Detecting Denial-of-Service Attack based on Multivariate Correlation Analysis

Syed. Khajabee[1], Shaik. Gouse John[2]

[1] M.Tech (CS), Nimra College of Engineering and Technology, A.P., India.

[2]Asst. Professor, Dept. of Computer Science & Engineering, Nimra College of Engineering and Technology, A.P., India.

Abstract — Servers, such as Web servers, database servers, cloud computing servers etc, are now under threads from network attackers. As one of most common and aggressive means, Denial-of-Service (DoS) attacks cause serious impact on these computing systems. In this paper, we present a DoS attack detection system based on 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 exploits 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 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 surpasses two other previously developed state-of-the-art approaches in terms of detection accuracy.

Keywords — Denial-of-Service attack, network traffic characterization , multivariate correlations, triangle area.

1.Introduction

One of the aggressive and dangerous intrusive behaviors to online servers is Denial of Service (Dos) attacks. DoS attacks severely compromise 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 network based 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.

Research community, therefore, started to explore a way to achieve novelty-tolerant detection systems and developed a more advanced concept, namely anomaly-based 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, machine learning [4] and statistical analysis [5]. However, these proposed systems commonly suffer from high false positive rates because the correlations between features/attributes are intrinsically neglected [6] or the techniques do not manage to fully exploit these correlations.

I. RELATED WORK

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 [7] 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 [8] to generate better discriminative features. However, this approach has dependency on prior knowledge of malicious behaviors. More recently, Jamdagni et al. [9] developed a refined geometrical structure based analysis technique, where Mahalanobis distance was used to extract the correlations between the selected packet payload features.

In this paper the DoS attack detection system 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. Our proposed DoS detection system is evaluated using KDD Cup 99 dataset [10] and outperforms the state-ofthe- art systems shown in [11].

II. PROPOSED WORK

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.

A. **Framework**

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) and is detailed in Section 2.2. 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 [10].

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. Our MCA method and the feature normalization technique are explained in Sections 3 and 5.2 respectively.

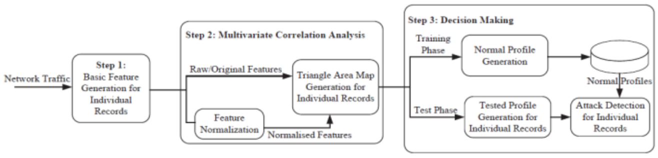

Network Traffic
Step 1: Basic Feature Generation for Individual Records
Step 2: Multivariate Correlation Analysis
Raw/Original Features
Triangle Area Map Generation for Individual Records
Feature Normalization
Normalised Features
Step 3: Decision Making
Training Phase
Normal Profile Generation
Normal Profiles
Test Phase
Tested Profile Generation for Individual Records
Attack Detection for Individual Records

Figure 1: Framework of the proposed denial-of-service attack detection system.

In Step 3, the anomaly-based detection mechanism [3] 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 Marking. 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

$$Tr^i_{j,k} = (\| (f^i_j, 0) - (0,0) \| \times \| (0, f^i_k) - (0,0) \|)/2,$$

employed in the “Attack Detection” module to distinguish DoS attacks from legitimate traffic.

B. **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. To better understand the merits, we illustrate them through a mathematical example

$$\begin{cases} P_1 = \displaystyle\int_{-\infty}^{\overline{\mu}} \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} dx, & (1) \\ P_2 = \displaystyle\int_{\overline{\mu}}^{+\infty} \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x-\mu_2)^2/2\sigma_2^2} dx, & (2) \end{cases}$$

$$\begin{cases} q_1 = \displaystyle\int_{-\infty}^{\overline{u}} (1/(\frac{1}{\sqrt{k}}\sigma_1\sqrt{2\pi}))e^{-(z-\mu_1)^2/\frac{2}{k}\sigma_1^2} dz, \\ q_2 = \displaystyle\int_{\overline{u}}^{+\infty} (1/(\frac{1}{\sqrt{k}}\sigma_2\sqrt{2\pi}))e^{-(z-\mu_2)^2/\frac{2}{k}\sigma_2^2} dz. \end{cases}$$

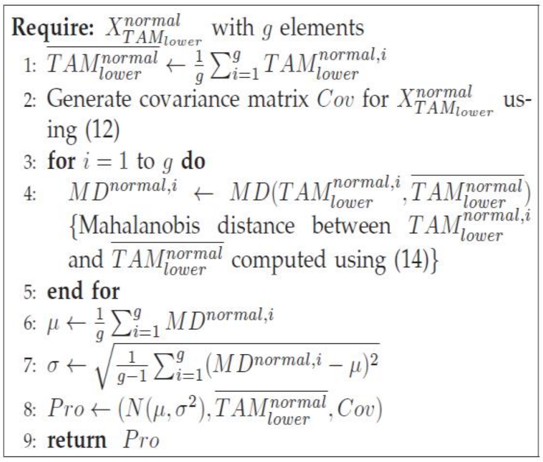
C. **Multivariate Correlation Analysis**

