A NOVEL PROPOSED RANDOM FOREST MODEL RESULTING IN SUPERIOR PERFORMANCE AS COMPARED TO SVM, NAIVE BAYES METHODS

[1]K.Koti Reddy, [2]P.Venkata Nagaveni, [3]I.Shasidhar, [4]Dasari Ashok

[1,2,3]Assistant Professor, [4]PGStudent, Dept. of Master of Computer Application, Newton's Institute of Engineering, Macherla, Andhra Pradesh, India.

ABSTRACT

The trends in network evolution include software-defined networking (SDN) and the internet of things (IoT). SDN primarily focuses on the administration and control of networks at a higher level, whereas IoT intends to link objects to enable sharing and real-time behavior monitoring. On the one hand, IoT gives us the ability to remotely monitor and manage devices and networks. The constantly expanding number of devices, however, poses management problems at the access and backbone layers and security issues about network assaults like Distributed Denial of Service (DDoS). SDN and IoT together create a promising strategy that might solve the management problem. Indeed, the flexibility and programmability of SDN could help in simplifying the network setup. However, there is a need to make a security enhancement in the SDN-based IoT network for mitigating attacks involving IoT devices. Therefore, this work develops various machine learning algorithms such as SVM, Random Forest, XGBOOST, ADABOOST, KNN and Naïve Bayes for detecting and predicting DDoS attacks in IoT environment. Here, CIC dataset (contains 10 different attacks in IoT and 1 normal class) is used to train the network. Finally, the simulations revealed that the proposed random forest resulted in superior performance as compared to SVM, naive bayes methods.

Keywords: Software defined network, Distributed Denial-of-Service, Internet of things

INTODUCTION

The creation of "smart cities" is being facilitated by the recent expansion of the Internet of Things (IoT), where billions of linked IoT devices are used to automate daily operations and create unique ubiquitous services (e.g., smart healthcare, smart home). Yet, since the number of vulnerable IoT devices continues to expand at a rapid rate, the effect of Distributed Denial-of-Service (DDoS) assaults is rising rapidly. With the introduction of IoT botnets like Mirai, the perception of IoT has transformed from being a smart city facilitator to being a potent amplification weapon for cyber attacks. This encourages the creation of novel methods that will enable attack collaborative decision-making to be flexible and effective in the environment of software defined networks (SDN). The new emerging technologies, such as SDN and blockchain, introduce new opportunities for low-cost, efficient, and flexible DDoS attacks collaboration for the IoT based environment. In this paper, we propose Co-IoT, a block chain-based framework for collaborative DDoS mitigation; it uses the concept of smart contracts (i.e., Ethereum's smart contracts) to facilitate the collaboration among SDN-based domains and transfer attacks information in a decentralized manner. The implementation of Co-IoT is deployedon Ethereum official test network Ropsten. The experimental results confirm that Co-IoT achieves flexibility, efficiency, security, and cost effectiveness making it a promising approach to mitigate large scale DDoS attacks. Software-Defined Networking (SDN) and Internet of Things (IoT) are the trends of network evolution.

SDN mainly focuses on the upper-level control and management of networks, while IoT aims to bring devices together to enable sharing and monitoring of real-time behaviours through network connectivity. On the one hand, IoT enables us to gather status of devices and networks and to control them remotely. On the other hand, the rapidly growing number of devices challenges the management at the access and backbone layer and raises security concerns of network attacks, such as Distributed Denial of Service (DDoS). The combination of SDN and IoT leads to a promising approach that could alleviate the management issue. Indeed, the flexibility and programmability of SDN could help in simplifying the network setup. However, there is a need to make a security enhancement in the SDN-based IoT network for mitigating attacks involving IoT devices.

