INTRUSION DETECTION WITH KNN CLASSIFICATION AND DS- THEORY

Deepika Dave
Lakshmi Narayan College of Technology,
Bhopal, Madhya Pradesh
INDIA

Prof. Vineet Richhariya
Lakshmi Narayan College of Technology,
Bhopal, Madhya Pradesh
INDIA

***Abstract* - Intrusion detection is a awfully exigent area of research in a current scenario. Now a days find a novel pattern of intrusion and detection of this pattern are exceedingly demanding job. our object is we affect a method for intrusion detection using KNN classification and Dempster theory of evidence. Through these manners we gathered a new revealed pattern of intrusion and classifies Category of pattern and apply event evidence logic with the help of DS- Theory. Finned pattern of intrusion compare with the existing pattern if intrusion and generate a new schema of pattern and update a list of pattern of intrusion detection and improved the true rate of intrusion detection. we have also accomplish some experimental task with KDD99Cup and DARPA98 databases from MIT Lincoln Laboratory show that the proposed method provides competitively high detection rates compared with other machine-learning techniques and crisp data mining. The experimental results clearly show that the proposed system achieved higher precision in identifying whether the records abnormal or attack one.**

Keywords:-Intrusion Detection, KNN, DS-Theory, KDD DATA SET 99.

1.INTRODUCTION:

Intrusion detection systems are hefty element for computer network. as one of the main security tools and organization of communication infrastructure. An IDSs the term for a mechanism which quietly listens to network traffic in order to detect abnormal or suspicious activity. There are two district major families of IDSs.(a) N-IDS (Network based) - handle security at the network level. (b) H-IDS (Host based) – handle the security at host level.Intrustion detection is a technique for protecting the system when a network is being used by an unauthorized person.

Traditionally intrusion detection technique divide into following ways, (1)Misuse detection : Misuse detection technique focus on developing model of known attacks i.e. in this we have predefined patterns of abnormal files which can be described by specific patterns or sequence of the data and elements.(2) Anomaly Detection: The main aim of anomaly detection is to identify cases that are abnormal within data that are apparently uniform, anomaly detection is an important tool for detecting network intrusion and other rare events that may have great impact but are difficult to find. Anomaly detection refers to the manner of finding patterns in data that do not be conventional to expected behavior. Intrusion detection has emerged as a significant field of research, because it is not theoretically possible to set up a system with no vulnerabilities .One main confrontation in intrusion detection is that we have to find out the concealed attacks from a large quantity of routine communication activities . Several machine learning (ML) algorithms, for instance Neural Network ,Support Vector Machine, Genetic Algorithm ,Fuzzy Logic, and Data Mining and more have been extensively employed to detect intrusion activities both known and unknown from large quantity of complex and dynamic datasets. Generating rules is vital for IDSs to differentiate standard behaviors from strange behavior by examining the dataset which is a list of tasks created by the operating system that are registered into a file in historical sorted order .Various researches with data mining as the chief constituent has been carried to find out newly encountered intrusions. The analysis of data to determine relationships and discover concealed patterns of data which otherwise would go unobserved is known as data mining. Many researchers have used data mining to focus into the subject of database intrusion detection in databases

we have designed intrusion detection with KNN Classification and DS-Theory with fuzzy logic. The input to the proposed system is KDD Cup 1999 dataset, which is separated into two subsets such as, training dataset and testing dataset. Initially, the training dataset is classified into five subsets so that, four types of attacks (DoS (Denial of Service), R2L (Remote to Local), U2R (User to Root), Probe) and normal data are separated. After that, we simply mine the 1-length frequent items from attack data as well as normal data. These mined frequent items are used to find the important attributes of the input dataset and the identified effective attributes are used to generate a set of definite and indefinite rules using deviation method. Then, we generate fuzzy rule in accordance with the definite rule by fuzzifying it in such a way, we obtain a set of fuzzy if-then rules with consequent parts that represent whether it is a normal data or an abnormal data. These rules are given to the fuzzy rule base to effectively learn the fuzzy system. In the testing phase, the test data is matched with fuzzy rules to detect whether the test data is an abnormal data or a normal data. we apply KNN classification and Dempster theory of evidence on classify data. Through these we gathered a new discovered pattern of intrusion and classifies Category of pattern and apply event evidence logic with the help of DS- Theory. Finned pattern of intrusion compare with the existing pattern if intrusion and generate a new schema of pattern and update a list of pattern of intrusion detection and improved the true rate of intrusion detection. we used concept of the Dempster theory, this work on event evidence and find the validity of data and reduce the rate of intrusion. Here we also used the patterns of design of schema and data conversion, in data conversion first type intrusion detection in MATLAB , But data of intrusion data in overall in string format ,now we has use classification method. We have face various difficulties classification of data conversion string through numeric format for suitability of classification. The process of data conversion we used the ratio mapping, the ratio mapping concept used by the machine learning recepotrary organization for mapping of data string to numeric format.

