DoS and DDoS attacks and Identification Techniques

Sai Prudhvi Valicherla,

Computer Science Department,

Georgia State University,

svalicherla1@student.gsu.edu.

***Abstract* - Denial of Service (DoS) is categorized as severe cyberattacks and considered a nightmare to the network security teams working for an organization. DoS attacks can take down even the most stable and well-established organizations by generating colossal traffic and flooding the system or network resources with unnecessary or falsified requests, thereby causing interruptions to the services offered and making the network resources unavailable to the customers. The attack can be intensified using the Distributed Denial of service (DDoS), where the intruder can remain undetected and launches the attack from the multiple compromised connected devices or zombie machines called botnets. The person in control of botnets is sometimes referred to as the botmaster. In some scenarios backtracking the botmaster is a tiresome or sometimes an impossible task. In the world of information, the internet is an essential requirement. The majority of the automated services and mission-critical functions are stored in the network, and considering the disruptive nature of the DoS DDoS attacks, they should be adequately addressed, and security must be considered a priority. This paper will discuss various DoS and DDoS attacks and state-of-art identification techniques against these attacks**.

Keywords-DoS, DDoS, falsified request, botnets, botmaster.

1. Introduction

We are in a world where everything and anything is available on the internet, and the number of internet users globally is increasing. Cybercrimes have been rising at an alarming rate than ever expected. Thousands of cyber-attacks are targeted to acquire user-sensitive information can be witnessed daily across the world. In this world driven by money, most, but not all, cybercrime is committed by cybercriminals or hackers who want to make money. Cybercriminals may also target users' private information and corporate data for theft and resale in the Black market. With the emergence of technology and Networking, network security issues have become more and more critical.

Denial-of-service (DoS) attack have always occupied the top spot in network security issues. In fact, every large-scale DDoS attack in the history of the internet can cause significant turbulence. DoS attacks are simple yet effective. In early 2000, there was an ample number of DDoS attacks witnessed over a week. Indeed, not the first DDoS attack, but that highly public and successful series of attacks transformed DDoS attacks from a minor nuisance to potent business disruptors. Since then, DDoS attacks have become an all too frequent menace, as they are commonly used for revenge, conduct extortion, as a means of online activism, and even wage cyberwar. DDoS attacks were limited to computers and internet-connected machines in the past, usually with a reasonable level of protection. Nevertheless, now, hackers consider the Internet of Things as a toyshop in the hands of a child: millions of devices, all too often left unprotected and unmonitored. The scale on which these attacks are possible is rising tremendously with the advancement of the Internet of Things.

The attackers' goal is to develop innovative ways to gain unauthorized and unprivileged access into networks and compromise the confidentiality, integrity, and availability [1] of data, building their targets from single individuals to small or medium-sized companies to business giants. Every year seems to bring an enormous number of attacks, but a considerable number of attacks defeating the security of huge companies, thus affecting information security, business continuity, and customers' trust.

2. Attacks and identifications

i. *"Review on the paper "Detection of Smurf Flooding Attacks Using Kullback-Leibler-Based Scheme"* [2]

Smurf attack

Metaphorically, a smurf attack can be described as a prank in which the prankster conceals one's identity as the company's CEO and calls the office manager. The cloaked CEO asks the manager to immediately set up a one-one meeting request with all the employees and executive manager with an update on the task at hand. Then the prankster sends the targeted victim call back number to the manager, hoping that the victim will receive unnecessary calls from the employees in the office.

