Grenze International Journal of Engineering and Technology, Special issue

An Improvement of Intrusion Attack Categorization and Recognition System

Sindhuja G[1] and Karthikeyan D[2]
[1-2]Kongunadu College of Engineering and Technology/Deportment of ECE, Trichy, India
Email;1031sindhu@gmail.coml;duraikkeyan@gmail.com

***Abstract*—Intrusion detection (ID) is a method of network security management system for computers. An IDS gathers the information from various area within a computer or a network to identify possible security breaches, which include both intrusions and misuse. ID uses vulnerability assessment, which is a method developed to assess the security of a computer system. The misuse detection, the IDS analyzes the information it gathers and compares it to large databases of attack signatures. In anomaly detection, the system administrator defines the baseline, state of the networks traffic load, breakdown, protocol, and typical packet size. In this project, we present an online updating technique for our osPCA. This updating technique allows to the efficiently calculate the approximated dominant eigenvector without performing Eigen analysis. Compared to the power Method algorithms, then the required computational costs and memory requirements are significantly reduced, and thus our method is especially preferable in online, streaming data, or large scale problems.**

***Index Terms*— Intrusion detection (ID), osPCA, NIDS, NNIDS, HIDS.**

I. INTRODUCTION

Network security consists of the provisions and policies. It's adopted by the network administrator to prevent and monitor unauthorized access, misuse, modification and network-accessible resources. Users choose an ID and password or other authenticating information that allows them access to their information and programs within their authority. Network security covers the variety of computer networks, both public and private. Security systems used in everyday job conducting transactions and communications among businesses, government agencies and individuals.

Intrusion Detection System helps the information systems prepare and deal with attacks. They accomplish by their collecting information from a variety of systems, network sources and analyzing the information for possible security problems. Intrusion detection provides the following:

- Monitoring and analysis the user and their system activities
- Auditing of system configurations and vulnerabilities
- Assessing the integrity of their critical systems and data files
- Statistical analysis of activities pattern based on their matching to known attacks
- Abnormal activity analysis
- Operating system audit

Grenze ID: 01.GIJET.5.1.18

There are three main components to Intrusion detection system

- **Network Intrusion Detection system (NIDS)**
- **Network Node Intrusion detection system (NNIDS)**
- **Host Intrusion Detection System (HIDS)**
- **Data mining approach used**

We were unable to configure K-Means Algorithm implemented in Weka in such a way that the distance between the centric of a cluster and data-point be used to calculate its probability of belonging to that cluster. On analysis this concluded that perhaps the attack that we used for their simulation (the SYN Flood attack) was affecting only a few fields of the data-points - like num bytes, num sync and the like. I tried to cluster on some of these specific attributes, and it increased accuracy too. However, this implied that the use of Domain Knowledge of this fields that would be affected, which is not the aim of anomaly detection methods. Next, i tried out the KNN technique for Outlier Detection. For each test data-point and distance from its Nearest Neighbor was computed. If this distance is found to be greater than some pre specified threshold, the point is marked as an Outlier. The next step was to establish this threshold. I obtained the threshold by finding out the distance of each training instance from its Nearest Neighbor (or average of distance from k- Nearest Neighbors). The maximum of this distance was estimated as the value for threshold.

II. System Design

In existing system, use the principal component analysis to identify the outliers. Principal component analysis (PCA) is a mathematical procedure that transforms a number of possible variables into a number of smaller variables. The objectives of the principal component analysis are reducing to dimensionality of the dataset but retain most of the original variability in the data. In this section we investigate the Eigen connection approach based on the principal component analysis. In our case, each connection and record corresponds to one vector of *n* variables. The procedure is following: The set of *n* different measures are collected in a vector is called as connection record vector representing the corresponding connection. But major drawback is we do not store the entire data matrix or covariance matrix, and thus our approach is especially of interest in online or large-scale problems.

- The major issue is to determine the trust metric based on a given set of attributes.
- It can't synchronize the trust level setting on various hubs when numerous ways cross with each other
- Threshold value to be treated as intruder and the path is rediscovered with the new threshold value and discarding the intruder node.

