International Journal of Scientific & Engineering Research, Volume 5, Issue 7, July-2014
ISSN 2229-5518

Anomaly Intrusion Detection Techniques: A Brief Review

Anurag Jain[#1], Bhupendra Verma[#2], and J. L. Rana[#3]
[#] RageevGandhi Technical University
[1]anurag.akjain@gmail.com, [2]bk_verma3@rediffmail.com, [3]jl_rana@yahoo.com

Abstract-In a broader sense detection of any unauthorized access of any information system is the basic aim of any intrusion detection system. However due to cost considerations it is practically impossible to provide total protection to an information system from intrusion for its entire useful life time. In this paper we provide a brief introduction to anomaly based intrusion detection systems that classify all reported techniques, including artificial immune systems (AIS), fuzzy logic (FL), swarm intelligence (SI), artificial neural networks (ANN), evolutionary computation (EC), and soft computing (SC). The various techniques of anomaly based intrusion detection system reported in the literature have been sorted out on the parameters like their strength and weakness. The important research contributions have been systematically compared and summarized to reflect the current status of research and challenges ahead. This will be helpful in knowing the new research directions. We also highlight the role of machine learning techniques for IDS. The contributions of research papers based on machine learning (ML) have also been considered. ML system have intrinsic properties like resilience to noisy data, adaptability, fault tolerance, robustness, low computational overhead etc, that provide a versatile tool in developing better intrusion detection techniques. We aim at providing a concise but comprehensive overview of research in progress and give direction to intrusion detection methods based on ensemble of ML techniques. This review work should be helpful and also provide critical insight into the current trend in IDS research especially in the application of ML approaches to IDS and related fields.

Index Terms— Intrusion detection system (IDS), Signature based intrusion detection system (S-IDS), Anomaly based intrusion detection system (A-IDS),Machine Learning based detection (ML-IDS), Knowledge based detection (K-IDS), Data Mining based detection (DM-IDS),Statistical Anomaly based detection (SA-IDS), Multi classifier approaches (MCA),Adaptive and Scalable intrusion detection scheme (ASIDS)

INTRODUCTION

Internet usage has percolated to all walks of human life. All those information and processes which otherwise needed protection also need more protection when internet is the main via media. The situation becomes even more difficult as independent of geographical location, internet supports any time any where connectivity. This has complicated the security problems and made it a major research area for business and personal networks. Intrusion detection systems (IDS) complement the prevention based security measures for safeguarding networks and connected systems from intrusive activities. Basically any deviation from the normal uses of the system is an intrusion requiring monitoring an analysis of the event occurring in the information system. Normally a general assumption is that these deviations are noticeably deferent from normal uses of system [1]. All the research work for Intrusion Detection is based on this assumption. First violation has to occur then only Intrusion detection system comes in the picture. If any Intrusion occurs, it may cause harm to the information system such as lose of confidentially and integrity, denial and illegal use of resources. Therefore IDS system is considered as second line of defense.

Most of IDS are software but it may work in combination with hardware. Audit data logs capture the behavior pattern of the users and details of connection uses. Using these information majority of IDSs try to detect intrusion in real time. As an example, for a TCP connection log records may contain number of attempts and duration of use for all source and destination. From the myriad of information, it is time consuming to discover such intricate relationships that exist between different features in the log. In real time detection, much less computational and memory recourses are available as such real time IDS should be computation and space wise efficient. System log may contain certain features of information that may be in other logs as well, making some data redundant. Such redundant data are not useful so it can be discarded. Similarly some of the features may contain noisy or false correlations at times which can cause problems in complex classification techniques. Removal of these redundant features resulting reduction in computational time as well as improvement in accuracy of IDS. By properly selecting feature subset that yields best classification in training data [2]. Classification results can be further improved.

Expert system and Fuzzy inference systems are some of the popular machine learning paradigms which have been reported in the literature for development of IDS. All though, it has been convincingly shown that some features do not contribute for better classification and can be left out without significant degradation in performance of the IDS. However no optimal methodology has been reported in the literature that can efficiently model IDS feature selection. More often than not the IDS tasks are modeled for

classification problem in the context of machine learning. Lately interest is more centered on use of fuzzy logic. Therefore more focus is given in presenting literature review of fuzzy based ensemble approach for IDS. In the section II, various types of intrusion detection systems and their pros and cons in anomaly and misuse detection are discussed. In section III, review of intrusion detection techniques is presented. In this section, various anomaly intrusion detection systems are reviewed briefly. In section IV a comparative table of various intrusion detection techniques is given. In section V some underlying advantages of multi classifiers approaches for IDS are elaborated. Various fuzzy based ensemble techniques for intrusion detection are discussed and compared. In section VI a methodology is proposed with brief discussion on the outcome, identify challenges and possible solutions are presented. The outcomes finally concluded in section VII.

II. INTRUSION DETECTION METHODS

A. Signature based intrusion detection system (S-IDS)

This method is for detection of misuse. Behavioral Pattern (signatures) of known attacks are stored in a database. The detection process matches the events pattern against the stored signatures. If a match is found an intrusion signal is generated. It has a major drawback that this method fails to identify new attacks whose patterns are not previously stored or same as known attacks [4].

B. Anomaly based intrusion detection system (A-IDS)

A-IDS detect unknown or novelty attacks. New or unknown attacks are known as novelty attacks. A-IDS detect such attacks which were not used for training the machine learning system [3].

C. Pros and cons of anomaly detection and misuse detection are given in table 1.

Table1: pros and cons of S-IDS and A-IDS

	Pros	**Cons**
S-IDS	Attacks which are similar to attack data used during training are detected with high accuracy.	Does not detect new or unknown attacks. Even common attacks with slight variations often go undetected.
A-IDS	It can detect new type or unknown attacks. Small changes in previously known attacks, including deviations from normal pattern of usage irrespective of privileged/ internal or an unauthorized external user can be detected.	At times it may fail to detect even well-known attacks, particularly if there is match with the usage profile of established user. Because of this, if once the attack is identified, for purpose of forensic investigation proper characterization of nature of attack becomes difficult.

