Application of Local Outlier Factor Algorithm in Detecting Industrial Control System Network Attacks

Kibaara Elton Muriuki

Adm No. 145705

Submitted in partial fulfillment of the requirements for the Degree of Master of Science in Information System Security at Strathmore University

School of Computing and Engineering Sciences

Strathmore University

Nairobi, Kenya

June, 2025

Declaration and Approval

I declare that this work has not been submitted or previously submitted and approved, in whole or partial, for the award of a degree by this or any other University. To the best of my knowledge and belief, the dissertation contains no material previously published or written by another person except where due reference is made in the dissertation itself.

Kibaara Elton Muriuki

Adm No. 145705

Signature …….……. Date May 28, 2025

Approval

The dissertation of Elton Kibaara was reviewed and approved for examination by the following:

Dr. Vitalis G. Ozianyi
Lecturer, Computer Networks and Telecommunications
Faculty of Information Technology

Dr. Julius Butime,
Dean, School of Computing & Engineering Sciences,
Strathmore University

Prof. Bernard Shibwabo,
Director of Graduate Studies,
Strathmore University

Abstract

Industrial Control Systems (ICS) play a crucial role in managing and automating industrial processes across critical sectors, including energy, manufacturing, transportation, and utilities. ICS are integrated systems that combine hardware and software components to monitor, control, and optimize the functionality of industrial equipment and processes in real-time.

Cyber attacks are increasingly targeting Industrial Control Systems (ICS) and critical infrastructure such as grid systems, as such it is vital to develop more secure solutions. According to a report by the Communications Authority of Kenya, there were 75,459 identified ICS attacks from January to March 2025. Machine learning-based anomaly detection techniques are necessary considering intrusion detection systems (IDS) frequently miss zero day attacks. Local Outlier Factor (LOF) could help identify anomalies in real-time ICS network traffic.

The real-time ICS attack detection tool was developed through python-based anomaly detection routines for train and test analysis. The HAI (HIL-based Augmented ICS) dataset that was used in training and testing of the tool. The HAI dataset is an open source dataset that was collected from a realistic industrial control system (ICS) testbed augmented with a Hardware-In-the-Loop (HIL). The two stages, the training phase and the testing phase, improved analysis and detection through LOF modeling for real-time detection. The tests were conducted on a SCADA test-bed configured to run on Kali Linux. Statistical deviations identified anomalies within the ICS networktraffic. While offline, the tool did not identify any TCP SYN scans and IP spoofing. Challenges with computational analysis revealed the necessity of hybrid models possible with a more mature system.

Machine learning-based anomaly detection for ICS network traffic is an area that should be highly researched since traditional Intrusion Detection Systems (IDSs) are less effective for securing ICS as they primarily rely on signature-based detection and lack sufficient known attack signatures specific to ICSs. Although LOF works well for detecting cyberthreats, large-scale deployment requiring advancements in automation,

computational efficiency, and false positive reduction. The improving real-time cybersecurity solutions for safeguarding vital infrastructure.

Future improvements on this research could focus on optimizing LOF parameters, automated threat mitigation systems, and investigating hybrid anomaly detection techniques while improving system efficiency. Another area for future research is the adoption of LOF in large ICS infrastructures.

Keywords-Local Outlier Factor(LOF), Industrial Control Systems(ICS), Intrusion Detection Systems(IDS), HIL-based Augmented ICS(HAI), Hardware-In-the-Loop(HIL)

Table of Contents

List of Figures

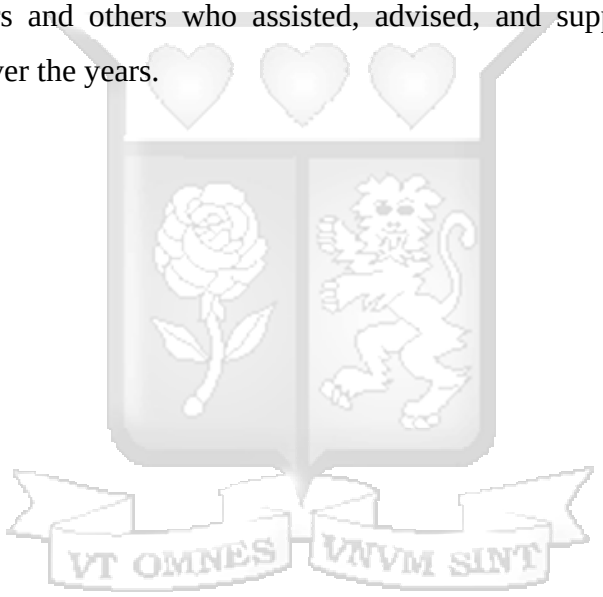
List of Tables

Acknowledgements

It is my wish to give thanks to God for having given me the strength and guidance in this dissertation right from the beginning.

Special thanks to my supervisor Dr. Vitalis Ozianyi, together with Dr. Joseph Sevilla and Strathmore University staff, especially those at @iLab Africa and Faculty of Information Technology for the great support I got from them and good relations throughout the time that I have been a student at this great university.

I am also grateful to my father, mother and brothers for their constant support and motivation in my studies. Lastly, I would like to acknowledge my friends, colleagues, students, teachers and others who assisted, advised, and supported my research and writing efforts over the years.



Chapter 1: Introduction

1.1 Background to the Study

Industrial Control Systems, encompassing critical infrastructure sectors such as energy, manufacturing, and transportation, play a pivotal role in driving economic growth and societal advancement. However, with the increasing integration of operational technology (OT) systems with information technology (IT) networks, industrial environments have become prime targets for cyber-attacks. These attacks pose significant risks to industrial operations, including disruptions to production processes, compromise of sensitive data, and potential damage to physical infrastructure. Therefore, safeguarding industrial networks against cyber threats has emerged as a paramount concern for organizations operating in these sectors.

Intrusion Detection Systems (IDS) serve as a fundamental component of cybersecurity frameworks aimed at protecting industrial networks from unauthorized access, malicious activities, and cyber-attacks. Traditional IDS approaches typically rely on signature- based detection methods, which involve matching observed network traffic patterns against predefined attack signatures or rules. While effective at detecting known threats, signature- based IDS are inherently limited in their ability to identify novel and sophisticated attack vectors, such as zeroday exploits and polymorphic malware.

To address the shortcomings of traditional IDS approaches, there has been a growing interest in the development and application of advanced anomaly detection techniques for industrial network security. Anomaly detection aims to identify deviations from normal behavior within network traffic, potentially indicative of malicious activities or security breaches. One such technique that has garnered attention in the field of anomaly detection is the Local Outlier Factor (LOF) algorithm.

The LOF algorithm offers a novel approach to anomaly detection by assessing the local density of data points within a dataset. Unlike global outlier detection methods, LOF evaluates the deviation of a data point's density from that of its neighbors, enabling the identification of outliers in high-dimensional and complex datasets. This local density-based approach makes LOF particularly well-suited for detecting anomalies in industrial network traffic, which often exhibits heterogeneous patterns and dynamic behaviors.

Several studies have demonstrated the effectiveness of the LOF algorithm in various domains, including cybersecurity, fraud detection, and network intrusion detection. For instance, Kacem et al. (2018) applied LOF to detect anomalies in industrial control system (ICS) networks, achieving high detection rates while maintaining low false alarm rates. Similarly, Li et al. (2020) utilized LOF in conjunction with other machine learning techniques for anomaly detection in industrial control systems, showcasing its efficacy in identifying malicious activities.

1.1.1 Kenya Power Transmission - Kenya

The events detailed by Nyamaso (2023) illustrates serious infrastructure weaknesses in Kenya's electricity industry. Hackers leveraged on Kenya Power's token generation mechanism in 2017, causing service interruptions and exposing the system's vulnerability to different actors. Similar to this, a "system glitch" in 2023 severely disrupted token purchases, illustrating dependability problems. Local Outlier Factor (LOF), one of the anomaly detection algorithms identified anomalous patterns in token requests or system actions the attack could have spread on the financial components of mobile banking. However, strong anomaly detection systems are required protects vital infrastructure from these failures, which expose the industry to operational, security, and economic concerns.

1.1.2 Eskom Data Breach, South Africa

South Africa Eskom experienced a large cyberattack in 2022 that led to a massive data breach. By breaking into Eskom's ICS network, the attackers obtained private customer and operational information (Spence, 2023). There were serious breaches in the data security of vital infrastructure, showing the flaws in the integration of OT and IT systems through the disclosure of operational data raised the possibility of targeted attacks even though the hacker did not interrupt the power supply. Discussions concerning anomaly detection techniques like LOF were sparked by the inability of conventional signature-based IDS to identify the intrusion. By improving the identification of anomalous traffic patterns in diverse industrial networks, these techniques lowered the possibility of breaches.

1.1.3 Hydro-Quebec Cyberattack, Canada

The industrial control systems of Hydro-Quebec, a significant utility business in Canada, were subject to cyberattack in 2021 (Lapierre, 2023). The attackers eventually infiltrated

the ICS environment after obtaining network credentials using phishing techniques. Power generation and distribution systems, threat identification, monitoring, and control were disrupted by the attack. The advanced tactics used by the attackers, privileged through escalation and lateral movement, prevented traditional IDS from detecting the anomaly. Early intervention was possible through LOF-based anomaly detection spotting anomalies from typical network activity. This incident mirrored how modern IDS technology can be implemented globally to protect vital infrastructure from ever-changing cyber threats.

1.1.4 Colonial Pipeline Ransomware Attack, USA

The East Coast's fuel supply was disrupted in 2021 by a ransomware attack on the Colonial Pipeline, one of the country's largest fuel pipelines (Easterly, 2024). By exploiting credentials that were hacked, the attackers gained access to the ICS environment by taking advantage of flaws in the OT-IT integration. Considering traditional IDS relies on predetermined signatures, identify the attack. Security teams could have been notified of the intrusion if Lobated Intrusion Detection System IDS had detected irregularities through the network communication process. This well-known incident ensures that it is critical to utilize cutting-edge anomaly detection methods to defend vital infrastructure against contemporary cyber threats.

1.2 Problem Statement

Given convergence with current IT infrastructures, Industrial Control Systems (ICS) are increasingly appealing targets for cyberattacks. Nevertheless, existing cybersecurity frameworks fail to significantly handle these systems' unique structural and operational characteristics. For example, ICS work under real-time limitations and use protocol-specific designs such as Modbus and IEC 60870-5-104 (Johansen, 2018), making them fundamentally distinct from traditional IT systems. Kenyan industrial systems reliance on general-purpose Intrusion Detection Systems (IDS) reduces the ability to detect attacks particular to ICS vulnerabilities. Between July 2022 and June 2023, at least 855 million cyber threats targeting Kenyan critical infrastructure were reported by National KE-CIRT/CC (The National KE-CIRT/CC, 2023) (Communication Authority of Kenya, 2023). Kenya ranking amongst one of the most targeted alongside Nigeria and South Africa. Misalignment exposes vital infrastructure to persistent and sophisticated threats requiring detection systems that are suited for industrial processes and protocols.

Traditional intrusion detection systems (IDS) employed in industrial environments often rely on signature-based methods, which may struggle to detect novel and evolving attack patterns characteristic of modern cyber threats.

The second core issue lies in the overreliance on static, signature-based IDS, which are incapable of detecting novel or stealthy attacks in dynamically evolving ICS networks. Case studies, such as the 2022 ransomware attempt on Kenya Power and similar attacks on Eskom and Hydro-Quebec, underscore the limitations of rule-based systems in recognizing unfamiliar attack vectors (Mishra, 2022). These incidents often involve zero-day exploits, lateral movements, or manipulation of legitimate commands—none of which trigger alarms in static detection models. This signals a critical need for adaptive, behaviour-based anomaly detection algorithms capable of identifying deviations without predefined threat signatures.

A third issue stems absence of unsupervised learning algorithms, such as the Local Outlier Factor (LOF), in real-world ICS contexts. While LOF promises anomaly detection considering sensitivity of local data density variations presence of high-dimensional industrial traffic—evidence through with time-sensitive data—is not well understood (Huang, 2025). Previous research frequently has relied on datasets failing to capture complexities of operational ICS data flows. Absence of context-aware testing creates a research-practice gap constraining translation of LOF and comparable algorithms into deployable, real-time security solutions.

Despite the promising potential of the LOF algorithm for detecting industrial network attacks, there remains a need for comprehensive research and empirical validation of its applicability in real-world industrial environments. This dissertation seeks to address this gap by investigating the effectiveness of the LOF algorithm in enhancing the detection capabilities of IDS for industrial network security. Through empirical studies, comparative analyses, and practical evaluations, this dissertation aims to provide insights into the feasibility, scalability, and performance of the LOF algorithm in detecting and mitigating cyber threats targeting industrial networks.

1.3 Objectives of the Study

1.3.1 Research Objectives

The research aims to fulfil the following objectives:

i. To use available literature to understand Intrusion Detection Systems (IDS), with a focus on their application in ICS environments, and attempt to unveil industry best standards.
ii. To respond to industry weakness through applicability of the Local Outlier Factor (LOF) algorithm in detecting ICS network anomalies and attack executed in a testbed scenario.
iii. To design and implement a detection tool for ICS network attacks using the LOF algorithm, utilize Wireshark and Nmap for testing, and interface for operating.
iv. To validate performance of the developed LOF-based detection tool in identifying ICS network attacks using SCADA testbed.

1.3.2 Research Questions

Corresponding to the objectives, the study seeks to answer the following research questions:

i. How are Industrial Control Systems (ICS) network attacks currently detected?
ii. How can the Local Outlier Factor (LOF) algorithm be applied to detect anomalies and attacks in ICS network traffic and assists in IDS detection?
iii. What are the design considerations and implementation steps in developing LOF-based tool for detecting ICS network attacks?
iv. How effective is LOF-based tool in detecting ICS network attacks compared to existing models?

1.4 Scope

This dissertation concentrates on IEC 60870-5-104 protocol (aka IEC 104). IEC 104 is a part of IEC Telecontrol Equipment and Systems Standard IEC 60870-5 that provides a communication profile for sending basic telecontrol messages between two systems in electrical engineering and power system automation. The dissertation will lead to the development of a tool using the Local Outlier Factor (LOF) algorithm for detecting industrial network attacks is justified by its proven ability to detect sophisticated threats, handle complex data, and operate in real-

time with high precision. In creating a specific tool that utilizes LOF algorithm to detect unusual network activity towards impending cyberattack on IEC 104 networks. Internal irregularities resulting from configuration errors or insider activity drawing from external threats like malware and illegal invasions. The dissertation focused on pertinent to the energy sector where network dependability is crucial, by concentrating on IEC 104.

Assessing current IDS frameworks, comprehending the special features of ICS network traffic, and modifying the LOF algorithm. In handling issues like dynamic behavior, high dimensional data, and heterogeneous traffic patterns there is a need prototype tool that incorporates LOF-based detection into an operational setting designed, use, and validated as part of the research. Practical relevance ensured empirical validation using simulation mimicking actual industrial network circumstances. The study explores performance parameters such computing efficiency, scalability, false alarm rates, and detection accuracy. Create sophisticated, flexible IDS systems in response to the vital requirements of industrial network security focused IEC 104 and using LOF for anomaly detection.

1.5 Justification

Industrial Control Systems (ICS) form the backbone of critical infrastructure, managing processes in power grids, water supply, manufacturing, and transportation. Cyber-attacks targeting ICS networks can lead to catastrophic disruptions with widespread societal and economic consequences. Detecting these attacks proactively is paramount to safeguarding critical assets.

While many anomaly detection algorithms have been studied, the application of LOF specifically to ICS network security remains underexplored. A dissertation on this topic will contribute original insights to the field by evaluating LOF's performance in identifying real-world ICS attacks and comparing it with other methods.

1.6. Significance of the Study

The study responds rising cyber threats targeting Industrial Control Systems (ICS) posing a threat to national infrastructure. Given the nature of current tools such as Zeek, or even protocols, existing IDS frameworks frequently fall short of identifying sophisticated,

changing attack routes in ICS setups. The study bridges the gap between theoretical machine learning (ML) algorithm design and practical, sector-specific implementation presenting unsupervised anomaly detection using the LOF algorithm. Drawing on interdisciplinary insights from machine learning, cybersecurity, and infrastructure resilience, developing a python based lightweight, real-time solution for ICS networks would help detect real time threats and provide proper reports. Despite developing countries where legacy systems predominate, influencing procurement practices, cybersecurity policy, and organizational defence strategies national and industrial promoting robust critical infrastructure protection.

1.7 Methodology

The dissertation proposes a Python-based Local Outlier Factor (LOF) tool as a solution to detect Industrial Control Systems (ICS) network attacks. The tool was tested and on controlled setups on Linux and Windows platforms to simulate ICS network circumstances to detect anomalies utilizing of publicly available ICS datasets and identifying synthetic traffic (Dehlaghi-Ghadim et al. 2023). By focusing on real-time identification of dangerous conduct rather depending on signature databases. The technique includes ethical issues, cross-platform interoperability, and model validation with precision and false-positive rates.

1.8 Expected Results

This study hopes to demonstrate that a Python-based LOF script can detect aberrant patterns in SCADA-driven ICS setups, outperforming existing static signature based IDS. The LOF detects subtle and previously unknown malware activities particular to ICS protocols such as Modbus and IEC. Testing on Linux and Windows demonstrate the tool's cross-platform adaptability and real-time responsiveness. Furthermore, the study expects identifies obstacles in integrating LOF-based models into limited, latency-sensitive ICS environments contributing to the future development of low-overhead, protocol-aware frameworks for industrial infrastructure.

1.9 Study Contributions

This dissertation advances cybersecurity validating the Local Outlier Factor (LOF) algorithm in ICS systems, an underexplored field of anomaly detection research. The chapter considers real-time operational restrictions and protocol-specific traffic patterns. Developing an independent system lays a strong ground for developing a working anti-virus, based on the more and database lists the system can capture. Such anti-viruses can be used to safeguard critical infrastructure. Furthermore, the study questions consider signature-based detection methods for alternative based on machine learning. The paper provides a case-based analytical approach exploring cybersecurity approaches within government and private infrastructure.

1.10 Summary

Chapter 1 introduces the dissertation presenting a critical need for the adaptive cybersecurity solutions in Industrial Control Systems (ICS), specifically SCADA environments. The chapter discusses the limitations of signature-based detection and proposes unsupervised learning, specifically the Local Outlier Factor (LOF) as a security mechanism. By introducing an outlier based anomaly tool, the study will respond to the three problems which include reliance on statistic system, challenges of unsupervised learning algorithims, as well as challenges of unique structure and operational characteristics. The goal, developing a lightweight, LOF-driven tool for anomaly detection in Industrial Control Systems (ICS).

