IDS alert analysis using clustering & Artificial Neural Network

Nanaso S. Bansode*
Department of Computer Engineering, University of Pune

Thaksen J. Parvat
Department of Computer Engineering, University of Pune

***Abstract*— Intrusion Detection System (IDS) is used to supervise all tricks which are running on particular machine or network. Also it will give you alert regarding to any attack. However now a day's these alerts are very large in amount. It is very complicated to examine these attacks. We intend a time and space based alert analysis technique which can strap related alerts without surroundings knowledge and provide attack graph to help the administrator to understand the attack on host or network steps wise clearly and fittingly for analysis. A threat evaluation is given to discover out the most treacherous attack, which decrease administrator's time and energy in calculating huge amount of alerts. We are analyzing the network traffic in form of attack using Entity Threat Evaluation (ETE) which find out which particular host is attacked, Gadget Threat Evaluation (GTE) which tells us within that host which device is attacked, Network Threat Evaluation (NTE) which tells us which network is attacked, Hit Threat Evaluation (HTE) by giving input as dataset of attack. Main idea is that the distribution of different types of attacks is not balanced. The attacks which are not repeatedly occurs, the learning sample size is too small as compared to high-frequent attacks. It makes Artificial Neural Network (ANN) not easy to become skilled at the characters of these attacks and therefore detection precision is much worse. To solve such troubles, we propose a new technique for ANN-based IDS, Fuzzy Clustering (FC-ANN), to enhance the detection precision for low-frequent attacks and detection stability.**

Keywords— ***IDS,ETE,GTE,HTE,NTE,ANN,FC-ANN***

I. Introduction

Once numbers of computers are connected to each other by some medium which may be connection oriented or connectionless and they can go halves information between them then we label these computers are in network. Some attackers attack to particular host or network which is called as unauthorized client. These unauthorized users are not valid user. They have no authorization to access any host. Only authorized user can use the utilities which are on particular host to access.

Among the near-term of Internet era, network security has grow to be the key structure to web applications, such as online retail sales, online auctions, etc. Intrusion detection attempts to detect computer attacks by analyzing various data records observed in processes on the network. Detection precision and detection ability is two key which shows to appraise intrusion detection systems [8] (IDS).There is many techniques, there to hit upon intrusions. Along with these techniques, Artificial Neural Network (ANN) is one of the broadly used techniques and has been successful in solving many multifaceted practical problems. And ANN has been successfully useful into IDS.

Conversely, the core dilemma of ANN-based IDS [5] exists in two concepts:
1. Weaker level discovery stability.
2. Inferior level detection precision, especially for low-frequent attacks, e.g., Remote to Local (R2L), User to Root (U2R)
In favor of the above two aspects, the main motive is that the distribution of different types of attacks is unnecessary. For low-frequent attacks, the receptiveness sample size is too small as compared to high-frequent attacks. It makes ANN not simple to study the lettering of these attacks and therefore detection accuracy is much lower.
To solve the above two troubles, we offer a narrative approach for ANN-based IDS [8], FC-ANN [5], to improve the detection accuracy for low-frequent attacks and detection loyalty.
In order to get better the detection accuracy and detection constancy, many researches have been completed. In the near the beginning stage, the research focus lies in using rule-based specialist systems and statistical approaches. But when encountering larger datasets, the results of rule-based specialist systems and statistical approaches become not as good as. Thus a lot of data mining techniques have been introduced to solve the problem. Among these techniques, Artificial Neural Network (ANN) is one of the broadly used techniques and has been triumphant in solving much composite reasonable harm.

The common process of FC-ANN [13] approach has the following three phases. In the first stage, a fuzzy clustering technique is used to produce dissimilar training subsets. In the basis of different training sets, dissimilar ANNs are trained in the second stage. In the third phase, in arrange to do away with the errors of different ANNs, a meta-learner, fuzzy aggregation module, is introduced to learn once more and stick together the different ANN's results. The whole come near reflects the famous viewpoint ''divide and conquer". By fuzzy clustering [5], the complete training set is divided into subsets which have less number and lower difficulty. Thus the ANN can study each subset more rapidly, vigorously and accurately, especially for low-frequent attacks, such as U2R and R2L attacks.

II. Literature Survey

There are number of methods to examine IDS performance but it is very hard get desired intrusion pattern. Most of these methods are autonomous. Each method has its own restriction so combing advantages of diverse methods which will give us better result than individual one. Hence we propose IDS alert result improvement model help experts without difficulty to create alert classification model using large number of alerts.

