*I.J. Information Technology and Computer Science,* **2013, 08, 75-90**
Published Online July 2013 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijitcs.2013.08.08

# Data Mining in Intrusion Detection: A Comparative Study of Methods, Types and Data Sets

**Chandrashekhar Azad**
Birla Institute of Technology Mesra, Ranchi, India
*E-mail: csazad@bitmesra.ac.in*

**Vijay Kumar Jha**
Birla Institute of Technology Mesra, Ranchi, India
*E-mail: vkjha@bitmesra.ac.in*

***Abstract***— In the era of information and communication technology, Security is an important issue. A lot of effort and finance are being invested in this sector. Intrusion detection is one of the most prominent fields in this area. Data mining in network intrusion detection can automate the network intrusion detection field with a greater efficiency. This paper presents a literature survey on intrusion detection system. The research papers taken in this literature survey are published from 2000 to 2012. We can see that almost 67 % of the research papers are focused on anomaly detection, 23 % on both anomaly and misuse detection and 10 % on misuse detection. In this literature survey statistics shows that 42 % KDD cup dataset, 20 % DARPA dataset and 38 % other datasets are used by the different researchers for testing the effectiveness of their proposed method for misuse detection, anomaly detection or both.

***Index Terms***— Anomaly Detection, Intrusion Detection, Misuse Detection, Data Mining

## I Introduction

The area of data mining is gaining importance due to availability of large volume of data, easily collected and stored in electronic format. The large amount of data gathered from the different sources may contain personal and sensitive data. During the process of data mining an important question is arises, does the mining of data violate the privacy and security of individuals or organizations?

From past decade data mining is paying more attention because of its enormous area of applications. For the security point of view, data mining may be helpful in confronting various types of security attacks to the universe. Data mining technology focuses on discovery of general or statistically significant patterns. In this sense, we believe that the real privacy concerns with unconstrained access to individual records, especially access to privacy-sensitive information such as credit card transaction records, health-care records, personal financial records, biological traits, criminal/justice investigations, and ethnicity. For the data mining applications that do involve personal data, in many cases, simple methods such as removing sensitive IDs from data may protect the privacy of most individuals. Nevertheless, privacy concerns exist wherever personally identifiable information is collected and stored in digital form, and data mining programs are able to access such data, even during data preparation. Improper or nonexistent disclosure control can be the root cause of privacy issues. With the rapid development of World Wide Web and computers during the past decade, security has become a crucial issue for computer systems. A secure network should provide data confidentiality, data integrity, and data availability. Intrusion is an action that tries to destroy data confidentiality, data integrality, and data availability of network information. Data mining may be used to detect and possibly prevent security attacks including cyber security, network security, social security, industrial security etc. For example, anomaly detection techniques could be used to detect unusual patterns and behaviors. Link analysis may be used to trace the viruses to the perpetrators. Classification techniques are used to group various cyber-attacks and then use the data mining to detect an attack when it occurs. Prediction techniques are used to determine potential future attacks.

Now the IT is in its full-fledged speed. Due to the development in IT our society has become technology dependent. People satisfying their need and wants with the help www, like to handle marketing, communication, online news, stock prices, email, online banking and online shopping etc. The integrity, availability and confidentiality of all these systems need to be defended against a number of threats. Hackers, terrorists and even foreign governments have the purpose and capability to

 

carry out attacks on communication network. Thus, the field of information security has also become important for safety and security. The rapid growth of electronic data processing and electronic business conducted through the use of the communication networks along with numerous occurrences of international terrorism raises the need for providing secure and safe information security systems through the use of authentication, encryption, firewalls, intrusion detection systems, and other hardware and software solutions.

This paper is organized in five sections. Section 2 gives the overview of the intrusion detection system. Section 3 describes the literature review and comparative study of intrusion detection systems. Section 4 deals with the discussion related to the type of intrusion detection system, methods and datasets used. Finally section five describes conclusion of this paper.

## II Intrusion Detection System

Intrusion is a set of actions that attempt to violate the integrity availability or confidentiality of data on a computing platform. An intrusion detection system is software; hardware or both that detect intrusions in the network [1]. Intrusion detection system can monitor all the network activities for detecting known or unknown attacks. The main objective of IDS is to alarm the system administrator if any suspicious activity happening. Intrusion detection techniques are classified in two categories: anomaly detection and signature detection or misuse detection.

