UTM
UNIVERSITI TEKNOLOGI MALAYSIA
INTERNATIONAL JOURNAL OF
INNOVATIVE COMPUTING
ISSN 2180-4370
Journal Homepage : https://ijic.utm.my/

Hybrid of Supervised Learning and Optimization Algorithm for Optimal Detection of IoT Distributed Denial of Service Attacks

Talha Farid[1]* & Maheyzah Sirat[2]
Department of Computing, Faculty of Engineering
Universiti Teknologi Malaysia
81310 UTM Johor Bahru, Johor, Malaysia
Email: faridtalha@graduate.utm.my[1]*, maheyzah@utm.my[2]

Submitted: 31/7/2021. Revised edition: 16/5/2023. Accepted: 16/5/2023. Published online: 30/5/2023
DOI: https://doi.org/10.11113/ijic.v13n1.329

***Abstract*—The high-speed internet has led to the development of Internet of Things (IoT) with a fundamental Three-Layer IoT architecture. However, small amount of un-indicative data captured at the end level of IoT network makes the edge IoT devices susceptible to cyber-security attacks aimed at its transport layer. The Distributed Denial of Service (DDoS) poses significant cyber-security threat to the heterogenous IoT devices which are rendered vulnerable by ineffectiveness of conventional cybersecurity softwares. The literature reveals numerous studies that employed machine learning for the mitigation of IoT DDoS attacks but they lack in terms of an extensive investigation on optimization of machine learning classifiers. Therefore, this study first evaluates the prediction performance of machine learning classification algorithms trained on an authenticated/validated real-time IoT traffic dataset. The results reveal Logistic Regression (LR) as the most effective supervised machine learning classifier for detecting IoT DDoS attacks with a prediction accuracy of 97%. Following this, another investigation on the hybridization of LR with optimization algorithms yields Grasshopper Optimizer Algorithms (GOA) as the most effective optimizer in improving its prediction accuracy to 99%. Hence, the LR hybridized by GOA is developed as the optimal IoT DDoS Attack detection solution. Thus, the study serves to lay the foundation of a data-driven approach for the mitigation of the emerging variants of malicious IoT DDoS attacks such as zero-day attacks.**

***Keywords*—Internet of Things (IoT) cybersecurity, Distributed Denial of Service (DDoS) Attack, Supervised Machine Learning**

I. INTRODUCTION

The advent of high speed internet as well as the influx of smart devices has led to the development of Internet of Things (IoT) paradigm. Originally, the term IoT was devised by Kevin Ashton [1] with reference to the domain of supply chain management. The technological advancements have led to the evolution of definition of "Things" into smart objects that not only ’sense’ the information in their surroundings while interacting with other objects but also employ internet- oriented communication channels to render communication, information transfer, and applications services. In short, the IoT is radically evolving to pave the way for a fully consolidated Future Internet [2]. In technical terms, the IoT is a network of smart objects, called "things", which are connected over the internet such that they allow remote human operation and control [3, 4].

A thorough review of the IoT development reveals that there is no standard architecture of IoT security. However, the fundamental, and, therefore, the most crucial architecture that renders an IoT network complete is the Three-Layer IoT Architecture [3-5]. This architecture comprises the perception/sensing layer, network/transport layer and application layer. The function of each of the IoT network layer is outlined as follows [5]:

1. Application Layer:
   - Combines the Internet of things technology to professional application, to realize the smart application.
   - Smart objects that provide service to the end user in the form of intelligent appliances/applications
2. Transport/Network Layer:
   - Centre of IoT network: process and relay the information received from sensing layer.

- Composed of communication gateways like internet, 3G network, Wi-Fi etc.

3. Perception/Sensing Layer:

- Collect information through sensing devices
- A variety of sensors, including temperature and humidity sensor, a two- dimension code label, RFID label, camera, GPS etc.

Overall, the IoT essentially drives smart applications in the four domains of Home, Enterprise, Utilities, and Mobile [6].

Generally, Internet-of-Things (IoT) devices encompassing both consumer (smart appliances) and industrial (critical infrastructures) usage, are categorized into two principal types in terms of cyber-security: gateway IoT devices and Edge IoT devices. Since the IoT network traffic is connected to the internet at the gateway level and that the associated traffic data is rich in terms of patterns and amount, so the gateway IoT devices are prioritized for cyber-security enhancements. On the other hand, small amount of un-indicative data captured at the end level makes the edge IoT devices susceptible to cyber-security attacks especially involving the transport layer of the IoT network. In this regard, there is a multitude of IoT cyber-attacks against this edge layer protocol where the protocol level attacks are of various types: Distributed Denial of Service (DDoS), Denial of Service attack on the Data Plane, Denial of Service Attack on the Devices, Replay Attacks, Key-based attacks, and Communication Privacy Attacks. In particular, DDoS attacks overpower the required network resources, especially services, with faked requests which ultimately cause the device failure thereby preventing the addressal of authentic requests. DDoS attacks are executed through botnets: internet-connected devices where each rogue device controls at least one zombie compute node to overwhelm the network resources. Keeping this in perspective, the DDoS attacks pose significant cyber-security threat on the IoT devices. This is primarily because the security of these highly heterogeneous IoT devices cannot be boosted by conventional cybersecurity softwares at the network edge, specifically the transport layer of IoT network, and that the IoT devices are usually equipped with inadequate compute resources.

In this regard, machine learning based solutions are playing a vital role in the cyber-security of IoT devices. Overall, machine learning can help to enhance the security of IoT devices in four multi-faceted domains: access control, secure offloading, authentication, and malware detection. As far as malware detection is concerned, the machine learning algorithms with their superior pattern recognition ability can supplement the limited computing resources of edge level IoT devices to classify large amounts of IoT network traffic data in real-time for rapid identification, and subsequent mitigation, of the DDoS attacks. For this purpose, various DDoS datasets pertaining to conventional network traffic are made available by academia, governments, and commercial organizations but these datasets substantially lack in network datasets specific to IoT devices in terms of benign and malicious traffic. Accordingly, the literature review reveals machine learning as an effective means to optimize the security of IoT devices. For instance, machine learning classifiers have been used for differentiating malicious traffic from benign network traffic for boosting smart-home IoT cyber-security against edge level IoT attacks with a high success rate in various contexts. Similarity, artificial neural network (ANN) has also been investigated as a promising supplement to explore and secure IoT network traffic in near-real-time against anomalies and intrusions especially in IoT gateway network traffic. Furthermore, deep learning solutions have also been investigated with encouraging performance outcomes in detecting DDoS attacks. Thus, these studies and many more establish the superiority of machine learning over traditional cyber- security techniques for securing IoT devices against DDoS attacks.

Although various types of numerous IoT cyber-security methods have been explored for the mitigation of DDoS attacks, most of them lack in terms of an extensive investigation of available machine learning techniques and their optimization into effective solutions for the detection and classification of unknown malicious traffic such as zero-day attacks which have null detection history. Furthermore, the machine learning datasets employed in these studies are inadequate in amount as well as in their relevance to IoT network traffic context.

