MASTER OF TECHNOLOGY (INFORMATION SYSTEMS)

NILUTPOL BORA

2023

SECURING INDUSTRIAL IOT: GCN-BASED IDS IMPLEMENTATION AND A REVIEW OF TESTING FRAMEWORKS

A DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS

FOR THE AWARD OF THE DEGREE

OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted by

NILUTPOL BORA

2K21/ISY/16

Under the supervision of

Mrs. ANAMIKA CHAUHAN

Assistant Professor

Department of Information Technology

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of engineering)

Bawana road, delhi-110042

JUNE, 2023

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of engineering)

Bawana road, delhi-110042

CANDIDATE'S DECLARATION

I hereby declare that the work which is being presented in this dissertation entitled, **"SECURING INDUSTRIAL IOT: GCN-BASED IDS IMPLEMENTATION AND A REVIEW OF TESTING FRAMEWORKS"** in partial fulfilment of requirements for the award of the degree of **Master Of Technology** in **Information Systems**, submitted in the Department of Information Technology, Delhi Technological University, is an authentic record of our own work carried out under the guidance of **Mrs. Anamika Chauhan**, Assistant Professor, Department of Information Technology, Delhi Technological University.

The content embodied in this report has not been submitted by me for the award of any other degree or diploma.

Place: Delhi — NILUTPOL BORA

Date: — 2K21/ISY/16

DEPARTMENT OF INFORMATION TECHNOLOGY

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi college of engineering)

Bawana road, delhi-110042

CERTIFICATE

The dissertation titled **"SECURING INDUSTRIAL IOT: GCN-BASED IDS IMPLEMENTATION AND A REVIEW OF TESTING FRAMEWORKS"** is hereby approved as a creditable study carried out by Nilutpol Bora (2K21/ISY/16), and presented in a manner satisfactory to warrant its acceptance as a prerequisite to the degree for which it has been submitted. It is understood by this approval that the undersigned do not endorse or approve any statement made, opinion expressed or conclusion drawn therein but approve only for the purpose for which it has been submitted.

Place: Delhi

Date:

Mrs. Anamika Chauhan

Project Supervisor

Assistant Professor

Department of Information Technology

Delhi Technological University

ACKNOWLEDGEMENT

I am grateful to **Prof. Dinesh K. Vishwakarma**, Head of Department (Department of Information Technology), Delhi Technological University (Formerly Delhi College of Engineering), New Delhi and all other faculty members of our department for their astute guidance, and constant encouragement and sincere support for this project work.

I would like to take this opportunity to express my profound gratitude and deep regard to my project supervisor **Mrs. Anamika Chauhan**, for her exemplary guidance, valuable feedback, and constant encouragement throughout the duration of the project. Her valuable suggestions were of immense help throughout our project work. Her perspective criticism kept us working to make this project in a much better way. Working under her was an extremely knowledgeable experience for us

I would also like to appreciate the support provided to me by my seniors and my peer group who aided me with all the knowledge they had regarding various topics. I hope that this project will serve its purpose to the fullest extent possible.

Place: Delhi NILUTPOL BORA

Date: 2K21/ISY/16

ABSTRACT

Cyber-attacks on Industrial IoT systems can result in severe consequences such as production loss, equipment damage, and even human casualties and hence security is of utmost concern in this application of IoT. This thesis, presents an approach for network security, intrusion detection that utilizes the spatial attributes of a network in attempt overcome the limitations discovered through literature review of various studies in Intrusion Detection and testing frameworks. For this graph-based neural network have been used that was seen promising in modelling complex relationships between graphical entities, making them a suitable approach for IDS in interconnected systems. Our approach leverages a graph representation of network traffic, that is used as an input for neural network through the use of convolution operation. Our approach makes use of flow features of the network in relation with the neighbouring flows in contrast to other machine learning models that uses flow features independent to each other. This work has been evaluated primarily on Edge-IIoT 2022, dataset and compared with existing well-known datasets and machine learning methods. The results show that our approach achieved average 5.49% improved F1-score, compared with other standard existing methods with our model having highest F1-Score of 0.996. Further research and development in this area will advance the field of IIoT security and enhance the resilience of industrial systems in the face of evolving threats.

TABLE OF CONTENTS

Candidate's Declaration ..... **ii**

Certificate ..... **iii**

Acknowledgement ..... **iv**

Abstract ..... **v**

Table of Contents ..... **vi**

List of Figures ..... **viii**

List of Tables ..... **ix**

CHAPTER 1 INTRODUCTION **1**

1.1 Research Problem 2

1.2 Purpose of the thesis 3

CHAPTER 2 LITERATURE REVIEW **5**

2.1 Industrial Internet of Things 5

2.1.2 Security concerns in IIoT 5

2.2 Network Security through Intrusion Detection Systems 8

2.2.1 Related works in IIoT-IDS 8

2.2.2 Limitations observed 11

2.3 Spatial Network Intrusion Detection Systems 12

2.3.1 Spatial characteristics of network and Graph-based IDS 12

2.3.2 Related works in spatial NIDS 13

2.4 Testing Frameworks 16

2.4.1 Need for Testing Frameworks 16

2.4.2 Review of Testing Frameworks in IoT security domain 17

CHAPTER 3 METHODOLOGY **24**

3.1 Processing Network Traffic Flow Records 25

3.2 Transforming network flows to graph object 25

3.3 Model design 26
3.3.1 E-Graph convolution layer 26
3.3.2 Classification layer 28
CHAPTER 4 EXPERIMENTAL EVALUATION 29
4.1 Network traffic dataset 29
4.2 Evaluation metrics 30
4.3 Experimental results 31
CHAPTER 5 RESULTS ANALYSIS 35
5.1 Interpretation of experimental results 35
5.2 Limitations of this study 36
5.3 Significance and Implications for future research 37
CHAPTER 6 CONCLUSION 39
REFERENCES 40
LIST OF PUBLICATIONS 45

LIST OF FIGURES

Fig. Number	Figure Name	Page Number
Figure 2.1	Graph Representation of a Network	12
Figure 3.1	Flowchart of proposed architecture	24
Figure 3.2	E-Graph Convolution Model Design	27
Figure 4.1	Confusion Matrix for proposed model a) Edge-IIOT; b) CICIDS-18; c) UNSW-NB15 dataset	33
Figure 4.2	Experimental Results for Edge-IIOT dataset	33
Figure 4.3	Experimental Results CIC-IDS 2018 dataset	34
Figure 4.4	Experimental Results for UNSW-NB 15 dataset	34

LIST OF TABLES

Table Number	Table Name	Page Number
Table 1.1	Research Questions	2
Table 2.1	Related studies in IIoT Intrusion Detection Systems	12
Table 2.2	Related studies in spatial Intrusion Detection Systems	14
Table 2.3	Studies using Testing Frameworks in IoT security domain	18
Table 4.1	Characteristics of network datasets used	30
Table 4.2	Confusion matrix	30
Table 4.3	Evaluation Metrics	31
Table 4.4	Experimental Results	32
Table 5.1	Rationale and Impact of stages in proposed model	36

CHAPTER 1

INTRODUCTION

The industrial revolution brought in a shift in focus towards the benefits of interconnectivity between the physical industrial machinery. Automating manufacturing and supply chains has been possible through smart sensory devices that enables industries to make optimized decisions. Industrial Internet of Things is the step towards this revolution, a system of devices interconnected through a communication channel that serves the purpose of making intelligent decisions reducing the dedicated human interventions to optimize the performance of an industrial ecosystem. However, as in any network of connected devices, there is always a possibility of an adversary that can affect the optimal utilization of the resources, and evidently, the need to detect such disasters from happening proper measures are to be placed.

The initial step to securing a network is to prepare for any adversaries and detect any suspicious activities in the system, this detection of such sceptical incidents is known as intrusion detection, and the set of procedures for the process is known as the intrusion detection system. This study focuses on the detection of attacks occurring in a network of Industrial Internet of Things by proposing a methodology that can identify anomalies in the network among the benign traffic. Studies on intrusion detection systems are abundantly available, ranging from traditional signature and anomaly-based, simple machine learning to deep learning-based techniques. Improvements were made over the time by researchers include a combination of multiple models and optimizing preprocessing techniques to improve the intrusion detection performance.

It was noticed during our survey in the majority of papers was the use of network flow records independent to each other to identify the inference based on the features of the individual record, making no assumption of the relation among the records. Although machine learning and deep learning methods have evolved tremendously in the past few decades to be able to give good results using the assumption, without taking into consideration the relation among the different flow records, we feel the spatial characteristics of the network can help in better identifying underlying patterns of a

newer generation of attacks. The understanding of the network flows at a global level to have better knowledge of how the relationships among the flow define an attack pattern, as the flow in the network inherently characterizes many attacks. Recent studies (Section 2.3) have been published where spatial features were used to represent the information of the network in order to better classify between anomalous and non-anomalous data traffic, which have shown significant improvement in terms of the performance of the models.

1.1 Research Problem

To address the security concerns in Industrial IoT networks, this study explores the Intrusion Detection as a proactive defence mechanism against potential security breaches. The main focus of this study includes providing a comprehensive analysis on existing works on intrusion detection for IIoT and proposing an architecture that could mitigate the limitations of the previous works.

To achieve our problem statement, we have identified the following research questions which also sets the dissertation structure.

Table 1.1: Research Questions

RQ#	**Research Questions**
RQ1	What are security concerns in IIoT and are they similar to IoT attacks?
RQ2	What is the role of IDS in IIoT security and are there any limitations to existing literature?
RQ3	Can inclusivity of spatial features in intrusion detection improve performance?
RQ4	What are the methods/frameworks to evaluate the effectiveness of the solution?

RQ1: What are security concerns in IIoT and are they similar to IoT attacks?

The IIoT architecture can be divided into multiple layers based on the services provided by them, where network layer being the focus of this study, security challenges and impact of a breach in the network are analysed in conjunction with Internet of Things,

and if there are similarities between the vulnerabilities in IoT and IIoT. The various network attacks observed during our study has been discussed in section 2.1.

RQ2: What is the role of IDS in IIoT security and are there any limitations to existing literature?

Intrusion Detection Systems monitors network traffic for any abnormal incidents, researches have been done intensively in network IDS by various researchers to improve on detection rates. We aim to analyze these studies for impact of intrusion detection on securing a network through literature review done in section 2.2 and identify any research gaps to improve upon in these studies.

RQ3: Can inclusivity of spatial features in intrusion detection improve performance?

Based on the research gaps identified, this study progresses towards inclusivity of spatial features for improved detection; however, it is to be reviewed beforehand the work done in the field of spatial-based IDS and if it has any improvement over traditional methods. This is discussed in literature review of various graph-based IDS that utilizes spatial attributes of a network in section 2.3. This study further proposes a spatial-based IDS that is evaluated and compared for its effectiveness through different IIoT and IoT based datasets over other popular models. The model implementation and results are discussed in section 3 and 4.

RQ4: What are the methods/frameworks to evaluate the effectiveness of the solution?

The applicability of any proposed architecture cannot be comprehended primarily on the basis of the theories contemplated. In such scenarios a feasible solution is the virtual testing of a simulated scenarios. Section 2.4 reviews various methods/frameworks to evaluate the effectiveness of the solution.

1.2 Purpose of the Dissertation

The purpose of this dissertation is to contribute to the advancement of security measures in industrial IoT environments by developing and implementing a spatial-based IDS solution. This study provides a comprehensive overview and analysis of the implementation of securing Industrial Internet of Things (IoT) environments through Intrusion Detection as a proactive defence mechanism. This research aims to address the

need for enhanced security measures in industrial IoT systems by leveraging the capabilities of spatial features in intrusion detection through understanding of specifications of network security through extensive literature survey of existing work on IDS. The development and implementation of Edge-Graph Convolution Network IDS, utilizing the graphs to extract the hidden patterns from the network traffic graph that can effectively model complex relationships within graph-structured data, offers promising potential for detecting anomalous behavior and identifying potential intrusions in IoT networks. By leveraging the power of graphs, the proposed IDS aims to enhance the security posture of industrial IoT systems by accurately and efficiently detecting and mitigating potential security breaches.