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 often translate to critical and actionable information in several domains. Anomalies are also referred to as outliers, change, deviation, surprise, aberrant, peculiarity, intrusion, etc. In misuse detection, the IDS analyze the gathered information and compare it to large databases of attack signatures. Essentially, the IDS look for a specific attack that has already been documented. Like a virus detection system. Misuse detection software contains the database of attack signatures and is used for network pocket monitoring. A Distributed IDS is one where data is collected and analyzed in multiple hosts, but in case of centralized intrusion detection system data is collected and analyzed in a centralized system. Both distributed intrusion detection system and centralized intrusion detection systems may use host based or network based data collection methods for intrusion detection, or most likely a combination of the both. Intrusion detection systems can response in two ways: Active - takes some action as a reaction to intrusion (such shutting down services, connection, logging user). Passive - generates alarms or notification. Audit information analysis in intrusion detection system can be done in two ways: on the fly processing (real time) and interval based (periodical). Intrusion detection system runs continuously for intrusion detection and gives result in real time is called real-time intrusion detection system. The term real-time does not indicate more than a fact that IDS reacts to an intrusion quick enough. Intrusion detection system runs periodical for intrusion detection, Interval based are also called periodical intrusion detection system.

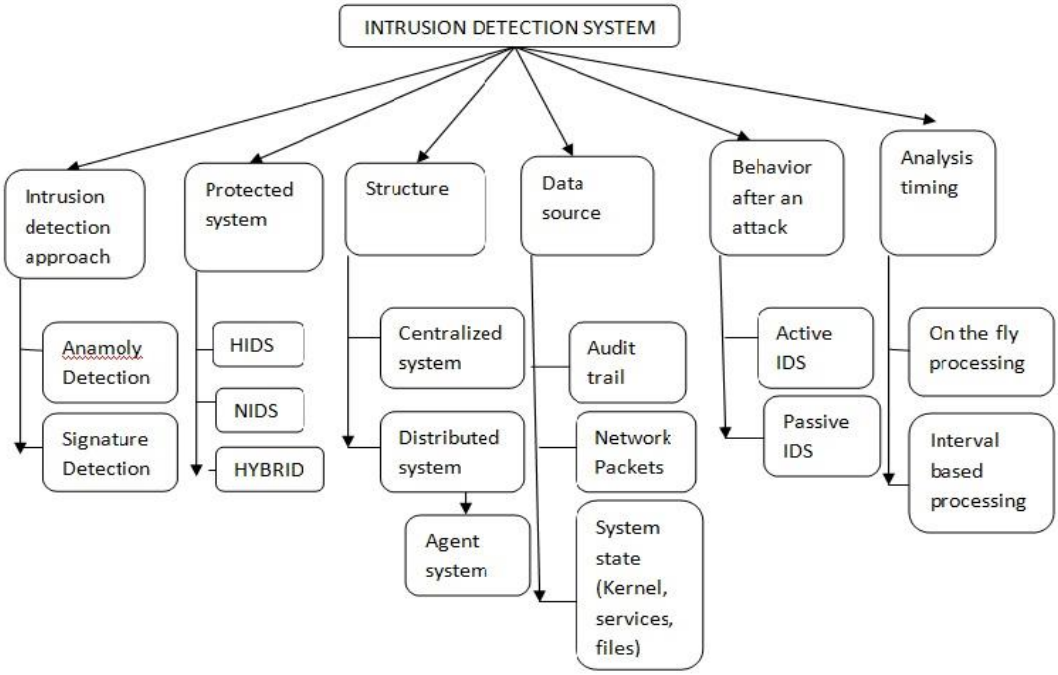


Fig. 1: Classification of intrusion detection systems [2]

