and limits
    - ISO norm sources
    - Structure
  - **Phase II**: Writing the thesis
    - Set up the access to the prototype
    - Portability of seminar work
    - Set up failover strategies in case of failure in different layers
    - Programmierprojekt
  - **Phase III**: Writing the thesis, Working on DigiMed prototype system
    - Balance between external work and writing
    - Scheduling possibilities of correction
    - Adjustments regarding complexity and limits
    - Programmierprojekt

- **RQ1:** How does the security attestation of TEEs work?
  - Guest requests an attestation report and finally verifies it
  - Interactions are done via VirTEE tools
  - Security attestation was not available as AMD SEV was not available to its full extent, SNP features were necessary
- **RQ2:** How is Usability affected when TEEs are implemented in a confidential HPCaaS?
  - **Frontend:** Users need to choose the right image to deploy a VM with SEV enabled
  - **Backend:** SEV partition is limited to 15 guests per host, SEV prerequisites need to be fulfilled on the OpenStack flavour or image
- **RQ3:** How is Performance affected when TEEs are implemented in a confidential HPCaaS?
  - Performance of SEV partition is slightly lower than the one of the non-SEV partition
  - Occasionally even better
  - Figures seem satisfying, the number of nodes is limited to ten

- **Next steps:**
  - Full implementation of AMD SEV-SNP
  - Migration of real use cases
  - Capacities of up-scale HPC cluster

References

[1] ISO/IEC/IEEE 15939:2017 - Systems and software engineering - Measurement process. IEEE, 2017, ISBN: 9781504448512.

[2] B. S. Institution, EN ISO 9241-11:2018 - Ergonomics of human-system interaction (BS EN ISO). London: British Standards Institution, 2018, vol. 9241-11:2018, ISBN: 9780580893285.

[3] B. S. Institution, ISO/IEC 22123-1:2023 - Information technology - Cloud computing - Part 1: Vocabulary. London: British Standards Institution, 2023.

[4] B. S. Institution, ISO/IEC 22123-2:2023 - Information technology - Cloud computing - Part 2: Concepts. London: British Standards Institution, 2023.

[5] B. S. Institution, ISO/IEC 22123-3:2023 - Information technology - Cloud computing - Part 3: Reference architecture. London: British Standards Institution, 2023.

[6] T. Geppert, S. Deml, D. Sturzenegger, and N. Ebert, “Trusted Execution Environments: Applications and Organizational Challenges,” English, Frontiers in Computer Science, vol. 4, p. 78, 2022. doi: 10.3389/fcomp.2022.930741. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fcomp.2022.930741/full.

@FI_CODE

http://www.unibw.de/code

References

[7] GitHub. “VirTEE.” (2024), [Online]. Available: https://github.com/virtee/.

[8] OpenStack. “Open Source Cloud Computing Infrastructure - OpenStack.” (2023), [Online]. Available: https://www.openstack.org/.

Q & A

Valentin Pfeil
Institute for Software Technology
Research Institute CODE
University of the Bundeswehr Munich
valentin.pfeil@unibw.de
https://www.unibw.de/code

Forschungsinstitut
Cyber Defence
Universität der Bundeswehr München