International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 01 | January - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

# CyberWatch: Deep Learning-Driven Network Intrusion Detection

Saksham Gupta, Aditya Sharma
sakshamgupta9656@gmail.com, adityasharma6816@gmail.com
Maharaja Agrasen Institute of Technology

***Abstract*—The present era is being dominated by the digital world which has also given rise to growing cyber threats and crimes, the project introduces an innovative intrusion detection system(IDS) that deploys deep learning's pattern recognition capabilities with a real-time algorithm approach to threats and network security. The project's framework involves two key components: Graph theory and Artificial Neural Network (ANN). The graph theory is used to represent the IoT network structure which helps represent relations between network structures and analyze the anomalies or malicious activities present in the network. Artificial Neural Network (ANN) model, on the other hand, is based on a human's brain neural network; which works as a machine learning algorithm to recognize and predict patterns. Cyberwatch is trained on various datasets such as CIC IDS 2018 and CIC IDS 2017. The project's main aim is to provide an effective solution to the ever-changing landscape of network intrusion.**

**Keywords—**Intrusion Detection System(IDS), Energy Efficiency, Internet Of Things(IoT), Network Forensics, Graph Theory Support Vector Machine(SVM), Genetic Algorithm, Artificial Neural Networks(ANN), Cybercrime

## I. Introduction

In the realm of cybersecurity, network forensics plays a significant role in restructuring network events, understanding user actions, decoding application behaviors, and comprehending device activities (Schwartz et Al., 2010)[12]. In this present epoch, the Internet and cybernetic networks are the essential elements of our evergreen society, having made monumental contributions to our economy and influencing people's work and way of life[10][13]. While both institutions and individuals resort more and more to web cyber networks, significant information that relates to institutions and individual activities is collected and stored. It leads to huge drawbacks if the network is been compromised. The recent emergence of the Internet of Things (IoT), identified by the integration of sensors, software, and network connectivity, gives rise to new challenges. The information stored within the IoT is often restricted and confidential, making it a prime target for cybercriminals (Alaba et Al., 2017)[2]. The minimization and energy optimisation of IoT devices render traditional network forensic technologies unsuitable, making it necessary to use specialized approaches tailored to IoT cybercrimes.

Several IDS, are classified into three types (Zarpelao et Al., 2017)[17] as shown in Fig. 1. Distributed IDS is used in devices having high computational capability. It involves detection mechanisms in every physical node. On the other hand, centralized IDS relies on dedicated components within the network for detection, while hybrid IDS is made by the combination of two or more approaches to the IDS.

In the context of IoT networks, threats can be extended beyond unauthorized intrusion, malware, and service outage to include the manipulation of data nodes and the generation of false or harmful information by the attackers (Ahmed, 2017)[1]. Especially, Denial of Service (DoS) attacks, particularly Energy Exhaustion Attacks (EEA), pose one of the biggest threats to IoT performance. Traditional intrusion detection methods may be insufficient in scenarios with mobility-constrained mobile sinks, such as railway or vehicle-based applications.

The study addresses the importance of intrusion detection in IoT environments (Azidine et Al., 2022)[5], using machine learning techniques or algorithms for anomaly detection. By analyzing the detection rate and thwarting attacks, the project aims to provide the security of IoT networks, crucial for safeguarding vulnerable IoT devices due to resource constraints.

A paper authored[16], introduces a IDS made especially for IoT applications with limited resources. The study uses graph theory to analyze IoT networks, displaying a hybrid system comprising Centralized and Active Malicious Node Detection (CAMD) and Distributed and Passive Energy-Efficient Authentication (DPER). CAMD is a genetic algorithm-based data-gathering scheme, that detects and traces malicious nodes which are being controlled by cyberpunk. DPER is used in communication protocols to reduce the impact of Energy Exhaustion Attacks (EEA). Simulation experiments conducted on the NS-3 platform showcases the system's effectiveness in precise detection without compromising energy efficiency, thereby contributing to the improved security and energy efficiency of IoT networks.

