Distributed Denial of Service Attack Detection Using Wallaroo-Based Time-Series Analysis

Farzana Zakaria, Mohd Fadzil Abdul Kadir, Mohamad Afendee Mohamed, Ahmad Faisal Amri Abidin, Ahmad Nazari Mohd Rose

Abstract: ***Nowadays, with the growth of computer technologies, there had been many problems arise regarding security issues. The hackers tend to try to break into any website they desired and affect it either by modified, steal information or shutdown the server. Distributed Denial of Service (DDoS) attacks falls into one of the category of critical at-tacks. DDoS attacks can be described as temporarily deny several services of the end users. In general, it usually consumes network resources and overloads the system with undesired request. Thus, the network can be protected against such attacks using an Intrusion Detection System. This paper presents the method of detecting DDoS attacks by using the Wallaroo-based by analyzing the change of the time series data obtained from the weighted mean and weighed standard deviation of data. Wallaroo-based is about the distributed data processing framework for building high-performance streaming data applications. A streaming DDoS attack detector is constructed, which consumes a stream of request logs from a large group of servers and uses statistical anomaly detection to alert user when a server is under attack.***

Index Terms: Distributed Denial of Service, Time-series Analysis, Intrusion Detection System, Wallaroo.

I. INTRODUCTION

Recently, the issues of the security attack had been arise due to the wide used of the Internet. Many services of the organizations or companies become the victims of the attack. Many categories of attacks, those to reveal the meaning of the data which normally been fought via cryptography [1]-[3], those meant for violating the integrity of data [4], and those to disrupt the availability of the service via Denial of Service (DOS) attacks [5]. Moreover, the issues of the much advanced version namely the Distributed Denial of Service (DDoS) attacks are drastically arise around the world. In network security, DDoS attacks are the kinds of the attacks that resulted as the most harmful attacks. This project focuses on streaming DDoS attack detection, which stream of request logs is consumed from a large group of servers by using two or more terminal/console. Anomaly detection is used to alert the detector when a server is under attack. There are two things is needed in order to detect DDoS which are a model for "normal" behaviour and a way to measure the current behaviour of the system in order to compare against the "normal" behaviour.

Anomaly detection is about the process when the difference between measured behaviour and the normal behaviour is too large and the measured data is signed as anomalous. The time-series analysis algorithm is used to measure the behaviour of the data. A weighted mean and a weighted standard deviation (for each server) for the requests/second and unique-clients/second is computed. The prediction of the final-value for the current second is made when the new data point arrives. When the prediction differs from the mean by more than 2 standard deviations, the server will mark as under attack because anomalous behaviour matches an attack pattern. At least, there must be 10 requests in the prediction sample to prevent flapping in order to change from "healthy" to "under attack" since the early data can be misleading.

Fig. 1: Wallaroo-based application

The detection of DDoS attacks apply by using the Wallaroo based application as shown in Fig. 1. Wallaroo-based is about the distributed data processing framework for building high performance streaming data applications. It consists of one or more pipelines that take in data from an external data source which is send data over TCP to an internal Wallaroo source. The streams of data that is sent was decoded by using Wallaroo source, then transformed it into a stream of internal data that used for computations. The output that produced is based on the computation of the input data. Generally the data from the external system taken to perform a series of computation based on that data and produced outputs which are sent to an external system based on the modules

II. RELATED WORKS

A. DDOS Attacks Detection

Nowadays, the Distributed Denial of Service (DDoS) attacks have been one of the most dangerous security threat to the Internet community [6]-[11]. DDoS attacks work by make the targets consume many system resources and that can cause

Revised Manuscript Received on February 11, 2019.

Farzana Zakaria, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.
Mohd Fadzil Abdul Kadir, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.
Mohamad Afendee Mohamed, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.
Ahmad Faisal Amri Abidin, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.
Ahmad Nazari Mohd Rose, Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Terengganu, Malaysia.

the victims cannot work normally. There are a few steps in order to launch the DDoS attacks. Firstly, the attackers exploited the technology of the client/server and established the botnet [6]. The botnet is established by combining a group of vulnerable computers. It enables the hackers to launch the DDoS attacks for one or more targets by send the commands to the botnet. This is why the attacker is hardly track-able. Usually, the DDoS attackers gather together a huge number of victim machines in order to mount the attacks directed towards the targets. It normally causes the exhausted resources of the victim side. DDoS attacks significantly reduce the performance of the machine as well as the network and obviously the bad guys are difficult to get detected and most of the time they get away. An intrusion detection system (IDS) has been designed for this and with the sophistication of tools and technologies, it becomes one of the biggest challenges.

