# A Scaling Approach for Effective Detection and Classification of DoS Attacks

V. Monica
CSE Dept., JNTUA College of Engineering, Anantapuramu, A.P, India.

B. Lalitha
Assistant Professor
CSE Dept., JNTUA College of Engineering, Anantapuramu, A.P, India.

**Abstract – Denial of Service (DOS) or distributed denial of service (DDoS) is an attempt to make a resource of the machine or network unavailable for its intended users. DOS attack can reduce server efficiency, in order to increase server efficiency, it is necessary to detect attacks back. In respect to this perspective, here in this paper, a novel mathematical scaling procedure is designed to approximate the network transactions on their security aspects. The recommended model uses bipartite graph strategy to approximate the powerful associability of the attributes. The attribute association is basically symbolized by the associability of that attributes with specific values. The outcomes explained from the empirical study identify that the recommended procedure effectively offer the consistency in identifying diverse attacks during the communication process.**

**Index Terms – Denial-of-service attack, bipartite graph strategy, novel mathematical scaling.**

## 1. INTRODUCTION

In previous decades the Internet has undergone a sudden growth. With the large expansion of new services, the quantity and impact of attacks are frequently increased. The amount of computer systems and their exposure was in rapid growth because of the level of sophistication and understanding the necessity to have an attack have become lower, which specify how the massive technical approach is commonly presented in websites.

Denial of service attack seriously degraded the efficiency of online services. The attackers in the way of cumulatively assigning user privileges too. In other words DDoS attack, prevent the legitimate users to use specific network resources such as web services, real-time services and cloud services.

A denial-of-service attack is an action that prevents damage or authorized use of networks, system or application by exhausting resources such as CPU, memory, bandwidth and disk space. In the case of, DOS attacks the user cannot access resources such as email and the internet. An attack can be directed to an operating system or network

DDoS inflicts intensive measurement tasks to the victim by taking advantage of the vulnerability of the system or flooding it with large amounts of unnecessary packets. The victim may be forced to leave the service for a day to even minutes. This intention profound damage to a service running on the patient. Consequently, effective detection of denial of service is important for the protection of online services. Work on finding denial of service is the main axes on the development of network -based discovery mechanisms.

Denial of Service (DOS) is an unlimited threat to Internet sites and among the trouble of difficult securing Internet today. The problem of denial of service has become well known, but it was difficult to get the Denial of service on the net. DDoS attack is a huge attack on a service availability of a victim or network resources structure, indirectly released through many compromises computers on the net. The researcher has no specific remedy to a DDoS problem

Discovery systems based on a single statistic strategy based intrusion detection abnormality is projected herein. Efforts to determine a proportion that the evaluations effect of a network transaction if it is protected, the suspect or the entry.

## 2. RELATED WORK

A lot of research work has been done on DDOS attack. There were many changes applied to the DDOS for detecting the attacks. Some of those efforts are shown below.

- Neuro-Fuzzy System
- Triangle Area Technique
- Traceback System
- TANN

### 2.1 Neuro-Fuzzy System

In "High accuracy detection of denial of service attack based on triangle map generation" [9] Analysis map generation technique and triangle statistical area facilitates our system to be able to distinguish DOS attacks legitimate network traffic. To produce an accurate detection of the fixed threshold value Neuro- Fuzzy systems have been proposed as subsystems of the whole. Type Sugeno Neuro - Fuzzy Inference System was chosen as the basis and use the classifier system gives a

highly accurate detection and low overhead calculation was obtained.