The paper encompasses three key elemental contributions. First, we combine data quality improvement techniques and SVM(Support Vector Machine) in the framework of an intrusion detection approach, through which we can attain precise and resilient results. Second, a method to improve features is being used, called graph theory is proposed, that can provide information (insights) on the internal structure of the network and allows straightforward, robust, and accurate assessment of network performance. Thirdly, there is strong adaptability in the proposed framework. It is adaptable and can be used for a variety of purposes when grouping objects into categories is necessary. Furthermore, it is easy to incorporate several efficient learning algorithms into our system.

Our project draws inspiration from the shared themes of IoT security and intrusion detection. The standout feature is the incorporation of deep learning technologies for advanced pattern recognition such as graph theory, Support vector machine (SVM), and ANN model. These techniques

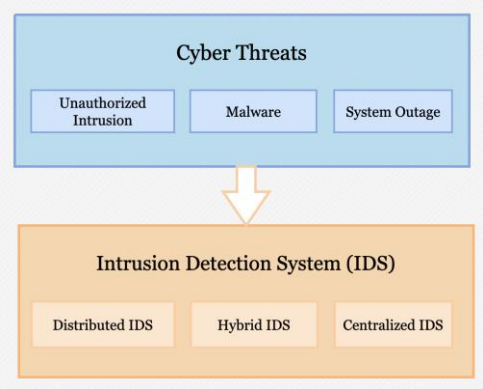


Fig. 1. Cyber threats and IDS categories

contribute valuable insights to our project's emphasis on real-time adaptability and detecting emerging threats to provide a unique and comprehensive approach to network security

## II. Background

In recent years, significant improvements have been made in developing anti-cybercrime technologies, leading to the proposal of various advanced Intrusion Detection Systems (IDSs). These IDSs are generally grouped into three types based on their placement within the networks.

Research efforts in the field of distributed Intrusion Detection Systems (IDSs) have predominantly centered around addressing the challenges associated with energy consumption when deploying IDSs extensively across network nodes. To tackle this issue, lightweight IDSs have emerged as effective solutions, aiming to mitigate additional energy consumption while efficiently carrying out intrusion detection tasks. For example, in a study conducted by (Oh et Al., 2014)[11], an IDS was introduced with a matching algorithm designed to identify malicious behaviors and analyze packet payloads. The authors successfully achieved lightweight functionality by reducing the iteration times of the pattern-matching algorithm. Similarly, Lee et Al.[9] explored ways to minimize IDS overhead by establishing a lightweight IDS under a low-energy-consuming communication protocol. However, the conventional distributed IDSs faced a significant challenge in resource-constrained IoT environments due to their demand for relatively high computational capabilities for each network node.

In contrast to this, centralized IDSs mainly relied on dedicated nodes for intrusion detection of a system. Wallgren et Al.[15]employed a heartbeat protocol for global monitoring within a centralized IDS framework. However, the considerable overhead associated with heartbeat packets raised major concerns, especially in energy-constrained IoT scenarios. Gunasekaran et Al. [6] proposed an anti-denial of Sleep (DoSL) system for IoT base stations, utilizing a genetic algorithm and encryption algorithm to identify DoSL attacks. Despite all these efforts, the delayed detection by dedicated nodes in centralized IDSs allowed cybercriminals to fulfill their goals before preventive measures could be implemented to stop them. Recognizing the urgent need for energy-efficient and real-time IDSs adapted for IoT applications, hybrid IDSs have emerged as a promising solution. Different approaches have been explored within this paradigm. Amaral et Al.[4] introduced a competition-based approach, selecting only robust nodes to monitor the adjacent nodes dynamically. However, this approach relied on resourceful nodes, introducing more energy consumption. Other methods, such as clustering, were utilized by Le et Al.[8] and Joby et Al.[7] to detect intrusions. While cluster heads detected malicious activities and behavior within their clusters, the extra computation led to the immediate depletion of cluster heads.

Azidine Guezzaz et Al.[5] addressed the need for robust intrusion detection in IoT environments, leveraging machine learning techniques for anomaly detection. Their project aims to fortify the security of IoT networks, providing critical protection for vulnerable devices due to resource constraints.

A survey has been conducted by Hashim et Al.[3] on cyber-attack prediction within the Network Intrusion Detection System(NIDS), focusing on alert correlation techniques. The research provides valuable insights that can enhance the intrusion detection capabilities. It also provides the knowledge for proactive measures against evolving cyber threats.