DoS attack traffic behaves differently from the legitimate network traffic, and the behavior of network traffic is reflected by its statistical properties. To well describe these statistical properties, we present a novel Multivariate Correlation Analysis (MCA) approach in this section. This MCA approach employs triangle area for extracting the correlative information between the features within an observed data object (i.e., a traffic record). The details are presented in the following.

Given an arbitrary dataset X = {x1, x2, · · · , xn}, where xi = [fi1 fi2· · · fim]T , (1 ≤ i ≤ n) represents the ith m-dimensional traffic record. We apply the concept of triangle area to extract the geometrical correlation between the jth and kth features in the vector xi. To obtain the triangle formed by the two features, data transformation is involved. The vector xi is first projected on the (j, k)-th two-dimensional Euclidean subspace as yi,j,k = [εj εk]T xi = [fij fik]T, (1 ≤ i ≤ n, 1 ≤ j ≤ m, 1 ≤ k ≤ m, j _= k). The vectors εj = [ej,1 ej,2 · · · ej,m]T and εk = [ek,1 ek,2 · · · ek,m]T have elements with values of zero, except the (j, j)-th and (k, k)-th elements whose values are ones in εj and εk respectively. The yi,j,k can be interpreted as a two dimensional column vector, which can also be defined as a point on the Cartesian coordinate system in the (j, k)-th two-dimensional Euclidean subspace with coordinate (fij , fik). Then, on the Cartesian coordinate system, a triangle ΔfijOfik formed by the origin and the projected points of the coordinate (fi j , fik) on the j-axis and k-axisis found. Its area Trij,k is defined as

D. **Detection Mechanism**

In this section, we present a threshold-based anomaly detector, whose normal profiles are generated using purely legitimate network traffic records and utilized for future comparisons with new incoming investigated traffic records. The dissimilarity between a new incoming traffic record and the respective normal profile is examined by the proposed detector. If the dissimilarity is greater than a pre-determined threshold, the traffic record is flagged as an attack. Otherwise, it is labeled as a legitimate traffic record. Clearly, normal profiles and thresholds have direct influence on the performance of a threshold-based detector. A low quality normal profile causes an inaccurate characterization to legitimate network traffic. Thus, we first apply the proposed trianglearea- based MCA approach to analyze legitimate network traffic, and the generated TAMs are then used to supply quality features for normal profile generation.



```
Require: $X^{normal}_{TAM_{lower}}$ with $g$ elements
$\overline{TAM^{normal}_{lower}} \leftarrow \frac{1}{g}\sum_{i=1}^{g} TAM^{normal,i}_{lower}$
Generate covariance matrix $Cov$ for $X^{normal}_{TAM_{lower}}$ using (12)
for $i = 1$ to $g$ do
    $MD^{normal,i} \leftarrow MD(TAM^{normal,i}_{lower}, \overline{TAM^{normal}_{lower}})$
    {Mahalanobis distance between $TAM^{normal,i}_{lower}$ and $\overline{TAM^{normal}_{lower}}$ computed using (14)}
end for
$\mu \leftarrow \frac{1}{g}\sum_{i=1}^{g} MD^{normal,i}$
$\sigma \leftarrow \sqrt{\frac{1}{g-1}\sum_{i=1}^{g}(MD^{normal,i} - \mu)^2}$
$Pro \leftarrow (N(\mu, \sigma^2), \overline{TAM^{normal}_{lower}}, Cov)$
return $Pro$
```

Figure 2: Algorithm for normal profile generation based on triangle-area-based MCA.

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] V. Paxson, “Bro: A System for Detecting Network Intruders in Realtime,”
Computer Networks, vol. 31, pp. 2435-2463, 1999

[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] 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.

[5] 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.

[6] S. T. Sarasamma, Q. A. Zhu, and J. Huff,
“Hierarchical Kohonenen Net for Anomaly Detection in Network Security,” Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, vol. 35, pp. 302-312, 2005.

[7] 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.

[8] S. Jin, D. S. Yeung, and X. Wang, “Network Intrusion Detection in Covariance Feature Space,” Pattern Recognition, vol. 40, pp. 2185- 2197, 2007.

[9] 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.

[10] 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.

[11] 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.