III. Module Description

A. Server Module

A server is a software and hardware of a system that responds to requests across a computer network to provide, or help to provide, a network service. If the servers can be run on a dedicated computer network, which is also often referred to as "the server", but many networked computers are capable of hosting servers operate within client-server architecture, if the server architecture computer programs running to serve the requests of other programs, the clients.

B. Gateway Module

In the network for an enterprise, a gateway node is often also acting as a proxy server and a firewall server. A gateway is often associated which knows where to direct a given packet of data that arrives at the gateway, and a switch, which furnishes the actual path in and out of the packet gateway. Gateways are also called as protocol converters, can operate at any network layer. The activities of gateway are more complex than that of the router as it communicates using more than one protocol.

- Lessens the quantity of dropped bundles

IV. System Description

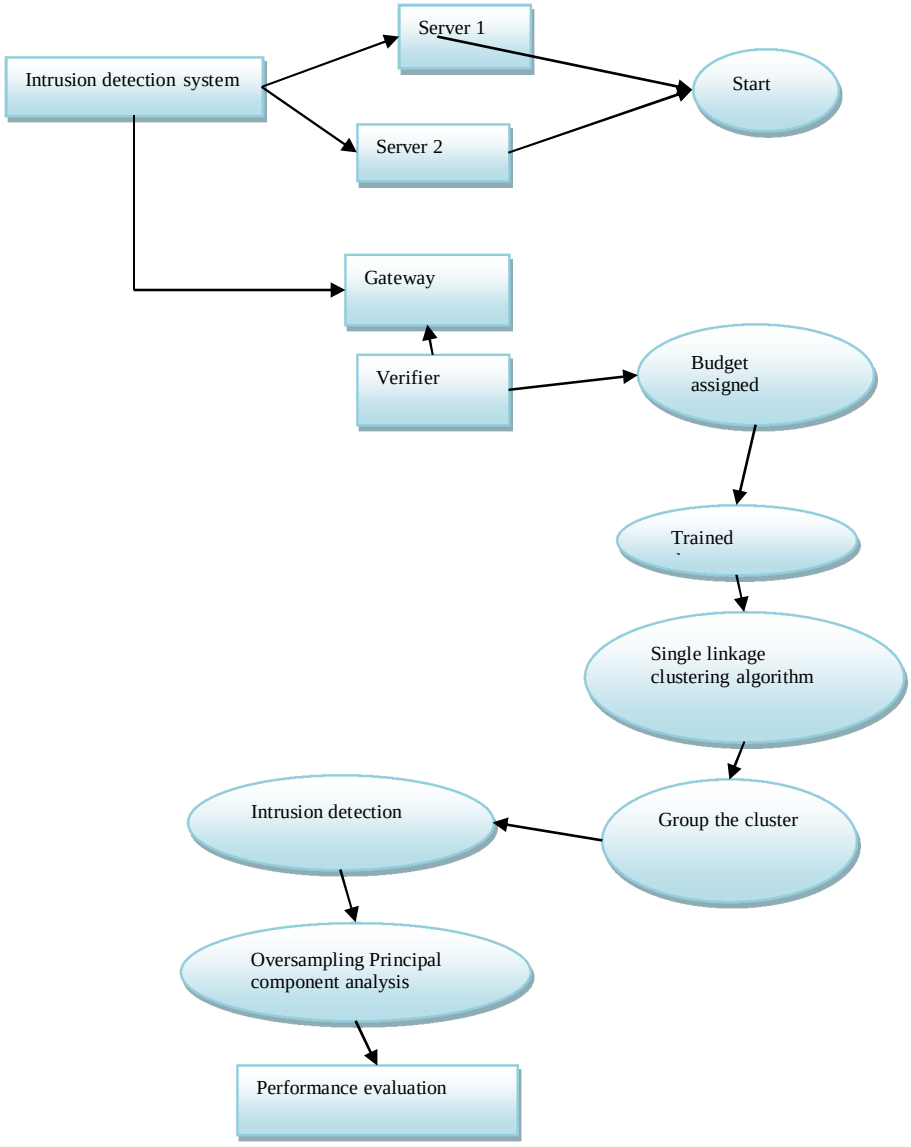

Fig. 3.1. Block diagram for PCA method

V. FLOW DIAGRAM

Level 0

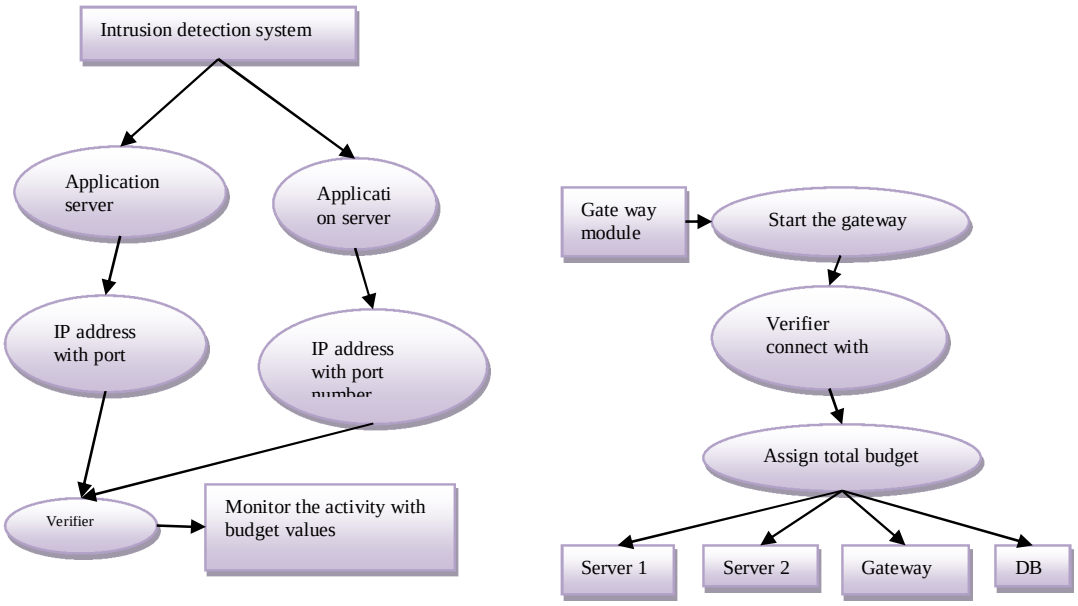

Fig. 3.2. Flow diagram for level

VI. METHODOLAGY

A. Cluster Analysis

Further used this datasets as synthetic datasets. This datasets analyzed with real network problems. These datasets are stored in the form of csv format. To create clusters from the input data instances and use the single linkage clustering. Single-linkage clustering is one of