Journal of Engineering and Applied Sciences 14 (Special Issue 3): 5979-5984, 2019
ISSN: 1816-949X

A Novel Dimensional Reduction Model for Detecting DDoS Attacks

[1]Alaa M. Hasan Abu Daym and [2]Bilal Majeed Abdulridha Al-Latteef
[1]Ministry of Education, Directorate General of Education of Karbala, Karbala, Iraq
[2]Ministry of Education, The General Directorate of Qadisiyah Education, Al-Qadisiyah, Iraq
Rosekaby@gmail.com, bil.pro84@gmail.com

Abstract: This study aims to present a simple yet powerful dimensional reduction approach for detecting DoS DDoS attacks. Two detection methodologies were hybridized and used to build up our system. The system is based on making use of these two methodologie's strength points and overcome their weakness points. The findings of this research indicates that hybrid systems can provide best detection rates, especially, when they have the capability of enhancing their detection rate as long as they are being used.

Key words: DoS/DDoS attacks, dimensionality reduction, signature-based detection, anomaly-based detection, capability, methodologies

INTRODUCTION

Internet development is leading the world towards a whole new digital world where most of our daily life operations can be executed online. This means that all the individual's critical information should be available for online services which increases the burden to provide secure online systems but as in real life no matter how much the system is secure, the threats of intrusions and all kinds of privacy invasion will always be there.

Denial of Service (DoS) attacks are new kinds of cyberattacks that has evolved during the last two decades as literature records. DoS attack aims to make network resources unavailable to its legitimate users (Tama and Rhee, 2015; Jia *et al*., 2016). Distributed Denial of Service (DDoS) attack is in fact multiple DoS attacks launched from multiple connected devices that are distributed across the internet. The motivation behind DoS/DDoS could either be hacktivism, cyber vandalism, extortion, personal rivalry, business competition and/or cyber warfare (Anonymus, 2018; Jia *et al*., 2016).

Detecting DoS/DDoS has become a huge challenge, therefore, many approaches were proposed. These approaches used different techniques and algorithms some of them used data mining techniques such as association, classification, clustering and hybrid methods. Dimensional reduction is a data mining method that has been used in building many detection systems for intrusion and DDoS attacks. Dimensionality reduction means reducing space dimensions to the possible minimum number of dimensions in order to get to the desired goal. For data dimensions represent the number of variables that are measured on each observation (Fodor, 2002). All the dimensionality reduction based approaches for DDoS attack's detection were build according to this concept. In our system we employed dimensionality reduction concept for detecting DDoS attacks.

MATERIALS AND METHODS

Dimensionality reduction: Dimensionality reduction is the process of reducing the number of variables and features in any dataset. Dimensionality reduction can be divided into two subcategories called feature selection and feature extraction (CC., 2017; Wang and Paliwal, 2003). Feature selection is filtering irrelevant or redundant features from a dataset. The key difference between feature selection and feature extraction is that feature selection keeps a subset of the original features while feature extraction creates a new smaller set of features that still captures most of the useful information (Anonymus, 2017; Buragohain *et al*., 2015).

Our system uses feature extraction method to calculate a set of features, the features include: number of packets, average of packet size, number of bytes, packet rate, bitrate, protocol type, num_failed_logins, duration, count and srv_count. These features obtained from studying the incoming network traffic and then calculating features to set them as a pattern for each connection in order to construct a complete profile for the monitored network system that will later be used in

Corresponding Author: Alaa M. Hasan Abu Daym, Ministry of Education, Directorate General of Education of Karbala, Karbala, Iraq, Rosekaby@gmail.com

classifying each incoming network connection to decide whether it's a normal connection or it's a DDoS attack.

Feature extraction: Feature extraction calculates the selected features for captured packets. We propose these features by observing the characteristics of each connection's packets. These features can be used to recognize and classify incoming packets. Experiments showed that these features contain significant information related to the presence of a DDoS attack. The following explains the proposed features.