The main reason of IDS alert is that collection and recognition are judgment more meaningful alert information and provide the information of relation between real alert to confirm system attacks. Some issues are derived from these purposes. How to decide fastidious analysis targets and data configure. How to unsoiled false alert correctly. How to find out attack methods and display particular data types for director to make policies.

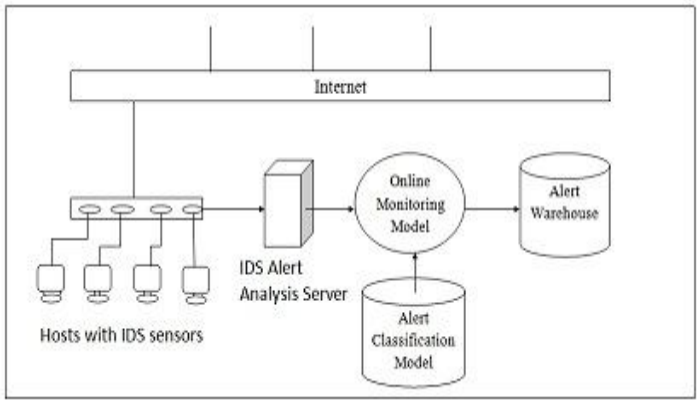

Fig.1: IDS alert analysis [12] architecture

In the online stage, IDS alert result development model for creating an alert result classification model consists of three phase alert preprocessing. In alert preprocessing phase we can collect all alert which are triggered by IDS [7] sensors. We can accumulate this alert into alert warehouse database, for alert preprocessing, we acquire one by one alert and convert it into alert sequence. In model construction phase, filtering and identification methods are planned for formation of classification rule classes to obtain away false alert and recognize each existing alert method. The standard alert behavior will receive firstly in online stage by sequential mining algorithm and used to decrease affect of noise, because normal behavior method happen periodically and frequently suspicious behavior method will happen in attacking environment.

The state Awareness give concentration to the information transforms technology between the conceptual data and the knowledge unstated by person. It has three levels to convert the data to the knowledge are perception, comprehension and prediction.

Its motive in network security is called Network Situation Awareness (NSA). predictable network security devices such as Intrusion Detection Systems (IDS), firewalls, and security scanners work autonomously and they cannot extend threat assessment of attacks which make a confront to administrators to understand how severe of an attack is. This lack of information result in numerous ambiguities when interpreting alerts and making decisions on proper responses.

To obtain in hand this problem, we suggest a time and space based alert analysis method which can associate related alerts without surroundings knowledge and offer attack graph to help the administrator understand the attack steps obviously and professionally. And threat evaluation is given to find the most hazardous attack, which further reduce administrator's time and energy in processing large amount alerts.

We present our own approach to automatically associate the alerts to generate easy attack graphs based on time and space constraint. In addition, we give an attack estimate method. We first recommend our own alert analysis method to correlate connected alerts and offer easy attack graph. Then, we provide an valuation function for likely threats. Via these considered methods, administrators can know the network condition and study how serious of an attack without checking personality alerts or estimate values.Here NSA wants to know where, when and how grave of an attack is, so we only need a little subset of alert fields. Using short alert message also reduce time and storage space.

Obtainable system of the IDS alert is shown in the below picture. Here IDPS[13] (Intrusion detection prevention System) is empowered with the sensors. Which sense the threat and finally send all to the IDPS[13] Management Server where we apply our threat evaluation and analysis method And Console server is for network admin to take decisions.

Method 1 Alert: An alert *al* is a seven tuple [13]
(*aid; srcip; dstip; srcport; dstport; type; time*)

autolid is an AUTO INCREMENT integer generated by database. It is used to classify each alert.

- ***Srcip:*** provides the source IP address. The operation Srcip(al) means obtain the source IP address of the alert *al*.
- ***Dstip:*** provides the destination IP address, the equivalent operation is Dstip(al).

- ***Srcport:*** provides the source port, the corresponding operation is Srcport(al)
- ***Dstport:*** provides the destination port, the equivalent operation is Dstport(a).
- ***Type:*** provides the alert's type,it is a short string which give a simple description of the attack, the corresponding operation is Type(a).
- ***Time:*** provides time of the alert generate, the corresponding operation is Time(a).