This study aims to evaluate the performance accuracy of the most popular machine learning algorithms for detecting and classifying the IoT DDoS cyber-attacks to determine which is the most effective supervised learning technique for IoT DDoS attack mitigation when deployed on real-time IoT network traffic. Once determined, the best-performing machine learning algorithm will then be hybridized with a viable optimization algorithm to develop an optimal IoT cyber-security solution against DDoS attacks.

The objectives of the study are consolidated in the form of measurable and sequential outcomes, as follows:

- Determine and obtain the most viable IoT-specific real-time network traffic machine learning dataset from amongst the multitude for IoT datasets made available to the public by organizations, academia, and industry.
- Pre-process the dataset by extracting features that are meaningful for machine learning detection and classification of IoT DDoS attacks
- Train and test the most popular supervised learning algorithms on the pre- processed dataset to evaluate their performance in terms of their DDoS attack classification accuracy to choose the best machine learning model
- Hybridize the best machine learning model with the most effective optimization algorithm to optimize the performance of the chosen machine learning algorithm for developing an optimal IoT cyber-security solution against DDoS attacks.

II. LITERATURE REVIEW

The continuous evolution of the Information & Communication Technologies (ICT) has seen the emergence of Internet-of-Things (IoT) as a futuristic paradigm of networking and communication [7]. This paradigm of IoT is based on driving internet application for inter-connection between physical objects involving human input to achieve a particular service(s) of interest. Conceptually, the IoT complements the elements of Internet domain, such as terminals, routers, and hosts, by employing smart objects (simply referred to as things) that are capable of identifiability by themselves as well as internet-based communication and interaction with other network entities and/or end user(s), hence, ensuring the accessibility of each smart object over the internet [8]. In practical terms, Internet of Things underpins the notion of direct communication from one machine to another over the internet. These machines include common network devices and services which employ their sensing, processing, and networking capacities to attain a useful communication objective via internet [9]. This is possible owing to the fundamental building blocks of IoT systems namely wireless sensor networks (WSNs) and radio-frequency identifications (RFIDs) [10]. In this way, the IoT serves to integrate the physical world with computer network architecture as well as applications/appliances that may or may not be controlled by end user(s). Subsequently, the IoT has driven extensive applications ranging from smart home appliances to cost-effective infrastructure development [6].

IoT is a revolutionary technology that has the ability to integrate communication between smart devices and machines by the virtue of Internet to automate the workflow. However, a multitude of cyber-threats at each layer of a fundamental three-layer IoT architecture presents a major obstacle to the evolution and widespread adoption of IoT technology [11].

The IoT perception layer, also known as the sensing layer, aims to collect information via sensor. This makes the perception layer susceptible to cyber-attacks that intend to fake the sensor data:

1. Eavesdropping: Eavesdropping is an exploitation attack that targets private communications to steal network-based information that is being transmitted over the associated network [12].
2. Nodes Capture: In this deadly cyber-attack, the attacker attains illegal control of a key node such that the key node enables the attacker to leak the information transmitted between the sender and receiver [13].
3. Fake Node and Malicious: This attack aims at destroying the IoT network by preventing the transmission of genuine information through the transmission of bogus data into network via fake node [14].
4. Replay Attack: In this attack, the intruder exploits the vulnerability of the authentic information transmitted by the sender to deceive the receiver into taking the action desired by the intruder [15].
5. Timing Attack: This attack takes advantage of the poor computing ability of the IoT devices to steal secrets stored in the IoT system security architecture [16].

The IoT network layer, also known as the transport layer, serves as a bridge for information transmission between the perception layer and application layer. This makes this layer a prominent target for IoT cyber-attacks that manipulate the integrity of the information being transmitted:

1. Denial of Service (DoS) Attack: This attack is aimed at preventing the genuine user from availing the IoT services by flooding the IoT network with bogus requests. The deadliest form of DoS attack is the Distributed Denial of Service (DDoS) attack [17].
2. Main-in-The-Middle (MiTM) Attack: In this attack, the privacy of the communication between the sender and receiver is at stake since the attacker manipulates the information between the sender and receiver in real time [18].
3. Storage Attack: The storage attacks target the user data stored on devices or cloud to replace it with fake or incorrect data that is detrimental to the information transmission within the IoT network [11].
4. Exploit Attack: As the name suggests, the attack exploits the system vulnerability to either gain unwanted control of the system resources or steal the information stored in it [11].

The IoT application layer comprising smart objects, or ’things’, that deliver the IoT services to the end-user such as smart homes, smart cities, smart government etc. However, in relaying smart services, the IoT application layer is prone to a multitude of cyber-security threats:

1. Cross Site Scripting: This attack is classified as an injection attack which manipulates the information displayed on smart applications in an illegal method [19].
2. Malicious Code Attack: This attack executed through coding in the device system/software produces unwanted damage to the system [11].

Given the pivotal role of IoT in integrating various network devices through internet-based communication to provide smart services, it is imperative that the privacy and protection of the end user data against cyber-security attacks is ensured. This is because, owing to the basis of the IoT architecture on smart objects, networks, and services, the IoT systems are vulnerable to network attacks such as jamming, denial of service (DoS) and spoofing attacks $. In particular, the most prominent IoT attack model is the DoS attack which involves overwhelming the target server with bogus requests to sabotage the services offered by IoT devices [10]. Amongst the distributed and ordinary types of DoS attacks, the Distributed Denial of Service (DDoS) cyber-attack commands the greatest security threat to IoT systems by deploying thousands and millions of service requests to distributed IoT devices via internet-based protocol addresses which deprives the system of its capability to differentiate genuine service requests from attacks, hence making the IoT systems susceptible to crash.

However, the cyber-security of IoT systems is faced with a multitude of challenges [20]. The outstanding challenges that make the security of IoT networks difficult to address are basically two-fold [21]: the heterogeneous nature of IoT networks - a variety of communication methods, various services/devices, and numerous system configurations involved - and the ever-increasing number of devices that usually comprise the current state-of-the-art IoT networks [22]. This means that the sheer volume of traffic data generated, especially in case of IoT DDoS cyber-attacks, makes it prohibitively expensive to execute cyber-security mitigation solutions effectively [23, 24]. Keeping this in perspective, machine learning is equipped with the ability to learn from massive datasets (similar in volume to the ones generated in IoT network traffic) such that, during data-driven learning, the predictive capability of the machine learning model continuously enhances for intelligent decision making [25]. Hence, machine learning has emerged as a cost- effective approach to complement IoT cyber-security techniques against DDoS attacks compared to conventional methods [26].

From cyber-security perspective, the overall architecture of the IoT system comprises three layers where each layer is vulnerable to a different attack based on its function, as explained by Xingmei, X *et al.* [5]. Firstly, the sensing layer, made up of sensing devices for information collection, is mainly susceptible to RFID and wireless sensor network security breach. Secondly, the transport layer, which serves to process and relay the information received from sensing layer, is composed of communication gateways like internet, 3G network, Wi-Fi etc. face the major security threat of DoS attacks, especially DDoS attacks. Finally, the application layer comprising of smart objects that provide service to the end user in the form of intelligent appliances/applications that often run the risk of user's privacy leakage and data exploitation. A number of machines learning based approaches involving unsupervised, supervised and reinforcement learning have been employed for IoT cyber-attack mitigation at different layers of the IoT network [27].

