Detection of Application Layer DDoS Attack Based on SIS Epidemic Model

DALIA NASHAT\(^1\), S. KHAIRY\(^1\), and MONTASER M. HASSAN,\(^2\)

\(^1\)Department of Information Technology, Faculty of Computers and Information, Assuit University, Assuit, Egypt
\(^2\)Department of Biology, Faculty of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

Corresponding author: Dalia Nashat (e-mail: dnashat@aun.edu.eg).

The current work was funded by Taif University Researchers Supporting Project number (TURSP-2020/119), Taif University, Taif, Saudi Arabia.

ABSTRACT
distributed Denial of Service attack (DDoS) is one of the major threats to network security. The HTTP flooding attack is the hardest type of DDoS attacks to detect since the malicious packets are hidden in a huge amount of normal traffic. In this work, we introduce a new detection scheme for HTTP flooding attack by using Susceptible-Infective-Susceptible (SIS) model of an infectious disease which used in dynamic systems. During any time interval, the server can measure various values of attributes for its users like number of total connections, number of open connections and number of closed connections. These values can be used to detect any abnormal behavior or infected connections in a server by mapping this attributes with SIS model. Thus we can get suspected and infected connections during every time interval. Extensive trace driven simulation has been conducted to demonstrate the efficiency of the proposed scheme in terms of its detection rate and probability of false positive.

INDEX TERMS
Intrusion detection, DDoS attack, HTTP protocol, SIS epidemic model.

I. INTRODUCTION

Nowadays due to the COVID-2019 pandemic, using internet becomes a living habit of most people where traditional ways of marketing, education, communication, and broadcasting are replaced by online systems. Therefore, the size of internet data increased rapidly every day and attackers try to benefit from this by hiding their malicious packets in the huge number of legitimate packets.

One of the powerful and harmful attacks in computer networks is distributed denial of service (DDoS) attacks. DDoS attack shut down some of world’s most high-profile websites. For example, YAHOO! and eBay Feb 2000, WeaKnees Oct 2003, China Networks May 2009 and a lot of the world web sites [1], [2]. More than 69 countries were affected by the DDoS attacks in 2015 and its size is growing by 200 percent in 2019 [3]. Recent DDoS attacks target websites of medical organizations, delivery services, gaming, and educational platforms at the beginning of 2020 [4].

The DDoS attacks prevent legitimate users from using their service by exhausting network resources such as routers, links, and servers [5]. The DDoS attack starts when the attacker compromises relay hosts called masters, which in turn compromise attack machines called agents. The attacker attacks the victim by ordering agents to send a stream of malicious packets to the victim at the same time to exhaust its resources. The DDoS flooding attacks can be launched into network-layer by using network protocols (i.e. TCP, UDP, ICMP, and DNS protocol packets) or into the application-layer by exhausting the server resources [6].

The HTTP flooding attack is the most prevalent type of DDoS attacks that targets the application layer and it is the hardest type to detect because the attacker takes advantage of the HTTP connections to make the malicious traffic be encapsulated within the huge amount of normal traffic [7]. The application layer DDoS refers to a type of malicious behavior designed to target the server using HTTP requests.

The HTTP flooding attack is executed as illustrated in Fig.1. During attack time, the attacker sends a high rate of HTTP requests to the target server by using any possible request method to exhaust server resources (e.g., Sockets, CPU, memory, disk/database bandwidth, and I/O bandwidth) [5], [8]. Thus the server can not respond to any incoming legitimate requests.

Legitimate users and system administrators are advised to close inward ports and use only trusted protocols for outgoing communication because of the growth of cybersecurity threats. Since the HTTP service is being widely used by most internet applications, it is not easy to close inward ports but...
it is highly desirable to propose efficient detection techniques for HTTP flooding attack.