several technology of agglomerative hierarchical clustering. In the beginning of the network security process, each element is in a cluster of its own. The clusters are sequentially combined into larger clusters, until all elements are end up being in the same cluster. At each step, the two clusters are separated by the shortest distance are combined. The algorithm starts with the training datasets and computes the centric of the cluster.

Mathematically, the linkage method – the distance D(X, Y) between clusters X and Y – is described by the expression

$$D(X,y) = \min_{x \in X, y \in Y} d(x,y)$$

Where X and Y are any two set of elements considered as clusters, and d(x, y) denotes the distance between the two elements x and y.

B. Intrusion Detection

In this module we using OsPCA algorithm to analysis the attackers

Input: the first principal direction w of the normal data, the weight β and threshold h, and the new arriving instance x

Output: x is an outlier or not

1. Compute the adjusted first principal direction

$$\tilde{w} = \frac{\beta(\sum_{i=1}^{n}(y_i + x_i) + y_t x)}{\beta(\sum_{i=1}^{n}(y_i^2) + y^2)}$$

2. Compute the cosine similarity of w and $\tilde{w}$
3. Check the cosine similarity of w and $\tilde{w}$ and see if it is higher than the specified threshold h

Based on this algorithm we identify the attackers as outliers and it is feasible for large-scale and online problems. The OsPCA allow to avoid the duplicate instances in multiple items and then online OsPCA method efficiently determine the intrusion of each instances with less computation energy.