In addition to the IDS implementation, this thesis also aims to provide a comprehensive review of testing frameworks used in the context of IIoT security. Testing frameworks play a crucial role in evaluating the performance and effectiveness of security solutions. By reviewing and analyzing existing testing frameworks, this research seeks to identify their strengths, limitations, and applicability in the context of industrial IoT security. This evaluation will enable researchers and practitioners to select appropriate testing frameworks for assessing the proposed solutions and gain insights into its effectiveness under various scenarios.

The outcomes of this research can inform the development of more robust and effective security strategies, ultimately enhancing the resilience and protection of industrial IoT systems in the face of evolving cyber threats.

CHAPTER 2

LITERATURE REVIEW

2.1 Industrial Internet of Things

The market for Internet of Things (IoT) has been on the rise since the arrival of low-cost devices accompanied by the growth of internet and internet-based services, where the reach of these devices is expanding over a varied range of applications from day-to-day necessities to some of the largest networks at smart cities. One such addition to these applications is the use of IoT in the industries integrated with industrial devices enabling real-time data exchange and analysis for optimization of overall throughput of the system. Industrial Internet of Things or IIoT can be said to be a sub domain of IoT where a system of devices is interconnected through a communication channel that serves the purpose of monitoring and analysing the information received from the sensory devices to make intelligent decisions based on the received data according to the requirement of the industrial systems reducing the dedicated human interventions so as to optimize the performance of an industrial ecosystem [1].

These systems are built over the existing industrial standards that enables the industries to take advantage of the internet-based services. Various devices being connected over the internet have empowered the organizations for improved management of these devices in the network. However, in regards to earlier generation of industrial systems connectivity to a wider global network comes forth with their own set of challenges, one such being the security of the data flowing in the network from outsiders. The following section describes challenges faced in protecting the IIoT network.

2.1.1 Security concerns in IIoT

Security concerns in Industrial Internet of Things (IIoT) environments are similar to those in general IoT but often have distinct characteristics and implications due to the critical nature of industrial systems. Some common security concerns in IIoT include:

- *Unauthorized Access*: IIoT devices and systems may be vulnerable to unauthorized access, allowing malicious actors to manipulate or disrupt critical

operations. This can lead to production losses, equipment damage, or safety hazards.

- *Data Breaches*: IIoT generates vast amounts of sensitive data, including proprietary information, customer data, and operational details. Data breaches can compromise confidentiality, integrity, and privacy, potentially leading to financial losses or reputational damage.
- *Malware and Ransomware Attacks*: IIoT devices can be targeted by malware or ransomware, which can disrupt operations, compromise data, or extort organizations for financial gain. Such attacks can result in downtime, financial losses, and operational inefficiencies.
- *Supply Chain Vulnerabilities*: IIoT systems often rely on a complex network of suppliers and vendors. Any vulnerability or compromise within the supply chain can impact the security and integrity of the entire IIoT ecosystem.
- *Lack of Security by Design*: Many legacy industrial systems were not originally designed with security in mind. This makes them susceptible to vulnerabilities and difficult to retrofit with robust security measures.
- *Interoperability and Standardization*: IIoT environments typically involve diverse devices and protocols, leading to interoperability challenges. Inconsistent security standards and protocols can create vulnerabilities and complexity in securing IIoT systems.

Industrial IoT architecture being closely related to IoT network systems, the security issues are similar to that of a IoT network, to identify the similarities and differences in the vulnerabilities of both systems, we have analyzed papers on security solutions in both IoT and IIoT systems. [2][3] The Industrial IoT and IoT do share similarities in terms of their architecture, means they also share some common security concerns as well:

- *Inadequate Security Measur*es: Both IIoT and IoT devices often suffer from inadequate security measures. They may lack proper authentication, encryption, or firmware updates, due to their limitations in hardware making them susceptible to unauthorized access, data breaches, and cyberattacks.
- *Device Heterogeneity*: Both IIoT and IoT ecosystems comprise a wide range of devices from different manufacturers, operating systems, and protocols. This

device heterogeneity introduces challenges in terms of standardization, compatibility, and security management.

- *Data Privacy Concerns*: Both IIoT and IoT generate vast amounts of data, raising concerns about privacy and data protection. Unauthorized access or mishandling of sensitive data can lead to privacy violations and potential misuse.

However, there are some key differences which makes an attack on IIoT system more dangerous [2][3]:

- *Critical Infrastructure*: IIoT typically involves industrial systems and critical infrastructure such as power plants, manufacturing plants, and transportation networks. The vulnerabilities in IIoT systems can have severe consequences, including disruptions to essential services and physical harm, making them high-value targets for malicious actors.
- *Legacy Systems*: Industrial environments often incorporate legacy systems that have been in operation for a long time. These legacy systems may lack modern security features and are not easily updated or replaced, making IIoT devices more vulnerable to attacks compared to IoT devices.
- *Impact of Attacks*: While attacks on IoT devices can cause significant harm to individuals and their privacy, attacks on IIoT systems can have broader implications. Disruptions in critical infrastructure or industrial processes can result in financial losses, operational downtime, and even impact public safety.
- *Security Prioritization*: Due to the potential consequences of attacks on IIoT systems, security is often a higher priority in industrial settings compared to consumer-oriented IoT devices. Industrial organizations typically invest more resources in securing their IIoT infrastructure, including implementing specialized security measures and conducting regular risk assessments.

While some security concerns in IIoT overlap with general IoT attacks, the consequences of security breaches in IIoT can be more severe. Industrial systems often involve critical infrastructure, such as power plants, transportation networks, or manufacturing facilities, which can have immediate and significant impacts on public safety, economy, and the environment. Therefore, securing IIoT requires specialized approaches that consider the unique characteristics, criticality, and operational requirements of industrial environments.

2.2 Network Security through Intrusion Detection Systems

Intrusion detection plays a significant role in network security by providing a proactive defence mechanism against unauthorized access, malicious activities, and potential security breaches. The significance of intrusion detection can be comprehended through the following key points [4]:

- *Early Threat Detection*: Intrusion detection systems (IDS) monitor network traffic, system logs, and user behavior in real-time, allowing for the early detection of potential security threats. This early detection enables prompt response and mitigation measures, minimizing the potential impact of security incidents.
- *Rapid Incident Response*: IDS can provide valuable information about the nature of the attack, compromised systems, and potential vulnerabilities. This helps in initiating a rapid incident response, allowing security teams to investigate and address the issue promptly.
- *Network Performance Optimization*: Intrusion detection systems not only focus on identifying and preventing security threats but also contribute to network performance optimization. By monitoring network traffic and analyzing patterns, IDS can identify potential bottlenecks, bandwidth utilization issues, and network anomalies that may affect the overall performance and efficiency of the network. This knowledge can be used to optimize network resources and ensure a smooth and reliable network operation.

Intrusion detection plays a crucial role in network security by providing early threat detection, facilitating rapid incident response. Implementing intrusion detection systems, industries can enhance their overall security posture, reduce the risk of security breaches, and protect their valuable assets and information from unauthorized access and malicious activities.

2.2.1 Related works in IIoT-IDS

In order to identify malicious actions and invasive behaviour in the system, the IDS can implement different methodologies, often categorised into: anomaly-based and misuse-based, where detecting any potential attacks via signature-based or misuse-based detection is the act of correlating signatures with observed events in order to identify

possible attacks, and identifying abnormal occurrences, anomaly-based detection compares observed activity with criteria of what is deemed normal. Intrusion detection systems generally follows one of these methods or combinations in order to identify intrusions, which has been further improved with the use of artificial intelligence as a tool that can help overcome shortcomings of each methodology.

Artificial intelligence being a recognized method that can extract patterns from a set of data, signature-based detection where, it is required for the system to have a set of recognized patterns to check from and anomaly-based detection where, the behavioural pattern of the network is to be known to the system, [4] an Artificial Intelligence model can extract such patterns without having explicit knowledge of the domain. This approach has been studied predominantly by the researchers, which has proven to outperform traditional ways given the relevant data to be processed and hence are considered for this paper. The studies involving AI-based methods for NIDS in IIoT or IoT are studied and tabulated in Table 2.1.

Table 2.1: Related studies in IIoT Intrusion Detection Systems

Paper	**Dataset Used**	**Algorithm Used**	**Results**
Awotunde et al. (2023) [5]	TON_IoT	XGBoost, Bagging, extra trees (ET), random forest (RF), and AdaBoost	Performed better than traditional methods with highest F-1 of 1.0 on XGBoost
Du et al. (2023) [6]	KDD CUP99, NSL_KDD, and UNSW_NB15	CNN-LSTM	Achieved higher accuracy for each dataset than individual model
Priya et al. (2021) [7]	WUSTL_IIOT-2018, N_BaIoT, and Bot_IoT	Ensemble Classifier	Achieved highest accuracy of 99.7%, and performed better than standalone classifiers.
Awotunde et al. (2021) [8]	NSL-KDD and UNSW-NB15	deep feedforward neural network	accuracy, FPR of 99.0%, 1.0%, for the NSL-KDD dataset, and 98.9%, 1.1%, for the UNSW-NB15 dataset

Kasongo et al. (2021) [9]	UNSW-NB15	Genetic Algorithm and Random Forest	Achieved 87.6% accuracy on proposed
Sarhan et al. (2021) [10]	UNSW-NB15, CSE-CIC-IDS2018, and ToN-IoT	deep feed-forward, random forest	Highest accuracy of 99.27% and 98.25% for respective datasets
Maharani et al. (2020) [11]	KDD Cup'99	Clustering and tree-based machine learning	Highest accuracy of 93% on K-means clustering
AL-Hawawreh et al. (2018) [12]	NSL-KDD and UNSW-NB15	Hybrid (Deep auto encoder and deep feedforward neural network)	Achieved highest detection rate of 99% and low false alarm rate (1.8%) on multiple attacks

The studies included in the table demonstrate the effectiveness of various approaches and techniques for detecting cyber-attacks in IIoT networks. The performances of the proposed models indicate the potential of IIoT IDS in improving the security of industrial systems. The papers studied were selectively tabulated in Table 2.1, so as to have diverse algorithms and datasets that are used in intrusion detection in the domain of IIoT. They range from machine learning techniques as in Maharani et al. [11] used tree and clustering based algorithms as K-means, Decision Tree to achieve an accuracy of 93% on KDD Cup'99 dataset.

Awotunde et al. [5] and Priya et al. [7] proposes an ensemble tree-based model for intrusion detection in IIoT networks. The model combines multiple decision trees-based ensemble models to improve accuracy and reduce false alarms. The authors evaluated their model on TON-IoT, a publicly available dataset and observed better performance compared through analysis on the results against base classifiers.

Du et al. [6], Awotunde et al. [8] and Sarhan et al. [10] proposed a deep learning-based approaches as CNN-LSTM and deep feed forward networks to classify network traffic as normal or malicious on IIoT networks. These models achieved high accuracies on datasets consisting of various network attacks.

Hybrid approaches where multiple algorithms were used in conjunction with one another, like Kasongo et al. [9] used GA as feature extraction method in addition to random forest for detection; whereas AL-Hawawreh et al. [12] used layered multi model approach to create a hybrid model that had high detection rate and low false alarm rate for datasets featuring multiple attacks.

These studies demonstrate the effectiveness of various approaches and techniques for detecting cyber-attacks in IIoT networks, indicating the potential of IIoT IDS in improving the security of industrial systems. However, it was observed though a lot of studies has been done on the field of intrusion detection, it was seen the lack of inclusion towards the topology of a network. This study aims to propose a network Intrusion Detection System that highlight the importance spatial characteristics in detecting the signature of the attack for which the following sections describe different papers studied which were utilizing spatial characteristics for intrusion detection.