Average packet size: DDoS attacks flood victim to consume system resources, then average packet size increases in attack time. We use this feature to identify DDoS attacks.

Number of packets: DDoS attacks send a great number of packets to the victim network. Therefore, the number of packets increases in comparison to normal case. We exploited this feature to detect DDoS attacks.

Number of bytes: Increase in number of bytes demonstrates launching DDoS attacks.

Packet rate: This feature shows the packet rate sent from a source address to a destination in a specific time span. Packet rate increases significantly in attack time.

Bitrate: A very high rate of this feature indicates launching DDoS attack (Karimazad and Faraahi, 2011).

Protocol type: Connection protocol (tcp, udp, icmp).

Num_failed _logins: Number of failed login attempts.

Duration: Duration of connection in seconds.

Count: Number of connections to same host as current connection in past 2 sec.

srv_count: Number of connections to same service as current connection in past 2 sec (Staudemeyer and Omlin, 2014).

Intrusion detection: An intrusion detection system is a software tool used to detect unauthorized access to a computer system or network. An intrusion detection system is capable of detecting all types of malicious network traffic and computer usage. This includes network attacks against vulnerable services, data driven attacks on applications, host-based attacks such as privilege escalation, unauthorized logins and access to sensitive files and malware. An intrusion detection system is a dynamic monitoring entity that complements the static monitoring abilities of a firewall. An intrusion detection system monitors traffic in a network in promiscuous mode, very much like a network sniffer. The network packets that are collected are analyzed for rule violations by a pattern recognition algorithm. When rule violations are detected, the intrusion detection system alerts the administrator (Patcha and Park, 2007).

According to Denning (1987), the idea of detecting intrusions could be presented by assuming that a network user would behave in a manner that enables automatic profiling. Which means that the behavior model could be constructed for the monitored network flow by the intrusion detection system and subsequent behavior could be verified such that any deviation from the norm will be considered an anomaly. Denning mentioned several models that are based on statics, Markov chains, time series, etc (Patcha and Park, 2007).

DoS/DDoS attacks: Denial of Service attack (DDoS) is one of the major types of intrusion attacks that has evolved during the last two decades as the literature records. As the name implies its concept is based on making internet based services or resources unavailable for their legitimate users. This prevention could take the form of disruption. So, DoS attack could disrupt user's connectivity by exhausting bandwidth, router processing capacity or network resources. These are essentially network/transport-level flooding attacks (Zargar *et al*., 2013). Or disruption could take the form of exhausting the server resources (e.g., sockets, CPU, memory, disk/database bandwidth and I/O bandwidth), these essentially include application-level flooding attacks.

Let's take a simplified example from our daily life events to explain DOS's concept in more clear way for example booking an appointment in a doctor clinic where a patient can call to get an a ppointment as shown in Fig. 1:

A DDoS attacker to this system will act as following: although he is not a patient he will continue to call on the clinic's number in order to keep this phone line busy for as much time as possible in this way the legitimate user which in, here, represents the patient will not be able to reach the resource which is represented in the clinic as shown in Fig. 2.

DDoS which denotes to distributed denial of service attack is in fact multiple DoS attacks distributed over multiple attackers. Let's go back again to our clinic

