Volume 2, Issue 12, December 2012 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

DDOS Attacks on Network; Anomaly Detection using Statistical Algorithm

[1]Saurabh Ratnaparikhi
[1]M.tech Scholar, Department of CSE
[1] CMJ,Universsity
Shilong

[2]Anup Bhange
[2] Asst Prof, Department of CSE
[2]KDK College of Engineering
Nagpur

Abstract: ***Distributed denial-of-service (DDoS) flood attack remains great threats to the Internet. This kind of attack consumes a large amount of network bandwidth or occupies network equipment resources by flooding them with packets from the machines distributed all over the world. To ensure the network usability and reliability, real-time and accurate detection of these attacks is critical. Flaws either in users' implementation of a network or in the standard specification of protocols has resulted in gaps that allow various kinds of network attack to be launched. Of the kinds of network attacks, denial-of-service flood attacks have caused the most severe impact. DoS/DDoS attacks are a virulent, relatively new type of Internet attacks, they have caused some biggest web sites on the world owned by the most famous E-Commerce companies such as Yahoo, eBay, Amazon became inaccessible to customers, partners, and users, the financial losses are very huge. To this end, our discussion on statistical evidence that the marginal distribution of real traffic is adequately modeled with alpha stable functions.***

Keywords: ***network security, statistical approach, α-Stable Model***

I. INTRODUCTION

Distributed denial-of-service (DDoS) attack has been one of the most regularly in the works attacks that badly intimidate the stability of the Internet. According to CERT Coordination Center (CERT/CC)[1], there are mainly three categories of DDoS attacks: flood attack, protocol attack and logical attack. This paper mainly focuses on flood and flash-crowds attack. In the DDoS flood attack, an intruder missiles attack packets upon a site (sufferer) with a huge amount of traffic so as to actually jam its entrance and block access by rightful users or appreciably demean its performance [2]. Therefore, a real-time and accurate detection of these attacks is critical to the Internet community usually, the attack recognition methods are classified into two categories. One is misuse recognition and the other is anomaly recognition. Misuse recognition is based on a library of known signatures to match against network traffic. Hence, unknown signatures from new variation of an attack mean 100% miss. Anomaly recognition does not suffer from this problem. Considering that DDoS flood attack is a process changing dynamically and frequently, Anomaly- based detectors play a key role in identify this kind of attack several studies show that DDoS flood attack can exert remarkable influence on the self-similarity of network traffic. Thus, this kind of attack can be effectively identified by monitoring the change of the Hurst parameter [3-4]. Existing flood attack recognition methods based on the self-similarity nature of network traffic divide the network traffic into non-overlapping segments. The Hurst parameter of each segment is estimated, once the Hurst parameter changes beyond a pre-defined fixed threshold, the loss of self-similarity (Loss) occurs and the DDoS flood attack is detected.

Network anomaly recognition is statistic based approach, and does not require prior knowledge about the signature of attacks. It operates by structure a statistic model of normal network traffic in learning phase, and informs this model in every discrete time interval. Any inconsistent deviation from the built model is considered as anomaly. While a wide range of anomaly recognition algorithms have been proposed to undermine attacks, the effectiveness of these models is largely depending on assumptions about the underlying traffic distributions parameters, and the built models to fit network traffic [5], [6]. They lack the capability of handling shape irregularities and unpredictable large fluctuations in real IP traffic.

When some of anomaly recognition algorithms are host based and not scalable for high speed network, most of existing network anomaly recognition algorithms are engaged in the detection of these anomalies as soon as possible. They are applied on the overall traffic, i.e. the whole stream of packets is aggregated in one flow, and a signal processing

algorithm is applied over the analyzed time-series (number of UDP packets, SYN packets, etc.). These algorithms only raise an alarm when there is a deviation larger than a predefined threshold in time-series analysis.

The purpose of change point recognition algorithms to the overall traffic tend to be inaccurate in verdict attacks, and do not reveal any information about attacker/victim for mitigation. Intuitively, if too many flows are aggregated, only the bigger anomalies will be visible. Furthermore, the identification task of victim or anomaly type without any information to separate malicious traffic from normal traffic is like searching for a needle in a haystack.