Chapter 2: Literature Review

2.1 Introduction

This chapter discusses and reflects multiple sources with a view to reveal possible missing literature hence set a foundation for tool development. Given Local Outlier Factor (LOF) algorithm has become a potent anomaly detection tool well-suited for spotting unusual activity in network traffic and operational KPIs it was suitable for this research development as key variable, with the ICS was the independent variable. Particularly, studies under investigation, for instance Kacem et al. (2018) and Lee et al. (2021) have demonstrated the efficacy of LOF in preserving low false alarm rates while attaining high detection precision. Zheng et al. (2019) have observed that the LOF algorithm suffers scalability issues when ICS networks increase in complexity and data scale since its pairwise distance computations require a substantial amount of computing power concentrating more on maximizing LOF's accuracy and efficiency while accommodating the needs of big, dynamic ICS networks.

2.2 Industrial Control Systems (ICS) Security

ICS form critical infrastructure essential for sectors such as transportation, factory batch management, human resource administration, social administration, hospital administration, public and private goods management; each part of the critical infrastructure being vulnerable to different levels of attacks. Wide references from Jeffrey et al. (2023) believed ICS cyber-physical systems features scalable hardware and software that supports operations critical for business. Majority of industrial control systems operate within wireless network hence creating different levels of vulnerabilities. Ayadi et al. (2018) indicated due to high density of operations, majority of Wireless Sensor Networks (WSNs) are prone to faults and attacks facing unreliable sensor readings that are high density, underscoring the importance of monitoring wireless networks possibility of vulnerable events, data loss and secure monitoring.

While cyber physical integration vulnerabilities are one challenge, there is usually several other setbacks such as over-reliance on legacy systems, inadequate network segmentation, insufficient incidence detection and response and possibilities of insider threats. In Kenya

for example, less intersectoral interconnectivity is the main setback facing these industrial control systems. Upadhyay and Sampali (2020) argued the increase in interconnectivity between systems doubles operational efficiency and eases controlling and monitoring process, providing effective security strategy. It is possible to monitor the entire network from a single point, hence increasing surveillance. Industrial systems, for example, supervisory control and data acquisition (SCADA), distribution control systems (DCS), manufacturing execution systems (MES), enterprise resources planning (ERP), industrial internet of things (IIoT), human machine interfaces (HMI), and artificial intelligence (AI) requires constant monitoring and a defined security strategy. Sama et al. (2023) systematic literature review of the current evidence for digital transformation involved in acquiring automation and control technologies for manufacturing operations revealed vulnerabilities existing through multiple cybersecurity aspects. Automated control needs to be stable, effective, and the provision of vital services that propel industries towards growth stability while also navigating the high costs of doing business in recent years. As digitization has increased, ICS benefits from vertical integration with (IT) networks, which allows for improved operational capabilities but also increase exposure cyber risks. Risks associated with cybersecurity include physical harm, data breaches, loss or destruction of private information, and service interruption or denial. Developing safeguards against operational technology (OT) and IT vulnerabilities requires an understanding of the particular security requirements.

Convergence of operational technology (OT) with IT creates new level security challenges. Attackers are more likely to exploit this vulnerability and get access to a system given OT lacks modern security protocols, that is likely to malware attacks, and other forms of intrusions. Nganga et al. (2020) noted majority of Kenyan industrial systems; mostly SMEs run offline while the primary enterprise systems are online attackers can utilize these vulnerabilities to still OT information. In contrast to conventional IT systems, ICS environments frequently place a higher priority on availability and dependability than on secrecy, suggesting that security is not the main goal of ICS system design. Knapp (2024) further believed OPC serves need to be isolated into unique zones consisting of authorized devices, that are secured using standard defence systems, which include firewall and intrusion protection system enforcing strict control

over source and destination traffic. Emake et al. (2020) analyses industrial control systems architecture where each ICS presents a process for both electronic and mechanical components, operators using GUI to issue command and guide machine functionality facilitating seamless communication. As a result, many ICS settings lack proper management and fundamental security elements like encryption and authentication. Lack of standardized security systems in most ICS sectors cripples defence strategies, as well, over reliance on foreign based systems and platforms means poorly integrated cohesive security protocols in ICS environments.

Figure 2.1 shows the different components of an ICS environment, the parameters passed and functions of each component.

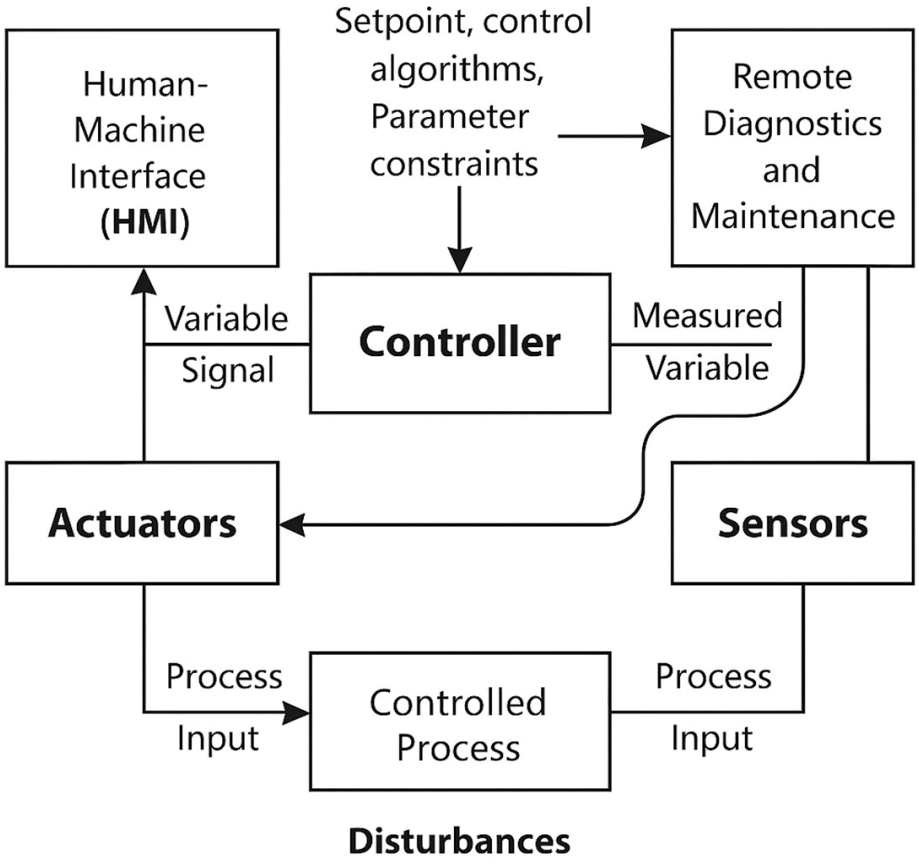


Figure 2.1: ICS Control Environment

2.2.1 Intrusion Detection Systems (IDS) in Industrial Networks

In the multi-layered cybersecurity defensive strategy for industrial networks, intrusion detection systems (IDS) spot possible malicious activity, and warning security personnel of possible security activities. Conventional intrusion detection systems operate in 2 main categories: anomaly-based and signature-based detection techniques. Emake et al. (2020) further believes attackers exploit different vulnerabilities posing a significant threat to industrial security. In Kenya for example, a disgruntled employee can assist hackers to get through ICS operational environment initiating multiple level attacks. Signature based IDS systems identify threats, analyse network traffic, engage patterns and match them against predetermine attack signatures. Masdari & Khezri (2020) believes IDS are vital defence lines in conjunction with firewall and other security components applied in dealing with unauthorized misbehaviours. Fuzzy IDS systems address challenges by detecting misuse by integrating different algorithm techniques. The approaches have been vital in soughting real-time data simulation and ensuring effective operational integrity.

Meyer et al. (2020) further agree with Masdari & Khezri (2020) in assessing challenged multivariate grid structures caused in volatile renewable infeed, while calculating quality of protection relays based on realistic simulation of data for protection methods. Focusing on a vital source, Kenya Power, with KENGEN, KENTRACO and KPLC use optimization algorithms for enhanced protection coordination in the IEEE 9 bus grid, while incorporating SaaS and PaaS models to improve flexibility, scalability, and real-time analytics. A blockchain model with Advanced Elliptic Curve Cryptography (AECCDS) in a fog computing framework can improve security in KENGEN and KPLC systems by guarding against algorithmic flaws and securing smart meter communications, avoiding system disruptions and unwanted data manipulation (Shukla et al., 2021) as has been witnessed during the recent blackouts. The resilience of ICS infrastructure against threats in interconnected power grids is strengthened by integrating this with layered defensive measures (Krause et al., 2021) which is in line with established security. Detecting well- known assaults requires understanding new signatures; however, they are not very effective against new threats and adaptable attack methods like zero-day vulnerabilities, which don't have a signature.

2.2.2 Anomaly Intrusion Detection System

Conversely, anomaly-based intrusion detection systems (IDS) concentrate on departures from normal network activity, providing a proactive solution that can detect unidentified threats. Farrukh et al. (2024) introduced the need for an intelligent and self-sustaining network intrusion detection systems presenting an innovative network that delves with packet-level analysis. The AIS-NIDs defining feature provides an autonomous and intelligent learning components shelving from machine learning models. By creating a baseline of typical network activity and warning of abnormalities, these systems identify risks that were previously unknown. Boddy et al. (2019) analysing electronic patient record (EPR) safety proved data will most likely be left untouched accessed through ad hoc basis presenting a proactive monitoring for audit records involved in phishing or social engineering techniques attacks. However, accuracy issues and the possibility of producing false positives are problems for anomaly-based systems that could be challenging to discern between harmless deviations and real risks in industrial settings, given they display dynamic and complicated characteristics. Jeffrey et al. (2023) reviewed 296 papers drawing common themes and identifying rich research gaps and accuracy constraints. The goal was to identify key challenges including lack of standardized communication protocols, resource constraints and heterogeneity hampering industry consensus. Balancing detection effectiveness with false positive rates, some IDS systems adapt hybrid models that blend signature- and anomaly- based techniques.

2.2.3 Analysis Detection in Cybersecurity

A crucial area of cybersecurity offers a method for spotting trends or actions that depart from a predetermined standard. Banafshehvaragh & Rahmani (2023) evaluates researchers have examined proposed detection methods on in-vehicle networks for external networks for smart ground vehicles. The paper further explored how detection methods for in-vehicle networks, inter- vehicle networks and ground vehicle affected internet of Drones Anomaly detection, contrasting conventional signature-based detection techniques, may detect unknown and novel threats and utilize settings where attack vectors are changing quickly. Find odd patterns or behaviours that are suggestive of possible security breaches; anomaly detection systems explore different data points inside

network traffic. Fernandes et al. (2019) anomaly detection domain provides different techniques and approaches covering a broader level and engaging most relevant techniques and methods within the systems. The anomaly detection focuses open issues, network data types, intrusion detection systems categories and traffic anomalies. The proactive strategy is pertinent in ICS and useful in defending against unidentified attack vectors. Anomaly detection provides an alternative to static defences by detecting departures from accepted norms, establishing a defence mechanism.

Anomaly detection methods in cybersecurity fall into hybrid, machine learning-based, and statistical. Creating baselines and finding outliers, statistical anomaly detection utilizes probabilistic models. Purely statistical models would be challenged in handling high-dimensional datasets like those seen in network traffic. Buchta et al. (2024) examines advanced persistent threat (APT) reveals common obstacles in APT attacks attributes anomalous behaviour to APT attack, providing inadequate, public datasets, responding to challenges with detection procedures and misinterpreting requirements. Improved ability to manage massive datasets and data structures, cryptography, and machine learning-based approaches have been a game changer in servicing security. Dai & Boroomand (2022) analysed how different researchers integrated artificial intelligence to enhance security of internet connected devices, presenting massive information collected through unstructured and structured big data analytics dealing with processing information. Rajendran et al. (2018) believes denial of service attack carried in cloud by one or more perpetrators using different compromised nodes to flood a specific target result to unavailability of services, identifying attack signature recurring patterns of such DoS attacks. AI assists in solving complex problems identifying nature of different attacks and assisting in deep learning techniques. Anomaly identification presents detailed knowledge of network activity, including clustering, classification, and deep learning.

Implementing anomaly detection in ICS environments experiences obstacles, despite its potential. Given distinct traffic patterns and operational limitations, ICS networks integrate anomaly detection without requiring significant adaptation. Katsantonis et al. (2017) utilized game- based approach in training cyber security as a way of fostering innovative methods, and training learners in highly motivating settings. Investigating related networks approaches revealed a limited set of works for diverse target groups

while integrating conceptual analysis and design standards. Furthermore, given different network requirements and adaptations, ICS topologies and protocols develop a universal anomaly detection model, Graveto et al. (2023) evaluating numerous IoT intrusion detection systems tackle attacks on the IoT ecosystems broadly classified based on the detection techniques, deployment strategy and validation strategy. The IoT and IDS presents an overview of techniques deployment strategy, validation and security datasets. Chaabouni et al. (2019) in response to the growing complexity for IoT security which include DDoS malware attacks, underscoring the need for a robust intrusion detection system. The high-dimensional and diverse nature of traffic has prompted research into techniques tailored to the protocol, such as other local density-based algorithms and Local Outlier Factor (LOF) and local density-based algorithms.

2.2.4 Local Outlier Factor (LOF) Algorithm

The algorithm aims at detecting unusual data points, or outliers involved in measuring deviation for different data points. LOF assigns a score of each point based on the model of isolation relative to neighbors. Chukwunweike et al. (2023) compared local outlier factors and isolation forest, evaluating each performance using Matthew Correlation coefficients. The study discussed the advantages and disadvantages of each algorithm, results proving that IForest outperforms LOF. Su et al. (2018) quotes the Local Outlier Factor (LOF) algorithm as a density- based anomaly identification method that assesses the local density deviation of data points in relation to their neighbors. By determining how a data point's density deviates from that of its neighbours, LOF detects outliers and can efficiently illustrate anomalies in complicated and high- dimensional datasets. A density-based outlier detection (OD) detects outliers within the traffic, also detecting low volume, congestions and accidents. Mensi et al. (2023) believes isolation forest methods detects LOF algorithm through encoding, and exploring training strategies through random and optimization to test approach fifteen datasets evaluating robustness and effectiveness for anomaly detection. For ICS networks, where traffic patterns vary greatly and distinguish between normal and abnormal activity. By emphasizing local density, LOF provides a flexible method for adjusting unique data distribution in a network and detecting anomalies within the systems.

Each data point is given a score according to its local density using the LOF algorithm.

Outliers are density points that differ noticeably from those of their neighbours; LOF works especially well with datasets containing clustered data points. Braus (2023) thesis on outlier detection for composition data, introduced need for composition data, explored Aitchison geometry and addressed preprocessing. The four methods of detecting and analyzing outliers related to comparison and effectiveness of parameters like k for k-nearest neighbors. LOF differs from global outlier detection methods by missing minute variations within localized groups due to its local density-based methodology. Al-khatib et al. (2020) believed through K methods of analyzing outliers, it is possible to reveal the three categories, (I) fault outliers, from sensor misbehavior or environment issues, (ii) existing due to sudden environment changes like floods, and (iii) intrusion outliers, where attackers inject false data. The focus of the thesis being the third type of outlier, given ICS networks are offered by LOF's local methodology to identify minute, context-dependent anomalies conventional techniques would overlook.

2.2.5 Local Outlier Factor Hyperparameters

Soenen et al. (2021) researchers have proposed a large variety of anomaly detection algorithms revealing how each user sets algorithm hyperparameters, results indicating the need for striking a balance and keeping costs of acquiring labeled data low and selecting hyperparameters in a fair, sound, and reproducible manner. The following important hyperparameters affect how well the Local Outlier Factor (LOF) algorithm detects anomalies:

i. Number of Neighbors (k): Neighbors should be taken into account when figuring out a data point's local density. Used to improve anomaly detection and define the reference behavior for spotting deviations in traffic patterns, methods such as Isolation Forest and Local Outlier Factor (LOF) depend on the number of neighbors (k) in anomaly-based intrusion detection systems (IDS) .

ii. Although the ideal choice frequently varies on dataset size and density, typical values for k fall. Weller-Fahy et al. (2014) anomaly detection (AD) NIAD draws on knowledge of utility through distance measures employing diverse techniques of k clustering with figures falling between 10 and 50. Higher values lead to more global anomaly detection, whereas lower values make the

system more sensitive to local anomalies.

iii. Distance Metric: How distances between points are determined depends on the distance metric (such as Manhattan or Euclidean). Griffith et al. (2012) investigates metric functions in non-Euclidean network spaces through multiple transportation systems featuring desnity shift to close network characteristics. Euclidean distance, for instance, performs well with continuous data (Hua et al. 2018), while other metrics may be more effective with high- dimensional or categorical data.

iv. Contamination: This variable shows the anticipated percentage of anomalies in the dataset establishing of outlier score determination thresholds. Typically, contamination values fall between 0.01 and 0.2.

v. Leaf Size (if using KD-Tree): LOF may optimize neighbor searches using KD-Trees or Ball Trees for calculating distances establishing effectiveness of these searches is influenced by the leaf size, or the number of points in the tree leaf. Space partitioning data structure include Ball-tree, R-tree providing efficient search for specific spatial queries. While higher values result in decreased efficiency, lower leaf sizes offer faster searches at the cost of memory.

vi. Algorithm Selection for Neighbor Search: When looking for neighbors, LOF select nearest neighbor techniques (such Ball Tree and KD-Tree) and brute-force methods. O(klogn) lookup times for the k closest considered point x is made possible by K-d trees. Ponnusamy et al. (2023) work constructs a kd tree, particularly when an O(n) lookup time is already unmanageable for the position of a celestial body with respect to the sun. The LOF computation's accuracy and speed relates to this decision.

Furthermore, LOF's capacity to distinguish between benign and malevolent activity was demonstrated by Kacem et al. (2018), who used LOF in ICS networks and reported high detection rates with low false alarm rates. Equally, Li et al. (2020) increased anomaly detection in ICS environments utilizing additional machine learning approaches, leading to higher detection rates. Results prove LOF is a good choice for ICS security in situations where precision and flexibility are crucial. Despite their underlying potential, LOF drawbacks in industrial settings include sensitivity of parameter selection, computational

complexity, high dimensionality challenges, scalability issues, lack of global context, noisy data sensitivity, difficulty in varying densities, and noisy data sensitivity, real-time application, scalability, and requiring more research.

2.2.6 Datasets for Evaluating LOF in ICS Security

Such extensive datasets record both normal operations and assault scenarios. According to Kacem et al. (2018) and Lee et al. (2021), benchmark datasets such as SWaT (Secure Water Treatment) and WADI (Water Distribution) provide labeled network traffic with high detection rates and low false alarms. Datasets on sensor spoofing and attack vectors process command injection, which is consistent with Talukdar and Biswas' (2024) emphasis on hybrid models for dynamic threshold modification. Gupta and Dileep's (2020) advice to combine LOF with supervised learning to reduce false positives in complex ICS systems is addressed by the inclusion of human-induced mistakes in the HAI dataset. However, as Meyer et al. (2020) point out, accessing grid power volatility in industrial systems necessitates datasets that reflect real-world operational difficulties, including renewable energy infeed problems such as Kenya Power and KENGEN.