The rest of the paper is organized as follows: In section 2 some related works are reviewed. Section 3 deals with KNN classifier. Section 4 Overview of DS-Theory. Section 5 KDD Dataset. Section 6 Describe our method and Section 7 shows Performance and results and Section 8 is Conclusion

2. LITERATURE SURVEY:

2.1 CLASSIFICATION METHOD BY FUZZY GNP –BASED CLASS ASSOCIATION RULES:

Ci Chen * ,Shingo Mabu* Chuan Yue [1] devise in the filled of intrustion detection with the approach As the fuzzy GNP based class association approach is designed for databases containig both discrete and continuous attribute as Network Connection Database,secific classification method is describe as a follows: The defination of the matching degree between the continous attribute A_i in rule r with q_i and testing data connection with value ai is:

$$\text{MatchDegree}(q_i, a_i) = Fq_i\,(a_i) \quad (1)$$

Where, Fq_i represent the membership function for linguistic term q_i.

And the matching between rule r (p continuos and q discrete attributes) and new unlabeled connection d is defined as:

$$\text{Match}_r(d) = \frac{1}{p+q}\Big(\sum_{i\in Ap} \text{MatchDegree}(q_i, a_i) + t\Big). \quad (2)$$

where.

i: index of continuous attribute in rule r;

Ap: set of suffixes of continuous attribute in rule r ;

p: number of continuous attribute in rule r;

q:number of discrete attribute in r;

t: :number of discrete attribute in new unlabeled connection d satisfying rulr r;

$\text{Match}_r(d)$ ranges from 0 to 1. If $\text{Match}_r(d)$ equals to 1.0, rule r matches coonection data d competly. While $\text{Match}_r(d)$ equals to 0, rule r does not matche connection d at all.Then the average matching between connection data d and all the rules in a certain rule pool is defined as:

$$\text{MATCHr}(d) = \frac{1}{|Rp|} \sum_{r\in Rp} \text{Matchr}(d) \quad (3)$$

Where R_p is the set of suffixes of extrated important class association rule in a cetain rules pool.

A. Classifier for misuse detection

The average matching betwen connection data d and all the rules in the normal rule in pool $\text{MATCH}_n(d)$ and the avearage matching between connection data d and all the rules in the intrustion rule pool $\text{MATCH}_i(d)$ are calculated and compared.

If $\text{MATCH}_r(d) \geq \text{MATCH}_r(d)$,connection data d is labaeld as normal. On the other hand if $\text{MATCH}_n(d) < \text{MATCH}_i(d)$ connection data d is labaeld as intrustion.

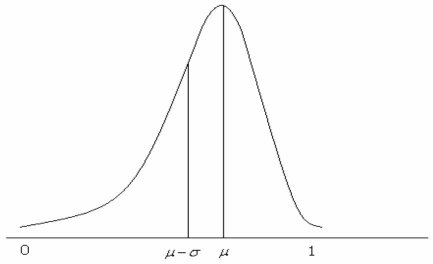
In summary, a new connection data is labeled according to their matching with normal and intrustion rule pools.Larger matching suggests the heigher possibilty of belonging to this class.

B. Classifier for anomaly detection

After getting matching between each connection data and rules in the normal rule pool. We can have the distribution of the matching with the mean value μ and standard deviation σ. Fig showes one example of the distribution.

In this testing peroid ,when a new unlabled connection data comes ,the matching between the data and the rules in normal rule pool is calculated. If $\text{MATCH}_n(d) < (\mu\text{-}k\sigma)$,label the connection as intrustion. On the hand,if $\text{MATCH}_n(d) \geq (\mu\text{-}k\sigma)$, label is normal. By adjusting parameter k, we can balance the PFR (Positive False Rate) and NFR(Negative False Rate).

In all, by using the improvrd Fuzzy GNP –based class association rule mining . we can find a large number of rules related to normal behaviour so as to explore the space of the normal connections. And any significant deviation from the normal space is viewed as an intrustion.



2.2 Probabilistic classification:

Nannan Lu, Shingo Mabu, Wenjing LI [2] devise in the filled of intrustion detection with the Nannan Lu, Shingo Mabu, Wenjing LI [2] devise in the filled of intrustion detection with the approach as: After extracting a number of important class association rules including normal and intrustion, a classifier is constructed to classify new connection data into normal ,misuse and anomaly intrusion correctly. The key points probabilistic classification concerns three aspects. First , the probability density function of the average matching degree of data with rules is used .Second, the probability that data is classified to anomaly intrustion also considerd.Third ,in order to improve the classification accuracy, weights are used to revise the probability approach as: After extracting a number of important class association rules including normal and intrustion, a classifier is constructed to classify new connection data into normal ,misuse and anomaly intrusion correctly. The key points probabilistic classification concerns three aspects. First , the probabilty density function of the average matching degree of data with rules is used .Second, the probability that data is classified to anomaly intrustion also considerd.Third ,in order to improve the classification accuracy.