Numerous studies have been conducted aimed at the cyber protection of application layer of IoT network using machine learning. In this regard, a study [28] established the effectiveness of machine learning in cost savings to detect the state of an IoT element through the combination of a perceptron network with multiple layers in conjunction with a probabilistic neural network. The study concluded the effectiveness of a probabilistic neural network for discovering the state of an IoT element through the exploration of related values in the past using perceptron network. Another investigation [29] has employed machine learning to IoT system in the context of edge computing based smart home system to classify mutated codes from the regular ones. This detection system is limited to only one classification algorithm namely support vector machine and does not involve DDoS attack detection to secure the IoT edge devices. Recently, machine learning models have also been trained on the application layer for intrusion detection with a performance accuracy of up to 97.2% [30].

A new dimension to the data-driven cyber security of IoT networks involves the mitigation of DoS attack in distributed IoT networks enhanced by fog nodes. In this regard, an investigation [31] establishes the effectiveness of deep learning models compared to the classical/shallow learning models for DoS attack detection in distributed IoT networks enhanced by fog nodes. This research endeavor presented a unique approach based on deep learning for detecting cyber-attack in Fog-to-things Computing. The authors have proved the superiority of shallow models in terms of scalability, detection accuracy and false alarm rate. Similar, results underpinning the prevalence of deep learning models over shallow machine learning models have also been conducted in another study [32]. However, both the studies are limited to DoS attack under of fog to things computing context and lack in further investigating the hybridization of the classical machine learning models like kNN, SVM etc. with optimizers to investigate their performance in classifying the IoT network traffic as malicious or benign.

Limited studies have been conducted aimed at the cyber security of IoT network against DDoS attacks. For instance, in [33] the author has investigated the performance of SVM in conjunction with deep learning-based auto-encoder for network intrusion detection that has not been extended to DDoS attack mitigation in transport layer of IoT networks. Another major study [21] drives machine learning application in IoT system by using Artificial Neural Networks (ANNs) for the detection of anomalous data sent from IoT edge devices with the goal of overall IoT system security instead of dedicated approach to tackle the deadliest cyber-attack of DDoS. Although successful endeavors have been conducted in the application of machine learning models for the DDoS attack detection, but these investigations do not incorporate optimizers for the development of optimum data driven IoT DDoS attack detection technique at the transport layer. For example, Christopher *et al.* [34] investigates machine learning based multi-class detection of DDoS attack vectors generated by mirai botnet such that the implemented machine learning approaches are constrained to deep learning models only. Likewise, Ruchi *et al.* [26] investigates the performance accuracy of the machine learning classifiers trained on honeypot generated dataset, without optimizers, for detecting IoT Botnet DDoS attacks. Along similar lines, Monika *et al.* [35] extensively explores the dedicated machine learning based mitigation solution against DDoS attacks by training deep learning models as well as machine learning classifiers on CICIDS2017. This study concludes the effectiveness of hybrid deep learning model, CNN+LSTM, compared other machine learning algorithms in terms of performance accuracy for detecting DDoS attacks. However, the study does not investigate the role of optimizers with data-driven models for exploring their effectiveness in enhancing the DDoS attack detection performance of machine learning algorithms especially in comparison to the deep learning models. Although, the study does not investigate feature election of dataset for machine learning training but recommends it for larger datasets with many attributes to ensure reduced training time and, therefore, computational cost savings.

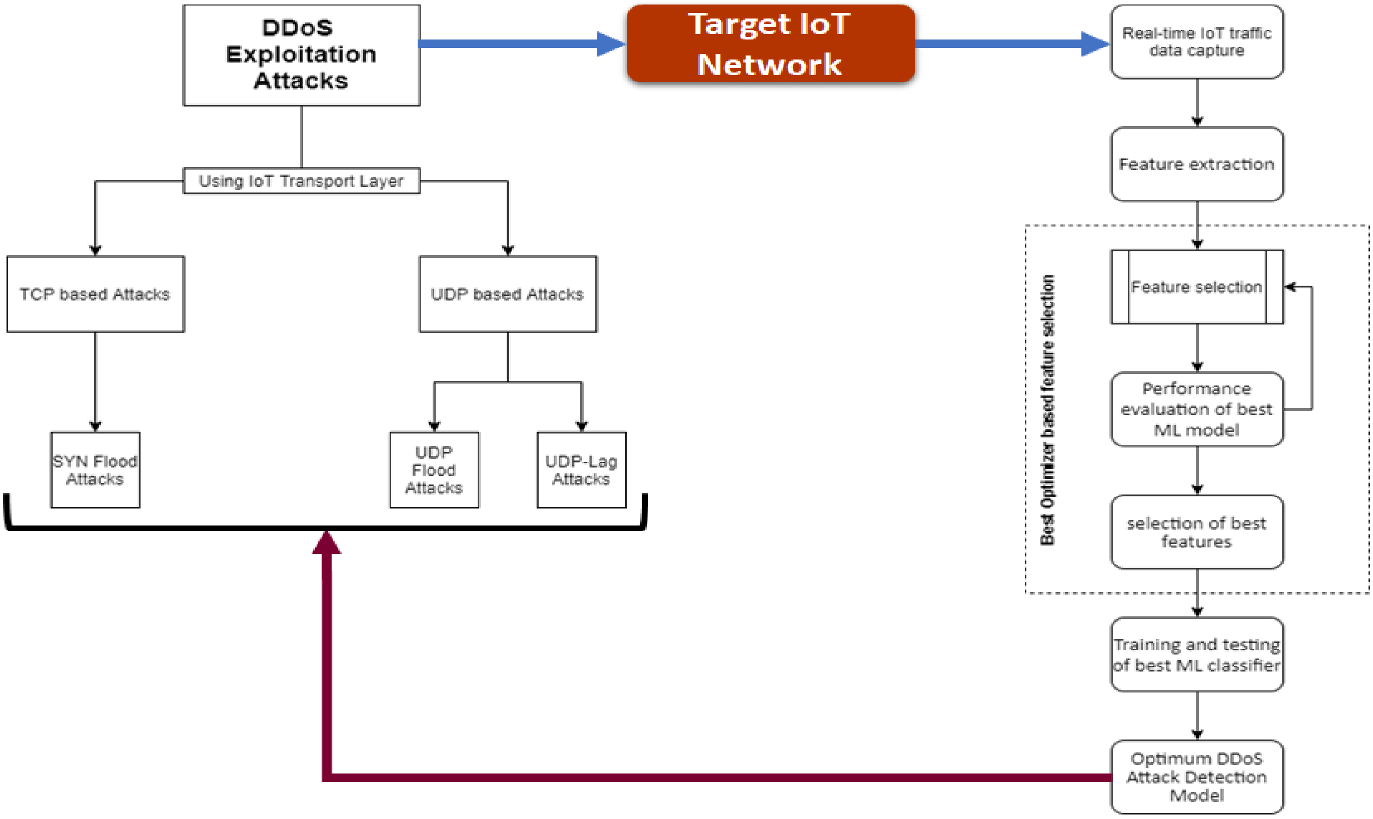
DDoS Exploitation Attacks
Target IoT Network
Real-time IoT traffic data capture
Using IoT Transport Layer
Feature extraction
TCP based Attacks
UDP based Attacks
Best Optimizer based feature selection
Feature selection
Performance evaluation of best ML model
SYN Flood Attacks
UDP Flood Attacks
UDP-Lag Attacks
selection of best features
Training and testing of best ML classifier
Optimum DDoS Attack Detection Model