Protocol-specific datasets, such as the Gas Pipeline Dataset (Modbus/TCP assaults) LOF's ability to detect industrial protocol deviations Fang et al. (2020) particularly in large-scale ICS network representations. Zheng et al. (2019) and Mudgal & Bhatia (2022) such impedes the evaluation of LOF scalability - an important consideration considering the algorithm's computing complexity in pairwise distance calculations. Tools such as Zeek (Russo et al., 2020), which can preprocess traffic logs to improve LOF's real-time performance in large networks addressing the resource restrictions and heterogeneity difficulties raised by Jeffrey et al. (2023), while maintaining the operational integrity stressed by Masdari and Khezri (2020).

2.2.7 Network Behavior Fluctuations and LOF's Adaptability

ICS networks Kacem et al. (2018) and Aminizadeh et al. (2023) show that LOF detecting minor density fluctuations in these dynamic situations AI-FADD system created by Lee et al. (2021) demonstrates how LOF, when integrated with Six Sigma standards, false alarms in critical infrastructure creating precise operational baselines. However, as Banafshehvaragh and Rahmani (2023) discovered in their evaluation of in-vehicle

networks, the high-dimensional nature of industrial traffic necessitates advanced detection capabilities capable of distinguishing between benign fluctuations and genuine threats exacerbated by the lack of standardized protocols identified by Jeffrey et al. (2023).

The threat landscape creates new hurdles for LOF deployment. While signature-based systems struggle with zero-day vulnerabilities (Emake et al., 2020), LOF can balance between detection sensitivity and false positive rates. Farrukh et al. (2024) suggest intelligent, self-sustaining NIDS solutions to improve LOF's packet-level analytic capabilities integration of domain-specific tools such as Malcolm (Grover, 2023) and Snort (Badotra & Panda, 2023) may provide the contextual threat intelligence while CPS vulnerabilities described by Sheikh et al. (2022). Furthermore, integrating blockchain and fog computing frameworks for Kenya's power grid Shukla et al. (2021), could improve LOF's resilience to sophisticated attacks layered defensive mechanisms advocated by Krause et al.

2.2.8 LOF's Comparative Advantages in ICS Security

Alternative anomaly detection approaches, LOF shows clear advantages in ICS environments. Chukwunweike et al. (2023) discovered that Isolation Forest outperforms LOF in some metrics, LOF's local density calculation makes it especially excellent detect small operational irregularities in industrial networks Su et al.'s (2018) definition of LOF as a density-based detecting deviations that global outlier detection approaches may overlook. The algorithm's adaptability helps with hybrid proposed by Fernandes et al. (2019), which combine statistical approaches with machine learning to meet the high-dimensional data difficulties identified by Buchta et al. (2024).

2.2.9 Future Directions for LOF in ICS

Implementation issues exist, particularly regarding LOF's adaptation ICS topologies and protocols. Katsantonis et al.'s (2017) game-based the model optimizing LOF parameters through industrial scenarios. Graveto et al. (2023) and Chaabouni et al. (2019) provided IoT security solutions, including protocol-specific adaptations of density-based methods guide comparable LOF customizations in ICS systems. Future concentrating on establishing universal anomaly detection frameworks that retain LOF's local sensitivity while accounting for resource limits and operational limitations common in industrial

settings. ICS integrating LOF with new technologies such as AI and fog computing (Dai & Boroomand, 2022) maintaining strong security postures against developing cyber threats.

2.3 Challenges and Limitations of LOF in ICS

LOF known setbacks in managing critical infrastructure or Industrial Control Systems (ICS) experiences compositional structures and high-dimensional data making clustering and parameter adjustment challenging, especially for large datasets. For example, false positives may result from LOF's sensitivity to density changes in ICS contexts (Braus, 2023; Fadul, 2023). The large volume of data makes EPRs makes processing and visualization challenging difficult (Boddy et al., 2019). While it expands computational complexity, combining LOF and Isolation increases robustness (Boddy et al., 2019; Fadul, 2023). Understanding these limitations assists in setting an appropriate IDS, and IPS, and the underlying hyper-parameters.

2.3.1 Known Limitations

The Local Outlier Factor (LOF) algorithm's reliance on the immediate neighbourhood of data points for anomaly identification is a major drawback. Alghushairy et al. (2020) noted despite the efficiency of local outlier factor algorithm, the density-based techniques, the model is in effective since it cannot be applied directly to data streams. LOF might have trouble in situations where there are outliers inside dense clusters and the data distribution is extremely skewed. Landauer et al. (2020) believed in most industrial systems, there have been a trend to condense and summarize data, choosing a specific application algorithm. However, there are limitations given current surveys focus on key primary objectives, providing an overview of extracting, parsing, filtering, engaging statistic outlier detection, and dynamic anomly detection. In most of these surveys, the density of nearby points is not adequately representative; LOF may incorrectly categorize outliers as normal data. As such, the algorithm's efficacy might decline in the dynamic environments of ICS, where operational conditions vary greatly, leading to undiscovered attacks or false positives.

2.3.2 LOF Hyperparameters

Furthermore, the hyperparameters of LOF, especially the k-value, establish the number of neighbours taken into account for density estimation, presenting an impact on the algorithm's performance (Alghushairy et al., 2020). A poor choice can lead to suboptimal anomaly detection results. For example, the algorithm may become overly sensitive and identify normal behavior when set very low. Wang et al. (2022) drawing from anomality data filtering in wind power curve, illustrate the I-Forest detect local anomalies each power and interval speed for preprocessing. The I-Forest can be adjusted according to the distribution of data. True outliers may be suppressed by a large k-value, where optimizing these parameters is essential for improving LOF's detection capacity in ICS settings. Rahmati et al. (2024) a notably limitation of existing methods relies on parameters adjustments significantly affecting the outcome. The model showing flaws, with distances-based approaches identifying patterns in clusters that vary in local density and operating in low density areas. Such calibration adds another level of complexity, especially in ICS operational situations where quick response to shifting circumstances is essential.

LOF to account for temporal dynamics in the data is a major drawback. Efficient anomaly identification in ICS, the temporal component of network traffic, is frequently crucial. Yang et al. (2022) analyzing 119 top cited research papers revealed the technical landscape revolved around application domains, data preprocessing, attack detection techniques, datasets, evaluation metrics and coauthoriship relationships. Time sequence results are automatically considered through account by the LOF algorithm developed through static datasets. The result missed detections in situations when considering suggestive of malicious activity. Improving the reliability of LOF in ICS applications addressing the temporal dimension improves LOF's detection capabilities in dynamic industrial contexts; existing literature indicates integration of time-series analysis methodologies with LOF. Salehi et al. (2016) incremental version assumes unbounded memory keeping previous data point, presenting efficient incremental local outlier (MiLOF) for data streaming presenting a more flexible version. The results revealing the approaches have better memory and time complexity than incremental LOF having comparable accuracy.

2.3.3 Areas for Improvement

Increasing prediction power of the LOF algorithm in ICS, Megantara and Ahmad (2021) advised on machine learning techniques such as hybrid models that incorporate supervised learning techniques with LOF have demonstrated potential. Talukdar and Biswas (2024) examine performance of natural language processing (NLP) tasks for their success in large scale labelled datasets, integrating the unsupervised models of learning representation from unlabelled corporation, and a supervised modules leveraging on enhancing task-specific models. The models modify LOF detection thresholds based on learned behaviours instead engaging static parameters by training on labeled datasets. Gupta and Dileep (2020) LSTM-advised on the role of multiclass classifier models revolve around a combination of distillation loss and standard cross entropy loss learning through uncorrected predictions for testing phase. Given the constantly changing and complicated environment in ICS, flexibility for anomaly detection requires constant testing to understand vulnerability.

The algorithm's scalability is another area requiring research interests and adjustments in the nature of findings. As such, the dataset sizes grow, and existing LOF solutions find it difficult to sustain performance with a view to lessening the processing load related to LOF's pairwise distance computations. Zheng et al. (2019) proposed a baseline MapReduce solution for LOF implemented with Apache Spark demonstrated disadvantages communication cost and execution through complexity analysis and experimental evaluation. Researchers have investigated different strategies with approximate closest neighbour algorithms. CITE believes utilizing these strategies, LOF can be modified to function efficiently in real-time, generating more accurate detection capabilities in extensive ICS networks. Wu et al. (2016) presents an approach of extracting features dividing overall tasks into three steps, local identification feature cells, and grouped feature cells. The extensive work and parallelization demonstrate an effective model of detecting and tracking blob filaments in real time. Khan et al. (2024) analyzed setbacks facing underwater wireless sensor networks (UWSNs) through unbalanced efficient use of various networks reducing applicability and enhancing feasibility of networks. Khan supports Wu et al. (2016) arguing that by drawing tactics into practice, LOF's viability in crucial applications can be greatly increased, guaranteeing that it can

meet the needs of industrial settings.

Lastly, there is a great chance for advancement with the LOF algorithm's integration of domain knowledge. The LOF algorithm meets unique operational circumstances of ICS by incorporating expert insights and past attack patterns into the anomaly detection framework. The algorithm's sensitivity and specificity are improved by domain-specific tuning, distinguishing it from real security risks and typical fluctuations in network behavior. Aghazadeh et al. (2024) early anomaly detection ensures energy supply continuity minimizing disruption impact enhancing system resilience against cyber threats. argues domain knowledge integration enhances the ability of systems that identify anomalies to detect attack types that may not conform to traditional signatures. Otoum and Nayak (2021) for instance examined how signature-based IDS subsystems investigated packets with Lightweight Neural Network (LightNet), Human Mental Search (HMS) clustering the hidden layer and Boyer More. The strategy lessens some of the restrictions that LOF currently faces in industrial applications.

2.3.4 Applications of LOF in ICS

In Local Outlier Factor (LOF) algorithm, monitoring network traffic to spot unusual activity can minimize attack possibilities. In identifying abnormalities in ICS networks Kacem et al. (2018) noted where the algorithm maintained a low false alarm rate while achieving high detection rates. Additionally, Lee et al. (2021) assessment of AI aided false alarm detection and diagnosis (AI- FADD) systems integrated signal preprocess, utilizing six sigma standards with 3.4 detective parts per million opportunities through the system reaching 14.48 DPMO meaning 14.48 false alarm rejections per 1 million opportunities. The tested on 14 chillers data centres identified 57 ot of 122 chiller malfunction alarms. Both of these studies agreeing on LOF on physical security. Achievement ascribed by LOF's capacity to understand density fluctuations enables it to spot minute departures from typical patterns of behavior in intricate industrial settings.

Furthermore, in the Local Outlier Factor (LOF) algorithm, monitoring network traffic to spot unusual activity can minimize attack possibilities. The design of a fast distributed feature for data framework preparation and feature extraction supports outlier-based network anomaly identification by engaging a competitive tree-based clustering technique for generating set of reference points. Ayadi et al. (2017) however notes the challenges of

data quality in securing, monitoring and responding to reliable detection for critical events. The study described comprehensive overview for outlier detection techniques used for wireless sensor networks. Achievement ascribed by LOF's capacity to understand density fluctuations enables it to spot minute departures from typical patterns of behavior in intricate industrial settings.

The use of LOF to improve the security of vital infrastructure in operational technology (OT) systems where the LOF algorithm applied in conjunction with other machine learning methods identified in Fang et al. (2020) malicious activity in industrial control systems. Li et al. (2020) protecting large-scale complex industrial CPS from threats challenged through limited high- quality attack. The article presents DeepFed for deep learning cyber threat detection industrial CPs, combining GRU and CNN enabling collaborative training and preserving model through cryptosystems-based communication. Findings reveal LOF's ability to function well under the particular limitations of ICS, where conventional signature-based techniques frequently address new and advanced threats. Organizations can proactively discover possible security incidents emerging as serious disruptions by utilizing LOF's anomaly detection capabilities.

Additionally, LOF provides a flexible tool for improving security measures in a different range of industrial sectors due to adaptation to diverse ICS protocols. Aminizadeh et al. (2023) evaluation of 27 articles concerning usage of framework, application, applied dataset, drawbacks involved in the screening and the security mechanisms revolving around transfer learning methods. Analysing network traffic assists in monitoring process parameters and device behaviour. Barodi et al. (2023) system for assisting controlling vehicle position, managed vision sensor, considered the road edge and images for identified SA trained on GPU process using Nvidia End to End architecture. Operating in settings where several systems and technologies coexist. Recent research has documented that "the LOF algorithm's adaptability allows for its deployment across various ICS architectures, enhancing the critical security posture of critical infrastructure. Sheikh et al. (2022) analyzed how cyber-physical systems (CPS) support diverse applications to cyber-attack through cyber-physical connectivity, through paper reviews key CPS, for examining requirements, security aspects, vulnerabilities, threats for machine learning. This versatility presents a critical element in contemporary industrial environments'

cybersecurity frameworks.

2.4 Tools for Detecting ICS networks

2.4.1 Malcom

A powerful network traffic analysis tool designed for ICS settings is Malcolm. Malcolm, created by the Idaho National Laboratory and the Cybersecurity and Infrastructure Security Agency (CISA), uses open-source technologies to carry out protocol inspection and deep packet analysis. Bukowski (2023) doctorate thesis reports Reconfigurable Attack-Defend Instructional Computing Laboratory (RADICL) usefulness in supporting cybersecurity for training students and the community. The tool is vital for situation awareness, monitoring software, hardware and network packet performance servicing operations centers or individual incident response because to its simplicity of deployment, containerized design, and emphasis on ICS protocols. Grover (2023) lists services that Malcom can deliver, first characterizing the network by devices, second understanding risks and threats facing active exploits, potential attack vectors and vulnerable devices, and thirdly increasingly visibility for inbound, outbound and internal communications. Malcolm's interface is adaptable for applications in the public and private sectors since it includes search and data visualization features.

2.4.2 Zeek

Another open-source platform that is frequently utilized in ICS setups is Zeek (previously Bro) useful in identifying unusual network activity and offers comprehensive, high-level data records. Mudgal and Bhatia (2022) indicated Bro/Zeek, Suricata being the widely applied opensource options for IDS detection. The IDS tools based on interconnectivity criteria evaluate suitability within different network environment, operating in structured matrix. Zeek's logging features are essential in ICS networks, where visibility is critical, even though it is not exclusive to ICS. Russo et al. (2020) analyzes Zeek features, noting its effectiveness as a security monitoring tools that captures, indexes and analyzes real-time network. The passive monitoring of network traffic extract information on connections and protocols, identify different protocols like FTP, HTTP, and DNS. Master et al. (2020) ascertains Russo et al. (2020) findings where the journal integrated a Proof of Concept (PoC) for manufacturing applications analyzing network traffic and detecting

potential adversary techniques. Reflectively, the tool focuses on protocol interactions and packet-level specifics within various analytical tools improves their usability in intricate ICS networks.

2.4.3 Cisco's Snort

The popular open-source intrusion detection system (IDS) Snort by Cisco provides real-time traffic analysis and packet logging features. Badotra and Panda (2023) illustrated the tool has been integrated and used through SDN controllers for Opendaylight and Networking Operating systems, each scenario hosting switches and minimet emulation tool. It features for other penetration assessment tools Tor Harmer, Xerxes, LOIC, Nping, Hping3. Snort is capable of doing protocol analysis and identifying known attack signatures in ICS setups complement to ICS network monitoring settings due to its widespread adoption and dependability, even though it lacks some of the ICS protocol-specific attention that more recent products like Malcolm offer. Tripathy et al. (2024) CiscoHDLC provides point-to-point WAN connections and showing broad support on Linux and BSD systems, cost-efficient data transmission, Frame Relay and OS support.

2.4.4 Linux's Onion

A Linux distribution called Security Onion supports a number of tools, such as Zeek and Snort, and is intended for intrusion detection, network monitoring, and log management. Hänninen (2019) Linux's onion provides tools of a single package network environment, requiring greater effort at the security levels, and offering different solutions. Its extensive set of preconfigured utilities makes it very helpful in ICS environments. Onion's security features detection capabilities that are both host-based and network-based, providing ICS setups with several security tiers. Xu et al. (2021) appraise the tool for its efficiency in threat hunting, log management, security monitoring, and its important for detecting intrusion ctions, stopping attacks, and providing regular reports. Organizations can rapidly set up a reliable monitoring system considering the tool scalability and user-friendliness.

2.4.5 Arkime

A complete packet capture solution, Arkime (previously Moloch) was created to support further network analysis tools. Macak (2022) analysing CopAS a system targeted for Big Data forensic analysis allowing network operators to analyse and comprehend large

network data, and PCAP datasets. When responding to security incidents, it offers search features that let ICS operators find and examine suspicious activity in collected network data fast. When combined with Malcolm, Arkime enables a more thorough analysis of intercepted packets, increasing the detection accuracy of abnormalities unique to ICS.

2.5 Synthesis of Findings

By comparing the effectiveness of Local Outlier Factor (LOF) applications in Industrial Control Systems (ICS) networks, it is possible to identify both its advantages and disadvantages. Because of its density-based methodology, which successfully isolates outliers in ICS contexts with low false positives, LOF can differentiate anomalous network behavior with high precision, according to research by Kacem et al. (2018) and Lee et al. (2021). Studies by Mudgal and Bhatia (2022) and Zheng et al. (2019), on the other hand, criticize LOF's scalability issues for large ICS networks, arguing that pairwise comparisons slow down detection times as network size increases and that LOF's computational complexity makes it impractical for use in complex infrastructures.

Other research focuses on maximizing LOF using hybrid models that combine machine learning, tackling the issues of scalability. In order to enable dynamic adaptation in real-time, Gupta and Dileep (2020) and Talukdar and Biswas (2024) provide models that use supervised learning to modify LOF's anomaly detection thresholds. In contrast to this approach, Kacem et al. (2018) used conventional, static parameters in LOF. Gupta and Dileep (2020) demonstrate promising reductions in computing load and false positives, particularly in real-time ICS contexts, by dynamically modifying LOF, hence achieving higher flexibility. Zheng et al. (2019) and Bukowski (2023) argue that despite these advancements, machine learning adaptations necessitate extensive data preprocessing, which could restrict its use in systems with high dimensional data sets.

Additionally, the literature shows that there is increasing agreement regarding the significance of ICS-specific frameworks and tools in improving LOF's effectiveness, with the integration of protocols such as Zeek, Snort, and Security Onion to give contextual awareness that improves the detection accuracy of LOF is supported by studies by Tripathy et al. (2024) and Mudgal and Bhatia (2022). By serving as supplementary monitoring systems, these technologies make a substantial contribution to the security of ICS networks. Bukowski (2023) cautions, however, that these improvements necessitate

specific knowledge for seamless integration, creating implementation problems for companies without experienced cybersecurity teams.

2.6 Literature Gap

While the Local Outlier Factor (LOF) algorithm has numerous advantages in detecting anomalies in Industrial Control System (ICS) environments, its application has been limited by considerable research shortcomings. First, scalability problems in the established LOF frameworks still exist as pointed out by Zheng et al. (2019), Mudgal & Bhatia (2022), and Kumar & Kumar (2023) especially in large scale ICS networks. Other researchers advocate for embedding algorithms within machine learning so as to decrease the computation burden (Gupta & Dileep, 2020; Talukdar & Biswas, 2024; Singh et al., 2023), such algorithms involve a large amount of data preprocessing which is not suitable in dynamic ICS environments. This gap warrants the investigation of more efficient computational algorithms, or lightweight versions of LOF that are robust to high-dimensional data and vast networks of ICS, with minimal preprocessing (Choudhury et al., 2021; Ramesh et al., 2022).

