A MALLICIOUS BOT IOT NETWORK TRAFFICDETECTON USING MACHINE LEARNING

[1]*Dr. S. MaruthuPerumal* [2]*K. Satish* [3]*L. Spandana*[4]*L. Sai Ram* [5]*M. Mahesh*

[1]Professor & HOD, Department of Computer Science and Engineering, Bharath Institute of Higher Education and Research, Chennai, India- 600073.
[2,3,4,5]Students, Department of Computer Science and Engineering,
Bharath Institute of Higher Education and Research, Chennai, India- 600073 .

Abstract-
The rapid expansion of cloud computing has led to increasingly complex and dynamic open networks and service sharing scenarios, posing heightened security challenges. Anomaly network traffic detection has emerged as a crucial method for network protection, capable of identifying various known attacks. However, existing models face limitations in fully capturing the temporal and spatial features of network traffic, necessitating improvements in classification accuracy. This project introduces an anomaly network traffic detection model, Integrated Temporal and Spatial Features Network (ITSN), utilizing a three-layer parallel network structure. ITSN integrates temporal and spatial features through feature fusion technology to enhance the accuracy of network traffic classification. Additionally, an enhanced method for raw traffic feature extraction is proposed to reduce redundancy, expedite network convergence, and mitigate dataset imbalances. By comparing Convolutional Neural Network (CNN) as the existing system with Recurrent Neural Network (RNN) as the proposed system, this study demonstrates that RNNoutperforms CNN in terms of accuracy, highlighting its potential for advancing anomaly network traffic detection in cloud computing environments.

keywords— Cloud computing ,Anomaly detection ,Network security ,Deep learning, Temporal features, Spatial features Feature fusion,Convolutional Neural Network (CNN),Recurrent Neural Network (RNN),Accuracy Rawtraffic feature extraction

I. INTRODUCTION

The rapid expansion of computer networks has reshaped perspectives on network security, with easy accessibility rendering them vulnerable to numerous threats from hackers. These threats pose significant risks to network integrity and functionality. Researchers have responded by developing Intrusion Detection Systems (IDS) capable of identifying attacks across various network environments. A multitude of methods, encompassing both misuse and anomaly detection, have been proposed to bolster network security. These technologies often complement each other, with certain approaches proving more effective in specific environments. This project introduces a novel intrusion detection system designed to categorize and evaluate existing methodologies. The taxonomy employed encompasses both the detection principle and operational aspects of IDS. The project utilizes a diverse array of algorithms, including Dismissal of Conquered Movements, Lexicographic Game Method, Collaborative Game Method, Repeated Game Approach, Stochastic RandomProcess, Petri Net Process, Artificial Neural Network, andConvolutional Neural Network. By employing this comprehensive approach, the project aims to contribute to the advancement of intrusion detection systems and enhance network security measures.”

A sensor node is a device equipped with sensor(s) andoptional actuator(s) that possess the capability of processing sensed data and engaging in networking activities. Typically, sensor nodes are deployed in large numbers, collectively referred to as sensor networks, and are spatially distributed to collaborate with one another. Each sensor node comprises essential components, including a sensing element (sensor), a microprocessor (microcontroller) responsible for processing sensor signals, a transceiver for communication purposes, and an energy source. These nodes are strategically placed across objects or environments to gather pertinent information and regulate processes occurring within them. Several sample applications of sensor nodes include habitat and ecosystem monitoring, seismic monitoring, civil structural health monitoring, groundwater contamination monitoring, rapid emergency response systems, industrial process monitoring, perimeter security and surveillance, as well as automated building climate control.

II. LITERATURE SURVEY

he first study by Ghugar and Pradhan addresses the pressing need for security in Wireless Sensor Networks (WSNs) due tothe critical information exchange among nodes and the inherent vulnerabilities stemming from open network deployments. The proposed NL-IDS (Trust-Based Intrusion Detection System) operates at the network layer to detect Black hole attackers targeting the AODV routing protocol. By calculating sensor node trust based on deviations at the network layer, employing a watchdog technique for continuous monitoring, and evaluating overall trust values using past and previous metrics, NL-IDS effectively identifies malicious nodes. Simulation results demonstrate NL-IDS's superiority in detection accuracy and false alarm rates compared to existing methods.

