International Journal on Emerging Technologies **12**(1): 116-121(2021)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

A Sustainable Framework for Preventing IoT Systems from Zero Day DDoS Attacks by Machine Learning

***Mubashir Ali*[1]*, Ayesha Siddique*[2]*, Aamir Hussain*[3]*, Farhad Hassan*[4]*, Amir Ijaz*[5] *and Aneela Mehmood*[6]**
[1]Department of Software Engineering, Lahore Garrison University, Lahore, Pakistan.
[2]Department of Computer Science, University of Agriculture, Faisalabad, Pakistan.
[3]Department of Computer Sciecne, MNS University of Agriculture, Multan, Pakistan.
[4]Department of Computer Science & Engineering, Air University Multan Campus, Pakistan.
[5]Department of Computer Engineering, HITEC University, Taxila, Pakistan.
[6]Department of Computer Computer Science, Lahore Garrison University, Lahore, Pakistan.

(Corresponding author: Aamir Hussain)
(Received 05 October 2020, Revised 26 December 2020, Accepted 29 January 2021)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Internet of things (IoT) is an emerging trend in computer science which brings the idea of everything connected. It leads to future smart systems in every field from education, health, industry to agriculture sector. IoT systems are generating data in exponential size from heterogeneous devices and now IoT security become main issue of research to address. IoT security is very crucial in nature and number of security techniques have been proposed to secure the IoT systems. Still, there is need of new and state of the art techniques to address the newly and day by day arising attacks known as Zero Day attacks. In this research paper, we have proposed a machine learning empowered sustainable framework to address the Zero Day DDoS attacks which is based on Honeypot. The framework in novel as it uses the specific honeypot generated dataset for underlying system. The system is efficiently trained on specific nature honeypot dataset to address the Zero Day DDoS attacks and sustainability feature will enable this framework to cope with future security requirements.

Keywords: IoT Security, Zero Day DDoS Attacks, Sustainable Framework, Machine Learning, Honeypots, IoT Honeypots

I. INTRODUCTION

Internet of things (IoT) is an emerging concept to interconnect different devices to form a network [1]. Along with various benefits, it has a shortcoming that it is vulnerable to DDoS attacks. IoT systems are more vulnerable compared to desktop PCs [2]. There has been an increase in botnet attacks targeting IoT networks [3]. Malware infections in IoT networks are known as botnet. A survey shows that there are more than 6 billion smart devices on the planet andmost of these devices are very vulnerable and are targeted by cyber criminals [4]. These malwares are increasing day by day in fact thousands of them were reported and half of them were detected from the year 2017 alone [5].

Attackers mindset and their methods can be analyzed and observed using a honeypot. The honeypot lures attackers so they can be studied and investigated. Honeypot enabled device acts as a vulnerable gateway for the attacker to get to the main server. This device is capable of collecting port numbers, IP addresses, MAC address, targeted devices, activities, commands and malware executables etc [6]. Honeypots are one of the most used mechanism for investigating the malware made by Fred Cohen in the year 1998and known as 'the deception Toolkit' [7]. It was available to general public for the defense against self-replicating malware known as worms. Honeypots have been classified into different variants that can be used for different applications [8]. They are normally differentiated depending upon the level of interaction with the intruder. The more the interaction gains the more data collected and they are divided into two main categories as low-interaction and high-interaction honeypots. They can also be categorized on the basis of their objectives. For example, research honeypots are used to collect knowledge regarding the threats in a system while production honeypots are used for security improvement to protect assets of the company [9]. In short, zero-day DDoS attacks can be effectively dealt by using honeypots without any compromise to the IoT devices [10].

Honeypots can also be classified as traditional and IoT honeypots [11]. Where one has x68 architecture and the other is heterogeneous because of the variety of IoT devices. The solution proposed uses honeypot primer 2 capture attack attempts done on IoT device. We can input the log files of attacker containing information to a machine learning model so it can be trained to deal with attacks. Power machine learning with available datasets is minimized because they can only provide us with the limited knowledge of data whereas the honeypot log files contain different unknown malware families [12]–[14].

