

DoS attack discovery framework using principle of anomaly based detection in attack recognition.

Md Gufran Khan

M.Tech Student

Department of CSE,

St.Peter's Engineering College,

Hyderabad, TS, INDIA.

Mr. Anjaiah Adepu

Assistant Professor

Department of CSE,

St.Peter's Engineering College,

Hyderabad, TS, INDIA.

K Rajesh

Assistant Professor

Department of CSE,

St.Peter's Engineering College,

Hyderabad, TS, INDIA.

Abstract: *In computing, a denial-of-service (DoS) attack is an attempt to make a machine or network resource unavailable to its intended users, such as to temporarily or indefinitely interrupt or suspend services of a host connected to the Internet. A distributed denial-of-service (DDoS) is where the attack source is more than one—and often thousands—of unique IP addresses. Criminal perpetrators of DoS attacks often target sites or services hosted on high-profile web servers such as banks, credit card payment gateways; but motives of revenge, blackmail or activism can be behind other attacks. In this paper, we present a DoS attack detection system that uses Multivariate Correlation Analysis (MCA) for accurate network traffic characterization by extracting the geometrical correlations between network traffic features. Our MCA-based DoS attack detection system employs the principle of anomaly-based detection in attack recognition. This makes our solution capable of detecting known and unknown DoS attacks effectively by learning the patterns of legitimate network traffic only. Furthermore, a triangle-area-based technique is proposed to enhance and to speed up the process of MCA. The effectiveness of our proposed detection system is evaluated using KDD Cup 99 dataset, and the influences of both non-normalized data and normalized data on the performance of the proposed detection system are examined. The results show that our system outperforms two other previously developed state-of-the-art approaches in terms of detection accuracy.*

Key Words: *DoS attacks, network traffic characterization, multivariate correlations, triangle area.*

Introduction:

Denial-of-Service (DoS) attacks are one type of aggressive and menacing intrusive behavior to online servers. DoS attacks severely degrade the availability of a victim, which can be a host, a router, or an entire network. They impose intensive computation tasks to the victim by exploiting its system vulnerability or flooding it with huge amount of useless packets. The victim can be forced out of service from a few minutes to even several days. This causes serious damages to the services running on the victim. Therefore, effective detection of DoS attacks is essential to the protection of online services. Work on DoS attack detection mainly focuses on the development of network-based detection mechanisms. Detection systems based on these mechanisms monitor traffic transmitting over the protected networks.

These mechanisms release the protected online servers from monitoring attacks and ensure that the servers can dedicate themselves to provide quality services with minimum delay in response. Moreover, network-based detection systems are loosely coupled with operating systems running on the host machines which they are protecting. As a result, the configurations of networkbased detection systems are less complicated than that of host-based detection systems.

Generally, network-based detection systems can be classified into two main categories, namely misuse based detection systems [1] and anomaly-based

detection systems [2]. Misuse-based detection systems detect attacks by monitoring network activities and looking for matches with the existing attack signatures. In spite of having high detection rates to known attacks and low false positive rates, misuse-based detection systems are easily evaded by any new attacks and even variants of the existing attacks. Furthermore, it is a complicated and labor intensive task to keep signature database updated because signature generation is a manual process and heavily involves network security expertise.

Existing System:

Generally, network-based detection systems can be classified into two main categories, namely misusebased detection systems [1] and anomaly-based detection systems [2]. Misuse-based detection systems detect attacks by monitoring network activities and looking for matches with the existing attack signatures. In spite of having high detection rates to known attacks and low false positive rates, misuse-based detection systems are easily evaded by any new attacks and even variants of the existing attacks. Furthermore, it is a complicated and labor intensive task to keep signature database updated because signature generation is a manual process and heavily involves network security expertise.

Disadvantages:

- i. Most existing IDS are optimized to detect attacks with high accuracy. However, they still have various disadvantages that have been outlined in a number of publications and a lot of work has been done to analyze IDS in order to direct future research.
- ii. Besides others, one drawback is the large amount of alerts produced.