2.2.2 Limitations observed

Despite the fact that machine learning approaches have achieved significant advances in the area of intrusion detection, the following issues still need to be addressed.

- The most widely used datasets for IDS for many years included KDD99, NSL-KDD, a more recent CICIDS 17 was also seen some studies. However, with the ever-changing networks, and newer attack types, makes the models of proposed studies on any older dataset questionable. There is a need of datasets, that are regularly updated.
- Decreased detection accuracy in real-world settings. Despite the fact that machine learning algorithms have a certain capacity to identify intrusions, they often do not perform well on data that is entirely novel to them. The majority of the available studies were carried out utilising labelled datasets as their starting point. As a result, good performance in real-world situations is not guaranteed.
- When it comes to practical IDSs, interpretability is critical for the detection made as the intrusions detected need to be analyzed by a human. An intrusion detection model which makes analysis difficult to understand, particularly deep learning models, having low interpretability, makes no sense. Every cyber security choice, on the other hand, should be taken with caution, not an unconvincing output result that can't be traced back to its source.

- Lack of consideration of spatial characteristics of the network and its changes over periods, for intrusion detection. Majority of the proposed models uses the rule-based detection or the inference is made on the network flow features to detect an anomaly, this though being able to draw fairly well conclusions on the majority of attacks, it performs poorly in detecting attacks that relies on multi flow strategies like DDoS attacks.

2.3 Spatial Network Intrusion Detection Systems

In the field of Intrusion Detection Systems (IDS) various approaches have been put forward by the researchers over decades, to keep up with the pace of ever evolving networks. However, there are few that takes into consideration the spatial features of a computer network. This study aims to propose an Intrusion Detection method that highlights the importance of spatial characteristics in detecting the signature of the attack, to which the following sections explores different research papers which were utilizing spatial characteristics for intrusion detection.

2.3.1 Spatial characteristics of network and Graph-based IDS

An Industrial Internet of Things network or a computer network in general, is a set of devices (nodes) in a plane that are connected with each other sharing resources over some communication protocol, however the topology in which the devices are connected can affect its throughput and reliability. The information on how these devices are connected and how they are dependent in the network helps in identifying the which device or traffic is affecting others in the network.

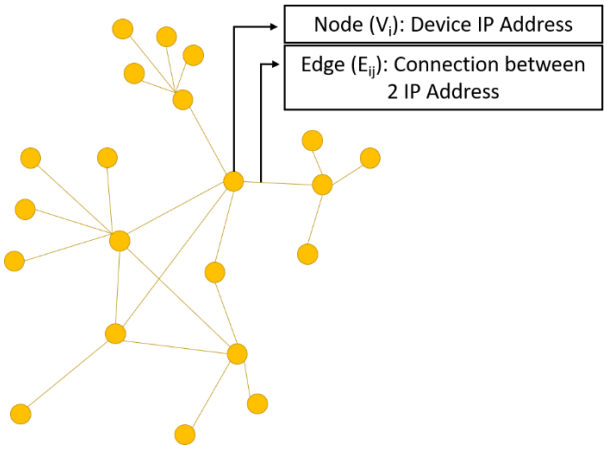

Figure 2.1: Graph Representation of a Network

This topological structure defining the dependency of devices on each other can be described as the spatial characteristics of the network. In its simplest form a computer network is nothing but a graph and hence an optimal way to represent the spatial characteristics of the network is a graph. The graph consists of vertices and edges, which can be mapped to the devices and its interconnection between them. We have used a graph object (G) to represent the spatial features where a node (V_i) in the graph is representing a host (IP Address) in the network and the edges ($E_{i,j}$) determines a connection between the corresponding hosts V_i and V_j. Graphs being data structures which can be essentially found in our everyday life, be it in transportation, molecular structures, the internet, and many more that represents a relationship among a set of entities providing the data structure much better expressive power than others. Researchers have been for the reason seen to be taking interest in graphs and combining it with artificial intelligence in the areas as social networks, molecular biology, transportation networks. [13]

The network traffic flow of the IIoT can be hence represented as a graph where the communicating devices can be given as the nodes and connections between the devices as the edges between the corresponding nodes in the graph. This allows us to have a structure for which meaningful information about flow of network traffic. The section ahead provides the literature review of the studies where the structural knowledge was used to classify an intrusion detection in a network.

2.3.2 Related works in spatial NIDS

Attackers have been evolving their methods for intruding the system, especially in cases of the Industrial IoT, that can become victims of espionage in the present day, where organizations are in ruthless competition for their share in the market. Attacks as distributed port scans, DNS amplification, botnet attacks, multi-flow attacks are more sophisticated in nature that methods as simpler machine learning models or deep learning models will fail to recognize. [14] The understanding of the network flow at a global level so as to have better knowledge on how the relationship among the flow can better define an attack pattern, as many attacks are inherently characterized by the flow in the network. The use of spatial features in intrusion detection systems has gained increased attention in recent years due to its potential to improve the accuracy of these systems.

Approaches to flag an intrusion varies from traditional signature and anomaly based, simple machine learning to deep learning based techniques, although it was noticed in the studies where the use of network traffic records were utilized independently, making no assumption of relation among the records performed comparatively lacking in accuracy than those using spatial characteristics of the network, to gather global perspective of the network. We have studied and tabulated in Table 2.2, this context of IDS to improve on the performance, different papers were discussed where novel approaches for incorporating spatial features into intrusion detection systems were used.

Table 2.2: Related studies in spatial Intrusion Detection Systems

Paper	**Dataset Used**	**Algorithm Used**	**Results**
Halbouni, A.H., et al. (2022) [21]	CIC-IDS2017	Convolution Neural Network	Achieved 99.55% detection rate on multiclass attack classification
Iacovazzi, A. et al. (2022) [15]	Mix-2022, Cui-2020	Graph-representation used in RF	Achieved 0.898, 0.846 macro F1 on multiclass attack classification
Zhu, H. et al. (2022) [18]	TON_loT, BoT-loT, and UNSW-NB 15	Line-Graph selective aggregation	Achieved 98.8, 99.9, 99.6 accuracy on respective datasets
Islam, R., et al. (2022) [16]	real rawCAN an OpelAstra dataset	Graph-based Gaussian naive Bayes	Achieved 98.1% and 99.57% on multiclass attack classification
Lo, Wai Weng et al. (2021) [19]	TON_loT, BoT-loT	Graph Neural Network selective aggregation	Achieved 1.0 and 0.87 F1-score in detecting different attack types
Chang, L. et al. (2021) [22]	UNSW-NB15, CIC-DarkNet, CSE-CIC-IDS, ToN-IoT	Graph-based Neural and Attention Network	Improved 2%, 3% F1-score for binary and multi attack over base model

Otoum, S., et al. (2020) [20]	NSL-KDD	DBSCAN	Achieved 95.6% accuracy on multiclass attack classification
Islam, R., et al. (2020) [17]	real CAN dataset	Graph-Based model	Achieved 97.53% accuracy on multiclass attack classification

Convolution Neural Network (CNN) is a technique that makes use of neighbouring information to give a spatial view of the network, this was used by [15] to utilize the topological information of the network in order to classify network traffic as benign or anomalous on CIC-IDS 2017 dataset. They were able to obtain high detection rate for various attack types in the dataset. However, CNN has drawbacks when it comes to represent a network topology, which were solved with the usage of graphs. Various researchers have used graphs in different fields [13] to detect anomalies in a system. In view of intrusion detection of a computer network, graph representation can be used with machine learning models as Gaussian Naive Bayes, Random Forest in [15][16][17], or deep learning models as in [18][19][20]. The graph being only a way of representing finds its applicability with various models has proved to improve on the performance of the system.

Graph is on the other hand is much computationally more expensive to process, and hence different researchers have been working to make it efficient for a wider range of applications. The study published by Lo et al. [19], proposed Graph selection and aggregation algorithm that classifies network graph edges or connections for anomalous behaviour, i.e., it can classify edges of a graph, where the edge features were also taking into consideration in the classification. This model was an extension to Graph SAGE network proposed by Hamilton et al. [23] where the authors extended the model to support edge features as well as classify them. The model works by creating a graph of nodes from the hosts and edges from the connection between them, having the same features as that of a network connection between the two communicating hosts. Liyan et al. [22] further proposed modified approach to E-Graph SAGE algorithm to add residual learning to this model aiming to improve on the minority attack classes performance by dealing with the high-class imbalance in datasets. They also proposed another algorithm E-ResGAT that uses attention mechanism in combination to the previous model, which resulted in better overall performance of the intrusion detection model as well as that of minority classes.

These studies demonstrate the potential of incorporating spatial features in intrusion detection systems to improve their accuracy. However, they also highlight some of the limitations of such approaches, such as the need for accurate information about the physical layout of the system, the requirement for a large amount of training data, and the need for significant computational resources. Nevertheless, the use of spatial features in intrusion detection systems holds great promise for improving the accuracy and effectiveness of these systems. The proposed approaches in the discussed papers offer new insights and potential solutions to the challenges of incorporating spatial features into intrusion detection systems.

2.4 Testing Frameworks

In the literature review from section 2.2, one of the limitations observed was the decreased detection accuracy in real-world settings. The cause for which seems to be inadequate testing of the model proposed by the researchers, to which we have explored different testing frameworks that are available in the market especially for the IIoT or IoT networks. Leveraging testing frameworks, industries can systematically assess the security of their IIoT systems, identify vulnerabilities, and implement necessary measures to enhance the overall security and resilience of their IIoT deployments.

2.4.1 Need for Testing Frameworks

With the advances in IoT technology and its improved accessibility made the IoT based devices market thrive now more than ever. However, with this new booming market, the companies have been facing the problem of substantiating the device's reliability, with more than millions of linked devices working simultaneously ensuring their expected performances is challenging when there are no set standards for testing strategy with a vast variety of IoT devices available. An IoT device which can range over healthcare devices, industrial equipment, smart homes, toys, etc. not necessarily following the same set of protocols, traditional approaches of testing over sets of input and validating the outputs becomes inefficient. Having varying combinations of sensors data and architectures can generate near impossible sets of testing inputs. Some of the key challenges in this are:

- Connecting and implementing a universal testing case scenario for multiprotocol devices can be challenging when the market for IoT based devices are ever evolving.
- Different devices in the market uses different deployment standards, configurations, software, power management systems, etc. which make a hardware-based simulation difficult.
- Testing thousands of linked devices in an IoT architecture for security vulnerabilities is a challenge when the cost of missing them could be very dangerous.

The complex and scalable nature of an IoT infrastructure needs to provide a comprehensive yet flexible approach to test the performance metric of such systems. This is where the virtual testing environments can come handy for IoT systems. A simulated real-world scenario capable of handling multiple linked devices, of varying protocols, can help assist solving these challenges. In this paper we have discussed different tools and platforms available in the market be open sourced, or commercially available, and the different evaluations per- formed on them by various researchers.

2.4.2 Review of Testing Frameworks in IoT security domain

The focus on research for better IoT algorithms, protocols, or techniques have been more prominent now than ever, however, such researches are not always viable to implement using the traditional hardware, countering which solutions based on simulations, multiple tools have been proposed over the years to simulate the process on a virtual environment to back up the theories proposed by the researchers.

One of the key concerns, nonetheless remains is securing the IoT network even with multiple studies proposing various security techniques as the applicability of it cannot be comprehended primarily on the basis of the theories contemplated. In such scenarios a feasible solution can be the virtual testing of a simulated real-world scenarios, for which Patel et al., [24] carried out a comprehensive review comparing such different tools available for IoT over different parameters and broadly classified them over simulators, emulators, and testbeds.

In this section, the different IoT tools that were found to be focusing on the security field of IoT are discussed briefly. The section is categorised as IoT simulators, emulators and testbeds, where the tools are briefed and related studies are analysed as to what different IoT tools provide in context of security. This comparison aims to help researchers and developers in choosing a better suited tool according to their needs.