The primary goal of the smurf attack is to break down or make the target server potentially sloppy to the point that either it is inaccessible or vulnerable. Smurf attack exploits the characteristics of the broadband network to boost the damage potential and act as an amplification attack is common in IPV4 and IPV6 networks. Network administrators use Internet Control Message Protocol (ICMP) to diagnose the networked hardware devices such as computers, printers, or routers. The control sends pings as an *echo request*, which requests a host to respond with the body of the message as *an echo reply*. In ICMPv4, *echo request* and *echo reply* messages are identified by types 8 and 0, respectively, and in ICMPv6, their types are 128 and 129, respectively. A ping is commonly used to see if a device is operational and track the amount of time it takes to go round trip from the source device to the target and back to the source. Unfortunately, the ICMP protocol does not have a handshake functionality, and therefore the devices receiving the requests are unable to verify the legitimacy of the request. The smurf attack utilizes this functionality and broadcasts many ICMP packet requests using the IP broadcast address with the intended victim's spoofed source IP. The hosts on the victim's network respond to the ICMP requests, creating a significant amount of traffic on the victim's network, resulting in bandwidth consumption and ultimately causing the victim's server to crash. Furthermore, there is still an additional step that makes the attack deadlier, which is to target the multiple hosts as a broadcast request, and with the default settings on the routers, all the hosts on that network will reply, allowing a single host to multiply itself to the number of hosts in that network. Considering a 200-fold multiplication on a 256 KB Digital Subscriber Line (DSL) line – transmit the high-speed digital data over telephone lines, a single host can saturate 10MB ethernet feed.

Lincoln Laboratory MIT and collaborated with Defence Advanced Research Projects Agency (DARPA) to develop and design the dataset and is called the DARPA99 dataset. DARPA 99 dataset is used as an intrusion detection benchmark for validating methods and mechanisms in detecting cyberattacks. DARPA99 dataset consists of 2 parts, off-line evaluation, and real-time evaluation, and simulates the real-world network of a military Air Force base. The off-line evaluation consists of 3 weeks of training data. The first and the third week's data do not contain any attacks, whereas the 2nd week's data consists of selected subsets of attacks during the period 1998 - 1999 and additional new attacks. The training data consists of 8GB of the following files: Outside and inside sniffing data is in Tcpdump format, BSM and NT audit data, a Long list of directory trees as well as dumps of selected directories. Fourth- and Fifth-weeks data is used as testing data which contains 201 instances of 56 types of attacks.

Kullback-Leibler Divergence (KLD) is a well-known metric to measure the distance between two probability distributions quantitatively.
When p1(x) and p2(x) are similar, KLD is close to zero due to measurement noise and uncertainty. Otherwise, large values of KLD reflect a notable deviation between the two distributions. In the absence of anomalies, KLD becomes closer to zero, whereas a larger KLD value is obtained under the presence of anomalies that lead to a potential attack. Shewhart chart, also known as the three-sigma rule, is used as a decision threshold.

$$KLD(P1//P2) = \int_{Rdx}^{*} p1(x) * \log\left(\frac{p1(x)}{p2(x)}\right)dx$$

$$CLKLD = \mu KLD\ 0 + 3\sigma KLD$$

The *echo reply* messages are extracted from the training data, which contains the ICMP traffic generated by the network devices when running diagnostic operations. The extracted testing data consists of anomalous ICMP traffic data and simulated high and low-intensity ICMP smurf attacks. The threshold is calculated using the training data, and if the decision statists in the new sample cross the threshold, it is recognized as the smurf attack.

Result

Two minutes of low-intensity ICMP smurf attack is injected every three hours into the attack-free DARPA99 training dataset; thus, the victim received 20 *echo reply* pings in the observed time.

Advantages and disadvantages

The proposed method KLD-Shewhart showed better performance compared to the Shewhart chart. However, the response time is extremely large. As the value p2(x) increases and p1(x) approaches 0, the value of KLD shrinks to 0. Maximum Likely hood estimation (MLE) gives the cases where the model produces samples that do not locate on the data distribution at a low cost. Consider the situation that p1(x)→0 means that x occurs infrequently; its likelihood could be pretty high as a

result of the lack of training for this group and therefore its high probabilities in the distribution; such uncommon samples that do not fit the data distribution may be produced when testing or validating. So, these failures require looking for different solutions.

ii. "*Anomaly Based Distributed Denial of Service Attack Detection and Prevention with Machine Learning*" [3]

Experimental setup for identifying the TCP, UDP, IMCP malicious packets

The incoming traffic load is inspected periodically over the network, and a threshold is determined based on the observations. Once the traffic over the network crosses the threshold and the anomalies are detected for a specific time interval, the data is collected accordingly. The data scaling techniques are applied to the collected dataset, and the clustering models are applied to detect the type of attack.