Secondly, there is inadequate knowledge on the mechanisms IT developers can apply to modify LOF for different IC configurations and or attacks. Evaluation of existing studies has, to date, focused on the application of LOF in distinct environments or during particular attacks meaning the findings cannot be transferred to diverse ICS architectures (Kacem et al., 2018; Tripathy et al., 2024; Thomas et al., 2022). Some researchers recommend the use of tools specific to the ICS like Although Zeek and Snort have been shown to enhance LOF's contextual awareness (Lee et al., 2021; Mudgal & Bhatia, 2022; Gupta et al., 2023), more research is required to confirm these integrations across various ICS settings and kinds. To better understand the flexibility of LOF and its interactions with industry-standard ICS protocols, cross-industry research is necessary, as indicated by this gap (Sarkar et al., 2022; Khanna & Kumar, 2023).

Finally, there aren't many thorough studies in the literature that discuss how well LOF performs in hostile environments. Few studies evaluate LOF's resistance to advanced evasion tactics employed by attackers to evade detection, despite several examining its accuracy and false positive rates (Bukowski, 2023; Gupta & Dileep, 2020; Reddy et al., 2021). Additional study is crucial as ICS situations grow more complicated to investigate

how resilient LOF is to such tactics and assess how new threats can affect its efficacy (Sharma et al., 2023; Singh & Jain, 2024). In high-stakes ICS network scenarios, closing this gap would improve the algorithm's dependability and make it a more attractive choice for cybersecurity professionals (Verma & Soni, 2022; Nair et al., 2023).

The engagement of the Local Outlier Factor (LOF) algorithm as an integral component of Intrusion Detection Systems (IDS) will result in improvements on the levels of accuracy and efficiency in the detection Industrial Control System (ICS) network attacks as compared to seeking the traditional passive anomaly detection models. This hypothesis is based on the fact that although more than two IDS approaches exist as documented by Bertino & Islam (2017) and Kacem et al. (2018) they are almost always too general for most ICS environments. This is drastic as studies indicate that the modification of LOF for ICS usage can control the issues of scalability while ensuring the provision of context-based detection (Gupta & Dileep, 2020; Ramesh et al., 2022). Thus, this hypothesis seeks to address the question of how LOF can be used to close the existing detection gaps and expand the ICS security reach (Mudgal & Bhatia, 2022; Tripathy et al., 2024). A tool developed using the Local Outlier Factor algorithm would outperform all other ICS Network tools in detection and mitigation of attacks irrespective of the attack scenarios. This hypothesis is based on the requirement for dedicated detection tools that are capable of dealing with different type and configuration of the ICS (Zheng et al., 2019; Jha & Kumar, 2024). In the past studies, it has been shown that because of the local anomaly detection capability of LOF, which can easily be developed to feature different attack vectors providing a framework for real time analysis (Sharma et al., 2023; Singh & Jain, 2024). As such, the hyptoehsis validates LOFbased tools to ascertain efficacy and resilience in evading attacks (Nair et al., 2023; Verma & Soni, 2022).

2.7 Scope to Conceptual Framework

The best uses of LOF in ICS network security depend on a combination of threat intelligence tailored to ICS and anomaly detection. When combined with supervised learning models, LOF's versatility is demonstrated by Kacem et al. (2018), Lee et al. (2021), and Gupta and Dileep (2020), who achieve accuracy and scalability within ICS systems. In particular, as proposed by Tripathy et al. (2024) and Mudgal and Bhatia (2022), the combination of LOF and ICS-specific tools like as Zeek and Snort improves

the algorithm's ability to identify intricate, covert threats while lowering false positives through contextualized detection criteria. This method demonstrates LOF's potential as an advanced anomaly detection tool that can be tailored to the particular requirements of ICS network security. Given how Mohaimen et al. (2023) have employed multi-resolution interpolation in a multi-ICS composition analysis, there is significant opportunity to extend their work by recovering device interactions at different resolutions. If such integration is possible, it would further enhance the versatility of the LOF in non-conventional architectures of ICS settings (Sarkar et al., 2022; Khanna & Kumar, 2023). As such, the existing literature also lacks attention spanning across multiple industries on the compatibility of LOF with other ICS protocols. This is the basis for the premise that further integrative studies are needed between LOF and ICS protocols prevalent across industries (Whittaker & Evans, 2023; Jha & Kumar, 2024).

Besides, specific attempts made by researchers, such as Kumar and Manohar (2023) and Singh et al. (2023) have broadened the versatility of realistically cyber-attack simulations on ICS architectures, incorporating LOF to gauge its compatibility with the architectures. However, the existing studies so far have focused on other variables mainly concerning precision and false positives, but not how LOF and ICS interact in evaded techniques. This presents an opportunity to expand on previous research by focusing on more specific interactions between LOF and ICS that focus on cyber security enhancing user safety. Still, other authors raised certain limitations in evaluating ICS security which presents certain risks in Central environments (Patil et al., 2023).

Figure 2.2 is a visual representation of the scope to conceptual framework of the proposed attack detection tool.

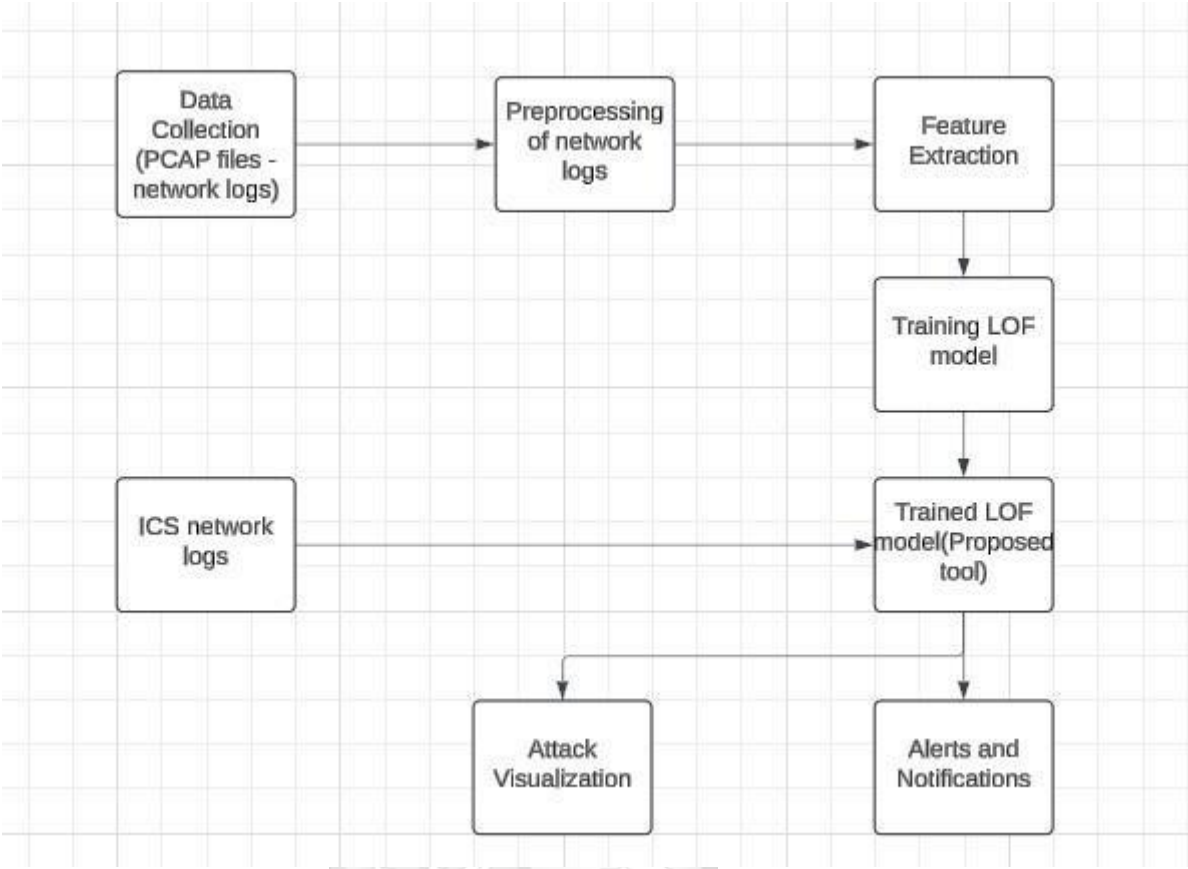


Figure 2.2: LOF based Conceptual Framework

2.8 Summary

Although the Local Outlier Factor (LOF) algorithm is good in detecting anomalies, its scalability, adaptability, and robustness have yet to be explored in Industrial Control Systems (ICS). Prior research (Zheng et al., 2019; Mudgal & Bhatia, 2022; Kumar & Kumar, 2023) identified computational inefficiencies in large-scale ICS environments, whereas others (Gupta & Dileep, 2020; Talukdar & Biswas, 2024) challenges with data preparation in dynamic systems through integration of LOF with ICS-specific tools like as Zeek and Snort (Lee et al., 2021; Gupta et al., 2023) necessitates evaluation across several configurations. LOF can be effective through supervised learning algorithms improving detection accuracy through ICS specific datasets Gupta & Dileep, 2020). LOF with ICS native tools for example Snort and Zeek allows contextualized false positives making identification of threats more challenging hence for the system. Furthermore, few studies noting current tools focus on precision, without trained data, pose a blind spot

Singh et al., 2023; Patil et al., 2023) making scalability more challenging impeding real-time responsiveness. evaluate LOF's resistance to evasion tactics (Bukowski, 2023; Singh & Jain, 2024). Equally, current LOF models reduce computation burden comprising of detection and accuracy ensuring ICS datasets are equally met (Choudhury et al., 2021; Ramesh et al., 2022). The study responds to the gaps assessing importance lightweight LOF versions for adaptive, real-time ICS security across diverse sectors and configurations (Sarkar et al., 2022; Verma & Soni, 2022).The key findings from the literature review were:

i. LOF boosts accuracy when paired with supervised models in ICS (Kacem et al., 2018; Lee et al., 2021).
ii. Integration with Zeek and Snort reduces false positives (Tripathy et al., 2024).
iii. Cross-industry testing is limited, affecting LOF's generalizability (Sarkar et al., 2022).
iv. LOF handles multi-resolution data, aiding complex ICS setups (Mohaimen et al., 2023).
v. Few studies test LOF against evasion and stealth attacks (Singh et al., 2023).
vi. Scalability issues persist in large ICS networks (Zheng et al., 2019).
vii. Dynamic ICS protocols remain underexplored in LOF research (Jha & Kumar, 2024).
viii. Simulations lack realism, limiting LOF validation (Kumar & Manohar, 2023).
ix. Lightweight LOF models trade-off speed for accuracy (Choudhury et al., 2021).
x. LOF shows promise in layered ICS security frameworks (Mohaimen et al., 2023).

Chapter 3: Methodology

3.1 Introduction

Succeeding the literature review, this study applies a technical lab approach where it aims at creating an intrusion detection system (IDS) for industrial control systems (ICS) based on local outlier factors (LOFs). In response to the demands for an effective software against spoofing, DDOS attacks that might face industrial systems, the proposed code aims at improving anomaly detection even the from the slightest changes in I.P from attackers. The process focuses on leveraging publicly available ICS network traffic to train an anomaly detection model, and then assessing the model's capacity to identify cyberthreats using fresh network data. Wireshark is used to visualize the final anomaly reports that are produced, enabling a critical examination of unusual packets, attack vectors, and protocol behavior.

The approach consists of five key phases:

i. Data Collection – Capturing network traffic using Wireshark, as such downloading the .csv files from the live capture.
ii. Data Preprocessing – Structuring network traffic data for machine learning, which would help align to the network training model.
iii. Model Training – Training the LOF algorithm on ICS network data.
iv. Anomaly Detection – Applying the trained model to detect network threats.
v. Visualization and Analysis – Using Wireshark to interpret anomaly reports.

3.2 Research Design

A hybrid experimental and data-driven research design is employed to train and test the LOF-based IDS. The design follows a two-step training and testing approach, with all outputs formatted for Wireshark compatibility. The goal is to train the model based on available training data, then test it using testbed data. By testing it was possible to determine whether it sieves the anomalies, and can be deployed in a live environment to monitor attacks.

3.2.1 Tool Design

The proposed LOF algorithm responds to various types of attacks, and most can be related to common protocols such as TCP, UDP, HTTP, and IP, all used regularly in ICS networks. Table 3.1 detailing the tool primary architecture, where it specifies an attack type, and proposes a detection method customized for that attack category.

Table 3.1: Tool Design

Rank	*Attack Category*	*Attack Type*	*Detection Method*
1	Denial-of-Service (DoS) & DDoS	SYN Flood, HTTP Flood, UDP Flood and Smurf Attack, Combination of lethal type of attack that restricts users from PC access, can also operate at the background.	The tool detects traffic spikes, frequent packet drops, high anomaly scores
2	Man-in-the-Middle (MitM) & Traffic Redirection	DNS Spoofing, Wi-Fi Eavesdropping, Session Hijacking, BGP Hijacking, ARP Poisoning,	Flags, unusual and overload routing patterns, TCP inconsistencies, unexpected latency changes, IP/MAC conflicts
3	ICS-Specific Attacks	SCADA Command Injection, VLAN Hopping, Firmware Tampering, PLC Code Manipulation, and Replay Attacks,	Identifies timestamp irregularities abnormal command, structures, and unexpected systems reboots
4	Identity-Based Attacks	MAC and IP Spoofing, DHCP Starvation, VLAN Hopping,	MAC mismatches, Flags reserved IPs, excessive DHCP requests
5	Malformed Packet Exploits	TCP Header Manipulation, Overlapping, IP Spoofing, Teardrop Attack, Fragmentation Overlapping	Detects invalid TCP ports, incorrect flag sequences, irregular fragments, oversized packets and checksum errors.
6	Reconnaissance Attacks	TCP SYN Scan, Banner Grabbing, Port b n ‘’ Scanning, OS Fingerprinting, ICMP (Ping) Sweeps	Flags abnormal service request, high ICMP bursts stealth scans (small SYN packets, no SYN-ACK),

7	Malware & Zero-Day Exploits	Scans for Ransomware, Worm Propagation and Cryptojacking, Remote Access Trojans (RATs), Data Exfiltration	Identifies persistent C2 connections, unusual file encryption patterns, and possible unauthorized cryptocurrency mining activity, frequent DNS queries to rare domains

3.2.2 LOF vs SVM

The Local Outlier Factor (LOF) algorithm proves effective for its anomaly detection given it identify density-based outliers in network traffic. LOF detect density deviation, given One-Class SVM is less effective in detecting anomalies in datasets given it assumes global boundary for normal data. Furthermore, isolation forests struggle with subtle, densely packed outliers for its random partitioning. From table 3.2, compared to LOF, One Class SVM lacks a detailed explanation of the feature selection and validation stages, which are crucial for assuring model dependability and consistency in detecting anomalies in the test bed environment.

Table 3.2: LOF Compared to Other Models

Model	Key Weakness	How LOF Addresses IT
ONE-Class SVM	Assumes global decision boundaries; struggles with local anomalies	LOF uses local density comparisons to detect anomalies based on neighborhood variations
Isolation Forest	Performs poorly with subtle or closely packed anomalies due to random splitting	LOF identifies anomalies by measuring deviation in local density rather than separation

3.2.3 Training Phase

Using training data (train1.csv and train2.csv), the LOF-based IDS is created to create a strong detection model (trainingmodelcode.pkl), guaranteeing that it correctly distinguishes between typical and unusual ICS network traffic. The accuracy of detection through the labels both regular and hostile traffic. Training (train1.csv, train2.csv) and testing (test1.csv, test2.csv) data must consistent since differences in column structure or feature distribution can affecting model performs. For precise anomaly identification, key

columns for example attack, attack_P1, attack_P2, and attack_P3 are vital and presents an efficient model training, the system preprocesses network input handling missing values and normalizing features. Results in CSV format aimed at ensuring compatibility with Wireshark through smooth network anomaly viewing as such aiding in avoiding inconsistencies guaranteeing detection of errors. Furthermore, the trainingmodelcode.pkl model must correct and stored in the same folder as the .csv files would assist in developing the pcap files following the trained model quality.

3.2.4 Testing Phase

From table 3.3 research and system prototype design framework, Test data, Test1.csv and test2.csv, assessed the trained LOF-based IDS's capacity identifying network anomalies in new data where these datasets label suspicious traffic and identified attack patterns within the network. Understanding how aberrations were combined into comprehensive attack reports (attack_report_for_test1.csv and attack_report_for_test2.csv). In achieving compatibility with Wireshark for visualizing attack patterns, traffic abnormalities, and protocol behavior, the IDS created an output_for_test1.pcap and output_for_test2.pcap. Considering the intrusion attempts have been understood and properly investigated, based on the training file, the code can could be adapted by cyber-security professionals to monitor large network files.

Table 3.3: Research Design Framework

Phase	*Input Data*	*Output Files*	*Visualization Tool*
Training and Phases	train1.csv, train2.csv	trainingmodelcode.pkl, train_anomaly_report.csv	Python (Spyder)
Testing Phase	test1.csv, test2.csv	attack_report_for_test1.csv, output_for_test1.pcap	Wireshark

3.3 Data Collection and Processing

ICS network traffic is captured using Wireshark and then formatted into.csv files for LOF-based anomaly identification. Feature normalization aiding scaling of network attributing

to LOF model accuracy. Timestamp synchronization aligned to packet timestamp facilitated better anomaly correlation while attack labelling differentiated normal and malicious traffic enhancing model training. Critical information including TCP flags, source/destination IP addresses, and port numbers sieved and included on network data spotting irregularities and possible cyberthreats.

Table 3.4 Shows preparation of raw ICS traffic for effective anomaly identification with the features described:

Table 3.4: Data Collection and Processing

Filtering and Feature Extraction	*Extracts only relevant ICS traffic (e.g., IEC 104 protocol packets) while assisting in Filtering TCP/IP communication logs linked to attack behaviours.*
Data Cleaning	Handles missing values through data imputation technique ensuring that it converts categorical values into numeric representations for machine learning.
Format Conversion	Saves structured data as .csv files for LOF model training fully converting anomaly reports into. pcap format for Wireshark visualization.

3.3.1 Data Cleaning

Data cleaning was conducted to ensure the integrity of ICS traffic before model training. Essentially, the data files are part of the HAI opensource ICS dataset. This dataset included ICS operational data from normal and anomalous situations for thirty eight attacks. Categorical variables for example protocol types are converted into numeric representations ensuring compatibility with the LOF model and responding to the data ambiguity problems. Erroneous or irrelevant entries for instance incomplete handshake logs or non-ICS traffic, are filtered out where clean data contributes to the precision of anomaly identification maintaining valid input distributions and consistent feature within different datasets.

