Volume 3, Issue 6, June 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Detecting Network Intrusions Using new Feature Representations

Greeshma K , Merin Meleet
Department Of CSE&
Calicut university , India

Abstract— Our network is facing a rapidly evolving threat landscape full of modern applications, exploits and attack strategies that are capable of avoiding traditional methods of detection. Threats are delivered via applications that dynamically, use non-standard ports, tunnel within other applications or hide within proxies, other types of encryption. Additionally, enterprises are exposed to targeted and customized malware, which can easily pass undetected through traditional antivirus solutions. .To provide the effective result for detecting intrusions, this process introduces a new feature representation approach by cluster centers and nearest neighbors. In this process, two distances are measured and summed. The first one is based on the distance between each data sample and its cluster center, and the second distance is between the data and its nearest neighbor in the same cluster. Then, this new and one-dimensional distance based feature is used to represent each data sample for intrusion detection.

I. INTRODUCTION

An intrusion detection system (IDS) is a device or software application that monitors network or system activities for malicious activities or policy violations and produces reports to a management station. Some systems may attempt to stop an intrusion attempt but this is neither required nor expected of a monitoring system. Intrusion detection and prevention systems (IDPS) are primarily focused on identifying possible incidents, logging information about them, and reporting attempts. In addition, organizations use IDPSes for other purposes, such as identifying problems with security policies, documenting existing threats and deterring individuals from violating security policies. IDPSes have become a necessary addition to the security infrastructure of nearly every organization. IDPSes typically record information related to observed events, notify security administrators of important observed events and produce reports. Many IDPSes can also respond to a detected threat by attempting to prevent it from succeeding. They use several response techniques, which involve the IDPS stopping the attack itself, changing the security environment (e.g. reconfiguring a firewall) or changing the attack's content. Intrusion detection (ID) is a type of security management system for computers and networks. An ID system gathers and analyzes information from various areas within a computer or a network to identify possible security breaches, which include both intrusions (attacks from outside the organization) and misuse (attacks from within the organization). ID uses vulnerability assessment (sometimes refered to as scanning), which is a technology developed to assess the security of a computer system or network. Intrusion detection is the detection of user behaviors in the network that deviates from the organizations network security policy. The goals of network intrusion detection are to identify, categorize and possibly respond to malicious or suspicious activities. There are basically two types of intrusion detection systems namely anomaly detection and misuse detection. Anomaly detection system first learns normal system activities and then alerts all system events that deviate from the learned model and the misuse detection uses the signature of attacks to detect intrusions by modeling attacks.

Machine learning refers to a system capable of the autonomous acquisition and integration of knowledge. This capacity includes learning from experience, analytical observation, and so on, which results in a system that can continuously self-improved and thereby offers increased efficiency and effectiveness. Machine learning usually refers to the change in systems that perform tasks associated with artificial intelligence (AI). The main goal of machine learning is to design and develop algorithms and techniques that allow computers to learn. In general, there are two types of machine learning techniques, which are supervised and unsupervised learning techniques.

Therefore, in this paper, we propose a novel feature representation for effective and efficient intrusion detection. It is based on combining cluster centers and nearest neighbours. Particularly, given a dataset the k-means clustering algorithm is used to extract cluster centers of each pre-defined category. Then, the nearest neighbor of each data sample in the same cluster is identified. Next, the sum of the distance between a specific data and the cluster centers and the distance between this data and its nearest neighbor is calculated. This results in a new distance as the feature to represent the data in the given dataset. Consequently, the new dataset containing only one dimension (i.e. distance based feature representation) is used for weighted k-nearest neighbor classification, which allows for effective and efficient intrusion detectio

II. LITERATURE SURVEY

Writing a basic article on network security is something like writing a brief introduction to flying a commercial airliner. Much must be omitted, and an optimistic goal is to enable the reader to appreciate the skills required. The first question to address is what we mean by "network security." Several possible fields of endeavor come to mind within this broad topic, and each is worthy of a lengthy article. To begin, virtually all the security policy issues apply to network as

well as general computer security considerations. In fact, viewed from this perspective, network security is a subset of computer security. The art and science of cryptography and its role in providing confidentiality, integrity, and authentication represents another distinct focus even though it's an integral feature of network security policy. The topic also includes design and configuration issues for both network-perimeter and computer system security. The practical networking aspects of security include computer intrusion detection, traffic analysis, and network monitoring. This article focuses on these aspects because they principally entail a networking perspective