Table 2.3: Studies using Testing Frameworks in IoT security domain

Tools/ Platform	Refd. paper	Scope of paper	Type of network	Attacks	Paper description
Simulators					
Ns-3	Siddiqui et al. [25] 2021	Performance Analysis	7,15 MANET Based IoT	Blackhole and Wormhole Attack	compared the network affectibility under attack using NS-3
	Wu et al. [26] 2020	Intrusion Detection on constrained resources	240 node IoT network	Energy Exhaustion Attacks	hybrid IDS for IoT, experiments performed on NS-3
OMNet++	Gupta et al. [27] 2018	ensure security of data communication	40 node IoT network	N/A	Blockchain consensus model for data transmission, evaluation on OMNet++
	Alnuman et al. [28] 2020	DDoS detection	100 node IoT network	DDoS	Machine learning based DDoS detection, validation on OMNet++
QualNet	Govindasamy et al. [29] 2018	Performance Analysis	50 node IoT network	Wormhole Attacks	Analysis of various routing techniques in presence of wormhole attacks
	Almomani et al. [30] 2017	Performance analysis	50 node IoT network	N/A	Implementation and performance analysis of proposed routing protocol in Qualnet

TOSSIM	Sedjelmaci et al. [31] 2016	Anomaly Detection	300 node IoT network	DoS	Lightweight anomaly detection for IoT, demonstrated the viability using TOSSIM
Emulators					
Cooja	Ioulianou et al. [32] 2018	Intrusion detection	7 node IoT network	DoS	Signature-based IDS for IoT, and evaluated on Cooja
	Aiash et al. [33] 2016	Specification-based IDS	100 node IoT network	RPL topology attack	Intrusion detection for RPL based network, validated on Cooja
	Yavuz et al. [34] 2018	Routing attack detection	1000 node IoT network	Decreased Rank, Hello Flood, Version Number	deep-learning based continuous security monitoring analysis for IoT
NetSim	Prasadh et al. [35] 2019	Efficiency Analysis	10 node IoT network	Jamming attacks	Anti-jamming techniques efficiency were analyzed on NetSim network
	Remesh et al. [36] 2020	Intrusion Detection	IoT network	DoS, DDoS, Botnet	Network performance was analyzed using NetSim during intrusions
NCTUns 6.0	Saeedi et al. [37] 2019	DDoS detection and mitigation	IoT network	DDoS	Proposed machine learning based model was tested in emulator
Test-beds					
Fit-IoT lab	Antonio et al. [38] 2020	RPL security improvement	IoT Test-bed	N/A	Evaluation of security mechanisms in RPL protocol and experimental analysis on testbed
	Khadr et al. [39] 2020	Performance validation	IoT Test-bed	Jamming attacks	Performance validation of proposed model for CR-IoT applications under jamming attacks performed on testbed

Smart Santander	Sidra et al. [40] 2016	Security threats analysis in smart city	IoT Test-bed	Physical, data, software attacks	Various security threats and possible solutions in the testbed based smart city were analyzed

Simulators

- ***NS-3***: The NS-3 is a network simulator designed typically for the research community; the simulator can be used for wide range of simulations for replicating the perceptual layer of IoT. Being a network simulator, it lacks some features for simulating the networks for IoT, though the support can be added through manual extensions. NS-series simulators though mainly seen for modelling generic network structures, in recent studies IoT simulations were also observed as Wu et al. [26] used to simulate their work on intrusion detection systems where they were able to detect as well as trace malicious nodes in the network, where the model proposed was focused on detecting energy exhaustion attacks on a 240 node IoT network modelled on NS-3. MANET based IoT was also simulated in a study by Siddiqui et al. [25], to analyze the effects of blackhole and wormhole attacks on a low powered IoT network, this models' reliability was supported by the simulations performed on a NS-3 modelled network of 7 and 15 nodes respectively.
- ***OMNeT++:*** OMNeT++ is a free non-commercial simulation tool majorly used for building network simulations. Some of the papers that used this tool in their research implementation were seen as in simulating DDoS detection in IoT, where Alnuman et al. [28] used OMNeT++ to represent a 100-node home network to evaluate their algorithms accuracy. In another paper by Gupta et al. [27] blockchain applicability in IoT to secure the data transmission was also simulated using this tool, this approach was tested on a 40 node IoT network to simulate the process of data communication. However, one of the concerns is the limited number of built-in protocols supported, which can be solved by using various manual extensions available according to the user needs.
- ***QualNet:*** The QualNet simulator is a commercial version of Glomosim, primarily supporting built-in ZigBee protocol. IoT security-based papers where QualNet was implemented were as Ahuja et al. [30] where the authors studied the effect of wormhole attack on routing protocols on a 50 node IoT network to support their

proposed model. Apart from intrusion detections, researchers have also used QualNet for performance analysis, as Govindasamy et al. [29] proposed a study comparing the performance of different hybrid routing protocol in IoT and verified their stance by simulating them in a 50 node IoT network in QualNet simulator.

- ***TOSSIM:*** TOSSIM is an IoT (Internet of Things) simulator designed primarily for simulations of TinyOS smart devices. Even though TOSSIM is majorly used to simulate TinyOS applications, still some researches where TOSSIM was implemented in security field for internet of things one such study by Sedjelmaci [31] follows anomaly detection techniques in low-resource IoT devices and the proposed model was simulated on a 300 node IoT network using the TOSSIM simulator.

Emulators

- ***Cooja:*** Cooja is an emulator that is accessible in the Contiki operating system (OS), which is one of the more popular OSs for programming IoT sensors. This makes modelling a network for IoT is much easier thanks to the access of majority of standards and protocols provided by Contiki, allowing researchers to recreate or model simulations faster. Another point for researchers to consider Cooja is the ability to directly transfer simulations to physical models with minimum efforts. Cooja was seen to be implemented in multiple studies to create the virtual network model, where researchers used it to support their work in the IoT security as; intrusion detection systems for Internet of Things network were proposed by Ioulianou et al. [32] for DoS attacks detections which was designed using Cooja and evaluated on 7 node IoT network; Anhtuan et al. [33] for routing attacks detection on a 100 node IoT network; and a deep learning-based monitoring of routing protocol attacks by Yavuz et al[34], where the study validated their results on a large scale 1000 node IoT network modelled on Cooja.
- ***NetSim:*** NetSim is an IoT (Internet of Things) network emulation tool for protocol simulation and security applications. It covers a wide range of protocols for simulating IoT devices, and sensor networks. NetSim was seen to be implemented in various studies to create the virtual network model, where researchers used it to support their work. Regarding the security in IoT, intrusion detection systems were proposed by researchers as Prasadh et al. [35], where efficiency of anti-jamming techniques proposed were analyzed using the NetSim before and after the attacks to compare the affect of attacks on a 10 node IoT network; in another paper by Athira

et al., [36] proposed an architecture for detecting DoS, DDoS, and botnet attacks, which was evaluated using NetSim and the impact of these attacks on the network was analyzed.

- ***NCTUns 6.0:*** NCTUns 6.0 is an open-source network simulator cum emulator. It was seen to be implemented in various studies to create the virtual network model, where researchers used it to support their work. Regarding the security in IoT, intrusion detection systems were proposed by researchers as Kubra et al., [37] to evaluate their DDoS detection model.

Testbeds

- ***FIT IoT-LAB:*** FIT IoT-LAB is one of the more popular experimental test-bed for testing out a wide scale IoT or embedded device. The testing environment features more than 200 mobile robots and 3000 IoT nodes. It was seen to be implemented in various studies to solidify the proposal, in terms of the security in IoT, Antonio et al., [38] proposed improvement of RPL security scalability and used FIT IoT-Lab for the experiments to evaluate the efficiency of the proposed technique. Another study by Khadr et al., [39] used FIT IoT-LAB to validate their algorithm against jamming attacks on IoT devices.
- ***SmartSantander:*** SmartSantander is a testbed for evaluating IoT applications in smart city field. Consisting of more than 20000 IoT devices having sensor nodes, RFIDs, etc. is one of the biggest testbeds for smart city domain. The major advantage of SmartSantander is the variety of sensors that allows researchers from different area of interest in IoT can make use of the testbed. It was seen to be implemented in various studies majorly in smart city-based research. Regarding the security, Shah et al., [40] conducted a thorough review over the SmartSantander testbed security concerns and proposed the viable solutions to them.

The various simulations tools/frameworks used in the security domains of IoT (Internet of Things) have been discussed and related studies have been explored, categorizing them into simulators, emulators and test-beds. The choice of the framework to be used by the researcher for their study, whether it be a simulator, emulator or test-bed depends upon the IoT system, the level of simulation required by them.

A simulator can be used for Internet of Things research based on the scope of the study, i.e., it is better suited for studies where an initial abstract model of the IoT network is adequate. Instead of a complex physical system of IoT devices, a simulator is used to design a much ideal model of the IoT network much easily and effectively. These models help researchers quickly design the proposed architecture or technique and analyse the logic flow and if the proposed theory is actually a viable solution that should be developed further for deployment, that is simulators are useful as a proof-of-concept tool. An IoT network simulated in a simulator helps to figure any semantic flaw in the algorithm, which can reduce wastage of resources. As in case of a real hardware implementation of a multi node IoT network to test a theoretical approach can lead to greater expenditure of resources, which could be prevented using a simulator. However, the results from simulators are very much an ideal scenario and hence are not always a reliable standard for determining the performance of this model deployed in a real-world scenario.

A better alternative for researchers to look for if the focus is toward creating simulations closer to practical networks as well as features the advantages of simulators as configurability, scalability, control of the network, the middle ground could be the emulators. An emulator maps real IoT devices to corresponding simulated devices, executing parallel to real Internet of Things nodes. This helps researchers to port their IoT system or architecture directly to real world IoT system with minimal changes, even so the results produced from simulations in an emulator are more reliable than those of a simulator.

While both simulators and emulators are helpful when researchers are working limited scope or are in the initial stages of development of IoT system, the researchers working in the later stages, they expect results that are more practical to better optimize their systems for real world. The IoT test beds are the better alternative to simulators and emulators when it is not viable for physical IoT devices yet they require more reliable experimentation results. These test-beds provides access to researchers to a variety of readily accessible IoT network to conduct their experiments. However, they lack the configurability, scalability, control over the network as in case of simulators and emulators.

CHAPTER 3

METHODOLOGY

The proposed model of Edge-Graph Convolution model is a modification of Graph Convolutional Network where, in a Graph Convolutional Network, the term convolution is used to describe the operation that propagates information across the nodes of a graph. The concept of convolution in GCNs is inspired by traditional convolutional neural networks (CNNs) used for image processing, but adapted to work with graph-structured data. The graph convolution is used to generate better informed representation of data by considering the information of not only the node itself but its neighbours as well. This information can help in finding interesting hidden patterns in the spatial domain of the network that may not be inferenced otherwise using other deep learning models.