DDoS attacks are a common threat to the network, although the attacker typically does not aim to steal any data. Basically, DDoS attacks aim at consuming system resources until the target is not available to offer its services. DDoS attacks could be divided into three categories: application layer attack, protocol attack and volumetric attack. For volumetric attack, an attacker can deplete the available resources of the victim or bandwidth towards the target. Not only the data plane in the SDN, but also the controller and southbound interface, could suffer from this kind of attack as well; this is because a client host can trigger inquiry from the data plane to control plane. Although there have been a lot of discussions about DDoS attacks in the SDN and IoT networks, the large number of IoT gadgets is still a good chance to launch attacks, as well as the communication link between controllers and switches in SDN. Additionally, more validations in the real network are required. Moreover, the programmability and centralised control in the SDN give users more options to probe into this threat. DDoS attacks have been well investigated for SDN networks, where the target of DDoS attackers maybe the control plane or data plane, compromising the controller or SDN switches.

In recent years, a popularization of communications networks is witnessed, which has allowed users to be connected at any time and almost anywhere, thus generating growing traffic demand. The proliferation of different smart devices and applications, as well as the development of a wide range of network technologies, are generating an unprecedented amount of data traffic. Thus, the expected traffic growth in the global Internet in 2019 will exceed 200 Exabytes per month, reaching 396 Exabytes per month in 2022. In addition, the rise in the number of objects connected to the Internet has made Internet of Things (IoT) an increasingly growing topic in recent years and it is expected that it will exponentially increase in the coming years. Several forecasts project that the current number of connected devices at the end of 2019, around 1.3 billion, will reach 5 billion IoT devices in 2025. In this context, the explosive rise of IoT is leading to the creation of new advanced services with more stringent requirements such as low response time and low energy consumption. Services such as the industrial IoT, automotive IoT or e-health are typical IoT-enabled critical infrastructures that require the network to be ready to provide the proper network capabilities and also to cope with different security challenges. Indeed, to achieve the success of IoT, it is necessary to develop advanced mechanisms able to ensure proper security levels to detect cyber-attacks and mitigate cyber-threats whenever occur in the managed IoT network. This poses a great challenge as IoT devices may handle sensitive information and many commercial IoT low-end devices do not usually support strong security mechanisms, making them easy targets to conform the malicious network of devices for different attacks such as DoS (Denial of Service) and DDoS (Distributed Denial of Service).

LITERATURE SURVEY

Dao, et al. [1] define a table in the controller to track the packets by IP address during a DDoS attack. All the new packets are regarded as suspicious packets and assigned a small timeout value in the flow entry. The number of packets using that connection is also compared with a minimum value to determine if it is a normal request or an attack. From the simulation, this method effectively reduces flow entries in the switch, and the bandwidth of controller-switch channel is still available during DDoS attacks. However, this mechanism consumes a huge number of resources on the controller if the attacker modifies source address.

Mousavi, et al. [2] propose to use entropy for DDoS detection due to its ability to measure randomness, where two essential components are time period and threshold. Although it may improve detection

accuracy in the real network, the proposed techniques only address detection without providing countermeasures.

Dong et al. [3] suggest a statistical tool, called Sequential Probability Ratio Test (SPRT), to improve existing false positive and false negative issues. The evaluation of the DARPA Intrusion Detection Data sets shows its promptness and accuracy. However, the proposed method is evaluated using only mathematical results without simulations, where random variables can be introduced.

Yan et al. [4] propose a "Multislot" strategy to process requests in each time slot so that legitimate users can communicate to each other properly during DDoS attacks. However, if the subscriber and attacker share the same switch port, large flow latency will be introduced because it places the legitimate and malicious request in the same queue.

Dharma et al. [5] propose to use a "flow collector", which sits between the switch and the controller. When the number of invalid packets exceeds the threshold within a certain duration, the flow collector is triggered to further inspect those suspicious packets. However, this introduces a delay to legitimate users. Furthermore, there is no mathematical analysis, simulations or real implementation.

Shoeb and Chithralekha, [6] define a peak time and establish a trust level to defend control and data planes against DDoS attacks. The node's trust level is used to determine the priority of processes on the controller, where the value is set depending on the behaviour during normal time. During peak time, the controller discards requests from particular nodes whose number of requests already exceeds a certain threshold. Even for the normal nodes, the controller replies switch a new rule with a lower timeout value. However, the authors do not indicate how to define a peak time and the threshold of the peak time. Moreover, the proposed technique is not simulated or tested on the real equipment.