Fig. 1: Normal connection

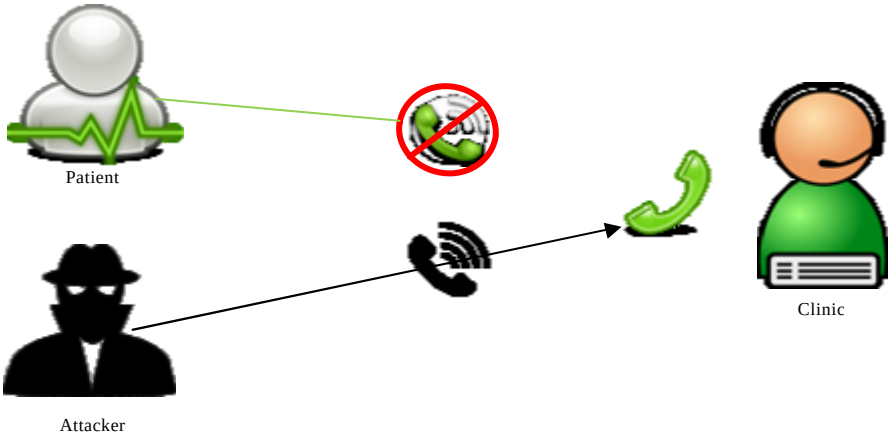

Fig. 2: DDoS attack

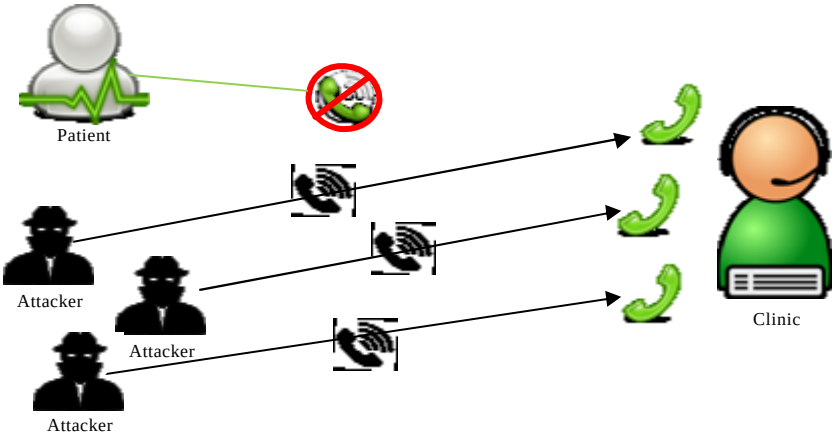

Fig. 3: DDoS attack (clinic booking system)

example, in DDoS attack case the clinic will provide more than one phone number in order to serve as much patients as possible, let's say three phone numbers. DDoS attack will include instead of one attacker there will be three attackers each one of them occupies a phone number and try to keep it busy as much as possible, preventing the patient from getting to clinic's booking system. This idea is explained in Fig. 3.

Although, widely known web sites such as Yahoo, CNN, eBay and Amazon.com were well-equipped in security, reports show that in 2000 they were damaged by DDoS attacks (Fengxiang and Abe, 2007). The DDoS attacks usually do not exploit of security vulnerabilities of network-connected systems but instead they aim to disrupt victim services by overwhelming the processing capacity of system or by flooding the bandwidth of the target. SYN flooding attacks, DNS flooding attacks and Smurf attacks are major DDoS attacks according to Arbor's survey in 2008 (ARBOR Networks, 2009).

There are two separate steps of DDoS attacks: compromising internet hosts and flooding the victim system. In the first step, attacker compromises a large

number of internet hosts using vulnerable software installed on them. Then, attacker installs attack software on compromised systems. These hosts called agent and attacker controls them via. handler systems. In the second step, attacker commands to the agents through handler systems to generate and send high volume of useless packets to victim simultaneously (Karimazad and Faraahi, 2011). The volume of sent packets is so high that the victim cannot response to them and then be exhausted. Using multiple agents and IP (Internet Protocol) spoofing techniques in DDoS attacks causes the detection becomes more difficult (Douligeris and Mitrokotsa, 2004).

For DDoS detection methodologies, many techniques were proposed each one of them has its advantages and drawbacks and its own usage conditions. These techniques can be divided into two major methodologies: signature-based and anomaly-based detection methodology.