Proposed System:

In this paper, we present a DoS attack detection system that uses Multivariate Correlation Analysis (MCA) for accurate network traffic characterization by extracting the geometrical correlations between network traffic features. Our MCA-based DoS attack detection system employs the principle of anomaly-based detection in attack recognition. The DoS attack detection system

presented in this paper employs the principles of MCA and anomaly-based detection. They equip our detection system with capabilities of accurate characterization for traffic behaviors and detection of known and unknown attacks respectively. A triangle area technique is developed to enhance and to speed up the process of MCA. A statistical normalization technique is used to eliminate the bias from the raw data.

Advantages:

- i. More detection accuracy
- ii. Less false alarm
- iii. Accurate characterization for traffic behaviors and detection of known and unknown attacks respectively.

Related Work:

Research community, therefore, started to explore a way to achieve novelty-tolerant detection systems and developed a more advanced concept, namely anomalybased detection. Owing to the principle of detection, which monitors and flags any network activities presenting significant deviation from legitimate traffic profiles as suspicious objects, anomaly-based detection techniques show more promising in detecting zero-day intrusions that exploit previous unknown system vulnerabilities[3]. Moreover, it is not constrained by the expertise in network security, due to the fact that the profiles of legitimate behaviors are developed based on techniques, such as data mining [4], [5], machine learning [6], [7] and statistical analysis [8], [9]. However, these proposed systems commonly suffer from high false positive rates because the correlations between features/attributes are intrinsically neglected [10] or the techniques do not manage to fully exploit these correlations.

Recent studies have focused on feature correlation analysis. Yu et al. [11] proposed an algorithm to discriminate DDoS attacks from flash crowds by analyzing the flow correlation coefficient among suspicious flows.

A covariance matrix based approach was designed in [12] to mine the multivariate correlation for sequential samples. Although the approach improves detection accuracy, it is vulnerable to attacks that linearly change all monitored features. In addition, this approach can only label an entire group of observed samples as legitimate or attack traffic but not the individuals in the group.

To deal with the above problems, an approach based on triangle area was presented in [13] to generate better discriminative features. However, this approach has dependency on prior knowledge of malicious behaviors. More recently, Jamdagni et al. [14] developed a refined geometrical structure based analysis technique, where Mahalanobis distance was used to extract the correlations between the selected packet payload features. This approach also successfully avoids the above problems, but it works with network packet payloads.

In [15], Tan et al. proposed a more sophisticated non-payloadbased DoS detection approach using Multivariate Correlation Analysis (MCA). Following this emerging idea, we present a new MCA-based detection system to protect online services against DoS attacks in this paper, which is built upon our previous work in [16]. In addition to the work shown in [16], we present the following contributions in this paper. First, we develop a complete framework for our proposed DoS attack detection system in Section 2.1. Second, we propose an algorithm for normal profile generation and an algorithm for attack detection in Sections 4.1 and 4.3 respectively. Third, we proceed a detailed and complete mathematical analysis of the proposed system and investigate further on time cost in Section 6. As resources of interconnected systems (such as Web servers, database servers, cloud computing servers etc.) are located in service providers' Local Area Networks that are commonly constructed using the same or alike network underlying infrastructure and are compliant with the underlying network model, our proposed detection system can provide effective protection to all of these systems by considering their commonality.

The overview of our proposed DoS attack detection system architecture is given in this section, where the system framework and the sample-by-sample detection mechanism are discussed.

Modules:

The project consists of four models they are as follows

1. Feature Normalization
2. Multivariate Correlation Analysis
3. Decision Making Module
4. Evaluation of Attack detection

Modules Description:

1. Feature Normalization Module:

In this module, basic features are generated from ingress network traffic to the internal network where protected servers reside in and are used to form traffic records for a well-defined time interval. Monitoring and analyzing at the destination network reduce the overhead of detecting malicious activities by concentrating only on relevant inbound traffic. This also enables our detector to provide protection which is the best fit for the targeted internal network because legitimate traffic profiles used by the detectors are developed for a smaller number of network services.