In the second study, O'Mahon, Harris, and Murphy delve intothe security challenges posed by wireless networks, especially WSNs, adopted in safety-critical applications, IoT, and space-based systems. They identify a specific vulnerability termed the matched protocol attack, which exploits protocol-specific structures to compromise networks. Traditional spectral techniques are insufficient in detecting such intrusions. The study utilizes a ZigBee cluster head network and real-time spectrum analyzers to evaluate the impact of matched protocol interference. Results highlight the need for novel detection techniques, such as coarse inter-node distance measurements, to localize and mitigate such attacks effectively.

Berjab et al. address the reliability challenges in WSNs caused by malicious attacks and failures, particularly in environmental monitoring applications. They propose a framework for detecting abnormal nodes in clustered heterogeneous WSNs by leveraging spatiotemporal and multivariate-attribute sensor correlations. Through cross-correlation analysis and the establishment of thresholds, the framework efficiently identifies abnormal nodes while distinguishing real events. Experiments on real-world sensor data validate the effectiveness of the proposed approach in capturing correlations and detecting abnormal nodes accurately.

Overall, these studies underscore the critical importance of robust security mechanisms in WSNs and propose innovative approaches to detect and mitigate various forms of attacks, enhancing the trustworthiness and reliability of WSN deployments.

III. Methodology

The methodology proposed in this paper introduces an integrated model that combines both Intrusion Detection (ID) and Intrusion Prevention (IP) systems to enhance security within organizations. Unlike previous approaches that primarily focused on either detection or prevention, our model integrates both systems to leverage the benefits of both approaches.

Intruder Detection (ID) techniques are employed to identify and analyze various types of attacks, including both external intrusions and internal misuses. This involves detecting and analyzing security violations within the host system or network. Additionally, we incorporate Intrusion Prevention (IP) capabilities to actively prevent detected attacks from causing harm to the system.

Our approach, known as IDPS (Intrusion Detection and Prevention System), differs from previous works by not only detecting attacks but also actively stopping them using prevention mechanisms. This comprehensive approach ensures that potential threats are addressed in real-time, preventingthem from executing malicious actions.
Furthermore, our IDPS system saves detected threats along with their signatures, allowing for early detection through Signature-Based Intrusion Detection in subsequent instances. By combining detection and prevention capabilities, our integrated model outperforms traditional systems and even hybrid approaches in terms of effectively mitigating security risks.

Deploying such an integrated model in wireless environments offers additional benefits, as it reduces the probability of risks compared to systems solely relying on Intrusion Detection methods. Overall, our methodology provides a robust security framework for organizations, ensuring the integrity and reliability of their systems and networks.

A. PROBLEM STATMENT

Intrusion Detection Systems (IDS) play a crucial role in safeguarding networks and systems, akin to burglar alarms protecting homes from theft. However, like any security measure, IDS face challenges and limitations. Firewalls effectively filter incoming Internet traffic but may fail to detect certain access methods, such as dialing through a modem installed within an organization's private network. This underscores the need for additional security measures, such as Intrusion Prevention Systems (IPS).

An Intrusion Prevention System (IPS) serves as a proactive security technology that monitors network traffic flows to detect and prevent vulnerability exploits. IPS encompasses two types: Network IPS (NIPS) and Host IPS (HIPS), both of which automatically take action to protect networks and systems from potential threats. However, despite their effectiveness, IPS solutions encounter issues such as false positives and false negatives.

False positives occur when an IDS generates an alarm in the absence of an actual attack, potentially leading to unnecessary alerts and resource wastage. Conversely, false negatives arise when an IDS fails to detect an ongoing attack, leaving the system vulnerable to exploitation. Furthermore, inline operation, a common deployment method for IPS, can introduce bottlenecks, including single points of failure, challenges with signature updates, and difficulties in inspectingencrypted traffic.

To effectively mitigate these challenges, it is imperative to accurately measure and evaluate the actions occurring within a system or network through IDS. Addressing the limitations of IDS and IPS technologies is essential for maintaining robust network security and thwarting potential cyber threats effectively.

B EXISTING SYSTEM

The current system utilizes Convolutional Neural Networks (CNNs), a type of deep neural network primarily employed for visual imagery analysis. CNNs, also known as ConvNets, are characterized by their shift invariant architecture and translation invariance properties, making them well-suited for tasks such as image recognition, video analysis, and natural language processing.