Signature-based detection: Every activity (legitimate and intrusive) over network has a unique pattern. These patterns can be used to detect which activities are going on the network. So, these patterns are also called as signatures. Signatures of known intrusive activities are defined and used to detect their existence. But there are two major problems in this approach both problems are related with the manual process usually carried out to create signature of attack. First, a detailed and precise knowledge about attacks process is required to define its signature. If defined signature is too simplified, it will generate high false positive rate. On other hand, if it is too specific, it will result in high false negative rate. Second, some time is required to gain detailed and precise knowledge about attack. This introduces delay between the first time attack is reported and generation of signatures to detect it. Thus 0 day or novel attacks are serious threat for computer systems. Olusola *et al.* (2010) proposed some approaches to make this process easier by correlating and thus reducing the number of alerts to analyze but major problems are still unsolved (Gangwar and Sahu, 2014).

Anomaly-based detection: Has always attracted many researchers, instead of focusing and tracking the malicious traffic, it tracks and monitors the normal traffic and tries to generalize their behavior with a general pattern or profile that contains the most significant information about the monitored network system. The network behavior is in conformity with the predefined behavior. Then it is accepted or else it triggers the event in the anomaly detection. The accepted network behavior can be prepared or learned by the specifications of the network administrators.

The major advantage of AD compared to signature based engines is at a novel attack for which a signature does not exist can be detected if it falls out of the normal traffic patterns (Zekri *et al.*, 2017). On the other hand, AD consumes more time than signature-based detection, since, it is dealing with the overwhelming class that is the normal legitimate class. In this study, we present a hybrid system that makes use of both signature-based and anomaly based detection methods in order to make use of both method's advantages and overcome their drawbacks as described in the following section.

RESULTS AND DISCUSSION

Deep mining approach based novel dimension reduction model for DDoS detection

Novel dimensional reduction model for DDoS detection: Our proposed system works on the idea of calculating difference by Difference Measuring Function (DFMF), then comparing this difference value within a previously specified range by Decision Making Function (DMF). This is the main concept and how and where it'll be applied is explained in details in this study.

The incoming traffic first enters into a signature based detection system which requires a previously generated data base DB_1 that contains signatures of all known attacks. If the incoming traffic matches specific signature in DB_1, then the administrator will be notified to take further action. If the incoming connection doesn't match any signature in DB_1, then this means that it's either a normal legitimate connection or a new connection the system has never exposed to. So, what will happen next is that it's features which include ten features and they are number of packets, average of packet size, number of bytes, packet rate, bitrate, protocol type, num_failed_logins, duration, count and srv_count as described previously will be extracted. After that these features will enter into the function (DFMF) that will calculate the difference between the incoming connection's features values and values stored in a data base DB_2 values in DB_2 are extracted from a set of normal traffic that the system has been exposed to.

After calculating the difference value for each feature, a set of ten values (according to feature count) will be generated, let's denote it with DS, each value in DS will be compared with a range in another set which we'll denote it with RS that contains ten ranges, a range for each feature type. Each range has been set by calculating the difference between each normal pattern's feature in DB_2 with all other pattern's features in DB_2. A set of values will be generated after difference calculations, the ranges are specified from these values and stored in DB_3 that will