Our framework uses machine learning as it makes it easier to detect the attacks automatically and is able to protect threats to different devices. There are two types of learning classified as unsupervised and supervised [15]. Supervised requires labels that helps and classification in the training phase this helps to matching the labels and features. Whereas unsupervised does not use labels and it matches different features itself during the training. Unsupervised learning is implied in proposed framework as there is no need of any human intervention with the learning process. Some of the well-known used unsupervised learning algorithms are clustering, neural networks and anomaly detection. There are two types of malware detection called classification problem or classification problem. Supervised learning is used in the classification problems as we have known data which is used to predict the problems. Illustration problems deal with unknown malwares and these malwares are clustered; then similarities can be identified using the algorithm. Machine learning algorithms also hold another great advantage, which is the ability to have less false negatives and positives when compared two different anomaly detection methods [16].

II. RELATED WORK

Various honeypot based techniques are proposed in literature to address the security of IoT systems. Comparatively, the security of IoT systems are breakable due to heterogeneous nature of interconnected devices and nodes. Honeypots rule are very crucial in defending the unrecognized nature of DDoS attacks. Kishore *et al* presented the turning concept of IoT to internet of vulnerabilities in context of IoT botnets. With newly emerged botnets, the devices within public IoT systems are greater under the cyber-attacks. They have presented the taxonomy of IoT botnets, along with well-known DDoS attacks, respective mitigation techniques and recommendations to secure the IoT systems [17]. Anirudh *et al* proposed a honeypot based approach to secure the primary server and IoT system from DoS attacks [6]. Quang et al suggested a honeypot based deception mechanism to get the information of attacker to handle the unexpected security breaches. They have implemented this mechanism to theoretical model of game and then Bayesian game with missing information. The deceptive actions of attacker are taken by defender to secure the system [18]. Runyu *et al* presented a honeypot empowered network defender model for game theory. They improved Bayesian model by considering the relative historical payoffs [19]. Wireless sensors networks providing base of communication to IoT systems. Ali et al reviewed the state of the art sinkhole attacks along with their mitigation techniques for wireless sensor networks [20].

Wei et al reviewed the state of the art IoT features in context of security and privacy, discussed new trends and existing solutions with challenges which needs to be addressed [5]. Ronald et al detailed surveyed the honeypot research domain along with their trends and opportunities [21].

Antara *et al* presented a honeynets and honeypots based deception techniques to counter the cyber-attacks [22]. Zobal *et al* described the honeypots in depth with their pros, cons, ethical and legal issues. Then they classified the honeypots in different classes based on characteristics, discussed recent developed honeypots with their impact and challenges [13]. Lik *et al* reviewed the use of honeypot in machine learning algorithms for investigating the malwares [15]. Christos et al critically explored the role of honeypots and honeynets in smart grids. They investigate the different strategies to attract the attackers, gain their information to build a forensic evidence that will be used in court proceedings[12].

III. SUSTAINABLE FRAMEWORK

In our proposed framework, we have considered the following points

- The factor of sustainability is considered to cope with future perspective.
- The framework is based upon honeypots to detect the unknown behavior of attackers which cause to Zero Day DDoS attacks.
- Different machine learning algorithms are implied for learning from attacker and predictions.

The framework is able to detect the type of malwares and categorized it in different known and unknown families. Various type of malwares which cause to DDoS attacks and yet the comprehensive defense against it is incomplete. Various techniques have been proposed in literature to address the DDoS attacks and honeypots are efficient in them. There are three main pillars of this framework as listed above. Honeypots are intentionally used to capture the data of attacker in log files. The log files will store the overall information about the style of attack to nature, severity, impact etc. Further the log files are converted into dataset which is used to train the machine learning model. The model is based on different machine learning algorithms as per scenario. The model will efficiently predict the suspicious activities based on log files which are already recorded in dataset. Whenever a new suspicious activity in recorded by honeypot, it will regenerate the log files which will be used again to update the machine learning model. The sustainability factor will help to cope with future perspective by considering state of the art machine learning models with time.