Aiming at the specific problems of IOT applications with a path-constrained mobile sink, this project provides a IDS designed to detect intrusion sources and provide digital evidence for network forensics. Especially, the proposed IDS performs within the energy constraints of IoT, ensuring minimal consumption of additional energy.

## III. Proposed Model

### A. *software requirements specification*

This document defines the requirements for the development of an Intelligent and Efficient IDS using an ANN model and graph theory. The system will be trained and tested on diverse datasets, including CIC IDS 2018 and CIC IDS 2017.

*1) Funtional Requirements:*

- Data Collection: The system shall utilize diverse datasets, including CIC IDS 2018 and CIC IDS 2017, for training and testing the ANN model. It should provide a method to import, preprocess, and store the datasets in a format suitable for training.
- ANN Model Implementation: Developing and implementing an Artificial Neural Network (ANN) model personalized for graph theory. The model should have the capability to learn and adapt to different patterns of network traffic.
- Training and Testing: The system shall assist in the training of the ANN model using the collection of various datasets. Deploying diverse methods for evaluating and validating the performance of the trained model through testing. Provide feedback on the accuracy, and precision of the model.

- User Interface: Designing an intuitive and user-friendly interface for system configuration and monitoring. Including dashboards and visualizations to present real-time and historical data related to intrusion detection. Implementing user controls for configuring system parameters, accessing reports, and adjusting alert settings.
- Logging and Auditing: Executing comprehensive logging mechanisms to record system activities, detected intrusions and model updates. Enable auditing features to trace user interactions, configuration changes, and system events for accountability.
- Integration with Existing Systems: Ensuring compatibility and smooth integration with existing network infrastructures, security protocols, and data storage systems. Developing APIs or interfaces for potential integration with third-party security tools or dashboards.
- Performance Metrics: Implementing mechanisms to measure and report the performance metrics of the intrusion detection system, including detection accuracy, processing time, and false positive rates.

*2) Non Funtional Requirements:*

- Response Time: This is the total amount of time a system takes to respond to a request for a particular service. The system should have low latency in processing and responding to real-time network events.
- Scalability: Scalability is the property of a system to handle a growing amount of work. The solution must be highly scalable to handle increasing volumes of network traffic and growing datasets.
- Availability: The system should maintain high availability, with downtime not exceeding than specified limit for routine maintenance.
- Fault Tolerance: The system must operate properly in the event of failure or any major dysfunction in one or more of its components.
- Data Encryption: All sensitive data, including training datasets and system logs, should be encrypted into cipherText to ensure confidentiality.
- Access Controls: Strict access restrictions must be put in place in order to protect a building, its people, information, and property. This is keeping visitors and authorised users access convenient while reducing the possibility of unwanted intrusion.
- Resource Utilization: The system should optimize the utilization of computational resources, to ensure the efficient usage of CPU, memory, and storage.
- Concurrency: The system must have the ability to handle multiple network events simultaneously without performance degradation.
- Modularity: System with a modular architecture, and modern user interfaces have easy updates, enhancements, and maintenance of individual components.
- Documentation: Maintain detailed documentation for the system, including information about its architecture, APIs, and configuration procedures.
- User Training: Develop user training materials and conduct necessary training sessions to ensure users can easily set up and use the system.
- Accessibility: The user interface should be accessible and user-friendly by security analysts with different levels of technical expertise.

*B. DFD*

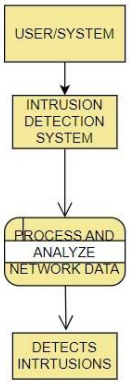


Fig. 2. Data Flow Diagram of IDS

The above Fig. 2. explains the following points:

- External Entity (User/System): This represents an external entity, which could be a user or another system interacting with the intrusion detection system. External entities act as sources or destinations of data.
- Intrusion Detection System: The core process responsible for detecting intrusions. It receives, analyzes, and processes network data to identify potential security threats or intrusions in the system.
- Process and Analyze Network Data: This is a subprocess within the intrusion detection process signifying the detailed steps involved in handling network data. It entails tasks such as data preprocessing, feature extraction, and analysis.
- Detects Intrusion: This provides the outcome of the intrusion detection process. If the analysis of network data reveals any patterns or behaviors indicating an intrusion, this stage generates an alert or notification.