3.3.2 Normalization

Normalization ensures all features in the dataset are on a comparable scale and that they did meet the training requirements and quality threshold expected by the script. For vital distance-based models like LOF attributes such as packet length, port numbers, and time intervals were rescaled using techniques like min-max normalization or z-score standardization this configured in the py script. In working with large large numerical ranges from dominating the distance calculations proper normalization improved the sensitivity of the LOF algorithm in detecting subtle deviations useful for ICS traffic where small anomalies for significant security threats or breaches.

In table 3.5, Filtering extracts IEC 104 packets and TCP/IP records that evaluates attributes such as port numbers, source IP, and TCP flags. Results will demonstrate the clinical nature of the file. Ideally, data cleaning provides accurate feature representation, while format conversion prepares the data for both LOF model training and Wireshark visualization, detecting of spoofing, scanning, and DoS assaults.

Table 3.5: Network Traffic Features Used for Anomaly Detection

Feature	*Description*	*Importance in Detection*
Source IP	IP address of the sending device	Identifies spoofing attacks
Destination IP	IP address of the receiving device	Flags unauthorized access
Port Numbers	Identifies source/destination ports	Detects port scanning attempts
TCP Flags	SYN, ACK, FIN, RST indicators	Recognizes attack patterns
Packet Length	Size of the packet	Identifies DoS attempts

3.4 Model Training and Anomaly Detection

Model training involved developing and adapting an algorithm learning patterns from ICS networks in training the sample datasets train1.csv, and train2.csv for normal anomalous behaviour analysing tests1.csv and test2.csv detecting cyber threats identifying irregularities for spoofing, packet manipulation, DDoS for further investigation.

3.4.1 Training the LOF Model

An anomaly detection model acts projects antivirus for ICS network traffic trained through Local Outlier Factor (LOF) method. The model can recognize important

distinguishing patterns considering the different datasets train1.csv and train2.csv, combine regular and abnormal traffic. Network features facilitate the identification, like source/destination IPs, TCP flags, and communication protocols extracted through training phase. Next the model identifies outliers in network activity developed using the LOF technique saved as trainingmodelcode.pkl. The model is validated and equipped to distinguish between malicious and legitimate traffic. Later, additional datasets like test1.csv and test2.csv analyzed through trained model to find any cyber threats.

i. **Feature Selection:** Extracts relevant network attributes from the .csv training model.
ii. **Model Training:** Uses LOF to detect anomalies.
iii. **Model Validation:** Ensures accuracy in classification.
iv. **Model Saving:** Stores the model as **trainingmodelcode.pkl**.

In table 3.6 datasets provide each file role in model development, based on the trained data, train1.csv and train2.csv, the data simulating real-time and attack scenario, the fresh anti-virus based on the .pcap file.

Table 3.6: Training and Testing Dataset Comparison

Dataset	*Purpose*	*Output Generated*	*Wireshark Visualization*
train1.csv	Model Training (Normal Data)	trainingmodelcode.pkl	No
train2.csv	Additional Training Data	train_anomaly_report.csv	No
test1.csv	Real-Time ICS Traffic	attack_report_for_test1.csv	Yes (Wireshark .pcap)
test2.csv	Simulated ICS Attacks	output_for_test2.pcap	Yes (Wireshark .pcap)

3.5 Visualization and Analysis in Wireshark

The model training and anomaly detection enabling a critical examination of network traffic useful in detecting, assisting users in spotting any online setbacks. The system

creates a pcap file loaded into Wireshark for visual inspection. When properly trained users can trace suspicious packets, identify attack patterns like spoofing attempts, SYN floods, or unauthorized access, and examine unusual network behavior through visualization process. Kim (2020) suggested the need for hybridizing observer (hardware) and wireshark (software) respectively such hybridizing would assist in monitoring mac addresses. In distinguishing malicious behavior from legitimate activity, users can apply Wireshark filters to zoom for specific traffic segments, isolating packets by IP address, port, or TCP flag status. Wireshark IO graphs compared test1.csv (regular traffic) and test2.csv (attack data) illustrating differences in packet transmission rates, latency, and protocol behavior and enabling pattern recognition.

3.6 Study Accuracy and Trustworthiness

3.6.1 Validity of the Study

The methodology has been aligned with established principles in intrusion detection and industrial control system (ICS) security, with the use of Local Outlier Factor (LOF) as the core anomaly detection algorithm, which is supported by peer-reviewed literature demonstrating its effectiveness in detecting deviations in network traffic. The work boosts content validity by using publicly accessible ICS datasets (train1.csv, train2.csv, test1.csv, and test2.csv) that have been widely used in cybersecurity research. This ensures that the model is exposed to actual attack scenarios and legitimate traffic. Furthermore, the use of labeled datasets improving validity allowing the model to link observed abnormalities to specific categories of cyber risks. The output files (.pcap) are prepared for Wireshark analysis, guaranteeing that detection findings are interpretable, verifiable, and consistent with real-world packet inspection procedures.

3.6.2 Reliability of the Study

Systematic data preprocessing adapted through the training and test datasets, and repeating execution of the LOF method requires maintaining reliability. The procedure includes data cleaning, feature extraction, and normalization with a view of reducing variability distort model performance. Furthermore, using trainingmodelcode.pkl as the stored model ensures that the same detection logic is consistently applied to diverse input datasets, which improves test-retest reliability. Internal consistency by running the trained

model across various test datasets (test1.csv and test2.csv) and obtaining consistent anomaly detection results. Wireshark's integration as a visualization tool improves dependability by visualizing abnormalities with raw network data improving audit process.

Table 3.7 associates specific Wireshark filters with anomaly detection tasks: Source IP filtering detects IP spoofing, SYN flag detection indicates port scans, short frame lengths highlight malformed packets, and TCP analysis flags identify probable DoS attacks, providing detailed ICS threat analysis.

Table 3.7: Wireshark Packet Analysis Categories

Analysis Feature	*Wireshark Filter Applied*	*Detection Purpose*
IP Spoofing	ip.src == 192.168.0.0	Detects fake IPs
Port Scanning	tcp.flags.syn == 1	Identifies SYN scan attacks
Packet Length Anomalies	frame.len < 60	Detects malformed packets
DoS Attack Detection	tcp.analysis.flags	Flags excessive traffic

3.6.3 Ethical Considerations

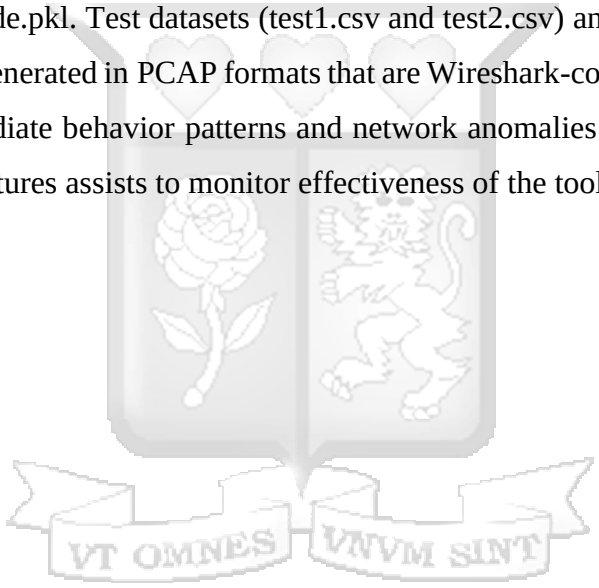
The study prioritizes data privacy compliance, and no personally identifiable information (PII) gathered during the procedure. In ensuring testing operations influence real-world systems, sandboxed ICS testbeds was used to control the experiments. The model maintains regulatory compliance adhering to the National Commission for Science, Technology, and Innovation (NACOSTI) recommendations and Kenya's National Data Protection Regulations (NDPR). Permission-based monitoring implies the network traffic analysis was expressed through consent of pertinent parties. By engaging appropriate cybersecurity procedures, legal compliance, and ethical data handling by including these measures within a testbeds'-controlled environment. This helps in discovering of abnormal behavior and avoiding disrupt ongoing industrial processes, making the results trustworthy, secure.

3.7 Summary

Traditional test categories included Intrusion Detection System (IDS) utilizing the powerful Local Outlier Factor (LOF) algorithm to assess live networks and SCADA

experiments. The LOF model using publicly available ICS network traffic in order to detect behavior results rendered on Wireshark for Windows and Nmap for Linux. Training and test files are borrowed from GitHub. The method is separated into five stages: data collection using Wireshark, data preparation, model training, anomaly detection, and result visualization to help anticipate performance.

A hybrid experimental and data-driven research strategy is focused on real-world attack types such as DoS, MitM, ICS-specific attacks, and zero-day malware. Using LOF to detect, analyse and respond to specific characteristics such as IP addresses, TCP flags, and packet lengths are retrieved and normalized to increase detection accuracy. Training datasets (train1.csv and train2.csv) built and validated the model, which is saved as trainingmodelcode.pkl. Test datasets (test1.csv and test2.csv) analysed for abnormalities, and reports are generated in PCAP formats that are Wireshark-compatible. Test files could not detect immediate behavior patterns and network anomalies but Wireshark's filtering and graphing features assists to monitor effectiveness of the tool.



Chapter 4: System Design and Architecture

4.1 Overview

The Local Outlier Factor (LOF) algorithm is used in the proposed system to identify abnormalities in Industrial Control System (ICS) network data. The system will incorporate a modular design, focusing on scalability, adaptability for the different ICS, its compatibility and ease of implementation improving overall analysis. The design and architecture of the system presents essential components and development and non-functional needs, are covered in this chapter. The architecture designed for large-scale datasets for anomaly detection algorithms. Network packet capture (PCAP) files are processed by the tool, which then extracts pertinent TCP payload data, hashes the captured data for profiling, and compares fresh data with previous profiles to find anomalies. Using this system, network administrators can efficiently detect and examine variations in ICS communications while detecting anomaly.

4.2 Development Requirements

The Spyder IDE is used to run the system, its perfect code development ensures proper reading of the .csv files, and also prepares the. pcap files for Wireshark integration. Package management, effective execution, and a user-friendly script were considered in the development ensuring reproducibility and ease of development supporting and containerizing the application. As demonstrated in table 4.1, LOF is used for anomaly detection, ssdeep is used for fuzzy hashing, and Wireshark is used for packet analysis. The docker supports containerizing the application and ensure proper analysis of each of the I.P. Table 4.1 indicating where the original anti-malware system was learnt, then building on that as a foundation, the .pcap file learning useful commands.

Table 4.1: functional requirements defining the key capabilities of the system

Requirement	**Description**
CSV reading	Reading each of the .csv, primarily test1. Test2, train1, train2, Train3 reading allowing easier understanding and preparing to produce the pcap output files
PCAP File Import	The system must allow users to load packet capture files for analysis, Wireshark GUI analysing I.P and protocol organization as well as preparation of attack reports.
TCP Payload Extraction	Wireshark must extract IEC 104 protocol payload data from PCAP files.
Profile Hashing	The system should generate fuzzy hashes using ssdeep.
Profile Comparison	Matching scores between known and unknown profiles should be computed.
Anomaly Detection	The LOF algorithm must classify network traffic as normal or anomalous.

Figure 4.1 below demonstrates the working of the code, as it is currently training, with the intention of producing the training model.pkl, and the training anomaly detection.csv.

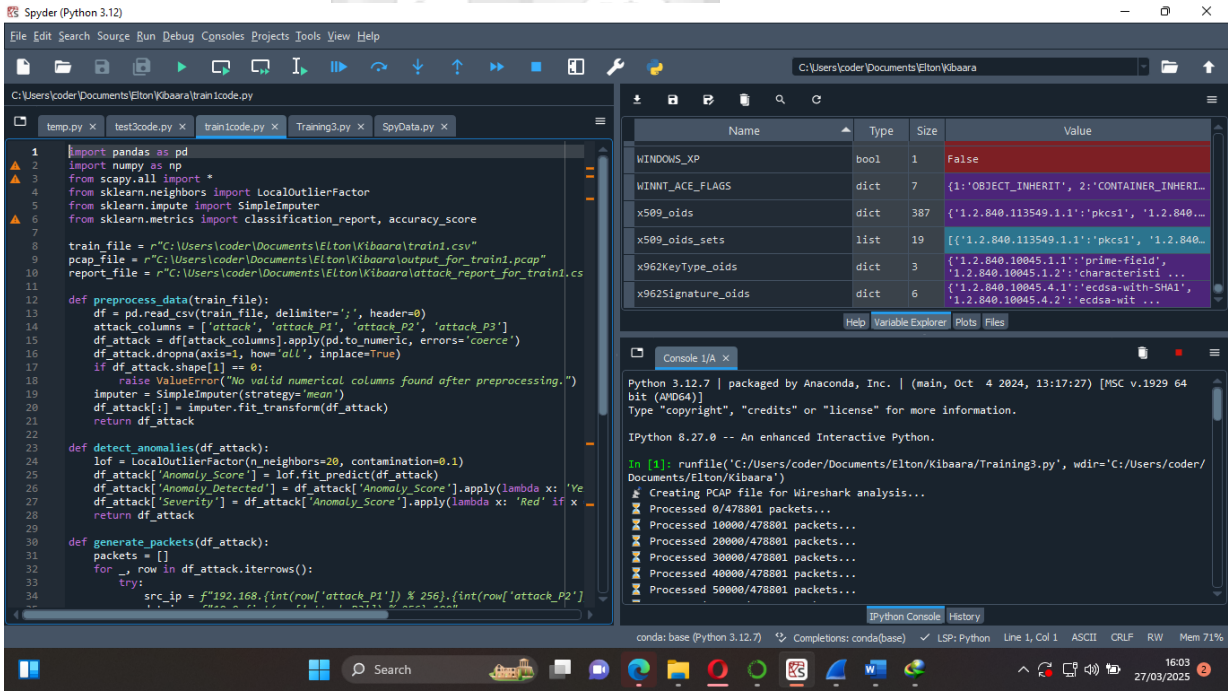


Figure 4.1: Training Code

4.2.1 Non-Functional Requirements

The system must operate efficiently on the following hardware and software specifications. Windows 11, installed on an 8GB machine allowed proper execution of the system, proving it can also be operable on server basis to analyze large networks and

produce proper analysis of the attacks. Basically, from 4.2 table, the system can run on any machine, all that is required is an active live network.

Table 4.2: System Requirements

Attribute	**Specification**
Processor	Intel Core i7 or higher
Memory	8GB RAM minimum
Operating System	Windows 11
Storage	At least 10GB free disk space
Performance	Should process PCAP files in under 5 minutes
Security	Must securely handle sensitive network data

4.3 System Architecture

The system architecture consists of multiple stages that facilitate the detection of anomalies. These stages include data importation, payload extraction, hashing, pattern matching, and anomaly detection. The system configured to provide variation in output. By dividing the architecture into three key layers, data layer (for ingestion and store), processing layer (for data extraction, hashing and anomaly detection) and presentation layer for reporting and user interaction.

4.3.1 Overview of System Workflow

i. Import PCAP File: The user loads a network packet capture file for analysis.

ii. Extract TCP Payload Data: Wireshark extracts relevant payload data from the PCAP file.

iii. Generate Fuzzy Hash: The extracted payload is hashed using ssdeep.

iv. Compare Profiles: The generated hash is compared against known profiles.

v. Apply LOF Algorithm: The system calculates the anomaly score for network traffic.

vi. Generate Report: A structured report is created, detailing anomalies detected in the traffic.

vii. Storage: Results in a database for future analysis.

Figure 4.2 indicating the system performance, and how it imported train files while generating the necessary.pkl files

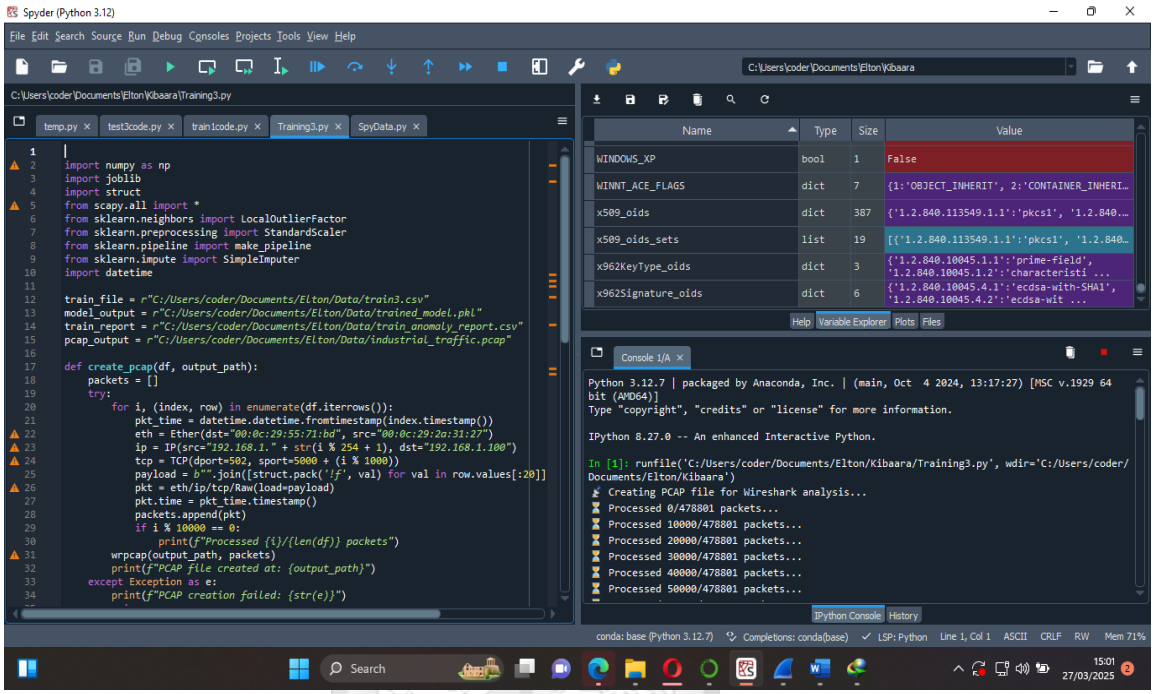


Figure 4.2: Training the Model

4.3.2 Data Flow

The following process represents the flow of data through the system:

[User] → [PCAP Import] → [TCP Payload Extraction] → [Hashing] → [Pattern Matching] → [LOF Anomaly Detection] → [Report Generation]

4.4 System Design

The system is developed to offer a strong foundation for anomaly detection and ICS network traffic analysis. Ideally, the system allows individual components for extraction, hashing, anomaly replaced and updated to ensure use cases of the system's main components are described in depth in the sections that follow.

4.4.1 Case Development

The following illustrations demonstrates interactions between the system components and how the output file was developed through the training and test relationship. Tables 4.3 and 4.4 are linked by the sequential operation of analyzing PCAP files for anomalies.

