A Hybrid Approach for Intrusion Detection Based on Pattern Matching and Clustering using Neural Network

Rajni Tewatia
M.Tech Scholar of computer science & Engineering, BSAITM, Faridabad, India.

Asha Mishra
Department of computer science &Engineering, BSAITM, Faridabad, India.

Abstract – Network security is becoming an important issue. Due to increase in the knowledge of intruders and attackers different security tools fails. With the advance of technology new tools are developed and new approaches are embedded in previously developed approaches. So, Intrusion Detection System (IDS) comes into role.IDS provide second line of defense which is not in the limit of firewall and others. . In this thesis a Hybrid intrusion detection method is proposed that is based on Misuse Detection Approach and Anomaly Detection. For Misuse Detection Pattern Matching algorithm is used. Attack pattern are matched with already stored pattern. In second phase for anomaly detection approach clustering is used. But the clustering is based on Neural Network. Through the Neural Network the IDS become self-adapting in nature.

Index Terms – Intrusion Detection, Clustering, Neural Network, Network Security.

1. INTRODUCTION

At present networking revolution is main part of the communication era and internet is changing the computing. The goal of interconnected computer system is to more efficiency and better information exchange. The number of attacks is increases because of the integration of the computer system, which faces some attacks. An attack is a realization of threat, to find and exploit the system vulnerability. Security of a network is always important, which monitors all network traffic passing on the segment. The act of detecting actions that attempts to compromise the confidentiality, integrity or availability of a resource to bypass its security mechanisms. Intrusion is the act of violating the security policy or legal protections that pertain to an information system. In network security basic two types of tools and techniques are used. Reactive tools and techniques are those that generate some kind of response or alert after detecting some unidentified behavior. They are called reactive because first activity occurs and they react accordingly e.g. Firewall and IDS. Proactive tools and techniques detect attack before it happened. Intrusion Prevention Systems are the example of such tools. A solution to enhance the overall security of the networks is to increase the security layers with intrusion detection systems. An intrusion detection system (IDS) is a device, typically a designated computer system, which monitors activity to identify malicious or suspicious alerts. It is placed inside an organization to monitor what occurs within the network of the organization. Basic components of IDS are Data Collection, Analyzer and Response Engine. Various technologies are used in analyzer part of IDS like Clustering, Neural Network, Pattern Matching, Data mining, and Rule Based System. Intrusion Detection is of two types Misuse and anomaly detection. Misuse detection systems compare current activities of the host or the network monitored with "signatures" of known attacks. So, this system detects only known attacks [methodologies]. Anomaly detection assumes that an attack will always reflect some deviation from normal patterns, which is designed to capture any deviations from the established profiles of the system normal behavior. Anomaly detection can detect new and unknown intrusion, but it has the shortcoming of false alarm rate [expert and clustering]. So, these two are combined to get better performance. In this a hybrid system is proposed which work in two phases, in first phase Pattern Matching is used and in second phase clustering using neural network(NN) is used.

2. RELATED WORK

De-gang Yang, Chun-yan Hu, Yong-hong Chen [1] presented a framework of cooperating intrusion detection based on clustering analysis and expert system. This framework integrates misuse detection that has higher detection rate and misuse detection that can detect new and unknown intrusion.

K. Prabha, S.Sukumaran[2] explained that the IDS uses string matching to compare the payload of the network packet and/or flow against the pattern entries of intrusion detection rules.

G.V. Nadiammai, M.Hemalatha [3] proposed that data mining has been popularly recognized as an important way to mine useful information from large volumes of data that are noisy, fuzzy & random and different clustering algorithms with their complexity.

Karuna Katariya, Rajanikanth Aluvalu [4] explained requirement of clustering is a Scalability, Ability to deal with different types of attributed, Discovery of clusters with arbitrary shape; Minimal requirements for domain knowledge of determine input parameters, Ability to deal with noisy data.

ShengYi Jiang, Xiaoyu Song, Hui Wang, Jian-Jun Han, Qing-Hua Li [5] presented a clustering-based method for unsupervised intrusion detections[5]The data classification is performed by an improved nearest neighbor (INN) method. A powerful clustering-based method is presented for the unsupervised intrusion detection.

