ISSN 2047-3338

Implementation of Anomaly Detection Technique Using Machine Learning Algorithms

K. Hanumantha Rao[1], G. Srinivas[2], Ankam Damodhar[3] and M. Vikas Krishna[4]
[1,2,3,4]Sri Indu College of Engineering and Technology, Hyderabad, India

***Abstract*— Data mining techniques make it possible to search large amounts of data for characteristic rules and patterns. If applied to network monitoring data recorded on a host or in a network, they can be used to detect intrusions, attacks and/or anomalies. In this paper, we present "machine learning" a method to cascade K-means clustering and the Id3 decision tree learning methods to classifying anomalous and normal activities in a computer network. The K-means clustering method first partitions the training instances into two clusters using Euclidean distance similarity. On each cluster, representing a density region of normal or anomaly instances, we build an ID3 decision tree. The decision tree on each cluster refines the decision boundaries by learning the subgroups within the cluster. Our work studies the best algorithm by using classifying anomalous and normal activities in a computer networks with supervised & unsupervised algorithms that have not been used before. We analyses the algorithm that have the best efficiency or the best learning and describes the proposed system of K-means&ID3 Decision Tree.**

***Index Terms*— Data Mining, Intrusion, Anomaly Detection, Anomalies, k-means and Decision Tree**

I. INTRODUCTION

ANOMALY Detection System (ADS) monitors the behavior of a system and flag significant deviations from the normal activity as an anomaly. Anomaly detection is used for identifying attacks in a computer networks, malicious activities in a computer systems, misuses in a Web-based systems. A network anomaly by malicious or unauthorized users can cause severe disruption to networks. Therefore the development of a robust and reliable network anomaly detection system (ADS) is increasingly important.

Traditionally, signature based automatic detection methods are widely used in intrusion detection systems. When an attack is discovered, the associated traffic pattern is recorded and coded as a signature by human experts, and then used to detect malicious traffic. However, signature based methods suffer from their inability to detect new types of attack.

Furthermore, the database of the signatures is growing, as new types of attack are being detected, which may affect the efficiency of the detection. We explored a number of techniques like Association Rule Mining and Frequent Episode rules [2]. Association Rule mining usually is very slow and though once a popular technique, it's being replaced by other powerful techniques like clustering and classification.

Then we came across a recent paper [1], which advocated the use of outlier detection technique for detecting the anomalous data points in datasets. Clustering was the first choice because the dataset was huge and multidimensional. We used the K-means algorithm for this. The idea was to train a K-Means cluster using Normal datasets and cluster the normal behavior points. For the test data set, the probability of its belonging to the most probable cluster was computed. If this was below a threshold, the instance was flagged as anomalous.

This approach did not give us very good results. As a consequence, even the data points corresponding to attack data were being assigned to clusters with a very high probability. The technique we adopted for anomaly detection was prediction of the ith system call for a record containing a sequence of n system calls. The predicted value was compared with the actual value. If the value was found to be different, then the confidence of prediction of the value is taken into consideration. All these confidence scores are added up to compute the total misclassification score.

If this misclassification score crosses a threshold, then the region is classified as an anomalous region. We used classification technique for prediction since the data had few dimensions, equal to the size of the sliding window. The different options we considered for classification were decision trees, SVM, naive bayes and meta-learners formed by the combination of these techniques. Out of these, decision trees gave us the best results. However, this may be due to the lack of tuning of the other classification models such as SVM.

A. Plan of the Paper

This paper focuses on a detailed introduction about several anomaly detection schemes for identifying normal and anomalies in a network anomaly data.

Section 2 describes intrusion detection and types of intrusion detection, categories of intrusion detection system.

Section 3 describes anomalies and several supervised and unsupervised anomaly detection techniques. Section 4 describes individual usage on the K-means & Id3 decision Tree.

Section 5 discusses about the comparative study. Section 6 describes combined approach of the proposed system. Section 7 is finally describes the conclusion and future work.

