Improving monolithic operating system kernel security and robustness through kernel subsystem isolation

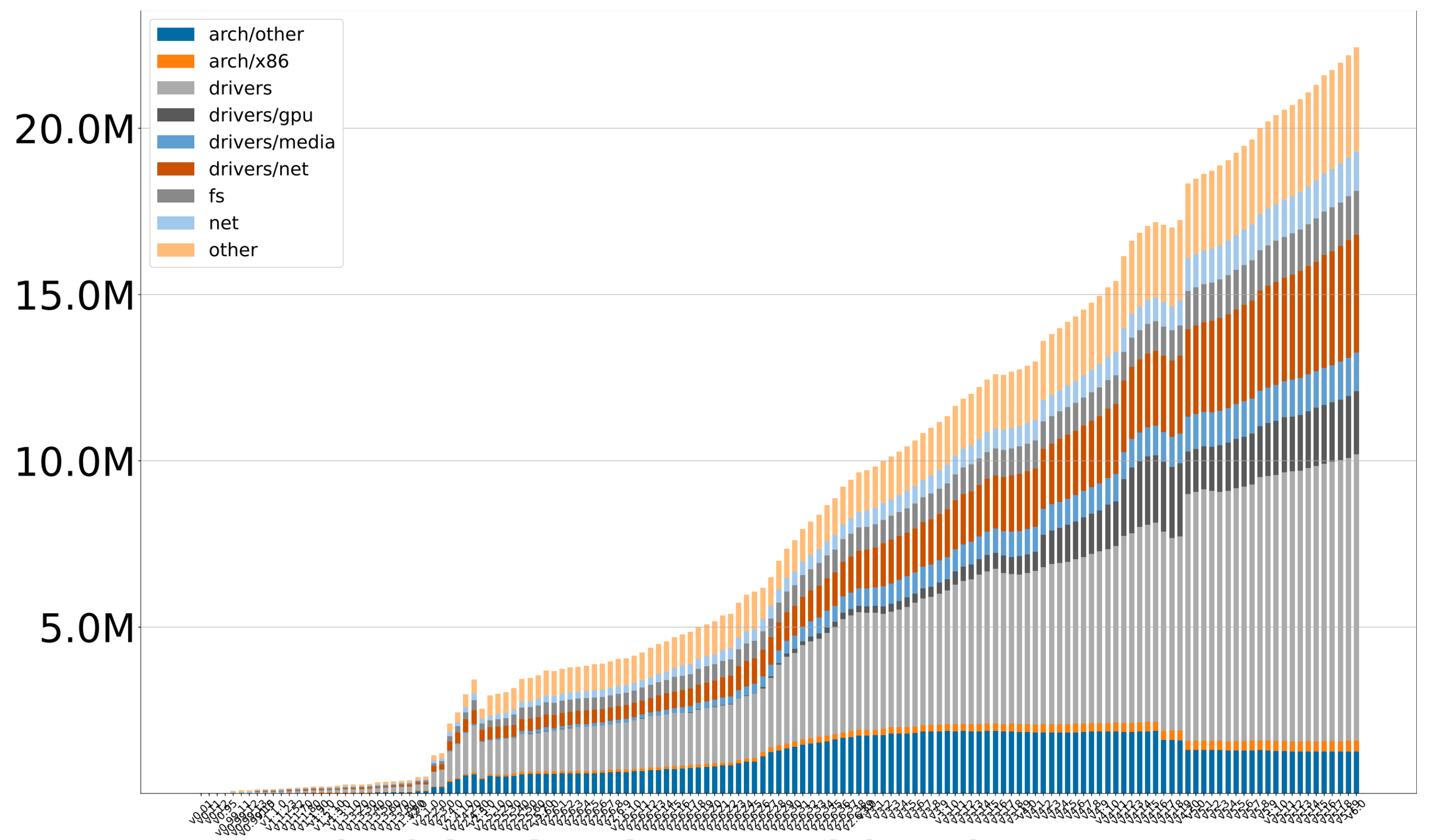
Bojan Novković, mag. ing. comp.

mentor: Prof. Marin Golub, PhD.
University of Zagreb Faculty of Electrical Engineering and Computing

SVEUČILIŠTE U ZAGREBU
Fakultet elektrotehnike i računarstva

1. Introduction

The structure of commodity operating systems kernels remains largely unchanged despite radical changes in underlying hardware and security risks. Millions of lines of code are executed at the highest level of privilege in a shared address space with no isolation between individual kernel subsystems, leading to serious security vulnerabilities and machine-crashing bugs.