A. ***Hybrid flexible neural-tree-based intrusion detection systems***

An intrusion is defined as a violation of the security policy of the system, and, hence, intrusion detection mainly refers to the mechanisms that are developed to detect violations of system security policy. Current intrusion detection systems (IDS) examine all data features to detect intrusion or misuse patterns. Some of the features may be redundant or contribute little (if anything) to the detection process. The purpose of this study is to identify important input features in building an IDS that is computationally efficient and effective. This article proposes an IDS model based on a general and enhanced flexible neural tree (FNT). Based on the predefined instruction/operator sets, a flexible neural tree model can be created and evolved. This framework allows input variables selection, overlayer connections, and different activation functions for the various nodes involved. The FNT structure is developed using an evolutionary algorithm, and the parameters are optimized by a particle swarm optimization algorithm. Empirical results indicate that the proposed method is efficient.

B. ***Hierarchical Kohonenen net for anomaly detection in network security***

A novel multilevel hierarchical Kohonen Net (K-Map) for an intrusion detection system is presented. Each level of the hierarchical map is modeled as a simple winner-take-all K-Map. One significant advantage of this multilevel hierarchical K-Map is its computational efficiency. Unlike other statistical anomaly detection methods such as nearest neighbor approach, K-means clustering or probabilistic analysis that employ distance computation in the feature space to identify the outliers, our approach does not involve costly point-to-point computation in organizing the data into clusters. Another advantage is the reduced network size. We use the classification capability of the K-Map on selected dimensions of data set in detecting anomalies. Randomly selected subsets that contain both attacks and normal records from the KDD Cup 1999 benchmark data are used to train the hierarchical net. We use a confidence measure to label the clusters. Then we use the test set from the same KDD Cup 1999 benchmark to test the hierarchical net. We show that a hierarchical K-Map in which each layer operates on a small subset of the feature space is superior to a single-layer K-Map operating on the whole feature space in detecting a variety of attacks in terms of detection rate as well as false positive rate.

C. ***Anomaly-based network intrusion detection***

With the advent of anomaly-based intrusion detection systems, many approaches and techniques have been developed to track novel attacks on the systems. High detection rate of 98% at a low alarm rate of 1% can be achieved by using these techniques. Though anomaly-based approaches are efficient, signature-based detection is preferred for mainstream implementation of intrusion detection systems. As a variety of anomaly detection techniques were suggested, it is difficult to compare the strengths, weaknesses of these methods. The reason why industries don't favor the anomaly-based intrusion detection methods can be well understood by validating the efficiencies of the all the methods. To investigate this issue, the current state of the experiment practice in the field of anomaly- based intrusion detection is reviewed and survey recent studies in this. This paper contains summarization study and identification of the drawbacks of formerly surveyed works.

III. PROPOSED METHOD

To avoid the problem in existing system, here we introduce the new system of a feature representation for effective and efficient intrusion detection. The proposed approach is based on two distances as the new features between a specific data and its cluster center and nearest neighbor respectively. This contains the processes of extracting cluster centers and nearest neighbors and new data formation. The purpose of this algorithm is to assign an unlabeled data to the class of its k nearest neighbors. Thus process provides the effective results.

A. Feature selection: KDD cup 99 dataset has been used to examine this technique. After loading the dataset, the dataset moves to the feature classification. In feature classification step, the features are classified by five classes. After completing the feature classification process, the classified features are separated into 41 feature set. Feature set are detect as normal and anomaly. Feature selection algorithm is used to eliminate the unimportant features.

***B. Distance of cluster center*:** To extract cluster centers, some clustering technique can be applied in this stage. In this paper, the k-means clustering algorithm is used. The chosen dataset consisting of 12 data samples (N1 to N12) is a five-class classification problem. Then, the number of clusters is defined as five (i.e. k = 5) for the k-means clustering algorithm. As a result, there are five clusters, in which each cluster contains a cluster center (i.e. C1, C2, C3, C4, and C5).