Recently, Many detection techniques were proposed to this attack. Most of these detection methods for application-layer DDoS attacks tried to find common characteristics of DDoS traffics and this leads to a false positive probability. Also, using a fixed threshold for detection is another shortcoming of recent methods. Therefore, new detection techniques do not require threshold setting or previous knowledge of DDoS characteristics are highly desirable. The proposed method deals with the problem of detecting attack HTTP requests as the problem of detecting infected individuals during diseases spreading in any community. We use SIS epidemic model as one of the most trusted models to study the mechanisms of disease spreading. We can guarantee high detection rate and zero false positive probability without prior knowledge of normal traffic characteristics or adopt fixed threshold.

The objective of the mathematical model of an infectious disease is to simulate the transmission process of the disease [9]–[11]. There are a lot of mathematical models which can be used to represent dynamic systems like SIS. In the proposed method, we use SIS mathematical model since it is the most suitable model for network attributes. We map HTTP connection attributes with SIS parameters and solve its equations, so suspected and infected connections during every time interval can be detected accurately.

The rest of this paper is organized as follows. Section 2 reviews the literature related to this work. Section 3 presents the The SIS epidemic model. The proposed method is described in Section 4. Experimental results are presented in Section 5. Section 6 concludes this work and introduces the future directions.

II. RELATED WORK

The DDoS attacks are a hard problem to solve because there are no common characteristics of DDoS streams that can be used for their detection. Also, the attacker can simulate legitimate users behavior to avoid detection or prevention. DDoS attacks detection techniques can be classified into Signature based detection and Anomaly based detection. The signature based detection techniques work by using the database of known attack signatures. In the other hand, the anomaly based detection techniques works by recognizing anomalies in system behaviors.

A fast all-packets-based DDoS attack detection approach (FAPDD) is proposed in [12]. This approach used a new time series network graph model to simplify the processing of network traffic handling. Also, authors used a dynamic threshold and freezing mechanism to display standard traffic changes. An attempt to model the behavioral dynamics of legitimate users is presented in [13]. This model used the simple annotated probabilistic timed automata (PTA) and the suspicion scoring mechanism to distinguish between legitimate and malicious users. The authors conducted their experiments on public datasets to evaluate the model performance in terms of detection rate and false-positive probability.

A machine learning-based intrusion detection system is proposed in [14]. It used the Tree-CNN hierarchical algorithm and the soft root sign activation function for detecting DDoS. The performance evaluation of this detection system shows that it has low complexity and less processing time in comparison with other machine learning methods.

Authors in [15] present DDoS detection method using the Matching Pursuit. They use some characteristics of network traffic for detecting low rate DDoS attacks efficiently. This method deployed the K-SVD algorithm on the parameters of the network traffic. They also use Matching Pursuit and Wavelet techniques and propose a hybrid DDoS detection framework that combines these methods with an artificial neural network. their approaches achieved 99% true positive rate and 0.7% false positive rate. Authors in [16] proposed an online detection algorithm for DDoS attack by using a kernel-based learning algorithm, the Mahalanobis distance and a chi-square test. The main disadvantage of this supervised algorithm that it is not able to detect different types of DDoS attacks.

A new detection method based on conditional probability and Bayes theorem is presented in [17]. This method first calculated the probability value for every normal traffic attribute. Then, authors compute the conditional probability for the same attribute in any incoming connection given the occurrence of the same value in the previous normal traffic. Finally, the total probability is calculated by using the Bayes theorem to classify it either as normal or abnormal connection.

A statistical model called the RM (rhythm matrix) was proposed in [18] to detect application layer DDoS attack. The
RM model can detect DDoS attacks according to the increase of the abnormality degree in the RM and distinguish flash crowds from application layer DDoS attacks. Authors in [19] proposed an entropy rate measurement (ERM) for detecting DDoS attacks. This is based on the differences between the probability distributions and the number of flows.