Fig. 1. Process flow of machine learning hybridized with optimizer for optimal IoT DDoS attack detection

Although various types of data-driven IoT cyber-security methods have been explored for the mitigation of DDoS attacks, most of them lack in terms of an extensive investigation of available machine learning techniques and their optimization into effective solutions for the detection of unknown malicious traffic such as zero-day attacks which have null detection history. Furthermore, the machine learning datasets employed in these studies are inadequate in amount as well as in their relevance to IoT network traffic context. Keeping the previous works in perspective, this study drives its novelty in complementing the transport layer cyber-security with optimized machine learning approach against the deadliest security threat of DDoS attack which directly targets the center of the whole IoT network using internet. This investigation aims to address the following research gaps considering the limitations highlighted in the literature, as discussed in the previous section:

- An extensive investigation into the machine learning based DDoS attack detection in IoT network is conducted through critical performance evaluation of machine learning models, namely k-Nearest Neighbor (k-NN), Support Vector Machine (SVM) and Logistic Regression (LR), for classification of real- time IoT traffic.
- Development of an optimal IoT DDoS attack detection solution through hybridization investigation, involving Ant Colony Optimization (ACO), Artificial Bee Colony algorithm (ABC), Crow Search Algorithm (CSA), and Grasshopper Optimization Algorithm (GOA) for the juxtaposition of the best machine learning classifier with an optimizer algorithm.

This development of hybrid supervised learning and optimization algorithm for optimal detection of IoT DDoS Attacks is carried out in four steps which are explained in the subsequent sections: real-time IoT traffic dataset generation, feature extraction, supervised machine learning training and testing, and optimization. In this way, this study serves to lay the foundation of a data-driven approach for the of mitigation of the emerging variants of malicious IoT DDoS attacks namely zero-day attacks.

III. PROPOSED IOTD DOS ATTACK DETECTION SOLUTION DEVELOPMENT

The development of hybrid of supervised learning and optimization algorithms for optimal detection of IoT DDoS Attacks is carried out in four steps which are explained in the subsequent sections:

1. Real-time IoT traffic dataset generation
2. Feature Extraction
3. Supervised Machine Learning: Training and Testing Optimization

Fig. 1 illustrates the process flow of optimal machine learning implementation for the detection of IoT DDoS attacks The proposed methodology conducts the comparison of the three machine learning classification models: k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), and Logistic Regression (LR) to detect the IoT traffic. The best performing model, gauged using estimation error, is then hybridized with an optimization algorithm, or simply an optimizer, which yields an optimal DDoS attack detection solution.

A. IoT Traffic Dataset Generation

Distributed Denial of Service (DDoS) is a cyber-attack that exhausts the IoT network with bogus traffic. In order to detect this bogus traffic from normal or benign traffic using machine learning, the IoT network traffic data must be well-designed yet collected with low computational overhead. Therefore, two real-time networks, named as ‘Target IoT network’ and ‘DDoS Attack network’, have been implemented to generate an IoT traffic dataset for machine learning in [36]. Since, this dataset has been validated and authenticated [37] so it has been employed in this study for the development of optimal data driven IoT DDoS attack detection solution.

The Target IoT network is a remotely deployed depiction of IoT network. It comprises a web-based server, two switches and four remotely deployed PC sessions running Windows®/Ubuntu® operating systems. All the send/receive traffic to the Target IoT network is completely captured through its main switch which, accordingly, has been configured as the mirror port. The benign background traffic from the Target IoT network is generated by profiling the abstract behavior of human interaction with each of the remotely deployed PC sessions based on FTP, HTTP, SSH, NETBIOS, LDAP, DNS and MSSQL protocols. In this way, the network events for each user, including DDoS attacks, are encapsulated in log files in terms of features pertaining to the packet size distribution, payload size, packet quantity and distributions of protocols’ request time.

On the other hand, the DDoS Attack network is used to deploy DDoS attacks from the transport layer of the Target IoT. A number of different DDoS attack profiles have been created from the DDoS Attack network to exploit the Target IoT network. The execution of these DDoS attacks on the Target IoT network has been carried out through related third-party tools/packages.

B. Feature Extraction

The captured traffic dataset of the Target IoT network in the form of logfiles is subject to FlowMeter, a specialized script integrated with hping [38], for the extraction of IoT traffic features in CSV file.

C. Supervised Machine Learning: Training and Testing

Although, initially, a number of IoT traffic features have been extracted but the most influential or the characteristic feature set for the DDoS attack detection is determined through the examination of training and testing performance of machine learning models on the IoT traffic dataset. These machine learning models train on the IoT traffic data from the Target IoT network to predict the classification of the IoT network traffic. In this regard, a brief over-view of each of the machine learning model is presented as follows.

1) k-Nearest Neighbor

k-Nearest Neighbor (k-NN) [39] is one of the simplest supervised machine learning technique that stores all data points (or cases) for classification into available categories (i.e. labels) based on a similarity measure in the form of distance functions. It is widely used in pattern recognition applications as a non-parametric algorithm. In this technique, the data point is assigned to a category based on the majority category of its k-nearest neighboring data points. The nearness of the data point to its k-nearest neighboring data points is quantified through either of the distance functions, provided the continuous variables are employed. These distance functions are named as Manhattan, Minkowski and Euclidean. The value of the parameter k is determined through cross-validation. The optimal value of k employed historically for majority of datasets usually falls between 3-10.

2) Support Vector Machine

Support Vector Cas (SVM) [40] is a supervised machine learning classification technique. In this technique, each data point is plotted in an n-dimensional space such that n denotes the number of features employed for classification and that the value of each feature is denoted by the value of a particular coordinate. In this way, the binary classification is performed by computing the hyper-plane(s), differentiating the two categories.

3) *Logistic Regression*

Logistic Regression (LR) [41] is supervised machine learning classification technique which is employed when the target variable is categorical in nature. In algorithmic terms, LR computes the probability of the output category in terms of the input features. In this way, it is utilized to develop a binary classifier by determining a cutoff value; if the probability of the input features is above the cutoff value, then it belongs to one class but if it is below then it belongs to the other class. The logistic function forms the basis of the probability computation for the LR binary classification.