Selvi, et al. [7] propose a model to use a Bloom filter to deal with the detection of link flooding attack in the SDN. This model contains two subsystems: (i) collector and (ii) detector. When the link utilisation is abnormal, the collector scans the flow table on the switch and finds the abnormal flows from the statistics of flow entries. The detector monitors the entire network using a controller; therefore, it can sniff packets. The classification of these packets are sent to the Bloom filter to determine whether it is abnormal, because relevant IP features are stored in the Bloom filter. However, there is no definition of abnormal link utilisation, and how to detect this issue on the controller is also unmentioned.

Kokila et al. [8] propose to detect DDoS attacks using a Support Vector Machine (SVM) classifier. The SVM learns the pattern with training samples and predicts the unknown traffic sample to be normal or attack. The 2000 DARPA intrusion detection scenario specific dataset is taken to instruct the SVM. The SVM has a higher accuracy and lower false positive comparing with other methods from the simulation. However, the performance of SVM is deeply based on the training dataset.

Phan et al. [9] propose to use the combination of SVM and SOM to classify DDoS attacks. SVM and SOM are trained by ready-made datasets before the model is used for testing. Each protocol has a dedicated SVM to filter traffic in the control plane. If a specific flow is in the attack region according to the SVM, it is then sent to the classifier. If the flow is in the vogue region, it is sent to SOM to make the decision. The simulations indicate that the combination of SVM and SOM has better performance than deploying them individually.

Lim et al. [10] propose to modify the IP address of the victim to mitigate DDoS attacks. The DDoS Blocking Application (DBA) running on the controller has a secure channel, which is directly connected to the server. Once the server detects DDoS attacks using some metrics, DBA assigns the server a new IP address and asks switches to redirect packets to this new address. After updating the IP address, if a host still sends packets to the previous address and the number exceeds a predefined threshold, the host is blocked as a bot. The simulation results show that DDoS attacks from bots are blocked. However, how to define the metric and threshold to trigger defence and drop action is not mentioned.

Lenka, et al. [11] proposed building scalable cyber- physical-social networking infrastructure using IoT. Wireless sensors are an important component to develop the Internet of Things (IoT) Sensing infrastructure. There are enormous numbers of sensors connected with each other to form a network (well known as wireless sensor networks) to complete the IoT Infrastructure. These deployed wireless sensors are with limited energy and processing capabilities. The IoT infrastructure becomes a key factor

to building cyber-physical-social networking infrastructure, where all these sensing devices transmit data toward the cloud data center. Data routing toward cloud data center using such low power sensor is still a challenging task. In order to prolong the lifetime of the IoT sensing infrastructure and building scalable cyber infrastructure, there is the requirement of sensing optimization and energy efficient data routing.

Sahay, et al. proposes [12] an autonomic DDoS defense framework, called ArOMA, that leverages the programmability and centralized manageability features of Software Defined Networking (SDN) paradigm. Specifically, ArOMA can systematically bridge the gaps between different security functions, ranging from traffic monitoring to anomaly detection to mitigation, while sparing human operators from non-trivial interventions. It also facilitates the collaborations between ISPs and their customers on DDoS mitigation by logically distributing the essential security functions, allowing the ISP to handle DDoS traffic based on the requests of its customers. The experimental results demonstrate that, in the face of DDoS flooding attacks, ArOMA can effectively maintain the performance of video streams at a satisfactory level.

Galeano-Brajones, et al. [13] proposes the expected advent of the Internet of Things (IoT) has triggered a large demand of embedded devices, which envisions the autonomous interaction of sensors and actuators while offering all sort of smart services. However, these IoT devices are limited in computation, storage, and network capacity, which makes them easy to hack and compromise. To achieve secure development of IoT, it is necessary to engineer scalable security solutions optimized for the IoT ecosystem. To this end, Software Defined Networking (SDN) is a promising paradigm that serves as a pillar in the fifth generation of mobile systems (5G) that could help to detect and mitigate Denial of Service (DoS) and Distributed DoS (DDoS) threats. In this work, we propose to experimentally evaluate an entropy-based solution to detect and mitigate DoS and DDoS attacks in IoT scenarios using a stateful SDN data plane. The obtained results demonstrate for the first time the effectiveness of this technique targeting real IoT data traffic.