$\text{MatchDegree}_k(Q_i, a_i) = F_{Qi}(a_i)$

Where F_{Qi} represents the membership function of linguistic term Q_i. Then, the matching degree between data and rule r (including p continuous attributes and q discrete attributes) is defined as:

$$Match_k(d, r) = \frac{1}{p+q}\ (\sum_{i\in CA} MatchDegree_k\,(Q_i,a_i)+t),\quad (5)$$

Where, I is the suffix of continuous attributes in rule r; CA denotes the set of suffix of continuous attributes in rule r; p and q represent the number of continuous attribute and discrete attributes in rule r, respectively, and t is the number of matched discrete attributes in rule r with data. Then, the average matching degree can be defined as

$$m_k(d) = \frac{1}{|Rk|}\quad (\sum_{k\in C} Match_k\,(d, r),\quad (6)$$

where, R_k is the set of suffixes of the extracted rules in class k in the rule pool(normal rules or misuse rules). Finally, the marginal probability density function $f_1(x_1)$, $f_2(x_2)$,…$f_K(x_k)$ can be generated by calculating the distribution of the average matching degree of training data d ε Dtrain(k) with r ε R_k, where, Dtrain(k) is the set of suffix of training data in class k. K=2 is used in this paper.

A. Building a Classifier

After creating the probability density function $f_K(x_k)$ of the average matching degree between training data d εD_{train}(k) and rule r ε R_k, the probability that new connection data d εD_{test} belongs to class k is represented as follow:

$$P_k(d)= \int_{mK(d)}^{1.0} f_K(x_K)dx_k \ldots \sum_{k\in C} f_K(x_K)dx_k.\ .$$

$$\int_{ml(d)}^{1.0} f_1(x_1)dx_1,\ (7)$$

where, Dtest is the set of suffix of testing data. Actually, the probability that d Sigma Dtest belongs to anomaly class is defined as:

$$P_0(d) = \sum_{k\in C} 1- P_k(d)\quad (8)$$

Where, C is the set of suffix of classes having training data. In the case of two classes, the probabilities of the first class and the second class can be calculated by the following equations.

$$P_1(d)= \int_{m2(d)}^{1.0} f_2(x_2)dx_2 \int_{0}^{ml(d)} f_1(x_1)dx_1\quad (9)$$

$$P_2(d)= \int_{0}^{m2(d)} f_2(x_2)dx2 \int_{ml(d)}^{1.0} f_1(x_1)dx_1\quad (10)$$

Then, the probability that a new connection data belongs to anomaly class is calculated by $P_0(d) = 1-\sum_{k\in C} P_k(d)$. Based on the calculation of these probabilities, d is assigned to the class with highest probability.

3. K-NN (KNOWN NEAREST NEIGHBOR)

KNN is a *non parametric lazy learning* algorithm. That is a pretty concise statement. When you say a technique is non parametric, it means that it does not make any assumptions on the underlying data distribution. This is pretty useful, as in the real world, most of the practical data does not obey the typical theoretical assumptions made (eg gaussian mixtures, linearly separable etc) . Non parametric algorithms like KNN come to the rescue here.

It is also a lazy algorithm. What this means is that it does not use the training data points to do any *generalization*. In other words, there is *no explicit training phase* or it is very minimal. This means the training phase is pretty fast . Lack of generalization means that KNN keeps all the training data. More exactly, all the training data is needed during the testing phase. (Well this is an exaggeration, but not far from truth). This is in contrast to other techniques like SVM where you can discard all non support vectors without any problem. Most of the lazy algorithms – especially KNN – makes decision based on the entire training data set (in the best case a subset of them).

There are various methods which can be used to determine nearest neighbor. Figure 3.1 shows the way in which decision is taken to decide the category of new point.

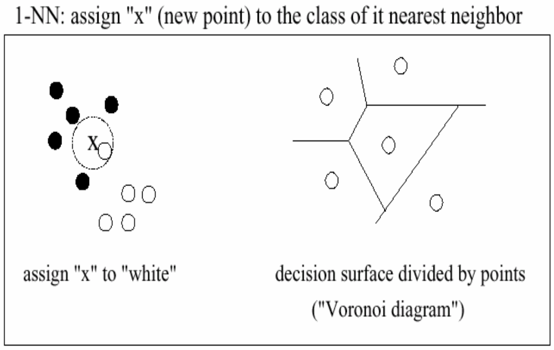

1-NN: assign "x" (new point) to the class of it nearest neighbor
x
assign "x" to "white"
decision surface divided by points
("Voronoi diagram")

Figure 3.1 Decision of nearest neighbor

Figure 3.2 and 3.3 shows various methods for deciding the nearest neighbor.