## III Related Work

### 3.1 Anomaly Detection:

**Lippmann R P et al. [3]** has proposed an intrusion detection evaluation test bed. In this six research groups participated for evaluation and the results were analyzed for probe, DoS, R2L, and U2R attacks. The performance is evaluated based on the parameters detection rate and false alarm rate. It gave detection rates ranging from 63% to 93% and 10 false alarms per day. **Gomez et al. [4]** has proposed a technique to generate fuzzy classifier using genetic algorithm that can detect anomalies and some specific intrusion, two elements false alarm rate and undetected attack rates define the cost function of intrusion detection system. The average performance of the proposed approach is good as compared to others. **Balajinath et al. [5]** has proposed an algorithm for intrusion detection called genetic algorithm based intrusion detector (GBID) based on learning the individual user behavior. The performance of the proposed system was tested using command history of 77 users. The performance of the proposed approach has evaluated using false alarm rate and accuracy of intrusion. The proposed approach has detected intrusion with accuracy of 96.8 % and a false alarm rate of 3.2 % on real life command history of 25 users. **Ye N [6]** has proposed a clustering and classification algorithm-supervised (CCA-S) for intrusion detection in computer network. CCA-S gave better result than two decision tree algorithms. It gave 100% hit rate at the 0% false alarm rate on the testing data set. **Zhang et al. [7]** has proposed a Hierarchical Intrusion Detection (HIDE) system, which detects network-based attacks as anomalies using statistical preprocessing and neural network classification. Proposed framework gave efficient result when tested on five different neural networks. It can reliably detect UDP flooding attacks with traffic intensity as low as five to ten percent of the background intensity. **Hoang et al. [8]** has proposed a fuzzy based scheme for the integration of HMM anomaly intrusion detection engine and normal-sequence database detection engine for program anomaly intrusion detection using system calls. Experimental design of the proposed method was based on the detection rate and the false positive rate. Experimental results show that the proposed detection scheme reduced false positive alarms by 48% and 28%, compared to the normal-sequence database scheme and the two layer scheme respectively. The proposed detection scheme also generated much stronger anomaly signals compare to the normal-sequence database scheme and the two-layer scheme. The HMM training time was reduced by four times and the requirement of memory was also decreased significantly. These improvements have made a good progress towards online and real-time intrusion detection. **Jha et al. [9]** has presented a rule based intrusion detection system for Linux platform. In this dual approaches for intrusion detection are: pre-emptory and reactionary. **Florez et al. [10]** has proposed an improved algorithm for fuzzy data mining for intrusion detection. The experimental result of the proposed method is better and accurate compare to the others. **Helmer et al. [11]** has proposed an automated discovery of concise predictive rules for intrusion detection. In this proposed method feature vector representation to describe the system calls is executed by privileged processes. The feature vectors are labeled as good or bad depending on whether or not they were executed during an observed attack. **Ye et al. [12]** investigates a multivariate quality control technique to detect intrusions by building a long-term profile of normal activities in information systems (norm profile) and using the norm profile to detect anomalies. Propsed technique is based on Hotelling's T2 test that detects both counter relationship anomalies and mean-shift anomalies. The performance of the Hotelling's T2 test is examined on two sets of computer audit data and a large multiday data set. Both data sets contain sessions of normal and intrusive activities. For the small data set, the Hotelling's T2 test produces no false alarms for the normal sessions and also for the large dataset. **Yeung et al. [13]** has proposed a host based intrusion detection for anomaly detection, they used system call and shell command for performance evaluation. Dynamic modeling approach is better than the static modeling approach for the system call datasets. The dynamic modeling approach is worse for the shell command datasets. **Jun et al. [14]** has proposed a novel intrusion detection technique that integrates the detection methods. The proposed method uses multiple measure and modeling methods. It integrates the results of individual detection methods with rule based approach to overcome the drawbacks of the conventional anomaly detection techniques. The performances of each detection method are compared with the integrated method. The integrated method shows 5.761% false-positive error rate at 100% detection rate whereas each element method produces more than 80% false positive error rate at the same detection rate. **Feng et al. [15]** has proposed a method for predicting intrusion intentions by observing system calls sequences. The performance of this method is effective in predicting the goals of many normal or anomalous system call sequences of UNIX operating systems. This method provides an effective way for analyzing and predicting attack attempts, to issue a short term early warning with direct evidences, and also forms the basis for developing defense and control strategies to respond to coordinated attacks with near real time. **Estevez-Tapiador et al. [16]** has presented a statistical analysis of both normal and hostile traffic and proposed a new anomaly-based approach to detect attacks carried out over HTTP traffic. The detection results provided by the proposed method shows important improvements in detection ratio and false alarms, in comparison with other current techniques. **Wang et al. [17]** has proposed a new intrusion detection method based on Principle Component Analysis (PCA) with low overhead and high efficiency. System call data and command