VII. Implementation and Results

A. Experimental Results

In our experiments, we evaluate our methods on real time datasets. We compare our methods PCAOD (only removing one instance in LOO), OPCAOD1 (with power method), and OPCAOD2 (with past algorithm) with LOF and fast ABOD. In our experiments,I use AUC to evaluate the suspicious outlier ranking in outlier detection phase. In our experiments, I have tried several parameters for these methods (r in OPCAOD1, k in LOF and fast ABOD, and β in OPCAOD2) and used the best parameter for each method respectively.

TABLE I. TABLE OF EXPERIMENTAL RESULTS

Attack type	Testing data size		TP rate	FP rate
	Normal	attack		
DOS	100	25	0.96	0.06
R2L	100	25	0.98	0.01
Probe	100	25	0.99	0.05

VIII. Coding and Screenshots

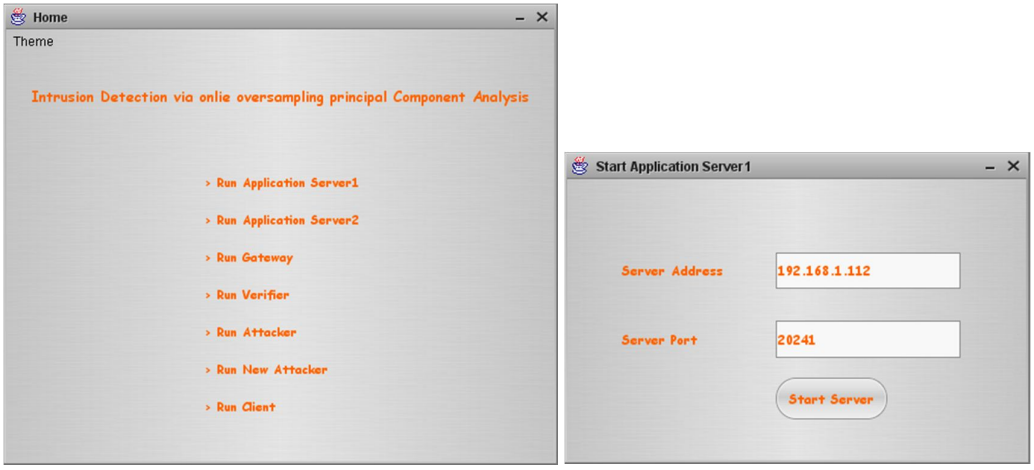

Fig.7.2 Run application server 1

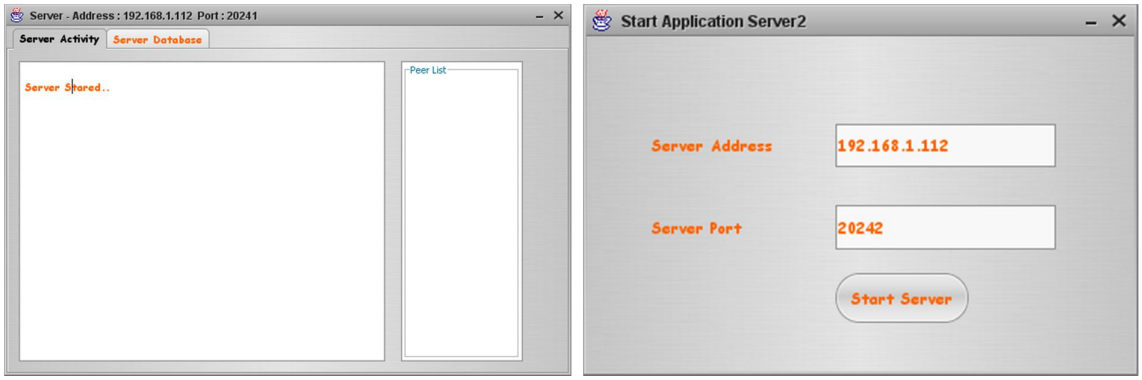

Fig.7.3 Run application server 2

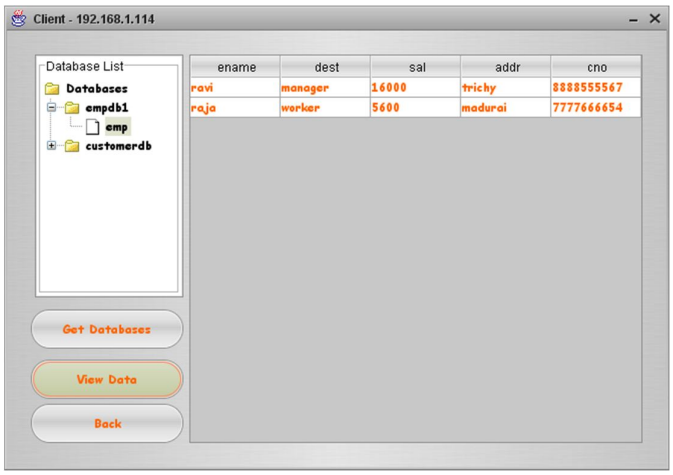

Fig.7.12 Output of Server 1 Database

IX. CONCLUSION

We conclude that the variation of principal directions caused by outliers indeed can help us to detect the anomaly. The proposed the over-sampling PCA can be used to enlarge the outlierness of an outlier. The online estimating principal directions in LOO is also proposed for reducing the computational loading and satisfying the on-line detecting demand The proposed OSPCA based anomaly detection is suitable for the extremely unbalanced data distribution in network security. Furthermore, our technology does not need to keep the entire covariance matrices during the online detection process.

REFERENCES

[1] Chapman and Hall, 1980. D. M. Hawkins, Identification of Outliers.
[2] V. Barnett and T. Lewis,1994. Outliers in statistical data. John Wiley& Sons
