Anomaly Detection on Cloud Native Systems using eBPF probes

Felix Schäfer - s1147681 - Cyber Security and AI
Supervised by: Prof. Lejla Batina, Zhuoran Liu
Start and End Date: 01.02.2026 - 01.08.2026

Short Description

This project develops and validates a two component system that records predefined triggers with eBPF probes and applies an AI model to detect anomalous behavior in those recordings. The system will be deployed in a realistic cloud native environment and evaluated on real or carefully constructed datasets to measure detection quality and operational feasibility. The goal is not to reach state of the art accuracy, but to deliver a practical detector that is accurate enough to justify higher log levels and to provide forensic indicators for later human analysis. The expected outcome is a reproducible pipeline from probe design and data collection to model training, evaluation, and deployment guidance.

Research Questions

- Can we use eBPF probes to extract meaningful information from cloud native systems?
- Which AI methods are suitable for anomaly detection based on the collected data and constraints of online monitoring?
- What detection accuracy and false positive rate can the system achieve under realistic workloads?
- How does the trade off between model complexity and operational overhead affect practical deployment?

Methods

- **Data Collection:** Define probe points and collect a dataset of cloud native system logs and events via eBPF probes. Acquire data from an industry partner if possible, otherwise construct a realistic synthetic workload that includes normal and anomalous behavior.
- **Data Processing:** Normalize, enrich, and structure the collected data into features suitable for model training and evaluation, including temporal aggregation and context features where relevant.
- **Model Training:** Train one or more anomaly detection models and document the effect of feature choices, model family, and training strategy on detection quality.
- **Model Evaluation:** Evaluate the model using precision, recall, false positive rate, latency, and resource overhead. Compare baseline models to justify the chosen approach.

Literature Overview

- **eBPF:** eBPF is a technology that allows you to write code that runs in the kernel of the operating system. We will use the bcc library for interaction. [1]
- **Cloud Native Systems:** Cloud native systems are systems that are designed to be deployed in a cloud environment. We will be using a Kubernetes k3s cluster to deploy the system. [2]
- **Anomaly Detection Models:** There are many different models for anomaly detection. I will consider models like the following:
  - Anomaly detection in streaming data: A comparison and evaluation study [3].
  - MoniLog: An Automated Log-Based Anomaly Detection System for Cloud Computing Infrastructures [4].
  - BERT based natural language processing (for path and file name analysis) [5].
  - River model collection (for online anomaly detection) [6].

Rough Global Planning

- **Weeks 1-2:** Further literature review, refine research questions, define probe scope and success metrics.
- **Weeks 3-6:** Implement and validate eBPF probes, set up the k3s environment, and secure access to data sources.
- **Weeks 6-12:** Data collection, feature engineering, and baseline model experiments.
- **Weeks 12-17:** Model evaluation, error analysis, and iteration on probes or model design based on findings.
- **Weeks 17-24:** Write the thesis, integrate results, and finalize documentation and reproducibility materials.

References

[1] IOVisor Project, *Bcc - tools for bpf-based linux io analysis, networking, monitoring, and more*, https://github.com/iovisor/bcc, GitHub repository, 2024. [Online]. Available: https://github.com/iovisor/bcc

[2] K3s Project, *K3s - lightweight kubernetes*, https://k3s.io/, The certified Kubernetes distribution built for IoT & Edge computing, 2024. [Online]. Available: https://k3s.io/

[3] F. Iglesias Vázquez, A. Hartl, T. Zseby, and A. Zimek, "Anomaly detection in streaming data: A comparison and evaluation study," *Expert Systems with Applications*, vol. 233, p. 120 994, 2023, ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2023.120994 [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423014963

[4] A. Vervaet, "Monilog: An automated log-based anomaly detection system for cloud computing infrastructures," *arXiv preprint*, Oct. 2023. arXiv: 2304.11940 [cs.LG]. [Online]. Available: https://arxiv.org/pdf/2304.11940

[5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint*, Oct. 2018. arXiv: 1810.04805 [cs.CL]. [Online]. Available: https://arxiv.org/abs/1810.04805

[6] J. Montiel et al., "River: Machine learning for streaming data in python," 2021. [Online]. Available: https://www.jmlr.org/papers/volume22/20-1380/20-1380.pdf