II. INTRUSION DETECTION SYSTEM

Intrusion detection systems (IDS) process large amounts of monitoring data. As an example, a host-based IDS examines log files on a computer (or host) in order to detect suspicious activities. Network-based IDS, on the other hand, searches network monitoring data for harmful packets or packet flows.

A. Types of Intrusion Detection System

i) Network Intrusion Detection System: Network–based intrusion detection system (NIDS) [16] that tries to detect malicious activity such as denial of service attacks, port scan or even attempts to crack into computer by monitoring network traffic. NIDS does this by reading all incoming packets and trying to find number of TCP connection requests to a very large number of different ports is observed, one could assume that there is someone conducting a port scan of some or all of the computers in the network. It mostly tries to detect incoming shell codes in the same manner that an ordinary intrusion detection system does. Often inspecting valuable information about an ongoing intrusion can be learned from outgoing or local traffic and also work with other systems as well, for example update some firewalls blacklist with the IP address of computers used by suspected crackers.

ii) Host-based Intrusion Detection System: Host-based intrusion detection system (HIDS) [16] monitors parts of the dynamic behavior and the state of computer system, dynamically inspects the network packets. A HIDS could also check that appropriate regions of memory have not been modified, for example- the system-call table comes to mind for Linux and various v table structures in Microsoft Windows. For each object in question usually remember its attributes (permissions, size, modifications dates) and create a checksum of some kind (an MD5, SHA1 hash or similar) for the contents, if any, this information gets stored in a secure database for later comparison (checksum-database). At installation time- whenever any of the monitored objects change legitimately- a HIDS must initialize its checksum-database by scanning the relevant objects. Persons in charge of computer security need to control this process tightly in order to prevent intruders making un-authorized changes to the database.

iii) Protocol-based Intrusion Detection system: Protocol-based intrusion detection system (PIDS) [16] typically installed on a web server, monitors the dynamic behavior and state of the protocol, and typically consists of system or agent that would sit at the front end of a server, monitoring the HTTP protocol stream. Because it understands the HTTP protocol relative to the web server/system it is trying to protect it can offer grater protection than less in-depth techniques such as filtering by IP address or port number alone, however this greater protection comes at the cost of increased computing on the web server and analyzing the communication between a connected device and the system it is protecting.

iv) *Application Protocol-based Intrusion Detection System:* Application protocol-based intrusion detection system (APIDS) [16] will monitor the dynamic behavior and state of the protocol and typically consists of a system or agent that would sit between a process, or group of servers, monitoring and analyzing the application protocol between two connected devices.

B. Categories of Intrusion Detection System

Intrusion detection is classified into two types, i) Misuse detection, and ii) Anomaly detection. Misuse detection uses well-defined patterns of the attack that exploit weakness in system and application software to identify the intrusions (Kumar and Spafford 1995).

These patterns are encoded in advance and used to match against user behavior to detect intrusions. Anomaly detection identifies deviations from the normal usage behavior patterns to identify the intrusion. The normal usage patterns are constructed from the statically measures of the system features, for example the CPU and I/O activities by a particular user or program. The behavior of the user is observed and any deviation from the constructed normal behavior is detected as intrusion.

III. INTRODUCTION TO ANOMALY

A. What is Anomaly?

Anomaly detection refers to detecting patterns in a given data set that do not conform to an established normal behavior. The patterns thus detected are called anomalies and translate to critical and actionable information in several application domains. Anomalies are also referred to as outlier, surprise deviation etc.

Most anomaly detection algorithms require a set of purely normal data to train the model and they implicitly assume that anomalies can be treated as patterns not observed before. Since an outlier may be defined as a data point which is very different from the rest of the data, based on some measure, we employ several detection schemes in order to see how efficiently these schemes may deal with the problem of anomaly detection. The statistics community has studied the concept of outliers quite extensively. In these techniques, the data points are modeled using a stochastic distribution and points are determined to be outliers depending upon their relationship with this model. However with increasing dimensionality, it becomes increasingly difficult and inaccurate to estimate the multidimensional distributions of the data points.