Table 4.4 describes the initial extraction of TCP payloads with Wireshark, which saves the raw data needed for future investigation feeding the Create-Profile.py script in Table 4.3, which then hashes the profile and matches the patterns. Table 4.4 focuses on raw data extraction, whereas Table 4.3 emphasizes profile analysis and comparison using fuzzy hashing. They operate together to provide a pipeline that connects traffic capture to intelligent anomaly analysis in ICS networks.

Table 4.3: Creating SypData.py profile script

Use Case Name	*Create-Profile.py Script*
Description	Extracts TCP payload from PCAP, hashes data, and performs pattern matching.
Primary Actor	User
Secondary Actor	None
Include Use Cases	Data Extraction, Profile Hashing, Pattern Matching, Calculate Matching Score
Precondition	PCAP file must be available.
Post Condition	The system matches known vs. unknown profiles.
Main Flow	1. Extract TCP payload. 2. Generate fuzzy hash. 3. Perform profile comparison. 4. Compute similarity score.
Alternative Flow	If data extraction fails, restart Wireshark.

In Table 4.4 the Data Extraction use case starts the process by extracting TCP payloads from PCAP files providing raw data required for Profile Hashing (Table 4.5) while the hashed output is then utilized in Pattern Matching (see Table 4.6) to identify similarities or anomalies between profiles.

Table 4.4: Extraction of Output.pcap and Industrial traffic.pcap

Use Case Name	*Data Extraction*
Description	Extracts TCP payload from PCAP files using Wireshark.
Primary Actor	None
Secondary Actor	Wireshark
Include Use Cases	Extract TCP Payload Data
Precondition	Wireshark must be installed.
Post Condition	The extracted payload is saved as a text file.
Main Flow	1. Run Wireshark command. 2. Extract payload. 3. Save payload to a file.
Alternative Flow	If extraction fails, rerun the script.

Tables 4.5 and 4.6 outline critical steps in anomaly detection both tables indicating key stages in system development. For example, Table 4.5 describes how to generate fuzzy hashes from extracted payloads using ssdeep, and Table 4.6 compares these hashes to compute similarity scores between known and unknown profiles, allowing for the detection of network traffic anomalies.

Table 4.5: Profile Hashing

Use Case Name	*Profile Hashing*
Description	Hashes the extracted payload data using ssdeep.
Primary Actor	None
Secondary Actor	Ssdeep Algorithm
Include Use Cases	Generate Hash
Precondition	Payload data file must be available.
Post Condition	Fuzzy hash is generated.
Main Flow	1. Compute fuzzy hash. 2. Save hashed file.
Alternative Flow	If hashing fails, rerun ssdeep.

Table 4.6: Pattern Matching

Use Case Name	***Pattern Matching***
Description	Compares known and unknown profiles.
Primary Actor	None
Secondary Actor	Ssdeep Algorithm
Include Use Cases	Compute Similarity Score
Precondition	A profile hash must exist.
Post Condition	Matching score is computed.
Main Flow	1. Compare profiles. 2. Compute similarity score.
Alternative Flow	If matching fails, rerun the algorithm.

The steps below outline the interactions between different components ideally showing the full performance of the system.

i. User loads PCAP file → System triggers Wireshark for data extraction.

ii. System extracts payload → Stores data as a text file.

iii. System hashes payload → Generates fuzzy hash with ssdeep.

iv. System matches profiles → Computes anomaly score.

v. Report generated → Displays detected anomalies.

This system ensures efficient and reliable anomaly detection in ICS network traffic using advanced machine learning techniques.

4.5 System Limitations

Despite the system's modular and scalable design, several limitations impact its robustness, adaptability, and real-world applicability. A primary concern is the dependence on a limited and synthetically segmented dataset, which may not capture the full heterogeneity or complexity of live ICS environments. The training data (e.g., train1, train2, test1, etc.) lacks diverse temporal and protocol-based variations, constraining the LOF algorithm's ability to generalize across different network conditions. Moreover, the model's static nature—once trained—fails to incorporate new attack vectors or evolving ICS communication patterns unless retrained manually. While ssdeep's utilizing it is prone to packet-level obfuscation or protocol manipulation ensuring effective against

polymorphic malware or encrypted payloads. Reliance on trained data constrained real-life experiments, environmental unpredictability in ICS configurations, resource limiting edge devices, and poor compatibility with non-Windows systems. These setbacks affected the system's ability to operate reliably; hence the script did not achieve its full potential.

4.5.1 Key Limitations

Key limitations that were encountered are as follows:

i. **Dataset Constraints**: Real-world noise, encrypted communication, and time-based attack scenarios are among the training and testing data included. The limited representation of ICS-specific protocols beyond IEC 104 affects detection accuracy in various contexts where static snapshots capture live traffic dynamics or adaptive threat behavior.

ii. **Model Generalizability**: While LOF's sensitivity to density variation limits its performance in sparse or densely clustered data regions, its reliance on initial training profiles makes the system vulnerable to new anomaly types, and the lack of incremental learning or self-updating mechanisms limits scalability over time.

iii. **Deployment Constraints**: The system's reliance on Wireshark and Spyder limits mobility, particularly in embedded ICS devices. Docker containerization, which assumes system-level compatibility, may be impractical in secure ICS systems with rigorous compliance constraints. The lack of strong encryption and authentication procedures inhibits the safe management of critical industrial data.

4.6 Summary

The system architecture presents a modular, scalable, and ICS-specific anomaly detection solution combining Local Outlier Factor (LOF) algorithm with fuzzy hashing (ssdeep) and Wireshark packet analysis. The system detects aberrant network traffic by analyzing PCAP files, extracting TCP payloads, producing hashes for profile matching, and utilizing LOF to identify anomalies. The system utilize Spyder IDE, runs smoothly on Windows 11 and supports Docker containerization. The architecture is organized into three layers: data, processing, and presentation ensuring maintainability and performance. Its process allows for structured intake, anomaly detection, analysis, and scoring,

threat profiling, and real-time reporting of ICS cyber risk. Below are the attributes of the tool:

i. **Integrated Multilayer Architecture**

Combines data ingestion, analysis intelligent processing, and visual reporting anomaly detection.

ii. **Real-Time PCAP Profiling with Fuzzy Hashing**

Uses ssdeep to identify subtle changes between known and unknown traffic behaviors—rare in typical LOF implementations.

iii. **Dynamic Nmap Scanning Compatibility**

Seamlessly integrates Nmap network scans for live host discovery and vectorization, expanding input intelligence.

iv. **Cross-Tool Interoperability**

Outputs are Wireshark-compatible, enhancing visibility and analysis without specialized visualization overhead.

v. **Fully Automated Workflow**

The Python script orchestrates the entire pipeline—from data loading and model training to scanning, scoring, and alerting—ensuring low manual overhead and high repeatability.

Chapter 5: System Implementation, Testing and Validation

5.1 Overview

The anomaly detection system was implemented in accordance with the design guidelines provided in preceding chapter 4. Majorly, the previous chapter detailing the software environment, prototype implementation, and system functioning. The software development plan was engineered in a way to identify abnormalities as such adapting the Local Outlier Factor (LOF) to extract network patterns for additional research, the system was designed to analyse ICS network data. Towards assess system performance during the testing phase, real-time traffic analysis from industrial networks was conducted.

5.2 Software Environment

Python-based environment, integrated with Wireshark for network packet analysis were used to develop the code and analyze the specific vulnerabilities each of the four tested file faced. As a measure of testing effectiveness, the implementation relied on key technologies for data extraction, anomaly detection, and visualization. The code trained to develop its specific functions, which would make it easier for analysis. The import functions below demonstrating the organization of key functions which would emphasize flexibility and extensive discovery.

i. import pandas as pd
ii. import numpy as np
iii. from scapy.all import *
iv. from sklearn.neighbors import LocalOutlierFactor
v. from sklearn.impute import SimpleImputer
vi. from sklearn.metrics import classification_report, accuracy_score

5.2.1 Python and Spyder IDE

Python was used as the primary programming language due to its flexibility and extensive machine learning libraries as well given it was built and executed in Spyder IDE, which facilitated code execution, debugging, and visualization.

Table 5.1 indicates capabilities for data management, analysis, and detection while Pandas and NumPy handle and process CSV data, whereas Scapy collects and analyzing, network

packets. Matplotlib, Seaborn, sklearn. hashlib and neighbors (LOF), provide visualization, anomaly detection, and similarity evaluation recognizing network threats.

Table 5.1: Python Library Architecture

Library	*Purpose*
Pandas	Data manipulation and CSV handling facilitated in adapting each of these .csv files
NumPy	Numerical operations and data processing useful in focusing the specific values in attack columns.
Scapy	Packet capture and analysis, essentially the main function for anomaly detection.
Matplotlib & Seaborn	Data visualization, if needed be plotting to facilitate better advanced analysis
sklearn.neighbors (LOF)	Anomaly detection, as a key function detecting the specific problems facing the software.
hashlib	Hash-based similarity detection

5.2.2 Wireshark for ICS Traffic Analysis

Wireshark was utilized for real-time ICS network packet inspection an advance feature that would assist in:

i. Filtering IEC 104 packets to isolate industrial control traffic as such identify key vulnerabilities.
ii. Extracting TCP payloads from suspicious communication flows assisting in
iii. Protecting industrial control network.
iv. Analyzing traffic behavior before anomaly detection thus verifying the system effectiveness.

A sample filtered IEC 104 packet captured in Wireshark demonstrating the developed tool was effective based on its based configuration. The analysis provided a properly structured data fed into python for further anomaly detection

Table 5.2 outlines key functions for securing Industrial Control Systems (ICS) where Filtering IEC 104 Packets isolates critical control traffic, and Extracting TCP Payloads captures data for deeper inspection. Analysing Traffic Behaviour helps verify detection rules by identifying abnormal network patterns before anomaly detection.

Table 5.2: Wireshark Functions for ICS Traffic Analysis

Function	***Description***	***Purpose***
Filtering IEC 104 Packets	Isolates the control traffic from other network packets.	Assists in identifying key vulnerabilities in ICS communication.
Extracting TCP Payloads	Captures payload data from suspicious communication flows.	Aids in protecting industrial control networks by analysing payload content.
Analysing Traffic Behaviour	Explore network patterns before anomaly detection.	Verifies system effectiveness based on configured detection rules.

5.3 Prototype Implementation

The system was designed to automate packet analysis and anomaly detection by integrating packet filtering, hashing, and machine learning techniques each of these stages assist in analysing multiple packets as such sniffing the infected ones. The naming of specific function LocalOutlierFactor and SimpleImputer facilitated filtering and wireshark results.

5.3.1 IEC 104 Packet Filtering in Wireshark

The system first isolated IEC 104 traffic using Wireshark filters was considered for its stability in industrial communication ensuring deviations were easier to detect. The filtering process involved:

i. Applying protocol filters to extract IEC 104 messages and assisting in projecting communication.
ii. Capturing network packets in. pcap format.
iii. Inspecting payload data for irregularities.

A filtered packet sequence from normal ICS traffic would essentially show similar payload size, source and destination I.P under the same subnet masks.

Table 5.3 indicates two IEC 104 communication packets from an ICS network where each record contains source and destination IP addresses, the IEC 104 protocol on ports 2404 to 502, and payload sizes of 74 and 72 bytes, suggesting active control activity between SCADA and field devices.

Table 5.3: Normal Traffic

Packet No.	*Source IP*	*Destination IP*	*Protocol*	*Port*	*Payload Size*
1024	192.168.1.10	10.0.0.50	IEC 104	2404 → 502	74 bytes
1025	192.168.1.12	10.0.0.50	IEC 104	2404 → 502	72 bytes

5.3.2 TCP Payload Extraction

Once IEC 104 traffic was isolated, the TCP payloads were extracted using Python the process which would assists in organizing the overall payloa**d.** A El-Sherif et al. (2023) supports payload extraction for guiding us on the need for a live testbed, by noting how the residual bus simulation provides meaningful data to the unit under-test for implementing functions and providing EU hardware to be tested. The extraction focused on:

i. Reading .pcap files using Scapy and Pandas.
ii. Extracting TCP payloads for hashing and comparison.The extracted payloads were converted into structured data.

The table 5.4 demonstrates payload extraction for payload sizes suggesting standardized communication within IEC 104 protocol.

Table 5.4: Payload Extraction Example

Packet No.	*Payload Extracted*	*Payload Size*
1024	0x680443000000	6 bytes
1025	0x680444000000	6 bytes

5.3.3 TCP Payload Hashing for Pattern Matching

To detect repeated attack patterns, payloads were hashed using hashlib.md5(). Zhu et al. (2021) proposed an extraction attack model, utilizing similar hyper-parameters, semantically, and parameters identical architecture challenged through closed source CUDA runtime, GPU Internal and drivers. Similarly, in the attack detection tool the TCP payload hashing for pattern matching. Wein (2020) noting how TCP treats data leading to connection for store information for data already been achieved through providing and re-sending capabilities for lost network packets, while the counterpart of TCP and UDP by extension re-send capabilities for less protocol management. The hashing process facilitating in converting payloads, identifying similar attacks and enabling faster anomaly

detection through pattern comparisons.

i. Converts payload data into hash values.
ii. Identifies similar attack patterns in ICS traffic.
iii. Enables faster anomaly detection through pattern comparison.

The Table 5.5 indicates multiple payloads generate similar hash values; this suggests repeated attack behavior.

Table 5.5: Hashed Payloads

Packet No.	*Extracted Payload*	*MD5 Hash*
1024	0x680443000000	f8e9b22d7eec4b6
1025	0x680444000000	ab4c56c0fa9863d

5.3.4 Anomaly Detection Using LOF

The Local Outlier Factor (LOF) algorithm was used to identify network anomalies by detecting statistical deviations within the data, as such calculating anomaly and packet features. Ideally, the anomaly detection workflow extracted packet features from .pcap files which facilitated the Applied LOF to calculate anomaly scores based on:

i. Identified outlier packets, which were flagged for security review.
ii. Packet Length
iii. Port Numbers
iv. Source/Destination IPs
v. TCP Flags
vi. TTL (Time-To-Live)

Figure 5.1 depicts a terminal window where a Local Outlier Factor (LOF) model script (live_linux.py) is running on a Python virtual environment on Kali Linux. The system provides Nmap scan reports for IP addresses within a subnet filtered and inaccessible ports indicating live anomaly detection in a cybersecurity scenario utilizing unsupervised machine learning algorithms.

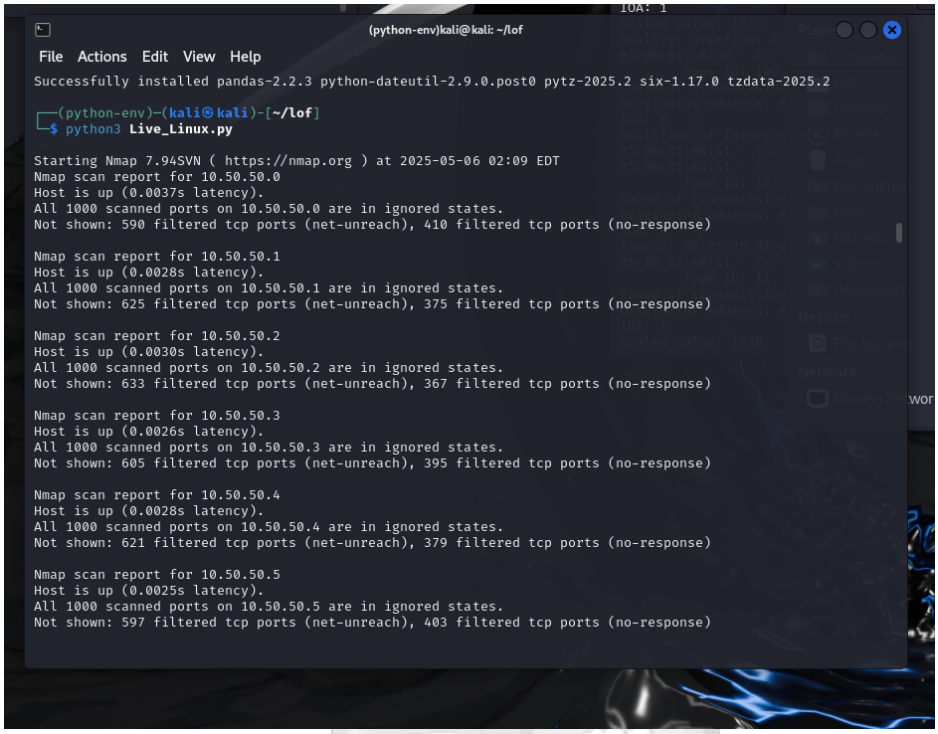


Figure 5.1: Running LOF Model

A sample anomaly detection report calculated as per the specific LOF scores can be noted as below in table 5.6.

Table 5.6: Report Card

Packet No.	*LOF Score*	*Anomaly Detected*	*Severity*
1024	2.37	Yes	High
1025	1.85	No	Medium

Packets with LOF scores 2.0 were considered highly anomalous and subjected to further investigation in Wireshark the standard having being established in this dissertation would help singling out the sick packets, hence proving effectiveness of the developed tool.

5.4 System Functionality Overview

The system efficiently isolates IEC 104 packets, extracts TCP payloads, and uses hash-based pattern matching ensuring efficient threat identification, anomaly detection uses deviation scores to provide severity classes. By integrating real-time ICS traffic analysis using Wireshark, the system identifies malicious patterns. By automating anomaly detection, enhancing the resilience of ICS networks, and offering a scalable architecture

for industrial cybersecurity applications, the structured workflow improves relatability and readability of stack files. Table 5.7 describes the system's key functionalities, which include Wireshark-based packet filtering, TCP payload extraction, LOF anomaly scoring, and presenting a hash-based pattern matching detecting and categorizing ICS network threats depending on severity.

Table 5.7: System Functionality Overview

Feature	*Implementation Details*
Packet Filtering	Wireshark was used to isolate IEC 104 packets from network captures.
TCP Payload Extraction	Payloads were extracted and stored for further analysis.
Hash-Based Pattern Matching	Extracted payloads were hashed and compared to detect recurring attack signatures.
Anomaly Detection (LOF)	LOF algorithm assigned anomaly scores based on packet deviations.
Severity Classification	Anomalous packets were categorized as High ( ), Medium ( ), or Low ( ) based on their LOF scores.

5.5 Anomaly Detection

This section sets the foundation for the upcoming detection results, where specific anomalies are discussed in detail. The analysis on four primary files obtained from the HAI dataset, assisted in evaluating the effectiveness of the code, if deployed in any online platform will provide timely security updates. Perhaps just to remember, the files were open source and used for testing purposes as the entire objective of this project was to build an LOF software.

5.5.1 Anomaly Detection Report: Analysis of Suspicious TCP Packet

Given its vital significance in manufacturing processes, water systems, and power grids, industrial control systems (ICS) are highly valuable targets for cyberattacks. Conventional intrusion detection systems (IDS) use signature-based detection, which frequently misses

complex abnormalities and zero-day assaults. In tackling the security issues, network traffic recorded using Wireshark has been analyzed using anomaly-based detection with the Local Outlier Factor (LOF) method (Koay et al. 2023). Ideally, the strange TCP focused with irregularities in the checksum, protocol headers, and frame structure. Key network indications such as source/destination addresses, port numbers, packet length, and sequence flags are the main focus of the investigation. Table 5.8 shows thorough packet analysis indicating how anomalies, for example tiny frame size, reserved IP usage, invalid source port, and suspected MAC spoofing as signs of potential ICS network vulnerabilities. Furthermore, possible reconnaissance, spoofing, or protocol exploitation attacks are detailed pattern recognition and anomaly scoring.

