ISSN (e): **2250 – 3005** || Vol, 04 || Issue, 8 || August – 2014 ||
International Journal of Computational Engineering Research (IJCER)

A New Way of Identifying DOS Attack Using Multivariate Correlation Analysis

R Nagadevi[1], P Nageswara Rao[2], Rameswara Anand[3]
1(Dept of Cse, Swetha Institute of Technology and Science, Tirupati)
2(Associate Professor & Head, Dept Of Cse, Swetha Institute Of Technology And Science, Tirupati)
3(Professor, Dept of Cse, Swetha Institute Of Technology And Science, Tirupati)

ABSTRACT
This paper talked about the results of MCA on the Distributed DoS detection and suggests an example, a covariance analysis model for detecting SYN flooding attacks. The imitation end results show that this method is highly accurate in detecting malicious system traffic in Distributed DoS attacks of different forces. This technique can effectively distinguish between ordinary and attack traffic. To be sure, this technique can identify even very fine attacks only a little different from normal behaviors. The linear difficulty of the method makes its immediate detection practical. The covariance model in this document to some area verifies the effectiveness of multivariate correlation analysis (MCA) for Distributed DoS detection. Some open problem still exists in this model for further research.

Key words: *Wireless DoS, MCA, malicious node, jammer, learning patterns.*

I. INTRODUCTION

In modern days, securities have been precedence throughout the transmission of data in wireless network, be it through ad hoc, Wi-Fi or wireless sensor network (WSN). As of the reality of hacking and further malicious activities that happen now like any other common day-to-day routine. Due to the growth in technology, wireless networks are coming into reality as they've become more inexpensive and easily reachable through the off-the-shelf machinery. So they are some tools to interrupt these developments. While wireless networks are easy handier for the use of internet in the next to past and future, it is more weak to attacks than wired network.

The broadly known authenticity about the wireless network is its easy accessibility and sharable nature of intermediate. This reality is together the in favor of and cheat when it comes to a wireless network i.e., it is very simple for the competitor to start an attack. This attack can be the disturbance of network functions and flooding the user and kernel buffers. It is termed as Denial of service attack or jamming, depending on whether one looks at the consequence or the cause of attack. A most common example of such an attack is while browsing the internet, the page that is to be unlocked is not catching loaded properly and the refresh push button is clicked a number of times than necessary. This is an example of jamming or the Denial of Service attack that is done accidentally. This attack can also be done purposely. For example, one can use a mobile device to send volume of SMS in hinterland. This is sufficient to block announcement among a few wireless nodes.

In fact, it has happen to extra like a contest between the enemy to attack a network and the security experts to invent efficient techniques to block the attack. The networks have to be able of broadcast of data between the valid nodes irrelevant of the attack encouraged by the enemy. There must not be any intermission between the genuine users. Intimation about the presence of an attacker must be given to the top of the network. It is also not decently and with honesty accepted if the legitimate node/ user communicate with the attacker. On such times the node complicated in such a trick must be recognized and advised of any other ambiguous behavior in the network could cooperate both the network and the data.

Our paper is organized as follows: In section II, we discuss the related theory about the jamming and various techniques. Section III comprises of the system design and proposed system. Section IV describes about the algorithms for Multivariate Correlation Analysis. Section V will conclude the paper.

II. RELATED THEORY

Denial of service attack is mainly done in categorize to block a node from receiving genuine data or to block the node entirely from another genuine node. This blocking is able to be done either with the data sent frequently or by sending radio signal indications or by some other means of transmission signal congestion. Manu authors who have discussed about the various congestion techniques and their detection and/ or prevention techniques.

In [1], the authors have analyzed the different types of denial of service attacks and the shown issues due to the DoS attack in all networks. They have provided a number of intrusion detection methods in their survey and have mentioned that there must be system implementation to avoid real world opponent. In all of the congestion techniques and the detection algorithms, throughput is 0 which successfully reduces the performance of the network.