The outline of the proposed model is demonstrated in the Figure 3.1. The flow from the network traffic flow records to the classification of traffic records into benign and attack data is divided into four stages as: Data preprocessing, creation of graph object, E-Graph Convolution model, and classification of edges for final output. These steps have been explained in the following subsections:

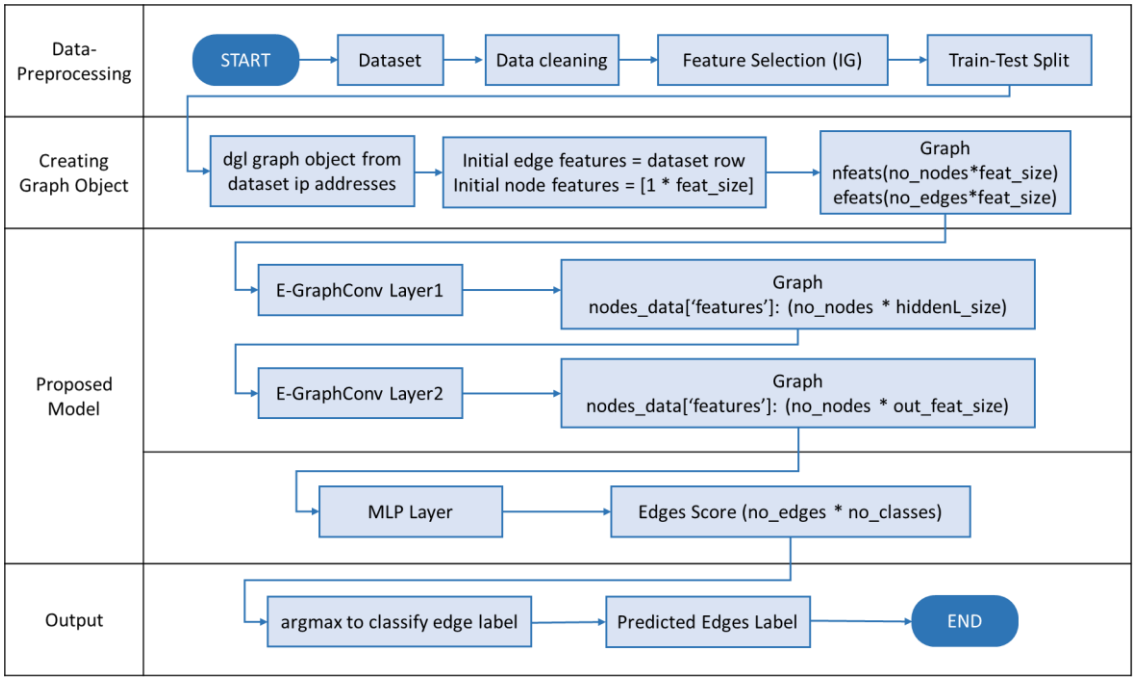

Figure 3.1: Flowchart of proposed architecture

3.1 Processing Network Traffic Flow Records

The traffic in a computer network is stored in the form of data flow, where a given flow identifies the source and destination of the data, and other fields that explains the data flow as flow duration, data size, protocols, etc. However, the recording of a network flow is not always perfect and some faults may occur during the process. If this unchecked data is forwarded to a learning model, the model can make incorrect inferences from such unintended variations in the traffic data and provide unreliable output, so as to avoid such cases the dataset is cleaned for any faulty data as missing, corrupted, or duplicate records. Further feature selection was performed on the dataset as for a given network traffic flow a record can have as much as hundreds of features, however there are only much that are relevant considering our problem statement. Selective features were used to train our model using Information Gain which provided with a ranking of features of a dataset, from this the features contributing to retain ninety percent of the information were used, which removed redundant features that only adds to model complexity. This processed data was split into train and test sub datasets for training the model and testing the performance of the model.

3.2 Transforming network flows to graph object

The network traffic as records of data flow is more widely used technique to capture the traffic where each record in the dataset is the set of attributes between the receiver and the sender IP addresses. However, being a set of sequential records, this kind of dataset is not ideal to capture any underlying spatial patterns that are more relevant to detecting attacks that uses multi flow strategies to deploy attacks on the network. Transforming the initial dataset into a graph object represents the data closer to real world network traffic where a node in the graph is representing a host (IP Address) in the network and the edges determines a connection between the corresponding hosts.

The dataset is to be first transformed into a graph object G (V, E, X_V, X_E), where V is the set of nodes (hosts), E is set of edges, X_V is set of features of node V, X_E is set of network features of edge E connecting the node V_i and V_j. To construct the graph the distinct set of private IP addresses in the dataset along with single merged IP address of external IP determines the set of nodes in the graph having initial features as zeroes of size that of number of attributes in the dataset. The edges between the nodes have features set of attributes of the record connecting the hosts. This translates our problem of intrusion

detection in a network to that of edge classification, where we have to determine whether an edge in the network graph is anomalous or not, i.e., a binary classification of the edges.

3.3 Model design

The proposed model of E-Graph Convolution model is a modification of Graph Convolutional Network where, in a Graph Convolutional Network, the term convolution is used to describe the operation that propagates information across the nodes of a graph. The concept of convolution in GCNs is inspired by traditional convolutional neural networks (CNNs) used for image processing, but adapted to work with graph-structured data.

In image-based CNNs, convolution involves applying a filter or a kernel to a local receptive field of pixels in the input image. This filter performs a dot product with the pixel values in the receptive field, producing a new feature representation that captures local patterns and spatial relationships. In GCNs, convolution is adapted to work with graph structures rather than regular grid-like image data. Instead of convolving with a fixed filter over local patches, GCNs perform convolution by aggregating and transforming information from neighbouring nodes in the graph.

The proposed model of E-Graph Convolution model is a simple neural network which is fed the transformed graph dataset, where we have incorporated edge features to be utilized in the model in contrast to graph neural networks that inherently works with nodes. This allows the model to learn from its edge features as the model trains, on which the final classification is performed through a Multi-Layer Perceptron layer. The proposed model can be sub divided into two parts as: e-graph convolution layer and classification layer shown in figure 3.1.

3.3.1 E-Graph convolution layer

E-graph convolution layer is the fundamental block of our model that makes use of information propagation to capture and encode the relational dependencies among edges in the graph. It operates on a graph structure and performs message passing and linear operation to update edge representations based on the information from neighbours.

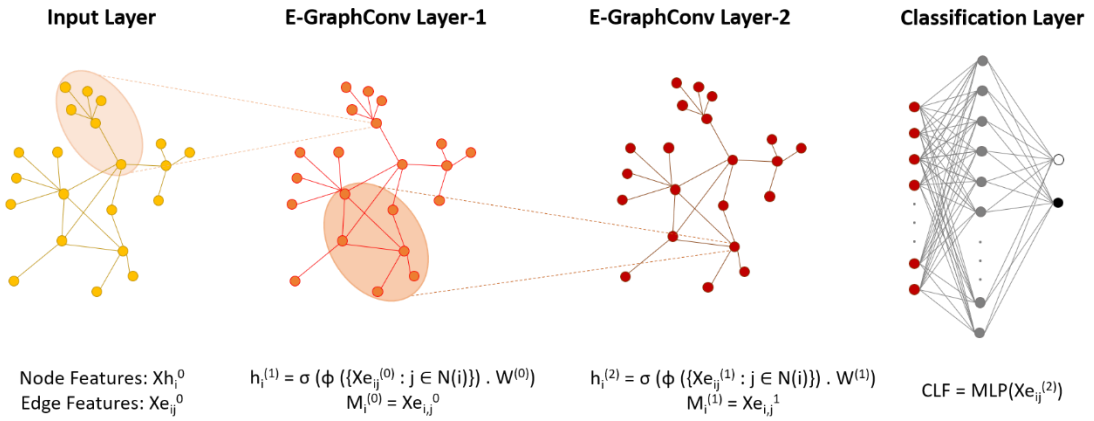

Figure 3.2: E-Graph Convolution Model Design

Our model uses the transformed graph G with multi-dimensional edge features, in order to learn from these features, the E-graph convolution layer uses update function and message passing function.

- *Update Function*: The update function computes the updated edge features based on the messages from neighbours and applies a non-linear activation function. The update function can be represented as:

$$h_i^{(l+1)} = \sigma(\varphi(\{Xe_{ij}^{(l)} : j \in N(i)\}) . W^{(l)})$$

where:

- $h_i^{(l+1)}$ is the updated feature vector for node I at layer $l+1$.
- φ is the linear transformation function the model uses to represent the information of the neighbouring nodes, for our proposed model we have used the mean of information from neighbouring nodes to create a summarized representation for each node.
- $Xe_{ij}^{(l)} : j \in N(i)$ represents the set of edge features from neighbouring edges.
- $. W^{(l)}$ represents the weight matrix that maps the aggregated information to a new feature space. The aggregated information is transformed using learnable parameters to generate new node features.
- σ is the nonlinear activation function applied element-wise to introduce non-linearities into the transformed node features. In our model we have used ReLu to introduce non linearity to the layer.

- *Message Passing Function*: Each edge in the graph uses information from its neighbouring nodes to update the edge representations which are passed using the message passing function. The message passing function can be represented as:

$$M_i^{(l)} = Xe_{i,j}$$

  where:

  - $M_i^{(l)}$ is the message sent from edge between nodes i and j at layer l to its neighbouring edges.
  - $Xe_{i,j}$ is the edge feature between nodes i and j at layer l.

Through stacking multiple layers, the model can capture information from multiple hops in the graph, allowing for the modelling of complex relationships and dependencies. Each layer propagates and aggregates information from neighbouring nodes, refining the node representations with each layer. By repeatedly applying this convolution operation across multiple layers, the model can capture increasingly complex patterns and dependencies in the graph structure, allowing for tasks such as edge classification. However, stacking up too many layers can result in traversing the whole graph making the information saturated and results unreliable.

Our proposed model uses two e-graph convolution layers, which allows the model to learn through the features of two hop neighbours of the network graph.

3.3.2 Classification layer

Multi-Layer Perceptron, or MLP layer is stacked in addition to our E-Graph Conv layer as the resulting edge embeddings from the above layer are of size the number of attributes in the network traffic flow, in order to classify them, scores are to be given to these edges. The MLP takes these inputs to update all the edges in the graph to a vector of size two, which implies the score towards the edge being benign or anomalous, i.e., the MLP layer classifies the edge embeddings into two classes, which can be used to compare with the labels provided in the network traffic dataset and tune the model during back propagation of the model.

CHAPTER 4

EXPERIMENTAL EVALUATION

The experimental evaluation of the proposed methodology for edge-based graph convolution network is performed on an Industrial IoT dataset along with other network traffic datasets, to substantiate our assumptions of improved detection rates while considering spatial knowledge of the IoT network. The experiments conducted to evaluate the performance of our model has been carried out in Google Colab tool with their free-to-use GPU, with Python as the coding language. This section presents the different Internet of Things network traffic datasets, performance measures used for evaluation and the results obtained for the method proposed.

4.1 Network traffic dataset

For our experiments on the methodology proposed, we require network traffic data. There are many datasets available for researchers to work on this field, however among the different network flow-based datasets, the dataset needs to be relevant to the scope of the paper. The parameters for selecting a dataset for our model was as follows:

- The network traffic is from a network of Industrial Internet of Things or Internet of Things devices, preferably having a heterogenous set of devices in the network.
- The dataset offers a varied range of attacks to better evaluate the model. The composition of the datasets selected has been listed in Table 4.1, that shows a variety of attacks available in them.
- The data streams collected are to be from an up-to-date network that uses current network methodologies.
- The dataset must have source and destination identifiers, like IP Addresses in order to create the graph object, as the graph object created have edges as traffic records between two hosts.

Based on the characteristics desired for evaluating our model the following datasets have been selected, as given in Table 4.1.

Table 4.1: Characteristics of network datasets used

Dataset	**No. of records**	**No. of classes**	**No. of Features**	**Attack Types**
Edge-IIoT [41]	2.2M	15	63	Backdoor, DDoS (HTTP, ICMP, TCP, UDP), Fingerprinting, MITM, Password, Port scanning, Ransomware, SQL, Uploading, Vulnerability scanner, XSS
CICIDS 2018 [42]	16.2M	7	80	BruteForce, Botnet, DoS, DDoS, Infiltration, Web Attacks
UNSW-NB15 [43]	2.5M	10	43	Analysis, Backdoor, DoS, Exploits, Fuzzers, Generic, Reconnaissance, Shellcode, Worms

4.2 Evaluation metrics

The performance of the proposed model is determined using the metrics used for evaluating the classification-based machine learning model. This is in relation to the model proposed in this study represents the output in the form of a boolean i.e., benign or attack, which is nothing but binary classification of the network traffic fed to the model. The performance metrics hence described using the confusion matrix for each dataset, using accuracy, defined as the ratio of number of correctly classified traffic to the total number of network traffic records; precision, as the correctly identified attacks to number of identified attacks in the traffic; Detection Rate as the number of correctly identified attacks to total number of attacks in the traffic; and F1-score as the harmonic mean of precision and Detection Rate.