2. Multivariate Correlation Analysis:

In this Multivariate Correlation Analysis, in which the "Triangle Area Map Generation" module is applied to extract the correlations between two distinct features within each traffic record coming from the first step or the traffic record normalized by the "Feature Normalization" module in this step. The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify the intrusive activities. All the extracted correlations, namely triangle areas stored in Triangle Area Maps (TAMs), are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher

discriminative information to differentiate between legitimate and illegitimate traffic records.

3. Decision Making Module:

In this module, the anomaly-based detection mechanism is adopted in Decision Making. It facilitates the detection of any DoS attacks without requiring any attack relevant knowledge. Furthermore, the labor-intensive attack analysis and the frequent update of the attack signature database in the case of misuse-based detection are avoided. Meanwhile, the mechanism enhances the robustness of the proposed detectors and makes them harder to be evaded because attackers need to generate attacks that match the normal traffic profiles built by a specific detection algorithm. This, however, is a labor-intensive task and requires expertise in the targeted detection algorithm. Specifically, two phases (i.e., the “Training Phase” and the “Test Phase”) are involved in Decision Making.

The “Normal Profile Generation” module is operated in the “Training Phase” to generate profiles for various types of legitimate traffic records, and the generated normal profiles are stored in a database. The “Tested Profile Generation” module is used in the “Test Phase” to build profiles for individual observed traffic records. Then, the tested profiles are handed over to the “Attack Detection” module, which compares the individual tested profiles with the respective stored normal profiles. A threshold-based classifier is employed in the “Attack Detection” module to distinguish DoS attacks from legitimate traffic.

4. Evaluation of Attack detection

During the evaluation, the 10 percent labeled data of KDD Cup 99 dataset is used, where three types of legitimate traffic (TCP, UDP and ICMP traffic) and six different types of DoS attacks (Teardrop, Smurf, Pod, Neptune, Land and Back attacks) are available. All of these records are first filtered and then are further grouped into seven clusters according to their labels. We show the evaluation results in graph.

Framework

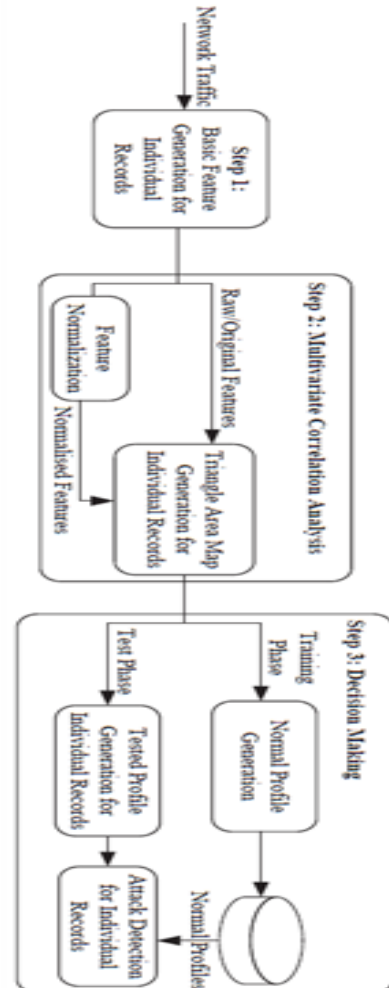


Fig No 1: System Architecture

The whole detection process consists of three major steps as shown in Fig. 1. The sample-by-sample detection mechanism is involved in the whole detection phase (i.e., Steps 1, 2 and 3). In Step 1, basic features are generated from ingress network traffic to the internal network where protected servers reside in and are used to form traffic records for a well-defined time interval. Monitoring and analyzing at the destination network reduce the overhead of detecting malicious activities by concentrating only on relevant inbound traffic. This also enables our detector to provide protection which is the best fit for the targeted internal network because legitimate traffic profiles used by the detectors are developed for a smaller number of network services. The detailed process can be found in [17].