There are two main advantages of A-IDS over S-IDS. Firstly it can detect novelty i.e. "zero days" attacks. This is because of its ability to detect abnormal behaviors. Secondly A-IDS can be customized for normal activities of different systems. Thus for an attacker it becomes difficult to be aware about what kind of his activities can go on undetected [5].

III. A BRIEF REVIEW AND COMPARISON OF VARIOUS A-IDS TECHNIQUES

Among various A-IDS techniques, the most important ones are machine learning or computational intelligence based detection (ML-IDS), knowledge based detection (K-IDS),Data Mining based detection(DM-IDS) and statistical anomaly based detection (SA-IDS) Figure-1 shows the proposed taxonomy of A-IDS.

1. *Machine learning or computational intelligence based intrusion detection system (ML-IDS):*

Machine learning is basically an ability of a program and/or a system to progressively improve their performance by learning over a time while performing a given task or tasks. Previous results are analyzed for correctness and accuracy based on this machine learning technique improves itself for better performance. This means recently acquired information is used to change the execution which is known as machine learning [6]. This feature makes it desirable to use Machine learning techniques in many situations. It has one major drawback that it is resource expensive. Various machine learning approaches can be listed as- decision tree learning, association rule learning, artificial neural networks, genetic programming, inductive logic programming, support vector machines, clustering, Bayesian networks, reinforcement learning, representation learning, sparse dictionary learning, etc. Among the above listed approaches, for intrusion detection more popular ones are - Bayesian network, artificial neural network, fuzzy, evolutionary computation (genetic algorithm), artificial immune system, swarm intelligence and soft computing. Researchers have classified ML-IDS classified into following categories:

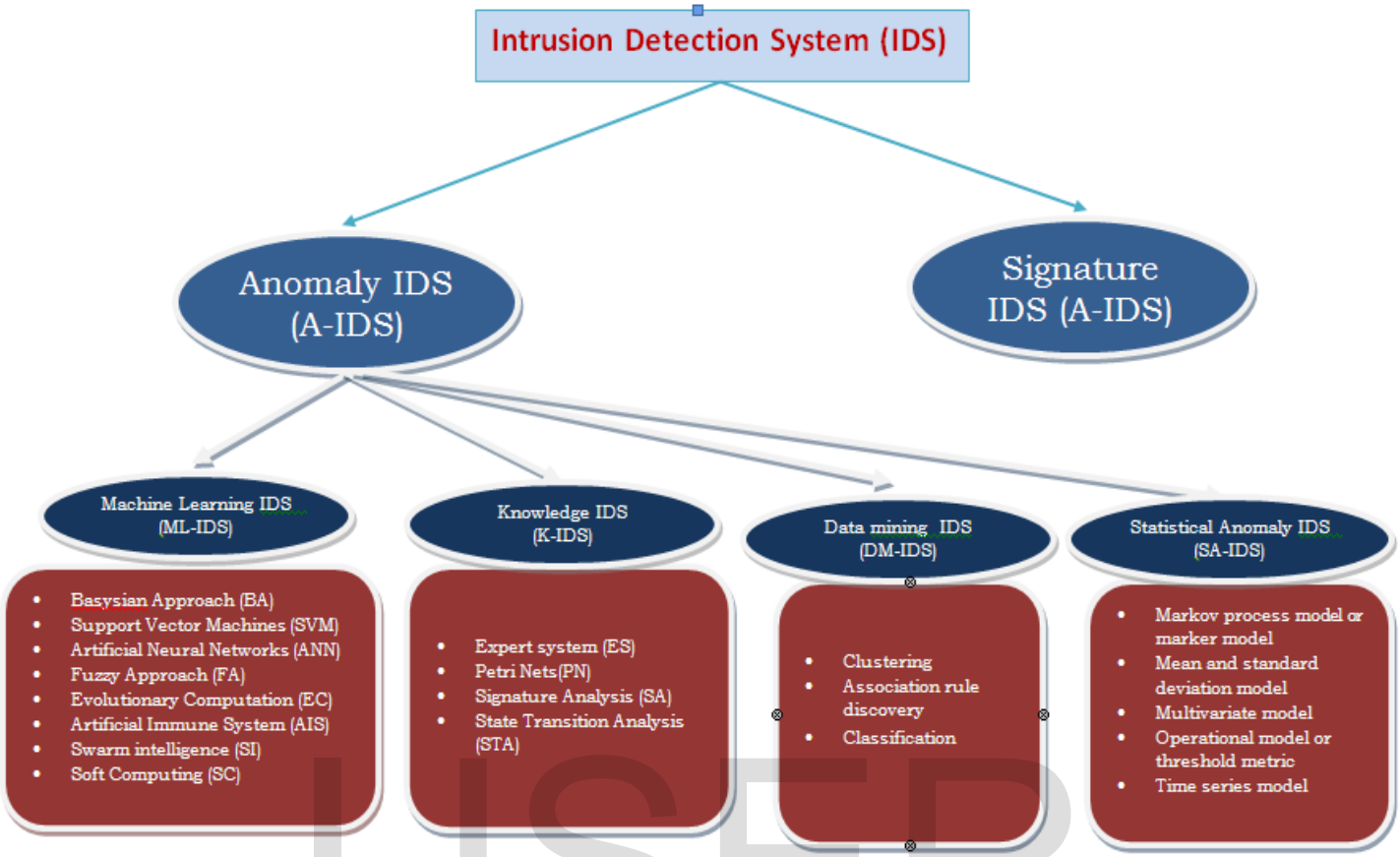

Fig1: Hierarchical Classification of IDS

1.1 Basysian Approach (BA): Probabilistic relationships of interest among variables are graphically modeled using a Basysian approach. In expert system also Basysian approach is convenient for encodings uncertain expert knowledge base. Techniques for learning Basysian networks from data have been recently proposed. These techniques have been proved to be remarkably reflective for certain type of data analysis problems. Many advantages are there when Bayesian approach is combined with other statistical methods for intrusion detection [8]. However when used directly, results of Bayesian networks are similar to results obtained from threshold-based intrusion detection systems but requiring relatively more computational efforts. A slightly different technique called pseudo-Bayes estimators has been proposed as an enhancement. This new method can detect new intrusions with lesser false alarm rate and lesser requirement of computational resources.