A semi-supervised weighted k-means detection method was proposed in [20]. Authors first used Hadoop-based hybrid feature selection algorithm to solve the problem of outliers and local optimal, then they deploy Semi-supervised K-means algorithm using hybrid feature selection (SKM-HFS) to detect the DDoS attacks. Article in [21] presents a comprehensive study of many machine learning detection algorithms like KNN, RF, MLP, Adaboost, and Naive-Bayes to extract the best features for detecting DDoS attacks.

III. THE SIS EPIDEMIC MODEL

The modeling of infectious diseases is a tool used to study the mechanisms of diseases spreading to predict the outcome of an epidemic and evaluate strategies to control it [22]. There are many mathematical models that can be used to represent dynamic systems like Susceptible and Infected (SI), Susceptible-Infected-Recovered (SIR), Susceptible-Infective-Susceptible (SIS) and Susceptible-Infective-Recovered-Susceptible (SIRS). In the proposed detection method, we use the SIS model since it is one of the simplest models that makes it suitable for computer networks. Also, the parameters of this model can fit HTTP connection attributes.

A disease is said to follow the SIS model if infected individuals do not have an exposed period and recover with no immunity. In this model a population with \( N \) individuals is classified into two categories: Susceptible \((S)\) and Infected \((I)\). Susceptible means individuals who do not have the disease but can be infected by disease. Infected means individuals who have the disease and can infect susceptible individuals. The SIS model assumes that new individuals are born in the susceptible class.

The flow diagram in Figure 2 describes the dynamics of the SIS epidemic model without demography [11]. The numbers of susceptible individuals at time \( t \) is denoted by \( S(t) \) and the numbers of infected individuals at time \( t \) by \( I(t) \).

The previous two equation should be solved to obtain the values of \( S \) and \( I \). The community is assumed to be infected if the value of \( I \) is greater than zero.

IV. OVERVIEW OF THE PROPOSED DETECTION SCHEME

In this section, we first show the overall architecture of the proposed scheme for detecting HTTP flooding attack. Fig.3 illustrates the diagram of the proposed detection scheme during one time interval.

In application layer DDoS attacks, there is no difference between an attack HTTP request and a normal HTTP request, and this makes the detection more difficult. The main idea of the proposed detection scheme is to detect the HTTP flooding attacks in any server based on the abnormal behavior of its users. The proposed detection scheme can deal with the problem of detecting attack HTTP requests as the problem of detecting infected individuals during diseases spreading.

\[
\frac{dS}{dt} = a(S + I) - bS - \beta SI \tag{1}
\]

\[
\frac{dI}{dt} = \beta SI - bI - \alpha I \tag{2}
\]

where

- \(a\) is the birth rate of the population;
- \(b\) is the mortality rate of the population;
- \(\alpha\) is the mortality caused by the disease;
- \(\beta\) is the disease transmission coefficient.
in any community. In other words, we can assume that all HTTP connections in any time interval as population in a community and attack like a disease. Therefore, we use SIS epidemic model as one of the most trusted models to study the mechanisms of disease spreading.

We assume that every interval time is a community and use the SIS model equations to decide if this time is a normal time interval or an attack time interval (i.e., detecting if the community is infected with disease or not). To achieve this goal, we first map the model parameters to HTTP protocol attributes during every time interval as described later in this section, then we solve the model equations to get the value of I which means infected community if its value is greater than zero.

During any time interval, the server can measure various values of HTTP attributes for its users like number of all connections, number of open connections (i.e., incomplete connections that are still open due to a server problem), and number of closed connections. These values can be used to detect any abnormal behavior or infected connections in a server by mapping these attributes with SIS model. Therefore, we need first to map these attributes with SIS model parameters. We can explain the steps of the proposed detection method in a time interval as follows:

**step1:** Map the HTTP attributes with the parameters of SIS model where \( a \) in HTTP attributes represents number of connection, \( b \) represents number of closed connection, \( \alpha \) represents the number of connections that are still open due to a server problem, then \( \beta \) can be calculated as:

\[
\beta = \frac{\text{Number of open connections}}{\text{Number of all connections}} \times 1000 \tag{3}
\]