However recent outlier detection algorithms that we utilize in this study are based on computing the full dimensional distances of the points from one another as well as on computing the densities of local neighborhoods. The deviation measure is our extension of the traditional method of discrepancy detection. As in discrepancy detection, comparisons are made between predicted and actual sensor values, and differences are interpreted to be indications of anomalies. This raw discrepancy is entered into a normalization process identical to that used for the value change score, and it is this representation of relative discrepancy which is reported.

The deviation score for a sensor is minimum if there is no discrepancy and maximum if the discrepancy between predicted and actual is the greatest seen to date on that sensor. Deviation requires that a simulation be available in any form

for generating sensor value predictions. However the remaining sensitivity and cascading alarms measures require the ability to simulate and reason with a causal model of the system being monitored. Sensitivity and cascading alarms are an appealing way to assess whether current behavior is anomalous or not is via comparison to past behavior. This is the essence of the surprise measure. It is designed to highlight a sensor which behaves other than it has historically. Specifically, surprise uses the historical frequency distribution for the sensor in two ways. It is those sensors and to examine the relative likelihoods of different values of the sensor. It is those sensors which display unlikely values when other values of the sensor are more likely which get a high surprise scores. Surprise is not high if the only reason a sensor's value is unlikely is that there are many possible values for the sensor, all equally unlikely.

B. Data Mining Classifications Methods for Anomaly Detection Systems

Anomaly detection builds models of normal data and then attempt to detect normal model in observed data. The broad categories of anomaly detection techniques exist

Supervised anomaly detection techniques learn a classifier using labeled instances belonging to normal and abnormal class and then assign a normal or anomalous label to a test instance.

Data Mining interfaces support the following supervised functions:

A classification task begins with build data (also known as training data) for which the target values (or class assignments) are known. Different classification algorithms use different techniques for finding relations between the predictor attributes' values and the target attribute's values in the build data.

Decision tree rules provide model transparency so that a business user, marketing analyst, or business analyst can understand the basis of the model's predictions, and therefore, be comfortable acting on them and explaining them to others Decision Tree does not support nested tables. Decision Tree Models can be converted to XML.

NB makes predictions using Bayes' Theorem, which derives the probability of a prediction from the underlying evidence. Bayes' Theorem states that the probability of event A occurring given that event B has occurred (P(A|B)) is proportional to the probability of event B occurring given that event A has occurred multiplied by the probability of event A occurring ((P(B|A)P(A)).

Adaptive Bayes Network (ABN) is an Oracle proprietary algorithm that provides a fast, scalable, non-parametric means of extracting predictive information from data with respect to a target attribute. (Non-parametric statistical techniques avoid assuming that the population is characterized by a family of simple distributional models, such as standard linear regression, where different members of the family are differentiated by a small set of parameters.)

Support Vector Machine (SVM) is a state-of-the-art classification and regression algorithm. SVM is an algorithm with strong regularization properties, that is, the optimization procedure maximizes predictive accuracy while automatically avoiding over-fitting of the training data. Neural networks and radial basis functions, both popular data mining techniques, have the same functional form as SVM models; however, neither of these algorithms has the well-founded theoretical approach to regularization that forms the basis of SVM.

Semi-supervised anomaly detection techniques construct a model representing normal behavior from a given normal training dataset, and then test the likelihood of test instances to be generated by the learnt model.

Semi-supervised learning is a class of machine learning techniques that make use of both labeled and unlabeled data for training - typically a small amount of labeled data with a large amount of unlabeled data. Semi-supervised learning falls between unsupervised learning without any labeled training data and supervised learning with completely labeled training data.

Semi-supervised is a combination of supervised and unsupervised

Machine learning is a scientific discipline that is concerned with the design and development of algorithms that allow computers to learn based on data, such as from sensor data or databases. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data. Hence, machine learning is closely related to fields such as statistics, probability theory, data mining, pattern recognition, artificial intelligence, adaptive control, and theoretical computer science.