2.2 Triangle Area Technique

In [3] propose a procedure called multivariate correlation analysis can be solved using a technique of statistical standardization to eliminate the bias data. This technique extracts the hidden geometric correlations in different pairs of two distinct functions within each network traffic recording, with true for further characterization behavior of network traffic. Furthermore Triangle Area Technical basis is used to accelerate the process of multivariate analysis Correlation (MCA). The proposed system can be evaluated using KDD Cup data set

2.3 Traceback System

In "Moderate Denial-of-Service attack detection based on Distance flow and Traceback Routing"[7] proposed IP Traceback system against DDoS attacks based on entropy changes. Here, the packet marking strategy is avoided, because it suffers from a number of disadvantages. It uses by storing information flow entropy changes at routers. Once a DDOS attack has been identified, it performs push search procedure. The algorithm first identified its upstream router investigation where the attack comes from streams and made the request to the router tracing related upstream

2.4 TANN

In "A triangular area based nearest neighbors approach to intrusion detection" [8] proposes a model based on triangle area based nearest neighbors (TANN) to detect attacks more effectively. In TANN, the k-means clustering is first used to obtain cluster centers corresponding to the classes of attack, respectively. Then, the triangle area by two cluster centers with a given data set is calculated and formed a new signature based on the data. Finally, the classifier k - NN is used to classify similar attacks based on the new functionality represented by the triangle areas

## 3. PROPOSED MODELLING

3.1 Intrusion Detection by Feature Association

The approach of PDDOS metric proposed in this paper is initially considered the records of the given training set and feature categorical values used in those records as two independent sets and further builds a bipartite graph between these two.

An Assumptions**:**

Let set of features

$\{f1, f2, f3, \ldots\ldots, fn \forall f_i = \{f_i v_1, f_i v_2, \ldots\ldots f_i v_m\}\}$ Which are having categorical values and used to form the $T$

Here $T$ is a set of network transaction records of the given training set such that $T = \{t_1, t_2, t_3, \ldots\ldots t_n \forall t_i = \{val(f_1), val(f_2), \ldots\ldots val(f_i), val(f_{i+1}), \ldots\ldots val(f_n)\}\}$

The set of categorical values of features belongs each network transaction will be considered as transaction value set $tvs$, and all transaction value sets are referred as '$STVS$'.

Here in above description $val(f_i)$ can be defined as $val(f_i) \in \{f_i v_1, f_i v_2 \ldots . f_i v_m\}$

Hereafter the term feature refers the current categorical value of the feature

Let two features '$val(f_i)$' and '$val(f_j)$', '$val(f_i)$ 'connected with '$val(f_j)$'if and only if $(val(f_i), val(f_j)) \in tvs_k$.

Build a weighted graph $WG$ with values of features as vertices and edges between values of features. An edge between any two features $val(f_1), val(f_2)$ will be weighted as follows

$$ctvs = 0;$$
$$foreach \{tvs \forall tvs \in STVS\} \quad (1)$$
$$ctvs+ = \{1 \forall (val(f_1), val(f_2)) \subseteq tvs\}$$

Here in the above equation $ctvs$ indicates the count of transactions, which contains both features $val(f_1), val(f_2)$. Then the edge weight between features $val(f_1)\ and\ val(f_2)$ can be measured as follows.

$$w(val(f_1) \leftrightarrow val(f_2)) = \frac{ctvs}{|STVS|} \quad (2)$$

In the process of building a weighted graph we consider that an edge between any two features exists if and only if $ctvs \geq 1$

B. Process

In regard to table1 explore the process by an example, let consider the total number of divergent values of features as 8 that represented as a set $V = \{val_1, val_2, \ldots . val_8\}$ and $|T|$ as 6, Here $|T|$ is size of the network transaction records

Table 1 binary representation of the association between $T$ and $V$

| | Val1 | Val2 | Val3 | Val4 | Val5 | Val6 | Val7 | Val8 | |
|---|---|---|---|---|---|---|---|---|---|
| tvs1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | (Val5, Val6, Val8) |
| tvs2 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | (Val2, Val3, Val6, Val8) |
| tvs3 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | (Val1, Val2, Val3, Val7) |
| tvs4 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | (Val7) |
| tvs5 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | (Val4, Val6, Val7, Val8) |
| tvs6 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | (Val1, Val2, Val3, Val4, Val7) |