**step2:** Solve the two equations of SIS model together to have \( S \) and \( I \).

\[
a(S + I) - bS - \beta SI = 0 \tag{4}
\]

\[
\beta SI - bI - \alpha I = 0 \tag{5}
\]

**step3:** From Eq.5

\[
I[-b + \beta S - \alpha] = 0 \tag{6}
\]

**step4:** From Eq.6 \( S \) can be computed as

\[
S = \frac{b + \alpha}{\beta} \tag{7}
\]

**step5:** Substituting Eq.7 in Eq.4 to get \( I \)

\[
I = \frac{(a - b)(b + \alpha)}{\beta(b + \alpha - a)} \tag{8}
\]

From the previous steps, the value of \( I \) in every time interval can be computed accurately. Thus, depending on this value we can decide if this interval is a normal interval or an attack interval. If this value is greater than zero, the HTTP traffic during this time is classified as an attack.

### V. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed scheme, a real-life Internet traces collected from Canadian Institute for Cybersecurity (CICIDS2017) are used which resembles the true real-world data (PCAPs). The data capturing period started at 9:00am, Monday, July 3, 2017 and ended at 17:00 on Friday July 7, 2017, for a total of 5 days. Monday is the normal day and only includes the benign traffic. The implemented attacks include Brute Force FTP, Brute Force SSH, DoS, Heartbleed, Web Attack, Infiltration, Botnet and DDoS. They have been executed both morning and afternoon on Tuesday, Wednesday, Thursday and Friday. CICIDS2017 are available in [23] and it’s analysis available through [21]. To further prove the detection efficiency of the proposed detection scheme, we compare the proposed method with previous detection methods using the same dataset and show that it achieves significantly higher performance than other schemes.

Before applying the proposed detection scheme, we calculated the number of HTTP requests in the used traffic traces CICIDS2017 over time intervals of 20 seconds. Fig.4 shows the number of HTTP requests on Wednesday from 10:25 to 10:47. We can easily see from this figure that the number of HTTP requests increases during attack time, which begins from time interval 54 to time interval 65. It is also noticed from this figure that the rate of HTTP flooding attack exceeds 5000 packets per the observation period of 20 seconds, and this differs from the normal behavior whose maximum rates does not exceed 500 packets per second as we mentioned in the previous section.

![FIGURE 4. Incoming request packets on Wednesday from 10:25 to 10:47](image)

Also, Fig.5 shows the number of HTTP requests for different times on the same day from 10:55 to 11:47. The attack starts at time interval 1 to time interval 14. We can easily see that the rate of HTTP flooding is about 6000 packets per the observation period of 20 seconds, and this differ from the normal behavior whose maximum rates does not exceed 1000 packets per second.
In the present work we fix the time interval as 20 seconds in our simulation. \( a \) in HTTP attributes represents number of connection, \( b \) represents number of closed connection, \( \alpha \) represents the number of open connections. Therefore, \( \alpha \) can be calculated by two different methods, in first method \( \alpha \) is calculated as:

\[
\alpha = \text{all connections} - Cc
\]  

(9)

Where \( Cc \) is the theoretical capacity of the server, which means the number of connections exceeds the server capacity during the time interval. The second method, \( \alpha \) is considered as the number of open connections due a server error, so \( Cc \) isn’t used in this method.

**TABLE 1.** The results of the proposed scheme by using \( Cc \)

| Duration Time | Detection Rate | false positive |
|---------------|----------------|----------------|
| 167 sec.      | 1.0            | 0.0            |
| 284 sec.      | 1.0            | 0.0            |
| 485 sec.      | 1.0            | 0.0            |

Table 1 shows the detection rate and false positive probability of the proposed scheme with different duration time 167 sec, 284 sec and 485 sec and using the first method for calculating \( \alpha \). It is clear that the proposed scheme achieves high detection rate and zero false positive probability. Moreover, our detection scheme can detect the attack in the first time interval of the flooding attack duration time.