Fig. 1 shows the architecture of our proposed framework. The system will show the loop holes and vulnerabilities to attract the intruder as show in figure. The attacker tries to enter the system and the corresponding log files will be created. This phenome is processed with IoT-Pot honeypot. Afterwards, a corresponding dataset will be created from log files based on selective features. The dataset will be used to train the machine learning model which will predict the abnormal intervention. Whenever, new type of suspicious attack is found, the framework will update the defense mechanism.

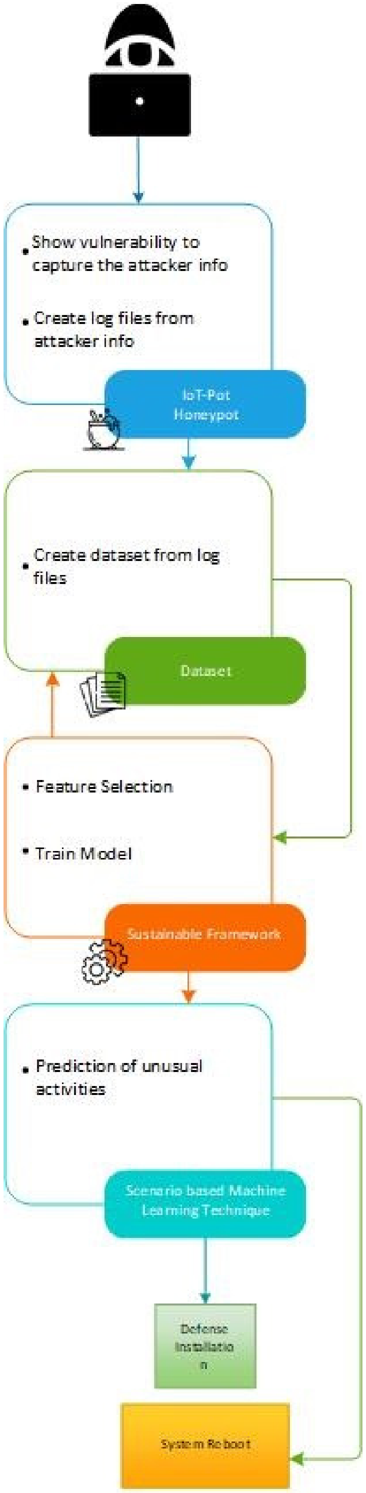
Show vulnerability to capture the attacker info
Create log files from attacker info
IoT-Pot Honeypot
Create dataset from log files
Dataset
Feature Selection
Train Model
Sustainable Framework
Prediction of unusual activities
Scenario based Machine Learning Technique
Defense Installatio n
System Reboot

Fig.1. Architecture of Sustainable Framework.

IV. IMPLEMENTATION

Without implementation, it is not possible to check the feasibility and calculate the efficiency with existing similar solutions, hence it plays a basic role in any approach. The proposed framework by us in the above section has series of steps. For the purpose of keeping our solution adaptable to the changes in technology, we can apply updated methods which are currently in use to handle the current challenges of IoT. Real time anomaly detection with machine learning and honeypots are important research domains for ensuring IoT security.