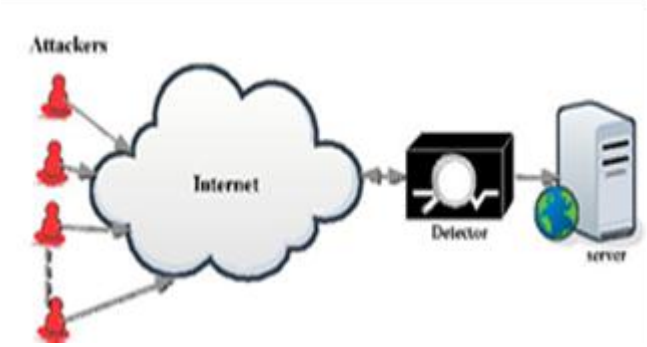

Fig. 2: DDoS detection

Since DDoS attacks involve the computer network, there are few protocols involved to launch an attack. According to the synchronization of Transfer Control Protocol (TCP) and User Datagram Protocol (UDP) behaviour, periodic sampling is conducted on every single IP for incoming and outgoing user traffic. Moreover, the traffic behaviour is decided whether it meets the synchronization or not. The vulnerabilities of network-connected systems do not exploit by the DDoS attacks [10]. But, instead they aim to disrupt victim services by overwhelming the processing capacity of system or by flooding the bandwidth of the target. Based on the research survey, the DNS flooding attacks, SYN flooding attacks and Smurf attacks are major DDoS attacks. Based on Fig. 2, there are a few types of intrusion detection which are anomaly detection and signature based detection.

B. Anomaly Based Detection

An anomaly-based detection uses the statistical approach to detect DDoS attacks. It builds the detection engine by learning from the network traffic prototype on a particular network. Based on this knowledge, this systems monitors the traffic pattern and filter the anomalies in the established real life network. The pattern representing normal activity is isolated, and any structure deviate from this is considered as intrusion in anomaly based approach. All packets are ranked by an anomaly score, in case the value is higher than preset threshold, an intrusion alert will be initiated [8]. There are many advantages for this approach. It is able to detect new unseen attacks such as DDoS attacks, virus and worm. Moreover, it is potentially easier to maintain, which is different to rule based approach because the record of signature does not need to be maintained and updated. However, the problems of this approach are the false positive and false negative, which result from non-optimal selection of threshold value. If the threshold is too low, the ratio of false positive will obviously increase, similarly the false negative increases with too high threshold value which means some attacks might have gone undetected.

C. Signature Based Detection

Signature-based detection which is also popular as rule-based works by separating existing attack signatures, or patterns, with the monitored traffic. An alarm for a potential attack is generated when a match is found. This detection technique has fairly short detection time, with an ability to detect high number of known attack. Moreover, it generally has a low false positive rate therefore only in a rare case where the system signal an alarm for legitimate traffic [6]. Signature based method extracts a few important features from a given set of activities labelled as malicious to form signature for the specific activities. If the new activity arrived, various groups of features are extracted and compared with those registered with the signature stored in the database. Any match is reported as a probable intrusion, however, this method cannot detect new intrusion which were not in its database. Misuse recognition stores a list of known signatures into its library which is always available for matching against incoming network traffic. Any new unknown signatures due to new attacks will cause 100% miss in detection. While it needs to be available at all time, it also needs to be kept up-to-date with current attacks.

D. Time-Series Analysis

The time series analysis is according on the assumption of the successive values in the data file that represent the consecutive measurements in equal time intervals. The sequence of observations is used to identify the nature of the phenomenon and hence predicting the future values of the time series variable via what is called as forecasting. The goals is, the future pattern of the observed past data is identified. The data can be interpreted and integrated among each other once the pattern is established.

According to [12], the HRPI time series is transformed into a multidimensional vector series by approximating the adaptive autoregressive (AAR) model. Then, a classifier based on trained support vector machine (SVM) is used to identify the attacks. Several databases were used during the experiments, and the results show this method to be able to successfully detect application-layer DDoS attacks [12], [13]. In application-layer DDoS attacks, attack sources have been programmed and worked according to their attack functions, so detection based on its pattern is possible [14], [15].

Based on [16], the researchers proposed a DDoS forecasting model for time-series analysis which provides assistance for the security officers during and after the DDoS attack. The model strives to predict the attack intensity rate (packet/rate) within minutes. The goal for using the forecasting model is to comprehend the future short term trend of the ongoing DDoS attacks and later able to recognize the current and similar situations of the future hence, it respond to the threat appropriately. According to

[17], there are some observable correlation between time-series of entropy values of the address and port distributions with clear and almost identical anomaly detection capabilities. The source (destination) port correlation with the source (destination) address distribution increases because of the nature of the underlying traffic patterns. However, the problem of identifying attack traffic is generally tricky due to the dynamicity of the pattern of the normal network flow which could result in those fixed thresholds to produce a high false positive rate [18], [19].