[3] M. Breunig, H-P. Kriegel, R. T. Ng, and J. Sander, 2000. 'LOF: Identifying density- based local outliers,' in Proceeding of the 2000 ACM SIGMOD International Conference on Management of Data.
[4] A. Lazarevic, L. Ert¨oz, V. Kumar, A. Ozgur, and J. Srivastava,2003. 'A comparative study of anomaly detection schemes in network intrusion detection,' in Proceedings of the Third SIAM International Conference on Data Mining.
[5] W. Wang, X. Guan, and X. Zhang,2004. 'A novel intrusion detection method based on principal component analysis in computer security,' in Proceeding of the International Symposium on Neural Networks.
[6] F. Angiulli, S. Basta, and C. Pizzuti, 2006. 'Distance-based detection and prediction of outliers,' IEEE Transactions on Knowledge and Data Engineering, vol. 18, no. 2, pp. 145–160.
[7] S. Rawat, A. K. Pujari, and V. P. Gulati,2006. 'On the use of singular value decomposition for a fast intrusion detection system,' Electronic Notes in Theoretical Computer Science, vol. 142, no. 3, pp. 215–228.
[8] W. Jin, A. K. H. Tung, J. Han, and W. W.ang, 2006 'Ranking outliers using symmetric neighborhood relationship,' in Proceeding of Pacific-Asia Conference on Knowledge Discovery and Data Mining.
[9] L. Huang, X. Nguyen, M. Garofalakis, M. Jordan, A. D. Joseph, and N. Taft, 2007. 'In-network pca and anomaly detection,' in Proceeding of Advances in Neural Information Processing Systems.
[10] Raghavendran, P. S., Asokan, R., & Vishnupriya, M. (2016) "Enhancing the Security against Flooding Attack in MANET for Medical Application", Asian Journal of Research in Social Sciences and Humanities, 6(6), 34-43.
[11] Karthikeyan, D., & Dharmalingam, M. (2013, February). Ant based intelligent routing protocol for MANET. in 2013, international conference on Pattern Recognition, Informatics and Mobile Engineering (pp. 11-16). IEEE.