In this paper, Our Discussion on a DDoS flood attack recognition method using discrete wavelet transform (DWT) and Schwarz information criterion (SIC) to verify the change point of self-similarity. The SIC [7] statistic is based on the maximum probability function for the model, and can be easily applied to change point detection by comparing the likelihood of the null hypothesis against the alternative hypothesis (i.e., a change is present). This paper presents the SIC algorithm working with the DWT to detect the change point of self-similarity in real-time. After the change point detection, To Recognize the attack fuzzy logic to adaptively decide the intensity of the DDoS flood attack. Also discussed a set of decision rules for the fuzzy logic to determine the strength of the DDoS flood attack. As a result, this proposed attack detection method can accurately detect not only the moment when the flood attack happens, but also the intensity of the attack.

II RELATED WORK.

Some anomaly recognition methods have been projected against DDoS flood attack in the literature [16-18]. In these methods, the network traffic activity is confine and then a profile instead of its stochastic behavior is created. This profile is mainly support on metrics such as the network traffic rate, the number of packets or bytes for each protocol, the rate of connections, the number of different IP addresses, etc. Any activity that deviates from the profile is treated as a possible attack. Much important assistance has been proposed to challenge anomaly in network traffic [10], [11], [12], [13],. When early approaches for anomaly recognition were listening carefully in the definition of models able to represent the traffic pattern, other advanced work aggregates the whole stream of packets in one flow, and apply change point detection algorithm to detect anomaly occurrence instant. There is a serious problem with statistical anomaly recognition methods. That is, it is hard to choose the suitable metric on the global scale, because the linear superposition of these micro-based recognition methods cannot cope with the complex behavior of entire network. The research done by Li[20] first mathematically proved that there is a statistically significant change in the average Hurst parameter under DDoS flood attack. Allen [15] et al. and W.Schleifer [14] et al. proposed a method using Hurst parameter to categorize attack, which causes a decrease in the traffic's self-similarity. Those methods consider the normal range of Hurst parameter to be [0.5, 0.99], and there is an attack when the Hurst parameter runs out of this range. The cut down of normal range of Hurst parameter can be more efficient in identify the low-rate DDoS flood attack. Nevertheless, all of these existing detection methods can only detect the presence of attack after the attack occurs, they cannot identify at what time the attack happened.

Fuzzy logic is one of the mainly popular techniques used in attack detection for it can deal with the vague and imprecise boundaries between usual traffic and different levels of attacks [10]. Wang [17] et al. proposed to use the fuzzy logic to examine the Hurst parameter and guess the time duration of DDoS attack. However, the work in [18] didn't consider the intensity of the attack traffic compared with the background traffic, therefore cannot accurately reflect the level of damage that is caused by the attack.

A. ANOMALY DETECTION:

The two mainly common techniques to detecting web-based attacks are signature-based detection and anomaly based detection. Signature-based detection relies on detecting patterns of known attacks to recognize malicious behavior. While they are correct, they have to be kept up-to-date with current attacks to be active. Any attacks that are not in the signature or pattern database will therefore not be detected. This weakness can be oppressed by creating diverse versions of a single attack.

Anomaly-based detection relies on statistical analysis of the data to and performance that deviates from the normal activity. One of the big advantages over signature based attacks, if used correctly, is that it is able to detect difference of attacks or even totally new attacks. However, this could also result in normal movement being standard as malicious.

B. Anomaly Detection of Web-based Attacks:

The anomaly detection explain by Kruegel and Vigna[19] works on personality requests. The center is largely on the detection of diverse data input related attacks, by analyzing various aspects of the request path of each request. The URI connected with each request (minus the domain name) is separated into three parts. The path, which consists of the resource path and program, and the limitation and their values. A program in this context, also called a resource, is dined by the last part of the path in the URI before the parameters start. Only HTTP GET requests that generated a response code by the web server indicating success1 were used. This dataset is further reduced by eliminating any requests that do not contain any query parameters.

C. Spectrogram:

Song et al.[20] describe a system, which is parallel to examine individual HTTP requests, but operating on a lower level. The major difference is that both HTTP GET and HTTP POST requirements are examine and the entire request path including query parameters are treated as a single object. For a POST request, the request body containing the POST data is also used. It uses a grouping of n-grams and Markov chains to calculate an anomaly score for this particular request. The given string is scanned and probabilities are considered for the succession of characters that occur in this string. It uses anticipation Maximization to and the optimal settings given the gram-size and the number of Markov chains to use in the training phase.