Unsupervised anomaly detection techniques detect anomalies in an unlabeled test data set under the assumption that majority of the instances in the data set are normal.

Unsupervised functions in data mining are association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Piatetsky-Shapiro [8] describes analyzing and presenting strong rules discovered in databases using different measures of interestingness. Based on the concept of strong rules, Agrawa [9]l et al. introduced association rules for discovering regularities between products in large scale transaction data recorded by point-of-scale (POS) systems in supermarkets. For example, the rule {onions, potatoes} => {beef} found in the sales data of a supermarket would indicate that if a customer buys onions and potatoes together, he or she is likely to also buy beef.

Clustering is a data mining (machine learning) technique used to place data elements into related groups without advance knowledge of the group definitions.

Association model is often used for market basket analysis, which attempts to discover relationships or correlations in a set of items. Market basket analysis is widely used in data analysis for direct marketing, catalog design, and other business decision-making processes. Traditionally, association models are used to discover business trends by analyzing customer transactions. However, they can also be used effectively to predict Web page accesses for personalization. For example, assume that after mining the Web access log, Company X discovered an association rule "A and B implies C," with 80% confidence, where A, B, and C are Web page accesses. If a user has visited pages A and B, there is an 80% chance that he/she will visit page C in the same session. Page C may or may not have a direct link from A or B. This

information can be used to create a dynamic link to page C from pages A or B so that the user can "click-through" to page C directly. This kind of information is particularly valuable for a Web server supporting an e-commerce site to link the different product pages dynamically, based on the customer interaction.

VI. USAGE OF K-MEANS +ID3 DECISION TREE FOR SOLVING ANOMALY

Data mining is extracting knowledge hidden information in large volumes of raw data, typical tasks of data mining are Detect fraud and abuse in insurance and finance, Estimate probability of an illness re-occurrence or hospital re-admission, Predict peak load of a network.

Data Mining-based anomaly Detection is become prevalent in essence. Network security is just network information security. In general, all technologies and theories about secrecy, integrality, usability, reality and controllable of network information are the research domain of network security. Intrusion is an action that tries to destroy that secrecy, integrality and usability of network information, which is unlicensed and exceed authority. Intrusion Detection is a positively technology of security defend, which gets and analyses audit data of computer system and network from some network point, and to discover whether there is the action of disobeying security strategy and whether be assaulted. Intrusion Detection System is the combination of software and hardware of Intrusion Detection Data mining can be supervised & unsupervised supervised learning is to use the available data to build one particular variable of interest in terms of rest of data.

A number of classification algorithms can be used for anomaly detection proposes the use of ID3 Decision tree classifiers to learn a model that distinguishes the behavior of intruder from the normal user's behavior.

Unsupervised learning is where no variable is declared as target the goal is to establish some relationship among all variables. Unsupervised learning [17] studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. The unsupervised learner brings to bear prior biases as to what aspects of the structure of the input should be captured in the output.

In this paper combination of Applications Supervised& Unsupervised has been combined together used to solve the problem of Network Anomaly Data.

A Very rare case both the Techniques have been combined Supervised Classification, Decision Tree, Bayesian Classification, Bayesian belief networks; neural networks etc are used in data mining based applications.

A. Classification Techniques

In Classification, training examples are used to learn a model that can classify the data samples into known classes. The Classification process involves following steps:

a. Create training data set
b. Identify class attribute and classes
c. Identify useful attributes for classification (relevance analysis)
d. Learn a model using training examples in training set
e. Use the model to classify the unknown data samples

Unsupervised (Clustering): Association Rules, Pattern Recognition, Clustering Technique. The paper clustering Technique is one of the media to Network Anomaly data.