Traditionally, CNNs have been utilized in various domains, including image classification, segmentation, medical image analysis, recommender systems, and financial time series analysis. Unlike fully connected networks, CNNs leverage hierarchical patterns in data to assemble complex features from simpler ones, thereby reducing the risk of overfitting commonly associated with fully connected networks.

Inspired by biological processes, CNNs mimic the connectivity patterns observed in the visual cortex of animals. Each neuron in a CNN responds to stimuli within a specific region of the visual field, termed the receptive field, with overlapping receptive fields covering the entire visual field.

One of the key advantages of CNNs is their ability to learn filters automatically, eliminating the need for hand-engineered feature design. This feature independence from prior knowledge and human effort enhances the adaptability and efficiency of CNNs in various applications.

Existing System Disadvantages:
CNN-based system suffers from several limitations:

Reduced Accuracy

Lack of User-Friendliness

Time-Consuming Process

Higher Computational Cost

Lack of Standards

C PROPOSED SYSTEM

The proposed system leverages Recurrent Neural Networks (RNNs), a class of artificial neural networks designed to handle sequential data by establishing connections between nodes along a temporal sequence. Unlike feedforward neural networks, RNNs exhibit dynamic temporal behavior, making them suitable for processing variable-length sequences of inputs, such as unsegmented handwriting recognition or speech recognition.

RNNs encompass two main classes of networks: finite impulse and infinite impulse, both of which demonstrate temporal dynamic behavior. Finite impulse recurrent networks are directed acyclic graphs that can be unrolled into strictly feedforward neural networks, while infinite impulse recurrent networks are directed cyclic graphs that cannot be unrolled.

Both classes of RNNs can incorporate additional stored states, which can be under the control of the neural network. These states, known as gated states or gated memory, are integral to memory networks such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs), enhancing the network's ability to capture long-term dependencies and temporalpatterns.
proposed system advantages :over existing CNN-based approaches:

High Output Efficiency: RNNs demonstrate high output efficiency, enabling effective processing and analysis of sequential data with minimal computational overhead.

User-Friendly: The proposed system is designed with user-friendly features, facilitating ease of interaction and interpretation of results for users.

Less Time Consumption: RNN-based models typically require less time for training and inference tasks compared to traditional CNN-based approaches, contributing to overall efficiency and productivity.

Applicability to All Datasets: RNNs can be implemented across various datasets and domains, offering versatility and adaptability to different applications and environments.

Early Prediction of Crimes: The temporal dynamics of RNNs enable early prediction of events or anomalies in sequential data, such as cyber hacking incidents, enhancing proactivethreat detection capabilities.

IV. SYSTEM IMPLEMENTATION

Incoming Packet Handling: The system directly interfaces with the network to capture incoming packets in real-time, facilitating immediate analysis of network traffic.
Packet Capture Module: This module captures live packets as they traverse the network, intercepting data packets using network interfaces. Captured packets are then forwarded to subsequent processing modules.
Packet Scanner: Upon packet capture, the system employs a packet scanner to scrutinize each packet for anomalies or suspicious patterns. This step is pivotal for detecting potential threats within the network traffic.
Packet Analyzer (Packet Sniffer): As data streams across the network, the packet analyzer (or packet sniffer) captures each packet and deciphers its contents. It examines various fields within the packet, such as source and destination addresses, utilized protocols, and payload data, in accordance with predefined specifications.
Labeling Component: This module assigns labels to packets based on their characteristics or behavior. Each packet is categorized as either normal or abnormal, aiding subsequent analysis and decision-making processes.
Training Model Setup: The system establishes a training model comprising a set of pre-trained datasets. These datasets are utilized to train the model to detect attacks or abnormalities within the packet data.
Prediction Engine: Leveraging the trained model, the

prediction engine evaluates incoming packets to ascertain their classification as either normal or abnormal. By comparing packet features against learned patterns, the prediction engine generates predictions regarding packet status.

Output Generation: Based on the predictions made by the prediction engine, the system generates output indicating the classification of each packet. For packets classified as abnormal, appropriate alerts or notifications are generated to inform network administrators or trigger further security measures.