1.2 Support Vector Machines (SVM): Vapnik 1998 [9] was first to proposed Support vector machines. In this approach input vector is first mapped into the higher dimension feature space and there from SVM obtains the hyper-plans which are optimally separated in the higher dimensional feature space. The versatility and power of SVM comes from the fact that the decision boundary (separating hyper plan) is determine by support vector instead of training sample making it extremely regimental to outliers. Basically SVM classifier works for binary classification i.e. separation of a set of training vectors belonging to different classes. Importantly in this approach support vector are same as training sample which are close to separating hyper plan. It uses a penalty factor a user defined parameter to improve the classification. Trade off between the separating hyper plan and number of misclassified sample is also allowed to be adjusted by the users. SVM has also been used for unsupervised learning in addition to clustering by Arnold et al. [10]. Researchers have reported that performance achieved by SVM to be better than other clustering methods. For identifying the malicious activities in KDD cup dataset, Mukkamala, et al. [11] used five SVMs in their approach, one to detect a normal traffic and four others to identify malicious activities. Using this approach every SVM could achieve 99% accuracy where as in the similar scenario using is much longer training time a neural network could achieve only 87.07% accuracy. In general SVMs are superior to neural nets in both accuracy and speed.

1.3 Artificial Neural Networks (ANN): Based on a sequence of commands given by a specific user, a system using neural network approach learns to predict next command thus neural networks solve the problem of modeling user's behavior in a continuous process which is used in anomaly detection because no explicit use model is needed. In expert system modeling researchers used neural networks as an alternative to statistical technique for intrusion detection [12]. It has been reported in the literature that the performance of both a basic signature matching system and a well trained neural network were similar Ghoes et al.[13].

Various neural networks like - multi layered perceptrons, radial basis function-based, Hopfield networks, etc. can be used for anomaly based IDS. A neural network based IDS basically has three phases:

- For a given period collect training data from audit log for each user. Construct a vector showing the frequency for command execution by user for each day.
- The user is to be identified by training the neural networks using command distribution vectors.
- Based on command distribution vector use neural network to check the user. An anomaly is detected in case the neural networks indicate a deferent user other than the actual user [12].

1.4 Fuzzy Approach (FA): Researchers are currently using fuzzy logic with data mining technique to build behavioral profile from audit data. Algorithm based on fuzzy sequentially rule was enhanced. Researchers using fuzzy association rules mined from new data were matched with the rules mined during the training phase for detecting anomaly in behavior. Florez et al.[14] proposed an algorithm for matching the data samples using fuzzy inference engine with fuzzy association rules. They also proposed an algorithm for calculating level of similarity between two different fuzzy association rules and used prefix trees to achieve improvement in computing time and detection accuracy. Different data mining techniques used fuzzy logic for anomaly detection by identifying outliers Yingbing et al. [15].Fuzzy C-Medoids algorithms and fuzzy C-Means algorithms identify outliers using common clustering technique. However these algorithms show performance degradation when applied to datasets with high dimensionality (dimensionality curse). Feature selections are therefore necessary step in data processing. For example, principal component analysis and rough sets are data preprocessing before clustering [16]. As a means of modeling the uncertainty of natural language Fuzzy logic constructs more abstract and flexible patterns for identifying intrusion. This results in better adaptability and robustness of intrusion detection system. At present researcher are active and applying fuzzy logic using two different approaches.

- Automatic design of fuzzy rules with learning and adaptation capabilities.
- Better understandability and through it simplification of some machine learning (SVMs/HMMs) algorithm.

Fuzzy logic helps in smoothing the abruptness in separation between normality and abnormality. The level of intensity of research work currently being undertaking clearly indicates that for some years to come fuzzy logic will play vital role in intrusion detection techniques.

1.5 Evolutionary Computation (EC): From the observation of evolution in nature, researchers have developed a creative process called Evolutionary computation for addressing complex real world problems. Some of these problems have complex non linear dynamics, randomness, and multimodal behavior which individually and collectively make these problems extremely difficult to solve using traditional algorithms [1]. Evolutionary computation uses iterative progress, such as growth or development in a population. This population is then selected in a guided random search using parallel processing to achieve the desired end. Such processes are often inspired by biological mechanisms of evolution. As evolution can produce highly optimized processes and networks, it has many applications in computer science including intrusion detection. In the field of intrusion detection researchers have used EC for automatic model design, optimization and even learning for classification. Simulation studies have amply demonstrated that EC is not only effective but accurate as well. However there are certain challenges for EC as listed. Performance of intrusion detection based on EC will show improvement if these problems are tackled.