Growth of the Linux kernel codebase in MLoC.

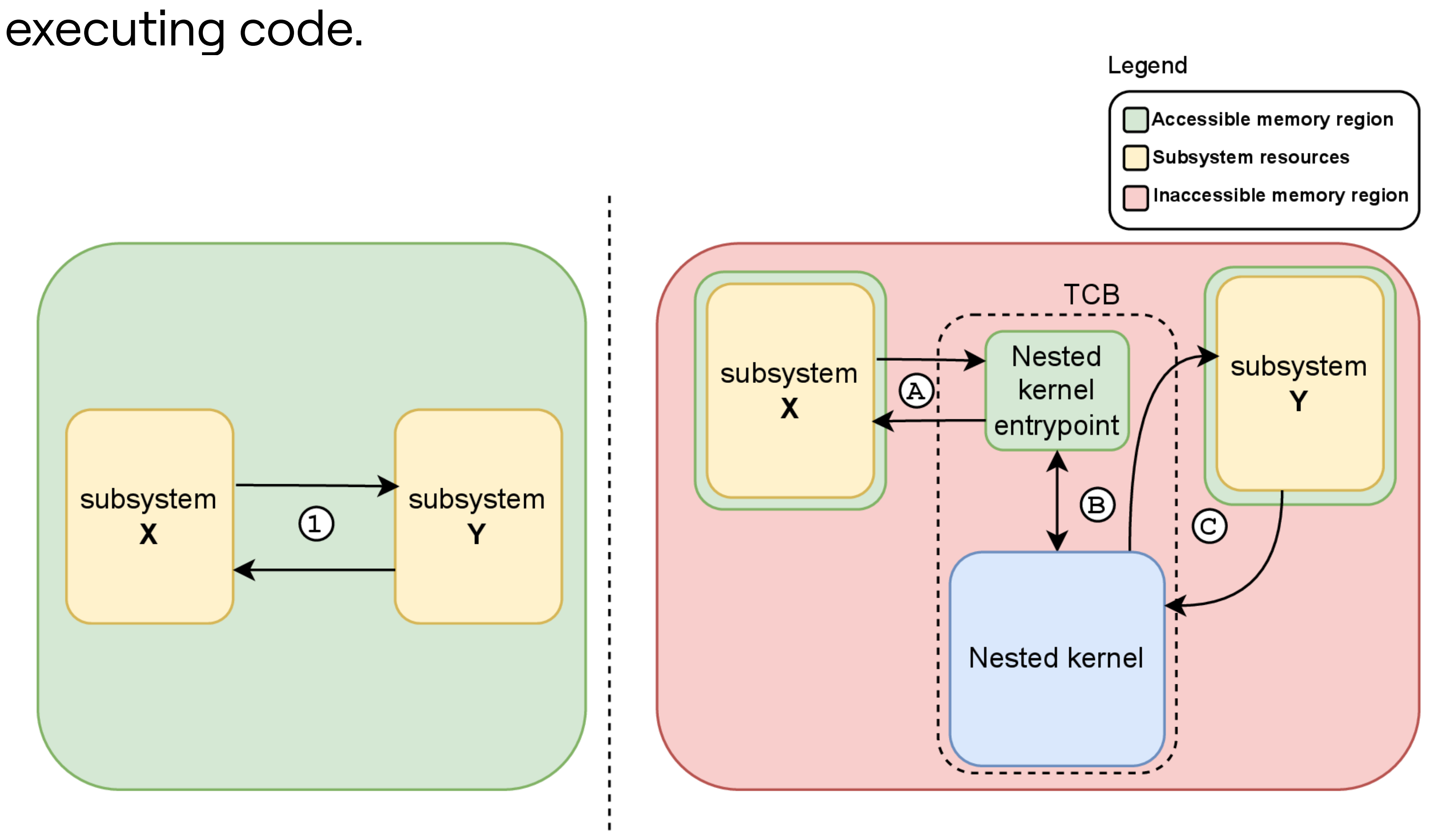
2. Background and motivation

The overall aim of this research is to design a novel, security-oriented monolithic kernel architecture. The research process consisted of six distinct phases:

1) Analysis of interactions and interdependence of kernel subsystems,
2) Analysis of runtime resources used by kernel subsystems,
3) Defining the basic *isolation unit*,
4) Designing a special, „nested" kernel used for enforcing isolation during runtime,
5) Designing a higher, software-backed level of privilege,
6) Designing a toolchain for specifying and integrating runtime subsystem isolation policies into the kernel.

3. Methodology

The crux of the proposed architecture is a form of intra-kernel *fault containment*, achieved using an SMP-capable nested kernel, a compiler pass that instruments all cross-execution context interactions, and a new, software-backed privilege level for executing code.

