INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

# TECHNICAL ANOMALY DETECTION SOLUTION FOR APACHE SERVER

A Project Proposal by

Mr. Pussella Kularathna (20200032/w1801027) only

Supervised by

Mr. Geethapriya Liyanage

Submitted in partial fulfillment of the requirements for the MSc in Advance software engineering degree at the University of Westminster.

**May 2022**

© The copyright for this project and all its associated products resides with Informatics Institute of Technology.

# Abstract

Today there are a lot of cyber attacks going on every day. Lot of resources, information, and financial frauds happen. One of them is directory traversal. What hackers do is randomly access directories via browser and download config files and steal data inside them. Anomaly detection is a key feature and most important aspect in many real-world applications, particularly for mobile platforms and the Internet of Things (IoT). Because of the growth of mobile devices and related app stores, the amount of new apps is rapidly increasing, necessitating a more effective method of analyzing all possible harmful actions. Anomaly directory traversal attacks have been always developing and growing. One approach to spot anomaly directory traversal is to look at the access logs and identify those anomaly folder accesses. These traffic patterns can be used to identify malicious users using machine learning. In this research an anomaly directory traversal detection solution using access logs with various access patterns and system performance stats will be investigated. Planning to construct Isolation Forest and train it based on access logs and system stats. Machine learning confronts two challenges: gathering a large enough training set of harmful and non-malicious data, and retraining the system as directory traversal evolves. This research will look at a method for overcoming these challenges by creating a detector that uses domains to train the system, which can then be used to analyze more detailed access logs using statistical and machine learning techniques.

Network traffic analysis focuses on extracting directory traversal communication patterns from HTTP access records (flows). Behavioral methods use characteristics extracted from access log fields to create a detector that can be used to any directory traversal that exhibits the desired behavior. System performance stats analysis on cpu, IO, memory, paging, number of child processes spanning etc.



**Keywords**: Directory traversal cyber attacks, Anomaly detection, Isolation forest, Machine learning, Apache server, Access logs.