A. IoT-Pot Honeypot:

In our proposed technique, our first target is to catch the attackers towards the IoT devices for the purpose of deliberately exploiting the vulnerability in them. To make it possible, a system is required which can mimic the behavior of an exploitable IoT architecture and grab the attention of attacker to make his move without any doubts about the intentional attraction. These systems are termed as IoT honeypots. In the above section we have already discussed that honeypots and these can be further classified on the basis of interaction. Their classification includes High-Interaction honeypots, Los-Interaction Honeypots, and Medium-Interaction Honeypots which is a combination of both. Preference should be given to the selection of medium interaction honeypot (MIH), as it is impractical to select a high interaction honeypot for resource limited IoT systems. Honeypot will be implemented virtually by adopting the IoT simulation platforms using IoT wireless communication protocols. Hence we named this as IoT-Pot honeypot. Honeypot can record the intruder information like payload, network traffic, malware samples, the device or system adopted for attack etc.
The honeypots which are developed recently for DDoS detection are listed below:

- IoTPOT [23]: It is responsible for emulating the services of Telnet for several IoT devices and is composed of a fronted low interaction virtual environment termed as IoTBOX which can operate at various architectures of CPU.
- Telnet IoT honeypot [24]: For Implementing the grabbing technique for IoT, telnet server is of great help.
- HoneyThing: A modem/router (having RomPager embedded web Server) and is TR-069 (CPE WAN Management Protocol) is vulnerable specific and emulated by this HoneyThinghoneypot.
- Dionaea [25]: Protocol used by this honeypot is MQTT and it is responsible for simulating the IoT behavior.
- ZigBee Honeypot [26]: For simulating the ZigBee Gateway, this honeypot is used.
- Multi-purpose IoT honeypot [27]: Telnet, SSH, HTTP, and CWMP is the pivot for this honeypot.
- ThingPot [28]: Not only a single application-layer communication protocol but the Complete IoT platform can be simulated by the use of this Honeypot.

The IoT Honeypot which fits best should have the capability to emulate the IoT devices by simulating the entire IoT platform along with all other application layer protocols which are acting as supporting protocols. XMPP (Extensible Messaging and presence Protocol), MQTT (Message Queue Telemetry Transport) by IBM, which is useful for providing basic instant messaging (IM) and presence functionality, CoAP (Constrained Application Protocol) designed for resource-constrained devices, AMQP (Advanced Message Queuing Protocol) which made its appearance from the financial industry, UPnP (Universal Plug and Play) group of network protocols responsible for the discovery of network devices and HTTP REST are the most used Application protocols for IoT communication. An architectural style that has been used on a large scale in Machine-to-Machine (M2M) communications and IoT platforms is termed as REST. To achieve our goal of intriguing number of possible malware attacks, the honeypot which can be used from our above mentioned list is ThingPot.

B. Machine Learning empowered Sustainable Framework

DDoS Detection Process includes machine learning based detection framework which is of great importance in current era. To perform the desired classification, many algorithms of machine learning are available. Here our purpose is not just to classify the malware but to give an appropriate machine learning solution for the accurate classification of malware features which will not generate number of false positives. A solution which can classify the malware and have the accuracy of 0.99 is proposed the field of real-time machine learning based detection in IoT devices[29]. IoT botnet attacks have increased to a great extent in the past recent years, and this solution is targeted to them.

Traditional laptops and smartphones have a different way of communication as compared to IoT devices hence the behavior of IoT traffic is quite different because the IoT devices communicate with endpoints within small range. To observe the behavior of these devices closely and in a precise manner, machine learning process can be used. There are several steps involved in this process initiated from data gathering, feature taking and binary classification at the end. The features which are extracted are mainly IoT- specific network behaviors and have network flow characteristics including protocol, packet length and inter-packet intervals. Different classifiers are compared against each other including random forests, K-nearest neighbors, support vector machines, neural networks and decision trees. The most effective ones among classifiers are random forests, K-nearest neighbors and neural net classifiers. For the purpose of performing feature selection process to attain the higher accuracy in detecting DDoS in IoT traffic with the support of several machine learning algorithms including neural networks, the IoT specific network behavior like the limited number of endpoints, the regular time interval between packets etc. is of great use.