Leonid B. Litinskii, Dmitry E. Romanov [6] presented an algorithm of clustering of many-dimensional objects, where only the distances between objects are used.

Centers of classes are found with the aid of neuron-like procedure with lateral inhibition.

Alan Bivens, Chandrika Palagiri, Rasheda Smith, Boleslaw Szymanski, Mark Embrechts [7] ids nn modular explored network based intrusion detection using classifying, self-organizing maps for data clustering and MLP neural networks for detection.

3. PROPOSED METHOD FOR INTRUSION DETECTION

In this section, a rough framework of hybrid intelligence intrusion detection based on clustering analysis and pattern matching is proposed that can improve detection rate and decrease false alarm rate.

3.1 Architecture of Proposed method

The architecture shown in Fig.1.1 has been developed with these goals in mind.

The architecture is consisting of Preprocessing of Data, Misuse Detection Based Pattern Matching phase, Knowledge Base, Preprocessing of data for clustering, Anomaly Detection based Neural Network Clustering Analysis, Pattern formation for detected Intrusions and Response Engine.

Each unit performs the following function.

- Preprocessing of Data

Data that come from Network stream is not directly used for analysis. It must be processed first in a format suitable for intrusion detection.

- Misuse Detection based Pattern Matching

Here the IDS uses Pattern matching to compare the payload of the network packet and/or flow against the pattern entries of intrusion detection rules stored in knowledge base. Pattern matching generally consists of finding a substring (called a pattern) within another string (called the text).These string matching algorithms are used to inspect the content of packets and identify the attacks signature in IDS.

- Neural Network based Clustering Analysis

This module performs clustering which is an anomaly based detection system. This method is used through this unknown intrusions can be detected. Neural network is used to make the system automatic and self learning. Neural Networks are good at being able to classify unseen data points. Clustering is done with the help of neural network to improve the performance

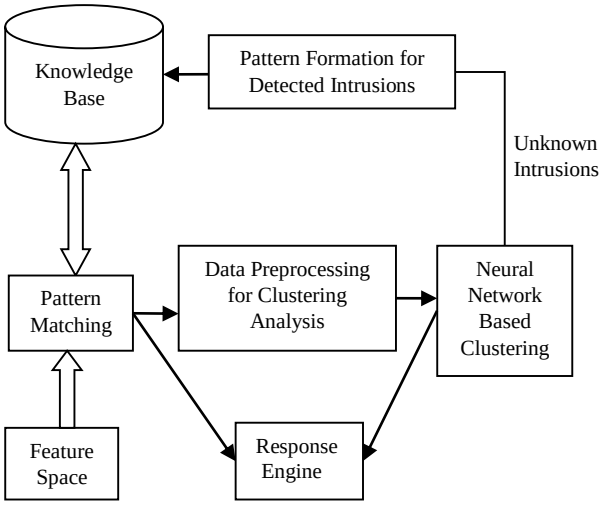

Fig. 1.1 Architecture of Proposed Method

- Pattern Formation for Detected Intrusion

As unknown intrusion are detected by clustering analysis. These unknown intrusions must be stored into knowledge base for future processing made by pattern matching. This step helps pattern matching phase a lot.

- Response Engine

If any Intruder is detected by the detection system then, this is reported by response engine. It is on the administrator that what action he would take, either he block the packet or destroy it depending upon the policy.

- Knowledge Base

This contains patterns for matching with incoming data. It contains two types of patterns one previously defined by the system security officer and second updated by intrusion detected by clustering analysis.

3.2 Single Keyword Pattern Matching Algorithms for IDS

The Brute Force algorithm requires no preprocessing of the pattern. In the Karp-Rabin Algorithm the main idea is that instead of using comparisons it involves mathematical computations which more specifically extends to the notion of hashing. So both of these are combined to compare two strings.

The two phases of the proposed algorithm are (i) preprocessing phase, (ii) searching phase.

Step1: In the first phase, algorithm performs the same preprocessing phase as in the existing two algorithms. It prepares the hash function used in KR algorithm and shift value for each character.

Step2: The process of computing hash functions for the patterns and text window are exactly the same as the process of creating them in the existing KR algorithm.