4) *Optimization*

Once the classification algorithms have been employed then the best performing machine learning algorithm is subject to optimization by hybridizing the algorithm with an optimization algorithm. Accordingly, the optimization algorithms are investigated in terms of their improvement in the classification performance accuracy through enhancement in the feature selection from the original IoT traffic data. These optimizers are enlisted as follows:

- Ant Colony Optimization (ACO) [42]
- Artificial Bee Colony algorithm (ABC) [43]
- Crow Search Algorithm (CSA) [44]
- Grasshopper Optimization Algorithm (GOA) [45]

The best performing optimization algorithm is then hybridized with the best performing machine learning model to develop an optimal DDoS attack detection model.

IV. RESULTS AND DISCUSSION

The real time authenticated/validated [37] IoT Traffic dataset [36] used in this study for training the machine learning algorithms is characterized by a total of 80 features. However, the data pre-processing and machine learning based investigations yielded that the most significant features for making data-driven detection of IoT DDoS attacks are Average Packet Size, Maximum Packet Length and Protocol Type. Accordingly, the data labels used for classification of IoT traffic are 1 for "Malicious Traffic (infected with UDP/TCP attacks)" and 0 for "Benign Traffic". Fig. 2 illustrates the data pre-processing. Following the training of machine learning algorithms on the real time IoT Traffic dataset, the performance of each machine learning algorithm as well as the performance of hybridization of the best performing machine learning algorithm is evaluated in the form of the Confusion Matrix.

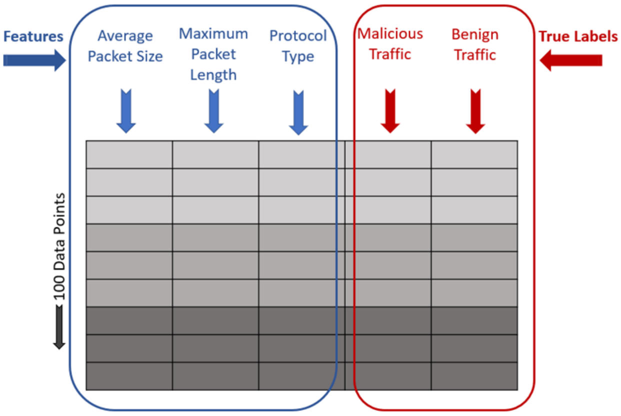

Fig. 2. Pre-processing of real-time IoT traffic experimental dataset

Overall, each of the machine learning algorithm is trained on the real time IoT Traffic Dataset such that the datapoints for training are increased until the prediction accuracy (conversely known as the estimation error) becomes independent of the number of data points. This is then followed by gauging the prediction performance of each algorithm using the Confusion Matrix.

A. k-Nearest Neighbor

k-Nearest Neighbor (k-NN) algorithm makes decision by looking at the nearest neighbors. KNN calculates the distance between the query example and the current example from the data then sorts the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances and, thus, it picks the value of K from sorted collection. The training k-NN on the real-time IoT Traffic Dataset yielded a prediction accuracy of around 86% with 500,000 samples when value of K was set to 20. With the increase in the value of K, the computation time increased without any significant impact on the prediction performance. Eventually, the data-independent prediction accuracy of 96% was attained after increasing the data points to approximately 1 million. Fig. 3 illustrates the achievement of the data-independent prediction accuracy for detecting IoT DDoS Attacks using the k- NN algorithm. Accordingly, the Confusion Matrix, as shown in Fig. 4, illustrates the prediction performance of the k-NN algorithm.

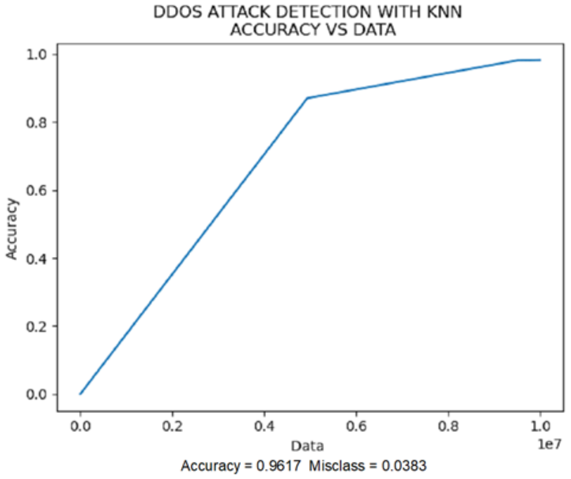

Fig. 3. Data-independent prediction accuracy results for detecting IoT DDoS Attacks using the k-NN algorithm

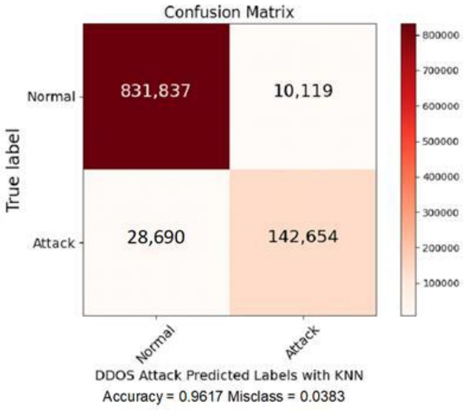

Fig. 4. Confusion Matrix for the prediction performance of k-NN

B. *Support Vector Machine*

Support Vector Machine (SVM) is a supervised machine learning model that uses classification algorithms for two-group or binary classification problems. After training the SVM model on the labelled real-time IoT Traffic Dataset [36] for each category, the algorithm is able to categorize the IoT Traffic as Benign or Malicious. Initially, the achieved prediction accuracy was 77% for a total of 500,000 data points but, with the increase in the datapoints for machine learning training, a data- independent prediction accuracy of 95% was achieved for 1 million samples. Fig. 5 illustrates the achievement of the data-independent prediction accuracy for detecting IoT DDoS Attacks using the SVM algorithm. Accordingly, the Confusion Matrix, as shown in Fig. 6, illustrates the prediction performance of the SVM algorithm.

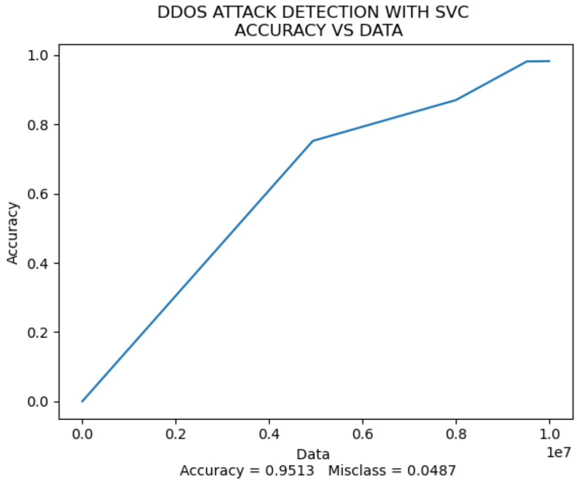

Fig. 5. Data-independent prediction accuracy results for detecting IoT DDoS Attacks using the SVM algorithm

