International Journal of u- and e- Service, Science and Technology
Vol.8, No.3 (2015), pp.295-300
http://dx.doi.org/10.14257/ijunesst.2015.8.3.28

Implementation of Support Vector Machines and Clustering of Intrusion Detection System for Computer Networks

Narges Salehpour

Department of Computer, Science, College of Lorestan, Islamic Azad University, Khorramabad, Iran
Department of Computer, Khorramabad Branch, Islamic Azad University, Khorramabad, Iran
Salehpour_narges@yahoo.com

Abstract

Considering that intrusion detection systems and anomaly clearly recognize malicious activity. Nowadays, data mining based intrusion detection systems, security and more rapidly detect attacks.Therefore,in this article we use a combination of k-means clustering algorithm and is used supervised support vector machine algorithm to find the best line separator. This is leading to the separation of normal and attack traffic.

Keywords*: intrusion detection systems, support vector machines, data mining techniques, clustering.*

1. Introduction

Due to the growth of computer networks, network security has been proposed as a major challenge. Intrusion detection systems to ensure the safe processing and storage of data on the network have been developed. As well considered as an essential component of network security. Using data mining techniques in order to increase the accuracy of intrusion detection systems which leads to increase the detection of intrusion detection systems.

2. Intrusion Detection System

Intrusion detection is the process of identifying and responding to destructive activities targeted, network resources are under attack. Although there are many levels of access to computer networks are protected. . However, hackers find ways to enter the network to major damages to create in the network. The goal of intrusion detection systems, but rather prevent the attack detection of attacks and detect Security Bugs in computer networks and finally, notify the system manager.
Tasks such as intrusion detection systems are [1]:
- Verification system errors,
- Check the integrity of the system and data files,
- Detection of anomalous behaviors and lack the normal system,
- Identify attacks and warnings.

Multiple alarm systems, intrusion detection systems, one of which is known as an expert system used for data analysis. Another intrusion detection system, NIDES that arises as a comprehensive intrusion detection system and anomaly detection and misuse implemented. NIDES anomalies Detected using a profile. In fact, the normal patterns of activity profiles offer system. Usually once daily profile in NIDES the new date. James Anderson's influence was divided into two categories: internal and external intrusion. Mr. Anderson on the set of records that would explain the unusual behavior of the system, such as the use of unauthorized, unregulated and often using unusual pattern of referrals

ISSN: 2005-4246 IJUNESST

to programs and the data focused. He also discussed the problems users are allowed access to critical information systems, and it was also warned that detects unauthorized use of security records, is a difficult task [2].

3. Clustering

Clustering includes the following methods:

3.1 Clustering with Only Link

This is one of the oldest and simplest methods of clustering and hierarchical clustering methods are component. This is called the clustering method nearest neighbor. First, the data is considered as a cluster and each iteration, one is the closest cluster. This process is repeated until the desired number of clusters to be lower.

$$d_{AB} = \min_{i \in A, j \in B} d_{ij} \quad (1)$$

The distance between two clusters i and j is given a sample i belongs to cluster A and j is a sample cluster B. The similarity between two clusters, the minimum distance between a member of one of the other members [3].

3.2 Clustering with Full Links

This method as Link the only component of the hierarchical clustering methods. This clustering technique called furthest neighbor. First, the data is considered as a cluster and each iteration, the closest clusters are merged And the new cluster to its distance from other clusters is equal to the maximum distance between the new cluster of other clusters obtained. This is repeated until the desired number of clusters to be lower. The method for calculating the similarity between two clusters A and B of the criteria used.

$$d_{AB} = \max_{i \in A, j \in B} d_{ij} \quad (2)$$

In the above equation, i, j, a data sample belongs to cluster A and cluster B is a sample of data. The similarities between the two methods mentioned maximum distance between a cluster member, a member of another example of a sample [3].

3.3 Mean Link Clustering Method

This link is only as part of a hierarchical clustering methods are exclusive. Since both single link clustering and full links are very sensitive to noise, but this method has the further calculations. The method used to calculate the similarity between two clusters A and B of the criteria used.

$$d_{AB} = \frac{\sum_{i \in A, j \in B} d_{ij}}{N_A N_B} \quad (3)$$

On the way mean the link, i samples belonging to the cluster A and cluster B j is a sample of data. N_A Number of clusters A and B is N_B number of cluster members. So in this way, the similarity between two clusters, mean distance between all members of a sample of all the other samples [3].