Attributes

Total count of incoming IMCP packets over a time interval (typically behavior - 500 IMCP packets per 10 sec, abnormal behavior - 5000 packets per 10 sec), Total size of the IMCP packets in the same interval, Length of the time frame during the abnormal behavior, count of distinct IMCP types. Addresses of the packets, type of the packet, the packet's Arrival time, Packet size. For measuring the network traffic, the attributes are

- the number of packets in an internal,
- the Total size of the packets in the interval,
- Bandwidth for that interval,
- Packets per second for that interval.

Dataset-1

- The time of the entry of a particular packet in epoch format.
- The number of TCP packets over the time interval.
- The number of UDP packets over the time interval.
- The number of IMCP packets over the time interval.
- Count of other types of packets over the time interval.

Dataset -2

- Time of entry in epoch format
- Bytes/sec: Rate the total size of the packets over the time interval.
- Total packet size.
- Interval Packets/sec - Number of packets captured over the time interval.
- Total packets - Total number of packets captured from the beginning.

Clustering

DBSCAN algorithm is very good at handling the dataset with numerous records and a vast number of attributes. The algorithm takes three main parameters,
Epsilon(eps) – the maximum distance between two points in the same cluster, Minimum sample points(min_pts) – minimum number of points in the cluster, and
Metric – distance function used to calculate the separation between the 2 points.
After performing the recurring tests, the values chosen for the parameters are
dataset - 1 eps = 0.08, min_pts= 15% of dataset and metric = correlation.

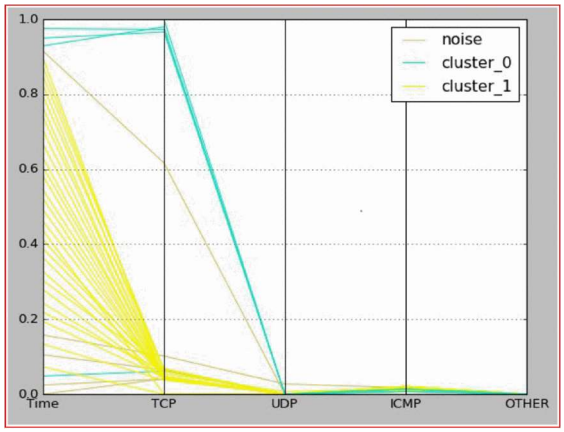

Figure 1. Multi-dimensional graph for dataset 1.

It is evident from the figure 1 that regular traffic is recorded between the time frame [0-0.9] as is classified as cluster_1. The attack is introduced with the TCP packets during the time frame [0.9 – 1.0], and there is a sharp increase in the attribute TCP, and noise is observed.

dataset - 2 eps = 0.03, min_pts= 15% of dataset and metric = correlation.

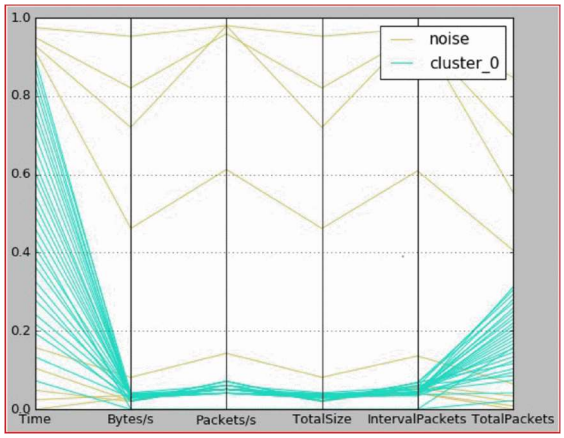

Figure 2. Multi-dimensional graph for dataset 2

It is evident from the figure 2 that regular traffic is recorded between the time frame [0-0.9] as is classified as cluster 0. The attack is introduced during the time frame [0.9 – 1.0], and there is a sharp increase in the attribute values, and noise is observed.

Advantages and disadvantages

DBScan is a density-based clustering algorithm that does an excellent job in clustering out the data points from the outliers and noises when the input features are the number of features is less DBScan does a perfect job in classifying the noise and regular data traffic. However, there are few drawbacks the algorithm fails when classifying the varying density clusters. If the number of input features is increasing, the algorithm fails to classify the points.

iii. "*Machine Learning DDoS Detection for Consumer Internet of Things Devices*" [4]

Mirai Botnet

Botnets such as Mirai took advantage of the insecure consumer IoT devices and scan over the internet for open teleports to perform distributed denial of service (DDoS) attacks on critical Internet infrastructure. With the development of new techniques, home gateway routers could automatically detect the DDoS attacks originating from the local IoT device using the combination of machine learning algorithms and traffic data that are flow-based and protocol independent. The designed model operates on network middleboxes and connected computers that could be part of an ongoing botnet to detect anomalous traffic.