**TABLE 2.** The results of the proposed scheme without using \( Cc \)

| Duration Time | Detection Rate | false positive |
|---------------|----------------|----------------|
| 167 sec.      | 1.0            | 0.0            |
| 284 sec.      | 1.0            | 0.0            |
| 485 sec.      | 1.0            | 0.0            |

The results of the proposed detection method by using the second method of calculating \( \alpha \) are illustrated in Table 2. As shown in this table, the presented scheme can still guarantee the same detection rate and false positive probability. Therefore, the method of calculating \( \alpha \) does not affect the detection rate or false positive probability.

To evaluate the performance of the proposed scheme, the results are compared with AdaBoost scheme, Deep Reinforcement Learning Algorithm and other algorithms mentioned in [21] which deployed on the same used dataset (i.e CICIDS2017). We use the same evaluation parameters: Precision (Pr) which denoted to the ratio of true classified attacks (TP), in front of all the classified connections (TP+FP). Recall (Rc) or Sensitivity that means the ratio of true classified attack (TP), in front of all generated connections (TP+FN) and F-Measure (F1) which means the combination of the precision and recall into a single measure.

Table 3 shows clearly that the proposed scheme success to detect all DDoS attacks with zero false positive probability. This table shows that our proposed scheme achieves significantly higher detection probability than other schemes. For example, when the detection probability of RF and ID3 is 98% the detection probability of the proposed scheme is 100%.

**VI. CONCLUSION**

In this paper, a new detection scheme for application layer HTTP flooding attack is presented by using SIS mathematical model. We use the SIS model since it is one of the simplest models that makes it suitable for computer networks. Also, the parameters of this model can fit HTTP connection attributes. Through mapping network attributes with SIS parameters and solve it’s equations, suspected and infected connections can be detected in every time interval accurately. The proposed scheme succeed to detect all HTTP flooding attacks with no false positive as previously shown. In addition, it outperforms other detection scheme in the detection rate.

**ACKNOWLEDGMENT**

The authors extend their appreciation to Taif University for funding current work by Taif University Researchers Supporting Project number (TURSP-2020/119), Taif University, Taif, Saudi Arabia.

**REFERENCES**

[1] D. Nashat, X. Jiang, and S. Horiguchi, “Router Based Detection for Low-Rate Agents of DDoS Attack,” in International Conference on High
A. Habibi and Ghorbani, “CIC-IDS-2017.”

D. J. Daley and J. Gani, Epidemic modelling: an introduction. No. 15,

Y. Gu, K. Li, Z. Guo, and Y. Wang, “Semi-supervised k-means ddos attacks,” IIEE transactions on communications, vol. 93, no. 5, pp. 1113–1121, 2010.

N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, “Botnet in DDoS Attacks:Trends and Challenges,” IEE Communications Surveys & Tutorials, vol. 17, no. 4, pp. 2242–2270, 2015.

H. Beitollahi and G. Deconinck, “Tackling Application-Layer DDoS Attacks,” Procedia Computer Science, vol. 10, pp. 432–441, 2012.

S. T. Zargar, J. Joshi, and D. Tipper, “A survey of Defense Mechanisms Against Distributed Denial of Service (DDoS) Flooding Attacks,” IEEE communications surveys & tutorials, vol. 15, no. 4, pp. 2046–2069, 2013.

F. Brauer and P. van den Driessche, “Models for Transmission of Disease with Immigration of Infectives,” Mathematical Biosciences, vol. 171, no. 2, pp. 143–154, 2000.

Y. Wang, J. Wang, and L. Zhang, “Cross Diffusion-Induced Pattern in an SI Model,” Applied Mathematics and Computation, vol. 217, no. 5, pp. 1965–1970, 2010.

S. Aly, H. B. Khenous, and F. Hussien, “Turing Instability in a Diffusive SIS Epidemiological Model,” International Journal of Biomathematics, vol. 8, no. 01, pp. 155–160, 2015.

