
# Threat Detection Based on Log Analysis for Automating Security Information and Event Management (SIEM) Functionality

C.A. Hewagama
MS22909282
M.Sc. in CS

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2025

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

_____

Dr. Dharshana Kasthurirathna

Approved for MSc. Research Project:

_____

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

_____

Head of Graduate Studies, FoC, SLIIT

# DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: Amal

C.A. Hewagama

Date: 03.11.2025

# ABSTRACT

## Threat detection based on log analysis for automating Security information and event management (SIEM) functionality

Chathura Amal Hewagama

MSc. in Cybersecurity

**Supervisor:** Dr. Dharshana Kasthurirathna

December 2025

The reliance of modern organizations on information systems continues to increase, making these infrastructures frequent targets for malicious activity. System logs represent a primary source of forensic evidence, yet the volume generated by large-scale environments renders manual inspection infeasible. Security Information and Event Management (SIEM) platforms automate log collection and correlation but remain limited in detecting evolving or previously unseen threats. As a result, there is growing interest in augmenting SIEM functionality through Natural Language Processing (NLP) and machine learning.

This study investigates lightweight Transformer models as candidates for log-based anomaly detection in SIEM contexts. Two compressed architectures, DistilBERT and TinyBERT, are evaluated under parameter-efficient adaptation strategies: frozen encoders with linear classification heads and Low-Rank Adaptation (LoRA). Log templates are extracted using the Drain algorithm to normalize unstructured log data, and experiments are conducted on two benchmark datasets, BGL and HDFS. A classical baseline using TF-IDF with Logistic Regression is also included for comparison. Evaluation covers both detection metrics (precision, recall, F1-score, PR-AUC, ROC-AUC) and efficiency metrics (latency, throughput, memory usage).

The scope of this research is limited to training and evaluation rather than live SIEM deployment. Its contribution lies in assessing the trade-offs between detection accuracy and computational efficiency across lightweight adaptation strategies, providing guidance on configurations most viable for integration into real-time SIEM pipelines.

# ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Sri Lanka Institute of Information Technology for providing me with the opportunity and resources to pursue my studies. I am especially grateful to Dr. Anuradha Jayakody, Dr. Harinda Fernando, and Dr. Prasanna Haddela for their invaluable guidance and expertise throughout this journey.

A special thank you to my supervisor, Dr. Dharshana Kasthurirathna, for his continuous support, insightful feedback, and encouragement, which were crucial to the completion of this thesis.

Lastly, I want to extend my heartfelt thanks to my parents for their unwavering love, patience, and support, without which this accomplishment would not have been possible.

# TABLE OF CONTENTS

# List of Figures

# List of Tables