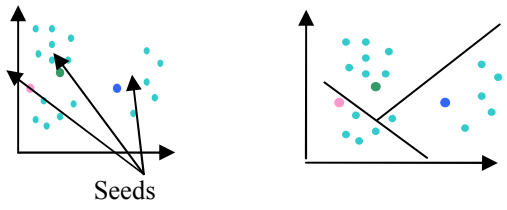
B. Clustering Technique

Cluster is a number of similar objects grouped together. It can also be defined as the organization of dataset into homogeneous and/or well separated groups with respect to distance or equivalently similarity measure. Cluster is an aggregation of points in test space such that the distance between any two points in cluster is less than the distance between any two points in the cluster and any point not in it. There are two types of attributes associated with clustering, numerical and categorical attributes. Numerical attributes are associated with ordered values such as height of a person and speed of a train. Categorical attributes are those with unordered values such as kind of a drink and brand of car.

Clustering is available in flavors of i) Hierarchical, and ii) Partition (non Hierarchical).

In hierarchical clustering the data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to n clusters each containing a single object [12]. Hierarchical Clustering is subdivided into *agglomerative* methods, which proceed by series of fusions of the n objects into groups, and *divisive* methods, which separate n objects successively into finer groupings.

For the partitional can be of K-means [15] and K-mediod. The purpose solution is based on K-means (Unsupervised) clustering combine with Id3 Decision Tree type of Classification (Supervised) under mentioned section describes in details of K-means & Decision Tree. K-means [3] [14] is a centroid based technique. Each cluster is represented by the center of gravity of the cluster so that the intra cluster similarity is high and inter cluster similarity is low. This technique is scalable and efficient in processing large data sets because the computational complexity is O(nkt) where n-total number of objects, k is number of clusters, t is number of iterations and k<<n and t<<n.



(a) Un-clustered Data Instances (b) Resultant Clusters

Fig. 1. Formation of clusters using seed points

C. K-mean Algorithm

1). Select k centroids arbitrarily (called as seed as shown in the Fig. 1) for each cluster C_i, i ε [1, k]

2). Assign each data point to the cluster whose centroid is closest to the data point.

3). Calculate the centroid C_i of cluster C_i, i ε [1, k].

4). Repeat steps 2 and 3 until no points change between clusters.

A major disadvantage of K means is that one must specify the clusters in advance and further the algorithm is very sensitive of noise, mixed pixels and outliers. The definition of means limit the application to only numerical variables. We choose k-means because it is data driven with relatively few assumptions on the distributions of underlying dat.

D. Decision Tree

Decision tree support tool that uses tree-like graph or models of decisions and their consequences [5][6], including event outcomes, resource costs, and utility, commonly used in operations research, in decision analysis help to identify a strategy most likely to reach a goal. In data mining and machine learning, decision tree is a predictive model that is mapping from observations about an item to conclusions about its target value. The machine learning technique for inducing a decision tree from data is called decision tree learning.

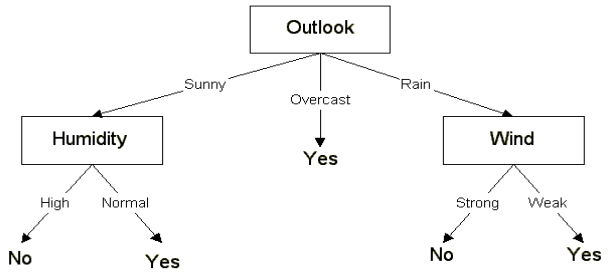


Fig. 2. The example of figure is taken from [11]

In above Fig. 2 tree is classified into five leaf nodes. In a decision tree, each leaf node represents a rule. The following rules are as follows in Fig. 2. Rule 1: If it is sunny and the humidity is high then do not play. Rule 2: If it is sunny and the humidity is normal then plays. Rule3: If it is overcast, then play. Rule 4: If it is rainy and wind is strong then do not play. Rule 5: If it is rainy and wind is weak then plays.