III. PROPOSED METHOD

Many offline-based intrusion detection approaches have been well studied. Offline IDS works with the stored data and check for the availability of unknown or attack patterns and the data which it holds would be classed into two categories such as training and testing data. In addition, offline models seem to have a quite disadvantage when the frequently changes of environment. The real time detection model continuously is trained based on new incoming data in order for adapts to the incoming traffic further and further. Previous research mention the anomaly detection is the best approach for this project. Initially, the system determines the model for the normal activity, and any structure that deviate from this model is reported as probable attacks. Time-series analysis uses and implement a powerful statistical and machine learning tools in order to predict future events based on past data. This method is very useful to make the prediction based on the streaming data processing. DDoS attack detection will be considered an anomaly system from which abnormal behaviour is generated and imposed on network traffic. Characterization of the network traffic with behaviour modelling could be a good indication of attack detection. The detection can be performed via abnormal behaviour identification.

It is very important to do the right configuration based on the system needs. So, Wallaroo application can be running easily. After setting up Ubuntu environment for Wallaroo, the Wallaroo application can be running which is the writing event-by-event data processing can be done. In this project, the synthetic data is used. It should produce the log file that contain a timestamp, a client IP address, the server IP address, and the resource requested. A timestamp is a sequence of characters in order to identify when the event occurs based on the date and time of the day. The data could have following characteristics:

a. 100 servers that serving 10000 resources.
b. 1000 clients making 1000 requests/second, distributed uniformly over all the servers (normal traffic).
c. 100000 clients making 25000 requests/second, with 90% of the traffic hitting only 10 servers (attack traffic).

After the data is generated, the sender command is started in the Ubuntu terminal in order to send the log file that have been produce. Then, the two separate terminal can be running in the Ubuntu to start the Wallaroo and start the listener to view the output in its own terminal. The listener is act as the DDoS detector when the sender port bind to the receiver port(listener) then able to send the packet and the Wallaroo acts as based for processing the data whether the server is under attack or not. When the listener is started, the output or the detection for the attacks can be view whether the server is 'healthy' or 'under attack' with at what time the state is changed.

IV. RESULTS AND DISCUSSION

Fig. 3 and 4 show the prediction on how the data is predicted as anomalous. The data that is measured are requested/second and unique/second. Based on the graph below, the normal behaviour which are the mean and standard deviation of the request/second and unique client/second for the healthy server is computed. The data then is compared with the expected data which are the mean and standard deviation of the current request/second and current unique client/second of the data. If the value of the expected data behaviour is less difference to normal data behaviour, then the data will not mark as the anomalous.

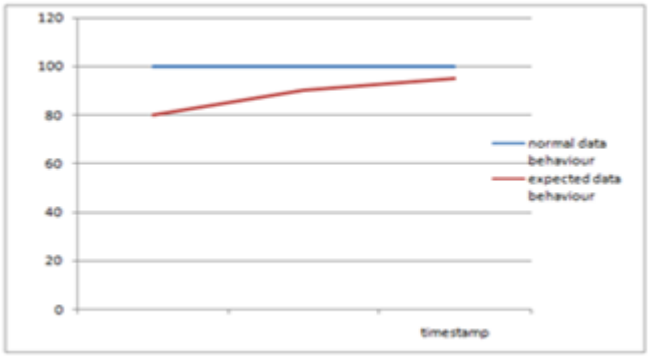

Fig. 3: No prediction to anomalous data

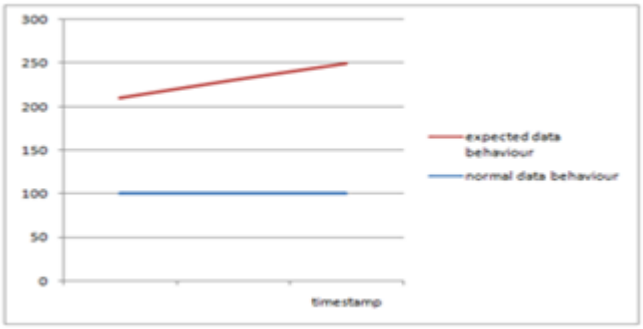

Fig. 4: Prediction to anomalous data

In statistics, the standard deviation represents the quantity of variation that exists in a given set of data values. Data points are said to be close to the mean if the standard deviation is small/low. Likewise, standard deviation with big/high value indicates that the difference between the expected value is bigger. The subtraction of the mean values from the predicted ones, and compare difference to their respective standard deviation. If the difference is greater than 2 multiples of standard deviation, it is called prediction anomalous.