Table 4.2: Confusion matrix

	Predicted Negative (0)	**Predicted Positive (1)**
Actual Negative (0)	True Negative (TN)	False Positive (FP)
Actual Positive (1)	False Negative (FN)	True Positive (TP)

Table 4.3: Evaluation Metrics

Performance Metric	**Equation**
Accuracy	(TP + TN) / (TP + TN + FP + FN)
Precision	TP / (TP + FP)
Detection Rate	TP / (TP + FN)
F1-score	2*(Precision*Detection Rate)/(Precision+Detection Rate)

where, in the equations:

- TP: True Positives (the number of correctly predicted positive instances)
- TN: True Negatives (the number of correctly predicted negative instances)
- FP: False Positives (the number of incorrectly predicted positive instances)
- FN: False Negatives (the number of incorrectly predicted negative instances)

4.3 Experimental results

In this section, we compare the performance of different models against our proposed model against various performance metrics mentioned in section 4.2. The dataset selected, Edge-IIoT (2022) [41], CICIDS-18 (2018) [42] , UNSW-NB15 (2015) [43-47] as in accordance to the requirements described in the previous section were served as input to the model, where each dataset was preprocessed to remove redundancies and noise and scaled using minmax to normalize the data values between zero and one. The data was normalized as the proposed model is using message passing function that aggregates the neighbouring edge feature values in intermediate steps, having high values could lead to overflow in certain scenarios.

The obtained processed data is given to the model in the form of a graph to train the model, as well as test the effectiveness of the model using the performance measures described in the above sub-section. It is also to be noted datasets used in this paper includes label in terms of the data being benign or anomalous in nature, which reduces the output as a binary classification problem and hence the model proposed is evaluated and compared in terms of performance metrics relevant to classification.

Table 4.4: Experimental Results

Dataset	Algorithm	Accuracy	Precision	Detection Rate	F1-Score
Edge-IIoT	**Proposed**	**0.997**	**0.998**	**0.997**	**0.996**
	CNN	0.95	0.97	0.91	0.94
	MLP	0.95	0.96	0.91	0.93
	RF	0.95	0.94	0.93	0.94
	DT	0.95	0.97	0.91	0.94
CICIDS 2018	**Proposed**	**0.986**	**1.0**	**0.973**	**0.986**
	CNN	0.98	0.99	0.98	0.98
	MLP	0.95	0.88	0.9	0.89
	RF	0.98	0.99	0.93	0.95
	DT	0.94	0.9	0.8	0.84
UNSW-NB15	**Proposed**	**0.99**	**0.997**	**0.982**	**0.990**
	CNN	0.98	0.99	0.98	0.98
	MLP	0.97	0.9	0.76	0.81
	RF	0.99	0.94	0.92	0.93
	DT	0.98	0.87	0.89	0.88

The performance measures obtained for the dataset is tabulated in the Table 4.4, and its confusion matrix for the testing portion is given in figure 4.1. To compare our achieved results the same dataset has been trained and tested for some of the well-known classification models as Naïve Bayes, Decision Tree, Random Forest, Multi-Layer Perceptron. The evaluation results have also been graphically represented in the figure 4.2-4.4, substantiating our model performance to be better than that of the popular classification algorithms.

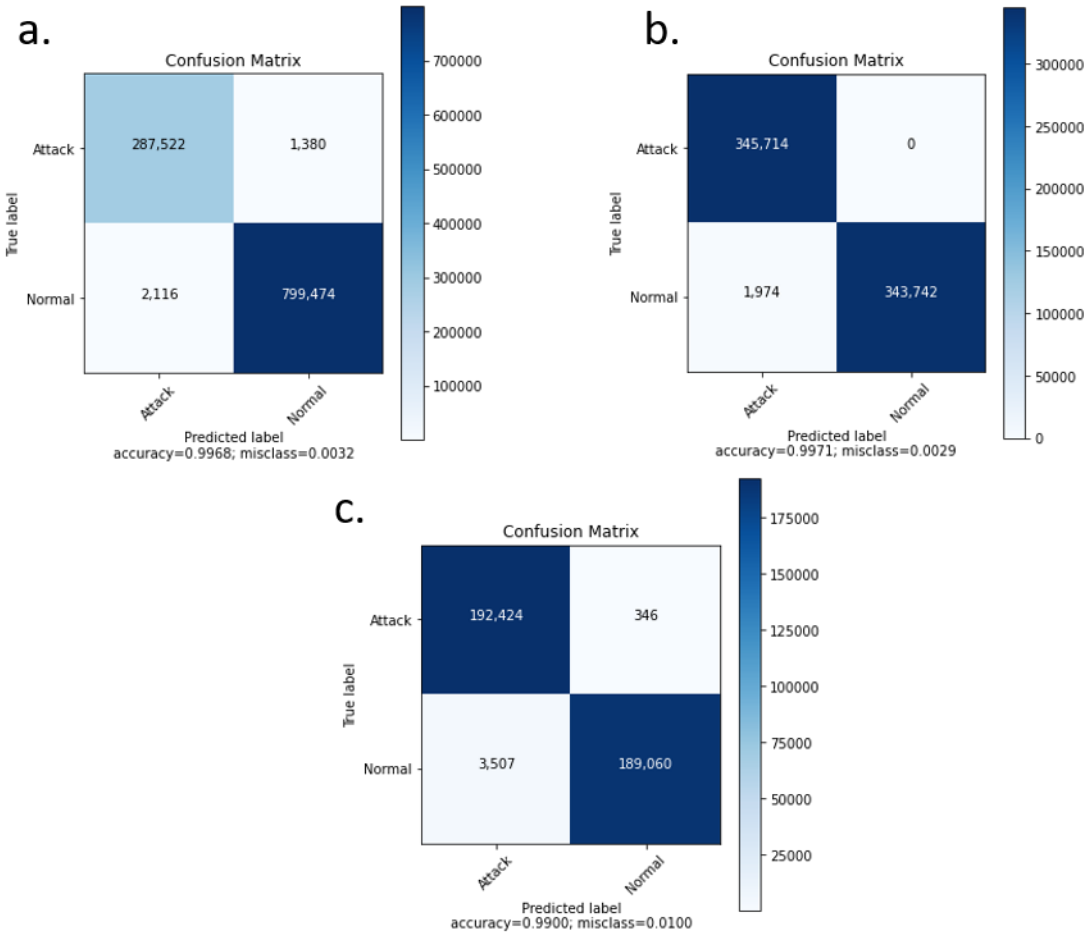

Figure 4.1: Confusion Matrix for proposed model a) Edge-IIOT; b) CICIDS-18 c) UNSW-NB15 dataset

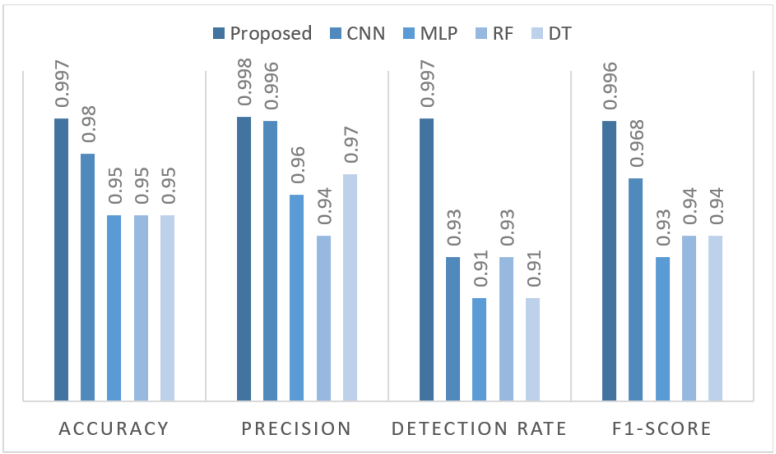

Figure 4.2: Experimental Results for Edge-IIOT dataset

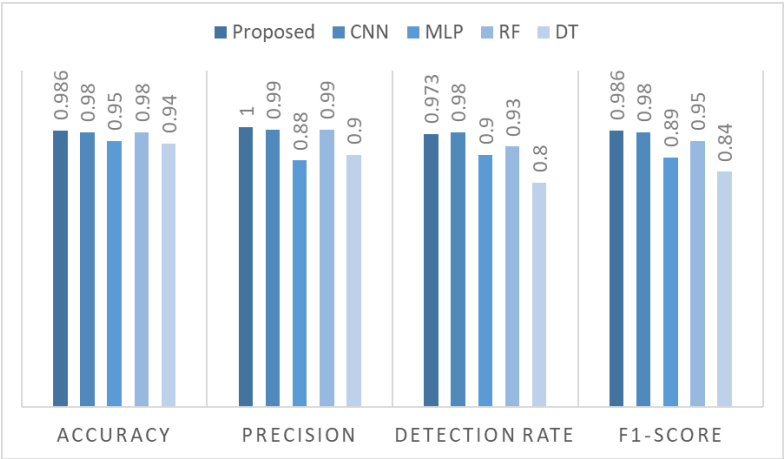

Figure 4.3: Experimental Results CIC-IDS 2018 dataset

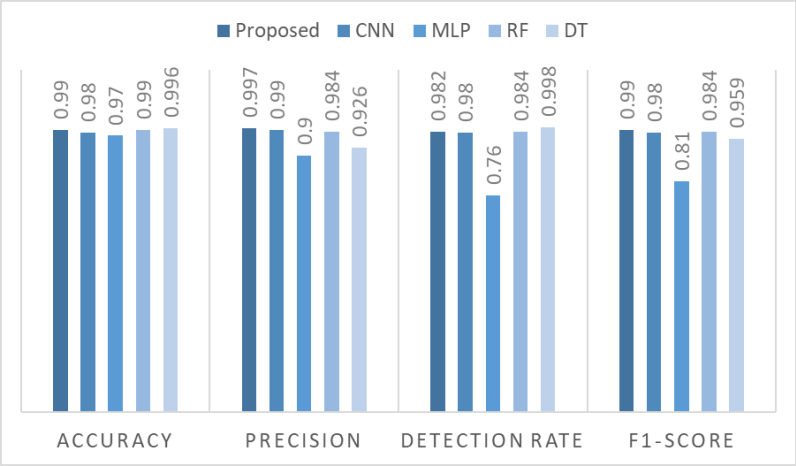

Figure 4.4: Experimental Results for UNSW-NB 15 dataset

CHAPTER 5

RESULTS ANALYSIS

5.1 Interpretation of experimental results

The key findings inferenced from the results are as follows:

- E-Graph convolution has demonstrated best performance in comparison with the highest F-1 score of 0.996, leveraging its ability to effectively capture and model graph-structured data. Its success can be attributed to its unique architecture, which combines graph convolutional layers with non-linear activation functions to extract meaningful representations from graph data. The incorporation of edge features aggregation and feature propagation further enhances its expressive power.
- CNNs are generally better at tasks that require local processing, such as image classification, whereas graph-based models as ours are generally better at tasks that require global processing. Its ability to capture both local and global graph structures enabling it to learn hierarchical representations of the underlying graph, verified from 2.89% improvement of F1 score and 7.2% detection rate. This hierarchical modelling facilitates the understanding of complex relationships and dependencies among graph elements, leading to enhanced predictive accuracy.
- Notably, E-GCN's performance surpassed traditional machine learning techniques in terms of accuracy, predictive power, and generalization ability, with 5.9% increase in tree-based model, and 7.1% in neural network model. This suggests popular models as Random Forest, Decision Tree, MLP which usually works on linear data structures struggles to exploit this structural information.