E. ID3 Decision Tree

Iterative Dichotomiser is an algorithm to generate a decision tree invented by Ross Quinlan, based on Occam's razor. It prefers smaller decision trees (simpler theories) over larger ones. However, it does not always produce smallest tree, and therefore heuristic. The decision tree is used by the concept of Information Entropy.

The ID3 Algorithm is:

1) Take all unused attributes and count their entropy concerning test samples

2) Choose attribute for which entropy is maximum

3) Make node containing that attribute

ID3 (Examples, Target _ Attribute, Attributes)

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
    - A = The Attribute that best classifies examples.
    - Decision Tree attribute for Root = A.
    - For each possible value, v_i, of A,
        - Add a new tree branch below Root, corresponding to the test A = v_i.
        - Let Examples(v_i), be the subset of examples that have the value v_i for A
            - If Examples(v_i) is empty common target value in the examples
        - Else below this new branch add the sub tree ID3 (Examples(v_i), Target_ Attribute, Attributes – {A}
- End
- Return Root

V. ARCHITECTURE OF PROPOSED SYSTEM

The futuristic approach of combining K-means+Id3 Decision Tree in a unique flavor is shown as: We choose K-means clustering because i) it is data driven method relatively few assumptions on the distributions of the underlying data and, ii) greedy search strategy of K-means guarantees at least a local minimum of the criterion function, thereby accelerating the convergence of clusters on large datasets. The first stage of K-means clustering is performed on training instances to obtain k disjoint clusters. Each k-means cluster represents a region of similar instances in terms of Euclidean distances between the instances and their cluster centroids The second stage of K-means+ID3 Decision Tree the K-means method is cascaded with the ID3 decision tree learning by building an ID3 decision tree using the instances in each K-means cluster. Computer is connected to Network; by connecting the network will get packets to search the packet whether it is normal or anomaly. Our proposed system using K-means and ID3 Decision Tree, input source is any one of dataset like KDD. We are giving input data to K-means Cluster to analyze, the k-means clusters grouping the data into clusters, it as different number of clusters K_1, K_2… Kn each cluster is having its own rules, for unknown data by default we consider N=2 (Number of clusters) depends on the Euclidean Distance the data is grouping into appropriate cluster. In k-means cluster anomaly type of data is not eliminated, if the overlapping is found, the id3 decision tree is used to classify each cluster. Finally the result of the k-means+ID3 is classifies the data into anomaly

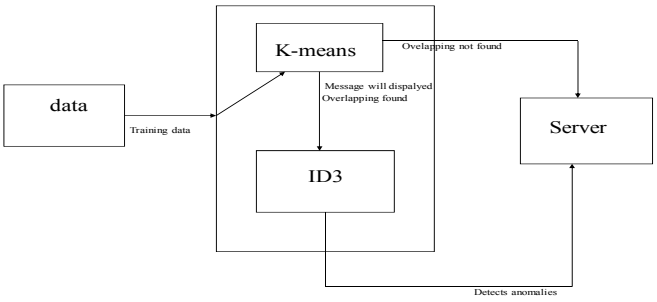


Fig. 3. The basic idea for the figure is taken from [13]

or normal, if the overlapping is not found in k-means cluster, the dataset is directly sends to the server. Overlapping may or may not occur for known data because based on the centroids in dataset only, we consider the N value.

Fig. 3 shows the Architecture of the K-means+ID3 Decision Tree using Anomaly detection. The K-means+ID3 method has two steps i) Training and, ii) Testing. During training step the K-means anomaly detection method are first applied to partition the training space into k disjoint clusters C1, C2 …. Ck then Id3 decision tree is trained with the instances in each k-means cluster. The K-means method ensures that each training instance is associated with only on cluster. However if there are any subgroups or overlaps within a cluster, the ID3 decision tree trained on that cluster refines the decision boundaries by partitioning the instances with a set of if-then rules over the feature space. The advantage of the proposed system is eliminates the anomalies from the k-means cluster using id3, disadvantage is Time taking process to integrate the K-means+ID3, because data format is total different. K-means is suitable for Numerical data & ID3 is Non-Numerical.