Akpakwu, et al. proposes [14] the state-of-the-art of the IoT application requirements along with their associated communication technologies are surveyed. In addition, the third-generation partnership project cellular-based low-power wide area solutions to support and enable the new service requirements for Massive to Critical IoT use cases are discussed in detail, including extended coverage global system for mobile communications for the Internet of Things, enhanced machine-type communications, and narrowband-Internet of Things. Furthermore, 5G new radio enhancements for new service requirements and enabling technologies for the IoT are introduced.

Yu, et al. proposes [15] design a platform to efficiently detect and rapidly respond to the DDoS attack in VNs based on software-defined networking (SDN). The proposed platform not only contains the trigger mechanism based on the message of OpenFlow protocol for a response not timely but also involves a flow feature extraction strategy based on the multi-dimensional information. Moreover, we construct an effective global network flow table feature values based on OpenFlow flow table feature and the entropy feature of flow table entry.

EXISTING SYSTEM

The mechanism is based on OpenState, an extension to current OpenFlow that exploits in-switch capabilities and has been proved to be a promising approach for network monitoring since it avoids sending packets to the controller. Based on this research, we have developed a proof-of-concept application at the top of the Ryu SDN controller that detects the DoS and DDoS attacks according to the entropy values. To analyse the effect of this metric under different conditions, we have evaluated the performance of the application in three scenarios. The first one is a general test bed in which the bandwidth and the entropy values are measured during the attack, whereas the second and third focuses on an IoT scenario. The experimental results demonstrated the benefits of using the correlation of the entropy values of different features to detect the attack and also the ability of SDN to mitigate easily just adding entries to the flow table of the switches. In the future, this work will be extended to generalize other types of (D)DoS attacks and also to include different statistical-based metrics that could help in the detection process. The current state of the application results in some issues when the

window size is large, which implies that the switches stop responding to the requests of states by the control plane. That is why an exhaustive study will be carried out to check if the problem corresponds to the architecture of the switch or to the SDN application. Besides, we will experiment with more complex cases outside Mininet and even make use of other proposals in the stateful SDN literature, experimenting also with entropy-based algorithms on real environments with hundreds of hosts generating network traffic to check the effectiveness of the solution in this context. Moreover, intelligent mechanisms will be introduced, such as Machine Learning techniques that allow self-configuring the algorithm parameters, that is, θ and the window size. Finally, we are working on a (D)DoS attack detector as a Network Function that could be deployed in the network as a standalone module. This would give flexibility to the detection process and would be in line with other 5G technology pillars such as NFV.

PROPOSED METHOD

DDoS attacks are a common threat to the network, although the attacker typically does not aim to steal any data. Basically, DDoS attacks aim at consuming system resources until the target is not available to offer its services. DDoS attacks could be divided into three categories: application layer attack, protocol attack and volumetric attack. For volumetric attack, an attacker can deplete the available resources of the victim or bandwidth towards the target. Not only the data plane in the SDN, but also the controller and southbound interface, could suffer from this kind of attack as well; this is because a client host can trigger inquiry from the data plane to control plane. Although there havebeen a lot of discussions about DDoS attacks in the SDN and IoT networks, the large number of IoT gadgets is still a good chance to launch attacks, as well as the communication link between controllers and switches in SDN [5,6]. Additionally, more validations in the real network are required. Moreover, the programmability and centralised control in the SDN give users more options to probe into this threat. In this paper, volumetric attack is implemented
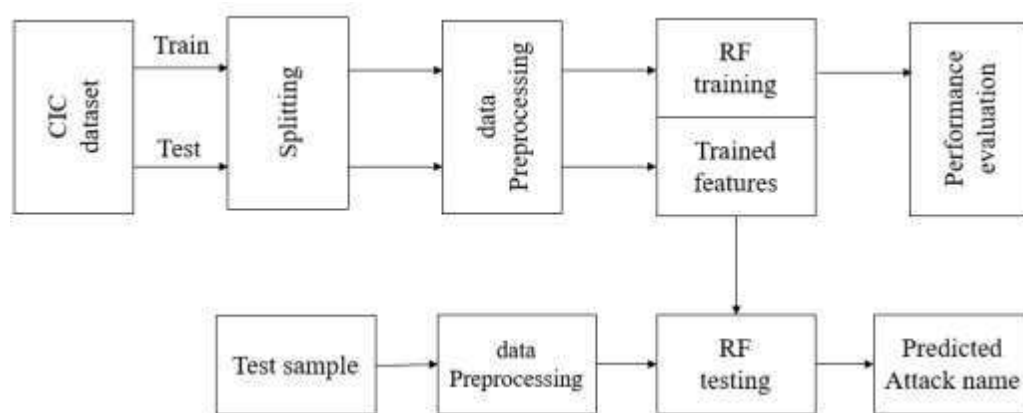