The different stages, data processing, graph creation and E-Graph convolution model of the proposed architecture have contributed towards the performance, the rationale and impact of stages are described in the Table 5.1.

Table 5.1: Rationale and Impact of stages in proposed model

Stage	**Algorithm/ Framework**	**Rationale**	**Impact**
Data Pre-processing	Data cleaning	Removing any redundant, noisy records	Reduces redundant information; improves model efficiency
	Minmax	Scaling data to avoid overflow	
	IG	Keeping more relevant feature attributes	
Graph Creation	DGL	Encode network traffic into graph	Extraction of spatial information
IDS Model	E-GraphConv	Learn spatial representations from the graph	Capture spatial dependencies and feature propagation
	MLP	Efficient and flexible neural network for classification	Classifies into benign and anomalous

By considering spatial relationships, through the source and destination of network traffic, helped in identifying abnormal communication patterns, unauthorized access attempts, or suspicious interactions between devices more accurately.

5.2 Limitations of this study

The limitations discovered during our study for the model are as follows:

- As our model operate on graph-structured data, which can be computationally expensive, especially for large-scale networks. The propagation of information through the graph and the iterative nature of layers can lead to high computational requirements, making it challenging to apply to resource-constrained environments.
- Graph-based models as our E-Graph Conv are sensitive to the choice of hyperparameters making it difficult to find a set of hyperparameters that work well for a particular dataset. For instance, in very large or dynamic graphs two hop neighbour might not be sufficient to capture enough details from far neighbours.
- In multi-layer GCNs, there is a risk of over-smoothing, where information gets overly diffused across nodes, leading to loss of discriminative power in the

learned node representations. This can affect the ability to distinguish between subtle variations in the graph structure.

- The performance relies on the underlying graph structure for information propagation that means changes in the graph topology, such as node reordering or edge modifications, can impact the model's predictions.
- Our model is not suited for dynamic graphs as it assumes a fixed graph structure. Adding or removing nodes or edges in a graph may require retraining the model from scratch.
- Neural networks like our model lack interpretability, i.e., it is challenging to understand the reasoning behind their decisions or feature importance. The complex and non-linear nature of the model layers makes it difficult to interpret the learned representations or provide human-readable explanations for detected intrusions

5.3 Significance and Implications for future research

The proposed model, Edge Graph Convolutional Network, is a graph-based model for intrusion detection which can aid in the future research in this area through the understanding of advantages of the model:

- *Capturing Spatial Dependencies:* IDS deals with complex networks where nodes represent devices, and edges represent their interactions or connections. Graph-based models have better capturing of spatial dependencies by considering the connectivity patterns and relationships between nodes in the graph. This allows our model to effectively model and analyze the dependencies between devices, aiding in the detection of intrusions and abnormal behavior.
- *Handling Complex Network Data:* Intrusion detection operates on network data that is inherently graph-structured. E-GCN is specifically designed to process and analyze graph-structured data. This makes GCNs suitable for intrusion detection in complex and dynamic network environments.
- *Improved Performance:* Our model has shown promise in effectively capturing spatial relationships and dependencies between network nodes. By incorporating spatial features into the E-GCN architecture, it becomes possible to model the topology, interconnections, and traffic patterns of the network more accurately.

This, in turn, can lead to improved performance in detecting intrusions and anomalous behavior within the network.

- *Defense against Advanced Attacks:* Intruders often employ sophisticated techniques, to evade detection. By leveraging spatial features, a GCN-based intrusion detection system can effectively capture and analyze the propagation of attacks within the network. This enables the system to detect advanced attack patterns and provide early warning signals for potential security breaches.

The potential of graph-based models in intrusion detection can be further explored in future, some of the implications for future research are as:

- Exploring different strategies for integrating spatial information, optimizing feature extraction processes, and enhancing the learning capabilities of GCNs.
- Research efforts can be directed towards improving the robustness and adaptability of graph-based intrusion detection systems to handle complex and dynamic IIoT environments.
- It is important to work on the explainability and interpretability of model for future research, exploring techniques for interpreting the learned representations and understanding the contribution of spatial features to intrusion detection decisions
- Future research should also focus on practical implementation through various testing frameworks and real-world environments. This involves on large-scale networks, considering resource constraints of IIoT devices.
- Graph-based models can also be utilized for anomaly detection in IDS. By learning the normal patterns and structural dependencies from unlabelled data, it can identify deviations from the learned representation as potential anomalies or intrusions. This unsupervised learning approach is particularly useful as in majority of cases the organization lacks labelled dataset to train the model, further it can also detect novel unseen attacks and abnormalities as well.

CHAPTER 6

CONCLUSION

In this study, we have presented a graph-based network intrusion detection model. The network traffic data was transformed to network graph of hosts and connections representing the features that was used to classify between an anomalous or benign edge. The focus on this paper was the intrusion detection on industrial internet of things owing to which the methodology was tested against datasets that can relate to IIoT applications.

The essence of the method proposed is the utilization of spatial features extracted from the graph that can help unveil attacks patterns in regards to features of a network record as well as its relation to its surrounding nodes. This approach achieved better performance results when compared with other standard classification models, which was verified through extensive evaluation through multiple datasets in the results section. The convolution method for recognizing the patterns in combination with deep learning allowed us to train a model that can identify network anomalies with the consideration of the global structural view of the network through convolution.

REFERENCES

[1] Tsiknas, Konstantinos, Dimitrios Taketzis, Konstantinos Demertzis, and Charalabos Skianis. "Cyber threats to industrial IoT: a survey on attacks and countermeasures." IoT 2, no. 1 (2021): 163-186.

[2] Jiang, Xingbin, Michele Lora, and Sudipta Chattopadhyay. "An experimental analysis of security vulnerabilities in industrial IoT devices." ACM Transactions on Internet Technology (TOIT) 20, no. 2 (2020): 1-24.

[3] Sengupta, Jayasree, Sushmita Ruj, and Sipra Das Bit. "A comprehensive survey on attacks, security issues and blockchain solutions for IoT and IIoT." Journal of Network and Computer Applications 149 (2020): 102481.

[4] Choudhary, Sarika, and Nishtha Kesswani. "A survey: Intrusion detection techniques for internet of things." International Journal of Information Security and Privacy (IJISP) 13, no. 1 (2019): 86-105.

[5] Awotunde, Joseph Bamidele, Sakinat Oluwabukonla Folorunso, Agbotiname Lucky Imoize, Julius Olusola Odunuga, Cheng-Chi Lee, Chun-Ta Li, and Dinh-Thuan Do. "An Ensemble Tree-Based Model for Intrusion Detection in Industrial Internet of Things Networks." Applied Sciences 13, no. 4 (2023): 2479.

[6] Du, Jiawei, Kai Yang, Yanjing Hu, and Lingjie Jiang. "NIDS-CNNLSTM: Network Intrusion Detection Classification Model Based on Deep Learning." IEEE Access 11 (2023): 24808-24821.

[7] Priya, V., I. Sumaiya Thaseen, Thippa Reddy Gadekallu, Mohamed K. Aboudaif, and Emad Abouel Nasr. "Robust attack detection approach for IIoT using ensemble classifier." arXiv preprint arXiv:2102.01515 (2021).

[8] Awotunde, Joseph Bamidele, Chinmay Chakraborty, and Abidemi Emmanuel Adeniyi. "Intrusion detection in industrial internet of things network-based on deep learning model with rule-based feature selection." Wireless communications and mobile computing 2021 (2021): 1-17.

[9] Kasongo, Sydney Mambwe. "An advanced intrusion detection system for IIoT based on GA and tree based algorithms." IEEE Access 9 (2021): 113199-113212.

[10] Sarhan, Mohanad, Siamak Layeghy, and Marius Portmann. "Feature analysis for ML-based IIoT intrusion detection." arXiv e-prints (2021): arXiv-2108.

[11] Maharani, Mareska Pratiwi, Philip Tobianto Daely, Jae Min Lee, and Dong-Seong Kim. "Attack detection in fog layer for IIoT based on machine learning

approach." In 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1880-1882. IEEE, 2020.

[12] Muna, AL-Hawawreh, Nour Moustafa, and Elena Sitnikova. "Identification of malicious activities in industrial internet of things based on deep learning models." Journal of information security and applications 41 (2018): 1-11.

[13] Ma, Xiaoxiao, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z. Sheng, Hui Xiong, and Leman Akoglu. "A comprehensive survey on graph anomaly detection with deep learning." IEEE Transactions on Knowledge and Data Engineering (2021).

[14] Pujol-Perich, David, Jose Suarez-Varela, Albert Cabellos-Aparicio, and Pere Barlet-Ros. "Unveiling the potential of graph neural networks for robust intrusion detection." ACM SIGMETRICS Performance Evaluation Review 49, no. 4 (2022): 111-117.

[15] Iacovazzi, Alfonso, and Shahid Raza. "Ensemble of Random and Isolation Forests for Graph-Based Intrusion Detection in Containers." In *2022 IEEE International Conference on Cyber Security and Resilience (CSR)*, pp. 30-37. IEEE, 2022.

[16] Islam, Riadul, Maloy K. Devnath, Manar D. Samad, and Syed Md Jaffrey Al Kadry. "GGNB: Graph-based Gaussian naive Bayes intrusion detection system for CAN bus." *Vehicular Communications* 33 (2022): 100442.

[17] Islam, Riadul, Rafi Ud Daula Refat, Sai Manikanta Yerram, and Hafiz Malik. "Graph-based intrusion detection system for controller area networks." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 3 (2020): 1727-1736.

[18] Zhu, Huidi, and Jialiang Lu. "Graph-based Intrusion Detection System Using General Behavior Learning." In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pp. 2621-2626. IEEE, 2022.

[19] Lo, Wai Weng, Siamak Layeghy, Mohanad Sarhan, Marcus Gallagher, and Marius Portmann. "E-graphsage: A graph neural network based intrusion detection system for iot." In *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, pp. 1-9. IEEE, 2022.

[20] Otoum, Safa, Burak Kantarci, and Hussein T. Mouftah. "A novel ensemble method for advanced intrusion detection in wireless sensor networks." In *Icc 2020-2020 ieee international conference on communications (icc)*, pp. 1-6. IEEE, 2020.

[21] Halbouni, Asmaa H., Teddy Surya Gunawan, Murad Halbouni, Faisal Ahmed Abdullah Assaig, Mufid Ridlo Effendi, and Nanang Ismail. "CNN-IDS: Convolutional Neural Network for Network Intrusion Detection System." In *2022 8th International Conference on Wireless and Telematics (ICWT)*, pp. 1-4. IEEE, 2022.

[22] Chang, Liyan, and Paula Branco. "Graph-based solutions with residuals for intrusion detection: The modified e-graphsage and e-resgat algorithms." *arXiv preprint arXiv:2111.13597* (2021).

[23] Hamilton, Will, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs." *Advances in neural information processing systems* 30 (2017).

[24] Patel, N. D., B. M. Mehtre, and Rajeev Wankar. "Simulators, emulators, and test-beds for internet of things: A comparison." In *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 139-145. IEEE, 2019.

[25] Siddiqui, Muhammad Nasir, Kaleem Razzaq Malik, and Tauqeer Safdar Malik. "Performance analysis of blackhole and wormhole attack in MANET based IoT." In 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2), pp. 1-8. IEEE, 2021.

[26] Wu, Chao, Yuan'an Liu, Fan Wu, Feng Liu, Hui Lu, Wenhao Fan, and Bihua Tang. "A hybrid intrusion detection system for iot applications with constrained resources." International Journal of Digital Crime and Forensics (IJDCF) 12, no. 1 (2020): 109-130.

[27] Gupta, Yash, Rajeev Shorey, Devadatta Kulkarni, and Jeffrey Tew. "The applicability of blockchain in the Internet of Things." In 2018 10th International Conference on Communication Systems \& Networks (COMSNETS), pp. 561-564. IEEE, 2018.