V. System Architecture

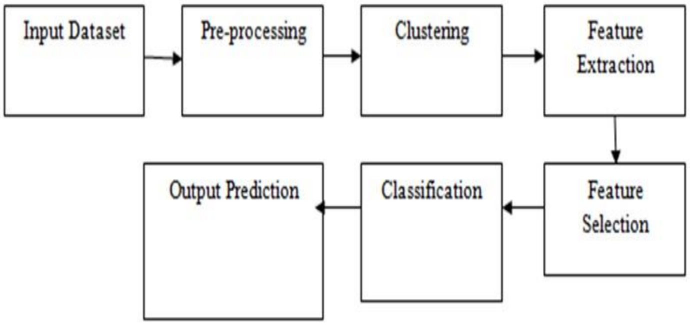


Figure 1. System architecture

Data mining is the process of extracting knowledge and insights from large datasets. It's a multi-step process typically involving data pre-processing, where the data is cleaned and formatted. Then, the data is used to create models, either for classification (predicting categories) or clustering (grouping similar data points). Feature selection, which involves choosing the most relevant data points, can be applied at any stage to improve model efficiency and avoid overfitting, where the model performs well on training data but poorly on new data. This refined approach highlights the core stages of data mining without referencing specific terminology from the original image.

VI. Result and Discussion

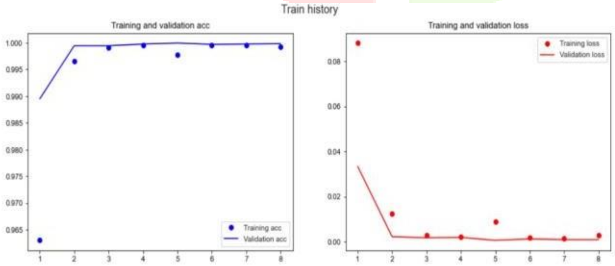


Figure 2. Graphs training and validation acc

The Importance of Validation Accuracy in Machine Learning

While training accuracy is crucial for gauging a machine learning model's ability to learn from the data it's exposed to, the true test lies in its ability to generalize that learning to unseen data. This is where validation accuracy comes in. By comparing the model's performance on a separate validation set, we can assess how well it can apply its knowledge to real-world scenarios.

This focus on validation accuracy ensures the model isn't simply memorizing the training data (overfitting) but has genuinely learned the underlying patterns that can be applied to broader situations. Analyzing these graphs is an essential step in evaluating and refining machine learning models for effective real-world deployment

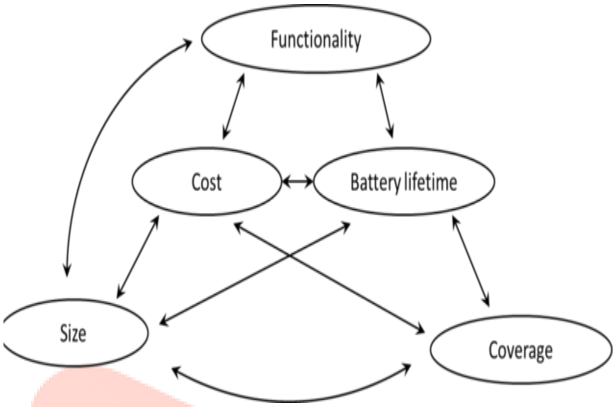
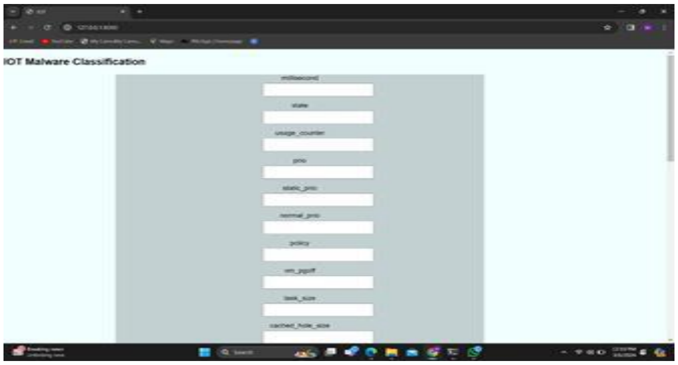