3.4 Clustering Using Group Mean the Link

This method as the link to only component of
hierarchical clustering methods and proprietary. This method is called the centroid distance. The method for calculating the similarity between two clusters A and B of the following criteria are used:

$$d_{AB} = d\left(\frac{\sum_{i\epsilon A} X_i}{N_A}, \frac{\sum_{j\epsilon B} X_J}{N_B}\right) \quad (4)$$

X_i Sample data that belong to cluster A, X_j sample data that belongs to cluster B. N_A number of members of the cluster A and cluster B is N_B members. The similarity between two clusters using the distance between all members of the sample mean vector with mean vector of all members is another sample.

3.5 Clustering by the distance between

This method is similar to the link, hierarchical clustering methods and proprietary components. The group mean the link method, if a small cluster combined with a large cluster point mean cluster result, the mean cluster will be larger point that is not desirable in some applications. In this distance between the clustering method, not the problem. In this way, the middle of a cluster as the center of the cluster is used [3].

3.6. Algorithm K -Means

Method K-Means, an unsupervised clustering method is repeatable in which a given set of data In a d- The event { c_1, c_2,c_k ...}With centers { s_1, s_2, s_K,... } dimensional space into k cluster {p_1, p_2, p_N,...}that the objective function to be minimized.

‖ ‖Criterion the distance between the points and the j-th cluster center c_i Algorithm K-Means, the first data point k as the initial centers. Criterion the distance between the points and the j-th cluster center c_i Algorithm K-Means, the first data point k as the initial centers.

The first part of a series or a series of random k points can be chosen from the data.

The algorithm is repeated as follows:

1. Each sample data to cluster the minimum distance to the center of the cluster is assigned.
2. All the data is then assigned to clusters, each cluster through the averaged points belonging to the cluster, cluster centers are updated.
3. Steps 1 and 2 are repeated until no change in cluster centers is not achieved.

Generally, the first stage of complexity o (ndk) and the second round complexity o(nd). This algorithm converges very fast, but does not guarantee overall optimality [5, 4].

4. Support Vector Machine

Suppose the number of training patterns as feature vectors or {x_1, x_2… ,x_i,…, x_N}Each having a d-dimensional feature vectors are present. With labels yi, such that is $y_i \in \{-1, +1\}$

Basic equations of two-class support vector machine is defined as the target, the optimal solution to a problem is categorized into two classes.

Data sets are separated by a line. Support Vector Machine classification as a risk that defines a set of numeric quantities. The minimum value of the account between them.. The MLP neural network when the resolution is correct, as well as minimizing the amount of error separation and considers it as a solution. Since the training data set and are labeled by separators, parts belonging to two classes are separated. These separators are known to the decision boundary. Figure 1 refers to the number of decision boundaries.

These separators are known to the decision boundary. Figure 1 refers to the number of decision boundaries[6].

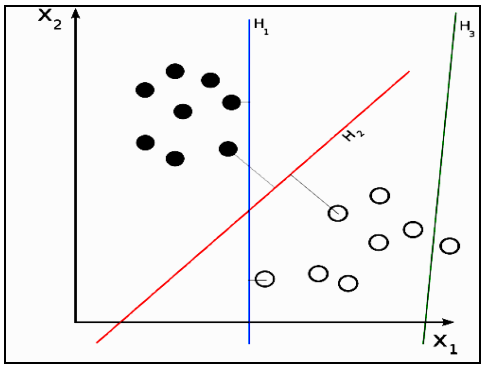

Figure 1. The Decision Boundary [6]

In Figure 1 separators H_3, the class is not properly classified. So separators H_3, separators are not desirable. But two separators H_1 and H_2 operational risk, which also has H_2 operational risk is less than H_1. In other words, H_2 in the figure above, the equilibrium point of the system. For each member of the class, called support vector. Nearest member of any class of its own to try to ward off the line separators.