Fig. 1: Proposed block diagram.

Fig. 1 shows the block diagram of proposed method. Initially, CIC dataset is spitted into 80% for training and 20% for testing. Then, dataset pre-processing operation is performed to normalize the entire dataset. Further, random forest classifier is used for prediction of DDoS attack from testsample. The performance evaluation is carried out to show supremacy of proposed method.

CIC dataset

CICDDoS2019 contains benign and the most up-to-date common DDoS attacks, which resembles the true real-world data (PCAPs). It also includes the results of the network traffic analysis using CICFlowMeter-V3 with labelled flows based on the time stamp, source, and destination IPs, source and destination ports, protocols, and attack (CSV files). Generating realistic background traffic was our top priority in building this dataset. We have used our proposed B-Profile system to profile the abstract behaviour of human interactions and generates naturalistic benign background traffic in the proposed testbed. For this dataset, we built the abstract behaviour of 25 users based on the HTTP, HTTPS, FTP, SSH and email protocols.

Pre-processing

Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data. And while doing any operation with data, it is mandatory to clean it and put in a formatted way. So, for this, we use data pre-processing task.

Splitting the Dataset

In machine learning data pre-processing, we divide our dataset into a training set and test set. This is one of the crucial steps of data pre-processing as by doing this, we can enhance the performance of our machine learning model. Suppose if we have given training to our machine learning model by a dataset and we test it by a completely different dataset. Then, it will create difficulties for our model to understand the correlations between the models. If we train our model very well and its training accuracy is also very high, but we provide a new dataset to it, then it will decrease the performance. So we always try to make a machine learning model which performs well with the training set and also with the test dataset. Here, we can define these datasets as:

Random Forest Algorithm

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.
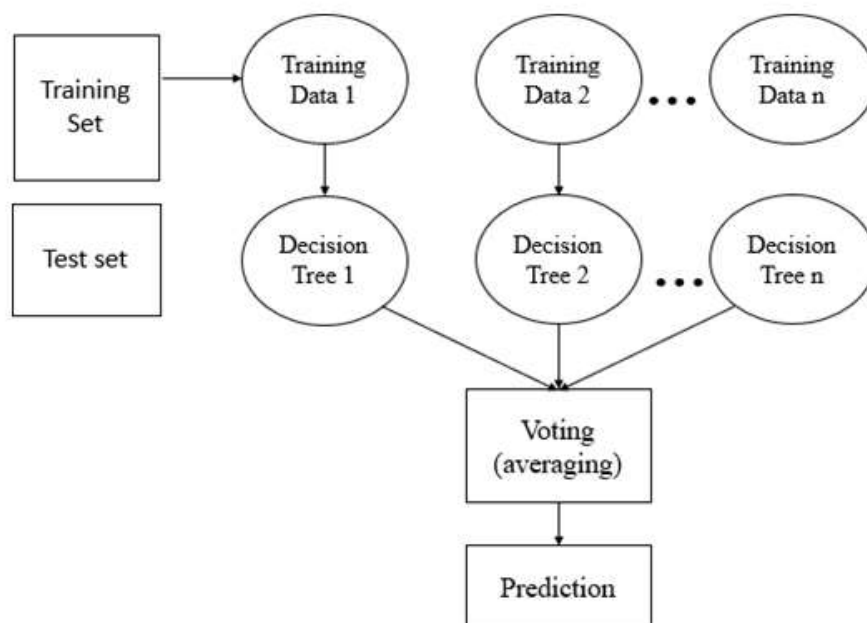

Fig. 2: Random Forest algorithm.