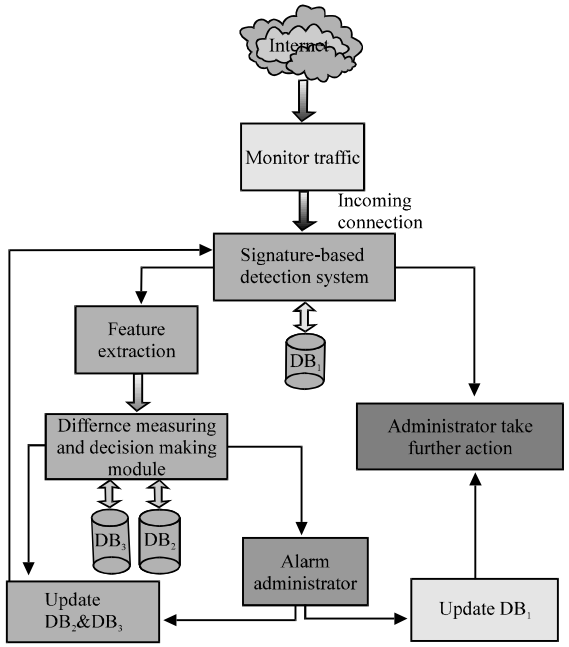

Fig. 4: System's model design

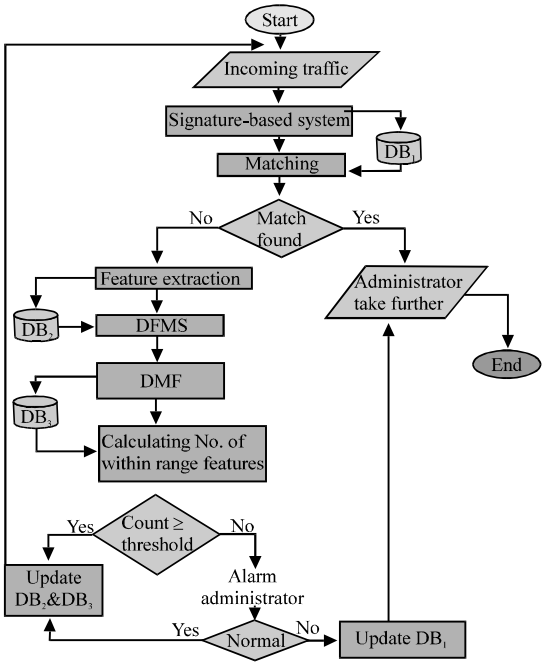

Fig. 5: Data-flow chart for the system model

later be used in (DMF) to make a decision about the incoming connection whether it is normal or not. Figure 4 shows the system's model.

Difference Measuring Function (DFMF): This function receives the set of previously described features, f_i, I = 1, ..., 10. DB_2 is a 2-dimensional array, each value in it is addressed as $F_{i,j}$, i = 1, …, 10; j = 1, ..., m; m = number of normal patterns the system has exposed to. The difference for ten features of the incoming connecton and m number of normal patterns in DB_2 will be measured as follows:

$$d_{i,j} = \text{Absolute value}(f_i - F_{i,j})$$
$$i = 1, ..., 10; j = 1, ..., M$$

After that the average value of each feature will be found:

$$D_i = \text{Average}(d_{i,j})$$

D_i will contain a set of ten average values for each feature. This array D_i will next enter to DMF.

Decision Making Function (DMF): The resulted difference values in D_i will be compared with all range's values in DB_3 by (DMF). These ranges have been specified previously by calculating the difference between each feature value for each pattern in DB_2 with the same feature values of all other patterns and the range for each feature will be specified accordingly. D_i will contain ten values, a value for each feature, DMF will calculate the count number of values that are within their range.

If count = 7, then the incoming connection is classified as normal and its pattern will be added to DB_2 and DB_3 values will be modified accordingly. Otherwise, the administrator will be alarmed to check and decide whether the incoming connection is normal or a possible DDoS attack the complete system's model is elaborated in Fig. 5 which shows the system's data-flow chart.

CONCLUSION

Machines are like human beings, they can't perform, act, detect or do anything until they are taught to just like what we do for our children when they are young. We try to give them the best teaching and learning that we could provide in order to make them able to face most of life's situations. But life is unpredictable and anything can go wrong and no man ever lived or alive knows everything about life and each one of us learns from his own experiences. This way is exactly the way that all machines and detection systems work.

Our system is about presenting simple yet powerful DDoS detection methods. Many DDoS detection methods have been presented, each one of them has its own advantages and drawbacks based on (Gangwar and Sahu, 2014) and many other references there are two major DDoS methodologies and all other methods are either derived from them or resulted from combining both of

them in a hybrid model such as in our presented model. However these methods are signature-based and anomaly-based detection methodologies. Regardless, the used method all methods share one of the two concepts, they are either based on the concept of measuring the correlation or the difference. All methods work according to one of these concepts or could combine both of them according to the type of the presented model. So, what we did in our presented system here, we simply used the difference as a standard to classify the incoming connection to normal or DDoS attack and in order to strengthen the system's detection rate, we hybridized it with a signature-based detection model to get the best detection rate.