1.6 Artificial Immune System (AIS): Artificial immune systems are similar to mimicking the principals of Human Immune System (HIS). The characteristics of an immune system, like uniqueness, distribution, pathogen recognition, imperfect detection, have been modeled to reinforce learning and memory capacity to compensate the weaknesses in traditional intrusion detection techniques. Use of AIS has made the intrusion detection self organized, autonomous, dynamic and distributive. While mimicking the hierarchical structure of HIS researchers have their own prospective in time of starting the modeling. It can be easily noted that other IDS algorithms have been investigated relatively more thoroughly and used widely compare to algorithms based on AIS approaches. Design of AISs has been benefited by emergence of danger theory. Now researchers are taking more interest in development of systems applying ideas inspired from AIS. Efforts have started a fresh to properly interpret and understand the terminology and immunological metaphor before reorganizing the research work or development of intrusion detection system based on AIS

[11].Though AIS has great potential and has shown many successes there are certain open questions which stile remain to be answered table1.

1.7 Swarm Intelligence (SI): It is a sub field of artificial intelligence (AI) that studies the collective behavior in decentralized systems [19]. In essence swarm intelligence simplifies design solutions of complex problem by computationally mimicking the social behavior of swarms. A large number of simple interactions arising out of complex system and patterns are collectively termed as emergent behavior. The basic idea of this approach is to solve complicated problem using large number of agents with simple functions without having to use centralized control or a global model. Global pattern or behavior that emerges as the results are local interaction of the agents with the environment. Thus two important features of SI are emergent strategy and highly distributed control. These two features make a system autonomous, flexible, scalable, adoptive, cost effective and self organizing. SI considers an individual in a population as potential solutions. Optimum solutions are obtained through iterative steps by these individuals working collaboratively. Unlike crossover or mutation operators of evolutionary computation in swarm intelligence, individuals and the search space communicate directly and indirectly to change their positions to reach optimal solutions. Particle swarm Optimization (PSO) and Ant Colony Optimization (ACO) are popular techniques inspired by swarm intelligence. Like social system of a flock of birds ACO is useful in nonlinear optimization problems with constant where as ACO mimics the behavior of ants and is useful in discreet optimization problem. The power of intelligence demonstrated by swarm through simple local interactions of local agents is useful for anomaly intrusion detection which requires high dimensional datasets for real time detection and constantly changing user behavior. Swarm intelligence works in the principal of divide and conquer by breaking a complex problem into several simple ones and assigning an agent to each one of them to work in parallel. These properties makes swarm intelligence based IDS cost-efficient, autonomous, adoptive and computationally fast.

1.8 Soft Computing (SC): It is an innovative approach to build a computationally intelligent system, analogous to the reasoning of human mind and ability to learn from environment under imprecision and uncertainty [17]. Basically soft computing encompasses various computational intelligence methodologies which may include Fuzzy logic, Artificial neural network, probabilistic computing, evolutionary computing, and also artificial immune system, belief networks etc. Each one of these methods are not independent and also does not compete with one another. Their working is cooperative complementary to each other's. These methods however can be loosely or tightly coupled with others. In the paradigm of soft computing hybrid systems are tightly coupled systems. Examples of tightly coupled system are - genetic-neuro systems, neuro-fuzzy systems, genetic-fuzzy-neuro systems, genetic-fuzzy systems and genetic-neuro systems[18]. Whereas ensemble systems are loosely coupled where they have collective approaches but each can be a distinct module.

2. Knowledge Based Intrusion Detection System (K-IDS)

It is useful for signature based IDS an also for detection of anomaly. Basically it gathers the knowledge about system vulnerabilities and specific attacks. This knowledge is used to identify the attacks. If an event is not identified as attack no action is taken. This makes knowledge based intrusion system relatively more accurate (less false alarms). The updating of knowledge of attack needs to be done regularly [20]. K-IDS classified into following categories

*2.1 Expert System (ES):*Intrusion detection techniques which are primarily knowledge based use Expert systems. In this contest set of rules described in text are contained in expert system. Events to be audited are first interpreted in to facts carrying semantic meaningful to the expert system, then using these rules and facts inference engine gives the results. This process used in an expert system enhances the abstraction label by attaching semantic to audit data.[21]

2.2 Petri Nets(PN): Researcher of Purdue University had developed a intrusion detection system using a knowledge base to describe attack profile. Later they enhanced there technique to Colored Petri Nets (CPN) for generality and ability to graphically represent the information. Facility was given to system administration in writing their own tech profiles and integrates the same to IDIOT (Intrusion Detection In Our Time) [7]. The advantage of generality of CPNs in easily describing complex attack patterns has the cost that it is computationally very expensive to match it with the audit trail. [21].

2.3 Signature Analysis (SA): Similar to expert system Signature analysis also uses knowledge -acquisition approach but in a different way. In this approach first attack description is converted in to semantic description which is again transformed in to the information similar to audit trail. It has the advantage that it needs much less level of semantic description of attack making it efficient to implement. Therefore it is mostly used in commercially available intrusion tools e.g. Haystack [21].

2.4 State Transition Analysis (STA): The concept of State transition analysis proposed by Porras and Kemmerer [22] is similar to model based reasoning. In this approach attacks are described using set of goal and transitions to construct a state transition diagram which graphically depicts a sequence of actions of an attacker. This makes it easy to identify and trace how an attack has taken place its requirement and extent of penetration. From this key action taken by attacker can be listed [21].

3. Data mining based approach (DM-IDS)

Only known attacks are identified by IDS. Even simple insiders attacks are also not detected by plain IDS. In fact solution for making IDS more effective is by making data mining as its core, which would accept the previously noticed pattern. One additional advantage of incorporating data mining technique in IDS is that it would reduce the data size needed for compression of historical network activities. This enables the creation of more meaningful data to detect anomaly [23]. Researchers have classified DM- into fallowing categories.

3.1 Clustering: Unlabeled data with many attributes (dimension) can be clustered for finding a pattern. Researchers have mostly used K-Mean clustering for sub grouping of similar data instances. If an observation shows unusual activities outside of these clusters can be part of a new attack. In clustering, approaches normally used are destiny-based methods, grid-based methods, model-based methods, patrioning methods and hierarchy methods [24].