[28] Alnuman, Ibrahim Ahmed, and Mousa Al-Akhras. "Machine learning DDos detection for generated internet of things dataset (IoT Dat)." In 2020 2nd International Conference on Computer and Information Sciences (ICCIS), pp. 1-6. IEEE, 2020.

[29] Govindasamy, Jegan, and Samundiswary Punniakody. "A comparative study of reactive, proactive and hybrid routing protocol in wireless sensor network under wormhole attack." Journal of Electrical Systems and Information Technology 5, no. 3 (2018): 735-744.

[30] Almomani, Iman, and Maha Saadeh. "S-FEAR: secure-fuzzy energy aware routing protocol for wireless sensor networks." KSII Transactions on Internet and Information Systems (TIIS) 12, no. 4 (2018): 1436-1457.

[31] Sedjelmaci, Hichem, Sidi Mohammed Senouci, and Mohamad Al-Bahri. "A lightweight anomaly detection technique for low-resource IoT devices: A game-theoretic methodology." In 2016 IEEE international conference on communications (ICC), pp. 1-6. IEEE, 2016.

[32] Ioulianou, Philokypros, Vasileios Vasilakis, Ioannis Moscholios, and Michael Logothetis. "A signature-based intrusion detection system for the Internet of Things." Information and Communication Technology Form (2018).

[33] A. Le, J. Loo, K. K. Chai, M. Aiash, A specification-based IDS for detecting attacks on RPL-based network topology, Information 7 (2) (2016) 25.

[34] Yavuz, Furkan Yusuf. "Deep learning in cyber security for internet of things." Master's thesis, Fen Bilimleri Enstitüsü, 2018.

[35] Prasadh, S. Kshipra, and Sumit Kumar Jindal. "Security and Efficiency Analysis of Anti-jamming Techniques." In International Conference on Internet of Things and Connected Technologies, pp. 251-259. Springer, Cham, 2019.

[36] Remesh, Athira, Divya Muralidharan, Neha Raj, J. Gopika, and P. K. Binu. "Intrusion detection system for IoT devices." In 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 826-830. IEEE, 2020.

[37] Saeedi, Kubra. "Machine learning for ddos detection in packet core network for iot." (2019).

[38] Arena, Antonio, Pericle Perazzo, Carlo Vallati, Gianluca Dini, and Giuseppe Anastasi. "Evaluating and improving the scalability of RPL security in the Internet of Things." Computer Communications 151 (2020): 119-132.

[39] Khadr, Monette H., Haythem Bany Salameh, Moussa Ayyash, Sufyan Almajali, and Hany Elgala. "Testbed Validation of Security-Aware Channel Assignment in Cognitive Radio IoT Networks." In 2020 IEEE 6th World Forum on Internet of Things (WF-IoT), pp. 1-6. IEEE, 2020.

[40] Ijaz, Sidra, Munam Ali Shah, Abid Khan, and Mansoor Ahmed. "Smart cities: A survey on security concerns." International Journal of Advanced Computer Science and Applications 7, no. 2 (2016): 612-625.

[41] Ferrag, Mohamed Amine, Othmane Friha, Djallel Hamouda, Leandros Maglaras, and Helge Janicke. "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning." IEEE Access 10 (2022): 40281-40306.

[42] Sharafaldin, Iman, Arash Habibi Lashkari, and Ali A. Ghorbani. "Toward generating a new intrusion detection dataset and intrusion traffic characterization." *ICISSp* 1 (2018): 108-116.

[43] Moustafa, Nour, and Jill Slay. "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)." In *2015 military communications and information systems conference (MilCIS)*, pp. 1-6. IEEE, 2015.

[44] Moustafa, Nour, and Jill Slay. "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set." *Information Security Journal: A Global Perspective* 25, no. 1-3 (2016): 18-31.

[45] Moustafa, Nour, Gideon Creech, and Jill Slay. "Big data analytics for intrusion detection system: Statistical decision-making using finite dirichlet mixture models." *Data Analytics and Decision Support for Cybersecurity: Trends, Methodologies and Applications* (2017): 127-156.

[46] Moustafa, Nour, Jill Slay, and Gideon Creech. "Novel geometric area analysis technique for anomaly detection using trapezoidal area estimation on large-scale networks." *IEEE Transactions on Big Data* 5, no. 4 (2017): 481-494.

[47] Sarhan, Mohanad, Siamak Layeghy, Nour Moustafa, and Marius Portmann. "Netflow datasets for machine learning-based network intrusion detection systems." In *Big Data Technologies and Applications: 10th EAI International Conference, BDTA 2020, and 13th EAI International Conference on Wireless Internet, WiCON 2020, Virtual Event, December 11, 2020, Proceedings 10*, pp. 117-135. Springer International Publishing, 2021.

LIST OF PUBLICATIONS

[1] Bora Nilutpol, Nandanwar Himanshu, and Chauhan Anamika. “Security in Internet of Things: A Comprehensive Review of Simulators, Emulators and Test-Beds.”, In *International Journal of Special Education* Vol.37, No.3, 2022.

[2] Bora Nilutpol, and Chauhan Anamika. “Edge-Graph Convolution Network: An Intrusion Detection Approach for Industrial IoT”, Accepted *in Springer Lecture Notes in Networks and Systems, 4th International Conference on Data Analytics & Management* 2023.

turnitin

Similarity Report ID: oid:27535:36095046

PAPER NAME
MasterThesis Nilutpol.pdf

AUTHOR
Nilutpol Bora

WORD COUNT
13237 Words

CHARACTER COUNT
74333 Characters

PAGE COUNT
49 Pages

FILE SIZE
1.6MB

SUBMISSION DATE
May 25, 2023 1:19 AM GMT+5:30

REPORT DATE
May 25, 2023 1:20 AM GMT+5:30

4% Overall Similarity

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

- 3% Internet database
- Crossref database
- 3% Submitted Works database
- 2% Publications database
- Crossref Posted Content database

Excluded from Similarity Report

- Bibliographic material
- Cited material
- Manually excluded sources
- Quoted material
- Small Matches (Less then 10 words)

International Journal of Special Education
ISSN : 0827 - 3383
Email : info@ardaconference.com
Website : www.internationaljournalofspecialeducation.com
Contact No : +91 93456 84472
IJSE

Date: 16th Mar 2022

Acceptance Letter

Dear Author(s): Nilutpol Bora, Himanshu Nandanwar, Anamika Chauhan

Paper ID	NIER_42013
Paper Title	Security in Internet of Things: A comprehensive review of Simulators, Emulators and Test-beds

This is to enlighten you that above manuscript reviewed and appraised by the review committee members of ARDA and it is accepted for the purpose of publication in the **"International Journal of Special Education"** with **ISSN: 0827-3383** that will be available at https://www.internationaljournalofspecialeducation.com/.

You have to send following documents at info@ardaconference.com on or before 20th Mar 2022.
1. Final Paper | Ms Word .doc/.docx file
2. Proof of Payment | Scanned | Online Received Email

Note: Please read carefully

1. Above manuscript will be published within 7 days from your Payment
2. International Journal of Special Education is a SCOPUS Indexed Journal.
3. Author(s) will receive Publication information and Published Paper link through ARDA
4. You may see more about the journal at: https://www.internationaljournalofspecialeducation.com/
5. You will receive Volume/ Issue information of your paper very soon.

Sincerely

Dr. Simpson Rodricks

President,
ARDA.

Scopus Preview

Source details

International Journal of Special Education

Scopus coverage years: 1995, from 2001 to Present

Publisher: SPED Ltd

ISSN: 0827-3383 E-ISSN: 1917-7844

Subject area: Medicine: Rehabilitation | Social Sciences: Education

Source type: Journal

View all documents > | Set document alert | Save to source list | Source Homepage

CiteScore 2021
1.1

SJR 2021
0.187

SNIP 2021
0.431

CiteScore | CiteScore rank & trend | Scopus content coverage

Improved CiteScore methodology

CiteScore 2021 counts the citations received in 2018-2021 to articles, reviews, conference papers, book chapters and data papers published in 2018-2021, and divides this by the number of publications published in 2018-2021. Learn more >

CiteScore 2021

$$1.1 = \frac{99 \text{ Citations } 2018 - 2021}{89 \text{ Documents } 2018 - 2021}$$

Calculated on 05 May, 2022

CiteScoreTracker 2022

$$0.5 = \frac{52 \text{ Citations to date}}{96 \text{ Documents to date}}$$

Last updated on 05 April, 2023 • Updated monthly

CiteScore rank 2021

Category	Rank	Percentile
Medicine Rehabilitation	#75/129	42nd
Social Sciences Education	#882/1406	37th

View CiteScore methodology > | CiteScore FAQ > | Add CiteScore to your site

Gmail Nilutpol Bora <pnilut@gmail.com>

ICDAM 2023: Paper Notification 878

2 messages

ICDAM Conference <icdam.conf@gmail.com> Thu, May 11, 2023 at 9:00 PM
To: Pnilut <pnilut@gmail.com>

Dear Author(s),

Greetings from ICDAM 2023!

We congratulate you that your paper with submission ID **878** and Paper Title **'Edge-Graph Convolution Network: An Intrusion Detection Approach for Industrial IoT'** has been accepted for publication in the Springer LNNS series [Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago; All books published in the series are submitted for consideration in Web of Science]. This acceptance means that your paper is among the top 20% of the papers received/reviewed. **Our registration process has started and we have left with a limited number of registrations. Kindly submit your registration fees as early as possible.**

You are requested to do the registration as soon as possible and submit the following documents to icdam.conf@gmail.com at the earliest.
1. Final Camera-Ready Copy (CRC) as per the springer format. (See https://icdam-conf.com/downloads)
2. Copy of e-receipt of registration fees. (For Registration, see https://icdam-conf.com/registrations)
3. The final revised copy of your paper should also be uploaded via Microsoft CMT.

The reviewers comments are given at the bottom of this letter, please improve your paper as per the reviewers comments. While preparing the final CRC manuscript, kindly check the following google link of proceedings of the previous International Conference on Data Analytics and Management:

https://scholar.google.com/citations?hl=en&authuser=2&user=9qFcrv0AAAAJ

and it is suggested to cite the relevant latest papers matching the area of your current research paper.

The paper prior to submission should be checked for plagiarism from licensed plagiarism softwares like Turnitin/iAuthenticate etc. The similarity content should not exceed 15%.

Scopus Preview

Source details

Lecture Notes in Networks and Systems

Scopus coverage years: from 2016 to Present

Publisher: Springer Nature

ISSN: 2367-3370 E-ISSN: 2367-3389

Subject area: Engineering: Control and Systems Engineering | Computer Science: Computer Networks and Communications | Computer Science: Signal Processing

Source type: Book Series

CiteScore 2021
0.7

SJR 2021
0.151

SNIP 2021
0.249

View all documents > | Set document alert | Save to source list | Source Homepage

CiteScore | CiteScore rank & trend | Scopus content coverage

Improved CiteScore methodology

CiteScore 2021 counts the citations received in 2018-2021 to articles, reviews, conference papers, book chapters and data papers published in 2018-2021, and divides this by the number of publications published in 2018-2021. Learn more >

CiteScore 2021

$$0.7 = \frac{9{,}307 \text{ Citations } 2018 - 2021}{14{,}098 \text{ Documents } 2018 - 2021}$$

Calculated on 05 May, 2022

CiteScoreTracker 2022

$$0.7 = \frac{18{,}643 \text{ Citations to date}}{25{,}876 \text{ Documents to date}}$$

Last updated on 05 April, 2023 • Updated monthly

CiteScore rank 2021

Category	Rank	Percentile
Engineering Control and Systems Engineering	#236/270	12th
Computer Science Computer Networks and Communications	#321/359	10th

View CiteScore methodology > | CiteScore FAQ > | Add CiteScore to your site