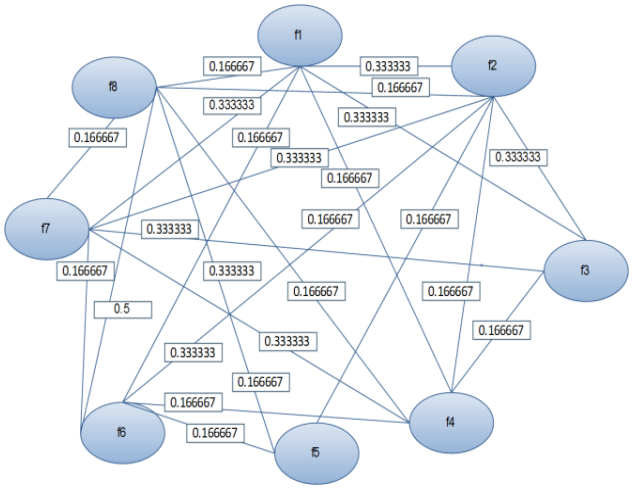


Fig 1: An example weighted graph of categorical values set of count 8.

Here in above table1 and Figure1 each element $\{val_1, val_2, ....val_8\}$ can be $f_i v_j$ such that $\{f_i v_j \exists i \in [1,2,.....n] \wedge j \in [1,2,.....m]\}$

The process of detecting the association of each feature categorical value $f_i v_j$ referred as $val_k$ with network transaction records, initially we build a bipartite graph between transaction value sets $STVS$ and the feature categorical values $V$ is been shown in figure 2.

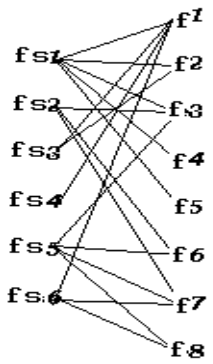


Fig 2: bipartite graph between STVS and V

If a feature categorical value $f_i v_j$ as shown in table 2 that referred as $val_1$ exists in $tvs_1$ then the weight of the connection between $val_1$ and $tvs_1$ will be the sum of the weights of the edges between $val_1$ and each feature categorical value $\{f_i v_j \exists f_i v_j \in tvs_1\}$ of $tvs_1$ that defined in weighted graph $WG$.

Table 2: matrix A as follows that represents the connection weights between a feature categorical value and each transaction value set

| | $val_1$ | $val_2$ | $val_3$ | $val_4$ | $val_5$ | $val_6$ | $val_7$ | $val_8$ |
|---|---|---|---|---|---|---|---|---|
| $tvs_1$ | 0.879761 | 0.593871 | 0.719431 | 0.888985 | 0.009925 | 0.951906 | 0.596128 | 0.140759 |
| $tvs_2$ | 0.487673 | 0.175749 | 0.473697 | 0.098931 | 0.672281 | 0.65032 | 0.688884 | 0.102504 |
| $tvs_3$ | 0.334752 | 0.194757 | 0.436627 | 0.005882 | 0.196037 | 0.337898 | 0.703112 | 0.58396 |
| $tvs_4$ | 0.737372 | 0.696016 | 0.147575 | 0.993947 | 0.865990 | 0.515846 | 0.629118 | 0.014945 |
| tvs5 | 0.071456 | 0.59152 | 0.712407 | 0.471292 | 0.328539 | 0.693965 | 0.889919 | 0.434219 |
| tvs6 | 0.00046 | 0.767607 | 0.056246 | 0.148297 | 0.924187 | 0.313737 | 0.506235 | 0.993027 |

Table 3: Transpose matrix $A'$ of matrix $A$ as fallows that represents the connection between a transaction and each transaction level feature set $fs$.