The network traffic in the IoT devices is different from the other connected devices. IoT devices communicate with a small set of finite endpoints and have repetitive network traffic patterns. For instance, a smart door could have limited states:" lock," "unlock," "emergency calls," and" connecting to Wi-Fi," with peculiar network traffic patterns.

The anomaly detection's objective is to identify the patterns and discern the attack traffic from the regular traffic. The limitations of the traditional threshold-based identifications and anomaly detection in IoT devices lead to implementing the machine learning algorithm in these domains. The standard time duration for the DoS attack to avoid detection [5] is roughly 1.5 minutes. Anyone of the connected devices in the network can launch a series of DoS attacks and vary the attack duration. In the given time interval, they are capable of sending the network and attack traffic.

Data collection:

As shown in figure -3, Raspberry Pi V3 is configured to be a Wi-Fi access point, and the following IoT devices are connected to the middlebox, YI home camera, Belkin WeMo Smart Switch, Withing's Blood Pressure Monitor is connected through Bluetooth to the android device. Non-DoS traffic is collected for 10 min from these devices. Kali Linux is used to stimulate the following DoS attacks, TCP SYN flood, a UDP flood, and an HTTP GET flood. The experiment was conducted at every 10 min time interval, and the attacks occurred uniformly between 90 to 110 sec. The final dataset consists of 491,855 packets, of which 459,565 are malicious packets and 32,290 benign packets.

Attributes

Stateful features observe the dynamic nature of the network traffic and require aggregating functions of multiple packets in a time window.

Bandwidth:

The dynamic average bandwidth with 10-sec windows is calculated by splitting and grouping the network traffic under the source device.

IP destinations:

The destination IP address of the IoT devices rarely changes.

Classified as two types

IoT devises as a victim - count of distinct destination IP addresses within a 10-second window. If there are comparatively more endpoints, then there is attack traffic.

IoT devises as an attacker- Count the change in the number of distinct destination IP addresses between time windows. If there are new endpoints, it might suggest that the device is conducting an attack.

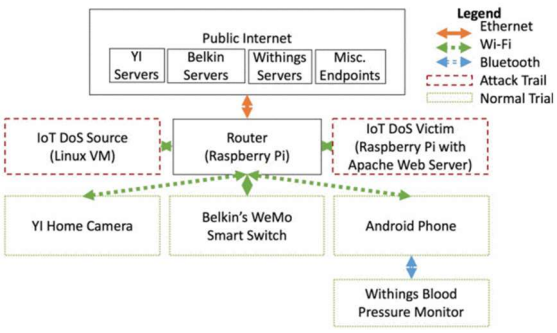

Figure -3 Experimental setup

Stateless features are derived without splitting the incoming traffic by IP source.

Packet Size:

The packet size will range from 100 to 1200 bytes during the regular traffic; on the contrary, the packet size is under 100 bytes during the attack. The TCP SYN Flood will try to establish as many connections as possible during the attack and thus try to lower the packet's size.

Inter Packet Interval:

There will be a significant time difference between the packets with regular traffic; on the other hand, the time difference between the packets is close to 0 when there is a DoS attack.

$$(\Delta T), \frac{d\Delta T}{dt}, \frac{d2\Delta T}{dt2}$$

Protocol:

To capture different protocols

IS TCP, IS UDP, and IS HTTP, IS OTHER

The following Machine learning models have been used

- K-nearest neighbor algorithm (KN)
- Support vector machine with the linear kernel (LSVM)
- Decision tree using Gini impurity scores (DT)
- Random Forest (RF)
- Neural Network (NN)

Table-1 shows the Precession, Recall, accuracy, and F1 scores while using the different machine learning algorithms.