IV. EXPERIMANTAL RESULTS

This screen represents the individual k-means cluster with input file attribute relational format type of iris.arff. Represents the number of iterations in the clusters is 7 with squared errors 62.1436 and the centroids of cluster 0 is versicolor and cluster 1 is setosa and shows the percentage of each cluster is cluster 0 is 67% and cluster 1 is 33%.

Fig. 4 represents the individual id3 decision tree with input file attribute relational format type of weather .nominal. It shows the result of weather dataset which shows confused matrix and classified matrix.

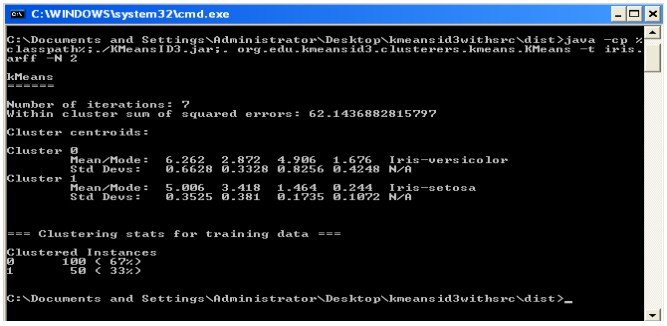


Fig. 4. Individual id3 decision tree with input file attribute

(a)

(b)

Fig. 5. K-Means & ID3 algorithms with iris.arff dataset

Fig.5 represents the K-Means & ID3 algorithms with iris.arff dataset, by default value of k is minimum number based on forced assignment and class dominance problems.

IIV. CONCLUSION AND FUTURE WORK

In this paper a general supervised and unsupervised for identifying anomaly instance in large and complex network datasets and an explanation mechanism to explain the normal or anomalies results was described. The specific approaches of the anomaly detection systems learning are characterized, we developed the K-Means&ID3 decision tree method is based on cascading two machine learning techniques i) the k-Means and, ii) the ID3 decision trees. The k-Means method is first applied to partitioning the training instance into k disjoint clusters. The ID3 decision tree built on each cluster learns the sub classifies within the cluster and partitions the decision space into classification regions; there by improving the system classification performance.

Our future direction is to utilize dependency measure and anomaly detection system for other purpose such as cascading the classifiers developed using different clustering methods like hierarchical clustering, adaptive resonance (ART) neural networks and kohonen's self-organizing maps and decision trees like C4.5 and classification and Regression trees (CART).

Motivated by issues of verification from false positives, our system not only identifies attacks but also be able to explain to an analyst normal or anomaly type of activities. To end this we designed and implemented a novel explanation mechanism for the problem of having high false alarms can also be resolved by one such approach that uses K-Means+ID3 is to detect the anomalies to achieve a high rate of accuracy in the case of unknown attacks in human- understandable can also improve the efficiency when analyzing the complex dataset.

REFERENCES

[1]. Lazarevic, A. Ozgur, L. Ertoz, J. Srivastava, and V. Kumar, "A Comparative Study of Anomaly Detection Schemes in Network Intrusion Detection," Proc. SIAM Int'l Conf. Data Mining, May 2003.

[2]. W. Lee, S. J. Stolfo *Data Mining Approaches for Intrusion Detection*.

[3]. Rui Xu,, Donald Wunsch II, Survey of Clustering Algorithms, IEEE in Neural Networks 16(3) (2005)

[4]. N. Ye, Y. Zhang, and C.M. Borror, "Robustness of the Markov- Chain Model for Cyber-Attack Detection," IEEE Trans. Reliability, vol. 53, no. 1, pp. 116-123, 2004.

[5]. D. Mutz, F. Valeur, G. Vigna, and C. Kruegel, "Anomalous System Call Detection," ACM Trans. Information and System Security, vol. 9, no. 1, pp. 61-93, Feb. 2006.