3.2 Association rule discovery: Although Association rule mining is a commonly used technique but it is computationally very slow. This is the reason that this technique is being replaced by more efficient technique such as clustering and classification. Lee, et al. in [25] initially introduced this technique and later extended it [25] [27]. This technique was initially used for market basket analysis and later for intrusion detection.

3.3 Classification: For Intrusion detection data mining technique can be used to solve a classification problem e.g. a class of normal instances or a particular kind of instances (Intrusion/attack). The main goal of classification is to learn from differently labeled classes of training data for identifying classes of new or previously unknown instances. Data and the training set must have well defined labels. Newly acquired data then can be classified on the basis of information gained from the training dataset. This basically forms a classification tree which has been contracted to classify (predict the category to which a particular dataset belongs) [25]. Top down and bottom up approaches are used to build classification tree. Two commonly used algorithms ID3 and C4.5 construct their decision tree in top down manner.

4. Statistical anomaly based intrusion detection system (SA-IDS):
The very early methods used SA-IDS for intrusion detection in information systems. Statistical properties and other statistical tests are used by it in anomaly detection by checking whether observed behavior is significantly different from expected behavior [29]. Statistical properties are used by SA-IDS to construct a normal profile based on normal activities in an information system. This profile and other statistical tests are applied to current usage pattern for checking any deviation. SA-IDS is a two step process- first for the normal and current activities a behavior profile is established. Later matching is done with established behavior profiles using different techniques and deviations are identified as intrusions. SA-IDS can be further classified into following categories:

4.1 Markov process model or marker model: The normal behavior of a particular event is determined by the Markovian model by using observed characteristics of immediately preceding events. In cases where a sequence of activities is especially important this technique is quite useful [29]. The system examines at fixed interval to keep the track of intrusion. State changes are checked by Markov chain to compute the probability of an event in a specified time interval. Low probability of event is taken as an anomaly [30]. for analyzing a model of sequence of information a Hidden Markove Model(HMM) is useful tool. In intrusion detection HMM is useful in maximizing detection rate and minimizing false positive error. This gives good performance at the cost of significant time required to model normal behaviors and determines intrusions. It is therefore not much suitable to real-time intrusions detection.

4.2 Mean and standard deviation model: In order to determine the normal behavior of an observation, it uses standard statistical methods which calculate the observed position relative to a specified confidence threshold range [29].

4.3 Multivariate model: It is based on correlation amongst features. It is effective in cases when multiple features are related to each other [29]. Based on the complexity of the situation it compares multiple parameters to identify potential anomalies. Several researchers have proposed different methods based on multi-vitiate model, e.g., chi-square statistic, Hotelling's test [30]. A profile of normal events in an information system is built in X2 test. For a normal profile the departure of events in the recent past from the

normal profile is computed. If a large departure is detected it is taken as an anomaly or a likely intrusion. Where as a multi-variate statistical process T2 test uses a control technique to analyze audit trail activities and there from identify host based intrusion. Both X2 test and T2 test statistically calculate the distance of an observation from the multi-variate mean vector of a data clusters. The chi square uses a distance and covariance matrix is used by X2 test and T2 test statistics respectively. The performance of X2 test which detects only mean shifts is better than Hotelling's T2 test for detecting Intrusions.

It is relatively more convenient to detect intrusion by mean shifts than by counter relationships. As such similar to multivariate cumulative sum (MCUSUM), multivariate process control techniques and multivariate exponentially weighted moving average (MEWMA) are normally used in monitoring as well as detection of anomalies in processes [29].

4.4 Operational model or threshold metric: In this model it is basically assumed that in an operation an anomaly is identifiable by comparing predefined limit with an observation. Cardinality of series of observations is collected over a time period. This is compared with a threshold value to raise an alarm. This type of model is mostly applicable to observation metrics and the past knowledge about certain values. Any deviations are normally called intrusions [29]. An adaptive threshold algorithm [30], over a predefined time interval checks traffic measurement against a set of particular thresholds whose values are set based on estimates computed from recent traffic observations.

4.5 Time series model: In time series model significant deviations of normal pattern of the data sequence are taken as anomalies. This model identifies anomalies by scrutinizing the particular order of activities in a specified time interval. If the probability of happening of an activity is low, then the event is termed as abnormal. Based on the activities of users this model has the inherent ability to evolve and become more accurate [29].

IV: COMPARISON OF VARIOUS INTRUSION DETECTION TECHNIQUES:

Comparisons are shown in table2.

Table 2: Comparison of various intrusion detection techniques

Approach	Strengths	Weaknesses
1. SA-IDS	1. Previous knowledge is not required. 2. accurate early warning is generated for long term attacks. 3. DoS attack are identified correctly. 4. Frequent Signature updates are not required. 5. Detects slow and low attacks. 6. Detects unusual activity. 7. It is capable of detecting the attack from part observation.	1. Correct statistical distribution is needed by statistical method give accurate result. However purely statistical method are not able to properly model all behaviors. 2. The basic assumption of quasi-stationary process is not fully met in real world intrusion detection system. 3. Even on relatively consistent network, SA-IDS learning process takes very long time to achieve accuracy and effectiveness. 4. It is a tricky problem to set threshold to proper value. 5. Even for legitimate changes in user behavior unacceptably high numbers of false alarms are generated.
2. DM-IDS	1. Researchers can focus on real intrusions because alarm data does not include data from normal activity. 2. "bad" sensor signatures and false alarm generators are identified. 3. Process narrows down to such anomalous activity that lead to a real intrusion. 4. Specially activities which continue for a long time are identified. (same activity ,different IP address)	1. Due to unpredictable changes in behavior patterns of users a large number of false alarms are produced. 2. Even for characterizing normal behavior of users, a large size of "training datasets" of system logs is often required.
3. K-IDS	1. The accuracy of the results produced by this technique is good. 2. False alarm rates is low. 3. It is flexible, scalable and robust. 4. Security officer can easily take preventive or corrective measures.	1. In order to keep this technique effective it is necessary that attack data is updated regularly. 2. Proper maintenance of knowledgebase is difficult because it requires proper analysis of each vulnerability. 3. Generalization is complex and time consuming task.
4. ML-IDS	1. Newly acquired information helps in Improving the performance by making appropriate execution strategy.	1. Resource expensive nature.