The Spectrogram system was tested using real data from two university web servers which was collected over a period of a month. These servers contained various scripts for the computer science department and personal homepages of students. Both of which can be interesting targets for attackers. Normalization is performed on the collected data by un-escaping strings, removing whitespaces and numbers and converting all characters to lowercase. A manual inspection of the data ensures that the dataset does not contain attacks of any kind. Finally all duplicate requests are removed to prevent creating a bias towards requests that occur more often than others. The resulting dataset was then used to train the model.

The attack-data includes remote le inclusion attacks, JavaScript and XSS, attacks, SQL injection and many unique shell code samples. The results were overall pretty good, with exceptional results in detecting worms, shell code attacks, SQL and XSS attacks.

D. Detecting Anomalous and Unknown Intrusion against Pro-grams:

Employing a neural network to detect malicious activity is proposed by Ghosh et al.[21]. A back propagation network is created which consists of a variable number of input nodes, ranging from 8 to 83, a single hidden layer with 125 nodes and one output node indicating positive or negative for the given input. The input dataset expected is a single string of data, in the case of this paper the input data to a printing program. given the similarities in input data. Like previous systems, the neural network has to be trained prior to usage.

Experiments are performed in two different situations:

-> For the black-box experiments, the authors use only data passed to the program, without having Access to the programs source or state.

-> For the white-box experiments, in addition to data used in the black-box experiments, they use internal program state data, which is only available when having access to the program source-code.

E. Flow based intrusion detection

All the approaches taken so far rely on the availability of detailed information inside a single request or network packet. Situations with limited amounts of information in a request or where most of the traffic is encrypted will not provide the data these algorithms require. Sperotto [22] focuses on network intrusion detection as opposed to web-based intrusion detection, by looking at SSH and DNS data. Since SSH traffic is encrypted, it is not possible as an observer to detect anomalous behavior by looking at the payload. The observed packets within a time frame are grouped together based on properties they might have in common, such as IP addresses, ports and protocol to form a flow. These flows have certain properties of their own, regardless of the payload contents of individual packets, including flows per second, packets per second, bytes per second and number of packets in a flow. In this case, the flows per second measurements are used to classify flows as benign or malicious.

A model consisting of two states is constructed based on Markov Chains. The two states indicate either activity, where SSH traffic was observer, or inactivity. The dataset consists of real traffic collected from the University of Twenty network. Only benign traffic is used to train the model. Based on this trained model, threshold values can be assigned to traffic flows. Classification of the flows is done based on these values, where flows exceeding a certain threshold are marked as malicious.

After training the model, two synthetic and two original data sets are used for testing. The original data is network traffic captured from the University of Twenty network. Each of these data sets contains both malicious and normal traffic; the malicious data is manually labeled for the datasets containing real network traffic. The results varied between the synthetic and original data sets, where the results were significantly better for the synthetic data sets. As previously mentioned, there is always a trade of between a good detection rate and a low false positive rate.

III ANOMALY DETECTION METHOD

A. Statistical Approach

In statistical-based approach it concludes normal network action and then all traffic that deviates from the normal is marked as anomalous. This approach is used to learn network traffic prototype on a particular network. By examine network traffic and processing the information with complex statistical algorithms, this systems look for anomalies in the established normal network traffic patterns. All packets are given an anomaly score and if the anomaly score is higher than a certain threshold, the intrusion recognition system will generate an alert.

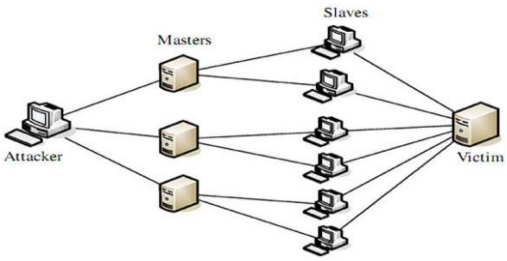


Fig. 1: Distributed denial of service attack

This approach has a number of advantages. It is capable of detecting new unseen attacks like denial of services attacks, worm or virus. It is also capable of detecting low intensity slow pace attacks. Another major benefit of this approach is that it is potentially easier to maintain than a rule based approach because we do not need to maintain and update any record of signature. The basic problem with this type of approach is the selection of appropriate threshold value. Problem of false positive and false negative occurs due to this value. If value is set low than ratio of false positive increase if value is set too high than the anomalous activities cannot be verify means false negative increases.