Random Forest algorithm

Step 1: In Random Forest n number of random records are taken from the data set having k number of records.

Step 2: Individual decision trees are constructed for each sample.Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting or Averaging for Classification and regression respectively.

Advantages of Random Forest
- It can be used in classification and regression problems.
- It solves the problem of overfitting as output is based on majority voting or averaging.
- It performs well even if the data contains null/missing values.
- Each decision tree created is independent of the other thus it shows the property of parallelization.
- It is highly stable as the average answers given by a large number of trees are taken.
- It maintains diversity as all the attributes are not considered while making each decision tree though it is not true in all cases.
- It is immune to the curse of dimensionality. Since each tree does not consider all the attributes, feature space is reduced.

Applications of Random Forest: There are mainly four sectors where Random Forest mostly used:
- Banking: Banking sector mostly uses this algorithm for the identification of loan risk.
- Medicine: With the help of this algorithm, disease trends and risks of the disease scan be identified.
- Land Use: We can identify the areas of similar land use by this algorithm.
- Marketing: Marketing trends can be identified using this algorithm.

RESULTS

The dataset was created in an on-premises testbed by the CIC, meaning it's a simulated attack happening in a controlled environment. Benign background traffic was also automated. Feature extraction from the raw .PCAP files was done using a tool developed in-house by the CIC called CICFlowMeter. All thanks and credit are due to the students and staff at UNB/CIC for making this dataset publicly available. The dataset analyses the SECOD algorithm to protect SDN-based IoT network in the real testbed. As IoT has a vast number of applications across different areas, this makes IoT prone to DDoS attacks, which have a huge impact on the SDN-based IoT networks.
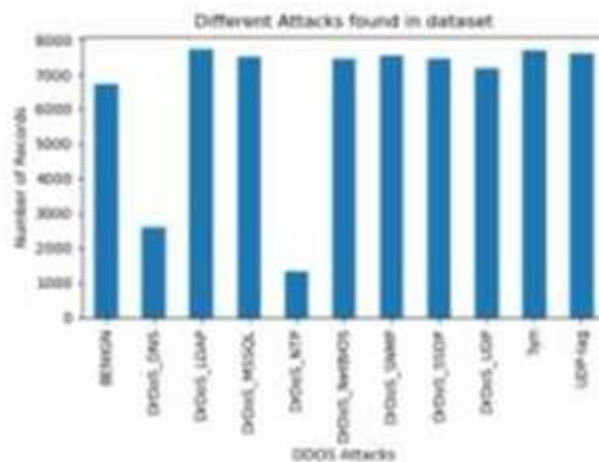

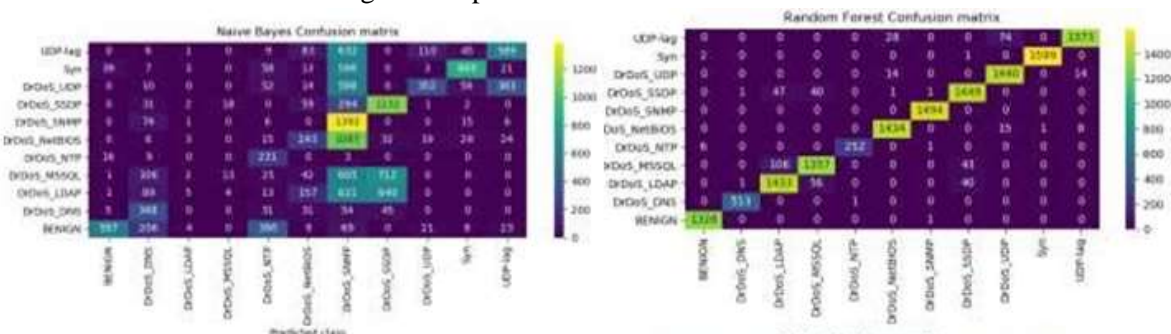Fig. 3: Sample attacks found in CIC dataset.