 

sequences data are used as information sources to validate the proposed method. The performance of the proposed method is promising in terms of detection accuracy, computational expense and implementation for real-time intrusion detection. **Xiang et al. [18]** has proposed a multi-level tree classifier for intrusion detection system. The performance of the proposed method is better as compared to the MADAM ID. **Depren O et al. [19]** has proposed a novel Intrusion Detection System (IDS) architecture for anomaly and misuse detection. Simulation results of both anomaly and misuse detection modules based on the KDD cup data set. Proposed method is evaluated using parameter data rate and false positive rate. Proposed method gave detection rate of 98.96% and a false positive rate of 1.01% for anomaly detection module and also a classification rate of 99.61% and a very low false positive rate of 0.20% are achieved for the misuse detection module. **Jiang S Y et al. [20]** has proposed a powerful clustering-based method for the unsupervised intrusion detection (CBUID). The time complexity of CBUID is linear with the size of dataset and the number of attributes. The experiments demonstrate that proposed method outperforms the existing methods in terms of accuracy and detecting unknown intrusions. **Perdisci R et al. [21]** has proposed a novel on-line alarm clustering system whose main objective is the reduction of the volume of alarms produced by today's IDS sensors. The clustering system has been devised to work in near real time. Experiments performed in different attack scenarios on a live network showed that the proposed algorithm effectively groups alarms related to the same attack, even though IDS produced alarms whose descriptions were erroneously referred to different types of attacks. **Tsang C et al. [22]** has proposed Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. Performance of the proposed system on the KDDCup99 intrusion detection benchmark data provides the highest detection accuracy for intrusion attacks and low false alarm rate for normal network traffic with minimized number of features. **Kayacik H G et al. [23]** has proposed a hierarchical SOM-based intrusion detection system, the performance of the proposed method is evaluated by using detection rate and false positive rate. The resulting false positive and detection rates were 1.38% and 90.4% respectively. **Powers S T et al. [24]** has proposed a hybrid artificial immune system and self Organising Map for network intrusion detection. Proposed method is evaluated on the basis of false positive rate, detection rate and classification rate. The results show that the proposed system is better at detecting and classifying DoS and U2R attacks than the three layer hierarchy of SOMs. **Ramachandran C et al. [25]** has proposed a FORK: A novel two-pronged strategy for an agent-based intrusion detection scheme in ad-hoc networks. Proposed method is effective in terms of the accuracy of rules formed and the simplicity in the content of the rules. **Xiang C et al. [26]** has proposed a multiple-level hybrid classifier, a novel intrusion detection system. The performance of the proposed method is evaluated using detection rate and false positive rate. The Performance of the proposed approach is measured using the KDD cup dataset and it achieves high detection rate and low false alarm rates. **Hoang X D et al. [27]** has proposed a program-based anomaly detection scheme using multiple detection engines and fuzzy inference. The performance of the proposed method was evaluated using HMM training cost, false positive rate, anomaly signals and the detection rate. Experimental results shows that the proposed detection scheme reduced false positive alarms by 48% to 28%, compared to the normal-sequence database scheme and the proposed detection scheme also generated much stronger anomaly signals, compared to the normal-sequence database scheme. **Tsai C et al. [28]** has proposed a hybrid learning model based on the triangle area based nearest neighbors (TANN) in order to detect attacks. The experimental result shows that TANN can effectively detect intrusion attacks. It provide high accuracy, high detection rates and low false alarm rate than three base- line models based on support vector machines, *k*-NN, and the hybrid centroid based classification model by combining k-means and k-NN. **Chen C et al. [29]** has proposed a Lightweight network intrusion detection system (LNID) for intrusion detection. According to the performance comparisons with other network-based IDS, LNID is the most efficient on detection rate and workload reduction. **Wang G et al. [30]** has proposed a new approach for intrusion detection using Artificial Neural Networks and fuzzy clustering. **Mok M S et al. [31]** has proposed Random effects logistic regression model for anomaly detection. Experimental on the proposed method is based on a sample of 49,427 random observations for 42 variables of the KDDCup 1999 data set that contains 'normal' and 'anomaly' connections. The proposed model has a classification accuracy of 98.94% for the training data set, and 98.68% for the validation data set. **Lee S et al. [32]** has proposed a Self-adaptive and dynamic clustering for online anomaly detection. The performance of the proposed approach is evaluated through experiments using the well-known KDD Cup 1999 data set and further experiments using the honeypot data recently collected from Kyoto University. It is shown that the proposed approach significantly increase the detection rate while the false alarm rate remains low. **Casas P et al. [33]** has proposed an unsupervised network intrusion detection system. Proposed method was evaluated on three different traffic datasets, including the well-known KDD dataset as well as real traffic traces from two operational networks. They particularly show the ability of UNIDS to detect unknown attacks, comparing its performance against traditional misuse-detection-based NIDSs. The performance of the proposed method is better with respect to different previously used unsupervised detection techniques. It shows that the algorithms used by UNIDS are highly adapted for parallel computation,

