

LB/TH/41/2025

TH6000

**REAL TIME ANOMALY DETECTION FOR
CONTAINERIZED ENVIRONMENTS**

Nirothipan Megalingham

219375K

Degree of MSc in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

REAL TIME ANOMALY DETECTION FOR CONTAINERIZED ENVIRONMENTS

Nirothipan Megalingham
219375K

Thesis submitted in partial fulfillment of the requirements for the degree
Degree of MSc in Computer Science

Department of Computer Science and Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

June 2024

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 01/09/2024

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Degree of MSc in Computer Science Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. Sunimal Rathnayake

Signature of the Supervisor:

Date: 23/09/2024

ACKNOWLEDGEMENT

I extend my deepest and most sincere gratitude to Dr. Sunimal Rathnayake of the University of Moratuwa, Faculty of Computer Science and Engineering, for his invaluable mentorship during this research endeavor. Without his unwavering support, navigating through this journey would have been nearly insurmountable. Dr. Rathnayake not only provided continuous guidance but also mentored me during the thesis passing the time zone barriers, offering assistance remotely and helped me maintain the momentum during the challenging phases.

A heartfelt thank you is also due to my family for their enduring patience and encouragement throughout my research program. Their unwavering support has been a constant source of strength, from the inception of my studies to the culmination of this endeavor. Furthermore, I express my deepest appreciation to my parents for their steadfast support throughout my academic journey.

I am also indebted to the University of Moratuwa for granting me the opportunity to participate in the MSc program and for furnishing the requisite resources to see this research to fruition. Additionally, I am grateful to my former workplace, WSO2, for accommodating my pursuit of this part-time MSc and research alongside my professional responsibilities. Their support has been indispensable in achieving this milestone.

ABSTRACT

Anomalies in containerized environments pose a significant threat, given their potential to escalate small failures into catastrophic outcomes. These anomalies, which can manifest in various forms, possess the capability to disrupt service level agreements and tarnish an organization's reputation irreversibly. Thus, it becomes imperative to detect and address these anomalies promptly to minimize their adverse effects on business operations. In this study, we delve into the landscape of anomalies prevalent in containerized environments, focusing on understanding their diverse nature and the substantial impact they can have.

In this comprehensive survey, our focus lies in the real-time detection of anomalies within Kubernetes environments, a critical aspect in ensuring the robustness and stability of modern containerized systems. To achieve this, we conducted an extensive literature review, delving into existing research and methodologies pertinent to anomaly detection within Kubernetes ecosystems.

Furthermore, we conducted practical experiments by deploying applications on Azure Kubernetes Services, leveraging the inherent Kubernetes metrics API and Prometheus API to gather pertinent data. Employing sophisticated feature selection techniques, we curated datasets and trained a decision tree model capable of discerning anomalous patterns in real-time metrics streams. Through rigorous experimentation, we validated the efficacy of our approach, achieving a remarkable accuracy rate of 95% in anomaly prediction. This research underscores the significance of real-time anomaly detection in Kubernetes environments and offers tangible insights for enhancing the resilience of containerized infrastructures.

Keywords: Real Time Anomaly Detection, Machine Learning, Containerization, Kubernetes

TABLE OF CONTENTS

Declaration of the Candidate & Supervisor	i
Acknowledgement	ii
Abstract	iii
Table of Contents	iv
List of Figures	vii
List of Tables	viii
List of Abbreviations	viii
List of Appendices	x
1 Introduction	1
1.1 Terminologies	1
1.2 Overview	2
1.3 Background	2
1.4 Motivation	4
1.5 Research Problem	5
1.6 Research Questions	6
1.6.1 Questions	6
1.6.2 Rationale	8
1.6.3 Scope and Boundaries	8
1.6.4 Connection to Research Objectives	8
1.7 Research Objectives	8
2 Literature Review	11
2.1 Overview	11
2.2 Anomaly Detection with Machine Learning or Modeling Techniques	12
2.3 Anomaly Detection with Statistical Model	18
2.4 Anomaly Detection with Feedback System	21
2.5 Anomaly Detection by evaluating Infrastructure Utilization	21
2.6 Anomaly Detection using Rule Based System	22