5. Clustering Algorithms in Intrusion Detection Systems

Clustering means putting the same data in a category. Among the methods that are used for anomaly detection, clustering using Euclidean distance. Euclidean distance similarity is measured using the formula QOS. Algorithm FPMAFIA based clustering is used for large data sets. Its main advantage of clusters with arbitrary shapes in which the Dytast KDD1998 used for classification. This clustering makes all the data in a cluster are to be treated in the same way. This clustering in statistics, machine learning and data mining applications. In an information system, the values of the normalized data are much more subversive data. Must therefore be larger than the normal cluster is a cluster of intrusion data. The normal behavior of intrusion detection systems and describe clusters grouped as normal as normal cluster signature is used for diagnosis [7].

In clustering, analysis, and grouping objects into clusters is performed, so that objects within a cluster are similar to each other. Clustering algorithms and machine vision applications such as image segmentation can be noted. K-Means clustering algorithms and spider algorithms are employed to improve the accuracy of intrusion detection system. The original idea of the K-Means algorithm to find k cluster centers and the best choice for the cluster, an object that is at least similar to other objects[8].

6. The Method Proposed

In the proposed method, first the attacks are considered as input data, normalized and repetitive attacks will be deleted. The operation is carried out attacks on clustering and k-means clustering based on cluster centers are specified. Finally, by using support vector normal traffic, and attacks are separated from each other. In this way the zero normal traffic and traffic with a given attack.

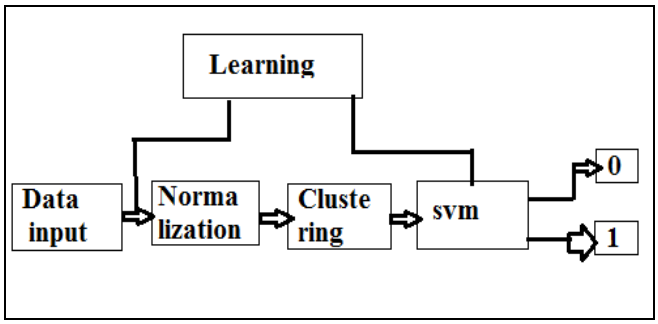

Figure 3. Proposed Method

7. Conclusion

Given that today's modern society is increasingly dependent on complex systems. Intrusion detection systems, destructive or unauthorized activities that are taking place diagnose. When the network intrusion attacks, often with separate attacks in network intrusion detection are done quickly.

As part of computer science and computer security problems in the field of computer security is to be included. In this paper, we use data mining technique support vector machine and k-means clustering algorithm is used to detect attacks (Normal, DOS,U2R, R2L). The results show that using k-means clustering algorithm and support vector machine more accurately detect attacks.

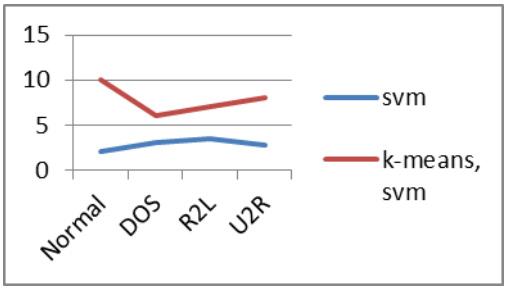

Figure 4. K-means and SVM Evaluation

References

[1] N. B. Idris and B. Shanmugam, "Artificial Intelligence Techniques Applied to Intrusion Detection", **(2005)**, pp. 52-55.
[2] J.P.Anderson, "Computer Security threat Monitoring And Surveillance", **(1980),** pp. 1-56.
[3] A. R. Webb, Statistical Pattern Recognition, **(2002)**.
[4] J. Macqueen, "Some Methods FOR Classification AND Analysis OF Multivariate Observations", pp. 281-297, **(1967)**.
[5] J. A. Hartigan and M. A. Wong, "Algorithm AS 136: A K-Means Clustering Algorithm", vol. 28, pp. 100-108, 1979.
[6] Prateek and D. S. K. Jena, "Intrusion Detection Using Self-Training Support Vector Machines", **(2013)**, pp. 1-35.
[7] G. Kumar, K. Kumar, and M. Sachdeva, "The use of artificial intelligence based techniques for intrusion detection: a review", **(2001)**, pp. 369–387.
[8] P. Scherer, M. Vicher, V. Snase, J. Martinovic, J. Dvorsky and V. Snasel, "Using SVM and Clustering Algorithms in IDS Systems", **(2011),** pp. 108-119.