| | $tvs_1$ | $tvs_2$ | $tvs_3$ | $tvs_4$ | tvs 5 | tvs6 |
|---|---|---|---|---|---|---|
| $val_1$ | 0.879761 | 0.487673 | 0.334752 | 0.737372 | 0.071456 | 0.00046 |
| $val_2$ | 0.593871 | 0.175749 | 0.194757 | 0.696011 | 0.59152 | 0.767607 |
| $val_3$ | 0.719431 | 0.473697 | 0.436627 | 0.147578 | 0.712407 | 0.056246 |
| $val_4$ | 0.888985 | 0.098931 | 0.005882 | 0.993941 | 0.471292 | 0.148297 |
| $val_5$ | 0.009925 | 0.672281 | 0.196037 | 0.865996 | 0.328539 | 0.924187 |
| $val_6$ | 0.951906 | 0.65032 | 0.337898 | 0.515843 | 0.693965 | 0.313737 |
| $val_7$ | 0.596128 | 0.688884 | 0.703112 | 0.629113 | 0.889919 | 0.506235 |
| $val_8$ | 0.140759 | 0.102504 | 0.58396 | 0.01492 | 0.434219 | 0.993027 |

Let consider $STVS$ as a database and depict it as a bipartite graph without loss of information. Let $STVS = \{tvs_1, tvs_2, ...., tvs_6\}$ be a list of network transactions with feature categorical values and $V = \{val_1, val_2, ....val_8\}$ be the corresponding set of feature correlation values. Then, clearly $STVS$ is equivalent to the duplex-graph $DG = (STVS, V, E)$

Here

$$E = \{(tvs_i, val_j) : val_j \in tvs_i, tvs_i \in STVS, val_j \in V\} \quad (3)$$

The bipartite graph (3) representation of the set of transaction value sets $SCFS$ is inspiring. It gives us the idea of applying link-based ranking models for the evaluation of connected sets. In this bipartite graph, the association support of a

transaction c is proportional to degree of all its features weight. However, it is crucial to have different closeness weights for different transaction value sets in order to reflect their different importance. The evaluation of influence connected sets $ics$ should be derived from these weights. Here comes the question of how to acquire weights in a set of transaction value sets. Intuitively, a transaction level feature set with high closeness weights should contain many of the features those belongs to the same transaction with high association support; at the same time, a transaction with high association support should be contained by less or zero other transaction value sets high closeness weights. The reinforcing relationship of transaction value sets and transactions is just like the relationship between hubs and authorities in the bipartite graph.

Further assuming transaction value sets as pure hubs and the feature categorical values as pure authorities, the hub and authority values can be measured as follows:

Let matrix representation of transaction value sets and feature connections as a matrix 'A' as shown in table 3. The value represents that a feature connected how many feature categorical values of the same transaction

If a feature $f_1$ exists in feature set $fs_1$ then the weight of the connection between $f_1$ and $fs_1$ will be the sum of the weights of the edges between $f1$ and each feature of $fs1$ that defined in weighted graph $WG$.

Consider the matrix $u$ that representing each hub initial value as 1.

As in table 4 initially consider the each recorded weights as 1 by default as fallow and represent them as matrix u.

| |
|---|
| 1 |
| 1 |
| 1 |
| 1 |
| 1 |
| 1 |

Transpose the matrix A as A' (see table 4)

Find Feature weights by multiplying $A'$ with $u$ as $v = A' \times u$ (Matrix multiplication between $A'$ and u gives a matrix v that represents the authority weights)

Now find the original recorded weights through matrix multiplication between $A and v$.

$$u = A \times v \tag{4}$$

Then the $Pddos$ of feature association value $f_i v_j$ can be measured as follows

$$Pddos(f_i v_j) = \frac{\sum_{k=1}^{|STVS|} \{u(tvs_k) : (f_i v_j \rightarrow tvs_k) \neq 0\}}{\sum_{k=1}^{|STVS|} u(tvs_k)} \tag{5}$$

Then the $Pddos$ between feature association values $f_i v_j$ and $f_{i'} v_{j'}$ can be measured as follows

$$pddos(f_i v_j \leftrightarrow f_{i'} v_{j'}) = \frac{\sum_{k=1}^{|STVS|} \{u(tvs_k) \exists (f_i v_j, f_{i'} v_{j'}) \subset tvs_k\}}{\sum_{k=1}^{|STVS|} u(tvs_k)} \tag{6}$$

Here in the above equation descriptions, the $|STVS|$ represents total number of transaction value sets.