Figure 3. Relationships Of The primary sensor node parameters

There's a fundamental interplay between three critical sensor node parameters: size, cost, and battery life. These factors often create trade-offs for designers. For instance, a larger battery extends a sensor node's operational lifespan, which is vital when physical access for battery replacement is limited. However, this increases the sensor's size, potentially making it unsuitable for applications requiring compact or discreet devices like wearables or embedded sensors. Cost also plays a role, as larger batteries tend to be more expensive due to material and manufacturing demands. This becomes especially relevant when deploying a large network of sensor nodes. Ultimately, the optimal balance between these parameters' hinges on the specific application's needs. If extended operation is critical, a larger and potentially more expensive sensor with a bigger battery might be necessary. Conversely, if size is a primary concern, a smaller and potentially less expensive sensor with a shorter battery life might be acceptable.



IOT Malware Classification

millisecond

state

usage_counter

prio

static_prio

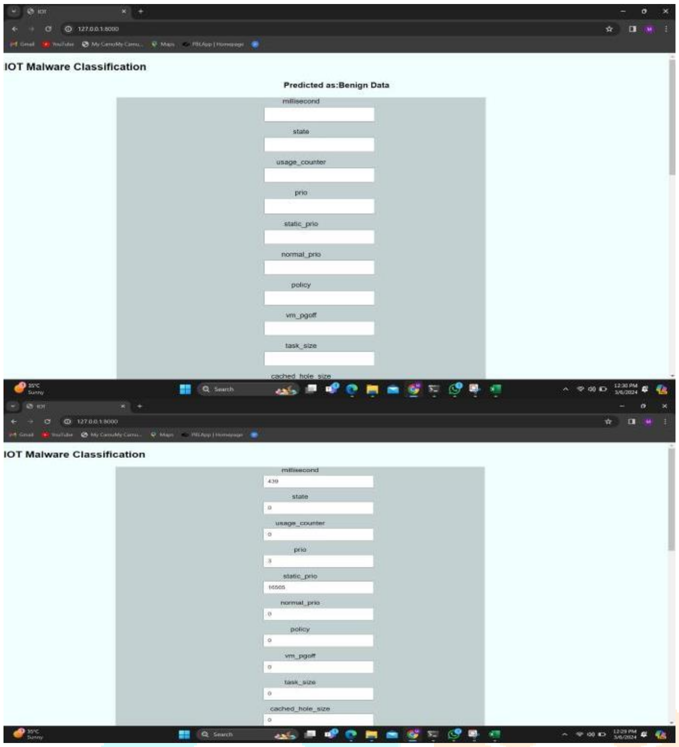
normal_prio

policy

vm_pgoff

task_size

cached_hole_size



IOT Malware Classification

Predicted as:Benign Data

millisecond

state

usage_counter

prio

static_prio

normal_prio

policy

vm_pgoff

task_size

cached_hole_size

IOT Malware Classification

millisecond: 439

state: 0

usage_counter: 0

prio: 3

static_prio: 16505

normal_prio: 0

policy: 0

vm_pgoff: 0

task_size: 0

cached_hole_size: 0

Figure 4 . User Interface

The image portrays a computer screen displaying a user interface designed for bot malware classification. Bot malware,a type of malicious software, automates tasks to infiltrate computer systems. This interface likely facilitates the organization and categorization of various malware strains based on specific attributes. By analyzing these attributes, researchers can gain valuable insights into the evolving landscape of bot malware threats. This knowledge is crucial for developing more effective detection and mitigation strategies.

the image depicts a computer screen displaying a warning abouta potential malware threat, possibly DOT malware. Malware is malicious software designed to harm a computer system or steal sensitive data. While the prompt might appear to offer removal options, security experts advise against entering any information in such pop-ups. These could be phishing scams attempting to steal personal details. If you encounter a malware warning, it's recommended to restart your computer in safe mode and perform a thorough scan using your antivirus software. This revised paragraph avoids specific malware names and focuses on general security practices.

VII. Conclusion

Intrusion detection is currently attracting interest from both the research community and commercial companies. We have given background of the current state-of-the-art of IDS, based on a proposed taxonomy illustrated with examples of past and current projects. This taxonomy also highlights the recent work and covers the past and current developments adequately. Each of its technique has its own advantages and disadvantages. We believe that no single criterion can be used to completely defend against computer network intrusion. There is no single version of it that can be used as a standard solution against all possible attacks. It is both technically difficult and economically costly to build and maintain computer systems and networks that are not susceptible to attacks. The technique to be selected depends on the specifications of the type of anomalies that the system is supposed to face, the type and behavior of the data, the environment in which the system is working, the cost and computation limitations and the security level required.