The Linear SVM has a minor performance, and thereby we can conclude that the data is linearly in-separable. On the other hand, KN, DT, RF, NN performed well with similar accuracies, suggesting consistency with the data and the algorithms.

	KN	LSVM	DT	RF	NN
Precision (normal)	0.998	0.992	0.996	0.999	0.983
Precision (attack)	0.999	0.991	0.999	0.999	0.999
Recall (normal)	0.993	0.87	0.993	0.998	0.989
Recall (attack)	0.999	0.999	0.999	0.999	0.998
F1 (normal)	0.999	0.999	0.999	0.999	0.998
F1 (attack)	0.999	0.995	0.996	0.999	0.999
Accuracy	0.999	0.991	0.999	0.999	0.999

Table -1 comparison metrics for various machine learning algorithms

As shown in table -2, The Gini impurity score is used to rank the features, and it is clear that the stateless features outperformed the stateful features.

Feature	Gini Score	Feature's type	Cumulative scores
Packet Size	0.51	Stateless	3.207
is_HTTP	0.177		
ΔT	0.7		
is_TCP	0.68		
is_OTHER	0.43		
is_UDP	0.41		
dΔT /dt	0.18		
d2ΔT /dt2	0.12		
Bandwidth	0.06	Stateful	0.13
# Destinations	0.04		
Δ # Destinations	0.03		

Table -2 Features and their scores using the Gini impurity

Advantages and disadvantages

Acknowledging the limited number of features, the packet-based machine learning models effectively distinguish the regular and DoS traffic in the IoT devices. The accuracy metrics of all the models show us that the algorithms can handle both the linear and non-linear data and effectively distinguish regular packets from malicious packets. Further, research is needed under this domain.

iv. *"Botnets and Internet of Things Security"* [6]

Bots are built for a comprehensive range of activities, including distributed cryptocurrency mining, denial-of-service (DDoS) attacks, keylogging, and crypto-currency mining. They can scan the entire network for vulnerabilities and weak passwords, install a small program on the compromised system, and await instructions from the botmaster. Botnet's architectures have emerged from the centralized to peer to peer(P2P) architecture. So, there is no need for the botmaster to depend on the central server. Few examples of P2P botnets are Sality, Kelihos. Domain fluxing is commonly used for communication in the botnet network; each botnet uses the domain-generation algorithm to compute a list of pseudorandom domain names and tries to establish the connection with the domains in a specific order until the connection is successful. On the other hand, the botmaster needs to register on few domains in the server as command-and-control servers.

IoT devices are at risk for the following reasons:

IoT devices are heterogeneous, self-sufficient to control other IoT devices, physically unguarded, and maintained by different organizations. The parameters for the IoT devices keep on changing with the user preferences, and the updates and installations of security patches are automatic and do not require permission from the user.

The Mirai malware attacked Brian Krebs' blog with the capacity of 620 gigabits per second, and the French web host OVH was attacked by another Mirai attack and is recorded as the most significant DDoS attack. Following this attack, US Computer Emergency Readiness Team (US-CERT) released the precautions, which includes:

- Default passwords should be changed to strong passwords
- IoT devices should be updated with the latest security patches
- Universal Plug and Play (UPnP) on the routers should be disabled.
- IP ports 2323/TCP and 23/TCP need to be monitored for unauthorized control over the IoT devices and port 48101 for anomalous traffic from other affected IoT devices.

Typically, defenses against the botnets can be broadly classified as prevention, monitoring, and responses.

Prevention:

The universal fact, prevention is always better than cure should be strictly administered against bot infections by practicing the below activities periodically: Installing the latest security patches along with antivirus software augmented with intrusion prevention and firewalls with content filtering. Lack of knowledge and awareness among the users are crucial factors and should be considered seriously to protect the users in the first place.

Monitoring:

The administrators from an organization or a network perform Network behavior analysis (NBA), which actively monitors the data transmission from the router and other sources and tag if there is any difference observed from the recommended values. Users should also actively participate and keep an eye on the signs such as delayed response times, excessive exhaustion of CPU cycles, random crashes, and more extended start-up and shutdown times.