IDS sensors will take all alerts coming from online stage and store it into alert warehouse. In analysis phase we are taking one by one alert from dataset, first convert that alert into alert sequence. Now we are mapping converted sequence with our rule classes which we will implement.

Time and Space Restriction Analysis (TSRA)[13]

It is known which is normal that a successful attack usually has several steps. The attacker may use scan tackle to get the target network information firstly. After finding weak point of the network, the attacker will concentrate on certain devices, and start certain attack steps. These attack steps are linked, thus their corresponding alerts are also associated. We correlate the related alerts to an attack scenario based on time and space relations.

Two alerts *alti; altj* , if they are related, they usually have
certain time and space relations as follows:
1 Srcip(*ai*)=Srcip(*aj*), Dstip(*ai*)=Dstip(*aj*),
Time(*ai*) <Time(*aj*).
2 Dstip(*ai*)= Srcip(*aj*), Time(*ai*) <Time(*aj*).

In the first state, a series of attacks has the same target, and the target device suffer from these attacks one by one. In the second situation, a former attack intruded a device, then the attacker uses the device as a launch pad to attack other device which maybe the final target or at a halt another facilitator.

We heave out these correlated attack-pairs based on the two limitations, and then fasten together together these pairs to a complete attack graph. Algorithm 1 shows the method of correlating two isolated alerts to an alert-pair.

Algorithm 1: Association of Isolated Alerts to Alert-Pair
INPUT: entity hyper-alerts a1,a2……an
OUTPUT: set of alert-pairs (alti,altj) denoted APs.
Let TW be the time-window which is set by administrator.
Let HyperAlert be the hyper-alert table.
Let AlertPairs be the alert-pairs table.
1. for all the hyper-alerts in HyperAlert do
2. if Srcip(alti) = Srcip(altj) and Dstip(alti) = Dstip(altj) and Time(alti) <Time(altj) and Time(altj) – Time(alti) < TW Then
3. put (alti,altj) into Alert-Pairs.
4. if Dstip(alti) = Srcip(altj) and Time(alti) < Time(altj) and Time(altj) – Time(alti) < TW Then
5. Put (alti,altj) into Alert-Pairs.
Then, we correlate these alert-pairs to an attack graph as

Algorithm 2 : Attack Graph Generation
INPUT: set of alert-pair (alti,altj) - APs.
OUTPUT: attack graph G(N,E)
Put every hyper-alert ai of APs into node set N;
Put every alert-pair (alti,altj) of APs into edge set E;
1. For every edge (ni,nj) do
2. If there is a indirect path ni,…..,nk,….nj then
3. Remove the edge (ni, nj) from edge set E
4. Return G(N,E)

III. Implementation details

Total implementation is divided into four phases as Alert analysis, Alert vector, ANN [13] function, and Prediction intrusion. In Alert Analysis phase we are passing dataset which contains the all information about arriving packets. We are momentary these dataset as input to alert analysis phase.

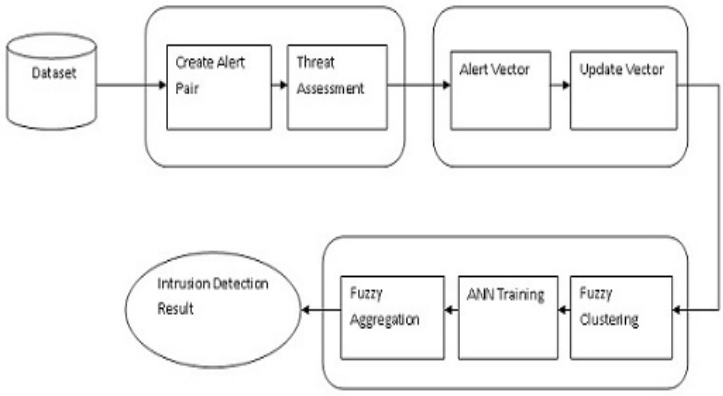

Fig.2: Proposed system architecture

By using this dataset we are creating alert pairs by allocating time slot for each arriving packet and create graph using generation of graph algorithm. Now we are ready to hit upon out where attack is going on. We have four threat evaluation techniques for evaluation of attack these are Entity Threat Evaluation, Hit Threat Evaluation, Gadget Threat Evaluation and Network Threat Evaluation. Just the once we found which network is attacked then we convert that attack into number of sets means just we are converting that attack in to sequence and then divide it into number of parts to get multiple sets.