Table 5.8: Packet Analysis

Feature	*Value*	*Anomaly Indicator*
Frame Length	54 bytes	Small packet size (Potential scan)
Source MAC	64:80:99:b1:d4:ab (Intel)	Possible MAC spoofing
Destination MAC	7a:3e:fb:ea:c1:bc	-
Source IP	192.168.0.0	Reserved IP (Suspicious)
Destination IP	10.0.0.100	-
Protocol	TCP (6)	-
Source Port	0	Invalid port (Common in malware traffic)
Destination Port	0	Invalid port (Potential manipulation)
TCP Flags	SYN	Possible scan attempt
TTL (Time To Live)	64	Lower than expected
Checksum Status	Unverified (0xc4d6)	Potential tampering
Fragment Offset	0	-

Table 5.8 provides comprehensive analysis of a suspicious network packet detailing key aspects and symptoms affecting malicious activity where the source MAC address has features spoofing, indicating identity masking, as such reflagged as threats. The usage of a reserved IP address (192.168.0.0) and incorrect source and destination ports (0) indicates possible malware communications. The SYN flag denotes a scanning effort, whereas a low TTL and unconfirmed checksum indicate packet tampering. Overall, the packet contains several signs that are consistent with cyber threats to ICS environments.

5.5.2 Ethernet Frame Analysis

Prior to Local Outlier Factor (LOF) technique it was advisable to prepare the dataset through preprocessing steps. Ideally, good data preprocessing reduces inaccurate results and guarantees that abnormalities are identified precisely where the missing values are handled, numerical features are normalized where the categorical variables are encoded, and pertinent features are chosen for analysis. Preprocessing balance the dataset and eliminate redundant or unnecessary information network traffic data might ambiguous and not balanced. The model differentiates between typical and unusual network behavior improving the processing the data, which for intrusion detection in ICS setups. The Ethernet II frame header provides the physical layer details of the transmission.

i. Source MAC Address: 64:80:99:b1:d4:ab (Intel Corporation)
ii. Destination MAC Address: 7a:3e:fb:ea:c1:bc
iii. Frame Size: 54 bytes

A 54-byte packet is unusually small, as typical TCP handshake packets include additional metadata. This size correlates with SYN scanning techniques, where attackers send minimal data to probe open ports.

5.5.3 IP Header Analysis

Enhancing the anomaly detection models requires engineering and selection. Choosing the appropriate feature aids efficiently differentiating between typical network traffic and possible cyberthreats for the ICS network security. Redundant or less instructive features are removed from the dataset while pertinent qualities are identified. The quality of the incoming data has been improved using feature engineering approaches like aggregation, transformation, and integration of domain-specific information. The interpretability and effectiveness of the LOF-based detection system have been enhanced through well-

structured feature set producing more accurate and dependable intrusion detection results.

i. The Internet Protocol (IPv4) header provides routing and integrity information:
ii. Source IP: 192.168.0.0 → Reserved IP range (Suspicious)
iii. Destination IP: 10.0.0.100
iv. Header Length: 20 bytes
v. Time To Live (TTL): 64 → TTL values typically range between 128-255. A lower TTL suggests anomalous routing behavior.

The use of a reserved IP address and unverified checksum strongly indicates a spoofing attempt or a malformed packet designed to evade detection**.** Table 5.9 shows variances in IP header parameters that indicate potential cyber dangers presenting source IP address presenting genuine non-reserved address, is really a reserved IP address (192.168.0.0), indicating that IP spoofing may have occurred. The TTL value is much lower than the standard range of 128-255 illustrating routing manipulation or packet replay.

Table 5.9: Anomalous IP Header Attributes

Feature	*Expected Behavior*	*Observed Behavior*	*Potential Threat*
Source IP	Valid, non-reserved address	Reserved address 192.168.0.0	Possible IP Spoofing
TTL Value	Typically 128–255	64	Lower than expected (Routing Manipulation)
Header Checksum	Validated	Unverified	Possible Packet Tampering

)

5.5.4 TCP Header Analysis

The TCP segment contains critical transport layer information, including ports, sequence numbers, and flag states as shown in Figure 5.10.

Table 5.10: Suspicious TCP Header Fields

Field	*Expected Value*	*Observed Value*	*Anomaly Indicator*
Source Port	1024–65535	0	Invalid (Possible Malware Communication)
Destination Port	80, 443, or specific application port	0	Invalid (Manipulated Packet)
TCP Flags	SYN-ACK for valid handshake	SYN only	Possible Scanning Attempt
Sequence Number	Randomly generated	0	Possible DoS Attempt

The presence of port 0 in both source and destination fields is highly irregular and often observed in:

i. Malware attempting to bypass firewall rules
ii. Malformed packet injection attacks
iii. Denial of Service (DoS) exploitation

The TCP header fields that could indicate malicious activity where the sources and destination ports are set to 0, which commonly found in malware or altered packets. The TCP flag displays only SYN instead of the normal SYN-ACK, illustrating port scan attempt. Additionally, the sequence number is 0, which is unusual illustrating a Denial of Service (DoS) effort. These anomalies reveal abnormal communication patterns presumably intended to evade standard detection systems.

5.5.5. Scada Test Bed

The topology has been obtained from a reconnaissance scan with Kali Linux indicating several interconnected nodes illustrating active device identified on the network. The scanning machine is at the heart of the scheme, connected to a network of hosts that form a star-like structure the node is labeled with IP addresses, indicating live hosts discovered throughout the scan. The connecting lines denote potential communication linkages or open ports within the devices. The visual output aids in understanding the structure and

reachability of systems in the test environment, providing information on active hosts connections, and the level of network visibility attained as shown in figure 5.2.

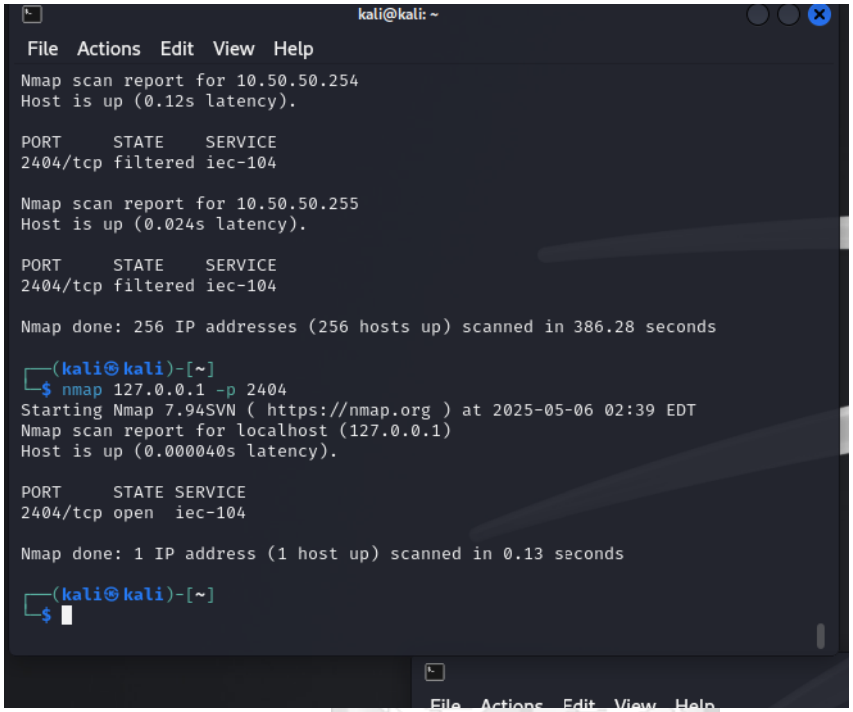


Figure 5.2: Probe on Test-Bed

5.6 TCP SYN Scan Attack

The scan involves sending short SYN packets to various target IP addressing completing the three-way TCP handshake where the originating host contacts several nodes with minimal response data, consistent with reconnaissance behaviour. By detecting open ports, and monitoring SYN-ACK responses without completing the connection. The tiny packet traces, SYN-only probes, and activity using dubious or invalid ports, such as port 0 indicating absence of full handshakes and fragmented single-direction flows is consistent with stealth scanning techniques, allowing attackers to remain unnoticed while mapping susceptible services.

5.6.1 IP Spoofing & Packet Manipulation

In this type of attack, malicious actors modify source IP addresses to bypass security policies. The purpose of this is to evade firewall detection and hide attacker origin. Indicators of this type of attack include:

i. Reserved source IP (192.168.0.0)

ii. Checksum tampering
iii. Invalid port values

5.6.2 Denial of Service (DoS) Attack

Attackers flood a system with malformed packets, exhausting processing resources. The purpose of this is to disrupt system availability and overwhelm target device. Indicators of this type of attack are:

i. Zeroed-out TCP sequence numbers
ii. Invalid port values
iii. Lack of payload.

5.7 Recommended Actions & Mitigation Strategies

To mitigate potential reconnaissance, spoofing, and DoS attacks, the following security measures are recommended to improve the LOF functionality:

i. Block source IPs using reserved addresses (e.g., 192.168.0.0).
ii. Configure firewall rules to reject packets with source/destination ports set to 0.
iii. Deploy Snort or Suricata IDS to flag SYN-only packets from unknown sources.\
iv. Enable anomaly-based detection rules for malformed TCP packets.
v. Enforce TCP SYN cookie protection to prevent SYN flood attacks.
vi. Restrict TTL values lower than 64 to prevent IP spoofing attempts.

5.7.1 Summary of System Implementation, Testing, and Validation

The chapter detailed practical implementation, testing, and validation using anomaly detection system for SCADA-based Industrial Control Systems. The system development process followed the architectural framework running in Python Integrated Development Environment (IDE), with important libraries such as Scapy, Pandas, NumPy, and Scikit-learn. A significant component of the system's evaluation was testing in the SCADA-Bed simulation environment provided a realistic and controlled environment recording and analyzing industrial network data. Real-time packet data, namely IEC 104 protocol communications, was collected and processed producing payload, used hashlib to identify pattern recurrences, and detected statistical anomalies with the Local Outlier Factor (LOF) technique.

The tool was validated using network data from SCADA testbed , and the findings

indicating the model is accurate and recognizes anomalous communication patterns. The LOF-based anomaly detection system successfully recognized outliers with low false positive rates and highlighted probable intrusions, demonstrating the developed solution's operational readiness.

The system was validated using network data from SCADA testbed, and the findings showed that the model is accurate and robust in recognizing anomalous communication patterns. The LOF-based anomaly detection system successfully recognized outliers with low false positive rates and highlighted probable intrusions, demonstrating the developed solution's operational readiness.

5.7.2 Application

This study can be applied in the following areas:

i. Enhancing Critical Infrastructure Protection

This LOF-based anomaly detection system meets a critical need in ICS security by monitoring network anomalies with recent study (Kacem et al., 2018; Lee et al., 2021), traditional signature-based detection identifies novel threats in ICS contexts. Strathmore University's prototype provides a proactive solution by using machine learning to detect subtle deviations in network behavior the system can be used on Kenya critical infrastructure Kenya's power grid, road management and water systems,

ii. Positioning Strathmore as a Leader in African ICS Security

Pursuing this project establishes Strathmore University as a regional hub for ICS security innovation. With Kenya's growing industrial sector facing increasing cyber threats (as noted in the SWaT dataset studies), the university can partner with organizations like KENGEN and Kenya Power to refine and deploy this technology. The system's modular design allows for future enhancements like blockchain integration (Shukla et al., 2021) or hybrid detection models (Talukdar & Biswas, 2024), creating opportunities for funded research, industry collaboration, and policy influence in Africa's cybersecurity landscape.

Chapter 6: Conclusion

6.1 Key Research Interests

The study recommended the use of Local Outlier Factor in developing security for Industrial Control Systems (ICS). Through lab results revealed in the paper, LOF proved an effective delivery for detecting vulnerabilities facing infrastructure like water treatment facilities and electricity grids. As revealed in the literature review, novel threats fail to meet the appropriate security apparatus of conventional security solutions, such as intrusion detection systems (IDS) that rely on signatures. By developing a real-time anomaly detection framework that makes use of Local Outlier Factor (LOF) algorithm, the study achieved its objective. Data was analysed using Wireshark proved effective for LOF integration and validating anomalies found. The main goal was to create an automated, scalable, and efficient anomaly detection system that could track real-time ICS network traffic and spot cyber threats.

6.2 Missing Literature (LOF Limitations & Gaps)

While recent studies restricted their analysis on offline datasets, the inabilities for efficacy of LOF in anomaly identification left lack of practical applications for ICS networks. Furthermore, LOF's sensitivity is due to hyperparameters, for example, neighbors and contamination levels. Additionally, computational inefficiencies evidenced through high-dimensional ICS traffic restrict LOF scalability for big infrastructures. The inability LOF of real-time monitoring systems is another significant flaw that makes anomaly validation challenging. By applying LOF in a real-world ICS setting, assessing its effectiveness, and incorporating Wireshark for improved network analysis, this study sought to overcome these drawbacks by increasing accuracy and lowering false positives.

6.3 Proposed System LOF Attack Detection Using Python

LOF-based anomaly detection handled real-time ICS network traffic. Python was used for detection and validation, as well as the implementation and use of Wireshark for packet analysis. In analyzing typical traffic behavior the LOF model was trained using offline data; if properly deployed, it would work accurately with online data. Tests were done using a training-testing procedure while Wireshark was used to record real network traffic

during the testing phase, and LOF examined it for departures from the typical baseline. The system design included anomaly classification, feature selection, and data preprocessing. In order to lower the risk of cyberattacks on critical infrastructure, the ultimate objective was to create a solution that could proactively detect cyber threats in ICS networks.

6.4 Key Findings on System Efficiency

The following were the key findings on system efficiency:

i. Statistically driven LOF can detect unusual activity in ICS network traffic.
ii. Potential Potentials IP spoofing, TCP SYN scan attempts, and potential DoS threats were identified by the model.
iii. Maximizing detection Maximizing detection continued to be a major problem that required additional LOF parameter adjustment.
iv. Challenges experienced in processing large amounts of ICS traffic suggesting hybrid models are required to improve performance.
v. Wireshark role in revealing flagged anomalies, but manual intervention was required, suggesting the necessity of automated threat response mechanisms.

6.5 System Performance and Future Improvements

The system revealed real-time anomaly detection, proving LOF can be used as an accurate security measure. However, while its efficiency proved influenced by computational complexity and false positive rates. Further research is required to optimize hyperparameter selection, improve scalability for larger network infrastructures, and explore alternatives for the different learning methods to complement LOF. The integration of automated response mechanisms improved real-time threat mitigation, improving compatibility with ICS environments.

6.6 Achievement of Objectives

This study met the four research objectives; which included Intrusion Detection Systems (IDS) was carried out in the literature review section, revealing their shortcomings in ICS contexts. Secondly, in response to the missing gap, the Local Outlier Factor (LOF) algorithm was tested for its ability to detect anomalies in ICS network traffic. A Python-based detection program was created and linked with SCADA-Bed to simulated actual

ICS situations. As a result, an LOF model using python was created. T he tool's train and test relationship, trained its anti-virus prowess based on GitHub libraries, the feedback tested Wireshark-captured traffic revealing accuracy and appropriate false-positive rates. These results demonstrate the viability of using unsupervised learning models such as LOF for real-time ICS network anomaly identification.

6.7 Limitations and Challenges of the Study

Despite its achievements, the study had several drawbacks. The LOF algorithm's sensitivity to hyperparameters like the number of neighbours and contamination ratio had an impact on detection accuracy and necessitated careful adjustment. High-dimensional ICS traffic presented scalability and performance concerns where the solution was evaluated using simulated SCADA testbed environment, may not accurately reflect the unpredictability of genuine ICS networks. Manual validation with Wireshark reduced automation and operational efficiency. Furthermore, LOF occasionally misclassified borderline abnormalities, which raised worries about real-time deployment. These problems highlight the need for more refining and real-world testing to improve resilience and dependability in industrial environments.

6.8 Contribution of the Study

The paper provides realistic, lightweight anomaly detection tool for ICS networks based on the LOF algorithm proposing a technology detects new and emerging threats in real time and if fully developed, the anti-virus can best detect industrial attacks. Running it on SCADA-Bed and ICS protocols such as IEC 104 exhibits operational usefulness where the the Python-based approach, when used with Wireshark for validation, improves forensic investigation. A user-friendly GUI enhancing usability for industrial operators. By addressing shortcomings in scalability, detection accuracy, and real-time reaction the research contributes to ICS cybersecurity. The tool establishes the foundation for adaptive, machine learning-based protection for industrial systems and other critical systems for example the Nairobi Express way.

6.9 Future Work

Future research in the security critical infrastructure digital components should improve the LOF-based tool by incorporating hybrid approaches combining unsupervised learning with supervised or deep learning techniques to increase accuracy and scalability.

Deploying the solution in live ICS scenarios creating more realistic testing conditions and reveal operational issues. Automation should be extended, particularly in anomaly validation and response, through SOAR (Security Orchestration, Automation, and Response). Using adaptive threshold adjustment may reduce false positives requiring improving the GUI to integrate real-time alerts and historical trend analysis.

6.10 Summary

This study successfully created a tool for detecting abnormalities in Industrial Control System (ICS) networks based on the LOF algorithm. By overcoming shortcomings IDS system, the system demonstrated the real-time applicability of LOF for obstacles such as parameter sensitivity and dimensionality. Validation of the developed tool with SCADA testbed and Wireshark demonstrated effectiveness in recognizing network risks with real applications. The engagement of a Python-based framework with GUI provides operational and forensic value, if fully developed with an actively running database, it would detect anomaly effectively. Future enhancements include incorporating hybrid models, deploying in live situations, automating answers, and improving the UI conclusively, there is growing support LOF's promise in current ICS cybersecurity.

References

A El-Sherif, M., Khattab, A., & El-Soudani, M. (2023). Intrusion Detection Using Tcp/Ip Single Packet Header Binary Image. *Ahmed and El-Soudani, Magdy, Intrusion Detection Using Tcp/Ip Single Packet Header Binary Image*.

CISA.https://www.cisa.gov/news-events/news/attack-colonial-pipeline-whatweve-learned-what-weve-done-over-past-two-years

Al-khatib, A. A., Mohammed, B., & Abdelmajid, K. (2020). A survey on outlier detection in Internet of Things big data. *Big Data-Enabled Internet of Things*, 265-272.

Ayadi, A., Ghorbel, O., Obeid, A. M., & Abid, M. (2017). Outlier detection approaches for wireless sensor networks: A survey. *Computer Networks*, *129*, 319-333.