C. Distance of nearest neighbor: To identify the nearest neighbor of a data point, *Di* for example, the weighted *k*-NN approach is used where the distance between *Di* and each of the other data points in the same cluster can be obtained. That is, the nearest neighbor of *Di* is based on the shortest distance identified by weighted *k*-NN.

D. Load new dataset:

After the cluster center and nearest neighbor for every data of the chosen dataset are extracted and identified, two types of distances are calculated and then summed. For the first distance type, they are based on each data point to the cluster centers. That is, if there are three cluster centers, then there are three distances between a data point to the three cluster centers respectively. The second distance type is based on each data point to its nearest neighbor. The distance between two data points is based on the Euclidean distance. Finally this distance provides the new dataset.

E. Weighted Knn classifier:

The new dataset is divided into the training and testing datasets to train and test a specific classifier respectively. Here, we consider the weighted k-NN classifier since it is easy to be implemented and widely used as a baseline classifier in many applications. The KDD-Cup 99 dataset containing 494,020 samples, which is the most popular and widely used in related work. Specifically, each data sample represents a network connection represented by a 41-dimensional feature vector, in which 9 features are of the intrinsic types, 13 features are of the content type, and the remaining 19 features are of the traffic type. Each pattern of the dataset is labeled as belonging to one out of five classes, which are normal traffic and four different classes of attacks, i.e. probing, denial of service (DoS), remote to local (R2L), and user to root (U2R).

The k−N-N classifier tries to estimate the value of $\eta(\mathbf{X})$ by taking the majority vote of the labels of the nearest neighbors of **X**. The implicit assumption here is that the value of $\eta(\mathbf{X}(i)n)$ for $i = 1, \ldots, k$ is close to the value of $\eta(\mathbf{X})$.However,the farther the neighbor, the "less likely" is that the value of η is close to the value at **X** .An approach to solving this problem is to assign different weights to different neighbors, namely, to weigh less the vote of farther neighbors than those of close neighbors. From the viewpoint of notation, call $w1 \geq w2 \ldots \geq wk$ the weights of $\mathbf{X}(1)n, \mathbf{X}(2)n, \ldots, \mathbf{X}(k)n$.

Weighted k−N-N Rule

Decide 0 if $\sum i Y(i)\wedge n\ wi < \sum i(1 - Y(i)\wedge n)wi$ *for* $1 \leq i \leq K$

Decide 1 if $\sum i Y(i)\wedge n\ wi > \sum i(1 - Y(i)\wedge n)wi$ *for* $1 \leq i \leq K$

For fixed k odd, let wi = 1/k; for k even, let w1 = 1/k + € all the other wi = 1/k − €/(k − 1), call R(k) the corresponding asymptotic risk. Call R(w1, . . ., wk) the asymptotic risk of the weighted rule with weights w1, . . ., wk. Then R(k) ≤ R(w1, . . ., wk) where equality holds if Pr {η(X) = 1/2} = 1, or if every numerical minority of the wi carries less than 1/2 of the total weight.

For example,MacLeod, Luk, and Titterington proposed the following class of weights,which seem to work well in many practical cases