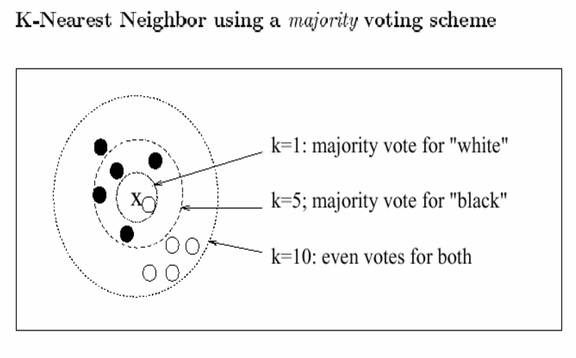

K-Nearest Neighbor using a *majority* voting scheme
X
k=1: majority vote for "white"
k=5; majority vote for "black"
k=10: even votes for both

Figure 3.2 Majority voting scheme

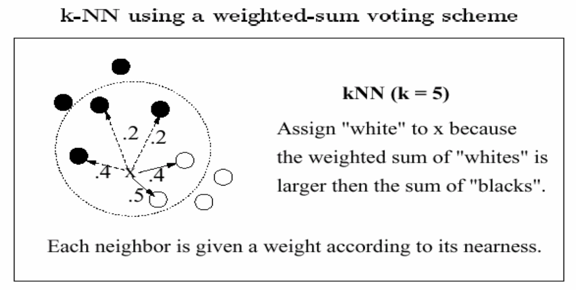

k-NN using a weighted-sum voting scheme
kNN (k = 5)
Assign "white" to x because the weighted sum of "whites" is larger then the sum of "blacks".
.2
.2
.4
x
.4
.5
Each neighbor is given a weight according to its nearness.

Figure3.3 Weighted-sum voting scheme

k-NN is a kind of example-based text categorization algorithm. However, the determination of the *k* has not yet got good solution. Moreover, the good selection of *k* most Similar texts also have bigger effect on categorization results. Also k-NN cannot effectively solve the problem overlapped category borders.

Statistical rules are used in general in the classification of textual information, which include several tasks in Information Retrieval. It includes not only the determination of good documents in terms of relevance attending to user needs but also the classification of documents into categories (topics) attending to prede.ned classes [18]. In the following, we include studies found in the literature about both the retrieval and the categorization tasks.

The use of rules for categorization comes from a process of classification of documents into different categories regarding their topics in order to optimize a posteriori retrieval process. One of the most relevant works of categorization using rules is the one of [20]. The general idea of this work is the discovery of classification patterns automatically for document categorization. The aim of the induction process is to and sets of decision rules to distinguish among different categories which documents belong to. The attributes of the rules can be one word or a pair of words constructing a dictionary where an elimination process of the less frequent words is carried out. Finally, association rules have been also used for categorization [21], where the authors propose a solution for text categorization based on the application of the best generated association rules to build a classifier.

4.THE DEMPSTER –SHEFER THEORY (DST)

The Dempster –Shefer theory(DST) of evidence originated in the work of [3,4]on theory of probabilities with upper and lower bounds. It has since been extended by numerous authors and popularized, but only to a degree, in the literature on Artificial Intelligence (AI) and expert systems , as a technique for modeling reasoning under uncertainty. In this respect it can be seen to offer numerous advantages over the more “traditional” methods of Statistics and Bayesian decision theory. Hajek [5] remarked that real, practical applications of DST methods have been rare, but subsequent to these remarks there has been a marked increase in the applications incorporating the use of DST. Although DST is not in widespread use, it has been applied with some success to such topics as face recognition [6], statistical classification [7] and target identification [8]. Additional applications centered on multi-source information, including medical diagnosis [9] and plan recognition [10]. An exception is the paper by cortes –Rello and Golshani [11], which although written for a computing science /AI readership does deal with the “knowledge domain” of forecasting and Marketing Planning. For those with even limited knowledge og these domains the paper appears rather naive , referring for example to rather naive. Referring for example to rather venerable old editions of standard texts such as[12].the aim of this paper is to suggest that there is a good deal of potential in the DST approach, which is as yet very largely unexploited. The origins of the mathematical theory of probability date back at least to the work of the eighteenth century scholar, The Reversed Thomas [13],whose work was published posthumously in 1763.it provides the foundations for the theory of statistical inference (involving both estimation and testing of hypotheses) and for techniques of design making under certainty. The roots of decision analysis lie in the 1930s and 1940s .Wald[14], included the “complete class theorem” ,which stated that any procedure in a statistical decision problem can be beaten or at least matched in performance by Bayesian procedure, defined as procedure based on the adoption of some set of prior probabilities . The fact that numerous statistical principles and techniques may be developed without using prior and posterior probability distribution involves no loss of generality, given that the special case of a uniform or rectangular prior distribution may be adopted. Decision analysis relies more on a subjectivist view of the use of probability, whereby the probability of an event indicates the degree to which someone believes it, rather than the alternative frequents approach .The latter approach is based only on the number of times an event is observed to occur .As savage [15,16] discusses ,the subjectivists have been responsible for much of the theoretical work into statically practice. He goes on to argue that the frequentists hold an uneasy upper hand over their Bayesian / subjective colleagues in the domain of mathematical statistics. Bayesian statisticians may agree that their goal is to estimate objective probabilities from frequency data, but they advocate using subjective prior probabilities to improve the estimates [17] . french questions savages’s theriry of subjective expected utility, which suggests that each of us has within us an exact subjective probability for each possible event in the small world (model) under consideration. For a much fuller discussion of subjective and frequentists approaches see the collection of papers in [18] who notes that the three defining attributes of the Bayesian approach are:

1.Reliance on a complete probabilistic model of the domain or "frame of discernment".

2.Willingness to aaccept subjective judgement as an expedient substitute for empirical data.

3.the use of Bayes Therom (conditionality) the primarly mechanism for updating belifes in ,light of new information. However,The Bayesian technique is not without its critics including among others Walleyl[19],as well as Caselton ans Luo[20] who discussed the difficulty arising when conventional Bayesian analysis is presented only with weak information sources. In such cases we have the "Bayesian domega of precision ",whereby the information concerning uncertain statistical parameters, no matter how vague, must be represented by conventional exactly specified ,probability distribution.

Some of the difficulties can be understood through the "principle of Insufficient Reason" as illustrated by Wilson [21].Suppose we are given a random device that randomly generates integer numbers between 1 and 6(its "frame of discernment") but with unknown chances. What is our belief in"1" being the next number? A Bayesian will use a symmetry argument, or the Principle of insufficient Reason to say that the Bayesian belief in a "1" being the next number, say P(1) should be 1/6. In general in a situation of ignorance a Bayesian is force to use this principle to evenly allocate subjective (additive) probabilities over the frame of discernment.

To further understand the Bayesian approach, especially with the regard to representation of ignorance, consider the following example, similar to that in [21]. Let a be a preposition that; "I live in Kings Road, Cardiff".

How could one construct P(a), a Bayesian belief in a? Firstly we must choose a frame of discernment, denoted by Θ and a subset A of Θ representing the preposition a; then would need to use the Principle of Insufficient Reason to arrive at a Bayesian belief. The problem is there are number of possible frames of discernment Θ that we could choose, depending effectively on how many Cardiff roads can be enumerated. If only two such streams are identifiable, then $\Theta=\{x_1,x_2\}$,$A=\{x_1\}$.The "Principle of Insufficient Reason" then gives P(a), to be 0.5,through evenly allocating subjective probabilities over the frame of discernment. If it is estimated that there are about 1000 roads in Cardiff, then $\Theta=\{x_1,x_{2,\ldots\ldots}\ x_{1000}\}$ with again $A=\{x_i\}$ and other $x_{i\ 's}$ representing the other roads. In this case the "theory of insufficient reason" gives P(A)=0.001. Either of these frames may be reasonable, but the probability assigned to A is crucially dependent upon the frame chosen. Hence once Bayesian belief is a function not only of the information given and one's background knowledge, but also of sometimes arbitrary choice of frame of discernment. To put the point another way, we need to distinguish between uncertainty and ignorance. Similar arguments hold where we are discussing not probabilities per se but weights which measure subjective assessments of relative importance. This issue arises in decision support models such as the Analytic Hierarchy Process (AHP), which requires that certain weights on a given level of decision tree to unity see [22].

5.KDD DATA SET 99

In 1998, DARPA in concert with Lincoln Laboratory at MIT launched the DARPA 1998 dataset for evaluating IDS [25]. The DARPA 1998 dataset contains seven weeks of training and also two weeks of testing data. In total, there are 38 attacks in training data as well as in testing data. The refined version of DARPA dataset which contains only network data (i.e. Tcpdump data) is termed as KDD dataset. The Third International Knowledge Discovery and Data Mining Tools Competition were held in colligation with KDD-99, the Fifth International Conference on Knowledge Discovery and Data Mining. KDD dataset is a dataset employed for this Third International Knowledge Discovery and Data Mining Tools Competition. KDD training dataset consists of relatively 4,900,000 single connection vectors where each single connection vectors consists of 41 features and is marked as either normal or an attack, with exactly one particular attack type [25]. These features had all forms of continuous and symbolic with extensively varying ranges falling in four categories:

- In a connection, the first category consists of the ***intrinsic*** features which comprises of the fundamental features of each individual TCP connections. Some of the features for each individual TCP connections are duration of the connection, the type of the protocol (TCP, UDP, etc.) and network service (http,telnet, etc.).
- The ***content*** features suggested by domain knowledge are used to assess the payload of the original TCP packets, such as the number of failed login attempts.
- Within a connection, the ***same host*** features observe the recognized connections that have the same destination host as present connection in past two seconds and the statistics related to the protocol behavior, service, etc are estimated.
- The ***similar same service*** features scrutinize the connections that have the same service as the current connection in past two seconds.