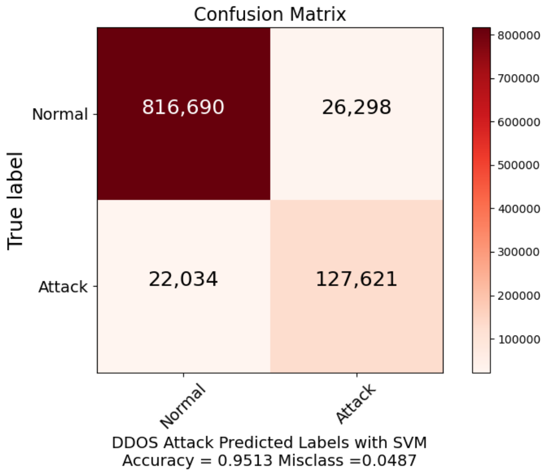

Fig. 6. Confusion Matrix for the prediction performance of SVM

C. *Logistic Regression*

The Logistic Regression (LR) algorithm is used to make binary classifications. The initial training of the LR on the real-time IoT Traffic Dataset [36] yields a prediction accuracy of 82% for a total of 500,000 samples to classify the IoT Traffic

as 'Benign' or 'Malicious'. The prediction accuracy increased with the increase in the training data points such that a data-independent prediction accuracy of 97% was attained for approximately 1 million data points. Fig. 7 illustrates the achievement of the data-independent prediction accuracy for detecting IoT DDoS Attacks using the LR algorithm. Accordingly, the Confusion Matrix, as shown in Fig. 8, illustrates the prediction performance of the SVM algorithm.

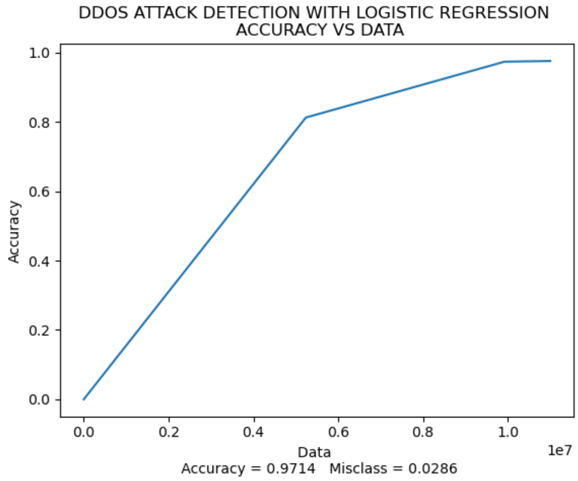

Fig. 7. Data-independent prediction accuracy results for detecting IoT DDoS Attacks using the LR algorithm

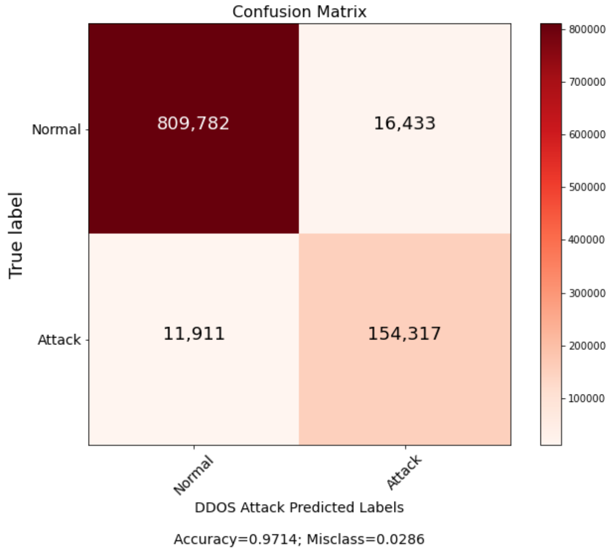

Fig. 8. Confusion Matrix for the prediction performance of LR

D. Optimization Algorithms

The investigation on the supervised learning classifiers for IoT DDoS Attack detection revealed that the best performing machine learning algorithm, in terms of prediction accuracy, is the Logistic Regression algorithm. Therefore, for the development of an optimal IoT DDoS Attack detection solution, the LR algorithm is hybridized with each of the following optimizers:

- Ant Colony Optimization (ACO)
- Artificial Bee Colony algorithm (ABC)
- Crow Search Algorithm (CSA)
- Grasshopper Optimization Algorithm (GOA)

The result of the hybridization of machine learning algorithms is shown in Figs. 9, 10, 11 and 12, respectively, which illustrate the Confusion Matrix, describing the prediction performance, pertaining to each of the optimizers.

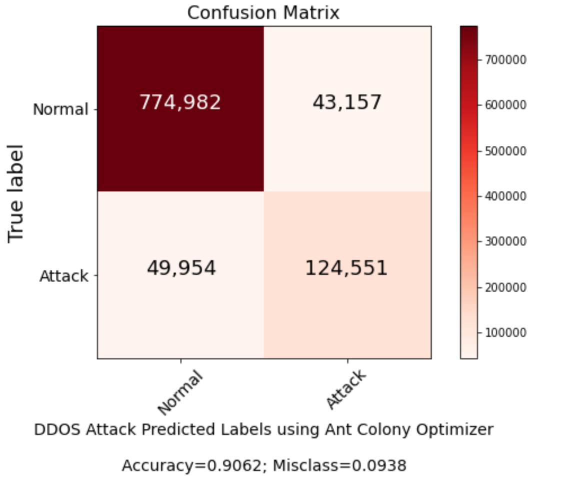

Fig. 9. Confusion Matrix for the prediction performance of Hybridized LR-ACO algorithm

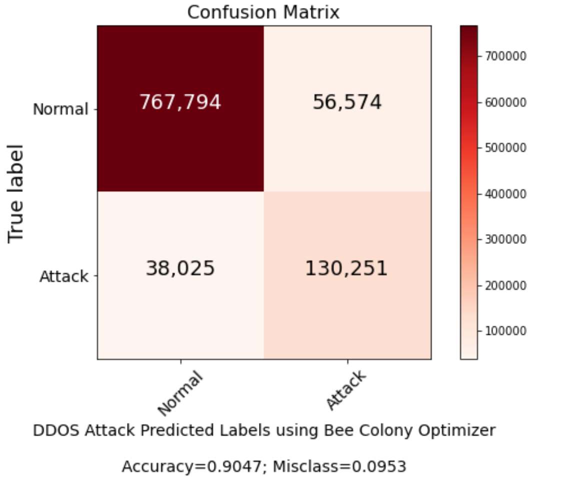

Fig. 10. Confusion Matrix for the prediction performance of Hybridized LR-ABC algorithm

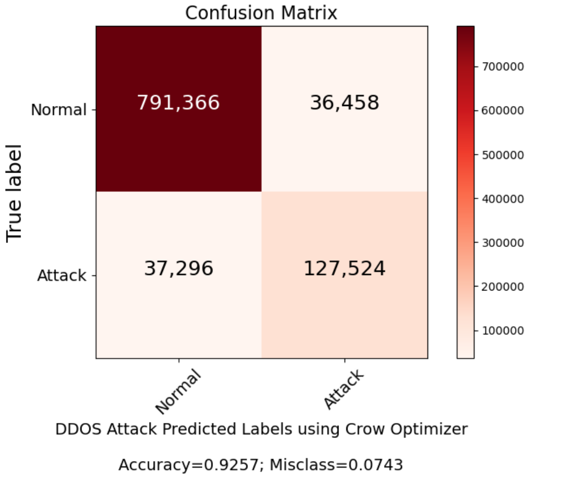

Fig. 11. Confusion Matrix for the prediction performance of Hybridized LR-CSA algorithm