In [2], the authors have detailed about the selective congestion where the opponent chooses the data to squeeze preferentially a high priority data when it concerns protection and privacy. They do so by performing packet classification at the physical layer. The authors have appraised the property of packet hiding by measuring the effective throughput of the
TCP connection in the following states:

1. No packet hiding (N.H.).
2. MAC-layer encryption with a static key (M.E).
3. SHCS (C.S.).
4. Time-lock CPHS (T.P.).
5. Hash-based CPHS (H.P.).
6. Linear AONT-HS (L.T.).
7. AONT-HS based on the package transform.

In [3], data forwarding without any delay in the defending congestion in a wireless sensor network is proposed. This offer consists of sensor nodes as clusters used for a exacting frequency rate. Now when a frequency rate where data promoting occurs is blocked, the cluster of sensor nodes in to frequency turn into inoperative and the other clusters act as backup.

[4] Discusses the system of game theory. This Game theory method offers powerful tools to form and evaluate such attacks. This technique talked about a class of such congestion games played at the MAC layer among a set of transmitters and squeezers. The stability strategies ensuing from these congestion games characterize the expected performance under DoS attacks and motivate robust network protocol design for secure wireless communications. A key characteristic of the distributed wireless access networks is that users do not have complete information regarding the other user's character, the traffic lively, the control channel characteristics, or the rates and rewards of other clients.

The whole detection process consists of three major steps as shown in Fig. 1. Step 1: The 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 period. Observing and analyzing at the destination network diminish the overhead of detecting cruel activities by concentrating only on related inbounded traffic. This as well allows our detector to give 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. Step 2: 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 removed relationships, that is to say triangle areas lay up in Triangle Area Maps (TAMs), are then used to replace the original

Basic features or the standardized features to represent the traffic report. This provides higher discriminative information to differentiate between legitimate and illegal traffic reports. Our MCA method and the quality normalization technique are explained in Sections 3 and 5 respectively. Step 3: The anomaly-based detection mechanism is adopted in result creation. It makes easy the detection of any DoS attacks without requiring any attack related knowledge. Also, the manual 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 manual task and needs expertise in the targeted detection algorithm. Particularly, 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 documentation. Next, the tested profiles are passed over to the "Attack Detection" section, which evaluates the individual tested profiles with the own stored normal profiles. A threshold-based classifier is employed in the "Attack Detection" section module to distinguish DoS attacks from legitimate traffic.

III. MULTIVARIATE CORRELATION ANALYSIS

DoS attack traffic behaves differently from the legitimate network traffic and the behavior of network traffic is reflected by its geometric assets. To well describe these statistical properties, here a novel Multivariate Correlation Analysis (MCA) moves toward in this section. This MCA approach use triangle area for remove the correlative data between the features within an observed data object (i.e., a traffic record).