 

which drastically reduces the overall analysis time of the system. **Devarakonda et al. [34]** has proposed an Intrusion Detection System using Bayesian Network and Hidden Markov Model. Performance of the proposed model is high for classification of normal and intrusions attacks. **Kavitha et al. [35]** has proposed an intrusion detection system using Neutrosophic Logic classifier which is an extension/combination of the fuzzy logic, intuitionistic logic, paraconsistent logic, and the three-valued logics that use an indeterminate value. The false alarm rate and the undetected attack rates are the two factors that define the cost function of proposed intrusion detection system. Proposed method gave the best result on KDD Cup data set. **Gong et al. [36]** has proposed an efficient negative selection algorithm with further training for anomaly detection. The experimental comparison among the proposed algorithm, the self-detector classification, and the V-detector on seven artificial and real-world data sets show that the proposed algorithm can get the highest detection rate and the lowest false alarm rate in most cases. **Pastrana S et al. [37]** has presents a comparison of the effectiveness of six different classifiers to detect malicious activities in MANETs. Results show that genetic programming and support vector machines may help in detecting malicious activities in MANETs. **Pereira C R et al. [38]** has proposed an Optimum Path Forest framework for intrusion detection in computer network. The experiments have been carried out on three datasets aiming to compare OPF against Support Vector Machines, Self Organizing Maps and a Bayesian classifier. Results show that the OPF is the fastest classifier and always with better results. Thus, it can be a suitable tool to detect intrusions on computer networks, as well as to allow the algorithm to learn new attacks faster than other techniques. **Sindhu et al. [39]** has proposed a Decision tree based light weight intrusion detection using a wrapper approach for anomalies detection in network. The proposed method has evaluated using detection percentage and error percentage. The proposed method gave better results as compare to Decision Stump, C4.5, Naive Bayes, Random Forest, Random Tree, and REP Tree. **Kang I et al. [40]** has proposed a new one class classification method with differentiated anomalies to enhance intrusion detection performance for harmful attacks. They also proposed new extracted features for host-based intrusion detection based on three viewpoints of system activity such as dimension, structure, and contents. Experiments with simulated dataset and the DARPA 1998 BSM dataset show that proposed differentiated intrusion detection method performs better than existing techniques in detecting specific type of attacks. The proposed method would benefit even other applications in anomaly detection area beyond intrusion detection. **Li Y et al. [41]** has proposed an efficient intrusion detection system based on support vector machines and gradually feature removal method. The accuracy of the proposed IDS achieves 98.6249%, and MCC value achieves 0.861161. The result shows that this IDS is a reliable one, which performs well in accuracy and efficiency. **Koc L et al. [42]** has proposed a network intrusion detection system based on a Hidden Naive Bayes multiclass classifier. Experimental results show that the HNB model exhibits a superior overall performance in terms of accuracy, error rate and misclassification cost compared to the traditional Naïve Bayes model, leading extended Naïve Bayes models and the Knowledge Discovery and Data Mining (KDD) Cup 1999 winner. Proposed model performed better than other models, such as SVM, in predictive accuracy. **Altwaijry H et al. [43]** has proposed a Bayesian based intrusion detection system. The proposed method was able to detect intrusion with a superior detection rate. **Jamdagni A et al. [44]** has proposed a RePIDS: A multi tier Real-time Payload-based Intrusion Detection System. Experimental results of the proposed system has better performance (high F-values, 0.9958 for DARPA 99 dataset and 0.976 for Georgia Institute of Technology attack dataset respectively, with only 0.85% false alarm rate) and lower computational complexity when compared against two state-of-the-art payload-based intrusion detection systems. Additionally, it has 1.3 time higher throughput in comparison with real scenario of medium sized enterprise network. **Chung et al. [45]** has proposed a hybrid network intrusion detection system using simplified swarm optimization (SSO). The testing on the proposed system shows that the proposed hybrid system can achieve higher classification accuracy than others with 93.3% and it can be one of the competitive classifier for the intrusion detection system. **Lin S et al. [46]** has proposed an intelligent algorithm based on feature selection and decision rules for anomaly detection. The performance of the proposed algorithm outperforms the winning entry of the KDD'99 contest and other existing approaches. **Zheng L et al. [47]** has proposed an anomaly detection system based on Filter-ary-Sketch. They demonstrate that the developed system can detect anomalies with high accuracy, low computation and memory costs. It can block the packets that are responsible for anomalies. **Chetan R et al. [48]** has proposed a Data Mining Based Network Intrusion Detection System: A Database Centric Approach. Database centric IDSs offers many advantages over alternative systems. These include integration of individual components, security, scalability, and high availability. **Brauckhoff D et al. [49]** has proposed an Anomaly Extraction in Backbone Networks Using Association Rules. The proposed anomaly extraction approach is generic and can be used with different anomaly detectors that provide meta-data about identified anomalies. It reduces the work-hours needed for the manual Verification of anomaly alarms. **Om H et al. [50]** has proposed a hybrid system for reducing the false alarm rate of anomaly detection system. This system can detect the intrusions and further classify them into four categories: Denial of Service (DoS), U2R (User to Root), R2L (Remote to Local), and probe. The performance of the method was evaluated using