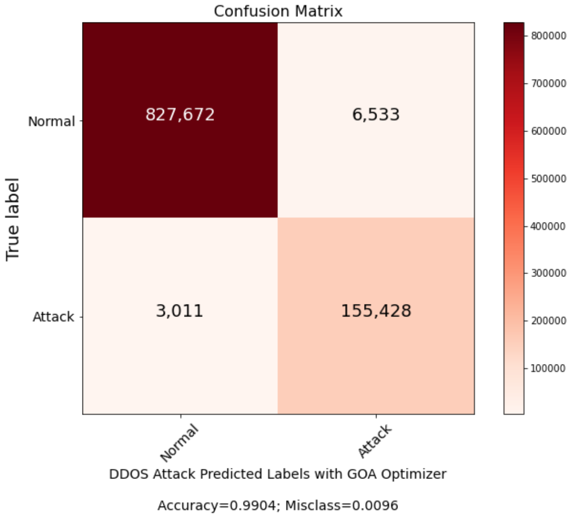

Fig. 12. Confusion Matrix for the prediction performance of Hybridized LR-GOA algorithm

Overall, the results establish the GOA Optimizer as the best performing hybridization algorithm to optimize the prediction performance of LR algorithm to a sheer 99% thereby resulting in the development of an optimal IoT DDoS Attack detection solution. It is worth mentioning that the GOA implemented on logistic regression selects the optimized features which give better accuracy than normal algorithm. The proposed GOA algorithm mathematically models and mimics the behavior of grasshopper swarms in nature for solving optimization problems. As the experimental results show, this significantly boost the performance of GOA and the proposed learning scheme can guarantee a more stable kernel extreme learning machine model with higher predictive performance compared to other optimizers.

V. CONCLUSION

The study reveals that the machine learning solutions in conjunction with the optimization algorithms provide invaluable solutions for the detection of DDoS Attacks especially in IoT networks prone to huge inflow of network data. Keeping this in perspective, this study serves to achieve two-fold primary research outcomes. The first research outcome revolves around an extensive investigation of machine learning algorithms for addressing the ever-increasing cyber-security threat of DDoS Attacks in the IoT networks. Although previous studies have been conducted along the same lines but, while building on this research goal, this study derives significant research value from its second research outcome: employing optimizers for hybridization with the machine learning algorithms for the development of an optimal IoT DDoS Attack detection solution. In fulfillment of the first research outcome of determining the best performing machine learning algorithm for the IoT DDoS Attack detection, the study compared and evaluated the prediction performance of three most effective supervised machine learning algorithms, namely k-NN, SVM and LR, for classification of the IoT ‘malicious’ traffic from ‘benign’ traffic. The results reveal that the most effective supervised 97%. Building upon this unique finding, the study further investigated the impact of optimization algorithms on the LR’s prediction accuracy, in wake of the achievement of the second research outcome. The investigation, involving four optimization algorithms namely ACO, ABC, CSA, and GOA, yielded GOA optimizer as the most effective optimizer in improving the prediction efficiency of the LR algorithm to a sheer 99%. Hence, the study established GOA-LR hybridized supervised machine learning algorithm as the optimal IoT DDoS Attack detection solution. After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

In light of this IoT DDoS Attack detection solution development endeavor, following recommendations are •presented for future work: Conduct an extensive evaluation of Deep Learning as well as Unsupervised Learning in comparison with the Supervised Learning for the IoT DDoS Attack detection

- Employ optimizer algorithms in conjunction with the Deep Learning, Unsupervised Learning as well as Supervised Learning to determine the most effective Hybridized DDoS Attack Detection Solution
- Expand the domain of investigation of this study by implementing it on other commercially available IoT traffic datasets especially Botnet IoT dataset [46]

ACKNOWLEDGMENT

The valuable feedback of Dr. Maheyzah MD Siraj and the members of the graduate evaluation committee at the UTM Department of Computing is acknowledged.

REFERENCE

[1] Ashton, K. (2009). That ‘internet of things’ thing. *RFID Journal*, *22*(7), 97-114.

[2] Buckley, J. (2006). The internet of things: From RFID to the next-generation pervasive networked systems.

[3] Atzori, L., Iera, A. and Morabito, G. (2010). The Internet of things: A survey. *Computer Networks*, *54*(15), 2787-2805. https://doi.org/10.1016/ j.comnet.2010.05.010.

[4] Giusto, D., Iera, A., Morabito, G. and Atzori, L. (2010). *The internet of things: 20th Tyrrhenian workshop on digital communications.* Springer Science &amp; Business Media.

[5] Xingmei, X., Jing, Z. and He, W. (2013). Research on the basic characteristics, the key technologies, the network architecture and security problems of the internet of things. *Proceedings of 2013 3rd International Conference on Computer Science and Network Technology*. 825-828. 10.1109/ICCSNT. 2013.6967233.

[6] Gubbi, J., Buyya, R., Marusic, S. and Palaniswami, M. (2013). Internet of things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems, 29*(7): 1645-1660. https://doi.org/ 10.1016/j.future.2013.01.010.

[7] Miorandi, D., Sicari, S., Pellegrini, F. D. and Chlamtac, I. (2012). Internet of things: Vision, applications and research challenges. *Ad Hoc Networks*, *10*(7): 1497-1516. https://doi.org/10.1016/j.adhoc.2012.02.016.

[8] Kortuem, G., Kawsar, F., Sundramoorthy, V. and Fitton, D. (2010). Smart objects as building blocks for the internet of things. IEEE Internet Computing, *14*(1), 44-51. 10.1109/MIC.2009.143.

[9] Whitmore, A., Agarwal, A. and Xu, L. D. (2015). The internet of things—A survey of topics and trends. *Information Systems Frontiers, 17*(2), 261-274.

[10] Andrea, I., Chrysostomou, C. and Hadjichristofi, G. (2015). Internet of things: Security vulnerabilities and challenges. *2015 IEEE Symposium on Computers and Communication (ISCC).* 180-187. 10.1109/ISCC.2015.7405513

[11] Burhan, M., Rehman, R. A., Khan, B. and Kim, B.-S. (2018). IoT elements, layered architectures and security issues: A Comprehensive survey. *Sensors (Basel, Switzerland), 18*(9), 2796. 10.3390/s18092796.

[12] Suo, H., Wan, J., Zou, C. and Liu, J. (2012). Security in the Internet of Things: A Review. *2012 International Conference on Computer Science and Electronics Engineering*, *3*, 648-651. 10.1109/ICCSEE.2012.373.