A variety of attacks incorporated in the dataset fall into following four major categories: **Denial of Service Attacks:** A denial of service attack is an attack where the attacker constructs some computing or memory resource fully occupied or unavailable to manage legitimate requirements, or reject legitimate users right to use a machine. **User to Root**

Attacks: User to Root exploits are a category of exploits where the attacker initiate by accessing a normal user account on the system (possibly achieved by tracking down the passwords, a dictionary attack, or social engineering) and take advantage of some susceptibility to achieve root access to the system.

Remote to User Attacks: A Remote to User attack takes place when an attacker who has the capability to send packets to a machine over a network but does not have an account on that machine, makes use of some vulnerability to achieve local access as a user of that machine. **Probes:** Probing is a category of attacks where an attacker examines a network to collect information or discover well-known vulnerabilities. These network investigations are reasonably valuable for an attacker who is staging an attack in future. An attacker who has a record, of which machines and services are accessible on a given network, can make use of this information to look for fragile points. Table5. 1 illustrates a number of attacks falling into four major categories :

Denial of Service Attacks	**Back, land, neptune, pod, smurf, teardrop**
User to Root Attacks	**Buffer_overflow, loadmodule, perl, rootkit,**
Remote to Local Attacks	**Ftp_write, guess_passwd, imap, multihop, phf, spy, warezclient, warezmaster**
Probes	**Satan, ipsweep, nmap, portsweep**

Table 5.1 Various types of attacks described in four major categories

6. METHOD:

The proposed system introduces intrusion detection system with KNN Classification and DS-Theory with fuzzy logic. The input to the proposed system is KDD Cup 1999 dataset, which is separated into two subsets such as, training dataset and testing dataset. Initially, the training dataset is classified into five subsets so that, four types of attacks (DoS (Denial of Service), R2L (Remote to Local), U2R (User to Root), Probe) and normal data are separated. After that, we simply mine the 1-length frequent items from attack data as well as normal data. These mined frequent items are used to find the important attributes of the input dataset and the identified effective attributes are used to generate a set of definite and indefinite rules using deviation method. Then, we generate fuzzy rule in accordance with the definite rule by fuzzifying it in such a way, we obtain a set of fuzzy if-then rules with consequent parts that represent whether it is a normal data or an abnormal data. These rules are given to the fuzzy rule base to effectively learn the fuzzy system. In the testing phase, the test data is matched with fuzzy rules to detect whether the test data is an abnormal data or a normal data. We apply KNN classification and Dempster theory of evidence on classifies data. Through these we gathered a new discovered pattern of intrusion and classifies Category of pattern and apply event evidence logic with the help of DS-Theory. Finned pattern of intrusion compare with the existing pattern if intrusion and generate a new schema of pattern and update a list of pattern of intrusion detection and improved the true rate of intrusion detection. We used concept of the Dempster theory, this work on event evidence and find the validity of data and reduce the rate of intrusion. Here we also used the patterns of design of schema and data conversion, in data conversion first type intrusion detection in MATLAB , But data of intrusion data in overall in string format ,now we has use classification method. We have faced various difficulties classification of data conversion string through numeric format for suitability of classification. The process of data conversion we used the ratio mapping, the ratio mapping concept used by the machine learning recepotrary organization for mapping of data string to numeric format. The above procedure is explained and depicted with the help of a flow chart mention below ;

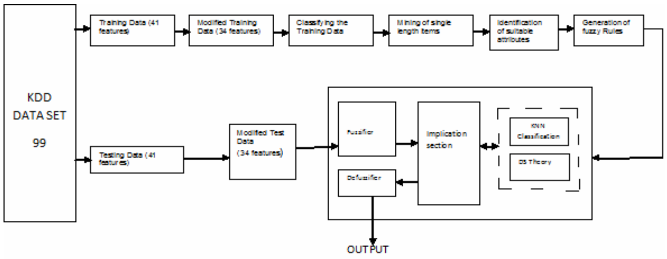

KDD DATA SET 99
Training Data (41 features)
Modified Training Data (34 features)
Classifying the Training Data
Mining of single length items
Identification of suitable attributes
Generation of fuzzy Rules
Testing Data (41 features)
Modified Test Data (34 features)
Fuzzifier
Implication section
KNN Classification
DS Theory
Defuzzifier
OUTPUT

Fig 6.1: The overall steps of the proposed intrusion detection system

7. EXPERIMENTAL RESULTS & PERFORMANCEANALYSIS

This section describes the experimental results and performance evaluation of the proposed system. The proposed system is implemented in MATLAB (R2010a) and the performance of the system is evaluated *False positive rate ,False negative rate, True positive rate, True negative rate and accuracy in respect of true positive and true negative rate*. For experimental evaluation, we have taken KDD cup 99 dataset, which is mostly used for evaluating the performance of the intrusion detection system. Here, we have used only 1000 instances of data of KDD Cup 99 dataset for training and testing.