Responses:

If signs of a possible DDoS attack or a compromised computer are discovered, intelligent and swift action must limit the damage and prevent the malware from spreading. Simple actions like disconnecting a suspicious computer from the network to monitoring, examining, and disabling botnets are all possible responses. NBA tools may perform certain mitigating activities, such as using the Border Gateway Protocol or other routing protocols to divert remotely malicious traffic to other hosts. However, specialist security firms or law enforcement agencies are required to incapacitate a botnet in the network.

v. *"Use of Honeypots for Mitigating DoS Attacks targeted on IoT Networks"* [7]

This paper presents a solution to the DoS attacks on IoT systems by implementing honeypots to redirect the attack and obtaining information about the intruder.
A honeypot is a device used to deceive possible hackers into gaining unwanted access to information systems by simulating the primary system. It is used to investigate hackers' activity who leave traces, as well as to improve server security in order to deter further attacks. It consists of programs and data that function as a decoy and simulate primary application behavior. It is typically used in a firewall so that it can be securely monitored. Honeypots are categorized as production or research honeypots based on their design and deployment. Production honeypots are installed in production networks to act as a decoy and as part of an intrusion detection scheme, while research honeypots are used to do a thorough investigation into the attacker and safety protection steps.

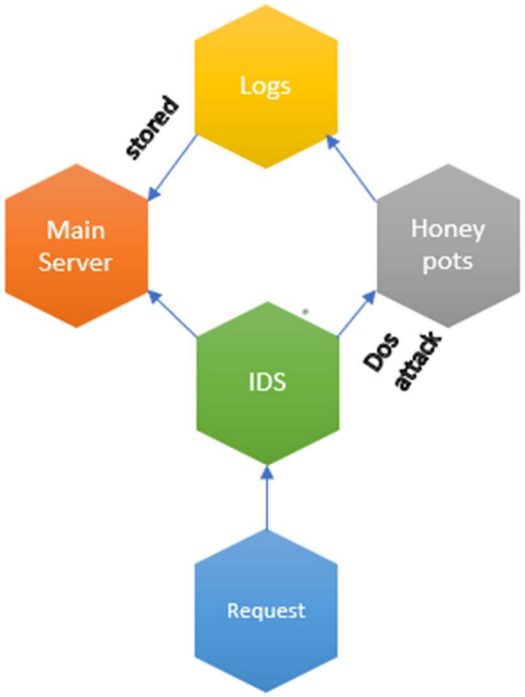

Figure – 4. Scenario 1

There exist two scenarios in the proposed model,
Scenario – 1

If the request from the client is receiving for the first time, then the request is passed through the Intrusion Detection System (IDS). If IDS is

suspicious about the request, the request is moved to the honey pots, and the underlying process completes. The information such as IP address, MAC address, and other details are collected and stored in the database.

Scenario – 2

The incoming request is cross-checked with the records in the database- contains the history of details about the spam. If there is a match, then the requests' source address is permanently blocked from the server; otherwise, the request is passed into the and is ready to be addressed.

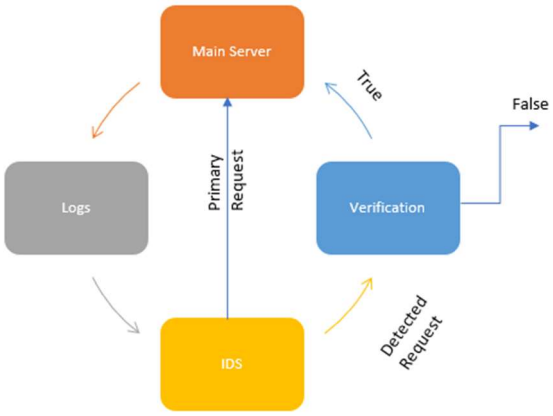

Figure- 5 Scenario 2

Experimental setup

The socket server-client model is used for simulation, and it replicates the IoT model with a central server that sends the request at varied periods and collects information such as temperature, pressure and humidity, and other types of data from the various nodes attached to it as IoT devices and bots. The number of bots connected varies from 0 to 100 in steps of 10.

Results

The models' efficiency is calculated as the ratio between the data sent from the client to the data received by the server.
The above scenarios are tested with and without the proposed model, and the results are recorded. The efficiency is increased by 60% using the honey pots.