The process of anomaly detection has various phases. First phase is Traffic Capture, the next one is grouping the packets by device and time, then moving on to the fracture extraction phase and the last one is Binary Classification Phase. The phase of traffic capture involves recording the source IP address, source port, destination IP address, destination Port, packet size and timestamp of all sent IP packets from IoT device that is a part of some smart home application. Because of the complexity and risk involved the task of collecting the DDoS traffic is quite tough. It has simulated the three most common variations of DDoS attack including a UDP flood, TCP SYN flood and HTTP GET flood to attain the goal of capturing the new coming variants in the properties of malware. Packets are based on Source IP address which is further divided into timestamps which do not overlap each other and recorded at the initial stage. Grouping is performed on these packets from IoT devices.

According to the behavior of IoT device, stateless and stateful features are being generated in the feature extraction phase. Lightweight features derived from flow independent characteristics of each sent packet are Stateless features and they do not split the traffic stream by IP source when generated. While stateful features are generated by capturing the aggregated flow data in the network traffic with respect to the short time spans. Packet size and Inter-packet interval comes under the category of stateless features whereas bandwidth and IP address cardinality and novelty comes under the category of stateful features. At the end, Binary Classification is performed using various classification algorithms like K-nearest neighbors, random forests, support vector machines and deep neural networks for the purpose of differentiating the DDoS traffic flow from the normal traffic. Deep learning classifiers work on additional data generated from the real-world deployments. Hence to get efficient results, use of deep learning classifiers will be much effective.

To conclude, we can carry out the implementation of proposed solution by using and IoT honeypot inspired by the ThingPotwhich has the capability of capturing several botnet binaries by imitating various IoT communication protocols along with Complete IoT platform behaviors and is an IoT-Pot honeypot. The virtual box helps to deploy it over every IoT device in a network for the purpose of keeping it separate from the original IoT platform. Considering the constraints of IoT, classifiers cannot be implemented on each device,the implementation is possible on router level. Traffic received by a specific IoT device is limited and inadequate to perform training over a machine learning model. To produce a sufficient amount of IoT traffic, IoT simulators are useful. IoT simulators can be of great help for generating an IoT environment for testing any IoT based application and in case if any storage facility is needed then using cloud it can be added. IoT simulators are not needed if we are using preferred honeypot, because in that case our honeypot will be responsible for all those functionalities. Bash scripts on Linux can be used to transform the log files into the desired format which is required as an input for machine learning model. To make the implementation of machine learning task possible, certain machine learning tools in a virtualized environment can be used like Microsoft Azure and MATLAB.

V CONCLUSION AND FUTURE WORK

With lot of benefits, IoT brings number of security challenges for smart and connected systems. The ratio of cyber-attacks has been increased with the deployment of embedded systems. The security of IoT systems becomes the main concern and continuous efforts are required to ensure it. IoT botnets are emerging threats which introduces new categories of attacks day by day. Various efforts are made to secure the IoT systems against botnets and DDoS attacks. In this research paper, we have proposed a honeypot based approach to tackle and mitigate the emerging Zero Day DDoS attacks in IoT systems. We have proposed a sustainable security framework empowered by machine learning techniques. Honeypot attracts the attacker by showing some vulnerability and take the attacker information in log files. After that, a dataset is created from log files and machine learning model is trained on created dataset which will predict the current and future attacks. Furthermore, the honeypots are stay active to regain the data of updated attacks and prediction model will be updated accordingly to fight against attackers.

CONFLICT OF INTEREST

With On behalf of all authors, the corresponding author states that there is no conflict of interest.

REFERENCES

[1]. Čolaković, A., & Hadžialić, M. (2018). Internet of Things (IoT): A review of enabling technologies, challenges, and open research issues. *Computer Networks*, *144*, 17-39.