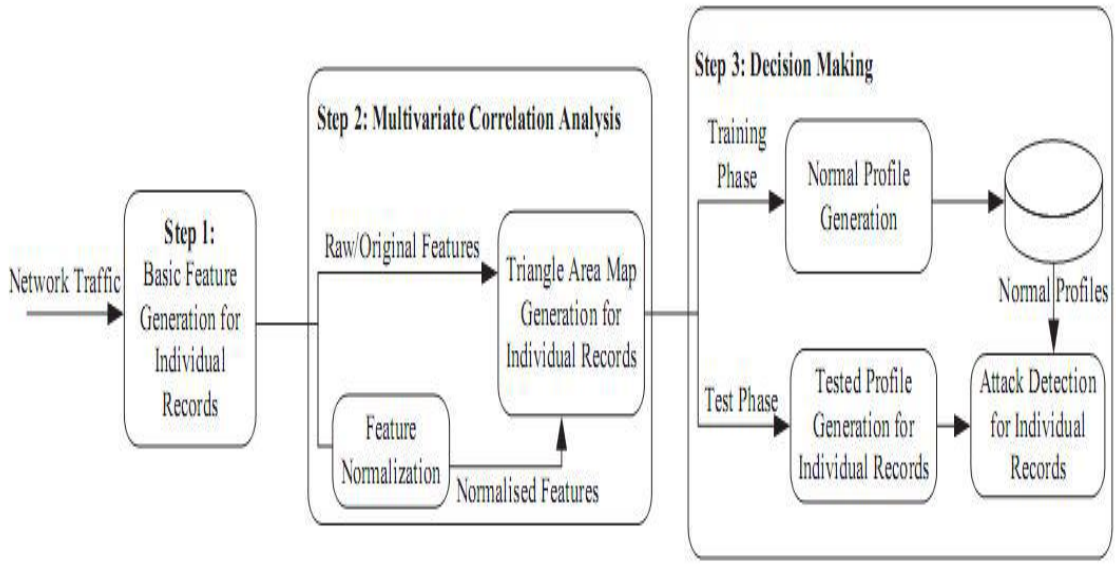

Figure 1: SYSTEM ARCHITECTURE

IV. DETECTION MECHANISM

In this section, we present a threshold-based on anomaly detector, whose regular profiles are produced using purely legal network traffic records and utilized for future comparisons with new incoming investigated traffic report. The difference between a fresh arriving traffic record and the individual normal outline is examined by the planned detector. If the difference is greater than a pre-determined threshold, then the traffic record is colored as an attack. If not, it is marked as a legal traffic record. Clearly, normal profiles and threshold points have direct power on the performance of a threshold-based detector. A low down quality normal shape origins an mistaken characterization to legitimate network traffic. Thus, we first apply the proposed triangle area- based MCA approach to analyze legitimate network traffic, and the created TAMs be then used to supply quality features for normal profile generation.

4.1 Normal Profile Generation

Assume there is a set of g legitimate training traffic records Xnormal = {Xnormal1, Xnormal2, · · · , Xnormalg}.The triangle-area-based MCA approach is applied to examine the records. The produced lesser triangles of the TAMs of the set of g legitimate training traffic records are denoted by Xnormal TAMlower ={TAMnormal,1lower , TAMnormal,2lower , · ·
· , TAMnormal,glower }.

Mahalanobis Distance (MD) is adopted to measure the dissimilarity between traffic proceedings. This is for the reason that MD has been successfully and widely used in cluster studies, categorization and multivariate outlier detection techniques. Unlike Euclidean distance and Manhattan distance, it assesses distance linking two multivariate information objects by taking the correlations between variables into account and removing the dependency on the scale of measurement during the calculation.

4.2 Threshold Selection

The threshold point is used to distinguish attack traffic from the legal one.

Threshold = μ + σ ∗ α. (16)

For a normal distribution, α is usually ranged from 1 - 3. This means that detection decision can be made with a certain level of confidence varying from 68% to 99.7% in association with the selection of different values of α. Thus, if the MD between an observed traffic record Xobserved and the respective normal profile is greater than the threshold point value, then it will be measured as an attack.

4.3 Attack Detection

To detect DoS attacks, the lower triangle (TAMobservedlower) of the TAM of an observed record needs to be generated using the proposed triangle-area-based MCA move toward. Next, the MD among the TAMobserved lower and the TAMnormal lower stored in the respective pre-generated normal profile Pro is computed. The detailed detection algorithm is shown in Fig. 2.

Require: Observed traffic record Xobserved, normal profile
Pro : (N(μ, σ2), TAMnormal
lower , Cov) and parameter
α
1: Generate TAMobserved
lower for the observed traffic
record xobserved
2: MDobserved ← MD(TAMobserved
lower ,TAMnormal lower)
3: if (μ − σ ∗ α) ≤ MDobserved ≤ (μ + σ ∗ α) then
4: return Normal
5: else
6: return Attack
7: end if

Fig. 2. Algorithm for attack detection based on Mahalanobis distance.