IV. NETWORK TRAFFIC MODELS

Conventionally, network traffic has been modeled as a Poisson process. Indeed, the Poisson model has been successfully used in telephone networks for many years, and so it was inherited when telecommunication networks became digital and started to send information as data packets. Also, this model has a simple mathematical expression [23], and has only one parameter, λ, which is in turn very intuitive (the mean traffic in packets per time unit). Several authors have proposed network traffic behavior and presented other models that overcome the limitations which are inherent to Poisson processes, the most notable ones probably being that the Poisson model has a fixed relationship between mean and variance values (both are equal to λ), and that it does not account for high variability or long-range dependence. Some proposed models are usually based on the assumption that network traffic is self-similar in nature, as originally stated in [23]. At this point, it should be clear that any model for instantaneous traffic marginal's must be flexible enough to adapt to some properties observed in traffic, namely:

1. Let C(t) be the amount of traffic accumulated at time t. Then, C(t) <=C(t+1) and C(t+1)-C(t)<=M, where M is the network maximum transmission rate.
2. The fact that at time t there is a certain amount of traffic C(t) does not imply in any way that at time t +1 the amount of traffic lies anywhere near C(t). This is equivalent to say network traffic exhibits the high variability property.

The latter property is also identified as the "Noah effect" or the infinite variance syndrome [24].

At the other side, the first aforementioned property states the clear fact that network traffic has compact support between 0 and the M. Compact support creates symmetric distributions (Gaussian distributions are symmetric) inappropriate. Accordingly, if traffic data concentrate near the maximum transmission rate, a symmetric model would allow traffic increments to be larger than physically possible, again, with a non-negligible probability. This also influences the Gamma distribution.

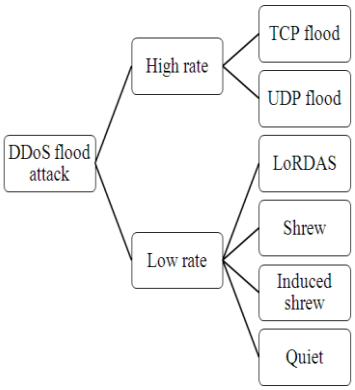
A. α-STABLE MODEL

α-stable distributions can be considered as a superset of Gaussian functions and originate as the solution to the Central Limit Theorem when second order moments do not present [25], that is, when data can suddenly vary by large amounts as time passes by. This fits nicely to the high variability property seen in network traffic. Moreover, α-stable distributions have an asymmetry parameter which allows their PDF to change from totally left-asymmetric to totally right-asymmetric, while genuine Gaussian distributions are always symmetric. This factor makes α-stable distributions naturally adaptable to the first traffic property (compact support) even when average traffic is virtually 0 or very near the maximum theoretical network throughput. In addition, α-stable distributions give an explanation to the restriction imposed in [23] about the requirement to aggregate many traffic traces for them to converge to a Gaussian distribution. According to the Generalized Central Limit Theorem [26], which includes the infinite variance case, the sum of n α-stable distributions is another α-stable distribution, although not necessarily Gaussian.

V. CLASSIFICATION of DDoS FLOOD ATTACK

Let { , } *i X x i* □ and { , } *i Y y i* □ be normal and abnormal traffic respectively, and { , } *i Z z i* □ be the attack traffic during transition process of attacking. *X* and *Z* are correlated so *Y* can be abstractly expressed by *Y=X+Z*. Figure 2 illustrates the components of normal and abnormal traffic. *xi*(*p*) represents the number of bytes sent out by node *p* at time

i for normal network services, and *zi*(*q*) represents the number of bytes sent out by node *q* at time *i* for network attack, and *yi* is the total traffic the target received at time *i*.



In the context of in order systems, a DoS attack happens when an attacker explicitly attempts to protect a service from being used by its legitimate users through many ways including by flooding a network With useless traffic to prevent legitimate network traffic.

VI. CONCLUSION

This paper has presented idea about the statistical anomaly recognition of network traffic. Here paper studied a statistical approach to analysis the distribution of network traffic to recognize the normal network traffic behavior. This Paper has also discussed flooding attacks. Most of the flood attacks reviewed in this study is the new type of flood attacks which are more furtive yet cause more severe impacts of denial of service, such as those attacks categorized under the low-rate DoS attacks. This paper also discussed a method to recognize anomalies in network traffic, based on a α-stable model and statistical hypothesis testing.

REFERENCES

[1] http://www.cert.or

[2] M. Li. An approach to reliably identifying signs of DDOS flood attacks based on LRD traffic pattern recognition. *Computers & Security*, 23(7): 549-558, 2004.