We have supervised five data set with each 1000 instances of data under .the result of ratio of attacks is represented in tabular format below:

CATEGORY	**DATA SET 1**	**DATA SET 2**	**DATA SET 3**	**DATA SET 4**	**DATA SET 5**
Normal	650	645	652	643	647
Probs	50	52	49	53	50
DoS	150	160	148	144	148
U2R	100	90	97	109	100
R2L	50	53	54	51	55

Table7.1: Result of Tested Data set

7.1PERFORMANCE ANALYSIS:

As seen from the output of performance on data sets it can be made out that when KNN is combined with DS method, the performance gets significantly improved.

Earlier application of isolated KNN on dataset has much greater Accuracy, than later by integrating both KNN and DS Methods. Also there is a considerable enhancement in the true positive and true negative detection ratio and false positive and false negative ratio .Thus this gives the direct improvised accuracy in the result. In this paper, we are showing the result for the parameters - **Accuracy**, **False positive rate (FP), False negative rate (FN), True positive rate (TP), True negative rate (TN)** only for one data set i-e for data set-1.Also, below we are showing the graph for that particular data set .Also below we are showing how to calculate these parameters by the suitable formulas.

$$\text{FP} = \frac{\textbf{correct detection}}{\textbf{Total Dtection}}$$

FN = Total detection-false positive

$$\text{Accuracy} = \frac{\mathbf{TP + TN}}{\mathbf{TP + FP + TN + FN}}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

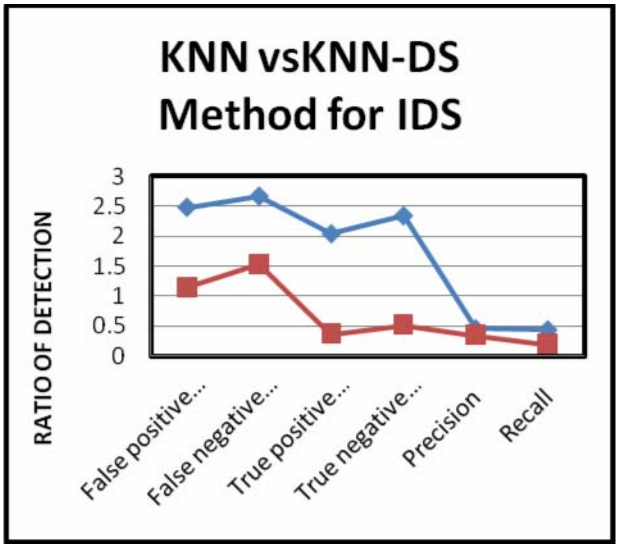
$$\text{Recall} = \frac{TP}{TP + FN}$$

Where,
TP: True Positive
TN: True Negative
FP: False Positive
FN: False Negative

Metric		Accuracy	FP)	(FN)	(TP)	(TN)	Preci-sion	Re-call
DATASET-1	KNN	92.144953	2.272320	2.749558	2.244299	2.618349	0.496942	0.449411
	KNN-DS	97.476417	1.258918	1.359074	0.721024	0.841194	0.3641642	0.346629
DATASET-2	KNN	89.906531	2.919854	3.533089	2.883848	3.64490	0.496891	0.445769
	KNN-DS	95.237995	2.375580	2.564574	1.360573	1.587335	0.364164	0.5029196
DATASET-3	KNN	88.206532	2.782502	2.8118753	2.942841	2.993842	0.514002	0.511379
	KNN-DS	94.250632	2.112374	1.553012	1.423994	0.849842	0.402671	0.478339
DATASET-4	KNN	91.233451	2.143543	2.654321	2.147892	2.978654	0.617017	0.447271
	KNN-DS	95.437654	1.145372	1.878653	0.927686	0.836398	0.447496	0.3305680
DATASET-5	KNN	92.226543	2.478562	2.664321	2.036743	2.336421	0.451075	0.433251
	KNN-DS	97.135122	1.136487	1.527689	0.362765	0.513754	0.341963	0.191893

Table7.2: Parameter Result of Tested Data set 1

Comparative Graph Chart of KNN and KNN-DS Result on given Dataset-1 for Intrusion Detection System