*C. Model*

Our security system has two main parts as illustrated in Fig. 3. The first part uses an algorithm called Genetic Algorithm (GA), which has a main function of collecting information. This method helps to save energy thus increasing the efficiency of the network. While it's collecting data, it's also on the lookout for any malicious or malware activities. It does this by comparing patterns by using a matrix.

Now, the second part works like a set of rules for communication, and these rules are named Advanced Shortest Path Tree (ASPT). These rules are resourceful – they not only make sure that different parts of the network use their energy wisely but also prevent cyber attacks. Specifically, they try to reduce the impact called Energy Efficiency Attack(EEA), which can cause exploitation of vulnerabilities in the software or hardware of a system. So, both parts make sure to keep the network safe and working efficiently.

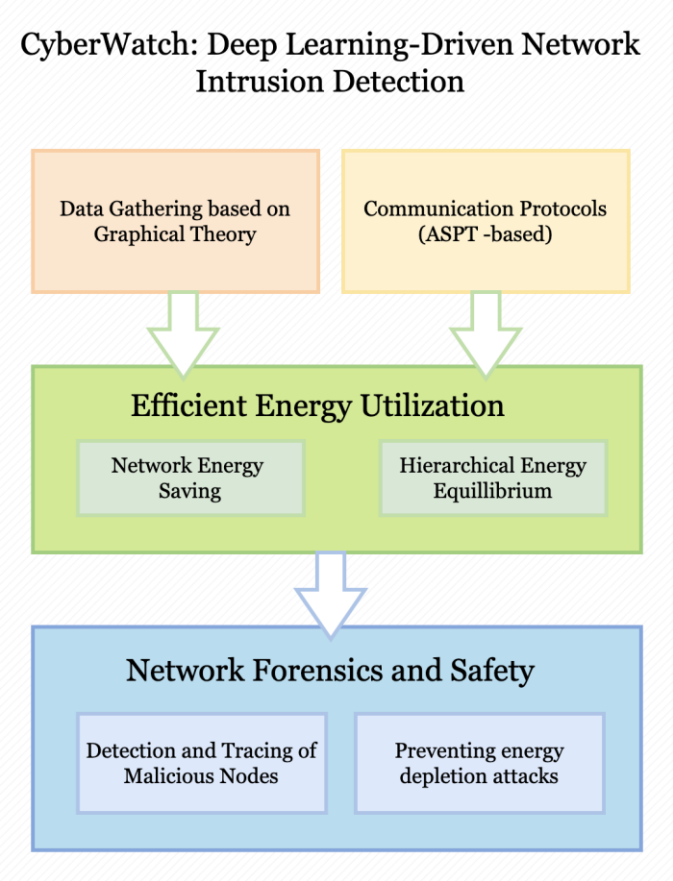


Fig. 3. Proposed IDS model

## IV. Algorithm

Our model detects and classifies network intrusions using the hybrid model combining traditional machine learning features of ANN and running input through graphical theorem. Our input vectors include:

- Network traffic data with selected features.
- Pre-trained neural network model (ids_ANN_imbalanced_plain_2.h5).
- Feature scaler (ANN_scaler_imbalanced_plain_2.dat).
- Class labels (LABELS).

The result predicts the class name(Benign or Intrusion), predicted class IDs, and confidence score for predictions. The initial steps are as follows:

1) Loading the pre-trained ANN model("ids_ANN_imbalanced_plain_2h˙5") and features scaler from the file('ANN_scaler_imbalanced_plain_2.dat').
2) Normalize the input features using the loaded feature scaler and make predictions using a pre-trained ANN model for intrusion probabilities while reshaping the input tensor to batch.

### *A. Post-Process Prediction*

1) Obtain class IDs by taking the argmax of predictions along the first axis.
2) Map class IDs to class names using the provided LABELS dictionary.
3) Calculate confidence scores by taking the maximum probability for each prediction.

The following Fig. 4. underlines the working algorithm.