At this moment we are going to next phase which is Alert Vector which has size unlimited. Its size is unlimited because if any new attack is occurred then we have to update the vector. Vector is nothing but is one type of array where we are like divide and conquer. First diving attack into number of parts and then take feedback from each part, at last combine result to get which attack is occurred in which network in which device. To recommend our approach we are using Data captured at real router in excel format dataset as an input our project. The Work is divided into mainly three parts as shown in below figure.

Storing the information about set which are created on Alert Analysis phase. Again those vectors are diving into sub vectors for further analysis.

Whatever we have sub vectors just we are passing these as a input to next phase which is ANN Function. ANN will take training set as input and will create small clusters for prediction of intrusions. The function of ANN is to divide available set into number of set as small cluster and will map each cluster with source and destination input packets and will take feedback from each cluster.

Task 1: Fuzzy cluster technique:

The main thing of fuzzy cluster technique is to dividing wall a known set of data into clusters, and it should have the following properties: homogeneity within the clusters, relating to data in same cluster, and heterogeneity between clusters, where data belonging to different clusters should be as dissimilar as possible.

All the way through fuzzy cluster technique, the training set is clustered into several subsets. Due to the fact that the size and complexity of every training subset is abridged, the effectiveness and efficiency of subsequent ANN module can be enhanced.

The fuzzy cluster is composed of the following steps:

Step 1: Initializing Data Sets.
Step 2: manipulative centers vectors
Step 3: Updating Vectors
Step 4: Creating Subset Vectors

Task 2: Artificial Neural Network

ANN component aims to learn the sample of every subset. ANN is a in nature stirred form of distributed estimation. It is collected of simple processing units, and links between them. In this study, we will employ classic feed-forward neural networks skilled with the back-propagation algorithm to imagine intrusion.

A feed-forward neural network has an input layer, an output layer, with one or more concealed layers in between the input and output layer. The ANN functions as follows we will see each node i in the input layer has a signal xi as network's input, multiplied by a weight value between the input layer and the hidden layer.

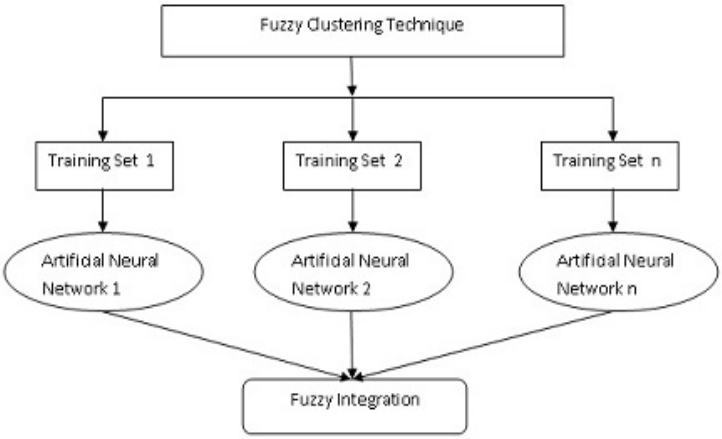

Fig.3: Fuzzy cluster Method

Task 3: Fuzzy Integration

The most important target of fuzzy aggregation[5] module is to aggregate dissimilar ANN's result and reduce the detection errors as every ANN_i in ANN module only understand from the subset TRi. Because the errors are nonlinear, in order to accomplish the purpose, we use another new ANN to study the errors as follows we see stepwise

Step 1: The total training set TR as data to input the every trained ANN_i and obtain the outputs

Step 2: Summarize the input for new ANN

Step 3: Plan the new ANN. We can use Y_{input} as input and use the absolute training set TR's class label as output to prepare the new ANN.

IV. Results

To compute the act of FC-ANN approach, a sequence of experiments on Excel data of type array where the data is in use at router side. The dataset contains about five million connection report as preparing data and about two million connection report as check data. And the dataset contain a set of 41 features consequential from each connection and a label which specify the position of connection report as either normal or specific attack type. These facial appearance have all forms of continuous, discrete, and symbolic variables, with significantly varying ranges falling in four categories as follows:

(1) The whole category contains of the intrinsic features of a connection, which include the essential features of individual TCP connections. The duration of the connection, the type of the protocol (TCP, UDP, etc.), and network service (http, telnet, etc.) are some of the features.