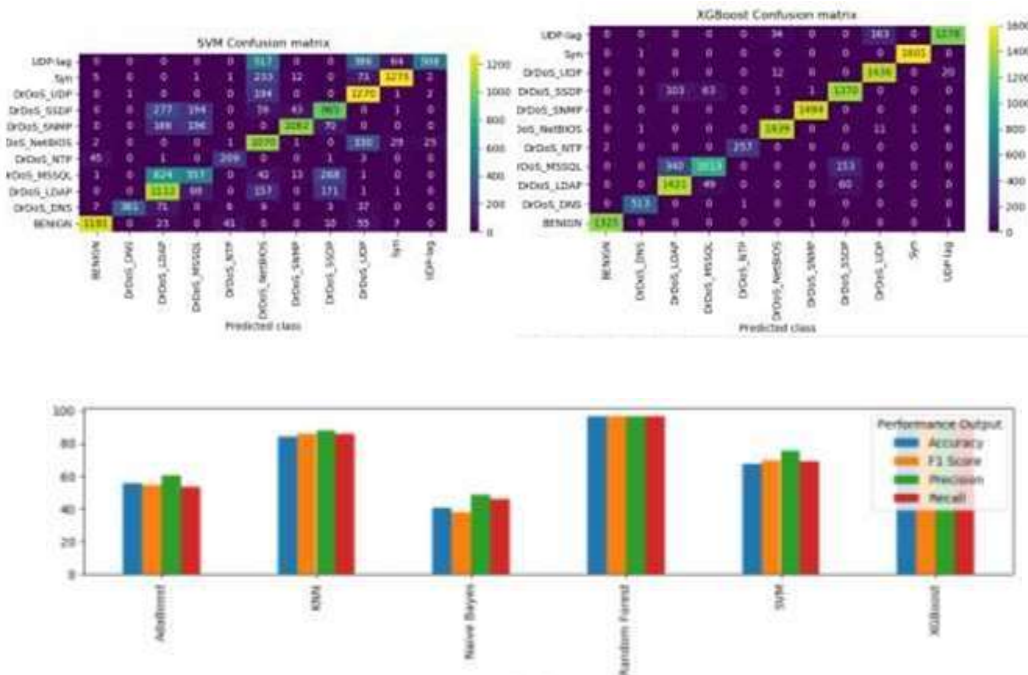Fig. 4: Obtained confusion matrices for various machine learning algorithms.

Fig. 5: Performance comparison of machine learning algorithms for DDoS attack detection.

CONCLUSION

Software-Defined Networking (SDN) and Internet of Things (IoT) are the trends of network evolution. SDN mainly focuses on the upper-level control and management of networks, while IoT aims to bring devices together to enable sharing and monitoring of real-time behaviours through network connectivity. On the other hand, the rapidly growing number of devices challenges the management at the access and backbone layer and raises security concerns of network attacks, such as DDoS. The combination of SDN and IoT leads to a promising approach that could alleviate the management issue. Therefore, this work developed machine learning algorithms for detecting and predicting DDoS attacks in IoT environment with CIC dataset (contains 10 different attacks in IoT and 1 normal class). The simulation results discloses that the higher accuracy is obtained for random forest classifier as compared to other machine learning classifiers.

REFRENCES

1. Dao, N.N.; Park, J.; Park, M.; Cho, S. A feasible method to combat against DDoS attack in SDN network. In Proceedings of the 2015 International Conference on Information Networking (ICOIN), Siem Reap, Cambodia, 12–14 January 2015; pp. 309–311. doi:10.1109/ICOIN.2015.7057902

2. Mousavi, S.M.; St-Hilaire, M. Early detection of DDoS attacks against SDN controllers. In Proceedings of the 2015 International Conference on Computing, Networking and Communications (ICNC), Anaheim, CA, USA, 16–19 February 2015; pp. 77–81. doi:10.1109/ICCNC.2015.7069319.