V. MULTI CLASSIFIER APPROACHES (MCA)

Using one or more classifiers, researchers are creating an adaptive and scalable intrusion detection scheme (ASIDS) with better accuracy for real time environment. In large databases the crucial problem is identifying the patterns and classifying then. Due to heavy processing requirement for inherent class ambiguity most of the schemes are not suitable for real time operation. Researchers are exploring the benefits of combining multiple classifiers for the intrusion detection system. The advantages of coupling these distinct techniques are well evident, in particular, in the case of exclusive classification requiring a low classification error, high performance rate. The following two approaches are used to combine one or more classifier.

*1. Ensemble Approach:*In this approach multiple hypotheses collectively form a better hypothesis. This makes ensemble approach capable to predict batter. Ensemble can be trained and as such it is a kind of supervised learning algorithm which combines many week learner in to a strong learner. In a way multiple classifier systems like ensemble approach can be differentiating in two parts according to adapting and non adapting approaches [31]. Adaptive approaches are better the non adaptive approaches. Some approaches have been divided as being adaptive and non adaptive approaches in table 3 [32].

Srilatha Chebrolua et. Al. [33] has used BN to build automatic intrusion detection system based on signature recognition. A major difficulty of signatures based intrusion detection is that signatures change over the time in the system must be retrained. This means IDS must be able to adapt to these changes. Authors have proposed a robust framework for an adaptive intrusion detection system

Mukamala et.al. [34] have proposed an adaptive intrusion detection system based on splines techniques. In addition to Multivariate Adaptive Regression Splines (MARS) authors have also explored neural networks and support vector machines. Procedure developed by the authors, fits separate splines of the predictor variable in to flexible regression models in distinct intervals.

Using Data mining technique Wanke Lee at al[35] have described a framework to compute inductively learnt classifiers based on set of relevant system features. Meta Learning has been used to build intrusion detection model which are effective and adaptive. Two techniques namely association rules and frequent episodes computed from information system have been clubbed together by the authors for data gathering and feature selection process. These algorithms have been further modified to use axis and reference attribute for computing only the relevant patterns. Additionally to identify important low frequency pattern authors have used a procedure which is approximate an iterative.

Table 3: Adaptive and Non-adaptive Ensemble approaches

Scheme	**Architecture**	**Trainable**	**Info-level**	**Comments**
Adaptive Scheme for ensemble classifiers				
Mixture of local experts(MLE)	Gated parallel	Yes	Confidence	Explores local expertise; joint optimization
Hierarchical MLE	Gated parallel hierarchical	Yes	Confidence	Same as MLE; hierarchical
Associative switch	Parallel	Yes	Abstract	Same as MLE, but no joint optimization
Voting	Parallel	No	Abstract	Assumes independent classifiers
Non-adaptive Scheme for ensemble classifiers				
Sum, mean, median	Parallel	No	Confidence	Robust assumes independent confidence estimators
Adaptive weighting	Parallel	Yes	Confidence	Good utilization of training data
Stacking	Parallel	Yes	Confidence	Good utilization of training data
Borda count	Parallel	Yes	Rank	Converts ranks into confidences
Logistic regression	Parallel	Yes	Rank confidence	Converts ranks into confidences
Class set reduction	Parallel cascading	Yes	Rank confidence	Efficient
Dempster-Shafer	Parallel	Yes	Rank confidence	Fuses non-probabilistic confidences
Fuzzy integrals	Parallel	Yes	Confidence	Fuses non-probabilistic confidences

Bagging	Parallel	Yes	Confidence	Needs many comparable classifiers
Boosting	Parallel hierarchical	Yes	Abstract	Improves margins; unlikely to over train, sensitive to mislabels; needs many comparable classifiers
Random subspace	Parallel	Yes	Confidence	Needs many comparable classifiers
Neural trees	Hierarchical	Yes	Confidence	Handles large numbers of classes

Dewan Md. Farid et at [36] have used a series of classifiers where each classifier individually contribute for classification of intrusions in to misuse or anomaly. There algorithm for each of probability set is generated using boosting and naive Bayesian classifier simultaneously updating the weights. The criteria for weight updating are done on the basis of above rate during training. Results of simulation using different network types, indicate better performance in terms of rate of intrusion detection and number of false positive.

2. Hybrid Approach: The conjunction of two or more classifiers is known as a hybrid technique. In hybrid approach, most intelligent classifiers require optimization of its parameters for proper classification. A weighted average combiner requires the determination of combiner weight. Such optimization challenges provided ideal for the use of evolutionary computing paradigms. Paradigms such as Genetic algorithm, Particle Swarm Optimization and k-nearest neighbor technique have successfully been used for such implementation.

Rangadurai et al [37] described a two stage architecture which uses a probabilistic classifier in the first stage to detect potential anomalies in the network traffic. In the second stage is a HMM based traffic model used for narrowing the potential attack on IP addresses. Authors have empirically shown that their model is capable of achieving good performance.

A hybrid approach based on a pattern matching engine and a neural network functioning in parallel is proposed by Amza, C et al [38] to improve the detection efficiency. Author's approach is based on a tool "Netpy", for network traffic monitoring and analysis tool which was developed by them earlier. It is claimed that the approach proposed by the authors is not only efficient but also superior to any of the two individual methods for identifying intrusion.