2.7	Anomaly Detection using Forecasts	23
2.8	Real Time Anomaly Detection System	24
2.9	Summary	26
3	Approach and Methodology	28
3.1	Overview	28
3.2	Data Collection	29
3.3	Data Preparation	30
3.4	Model Training	30
3.5	Summary	31
4	Implementation	33
4.1	Overview	33
4.2	Deployment and Data Collection	33
4.2.1	Data Set Creation	33
4.2.2	Data Set Details	39
4.2.3	Infrastructure	39
4.3	Anomaly Prediction System	41
4.3.1	Data Normalization and Feature Selection	42
4.3.2	Detection Model Building	43
4.4	Real Time Anomaly Detection System	44
5	Evaluation	49
5.1	Overview	49
5.1.1	Control Setup: Order Management MicroService	49
5.1.2	Experimental Setup: Stock Monitoring System	49
5.1.3	Experimental Setup: User Activity Tracking System	50
5.2	Experiment with Order Management Micro Service	50
5.3	Validation	51
5.3.1	Ten-Fold Validation	51
5.3.2	Cross-Validation	51
5.3.3	Results	52
5.3.4	Discussion	52
5.4	Test Results	52

5.4.1	Experiment Details	52
5.4.2	Test Results	53
6	Conclusion and Future Work	57
6.1	Conclusion	57
6.2	Future Work	57
	References	59
	Appendix A Snapshot of Applications K8s	63
	Appendix B Anomaly Detection Resource Group in Azure	64
	Appendix C Sample Application in Docker	65
	Appendix D Azure Kubernetes Service	66

LIST OF FIGURES

Figure	Description	Page
Figure 1.1	Containers provide encapsulation environment for applications and dependencies	3
Figure 2.1	Overall Methodology Using LSTM by Rao[1]	13
Figure 2.2	Overall structure of TopoMAD[2]	14
Figure 2.3	Classified learning and anomaly detection of CDL.	17
Figure 2.4	The layered microservice-based architecture of the proposed solution.	17
Figure 2.5	Block diagram of cryptominer pod detection	19
Figure 2.6	Implementation of the Anomaly Detection System by Du X	20
Figure 2.7	Anomaly Detection workflow with Prometheus	23
Figure 2.8	Architecture overview of ContainerGuard	24
Figure 2.9	Resilient host-based intrusion detection system logic flow diagram and architecture	25
Figure 3.1	Overall Real Time Prediction Flow	28
Figure 3.2	Data Collection System	30
Figure 3.3	Data Preparation Workflow	31
Figure 3.4	Model Training Flow	31
Figure 4.1	Data Collection and Deployment Infrastructure	39
Figure 4.2	Web UI	45
Figure 4.3	Spring APIs	46
Figure 4.4	Python APIs	46
Figure 4.5	Python code for Flask application	47
Figure 4.6	Java Backend	47
Figure 4.7	Python Micro service	48
Figure 5.1	Real Time Anomaly Detection with Stock Price Monitoring System	56

LIST OF TABLES

Table	Description	Page
Table 4.1	Summary of Data Points for Normal Class and Anomalies	40
Table 4.2	Accuracy Levels of Different Prediction Models	44
Table 5.1	Validation Results for Anomaly Detection Model	52
Table 5.2	Anomaly Detection Results for Stock Price Monitoring System	54
Table 5.3	Anomaly Detection Results for User Activity Tracking System	55
Table 5.4	Confusion Matrix for Stock Price Monitoring System	55
Table 5.5	Confusion Matrix for User Activity Tracking System	55

LIST OF ABBREVIATIONS

Abbreviation	Description
AKS	Azure Kubernetes Service
CICD	Continuous Integration and Continuous Deployment
HHMM	Hierarchical Hidden Markov Models
K8s	Kubernetes
LSTM	Long short-term memory
MSA	Microservice Architecture
NFV	Network Function Virtualization
S3	Amazon Simple Storage Service
VM	Virtual Machines

LIST OF APPENDICES

Appendix	Description	Page
Appendix -A	Snapshot of Applications K8s	63
Appendix -B	Anomaly Detection Resource Group in Azure	64
Appendix -C	Sample Application in Docker	65
Appendix -D	Azure Kubernetes Service	66