[13] Bharathi, M. V., Tanguturi, R. C., Jayakumar, C. and Selvamani, K. (2012). Node capture attack in Wireless Sensor Network: A survey. *2012 IEEE International Conference on Computational Intelligence and Computing Research*. 1-3. 10.1109/ICCIC.2012.6510237.

[14] Kozlov, D., Veijalainen, J. and Ali, Y. (2012). Security and privacy threats in IoT Architectures. *Proceedings of the 7th International Conference on Body Area Networks. Brussels, BEL: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).* 256-262.

[15] Puthal, D., Nepal, S., Ranjan, R. and Chen, J. (2016). Threats to networking cloud and edge datacenters in the internet of things. *IEEE Cloud Computing, 3*(3), 64-71. 10.1109/MCC.2016.63.

[16] Brumley, D. and Boneh, D. (2005). Remote timing attacks are practical. *Computer Networks, 48*(5), 701-716. https://doi.org/10.1016/j.comnet.2005. 01.010.

[17] Prabhakar, S. (2017). Network security in digitalization: Attacks and defence. *Int. J. Res. Comput. Appl. Robot*, *5*(5), 46-52.

[18] Conti, M., Dragoni, N. and Lesyk, V. (2016). A survey of man in the middle attacks. *IEEE Communications Surveys Tutorials*, *18*(3), 2027-2051. 10. 1109/COMST.2016.2548426.

[19] Gupta, S. and Gupta, B. B. (2017). Cross-site scripting (XSS) attacks and defense mechanisms: classification and state-of-the-art. *International Journal of System Assurance Engineering and Management*, *8*(1), 512-530.

[20] Mahmoud, R., Yousuf, T., Aloul, F. and Zualkernan, I. (2015). Internet of things (IoT) security: Current status, challenges and prospective measures. *2015 10th International Conference for Internet Technology and Secured Transactions (ICITST).* 336-341. 10.1109/ICITST.2015.7412116.

[23] Bhardwaj, K., Miranda, J. C. and Gavrilovska, A. (2018). Towards IoT-DDoS prevention using edge computing. {USENIX} workshop on hot topics in edge computing (HotEdge 18). Boston, MA: USENIX Association.

[24] Herzberg, B., Bekerman, D. and Zeifman, I. (2016). Breaking down mirai: An IoT DDoS botnet analysis. Incapsula Blog, Bots and DDoS, Security.

[25] Negnevitsky, M. (2011). *Artificial intelligence: A guide to intelligent systems*. Third Edition.

[26] Vishwakarma, R. and Jain, A. K. (2019). A Honeypot with machine learning based detection framework for defending IoT based Botnet DDoS attacks. *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).* 1019-1024. 10.1109/ICOEI.2019.8862720.

[27] Xiao, L., Wan, X., Lu, X., Zhang, Y. and Wu, D. (2018). IoT security techniques based on machine learning: How do IoT devices use AI to enhance security? *IEEE Signal Processing Magazine, 35*(5), 41-49. 10.1109/MSP.2018.2825478.

[28] Kotenko, I., Saenko, I., Skorik, F. and Bushuev, S. (2015). Neural network approach to forecast the state of the internet of things elements. *2015 XVIII International Conference on Soft Computing and Measurements (SCM*). 133-135. 10.1109/SCM.2015.7190434.

[29] Hou, S. and Huang, X. (2019). Use of machine learning in detecting network security of edge computing system. *2019 IEEE 4th International Conference on Big Data Analytics (ICBDA).* 252-256. 1109/ICBDA.2019. 8713237.

[30] Mahmood, M. T., Ahmed, S. R. A. and Ahmed, M. R. A. (2020). Using machine learning to secure IOT Ssystems. 4th *International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT),* 1-7. 10.1109/ISMSIT50672.2020.9254304. Conference on Security Technology (ICCST). 2019. 1-8. URL 10.1109/CCST.2019.8888419.

[31] Abeshu, A. and Chilamkurti, N. (2018). Deep learning: The frontier for distributed attack detection in fog-to-things computing. *IEEE Communications Magazine, 56*(2), 169-175. 10.1109/MCOM.2018.1700332.

[32] Vigneswaran, R. K., Vinayakumar, R., Soman, K. P. and Poornachandran, P. (2018). Evaluating shallow and deep neural networks for network intrusion detection systems in cyber security. *9th International Conference on Computing, Communication and Networking Technologies (ICCCNT).* 1-6. 10.1109/ICCCNT.2018.8494096.

[33] Al-Qatf, M., Lasheng, Y., Al-Habib, M. and Al-Sabahi, K. (2018). Deep learning approach combining sparse autoencoder with SVM for network intrusion detection. *IEEE Access, 6,* 52843-52856. 10.1109/ACCESS. 2018.2869577.

[34] McDermott, C. D., Majdani, F. and Petrovski, A. V. Botnet (2018). Detection in the internet of things using deep learning approaches. *2018 International Joint Conference on Neural Networks (IJCNN).* 1-8. 10.1109/IJCNN. 2018.8489489.

[35] Roopak, M., Tian, G. Y. and Chambers, J. 2019. Deep Learning Models for Cyber Security in IoT Networks.

(2019). *IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC).* 0452-0457. 10.1109/CCWC.2019.8666588.

[36] New Brunswick, U. (2019). DDoS Evaluation Dataset (CIC-DDoS2019), 2019. https://www.unb.ca/cic/datasets/ddos-2019.html.

[37] Sharafaldin, I., Lashkari, A. H., Hakak, S. and Ghorbani, A. A. (2019). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. *2019 International Carnahan Conference on Security Technology (ICCST).* 1-8. 10.1109/CCST.2019.8888419.

[38] hping. https://github.com/antirez/hping.

[39] Cunningham, P. and Delany, S. J. k-Nearest Neighbour Classifiers–2nd Edition.

[40] Pisner, D. A. and Schnyer, D. M. (2020). Chapter 6 - Support vector machine. https://doi.org/10.1016/B978-0-12-815739-8.00006-7.

[41] Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M. and Klein, M. (2002). *Logistic regression.* Springer.

[42] Dorigo, M., Birattari, M. and Stutzle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine*, *1*(4), 28-39. 10.1109/MCI. 2006.329691.

[43] Karaboga, D., Gorkemli, B., Ozturk, C. and Karaboga, N. (2014). A comprehensive survey: artificial bee colony (ABC) algorithm and applications. *Artificial Intelligence Review*, *42*(1), 21-57. 10.1007/s10462-012-9328-0.

[44] Zolghadr-Asli, B., Bozorg-Haddad, O. and Chu, X. Crow (2018). Search Algorithm (CSA). Bozorg-Haddad, O., ed. Advanced Optimization by Nature-Inspired Algorithms. Singapore: Springer Singapore. 143-149. 10.1007/ 978-981-10-5221-7_14.

[45] Saremi, S., Mirjalili, S. and Lewis, A. (2017). Grasshopper optimisation algorithm: Theory and application. *Advances in Engineering Software, 105*, 30-47. https://doi.org/10.1016/j.advengsoft.2017.01.004.

[46] Moustafa, N. (2019). The Bot-IoT dataset. *IEEE Dataport.*