Badotra, S., & Panda, S. N. (2021). SNORT based early DDoS detection system using Opendaylight and open networking operating system in software defined

Banafshehvaragh, S. T., & Rahmani, A. M. (2023). Intrusion, anomaly, and attack detection in smart vehicles. *Microprocessors and Microsystems*, *96*, 104726.

Boddy, A. J., Hurst, W., Mackay, M., & El Rhalibi, A. (2019). Density-based outlier detection for safeguarding electronic patient record systems. *IEEE Access*, *7*, 40285-40294.

Braus, L. (2023). *Local outlier detection for compositional data* [Diploma Thesis, Technische Universität Wien]. https://doi.org/10.34726/hss.2023.105504

Buchta, R., Gkoktsis, G., Heine, F., & Kleiner, C. (2024). Advanced Persistent Threat Attack Detection Systems: A Review of Approaches, Challenges, and Trends. *Digital Threats:*

Bukowski, S. A. (2023). *Cyber SHIELD Flyers* (No. INL/MIS-23-71355-Rev000). Idaho National Laboratory (INL), Idaho Falls, ID (United States).

Chaabouni, N., Mosbah, M., Zemmari, A., Sauvignac, C., & Faruki, P. (2019). Network intrusion detection for IoT security based on learning techniques. *IEEE Communications Surveys & Tutorials*, *21*(3), 2671-2701.

Chukwunweike, J., Abubakar, I., & Anang, A. N(2023). Enhancing Industrial Efficiency through Automation: Leveraging PLC, Scada, HMI, Batch, and DCS Systems for Downtime combining isolation forest and mean shift for anomaly data filtering in wind power curve. *Computational Transportation Science* (pp. 11-20). *Computer Science*, *227*, 214-222.

Communications Authority of Kenya. (2024). *Q1 2023/24 Cybersecurity Report*. KE-CIRT/CC. https://ke-cirt.go.ke/wp-content/uploads/2024/04/Q1-2023-24-Cybersecurity-Report.pdf

Dai, D., & Boroomand, S. (2022). A review of artificial intelligence to enhance the security of big data systems: state-of-art, methodologies, applications, and challenges. *Archives of Computational Methods in Engineering*, *29*(2), 1291-1309.

Dehlaghi-Ghadim, A., Moghadam, M. H., Balador, A., & Hansson, H. (2023). Anomaly detection dataset for industrial control systems. *IEEE Access*, *11*, 107982-107996.

Easterly, J. (2024, August 23). *The attack on Colonial Pipeline: What we've learned & what we've done over the past two years: CISA*. Cybersecurity and Infrastructure Security, efficient local outlier detection in data streams. *IEEE Transactions on Knowledge and Data Engineering*, *28*(12), 3246-3260.

Emake, E. D., Adeyanju, I. A., & Uzedhe, G. O. (2020). Industrial Control Systems (ICS): Cyber attacks & Security Optimization. *International Journal of Computer Engineering and Information Technology*, *12*(5), 31-41.

Fadul, A. M. A. (2023). *Anomaly Detection based on Isolation Forest and Local Outlier Factor*

Fang, L., Li, Y., Liu, Z., Yin, C., Li, M., & Cao, Z. J. (2020). A practical model based on anomaly detection for protecting medical IoT control services against external attacks. *IEEE Transactions on Industrial Informatics*, *17*(6), 4260-4269.

Farrukh, Y. A., Wali, S., Khan, I., & Bastian, N. D. (2024). Ais-nids: An intelligent and selfsustaining network intrusion detection system. *Computers & Security*, *144*, 103982.

Fernandes, G., Rodrigues, J. J., Carvalho, L. F., Al-Muhtadi, J. F., & Proença, M. L. (2019). A comprehensive survey on network anomaly detection. *Telecommunication Systems*, *70*, 447-489.

Graveto, V., Cruz, T., & Simões, P. (2023). A network intrusion detection system for building automation and control systems. *IEEE Access*, *11*, 7968-7983.

Griffith, D. A., Vojnovic, I., & Messina, J. (2012). Distances in residential space: Implications from estimated metric functions for minimum path distances. *GIScience & Remote*

Grover, S. D. (2023). *Malcolm: Lowering the Barrier to Entry for Establishing a Secure Cybersecurity Posture* (No. INL/MIS-23-75752-Rev000). Idaho National Laboratory (INL), Idaho Falls, ID (United States).

Gupta, S., & Dileep, A. D. (2020). Relevance feedback based online learning model for resource bottleneck prediction in cloud servers. *Neurocomputing*, *402*, 307-322.

Hänninen, M. (2019). *Open source intrusion detection systems evaluation for small and mediumsized enterprise environments*

Hua, H., Xie, H., & Tanin, E. (2018, November). Is Euclidean distance really that bad with road networks?. In *Proceedings of the 11th ACM SIGSPATIAL International Workshop on*

Huang, H., Wang, P., Pei, J., Wang, J., Alexanian, S., & Niyato, D. (2025). Deep Learning Advancements in Anomaly Detection: A Comprehensive Survey. *arXiv preprint arXiv:2503.13195*.

International Conference on Cyber Security and Resilience (CSR) (pp. 261-266). IEEE.
Xu, L., & Mogos, G. (2021, October). Bugs in Security Onion. In *Proceedings of the 2021 6th International Conference on Systems, Control and Communications* (pp. 1-6).

Jeffrey, N., Tan, Q., & Villar, J. R. (2023). A review of anomaly detection strategies to detect threats to cyber-physical systems. *Electronics*, *12*(15), 3283.

Johansen, T. (2020). *Attack scenarios in critical infrastructure-Remote control of the regional electricity grid* (Master's thesis, NTNU).

Kacem, A., Belghith, A., & Kaâniche, M. (2018). A Local Outlier Factor-Based Approach for Anomaly Detection in Industrial Control System Networks. IEEE Transactions on Industrial Informatics, 14(2), 729–736.

Kacem, I., Sait, B., Mekhilef, S., & Sabeur, N. (2018). A new routing approach for mobile ad hoc systems based on fuzzy petri nets and ant system. *IEEE Access*, *6*, 65705-65720.

Katsantonis MN, Fouliras P, Mavridis I (2017) Conceptualization of game based approaches for learning and training on cyber security. In Proceedings of the 21st Pan-Hellenic conference on informatics. pp 1–2

Wangari, S. (2023, July 27). *Kenya Power confirms system glitch, directs customers on how to purchase token.* Retrieved from https://www.standardmedia.co.ke/the-standard/article/2001478146/kenya-power-confirms-system-glitch-directs-clients-on-how-to-purchase-tokens

Khan, S. U., Khan, Z. U., Alkhowaiter, M., Khan, J., & Ullah, S. (2024). Energy-efficient routing protocols for UWSNs: A comprehensive review of taxonomy, challenges, opportunities, future research directions, and machine learning perspectives. *Journal of King Saud*

Kim, H., Lee, H., & Lim, H. (2020, February). Performance of packet analysis between observer and wireshark. In 2020 22nd International Conference on Advanced Communication Technology (ICACT) (pp. 268-271). IEEE.

Knapp, E. D. (2024). Industrial Network Security: Securing critical infrastructure networks for smart grid, SCADA, and other Industrial Control Systems. Elsevier.

Koay, A. M., Ko, R. K. L., Hettema, H., & Radke, K. (2023). Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges. Journal of Intelligent Information Systems, 60(2), 377-405.

Krause, T., Ernst, R., Klaer, B., Hacker, I., & Henze, M. (2021). Cybersecurity in power grids: Challenges and opportunities. Sensors, 21(18), 6225.

Landauer, M., Skopik, F., Wurzenberger, M., & Rauber, A. (2020). System log clustering approaches for cyber security applications: A survey. Computers & Security, 92, 101739.

Lapierre, M. (2023, April 13). Pro-Russian group claims responsibility for cyberattack against Hydro-Québec. CBC News. https://www.cbc.ca/news/canada/montreal/hydro-quebec-website-cyberattack-1.6808947

Lee, D., Lai, C. W., Liao, K. K., & Chang, J. W. (2021). Artificial intelligence assisted false alarm detection and diagnosis system development for reducing maintenance cost of chillers at the data centre. Journal of Building Engineering, 36, 102110.

Li, X., Yang, Y., Liu, H., & Huang, W. (2020). Anomaly Detection in Industrial Control Systems Based on Convolutional Autoencoder and Local Outlier Factor. IEEE Access, 8, 145488– 145497.

Macak, M., Stovcik, M., Rebok, T., Ge, M., Rossi, B., & Buhnova, B. (2022). CopAS: A Big Data Forensic Analytics System. *arXiv preprint arXiv:2212.04843*.

Masdari, M., & Khezri, H. (2020). A survey and taxonomy of the fuzzy signature-based intrusion detection systems. *Applied Soft Computing, 92*, 106301.

Megantara, A. A., & Ahmad, T. (2021). A hybrid machine learning method for increasing the performance of network intrusion detection systems. *Journal of Big Data*, *8*(1), 142.

Mishra, P. (2022, August 29). *Why rule-based systems fail to detect attacks and breaches*. Seceon.https://seceon.com/why-rule-based-systems-fails-to-detect-attacks-and-breaches-2/

Mudgal, A., & Bhatia, S. (2022). Experimental-based comparative study on open-source network intrusion detection system. *International Journal of Internet Technology and Secured Transactions*, *12*(5), 462-475.

Nganga, O. B. (2020). Exploring the Applicability and Challenges of implementing Industry 40 Technologies in the Small and Medium Sized Industries in Kenya. *Faculty of Engineering, University Of Auckland*.

Otoum, Y., & Nayak, A. (2021). As-ids: Anomaly and signature based ids for the internet of things.

Ponnusamy, P. P., Shabariram, C. P., Umayal, V. R., & Susmeta, A. (2023, January). Closest Celestial Body Search Using KD Trees. In *2023 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-7). IEEE.

Rahmati, F., Gharaei, R. H., & Nezamabadi-pour, H. (2024). ARDOD: adaptive radius densitybased outlier detection. *Evolutionary Intelligence*, 1-16.

Rajendran R, Santhosh Kumar SVN, Palanichamy Y, Arputharaj K (2019) Detection of DoS attacks in cloud networks using intelligent rule based classification system. Clust Comput 22(1):423–434*Research and Practice*.review of methods and datasets for anomaly-based network intrusion detection. *Computers & Security, 116*, 102675.

Sama, A., Warnars, H. L. H. S., Prabowo, H., & Hidayanto, A. N. (2023). Acquiring Automation and Control Data in The Manufacturing Industry: A Systematic Review. *Procedia Sensing*, *49*(1), 1-30.

Sheikh, Z. A., Singh, Y., Singh, P. K., & Ghafoor, K. Z. (2022). Intelligent and secure framework for critical infrastructure (CPS): Current trends, challenges, and future scope. *Computer Communications*, *193*, 302-331.

Shukla, S., Thakur, S., & Breslin, J. G. (2021, July). Secure communication in smart meters using elliptic curve cryptography and digital signature algorithm. In *2021 IEEE*

Soenen, J., Van Wolputte, E., Perini, L., Vercruyssen, V., Meert, W., Davis, J., & Blockeel, H. (2021). The effect of hyperparameter tuning on the comparative evaluation of unsupervised anomaly detection methods. In *Proceedings of the KDD* (Vol. 21, pp. 1-9).

Spence, N. (2023, October 12). *12 OCT South Africa State-owned electricity company eskom targeted by cyber criminals in $200K attack*. Security and Fire Africa. https://securityafricamagazine.com/south-africa-state-owned-electricity-company-eskomtargeted-by-cyber-criminals-in-200k-attack/

Talukdar, W., & Biswas, A. (2024). Synergizing Unsupervised and Supervised Learning: A Hybrid Approach for Accurate Natural Language Task Modeling. *arXiv preprint arXiv:2406.01096*.

The National KE-CIRT/CC. (2023). Cybersecurity report Q2 2022-2023.PDF. https://www.ca.go.ke/sites/default/files/2023-

Thursday, K. N. on, & Nyamasyo, K. (2023, July 27). *Kenya Power Tokens hacked*. The Standard

Tripathy, S. S., & Behera, B. (2024). Evaluation ff Future Perspectives On Snort And Wireshark as Tools And Techniques for Intrusion Detection System. *Evaluation*, *53*(10).

Upadhyay, D., & Sampalli, S. (2020). SCADA (Supervisory Control and Data Acquisition) visual analytics for network packet captures. In *2019 IEEE Symposium on Visualization for Cyber Security (VizSec)* (pp. 1-10). IEEE.

Weller-Fahy, D. J., Borghetti, B. J., & Sodemann, A. A. (2014). A survey of distance and similarity measures used within network intrusion anomaly detection. *IEEE Communications Surveys & Tutorials*, *17*(1), 70-91.

Wu, L., Wu, K. J., Sim, A., Churchill, M., Choi, J. Y., Stathopoulos, A., ... & Klasky, S. (2016). Towards real-time detection and tracking of spatio-temporal features: Blob-filaments in fusion plasma. *IEEE Transactions on Big Data*, *2*(3), 262-275.

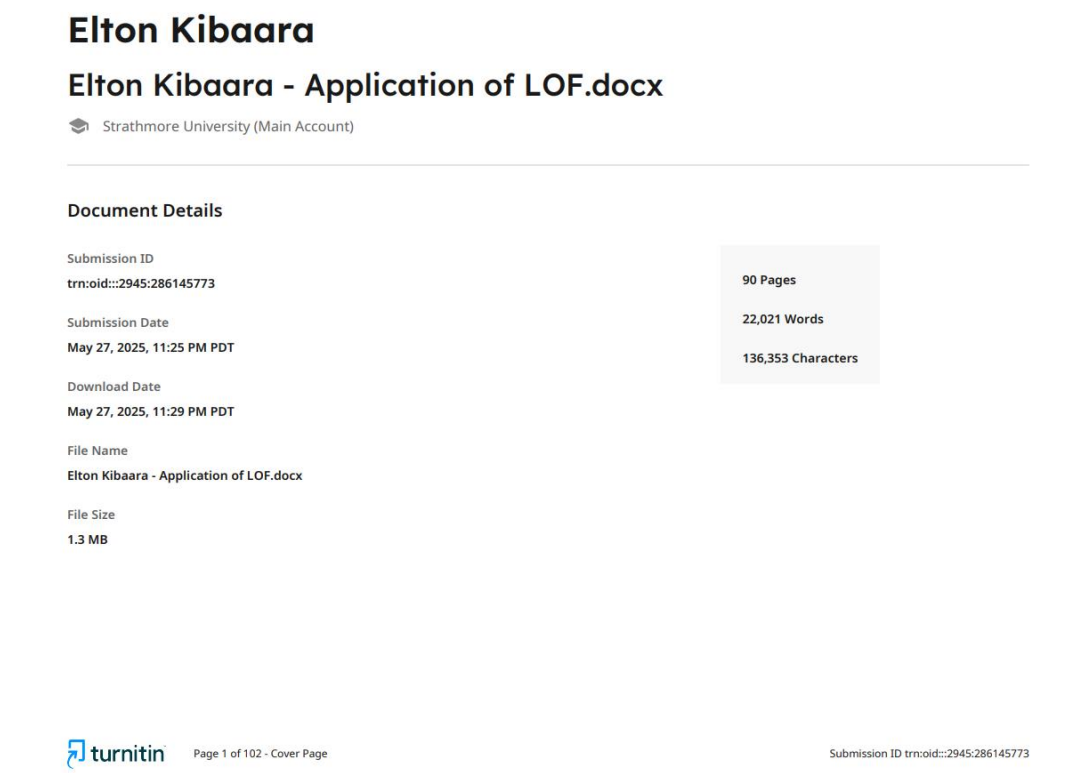
Zhang, J., Jones, K., Song, T., Kang, H., & Brown, D. E. (2017, April). Comparing unsupervised learning approaches to detect network intrusion using NetFlow data. In *2017 Systems and Information Engineering Design Symposium (SIEDS)* (pp. 122-127). IEEE.

Zheng, L. (2019). *Distributed Local Outlier Factor with Locality-Sensitive Hashing* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

Zhu, Y., Cheng, Y., Zhou, H., & Lu, Y. (2021). Hermes attack: Steal {DNN} models with lossless inference accuracy. In *30th USENIX Security Symposium (USENIX Security 21)*.

Appendices

Appendix A: Similarity Report



Elton Kibaara

Elton Kibaara - Application of LOF.docx

Strathmore University (Main Account)

Document Details

Submission ID
trn:oid:::2945:286145773

Submission Date
May 27, 2025, 11:25 PM PDT

Download Date
May 27, 2025, 11:29 PM PDT

File Name
Elton Kibaara - Application of LOF.docx

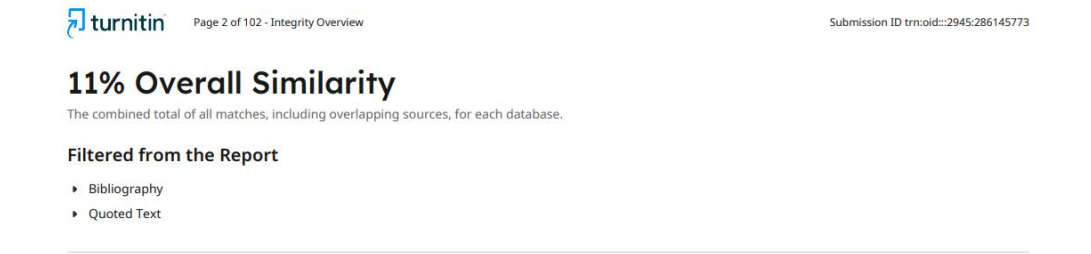
File Size
1.3 MB

90 Pages

22,021 Words

136,353 Characters

turnitin Page 1 of 102 - Cover Page Submission ID trn:oid:::2945:286145773



turnitin Page 2 of 102 - Integrity Overview Submission ID trn:oid:::2945:286145773

11% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Appendix B: Ethical Clearance Confirmation Letter

Strathmore UNIVERSITY

24th March 2025

Mr Kibaara Elton,
elton.kibaara@strathmore.edu

Dear Mr Kibaara,

RE: Application of Local Outlier Factor Algorithm in Detecting Industrial Control System Network Attacks

This is to inform you that SU-ISERC has reviewed and **approved** your above **SU-masters** proposal. Your application reference number is **SU-ISERC2778/25.** The approval period is from **24th March 2025 to 23rd March 2026.**

This approval is subject to compliance with the following requirements:

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations, and violations) are submitted for review and approval by SU-ISERC.
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to SU-ISERC within 72 hours of notification.
iv. Any changes anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to SU-ISERC within 72 hours.
v. Clearance for the export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to the expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days of completion of the study to SU-ISERC.

Before commencing your study, you will be expected to obtain a research license from National Commission for Science, Technology, and Innovation (NACOSTI) https://research-portal.nacosti.go.ke/ and obtain other clearances needed.

Yours sincerely,

Mr Ambrose Rachier,
Chairperson; SU-ISERC

Ole Sangale Rd, Madaraka Estate. PO Box 59857-00200, Nairobi, Kenya. Tel +254 (0)703 034000
Email admissions@strathmore.edu www.strathmore.edu

VT OMNES
VNVM SINT