Further the $Pddos$ of the each transaction value set $tvs_i$ can be measured as follows

$$pddos(tvs_i) = 1 - \frac{\sum_{j=1}^{m} \{pddos(\{val_j \exists val_j \in V\}) : (val_j \subset tvs_i)\}}{|tvs_i|} \tag{7}$$

$$pddost = \frac{\sum_{i=1}^{|STVS|} pddos(tvs_i)}{|STVS|} \tag{8}$$

Here in the above equation $|STVS|$ indicates the total number of transaction value sets

The standard deviation of the $Pddos$ of each transaction value set needs to be measured further, which is in regard to estimate the low, medium and high ranges of $pddost$. The exploration of mathematical notation of estimating standard deviation follows

$$sdv_{pddost} = \sqrt{\frac{\left(\sum_{i=1}^{|STVS|} (pddos(tvs_i) - pddost)^2\right)}{(|STVS| - 1)}} \tag{9}$$

The Feature Association Impact Scale range can be explored as follows

Lower threshold of $pddost$ range is

$$pddost_l = pddost - sdv_{pddost} \quad (10)$$

Higher threshold of $ddpt$ range is

$$pddost_h = pddost + sdv_{pddost} \quad (11)$$

A network Transaction $nt$ can be said as safe if and only if

$$pddos(nt) < pddost_l \quad (12)$$

A Network Transaction $nt$ can be said as suspected to be an intrusion if and only if

$$pddos(nt) \geq pddost_l \,\&\&\, pddos(nt) < pddost_h \quad (13)$$

The Network Transaction $nt$ can be confirmed as intrusion if

$$pddos(nt) \geq pddost_h \quad (14)$$

## 4. PRAGMATIC ANALYSIS OF THE PROPOSED MODEL

We considered the reliability of the projected system on a prepared network transaction dataset of NSL-KDD.

The preceding said data set possesses 125973 selections as preparing set, and 22544 selections are obtainable as test set. The working out set is used to calculate the showcase relationship affect scale threshold and its lower, medium and upper values. The test set is utilized to forecast the scalability of the projected model. Curiously, the scientific study provided promising results. The reports explained in table 5

Table 5: Statistics of the experiment results

| | |
|---|---|
| Total Number of Records | 148517 |
| Total Number of Fields in Record | 41 |
| Total Number of feature categorical values found | 18370 |
| Total number of edges determined | 146960 |
| Feature Association impact Scale Threshold Found | 0.802100787 |
| Feature Association impact Scale Threshold Upper Bound | 0.862553922 |
| Feature Association impact Scale Threshold Lower Bound | 0.741647652 |

Total records Tested 22544

Total number of records found with ' *fais* ' less than lower bound are 3502 (out of this false negatives are 1288)

Total number of records found with ' $fais$ ' greater than the lower bound are 21042 (true positives are 18692 and 2350 records are false positives)

As per the results explore in table 2 and 3, the projected model is perfect to the level of 92.73%. The failure percentage is 7.26%, which is supposed and occurred due to categorical principles of the features.