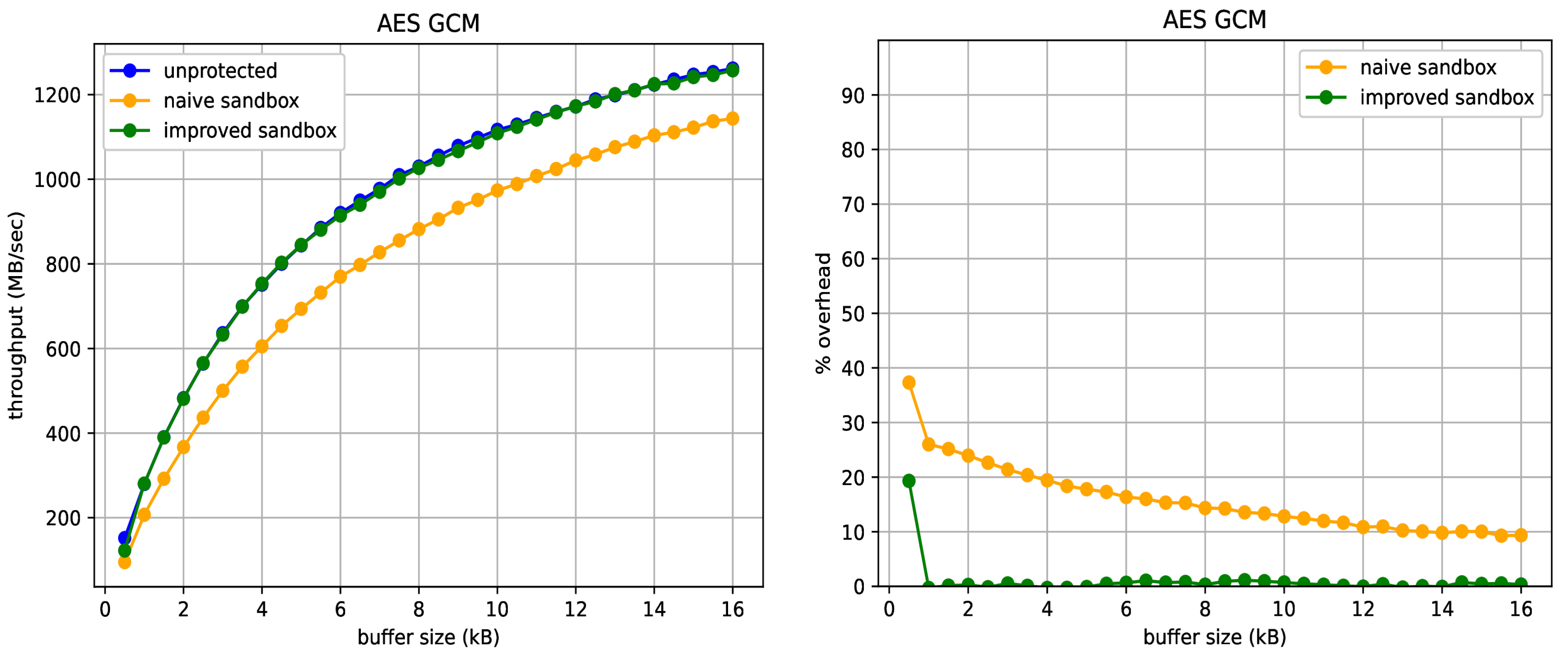


4. Results

A prototype of the proposed architecture was implemented using the **FreeBSD 14.0** operating system**.** The prototype was subjected to rigorous security and performance evaluations to test all aspects of the proposed architecture.

The performance of the prototype system was evaluated using three well-known benchmark suites – **PARSEC** [1], **BenchBase** [2], and **LMBench**. The isolation mechanism was evaluated by running throughput tests for various cryptographic algorithms provided by the **FreeBSD *cryptodev*** loadable kernel module which exposes kernel cryptographic facilities to userspace.

The architecture's security effectiveness was evaluated on a set of real-world vulnerabilities found in the FreeBSD kernel, which it successfully prevented.



Throughput and average overhead for the isolated *cryptodev* kernel module.

	wikipedia	ycsb	twitter
Baseline	29095.72	34327.71	32782.12
Prototype	28612.32	32963.23	31692.04
Overhead	**1.66%**	**3.97%**	**3.33%**

Averaged throughput (*requests per second*) for the *benchbase* database benchmarks.

The ***cryptodev*** results show that the throughput values of the isolated systems closely match those of the unsandboxed system,even dropping below 10% for larger buffer sizes. The database benchmark results show a modest performance overhead, showing that the proposed architecture has an acceptable influence on userspace programs.