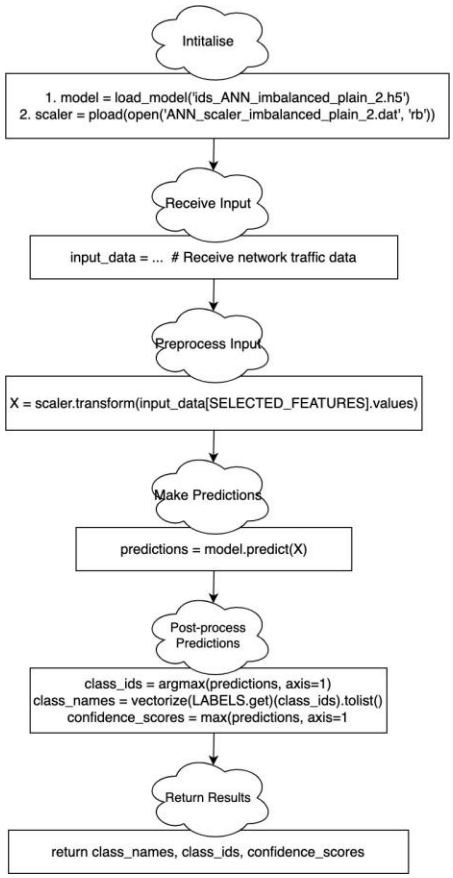


Fig. 4. Working Algorithm

## V. RESULTS

The subsection that follows provides the analysis of the results with the previous approaches for the hybrid intrusion detection system. The areas that are covered are DoS, Probe attacks, U2R, and R2L attack vectors. And a comparative analysis of IDS techniques.

The paper by author Vinayakumar R et Al.[14], examines the effectiveness of Integer only Recurrent Neural Network

and other Recurrent Neural Network variants for intrusion detection. The detection rates obtained had proven to be highly effective for 'DoS' and 'Probe' attacks because unique time series of network events is being formed . However, they also highlighted that the frequency can be enhanced by promoting training or stacking a few more layers to the existing architecture, or integrating new features into the existing data.

Following their analysis we proposed a model, which incorporated ANN for intrusion detection using a second layer of data presumed from graph theory. Overall, the usage of ANN has Significantly improved detection rates in relation to the IRNN variants from the previous studies. Below is the results obtained from the dataset CIC IDS 2018 and CIC IDS 2017.

| Packet ID | Source IP Address | Server IP | VM Name | Class ID | Predicted Class | Confidence Score |
|---|---|---|---|---|---|---|
| 1 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 2 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 3 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 4 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 5 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 6 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 7 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |
| 8 | 183.83.155.234 | 172.45.101.155 | third-eye: paperspace_rtx_2080 | 1 | Intrusion | 0.9999999404 |

Fig. 5. Detailed result of CIC IDS 2017 and CIC IDS 2018

We can assume the efficiency, of our model by averaging the confidence score which comes out to be 99.1%.

It is observed that the attacks on computers and their networks are dynamically evolving in modern days. Though IRNN variants prove to be modular for ID, it is considered to be outdated due to the fact that the attacks are common and inherent issues in connection records. Our approach can be considered as one of a future direction which is summarised with the below test results.

| Algorithm | MLP 1 | IRNN 2 | RNN 4 | ANN |
|---|---|---|---|---|
| Accuracy | 0.924 | 0.949 | 0.942 | 0.991 |

Fig. 6. Comparative analysis of test result of KDDCup-99 with CIC IDS

## VI. CONCULSION

In conclusion, our proposed hybrid IDS, rooted in graph theory, works as a robust solution for securing IoT applications. It uses the ANN model to represent the IoT network and helps in detecting malicious activities. On the other hand, SVM works as a supervised learning algorithm. It is used to help classification and regression of tasks.

This research contributes significantly to the ever-changing landscape of IoT security, addressing the unique challenges posed by resource-constrained environments. The application of advanced Shortest Path Tree-based communication protocols ensures a secure defense against cyber threats, providing a safer and more reliable IoT ecosystem. Our hybrid IDS provides a real-time and effective solution to fortify network defenses and preserve the privacy of sensitive data in IoT applications.