X. Liu, J. Ren, H. He, B. Zhang, C. Song, and Y. Wang, “A fast all-packets-based ddo attack detection approach based on network graph and graph kernel,” Journal of Network and Computer Applications, vol. 185, p. 103079, 2021.

A. Prased and P. S. Thilagam, “Modelling behavioural dynamics for asymmetric application layer ddos detection,” IEEE Transactions on Information Forensics and Security, vol. 16, pp. 617–626, 2020.

R. V. Mendoça, A. A. M. Teodoro, R. L. Rosa, M. Saadat, D. C. Melgarejo, P. H. J. Nardelli, and D. Z. Rodríguez, “Intrusion detection system based on fast hierarchical deep convolutional neural network,” IEEE Access, vol. 9, pp. 61024–61034, 2021.

D. Erhan and E. Ararim, “Hybrid ddo detection framework using matching pursuit algorithm,” IEEE Access, vol. 8, pp. 118912–118923, 2020.

S. Daneshgadeh Çakmakçı, T. Kemmerich, T. Ahmed, and N. Baykal, “Online ddo attack detection using mahalanobis distance and kernel-based learning algorithm,” Journal of Network and Computer Applications, vol. 168, p. 102756, 2020.

S. Khairy and D. Nashat, “Detection of application layer ddos attacks based on bayesian classifier,” Assiut Univ. J. of Mathematics and Computer Science, vol. 48, pp. 1–13, 2019.

H. Lin, S. Cao, J. Wu, Z. Cao, and F. Wang, “Identifying application-layer ddos attacks based on request rhythm matrices,” IEEE Access, vol. 7, pp. 164480–164491, 2019.

L. Zhou, K. Sood, and Y. Xiang, “Erm: An accurate approach to detect ddos attacks using entropy rate measurement,” IEEE Communications Letters, vol. 23, no. 10, pp. 1700–1703, 2019.

Y. Gu, K. Li, Z. Guo, and Y. Wang, “Semi-supervised K-means ddo detection method using hybrid feature selection algorithm,” IEEE Access, vol. 7, pp. 64351–64365, 2019.

I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization,” in ICISSP, pp. 108–116, 2018.

B. J. Daley and J. Gani, Epidemic modelling: an introduction. No. 15, Cambridge University Press, 2001.

A. Habibi and Ghorbani, “CIC-IDS-2017.”

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3132130, IEEE Access

---

**S. KHairy** Received her Bachelor sciences degree and Master degree in computer science from Assuit University, Faculty of Science, Department of Mathematics, Egypt. In 2012 and 2019 respectively. Her Master’s focused on Network Security. She was Demonstrator in Faculty of Science, Department of Mathematics, Assuit University, Assuit, Egypt, from Feb. 2013 to Nov. 2020. She is currently an Assistant Lecturer in Department of Information Technology, Faculty of Computers and Information, Assuit University, Assuit, Egypt.

**Dalia Nashat** received the PhD degree in Networks Security at Tohoku University, Japan in 2010. She was an Assistant Professor of Computer science at Faculty of Science, Assuit University, Egypt, from 2010 to 2020. She is currently an Assistant Professor in Department of Information Technology, Faculty of Computers and Information, Assuit University, Assuit, Egypt. Dr. Dalia was an Assistant Professor in Taif University, KSA, from 2011 to 2014. Her main research interests are: Networks security, Intrusion Detection and Signal Processing.

**MONTASER M. HASSAN** received his B.S., M.S. and Ph.D degrees in 1996, 2003, and 2006 respectively, all from Ain Shams University, Cairo, Egypt. Ph. D was fulfilled with channel grant to Prairie View Texas A&M University, Texas, USA. He is associate professor in Ain Shams University, Cairo, Egypt and working currently in Taif University, KSA. Research interests include Ecology and bio-control of pests, Zoology, Aquatic invertebrates and macrobenthos, Molecular biology.