## 5. PERFORMANCE ANALYSIS

We used interruption detection correctness (the portion of appropriate forecasts of the recommended) as the primary efficiency measure. As shown in table 6, In acquisition to calculating precision, the precision, recall, and F-measure were utilized to measure the efficiency; these are characterized using appropriate equations.

$$pr = \frac{t_+}{t_+ + f_+} \quad (15)$$

Here in above Equation the $pr$ indicates the precision, $t_+$ indicates the true positives and $f_+$ indicates the false positive

$$rc = \frac{t_+}{t_+ + f_-} \quad (16)$$

Here in exceeding Equation, the ' $rc$ ' indicates the recall, ' $f_-$ ' indicates the false negative.

$$F = \frac{2 * pr * rc}{pr + rc} \quad (17)$$

Here in the above Equation, ' $F$ ' indicates the F-measure.

Table 6: Precision, recall and F-measure values found from the results of the empirical analysis.

| Precision | Recall | F-measure |
|---|---|---|
| 0.888336 | 0.9355634 | 0.9113436 |

## 6. CONCLUSION:

A unique statistical strategy regarding anomaly based intrusion detection is projected on this paper. The endeavors to determine a proportion that assessments the effect of a network transaction if it is protected, suspicious or entrance is first in best of our information. The empirical results acquired from scientific study performed on NSL-KDD dataset is excellent and stimulating our analysis further. In upcoming a novel future connection evaluation strategy can be required that might lead to eliminate the deemed feature set and procedure difficulty, and also might stimulate the reliability towards intrusion detection scope.

## REFERENCES:

[1] Kemmerer, R.A., Vigna, G., Intrusion Detection: a Brief History and Overview, IEEE Security and Privacy (supplement to Computer, vol. 35, no. 4) pp 27-30, April 2002
[2] Ye, N., Yebin Zhang, Y., and Borror, C.M., Robustness of the MarkovChain for Cyber-Attack Detection, IEEE Transactions on Reliability, Vol. 53, no. 1, pp. 116-123, March 2004
[3] V. Paxson, "Bro: A System for Detecting Network Intruders in Real-Time," Computer Networks, vol. 31, pp. 2435-2463, 1999.
[4] Z. Tan, A. Jamdagni, X. He, P. Nanda, and R.P. Liu, "Triangle-Area-Based Multivariate Correlation Analysis for Effective Denialof- Service Attack Detection," Proc. IEEE 11th Int'l Conf.Trust, Security and Privacy in Computing and Comm., pp. 33-40, 2012.
[5] M. Craymer, J. Cannady, J. Harrell. "New Methods of Intrusion Detection using Control-Loop Measurement." In: Fourth Technology for Information Security Conference'96. May, 16, 1996.
[6] W. Lee, S. Stolfo. "Data Mining Approaches for Intrusion Detection." In: Proceedings of the 7th USENIX Security Symposium. 1998
[7] T.Senthil Prakash, D Yuvraj. "Moderate Denial-of-Service attack detection based on Distance flow and Traceback Routing" ISSN (P): 2349-3968, ISSN (O): 2349-3976 Volume 1 Issue 7, November 2014.
[8] 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.
[9] J. Welkin, S. Karthiprem, E. Thangadurai, "High accuracy detection of denial of service attack based on triangle map generation" Mobile Computing ISSN 2320–088X.

Authors

V. MONICA received B.Tech degree in Computer Science and Engineering from SRM University, Chennai, T.N, India, during 2009 to 2013. Currently pursuing M.Tech in Computer Science from JNTUA College of Engineering, Anantapuramu, A.P, India. Her Area of interests include Computer Networks, Network Security.

B. LALITHA is an Assistant Professor of Computer Science and Engineering at Jawaharlal Nehru Technological University College of Engineering, Ananthapuramu. She obtained her Bachelor degree in Computer Science Engineering from Sitams, Chittor, Master of Technology in Computer Science from Jawaharlal Nehru Technological University Anantapur and pursuing Ph.D. in Computer Science and Engineering from Jawaharlal Nehru Technological University Anantapuramu. She has published several Research papers in National International Conferences and Journals. Her research interests include Distributed Computing and Cloud Computing.