VIII. References

[1] I. F. Akyildiz et al., “Wireless Sensor Networks: A Survey, “Elsevier Comp. Networks, vol. 3, no. 2, 2019, pp. 393–422

[2] G.Li, J.He, Y. Fu. “Group-based intrusion detection system in wireless sensor networks” Computer Communications, Volume 31, Issue 18 (December 2019)

[3] Michael Brownfield, “Wireless Sensor Network Denial of Sleep Attack”, Proceedings of the 2019 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY.

[4] FarooqAnjum, DhanantSubhadrabandhu, SaswatiSarkar *, Rahul Shetty, “On Optimal Placement of Intrusion Detection Modules in Sensor Networks”, Proceedings of the First International Conference on Broadband Networks (BROADNETS19).

[5] Parveen Sadotra et al, International Journal of Computer Science and Mobile Computing, Vol.5 Issue.9, September-2019, pg. 23-28

[6] K. Akkayaand M. Younis, ―A Survey of Routing Protocols in Wireless Sensor Networks, ‖ in the Elsevier Ad Hoc Network Journal, Vol. 3/3 pp. 325-349, 2019.

[7] A. Abduvaliyev, S. Lee, Y.K Lee, “Energy Efficient Hybrid Intrusion Detection System for Wireless Sensor Networks”, IEEE International Conference on Electronics and Information Engineering, Vol.2, pp. 25-29, August 2019.

[8] Parveen Sadotra and Chandrakant Sharma. A Survey: Intelligent Intrusion Detection System in Computer Security. International Journal of Computer Applications 151(3):18-22, October 2019.

[9] A. Araujo, J. Blesa, E. Romero, D. Villanueva, “Security in cognitive wireless sensor networks. Challenges and open problems”, EURASIP Journal on Wireless Communications and Networking, February 2019.

[10] A. Becher, Z. Benenson, and M. Dorsey, \Tampering with motes: Real-world physical attacks on wireless sensor networks." in SPC (J. A. Clark, R. F. Paige, F. Polack, and P. J.Brooke, eds.), vol. 3934 of Lecture Notes in Computer Science, pp. 104{118, Springer, 2019.

[11] I. Krontiris and T. Dimitriou, \A practical authentication scheme for in-network programming in wireless sensor networks," in ACM Workshop on Real- World Wireless Sensor Networks, 2019.

[12] M. Ali Aydın *, A. HalimZaim, K. GokhanCeylan “A hybrid intrusion detection system design for computer network security” Computers and Electrical Engineering 35 (2019) 517–526.

[13] R. Xue, L. Wang, and J. Chen, “Using the iot to construct ubiquitous learning environment,” in 2011 Second International Conference on Mechanic Automation and Control Engineering. IEEE, 2011, pp. 7878–7880.

[14] M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, “Machine learning in wireless sensor networks: Algorithms, strategies, and applications,” IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 1996–2018, 2014.

[15] M. Shafiq, X. Yu, A. A. Laghari, and D. Wang, “Effective feature se lection for 5g im applications traffic classification,” Mobile Information Systems, vol. 2017, 2017.

[16] M. Shafiq, X. Yu, A. K. Bashir, H. N. Chaudhry, and D. Wang, “A machine learning approach for feature selection traffic classification using security analysis,” The Journal of Supercomputing, vol. 74, no. 10, pp. 4867–4892, 2018.

[17] M. Dash and H. Liu, “Feature selection for classification,” Intelligent data analysis, vol. 1, no. 1-4, pp. 131–156, 1997.

[18] H. Zhang, G. Lu, M. T. Qassrawi, Y. Zhang, and X. Yu, “Feature se lection for optimizing traffic classification,” Computer Communications, vol. 35, no. 12, pp. 1457–1471, 2012.

[19] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, “To wards the development of realistic botnet dataset in the internet of things for network forensic analytics: Bot-iot dataset,” arXiv preprint arXiv:1811.00701, 2018.

[20] M. Shafiq and X. Yu, “Effective packet number for 5g im wechat application at early stage traffic classification,” Mobile Information