Step 2 is Multivariate Correlation Analysis, in which the “Triangle Area Map Generation” module is applied to extract the correlations between two distinct features within each traffic record coming from the first step or the traffic record normalized by the “Feature Normalization” module in this step (Step 2). The occurrence of network intrusions cause changes to these correlations so that the changes can be used as indicators to identify the intrusive activities. All the extracted correlations, namely triangle areas stored in Triangle Area Maps (TAMs), are then used to replace the original basic features or the normalized features to represent the traffic records. This provides higher discriminative information to differentiate between legitimate and illegitimate traffic records.

In Step 3, the anomaly-based detection mechanism is adopted in Decision Making. It facilitates the detection of any DoS attacks without requiring any attack relevant knowledge. The labor-intensive attack analysis and the frequent update of the attack signature database in the case of misuse-based detection are avoided. Meanwhile, the mechanism enhances the robustness of the proposed detectors and makes them harder to be evaded because attackers need to generate attacks that match the normal traffic profiles built by a specific detection algorithm. This, however, is a labor-intensive task and requires expertise in the targeted detection algorithm.

Sample-by-sample Detection

Jin et al. [12] systematically proved that the group-based detection mechanism maintained a higher probability in classifying a group of sequential network traffic samples than the sample-by-sample detection mechanism.

Whereas, the proof was based on an assumption that the samples in a tested group were all from the same distribution (class). This restricts the applications of the group-based detection to limited scenarios, because attacks occur unpredictably in general and it is difficult to obtain a group of sequential samples only from the same distribution. To remove this restriction, our

system in this paper investigates traffic samples individually. This offers benefits that are not found in the group-based detection mechanism. For example, (a) attacks can be detected in a prompt manner in comparison with the group-based detection mechanism, (b) intrusive traffic samples can be labeled individually, and (c) the probability of correctly classifying a sample into its population is higher than the one achieved using the group-based detection mechanism in a general network scenario.

Conclusion:

This paper has presented a MCA-based DoS attack detection system which is powered by the triangle-area-based MCA technique and the anomaly-based detection technique. The former technique extracts the geometrical correlations hidden in individual pairs of two distinct features within each network traffic record, and offers more accurate characterization for network traffic behaviors. The latter technique facilitates our system to be able to distinguish both known and unknown DoS attacks from legitimate network traffic. Evaluation has been conducted using KDD Cup 99 dataset to verify the effectiveness and performance of the proposed DoS attack detection system. The influence of original (non-normalized) and normalized data has been studied in the paper. The results have revealed that when working with non-normalized data, our detection system achieves maximum 95.20% detection accuracy although it does not work well in identifying Land, Neptune and Teardrop attack records. The problem, however, can be solved by utilizing statistical normalization technique to eliminate the bias from the data. The results of evaluating with the normalized data have shown a more encouraging detection accuracy of 99.95% and nearly 100.00% DRs for the various DoS attacks. Besides, the comparison result has proven that our detection system outperforms two state-of-the-art approaches in terms of detection accuracy. Moreover, the computational complexity and the time cost of the proposed detection system have been analyzed and shown in Section 6. The proposed system achieves equal or better performance in comparison with the two state-of-the-art approaches. To be part of the

future work, we will further test our DoS attack detection system using real world data and employ more sophisticated classification techniques to further alleviate the false positive rate.

References:

- [1] Zhiyuan Tan, Aruna Jamdagni, Xiangjian He[‡], Senior Member, IEEE, Priyadarsi Nanda, Member, IEEE, and Ren Ping Liu, Member, IEEE, "A System for Denial-of-Service Attack Detection Based on Multivariate Correlation Analysis", IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. , NO. , 2013.
- [2] P. Garca-Teodoro, J. Daz-Verdejo, G. Maci-Fernndez, and E. Vzquez, "Anomaly-based Network Intrusion Detection: Techniques, Systems and Challenges," *Computers & Security*, vol. 28, pp. 18-28, 2009.
- [3] D. E. Denning, "An Intrusion-detection Model," *IEEE Transactions on Software Engineering*, pp. 222-232, 1987.
- [4] K. Lee, J. Kim, K. H. Kwon, Y. Han, and S. Kim, "DDoS attack detection method using cluster analysis," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1659-1665, 2008.
- [5] A. Tajbakhsh, M. Rahmati, and A. Mirzaei, "Intrusion detection using fuzzy association rules," *Applied Soft Computing*, vol. 9, no. 2, pp. 462-469, 2009.
- [6] J. Yu, H. Lee, M.-S. Kim, and D. Park, "Traffic flooding attack detection with SNMP MIB using SVM," *Computer Communications*, vol. 31, no. 17, pp. 4212-4219, 2008.
- [7] W. Hu, W. Hu, and S. Maybank, "AdaBoost-Based Algorithm for Network Intrusion Detection," *Trans. Sys. Man Cyber. Part B*, vol. 38, no. 2, pp. 577-583, 2008.
- [8] C. Yu, H. Kai, and K. Wei-Shinn, "Collaborative Detection of DDoS Attacks over Multiple Network Domains," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 18, pp. 1649-1662, 2007.
- [9] G. Thatte, U. Mitra, and J. Heidemann, "Parametric Methods for Anomaly Detection in Aggregate Traffic," *Networking, IEEE/ACM Transactions on*, vol. 19, no. 2, pp. 512-525, 2011.
- [10] S. T. Sarasamma, Q. A. Zhu, and J. Huff, "Hierarchical Kohonen Net for Anomaly Detection in Network Security," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 35, pp. 302-312, 2005.
- [11] S. Yu, W. Zhou, W. Jia, S. Guo, Y. Xiang, and F. Tang, "Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 23, pp. 1073-1080, 2012.
- [12] S. Jin, D. S. Yeung, and X. Wang, "Network Intrusion Detection in Covariance Feature Space," *Pattern Recognition*, vol. 40, pp. 2185- 2197, 2007.
- [13] C. F. Tsai and C. Y. Lin, "A Triangle Area Based Nearest Neighbors Approach to Intrusion Detection," *Pattern Recognition*, vol. 43, pp. 222-229, 2010.
- [14] A. Jamdagni, Z. Tan, X. He, P. Nanda, and R. P. Liu, "RePIDS: A multi tier Real-time Payload-based Intrusion Detection System," *Computer Networks*, vol. 57, pp. 811-824, 2013.
- [15] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "Denialof- Service Attack Detection Based on Multivariate Correlation Analysis," *Neural Information Processing*, 2011, pp. 756-765.
- [16] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R. P. Liu, "Triangle- Area-Based Multivariate Correlation Analysis for Effective Denialof- Service Attack Detection," *The 2012 IEEE 11th International Conference on Trust, Security and Privacy in*

Computing and Communications, Liverpool, United Kingdom, 2012, pp. 33-40.

[17] S. J. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. K. Chan, "Costbased modeling for fraud and intrusion detection: results from the JAM project," The DARPA Information Survivability Conference and Exposition 2000 (DISCEX '00), Vol.2, pp. 130-144, 2000.

[18] G. V. Moustakides, "Quickest detection of abrupt changes for a class of random processes," Information Theory, IEEE Transactions on, vol. 44, pp. 1965-1968, 1998.

[19] A. A. Cardenas, J. S. Baras, and V. Ramezani, "Distributed change detection for worms, DDoS and

other network attacks," The American Control Conference, Vol.2, pp. 1008-1013, 2004.

[20] W. Wang, X. Zhang, S. Gombault, and S. J. Knapskog, "Attribute Normalization in Network Intrusion Detection," The 10th International Symposium on Pervasive Systems, Algorithms, and Networks (ISPAN), 2009, pp. 448-453.

[21] M. Tavallae, E. Bagheri, L. Wei, and A. A. Ghorbani, "A Detailed Analysis of the KDD Cup 99 Data Set," The The Second IEEE International Conference on Computational Intelligence for Security and Defense Applications, 2009, pp. 1-6.

[22] D. E. Knuth, The art of computer programming vol I: Fundamental Algorithms Addison-Wesley, 1973.