The standard deviation can be calculated using formula

$$s = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}} \quad (1)$$

whereas the weighted standard deviation can be calculated using formula

$$sd_w = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x}_w)^2}{\frac{(N^{'}-1)\sum_{i=1}^{N} w_i}{N^{'}}}} \quad (2)$$

In most experimental data, effects are considered statistically significant if it falls much distant than two standard deviations away from earlier anticipated.

V. CONCLUSION

In this studies, time-series analysis was engaged for better understand and thus prediction of the future short term trend of the ongoing DDoS attack in order to offer behavior recognition capabilities of the current and future situations prior expecting an appropriate response. Usually, DDoS attacks aim to disrupt victim services instead of exploit the vulnerabilities of security of systems by over-whelming the processing capacity of system or by flooding the bandwidth of the target. For the detection method, the anomaly detection is the best approach.

ACKNOWLEDGMENT

The authors would like to thank the Government of Malaysia for funding this research under the Fundamental Research Grant Scheme (FRGS/1/2017/ICT03/UNISZA/02/1) and also the Research Management, Innovation and Commercialization of Universiti Sultan Zainal Abidin.

REFERENCES

1. R. L. Rivest, A. Shamir, and L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems,” Comm. ACM, 21(2), 1978, pp. 120-126.
2. N. Koblitz, “Elliptic curve cryptosystems,” Mathematics of Computations, 48(177), 1987, pp. 203-209.
3. M. A. Mohamed, “A survey on elliptic curve cryptography,” Applied Mathematical Sciences, 8(153-156), 2014, pp. 7665-7691.
4. L. Chi, and X. Zhu, "Hashing techniques: A survey and taxonomy," ACM Computing Surveys, 50(1), 2017, pp. 1-36.
5. T. Mahjabin, “A survey of distributed denial-of-service attack, prevention, and mitigation techniques,” International Journal of Distributed Sensor Networks, 13(12), 2017, pp. 1-33.
6. Y. Zhang, Q. Liu, and G. Zhao, "A real-time DDoS attack detection and prevention system based on per-IP traffic behavioral analysis," IEEE 3rd International Conference on Computer Science and Information Technology, 2010, pp. 163-167.
7. S. Ratnaparikhi, and A. Bhange, “DDOS attacks on network: Anomaly detection using statistical algorithm,” International Journal of Advanced Research in Computer Science and Software Engineering, 2(12), 2010, pp. 321-326.
8. W. Zhoua, W. Jia, S. Wen, Y. Xiang, and W. Zhou, “Detection and defense of application-layer DDoS attacks in backbone web traffic,” Future Generation Computer System Journal, 38, 2014, pp. 36-46.
9. M. Anjali, “Detection of DDoS attacks based on network traffic prediction and chaos theory,” International Journal of Computer Science and Information Technologies, 5(5), 2014, pp. 6502-6505.
10. M. Alenezi, and M. J. Reed, “Methodologies for detecting DoS/DDoS attacks against network servers,” 7th International Conference on Systems and Networks Communications, 2014, pp. 92-98.
11. C. Fachkha, E. B. Harb, and M. Debbabi, “Towards a forecasting model for distributed denial of service activities,” IEEE 12th International Symposium on Network Computing and Applications, 2013, pp. 110-116.
12. N. Tongguang, X. Gu, H. Wang, and Y. Li, “Real-time detection of application-layer DDoS attack using time series analysis,” Journal of Control Science and Engineering, 2013, 2013, pp. 1-6.
13. H. Liu, and M. S. Kim, “Real-time detection of stealthy DDoS attacks using time-series decomposition,” IEEE International Conference on Communications, 2010, pp. 1-6.
14. W. Q. Tao, and S. Z. Qing, “Detecting DDoS attacks against web server using time series analysis,” Wuhan University Journal of Nature Sciences, 11(1), 2006, pp. 165-180.
15. R. Fouladi, C. Kayatas, and E. Anarim,”Statistical measures: Promising features for time series based DDoS attack detection,” International Workshop on Computational Intelligence for Multimedia Understanding, 2018, pp. 1-9.
16. T. G. Ni, X. Q. Gu, and H. Y. Wang, “Detecting DDoS attacks against DNS servers using time series analysis,” TELKOMNIKA Indonesian Journal of Electrical Engineering, 12(1), 2014, pp. 753-761.
17. L. Li, and G. Lee, “DDoS attack detection and wavelets,” Telecommunication Systems, 28(3-4), 2005, pp. 435-451.
18. R. Karimazad, and A. Faraahi, “An anomaly-based method for DDoS attacks detection using RBF neural networks,” International Conference on Network and Electronics Engineering, 2011, pp. 44-48.
19. S. Kumarasamy, and R. Asokan, “Distributed Denial of Service (DDoS) attacks detection mechanism,” International Journal of Computer Science, Engineering and Information Technology, 1(5), 2011, pp. 39-49.