Step 3: After the preprocessing phase has finished, the comparison start between the text and pattern by comparing the numerical value of pattern hash and window text hash.

Step 4: Whether, if the two hash value are not identical then the algorithm perform the shifting.

Step 5: Next Shift to the right according to shift value if possible. This will speed up the algorithm during the comparison process and produce result if exist.

The process continues until all characters in the text are being compared and whether the mismatching or matching is found.

3.3 Clustering and Neural Network

The process of organizing similar objects into groups is called clustering. In a cluster there may be many groups according to the dataset it differs. Clustering is an unsupervised learning technique.

- Training of dataset through Neural Network

Before actual processing data is trained using neuron like procedure here. Data set are trained by neural network. In neural network input is provided, weights are adjusted through training and desired output is produced at the end.

Training algorithm is divided into the following steps.

Initial Conditions

For a given set of *m*-dimensional points

$\{X_i\}_1^N \in \mathbf{R^m}$ **i**= 1, 2,..., *N* .

Data clustering aims at identifying clusters as more densely populated regions in the space Rm. m is number of measurable features on the basic of which clusters are formed. Then calculate a quadratic (N× N) matrix of Euclidean distances between them: D = (D_{ij}). In what follows these distances D_{ij} are needed only. Suppose that in each point xi there is a neuron with initial potential $P_i(0)$, which will be defined later.

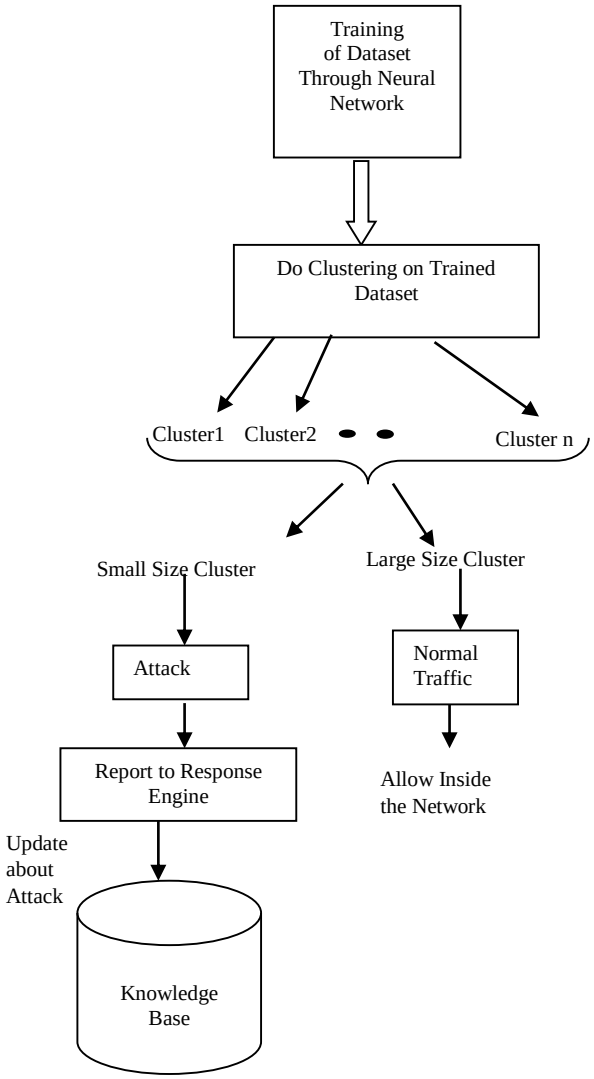

Fig. 1.2 Clustering Using Neural Network

The distance between points p and q is defined as

$$D(p,q) = \sqrt{\sum_{i=1}^{m} dif(p_i, qi)^2/m}$$

For numerical attributes, dif (pi, qi) = |pi-qi|

Step 1: For a fixed interaction threshold T>0 let us set the value of a connection w_{ij} between ith and jth neurons as

$$w_{ij} = \begin{cases} \dfrac{T^2}{D^2_{ij}+T^2} & \text{when } D_{ij}<=T, \\ 0 & \text{when } D_{ij}>T. \end{cases}$$

As there are no connections between neurons, if the distance between points is greater than T.