REFERENCES

ARBOR Networks, 2009. Worldwide infrastructure security report. http://www.arbornetworks.com/. research/ infrastructure-security-report.

Anonymus, 2017. Dimensionality algorithms: Strengths and weaknesses. EliteDataScience, San Francisco, California, USA.

Anonymus, 2018. Protect the pulse of your business. Imperva, Redwood Shores, Redwood city, California, USA. https://www.imperva.com.

Buragohain, C., M.J. Kalita, S. Singh and D.K. Bhattacharyya, 2015. Anomaly based DDoS attack detection. Intl. J. Comput. Appl., 123: 35-40.

CC., 2017. Machine learning-dimensionality reduction-feature extraction and selection. Cognitive Class, USA.

Denning, D.E., 1987. An intrusion-detection model. IEEE Trans. Software Eng., SE-13: 222-232.

Douligeris, C. and A. Mitrokotsa, 2004. Ddos attacks and defense mechanisms: Classification and state-of-the-art. Comput. Networks, 44: 643-666.

Fengxiang, Z. and S. Abe, 2007. A heuristic DDoS flooding attack detection mechanism analyses based on the relationship between input and output traffic volumes. Proceedings of the 2007 16th International Conference on Computer Communications and Networks, August 13-16, 2007, IEEE, Honolulu, Hawaii, USA., ISBN:978-1-4244-1250-1, pp: 798-802.

Fodor, I.K., 2002. A survey of dimension reduction techniques. Technical Report, pp: 1-18. https://computation.llnl.gov/casc/sapphire/pubs/148494.pdf.

Gangwar, M.A. and M.S. Sahu, 2014. A survey on anomaly and signature based Intrusion Detection System (IDS). Intl. J. Eng. Res. Appl., 4: 67-72.

Jia, B., Y. Ma, X. Huang, Z. Lin and Y. Sun, 2016. A novel real-time ddos attack detection mechanism based on MDRA algorithm in big data. Math. Prob. Eng., 2016: 1-10.

Karimazad, R. and A. Faraahi, 2011. An anomaly-based method for DDoS attacks detection using RBF neural networks. Proceedings of the International Conference on Network and Electronics Engineering, September 16-18, 2011, IACSIT Press, Singapore, pp: 16-18.

Olusola, A.A., A.S. Oladele and D.O. Abosede, 2010. Analysis of KDD99 intrusion detection dataset for selection of relevance features. Proceedings of the World Congress on Engineering and Computer Science (WCECS) Vol. 1, October 20-22, 2010, San Francisco, USA., ISBN:978-988-17012-0-6, pp: 1-7.

Patcha, A. and J.M. Park, 2007. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Comput. Networks, 51: 3448-3470.

Staudemeyer, R.C. and C.W. Omlin, 2014. Extracting salient features for network intrusion detection using machine learning methods: Research article. South Afr. Comput. J., 52: 82-96.

Tama, B.A. and K.H. Rhee, 2015. Data mining techniques in DoS/DDoS attack detection: A literature review. Inf. Japan, 18: 3739-3747.

Wang, X. and K.K. Paliwal, 2003. Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. Pattern Recognit., 36: 2429-2439.

Zargar, S.T., J. Joshi and D. Tipper, 2013. A survey of defense mechanisms against Distributed Denial of Service (DDoS) flooding attacks. IEEE. Commun. Surv. Tutorials, 15: 2046-2069.

Zekri, M., S. El Kafhali, N. Aboutabit and Y. Saadi, 2017. DDoS attack detection using machine learning techniques in cloud computing environments. Proceedings of the 2017 3rd International Conference of Cloud Computing Technologies and Applications (CloudTech), October 24-26, 2017, IEEE, Rabat, Morocco, ISBN:978-1-5386-1116-6, pp: 1-7.