5. Conclusion

This research presented a modified monolithic kernel architecture that enables intra-kernel fault containment and runtime kernel subsystem separation with support for SMP systems. It relies on a small, privileged nested kernel, a compiler plugin, and a separation policy framework to provide minimally invasive intra-kernel sandboxing.

The results of this research have been published as a research article in *"Computers & Security".*

Acknowledgment

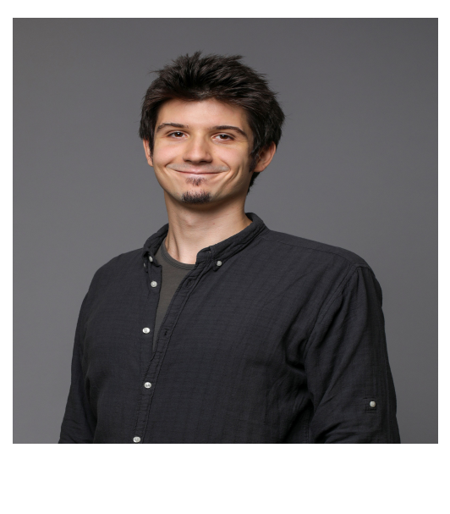
This work has been carried out within the project ***"Digital platform for ensuring data privacy and prevention of malicious manipulation of the personal data – AIPD2"***, funded by the *European Regional Development Fund* in the Republic of Croatia under the Operational Programme Competitiveness and Cohesion 2014-2020.

References

[1] Bienia, Christian, et al. "The PARSEC benchmark suite: Characterization and architectural implications." Proceedings of the 17th international conference on Parallel architectures and compilation techniques. 2008.

[2] Difallah, Djellel Eddine, et al. "Oltp-bench: An extensible testbed for benchmarking relational databases." Proceedings of the VLDB Endowment 7.4 (2013): 277-288.

Contact



Bojan Novković,
mag. ing. comp .
bojan.novkovic@fer.hr
+385917517801

PhD Day , Faculty of Electrical Engineering and Computing, 1 . June 2023