8. CONCLUSION:

Our dissertation presents the performance of Intrusion detection system on application of our new design technique. We have designed an Intrusion detection system using KNN Classification And Dempster Theory for detecting intrusion behavior within the network .As compared to the earlier technique used ,the combined use of KNN And Dempster theory ,it's found out that ,the performance get considerably enhanced. This improvised design technique gives more efficient results. it was observe that KNN And Dempster can perform better and almost situation, Which is further proven by comparing the result on KDD Data Set 99. Our Experiment on different dataset classifies the data using KNN classification (Normal Packet, DOS, R2L, U2R, Probes) and later the factor of evidence is formulated by using DS theory. The new pattern of intrusion is compared with the existing pattern of intrusion and generates a new schema of pattern and updates a list of pattern of intrusion detection and improved the true rate of intrusion detection.

9. FUTURE WORK:

The work can be extended by studying nitty-gritty of data mining techniques and the fundamentals of intrusion detection system and network behavior patterns. As a piece of future work, our design can be clubbed up with more optimize classification technique. This improvised structure will increase the efficiency and will give improvised result; also the design can be made more comprehensive by supervising data from varied data sources and examining more complicated intrusion network scenarios.

REFRENCES:

[1] JiaWei Han,Micheline Karnber, "Data Mining: Concept and Technology" [M]. China Machine Press, 2001.8

[2] Freund Y. "Boosting a Weak Learning Algorithm by Majority"[J]. Information and Computation, 1995,121(2):256-285

[3]The Dempster Shafer theory of evidence: an alternative approach to multicriteria decision modelling Malcolm Beynon, Bruce Curry*, Peter Morgan Cardi Business School, Colum Drive, Cardi, CF1 3EU, UK Received 1 December 1998; accepted 1 June 1999.

[4]Dempster AP. Upper and lower probabilities induced by a muilti-valued mapping. Ann Math Stat 1967;38:325-39.

[5] Hajek P. Systems of conditional beliefs in Dempster Shafer theory and expert systems. Int J General Systems 1994;22:113-24.

[6] Ip HHS, Ng JMC. Human face recognition using Dempster±Shafer theory. In: ICIP. 1st International Conference on Image Processing, vol. 2, 1994. p. 292-5.

[7] Denoeux T. A k-nearest neighbour classi®cation rule based on Dempster±Shafer theory. IEEE Transactions on Systems, Man and Cybernetics 1995;25(5):804-13.

[8] Buede DM, Girardi P. A target identi®cation comparison of Bayesian and Dempster±Shafer multisensor fusion. IEEE Transaction on Systems, Man and Cybernetics-
Part A: Systems and Humans 1997;27(5):569-77.

[9] Yen J. GERTIS: A Dempster±Shafer approach to diag- nosing hierarchical hypotheses. Commun ACM1989;32(5):573-85.

[10] Bauer M. A Dempster±Shafer approach to modeling agent preferences for plan recognition. User Modeling and User-Adapted Interaction 1996;5:317-48.

[11]Cortes-Rello E, Golshani F. Uncertain reasoning using the Dempster-Shafer method: an application in forecasting and marketing management. Expert Systems 1990;7(1):9-17.

[12] Kotler P. Marketing management: analysis, planning and control. Englewood Cliffs, NJ: Prentice Hall, 1980.

[13].Bayes T. An essay toward solving a problem in the doc- trine of chances. Phil Trans Roy Soc (London) 1763;53:370-418.

[14]. Wald A. Statistical decision functions. New York: Wiley, 1950.

[15] Savage LJ. The foundations of statistics. New York:Wiley, 1954 (2nd rev.ed., 1972 Dover).

[16] Savage LJ. The foundation of statistics reconsidered. In: Proceedings of the Fourth Berkeley Symposium on Mathematics and Probability 1. Berkeley: University of California Press, 1961.

[17] Good IJ. Good thinking: the foundations of probability and its applications. Minneapolis: University of Minnesota Press, 1983.

[18]Shafer G, Pearl J. Readings in uncertain reasoning. San Mateo, CA: Morgan Kaufman, 1990.

[19] Walley P. Belief-function representations of statistical evidence. Ann Stat 1987;10:741-61.

[20] Caselton WF, Luo W. Decision making with imprecise probabilities: Dempster±Shafer theory and applications. Water Resources Research 1992;28(12):3071-83.

[21] Wilson PN. Some theoretical aspects of the Dempster± Shafer theory. PhD Thesis, Oxford Polytechnic, 1992.

[22] Saaty TL. The Analytic Hierarchy Process: planning, pri-ority setting, resource allocation. New York: McGraw- Hill, 1980.

[23] R. Shanmugavadivu , Dr.N.Nagarajan "KDD CUP 99 DATASET NETWORK INTRUSION DETECTION SYSTEM USING FUZZY LOGIC ", 1998.