VI. Proposed Methodology

A new intrusion model is being developed, where exiting algorithms will be suitably modified to make entire intrusion detection system adaptive and scalable.

The architecture of the model will support two data bases one storing known intrusion pattern and other storing regular usage pattern. The model will have the following characteristics:

1. Model will be adaptive in the sense that it will monitor the load on the server and frequency of intrusion so as to deploy one or multiple agents for pattern(s) matching.
2. Usage pattern will be first matched with the known intrusion pattern base and there after regular usage pattern database. This will make the detection faster.
3. Any new pattern will be temporarily stored and later on check it for being normal or intrusive and accordingly pattern databases will be updated after every fixed time window.

- Known algorithms will be suitably modified to for integration in the model. As an intermediary process systematic data mining approaches will be used to select the relevant system features to build better detection models.
- We propose to use (meta) learning agent-based architecture to combine multiple models, and to continuously update the detection models.

VII. Conclusion

In the field of cyber security, intrusion detection system has become an activity of paramount importance, because it can detect and take timely action on an unauthorized access before any damage is done. In worst scenarios as a forensic measure also it is extremely useful for tracing and back auditing and also providing information to be used in new development of intrusion detection system so as to avoid the same or similar fallacies in future.

In this paper all aspect of intrusion detection has been covered. Various methods, framework and approaches used by researcher have been systematically grouped and strength and weakness have been identified. Taxonomy of broader classification of various intrusion detection methods has been proposed.

From these papers it becomes apparent that in addition to using very basic concepts, researches have also exploited old tools used in different application area to develop new tools for intrusion detection tools. Most of these tools have been tested using standard test data bases and some on real life test data. Data mining tools and technique have proved to be rich source of concepts but to make intrusion detection more versatile, other concepts from machine learning and fuzzy logic have also been used. For taking care of vast size of data and also to identify newer type of intrusion various concepts from areas of artificial intelligence, statistical methods and evolutionary computing have been incorporated in intrusion detection system.

Research on knowledge based intrusion detection is on the rise. But as there are varied types of ways and means of new attack, simple intrusion detection system may not be that effective. For combining the strength of individual intrusion detection and eliminating their weaknesses, researchers have started using data preprocessing and technique like ensemble and hybrid approaches.

Some more work needs to be done to intelligently adjust the weight in machine learning as well as in modifying the implemented result of individual technique in ensemble approach when final inferences are drawn. Whatever tools are used in intrusion detection, always results shall not be 100% correct. Therefore it is also necessary to incorporate cost factor in intrusion detection to penalize the wrong result and/or reward the correct ones.

At the end, a methodology have been very briefly discussed that would assimilate various aspects of intrusion detection. Normally clubbing multiple approaches consumes more resources thereby slowing down the system especially in real time intrusion detection. To avoid these in the proposed methodology a technique for load balancing is being incorporated so that the proposed intrusion detection tool does the self-adjustment and overloading of the system under scan is avoided.

.

REFERENCES

1. Herv´e Debar"An “introduction to intrusion-detection systems." *Proceedings of Connect* 2002 (2000).

2. Patel, Reema, Amit Thakkar, and Amit Ganatra. "A Survey and Comparative Analysis of Data Mining Techniques for Network Intrusion Detection Systems."*International Journal of Soft Computing and Engineering (IJSCE)* ISSN (2012): 2231-2307.

3. Wu, Shelly Xiaonan, and Wolfgang Banzhaf. "The use of computational intelligence in intrusion detection systems: A review." *Applied Soft Computing10*, no. 1 (2010): pp1-35.

4. Dorothy E. Denning, “An Intrusion-Detection Model”, *IEEE Transactions on Software Engineering*, Vol. SE-13, No. 2, pp. 222-232, Feb 1987.

5. Koch, Robert. "Towards next-generation intrusion detection." In *Cyber Conflict (ICCC), 2011 3rd International Conference on*, pp. 1-18. IEEE, 2011.

6. Wang, Jing-Hong, Yu-Feng Dong, and Hai-Feng Liu. "On intrusion detection matching algorithm from the perspective of multi-agent." In *Machine Learning and Cybernetics (ICMLC), 2012 International Conference on*, vol. 2, pp. 601-606. IEEE, 2012.

7. Gyanchandani Manasi, J. L. Rana, and R. N. Yadav. "Taxonomy of Anomaly Based Intrusion Detection System: A Review." *International Journal of Scientific and Research Publications, Volume 2, Issue 12, December 2012.*

8. Wang, Jing-Hong, Yu-Feng Dong, and Hai-Feng Liu. "On intrusion detection matching algorithm from the perspective of multi-agent." In *Machine Learning and Cybernetics (ICMLC), 2012*

9. Vapnik. Statistical Learning Theory. Springer, N.Y., 1998

10. A Arnold, Eskin E, M Preraua, L Portnoy, and S. J. Stolfo, “A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data”, In D. Barbar and S. Jajodia (Eds.), Data Mining for Security Applications, Boston: Kluwer Academic Publishers, May 2002.

11. Mukkamala, S., Janoski, G., & Sung, A. (2002). Intrusion detection using neural networks and support vector machines. In: Proceedings of the *IEEE international joint conference on neural networks* (pp. 1702–1707).

12. Zhao, Jinguo, Min Chen, and Qinyun Luo. "Research of intrusion detection system based on neural networks." In *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, pp. 174-178. IEEE, 2011.

13. Ghosh A, K. A Schwartzbard, and M Schatz, “Learning program behavior profiles for intrusion detection”, In Proceeding of 1st USENIX, 9-12 April, 1999.