(2) The given Category content features within a connection recommended by domain knowledge are used to assess the payload of the original TCP packets, such as the number of failed login attempts.

(3) The same host features observe established connections in the earlier period two seconds that have the same destination host as the current connection, and calculate the statistics related to the protocol behavior, service, etc.

(4)The alike same service features inspect the connections in the what went before two seconds that have the same service as the current connection.

Attacks go down into four categories:

(1) Denial of Service (DoS): making some computing or memory resources too busy to agree to legitimate users access these resources.

(2) Probe (PRB): host and port scans to get together information or get known vulnerabilities.

(3) Remote to Local (R2L): unauthorized access from a remote machine in order to utilize machine's vulnerabilities.

(4) User to Root (U2R): unauthorized access to local super user (root) privileges using system's vulnerability.

Excluding as the number of occurrence for the U2R, PRB, and R2L attacks in the training set and test set is every low, these quantities is not enough as a standard performance measure. Hence, if we use these quantities as a measure for testing the performance of the systems, it could be biased. For these reasons, we give the accuracy, recall, and F-value which are not dependent on the size of the training and the testing samples. They are definite as follows:

Precision = TP/ (TP+FP)
Recall =TP/(TP+FN)
Where TP, FP, and FN are the number of true positives, false positives, and false negatives.

V. Conclusion & Future work

Prevention of security breaches completely using the on hand security technologies is unfeasible. As a result, intrusion detection is an essential component in network security. IDS offers the possible advantages of reducing the manpower needed in monitoring, increasing detection effectiveness, providing data that would otherwise not be obtainable, helping the information security community learn about new vulnerabilities and providing officially authorized proof. In this paper, we propose a new intrusion detection approach, called FC-ANN, based on ANN and fuzzy clustering. Through fuzzy clustering technique, the mixed training set is divided to several homogenous subsets. Thus difficulty of each sub training set is reduced and accordingly the detection performance is improved. The experimental results using the KDD CUP 1999 dataset demonstrates the efficiency of our new approach specially for low-frequent attacks.

References

[1] Anderson, J. P. (1980). Computer security threat monitoring and surveillance. Technical Report, Fort Washington, PA, USA.

[2] Anderson, J. (1995). An introduction to neural networks. Cambridge: MIT Press. Axelsson, S. (2003). The base-rate fallacy and the difficulty of intrusion detection. ACM Transaction on Information and System Security, 3, 186–205.

[3] Barbard, D., Wu, N., & Jajodia, S. (2001). Detecting novel network intrusions using Bayes estimators. In: Proceedings of the first SIAM international conference on data mining (pp. 1–17).

[4] Beghdad, R. (2008). Critical study of neural networks in detecting intrusions. Computers and Security, 27(5-6), 168–175.

[5] Bezdek, J. C. (1973). Fuzzy mathematics in pattern classification. PhD thesis, Applied Math. Center, Cornell University Ithaca.

[6] Chen, Y. H., Abraham, A., & Yang, B. (2007). Hybrid flexible neural- tree-based intrusion detection systems. International Journal of Intelligent Systems,22(4), 337–352.

[7] Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation. Journal of Intelligent and Fuzzy Systems, 2, 267-569

[8] Depren, O., Topallar, M., Anarim, E., & Ciliz, M. K. (2005). An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks. Expert Systems with Applications, 29(4), 713–722.

[9] Dokas, P., Ertoz, L., Lazarevic, A., Srivastava, J., & Tan, P. N. (2002). Data mining for network intrusion detection. Proceeding of NGDM, 21–30.

[10] Endorf, C., Schultz, E., & Mellander, J. (2004). Intrusion detection and prevention. California: McGraw-Hill.

[11]Gordeev,M(2000).Intrusion detection: Techniques and approaches.<http://www.gosecure.ca/SecInfo/library/IDS/ids2.pdf (accessed March 2009).

[12] Yan Zhang, Shuguang Huang, Yongyi Wang" IDS Alert Classification Model Construction Using Decision Support Techniques" 2012 International Conference on Computer Science and Electronics Engineering.

[13] Prof.TambeShital B.,Prof. Sonkar S.K." Analysing Various Alerts & Evaluating Threat Techniques In NSA" International Journal of Scientific & Engineering Research Volume 4, Issue 2, February-2013 ISSN 2229-5518.