[2]. Porras, J., Khakurel, J., Knutas, A. and Pänkäläinen, J. (2018). "Security Challenges and Solutions in theInternet of Things," Nord. Balt. J. Inf. Commun. Technol., vol. 2018, no. 1, pp. 177–206, Sep. 2018, doi: 10.13052/nbjict1902-097X.2018.010.

[3]. Silva, S. S., Silva, R. M., Pinto, R. C., & Salles, R. M. (2013). Botnets: A survey. *Computer Networks*, *57*(2), 378-403.

[4]. Kolias, C., Kambourakis, G., Stavrou, A., & Voas, J. (2017). DDoS in the IoT: Mirai and other botnets. *Computer*, *50*(7), 80-84.

[5]. Zhou, W., Jia, Y., Peng, A., Zhang, Y., & Liu, P. (2018). The effect of iot new features on security and privacy: New threats, existing solutions, and challenges yet to be solved. *IEEE Internet of Things Journal*, *6*(2), 1606-1616. doi: 10.1109/JIOT.2018.2847733.

[6]. Anirudh, M., Thileeban, S. A., & Nallathambi, D. J. (2017, January). Use of honeypots for mitigating DoS attacks targeted on IoT networks. In *2017 International conference on computer, communication and signal processing (ICCCSP)* (pp. 1-4). IEEE. doi: 10.1109/ICCCSP.2017.7944057.

[7]. "Fred Cohen's The Deception ToolKit." https://cypherpunks.venona.com/date/1998/03/msg00116.html

[8]. Mokube, I., & Adams, M. (2007, March). Honeypots: concepts, approaches, and challenges. In *Proceedings of the 45th annual southeast regional conference* (pp. 321-326). doi: 10.1145/1233341.1233399.

[9]. Razali, M. F., Razali, M. N., Mansor, F. Z., Muruti, G., & Jamil, N. (2018, November). IoT honeypot: A review from researcher's perspective. In *2018 IEEE Conference on Application, Information and Network Security (AINS)* (pp. 93-98). IEEE., doi: 10.1109/ains.2018.8631494.

[10]. J. Krupp, M. Backes, and C. Rossow, "Identifying the scan and attack infrastructures behind amplification DDoS attacks," in Proceedings of the ACM Conference on Computer and Communications Security, Oct. 2016, vol. 24-28-October-2016, pp. 1426–1437, doi: 10.1145/2976749.2978293.

[11]. M. Wang, J. Santillan, and F. Kuipers, "ThingPot: An interactive internet-of-things honeypot," arXiv. arXiv, Jul. 11, 2018.

[12]. C. Dalamagkas et al., "A Survey on honeypots, honeynets and their applications on smart grid," in Proceedings of the 2019 IEEE Conference on Network Softwarization: Unleashing the Power of Network Softwarization, NetSoft 2019, Jun. 2019, pp. 93–100, doi: 10.1109/NETSOFT.2019.8806693.

[13]. L. Zobal, D. Kolář, and R. Fujdiak, "Current State of Honeypots and Deception Strategies in Cybersecurity," in International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, Oct. 2019, vol. 2019-October, doi: 10.1109/ICUMT48472.2019.8970921.

[14]. M. L. Bringer, C. A. Chelmecki, and H. Fujinoki, "A Survey: Recent Advances and Future Trends in Honeypot Research," Int. J. Comput. Netw. Inf. Secur., vol. 4, no. 10, pp. 63–75, Sep. 2012, doi: 10.5815/ijcnis.2012.10.07.

[15]. I. M. M. Matin and B. Rahardjo, "The Use of Honeypot in Machine Learning Based on Malware Detection: A Review," Oct. 2020, doi: 10.1109/CITSM50537.2020.9268794.

[16]. E. Bertino and N. Islam, "Botnets and Internet of Things Security," Computer (Long. Beach. Calif)., vol. 50, no. 2, pp. 76–79, Feb. 2017, doi: 10.1109/MC.2017.62.