[3] C.S. Sastry, S. Rawat and A.K. Pujari. Network traffic analysis using singular value decomposition and multiscale transforms. *Information Sciences*, 177(23): 5275-5291, 2007.

[4] M.F. Rohani, M.A. Maarof and A. Selamat. Continuous LoSS detection using iterative window based on SOSS model and MLS approach. *In Proceedings of the International Conference on Computer and Communication Engineering*, Kuala Lumpur, Malaysia, May 2000

[5] H. Hajji. Statistical analysis of network traffic for adaptive faults detection. *IEEE Transactions on Neural Networks*, 16(5):1053–1063, September 2005.

[6] J. D. Brutlag. Aberrant behavior detection in time series for network monitoring. In *LISA '00: Proceedings of the 14th USENIX conference on System administration*, pages 139–146, Berkeley, CA, USA, 2000.

[7] D. Rincón and S. Sallent. On-line segmentation of non-stationary fractal network traffic with wavelet transforms and Log-likelihood-based statistics. *LNCS*, 3375: 110-123, 2005

[8] C. Douligeris and A. Mitrokotsa. DDoS attacks and defense mechanisms: classification and stateof- the-art. *Computer Networks*, 44(5): 643-666, 2004.

[9] P. García-Teodoro, J. Díaz-Verdejo and G. Maciá- Fernández. Anomaly-based network intrusion detection: techniques, systems and challenges. *Computers & Security*, 28(1-2): 18-28, 2009.

[10] V. A. SIRIS and F. PAPAGALOU. Application of anomaly detection algorithms for detecting syn flooding attacks. In *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '04)*, volume 4, pages 2050–2054, Dallas, USA, 2004

[11] H. Wang, D. Zhang, and K. G. Shin. Syn-dog: Sniffing syn flooding sources. In *Proceedings of the 22th International Conference on Distributed Computing Systems (ICDCS'02)*, pages 421–429, Washington, DC, USA, 2002. IEEE Computer Society.

[12] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming (ICALP '02)*, pages 693–703, London, UK, 2002. Springer-Verlag.

[13] J. D. Brutlag. Aberrant behavior detection in time series for network monitoring. In *LISA '00: Proceedings of the 14th USENIX conference on System administration*, pages 139–146, Berkeley, CA, USA, 200

[14] V. Paxson. Bro: A System for Detecting Network Intruders in Real- Time. In *Computer Networks*, volume 31 (23–24), pages 2435–2463,1999.

[15] E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.
[16] S.X. Wu and W. Banzhaf. The use of computational intelligence in intrusion detection systems: A review. *Applied Soft Computing*, 10: 1- 35, 2010.
[17] O. Salem, S. Vaton, and A. Gravey. An efficient online anomalies detection mechanism for high-speed networks. In *IEEE Workshop on Monitoring, Attack Detection and Migitation (MonAM 2007)*, November 2007.
[18] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE/ACM Trans. Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994
[19] Christopher Kruegel and Giovanni Vigna. Anomaly detection of web-based attacks. In Proceedings of the 10th ACM conference on Computer and communications security, CCS '03, pages 251{261, New York, NY, USA, 2003. ACM.
[20] Yingbo Song, Angelos D. Keromytis, and Salvatore J. Stolfo. Spectrogram: A mixture-of-markov-chains model for anomaly detection in web traffic.
[21] Anup K. Ghosh, James Wanken, and Frank Charron. Detecting anomalous and unknown intru-sions against programs. In Proceedings of the Annual Computer Security Application Conference(ACSAC98), pages 259{267, 1998
[22] Anna Sperotto. Flow-Based Intrusion Detection. W□ohrmann Print Service, 2010.
[23] A. Papoulis, Probability, Random Variables, and Stochastic Processes, third ed., McGraw-Hill, 1991
[24] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-Similar Nature of Ethernet Traffic (Extended Version)," IEEE/ACM Trans. Networking, vol. 2, no. 1, pp. 1-15, Feb. 1994
[25] O. Salem, S. Vaton, and A. Gravey. An efficient online anomalies detection mechanism for high-speed networks. In IEEE Workshop on Monitoring, Attack Detection and Migitation (MonAM 2007), November 2007
[26] G.R. Arce, Nonlinear Signal Processing: A Statistical Approach. John Wiley and Sons, 2005.