3. Dong, P.; Du, X.; Zhang, H.; Xu, T. A detection method for a novel DDoS attack against SDN controllers by vast new low-traffic flows. In Proceedings of the 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 1–6. doi:10.1109/ICC.2016.7510992.

4. Yan, Q.; Gong, Q.; Yu, F.R. Effective software-defined networking controller scheduling method to mitigate DDoS attacks. Electron. Lett. 2017, 53, 469–471.

5. Dharma, N.I.G.; Muthohar, M.F.; Prayuda, J.D.A.; Priagung, K.; Choi, D. Time-based DDoS detection and mitigation for SDN controller. In Proceedings of the 2015 17th Asia-Pacific Network Operations and Management Symposium (APNOMS), Busan, Korea, 19–21 August 2015; pp. 550–553. doi:10.1109/APNOMS.2015.7275389.

6. Shoeb, A.; Chithralekha, T. Resource management of switches and Controller during

saturation time to avoid DDoS in SDN. In Proceedings of the 2016 IEEE International Conference on Engineering and Technology (ICETECH), Coimbatore, India, 17–18 March 2016; pp. 152–157.doi:10.1109/ICETECH.2016.7569231.

7. Xiao, P.; Li, Z.; Qi, H.; Qu, W.; Yu, H. An Efficient DDoS Detection with Bloom Filter inSDN. In Proceedings of the 2016 IEEE Trustcom/BigDataSE/ISPA, Tianjin, China, 23–26 August 2016; pp. 1–6. doi:10.1109/TrustCom.2016.0038

8. RT, K.; Selvi, S.T.; Govindarajan, K. DDoS detection and analysis in SDN-based environment using support vector machine classifier. In Proceedings of the 2014 Sixth International Conference on Advanced Computing (ICoAC), Chennai, India, 17–19 December 2014; pp. 205–210.

9. T.Phan.; Bao, N.; Park, M. A Novel Hybrid Flow-Based Handler with DDoS Attacks in Software-Defined Networking. In Proceedings of the IEEE Conferences on UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld, Toulouse, France, 18–21 July 2016; pp. 350–357.

10. Lim, S.; Ha, J.; Kim, H.; Kim, Y.; Yang, S. An SDN-oriented DDoS blocking scheme for botnet-based attacks. In Proceedings of the Conference on Ubiquitous and Future Networks, Shanghai, China, 8–11 July 2014; pp. 63–68. doi:10.1109/ICUFN.2014.6876752.

11. Lenka, R.K.; Rath, A.K.; Tan, Z.; Sharma, S.; Puthal, D.; Simha, N.V.R.; Prasad, M.; Raja, R.; Tripathi, S.S. Building Scalable Cyber-Physical-Social Networking Infrastructure UsingIoT and Low Power Sensors. IEEE Access 2018, 6, 30162–30173.

12. Sahay, R.; Blanc, G.; Zhang, Z.; Debar, H. ArOMA: An SDN based autonomic DDoS mitigation framework. Comput. Secur. 2017, 70, 482–499.

13. Galeano-Brajones, J.; Carmona-Murillo, J.; Valenzuela-Valdés, J.F.; Luna-Valero, F.Detection and Mitigation of DoS and DDoS Attacks in IoT-Based Stateful SDN: An Experimental Approach. Sensors 2020, 20, 816.

14. Akpakwu, G.A.; Silva, B.J.; Hancke, G.P.; Abu-Mahfouz, A.M. A Survey on 5G Networks for the Internet of Things: Communication Technologies and Challenges. IEEE Access 2018, 6,3619–3647.

15. Yu, Y.; Guo, L.; Liu, Y.; Zheng, J.; Zong, Y. An efficient SDN-based DDoS attack detection and rapid response platform in vehicular networks. IEEE Access 2018, 6, 44570–44579.