[17]. K. Angrishi, "Turning Internet of Things(IoT) into Internet of Vulnerabilities (IoV) : IoT Botnets," arXiv, Feb. 2017, [Online]. Available: http://arxiv.org/abs/1702.03681.

[18]. Q. D. La, T. Q. S. Quek, J. Lee, S. Jin, and H. Zhu, "Deceptive Attack and Defense Game in Honeypot-Enabled Networks for the Internet of Things," IEEE Internet Things J., vol. 3, no. 6, pp. 1025–1035, Dec. 2016, doi: 10.1109/JIOT.2016.2547994.

[19]. R. Guan, L. Li, T. Wang, Y. Qin, W. Xiong, and Q. Liu, "A bayesian improved defense model for deceptive attack in honeypot-enabled networks," in Proceedings - 21st IEEE International Conference on High Performance Computing and Communications, 17th IEEE International Conference on Smart City and 5th IEEE International Conference on Data Science and Systems, HPCC/SmartCity/DSS 2019, Aug. 2019, pp. 208–214, doi: 10.1109/HPCC/SmartCity/DSS.2019.00043.

[20]. M. Ali, M. Nadeem, A. Siddique, S. Ahmad, and A. Ijaz, "Addressing Sinkhole Attacks In Wireless Sensor Networks - A Review," Int. J. Sci. Technol. Res., vol. 9, no. 08, 2020.

[21]. R. M. Campbell, K. Padayachee, and T. Masombuka, "A survey of honeypot research: Trends and opportunities," in 2015 10th International Conference for Internet Technology and Secured

Transactions, ICITST 2015, Feb. 2016, pp. 208–212, doi: 10.1109/ICITST.2015.7412090.

[22]. A. D. Oza, G. N. Kumar, and M. Khorajiya, “Survey of Snaring Cyber Attacks on IoT Devices with Honeypots and Honeynets,” Nov. 2018, doi: 10.1109/I2CT.2018.8529510.

[23]. Y. M. P. Pa, S. Suzuki, K. Yoshioka, T. Matsumoto, T. Kasama, and C. Rossow, “IoTPOT: A novel honeypot for revealing current IoT threats,” J. Inf. Process., vol. 24, no. 3, pp. 522–533, 2016, doi: 10.2197/ipsjjip.24.522.

[24]. “GitHub - Phype/telnet-iot-honeypot: Python telnet honeypot for catching botnet binaries.” https://github.com/Phype/telnet-iot-honeypot.

[25]. “GitHub - DinoTools/dionaea: Home of the dionaea honeypot.” https://github.com/DinoTools/dionaea.

[26]. “A ZigBee honeypot to assess IoT cyberattack behaviour - IEEE Conference Publication.” https://ieeexplore.ieee.org/document/7983603

[27]. V. A. Memos and K. E. Psannis, “AI-Powered Honeypots for Enhanced IoT Botnet Detection,” in 2020 3rd World Symposium on Communication Engineering, WSCE 2020, Oct. 2020, pp. 64–68, doi: 10.1109/WSCE51339.2020.9275581.

[28]. M. Wang, J. Santillan, and F. Kuipers, “ThingPot: an interactive Internet-of-Things honeypot,” arXiv, Jul. 2018, [Online]. Available: http://arxiv.org/abs/1807.04114.

[29]. R. Doshi, N. Apthorpe, and N. Feamster, “Machine learning DDoS detection for consumer internet of things devices,” in Proceedings - 2018 IEEE Symposium on Security and Privacy Workshops, SPW 2018, Aug. 2018, pp. 29–35, doi: 10.1109/SPW.2018.00013.

How to cite this article: Mubashir Ali, Ayesha Siddique, Aamir Hussain, Farhad Hassan, Amir Ijaz and Aneela Mehmood (2021). A Sustainable Framework for Preventing IoT Systems from Zero Day DDoS Attacks by Machine Learning, **12**(1): 116–121.