[6]. M. Thottan and C. Ji, "Anomaly Detection in IP Networks," IEEE Trans. Signal Processing, vol. 51, no. 8, pp. 2191-2204, 2003.

[7]. C. Kruegel and G. Vigna, "Anomaly Detection of Web-Based Attacks," Proc. ACM Conf. Computer and Comm. Security, Oct. 2003.

[8]. Modeling intrusion detection system using hybrid intelligent systems, Sandhya Peddabachigaria, Ajith Abrahamb, Crina Grosanc, Johnson Thomasa, A Computer Science Department, Oklahoma State University, OK 74106, USA School of Computer Science and Engineering.

[9]. Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, in G. Piatetsky-Shapiro & W. J. Frawley, eds, 'Knowledge Discovery in Databases', AAAI/MIT Press, Cambridge, MA.

[10]. R. Agrawal; T. Imielinski; A. Swami: *Mining Association Rules* Between Sets of Items in Large Databases", SIGMOD Conference 1993: 207-216.

[11]. Intrusion Detection System using Data mining Techniques Sanket Patle 07305034, Raviraj Vaishampayan 07305040 Kumar Avinav Dubey 07305044,Ganesh Wagle 07305805 under the guidance of Prof. Bernard L. Menezesection IIT Bombay Department of Computer Science and Engineering Indian Institute of Technology, Bombay Nov 2007.

[12]. Text Book of Data mining Techniques by Arun K Pujari Universities Press (India) Private Limited.

[13]. Intoduction to hierarchical clustering, A tutorial on clustering, A Tutorial on Clustering Algorithms **...** Hierarchical Clustering - Interactive demo.

[14]. K-Means+ID3: A Novel Method for Supervised Anomaly Detection by Cascading K-Means Clustering and ID3 Decision Tree Learning Methods, Shekhar R. Gaddam, Vir V. Phoha, and Kiran S. Balagani, IEEE Transactions on Knowledge and Data Engineering, VOL. 19, NO. 3, March 2007.

[15]. Karl-Heinrich Anders, A Hierarchical Graph-Clustering Approach to find Groups of Objects, IEEE Transactions.

[16]. A. Macgregor, M.Hall, P.Lorier and J.Bruskill, "Flow clustering using machine learning techniques", In PAM 2004, Antibes-Juan-Les-Pins, France, LNCS. pp. 205-214, 2004.

[17]. Intrusion Detection Systems An intrusion detection system is used to detect several types of Intrusion detection system evasion techniques bypass detection by creating different states on the IDS and ... The Network anomaly detection and intrusion reporter (NADIR), The Audit data analysis and mining (ADAM) IDS in 2001.

[18]. Learning intrusion detection: supervised or unsupervised? Pavel Laskov, Patrick D¨ussel, Christin Sch¨afer and Konrad Rieck Fraunhofer-FIRST.IDA, Kekul´estr. 7, 12489 Berlin, Germany.

[19]. An Implementation of ID3: Decision Tree Learning Algorithm Wei Peng, Juhua Chen and Haiping Zhou Project of Comp 9417: Machine Learning University of New South Wales, School of Computer Science & Engineering, Sydney, NSW 2032, Australia.

K. Hanumantha Rao, working as Asst. Prof. at Sri Indu College of Engineering & Technology, Hyderabad, He has guided many PG level and Engineering students, areas of interest are Operating System, Mobile Computing and Information Security.

G. Srinivas, pursuing M.Tech CS at Sri Indu College of Engg & Tech from JNTU Hyderabad, areas of interest are Network Security, Wireless Sensor Networks and Mining.

Ankam Damodhar pursuing M.Tech CS at Sri Indu College of Engg. & Tech. from JNTU Hyderabad, areas of interest are Information Security, Mobile Computing, and Data Warehousing.

M. Vikas Krishna pursuing M.Tech CS at Sri Indu College of Engg. & Tech. from JNTU Hyderabad, areas of interest are Information Security, Mobile Computing, Data Warehousing and Mining.