## REFERENCES

1 Abdulghani Ali Ahmed. Investigation approach for network attack intention recognition. In *Digital Forensics and Forensic Investigations: Breakthroughs in Research and Practice*, pages 185–208. IGI Global, 2020.
2 Fadele Ayotunde Alaba, Mazliza Othman, Ibrahim Abaker Targio Hashem, and Faiz Alotaibi. Internet of things security: A survey. *Journal of Network and Computer Applications*, 88:10–28, 2017.
3 Hashim Albasheer, Maheyzah Md Siraj, Azath Mubarakali, Omer Elsier Tayfour, Sayeed Haider Salih, Mosab Hamdan, Suleman Khan, Anazida Zainal, and Sameer Kamarudeen. Cyber-attack prediction based on network intrusion detection systems for alert correlation techniques: A survey. *Sensors (Basel, Switzerland)*, 22, 2022.
4 Joao Amaral, Luis Oliveira, Joel Rodrigues, Guangjie Han, and Lei Shu. Policy and network-based intrusion detection system for ipv6-enabled wireless sensor networks. pages 1796–1801, 06 2014.
5 Azidine Guezzaz, Mourade Azrour, Said Benkirane, Mouaad Mohy-Eddine, Hanaa Attou, and Maryam Douiba. A lightweight hybrid intrusion detection framework using machine learning for edge-based iiot security. *Int Arab J Inf Technol*, 19(5), 2022.
6 Mahalakshmi Gunasekaran, Subathra Periakaruppan, et al. Ga-dosld: genetic algorithm based denial-of-sleep attack detection in wsn. *Security and Communication Networks*, 2017, 2017.
7 PP Joby and P Sengottuvelan. A localised clustering scheme to detect attacks in wireless sensor network. *Int. J. Electronic Security and Digital Forensics*, 7(3):211, 2015.
8 Anhtuan Le, Jonathan Loo, Kok Keong Chai, and Mahdi Aiash. A specification-based ids for detecting attacks on rpl-based network topology. *Information*, 7(2), 2016.
9 Tsung-Han Lee, Chih-Hao Wen, Lin-Huang Chang, Hung-Shiou Chiang, and Ming-Chun Hsieh. A lightweight intrusion detection scheme based on energy consumption analysis in 6lowpan. In *Advanced Technologies, Embedded and Multimedia for Human-centric Computing: HumanCom and EMC 2013*, pages 1205–1213. Springer, 2014.
10 Jun Luo, Senchun Chai, Baihai Zhang, Yuanqing Xia, Jianlei Gao, and Guoqiang Zeng. A novel intrusion detection method based on threshold modification using receiver operating characteristic curve. *Concurrency and Computation: Practice and Experience*, 32(14):e5690, 2020.
11 Doohwan Oh, Deokho Kim, and Won Woo Ro. A malicious pattern detection engine for embedded security systems in the internet of things. *Sensors*, 14(12):24188–24211, 2014.
12 Eddie Schwartz. Network packet forensics. *CyberForensics: Understanding Information Security Investigations*, pages 85–101, 2010.
13 Chih-Fong Tsai, Yu-Feng Hsu, Chia-Ying Lin, and Wei-Yang Lin. Intrusion detection by machine learning: A review. *Expert Systems with Applications*, 36(10):11994–12000, 2009.
14 R Vinayakumar, KP Soman, and Prabaharan Poornachandran. A comparative analysis of deep learning approaches for network intrusion detection systems (n-idss): deep learning for n-idss. *International Journal of Digital Crime and Forensics (IJDCF)*, 11(3):65–89, 2019.
15 Linus Wallgren, Shahid Raza, and Thiemo Voigt. Routing attacks and countermeasures in the rpl-based internet of things. *International Journal of Distributed Sensor Networks*, 9(8):794326, 2013.
16 Chao Wu, 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(1):109–130, 2020.
17 Bruno Bogaz Zarpela˜o, Rodrigo Sanches Miani, Cla´udio Toshio Kawakani, and Sean Carlisto de Alvarenga. A survey of intrusion detection in internet of things. *Journal of Network and Computer Applications*, 84:25–37, 2017.