$$w_j = \frac{(d_s - d_j) + \alpha(d_s - d_1)}{(1+\alpha)(d_s - d_1)}$$

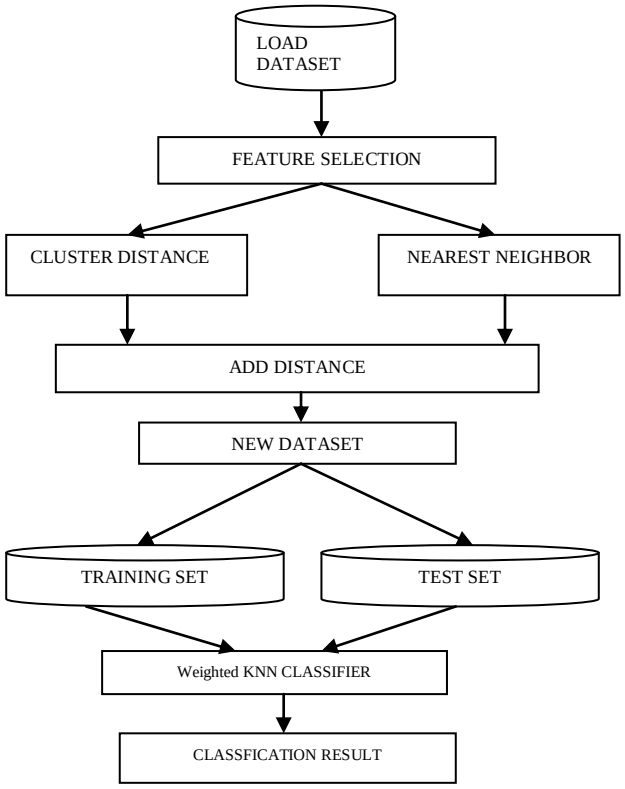
for $s > 1, s \geq k$. Here,

$$d_j = \left\| \mathbf{X} - \mathbf{X}_n^{(j)} \right\|, \ \alpha \geq 0.$$

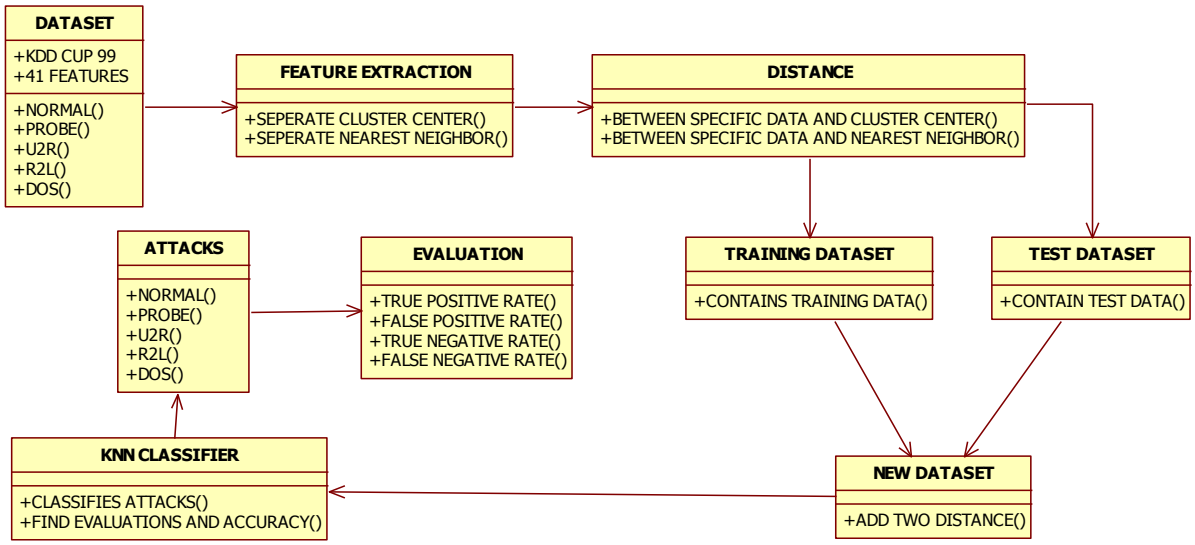
F.Accuracy:

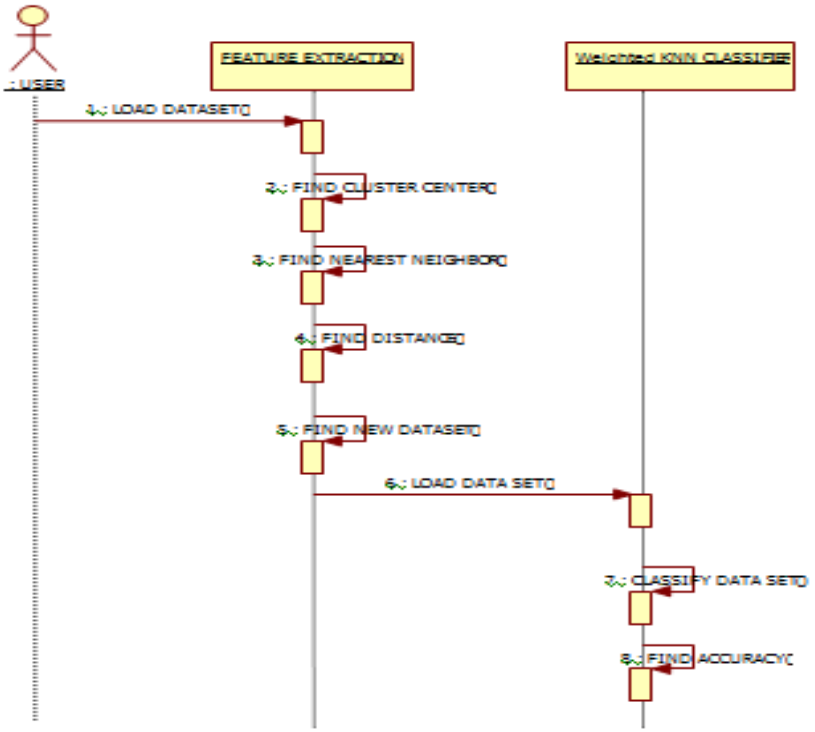
Accuracy process provides the accuracy result of the attacks in the dataset. This accuracy process contains the information of true positive rate and false positive rate. Accuracy and positive rate are getting from the attack detection results. So after detecting the attack, then only the accuracy and other positive rate are calculated. After calculating the accuracy, the graph is provided from the modified mutual information feature selection.

IV. SYSTEM ARCHITECTURE



V. CLASS DIAGRAMS AND ACTIVITY DIAGRAMS





VI .CONCLUSION

A new feature representation approach is proposed in this paper for effective and efficient intrusion detection. This approach first transforms the original feature representation of a given dataset into one dimensional distance based feature. Then, this new dataset is used to train and test a weighted k-NN classifier for classification. Our experimental results show that this performs similar to the k-NN and SVM classifiers using the original dataset in terms of accuracy, detection rates, and false alarm rates. However, the important strength of this method is that it needs less computational effort than the k-NN and SVM classifiers trained and tested by the original datasets. That is, although this requires additional computations for extracting the proposed distance based feature, it largely reduce the training and testing (i.e. detection) time since the new dataset only contains one dimension.

REFERENCES

[1] Cannady J. (1998). Artificial neural networks for misuse detection. Proceedings of the 1998 National Information Systems Security Conference (NISSC'98), 443-456, Arlington, VA.

[2] Lippmann, R.; Haines, J.; and Zissman, M. (2003). An overview of issues in testing intrusion detection systems. National institute of standards and technology (NTIS).

[3] Chen, W.H.; Hsu, S.H.; and Shen, H.P. (2005). Application of SVM and ANN for intrusion detection. Computers & Operations Research, 32(10), 2617–2634.

[4] Lorenzo-Fonseca, I.; Maciá-Pérez, F.; Mora-Gimeno, F.; Lau-Fernández1, R.; Gil-Martínez-Abarca, J.; and Marcos-Jorquera, D. (2009). Intrusion detection method using neural networks based on the reduction of characteristics. LNCS, 5517, 1296–1303.

[5] Mukkamala, S. (2002). Intrusion detection using neural networks and support vector machine. Proceedings of the 2002 IEEE International Honolulu, HI.

[6] Sammany, M.; Sharawi, M.; El-Beltagy, M.; and Saroit, I. (2007). Artificial neural networks architecture for intrusion detection systems and classification of attacks. Accepted for publication in the 5th international conference INFO2007, Cairo University.

[7] Selvakani, S.; and Rajesh, R.S. (2009). Escalate intrusion detection using GA–NN. International Journal of Open Problems in Computer Science and Mathematics, 2(2), 272-284.

[8] Morteza, A.; Jalili, R.; and Hamid R.S. (2006). RT-UNNID: A practical solution to real-time network-based intrusion detection using unsupervised neural networks. Computers & Security, 25(6), 459 – 468.

[9] Tran. T.P.; Cao, L.; Tran, D.; Nguyen, C.D. (2009). Novel intrusion detection using probabilistic neural network and adaptive boosting. International Journal of Computer Science and Information Security (IJCSIS), 6(1), 83-91.

[10] Chen, R.C.; Cheng, K.F.; and Hsieh, C.F. (2009). Using rough set and support vector machine for network intrusion detection. International Journal of Network Security & Its Applications (IJNSA), 1(1), 1-13.