V. V EVALUATION OF THE MCA-BASED DOS ATTACK DETECTION SYSTEM

The estimate of our projected DoS attack detection system is conducted using KDD Cup 99 dataset [17]. Despite the dataset is criticised for redundant records that prevent algorithms from learning infrequent harmful records [21], it is the only publicly available labeled benchmark dataset, and it has been widely used in the domain of intrusion detection research. Testing our approach on KDD Cup 99 dataset contributes a convincing evaluation and makes the comparisons with other state-of-the-art techniques equitable. Additionally, our detection system innately withstands the negative impact introduced by the dataset because its profiles are built purely based on legitimate network traffic. Thus, our system is not affected by the redundant records. During the evaluation, the 10 percent labeled data of KDD Cup 99 dataset is worn, here we have 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 (see Table 9 in Appendix 4 in the supplemental file to this paper for details).

The general evaluation procedure is in depth as follows.

First, the proposed triangle-area-based MCA approach is assessed for its capability of network traffic characterization. Second, a 10-fold cross-validation is conducted to evaluate the detection performance of the proposed MCA-based detection system, and the entire filtered data subset is used in this assignment. In the training stage, we utilize only the Normal records. Normal profiles are built with respect to the different types of legitimate traffic using the algorithm presented in Fig. 2. The corresponding thresholds are determined according to (16) given the parameter α varying from 1 to 3 with an addition of 0.5. In the test phase, together the Normal records and the attack records are taken into account. As given in Fig. 3, the observed samples are examined against the respective normal profiles which are built based on the legitimate traffic records carried using the same type of Transport layer procedure. Third, four metrics, namely True Negative Rate (TNR), Detection Rate (DR), False Positive Rate (FPR) and Accuracy (i.e. the proportion of the overall samples which are classified correctly), are used to evaluate the proposed MCA-based detection system. To be a good candidate, our proposed detection system is required to achieve a high detection accuracy.

5.1 Problems with the Current System and Solution

Even though the detection system reaches a moderate overall detection performance in the above evaluation, we want to explore the causes of degradation in detecting the Land, Teardrop and Neptune attacks.

Our analysis shows that the problems come from the data used in the evaluation, where the basic features in the non-normalized original data are in different scales. Therefore, even though our triangle-area-based MCA approach is promising in characterization and clearly reveals the patterns of the various types of traffic report, our detector is silent ineffective in various of the attacks. For instance, the Land, Teardrop and Neptune attacks whose patterns are different than the patterns of the legitimate traffic. However, the level of the dissimilarity between these attacks and the respective normal profiles are close to that between the legitimate traffic and the respective normal profiles. Moreover, the changes appearing in some other more important features with much smaller ideals can only just take effect in unique the DoS attack traffic from the legal traffic, since the overall variation is subject by the features with large values. Nevertheless, the non-

normalized original data have zero values in several of the features (both the important and the less important features), and they confuse our MCA and make many new generated features equal to zeros. This vitally degrades the discriminative power of the new feature set (TA lower), which is not supposed to happen. Apparently, an appropriate data normalization technique should be employed to eliminate the bias. We adopt the statistical normalization technique [20] to this work. The statistical normalization takes both the mean scale of attribute values and their statistical distribution into description. It exchanges data derived from any normal allocation into standard normal distribution, inside which 99.9% samples of the attribute are scaled into [-3, 3]. In addition, statistical normalization has been proven improving detection performance of distance-based classifiers and outperforming other normalization methods, such as mean range [0, 1], ordinal normalization etc

VI. CONCLUSION AND FUTURE WORK

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. In addition, the computational complexity and the time cost of the proposed detection system have been analyzed and shown in Section 6. The proposed system realizes 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 actual world data and spend more sophisticated arrangement performances 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] 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 Kohonenenet 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.