*Step 2:*Let us set initial Potential (energy)of each neuron to be

$$P_i(0)= \sum_{i=1}^{N} W_{ij}>=1$$

Neurons, which are inside whole region of the points, have large initial potenial, because they have more nonzero connections than neurons at the periphery of region.

*Step 3:*Start the lateral inhibition process:

Where α is a parameter that characterizes the speed of inhibition. It is easy to see that during process a neuron with large initial potential "takes away" potential from neurons with whom it interacts and whose potential are less. The potential of these surrounding neurons decrease steadily.

Step 4: If during the transmitting process the activity of a neuron becomes negative $P_i(t)<0$, get $P_i=0$, and eliminate this neuron from the transmitting process (it has nothing to give away).It is clear that little by little the neurons from the periphery of region shall drop. This means that step by step the neurons from the periphery will leave the field. Gradually, there is a situation, when only some far from each other non-interacting neurons remain. Subsequent inhibition is impossible and the procedure stops.

Step 5: Suppose as a result of the inhibition process K neurons remain far away from each other. The input points xi corresponding to these neurons will be called the centers of the cluster.

- Descriptions of the steps of Training through Neural Network

In Step1: Weights are adjusted according to the formula given above. In neural network learns through the process of adjustment. Threshold is defined.

In Step 2, Step 3 and Step 4: the process of lateral inhibition is discussed. Lateral inhibition is the capacity of an excited neuron to reduce the activity of its neighbors. Lateral inhibition disables the spreading of action potentials from excited neurons to neighboring neurons in the lateral direction.

Step 5: At the end of the training, a set of clusters is formed, with the same center of cluster x_i which is called as threshold center.

- Clustering Algorithm

Step 1: Initialize the set of clusters S, containing previously formed clusters.

Step 2: Read a new data, If no data are left in the database, go to step 6, otherwise read a new data p.

Step 3: Find the cluster Ci in S that is closest to p. namely, find a cluster Ci in S, such that

$d(p,C_i)<=d(p,\overline{C})$ for all $\overline{C}$ in S. C is set of S-C_i clusters(other than C_i clusters in the set S), d is Euclidean distance as defined in training step.

Step 4: If $d(p, C_i) > x_i$, go to step 3. x_i is previously found center of cluster trough training.

Step 5: Merge p into cluster C_i and modify the cluster C_i. Go to step 2.

Step 6: Stop.

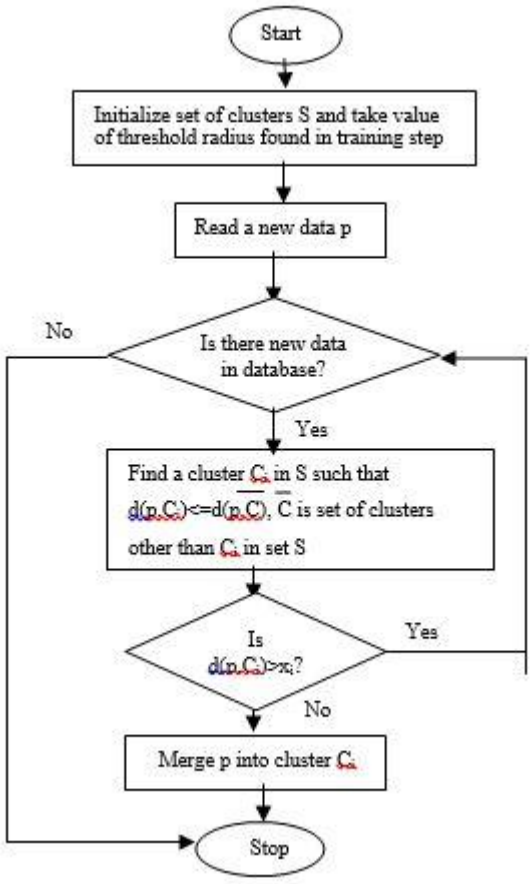

Fig. 1.3 Flow Diagram of Clustering Algorithm

4. CONCLUSION

Although a lot has been done in the field of intrusion detection system but a lot of work still has to be done. As with recent advances and emerging technologies, increase in the types of intruder goals, intruder ability, tool sophistication and diversity as well as the use of more complex and new attack scenario to the IDS increases. Here two methods are used in the detection

part of intrusion detection system. These two approaches are used because a hybrid system is proposed. In hybrid system two different approaches are combined to take benefits from both the approaches. Because alone with one approach there will be no better performance and better detection. With this proposed architecture both known and unknown attacks are detected.