Advantages and disadvantages

Generates the hacker's profiles when they attempt to intrude the system, simulates multiple hosts, and improves the security posters, and wastes the intruder's resources. Identify when a hacker is active in the personal or organizational network, then stimulates diverse TCP/IP packets. Honeypots have high true-positives and high-false positives values and low false-positive and false-negative, making the model highly reliable.
Honeypots can work as the zombie or bots and can attack other devices. They have a narrow view, and honeypots can identify the attack if directed towards the individual system. Honeypots need to be built perfectly. If there are any imperfections, they can leave fingerprints, and the attacker can use them.

No of Bots	Efficiency with Honeypots (%)	Efficiency without Honeypots (%).
10	99.64	42.17
20	99.24	41.25
30	98.79	41.25
40	98.3	39.09
50	97.75	37.82
60	97.14	36.42
70	96.47	34.88
80	95.73	33.18
90	94.92	31.31
100	94.04	29.29

Table- 3 shows the analysis of the proposed system with and without honeypots

Table- 3 shows that the number of bots in the request is increased in steps of 10, and the efficiencies are observed using the honeypots and without using the honeypots.

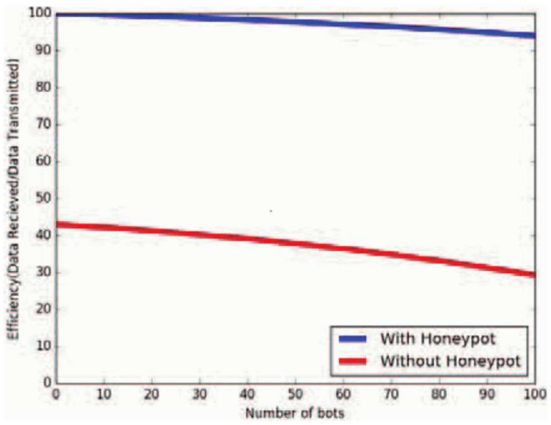

Figure-6 The behavior of the IoT systems with the proposed model.

Figure-6 shows the graphical representation of the efficiencies in the proposed system with and without the honeypots.

REFERENCES

[1]. Kumar, Virender, et al. "Neutralizing the Impact of Account or Service Traffic Hijacking in Cloud Computing." International Journal of Electronics Communication and Computer Engineering, vol. 5, no. 1, International Journal of Electronics Communication and Computer Engineering (IJECCE), Jan. 2014, p. 206.

[2]. B. Bouyeddou, F. Harrou, Y. Sun and B. Kadri, "Detection of smurf flooding attacks using Kullback-Leibler-based scheme," 2018 4th International Conference on Computer and Technology Applications (ICCTA), 2018, pp. 11-15, DOI: 10.1109/CATA.2018.8398647.

[3]. U. Dincalp, M. S. Güzel, O. Sevine, E. Bostanci, and I. Askerzade, "Anomaly Based Distributed Denial of Service Attack Detection and Prevention with Machine Learning," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2018, pp. 1-4, DOI: 10.1109/ISMSIT.2018.8567252.

[4]. R. Doshi, N. Apthorpe, and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Security and Privacy Workshops (SPW), 2018, pp. 29-35, DOI: 10.1109/SPW.2018.00013.

[5]. (2016) Threat advisory: Mirai botnet. Akamai. [Online]. Available: https://www.akamai.com/us/en/multimedia/documents/state-of-the-internet/akamai-mirai-botnet-threat-advisory.pdf

[6]. E. Bertino and N. Islam, "Botnets and Internet of Things Security," in Computer, vol. 50, no. 2, pp. 76-79, Feb. 2017, DOI: 10.1109/MC.2017.62.

[7] M. Anirudh, S. A. Thileeban and D. J. Nallathambi, "Use of honeypots for mitigating DoS attacks targeted on IoT networks," 2017 International Conference on Computer, Communication and Signal Processing (ICCCSP), 2017, pp. 1-4, DOI: 10.1109/ICCCSP.2017.7944057.

[8]. Andreea Bendovschi, Cyber-Attacks – Trends, Patterns, and Security Countermeasures, Procedia Economics and Finance, Volume 28, 2015, Pages 24-31, ISSN 2212-5671, https://doi.org/10.1016/S2212-5671(15)01077-1.