14. Florez G., Bridges S. M., Vaughn R. B., "An Improved Algorithm for Fuzzy Data Mining for Intrusion Detection". *Fuzzy Information Processing Society, Proceedings. NAFIPS, 2002 Annual Meeting of the North American,* IEEE. (2002).

15. Yu, Yingbing, and Han Wu. "Anomaly intrusion detection based upon data mining techniques and fuzzy logic." In *Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on*, pp. 514-517. IEEE, 2012.

16. Liangxun, Shuo, Zhao Jinhui, and Wang Xuehui. "An Adaptive Invasion Detection Based on the Variable Fuzzy Set." In *Network Computing and Information Security (NCIS), 2011 International Conference on*, vol. 2, pp. 115-118. IEEE, 2011.

17. Sampada Chavan, Khusbu Shah, Neha Dave and Sanghamitra Mukherjee " Adaptive Neuro-Fuzzy Intrusion Detection Systems" *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04)* IEEE 2004

18. Gaffer, Salma Mahgoub, Moawia Elfaki Yahia, and Khaled Ragab. "Genetic fuzzy system for intrusion detection: Analysis of improving of multiclass classification accuracy using KDDCup-99 imbalance dataset." In *Hybrid Intelligent Systems (HIS), 2012 12th International Conference on*, pp. 318-323. IEEE, 2012.

19. Xie, Li-Ming, and Jing-Li Gao. "A concept lattice-based adaptive intrusion detection algorithm." In *Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), 2011 2nd International Conference on*, pp. 2699-2702. IEEE, 2011.

20. Sumit More, Mary Matthews, Anupam Joshi, Tim Finin" A Knowledge-Based Approach to Intrusion Detection Modeling" *IEEE Symposium on Security and Privacy Workshops* 2012 PP. 75-81.

21. Herve Debar, Marc Dacier, Andreas Wespi, "Towards a taxonomy of intrusion-detection systems", *Elsevier, Computer Networks*, Vol. 31, pp. 805–822, 1999.

22. P.A. Porras and R.A. Kemmerer, "Penetration State Transition Analysis: A Rule-Based Intrusion Detection Approach", In *Proceedings of the Eighth Annual Computer Security Applications Conference*, San Antonio, TX, December, 1992.

23. Shetty, Monali, and N. Shekokar. "Data Mining Techniques for Real Time Intrusion Detection Systems." *International Journal of Scientific & Engineering Research* volume 3, issue 4, April (2012)pp764-770.

24. Idowu, S. A., and Ajayi Adebowale. "Characterization and Trends of Development in Data Mining Techniques For Intrusion Detection Systems (IDS)." *The International Journal Of Engineering And Science (IJES) Volume 2 Issue 6* Pages 19-25, 2013 ISSN(e): 2319 – 1813 ISSN(p): 2319 – 1805.

25. Amudha, P., and H. Abdul Rauf. "Performance Analysis of Data Mining Approaches in Intrusion Detection." In *Process Automation, Control and Computing (PACC), 2011 International Conference on*, pp. 1-6. IEEE, 2011.

26. K. Ilgun, R. Kemmerer, P. A. Porras, "State Transition Analysis: A Rule-Based Intrusion Detection Approach", *IEEE Transaction on Software Engineering*, 21(3):pp. 181-199. 1995.

27. Moorthy, M., and S. Sathiyabama. "A study of Intrusion Detection using data mining." In *Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on*, pp. 8-15. IEEE, 2012.

28. Davide Ariua, Roberto Troncia, Giorgio Giacinto "HMMPayl: an Intrusion Detection System based on Hidden Markov Models" *Elsevier Computers & Security*, February 2011

29. V. Jyothsna, V. V. Rama Prasad, K. Munivara Prasad "A Review of Anomaly based Intrusion Detection Systems" *International Journal of Computer Applications* Volume 28– No.7, August 2011 pp 28-34

30. Bhuyan, M.H. ; Bhattacharyya, D.K. ; Kalita, J.K. " Network Anomaly Detection : Methods, Systems and Tools" *Communications Surveys & Tutorials*, IEEE Volume:16, Issue: 1, 2014, pp 303 – 336.

31. Chebrolu, Srilatha, Ajith Abraham, and Johnson P. Thomas. "Feature deduction and ensemble design of intrusion detection systems." *Computers & Security*24, no. 4 (2005): 295-307.

32. Martin Sewell" Ensemble learning" UCL department of computer science Jan 2011.

33. Srilatha Chebrolua, Ajith Abrahama,b, Johnson P. Thomasa "Feature deduction and ensemble design of intrusion detection systems" *Computers & Security Elsevier Ltd*. (2005) 24, pp 295-307.

34. Srinivas Mukkamala, Andrew H. Sung, Ajith Abraham, Vitorino Ramos," Intrusion Detection Systems Using Adaptive Regression Spines" *Enterprise Information Systems VI Springer Netherlands* 2006 pp 211-218.

35. Wenke Lee and Salvatore J. Stolfo "Data Mining Approaches for Intrusion Detection" *Proceedings of the 7th USENIX Security Symposium San Antonio, Texas*, Jan 26-29, 1998.

36. Dewan Md. Farid, Mohammad Zahidur Rahman, Chowdhury Mofizur Rahman" Adaptive Intrusion Detection based on Boosting and Naïve Bayesian Classifier" *International Journal of Computer Applications* Volume 24,No.3, June 2011 pp12-19.

37. R Rangadurai Karthick, Vipul P. Hattiwale, Balaraman Ravindran" Adaptive Network Intrusion Detection System using a Hybrid Approach" *Communication Systems and Networks (COMSNETS)*, 2012 Fourth International Conference IEEE jan2012 pp 1-7.

38. C. Amza, C.Leordeanu, V. Cristea, "Hybrid network Intrusion Detection ", *IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2011, pp 503 – 510.