It integrates the virtues of both misuse detection and anomaly detection to improve the detection performance. Moreover it converts unknown intrusion to known intrusion, hence improves the detection accuracy and efficiency. So, intrusion Detection system must not be done at the cost of system performance.

REFERENCES

[1] De-gang Yang, Chun-yan Hu, Yong-hong Chen "A framework of cooperating Intrusion Detection based on Clustering analysis and expert system", lnfoSeeu04, November14-16,2004, Pudong, Shanghai, China. Copyright 2004 ACM ISBN: 1-58113-955-1.

[2] K. Prabha, S.Sukumaran, "Improved Single Keyword Pattern Matching Algorithm for irntrusion Detection System", International Journal of Computer Applications (0975 – 8887) Volume 90 – No 9, March 2014.

[3] G.V. Nadiammai, M.Hemalatha," An Evaluation of Clustering Technique over Intrusion Detection System", ICACCI'12, August 3-5, 2012, Chennai, T Nadu, India. Copyright 2012 ACM 978-1-4503-1196-0/12/08.

[4] Karuna Katariya ,Rajanikanth Aluvalu "Agglomerative Clustering in Web Usage Mining: A Survey", International Journal of Computer Applications (0975 – 8887) Volume 89 – No 8, March 2014.

[5] ShengYi Jiang, Xiaoyu Song, Hui Wang, Jian-Jun Han, Qing-Hua Li "A clustering-based method for unsupervised intrusion detections", www.elsevier.com/locate/patrec, 2005 Elsevier B.V.

[6] Leonid B. Litinskii, Dmitry E. Romanov," Neural Network Clustering Based on Distances Between Objects" Institute of Optical-Neural Technologies Russian Academy of Science, Moscow litin@iont.ru, demaroman@yandex.ru

[7] Alan Bivens, Chandrika Palagiri, Rasheda Smith, Boleslawszymanski, Mark Embrechts "Network-Based Intrusion Detection eUsing Neural Networks" Proc. Intelligent Engineering Systems through Artificial Neural Networks ANNIE-2002, St. Louis, MO, vol. 12, ASME Press, New York, NY, 2002, pp. 579-584

[8] Wenke Lee Salvatore J. Stolfo, "A Framework for Constructing Features and Models for Intrusion Detection Systems" ACM Transactions on Information and System Security, Vol. 3, No. 4, November 2000.

[9] G.V. Nadiammai, M.Hemalatha," An Evaluation of Clustering Technique over Intrusion Detection System", ICACCI'12, August 3-5, 2012, Chennai, T Nadu, India. Copyright 2012 ACM 978-1-4503-1196-0/12/08.

[10] A.S. Aneetha, T.S. Indhu, Dr. S. Bose," Hybrid Network Intrusion Detection System Using Expert Rule Based Approach" CCSEIT-12, October 26-28, 2012, \Coimbatore [Tamil nadu, India] Copyright © 2012 ACM 978-1-4503-1310-0/12/10.

[11] Baykara, Muhammet, and R. Das. "A survey on potential applications of honeypot technology in intrusion detection systems." *International Journal of Computer Networks and Applications* 2.5 (2015): 1-9.

Authors

Rajni Tewatia received B.TECH degree in Computer Science and Engineering with Hons. from Maharshi Dayanand University in 2013 and is persuing M.Tech. in Computer Engineering. Presently, her areas of interests are Artificial Neural Network, Soft Computing, Analysis & Design of Algorithm, Network Security and Operating System.

Asha Mishra received B.E. degree in Computer Science & Engineering from NIT, Assam and M.Tech in Computer Science from A.I.T.M, Palwal. Presently, she is working as Senior Lecturer in Computer Engineering department in B.S.A. Institute of Technology & Management, Faridabad. Her areas of interests are DBMS, Analysis & Design of Algorithm and Network Security.