DR, FPR AND ACCURACY parameters. **Sharma M et al. [51]** has proposed a Pre-Clustering algorithm for anomaly detection and clustering that uses variable size buckets. The experimental result about the algorithm shows that the proposed algorithm clusters the data more efficiently and detects the anomalies efficiently. **Sharma S K et al. [52]** has proposed an improved network intrusion detection technique based on k-Means Clustering via Naive Bayes Classification. Proposed technique performs better in terms of detection rate as compared to a naive Bayes based approach.

### 3.2 Misuse Detection

**Barbara et al [53]** has proposed a testbed ADAM (Audit Data Analysis and Mining) for Exploring the Use of Data Mining in Intrusion detection. **Ning et al. [54]** has present a design and implementation of decentralized prototype intrusion detection system, named coordinate attacks response and detection system for detecting distributed attacks. **Guan et al [55]** has proposed a K-means based clustering algorithm, named Y-means, for intrusion detection. The result of the proposed method on KDD-99 data set shows that Y-means is an effective method for partitioning large data space. A detection rate of 89.89% and a false alarm rate of 1.00% are achieved with Y-means. **Abadeh M S et al. [56]** has proposed a parallel genetic local search algorithm for intrusion detection in computer networks. The performance of the proposed method is evaluated on the basis of false positive rate and data rate. **Sangkatsanee P et al. [57]** has proposed a practical real-time intrusion detection using machine learning approaches. The performance of the proposed method is efficient in terms of real-time detection speed and consumption of CPU and memory. It takes only 2s to detect and classify the incoming network data. The classification results can be improved with an optional data post-processing procedure, which can additionally reduce the false alarm rate. **Abadeh M S et al. [58]** has proposed a design and analysis of genetic fuzzy systems for intrusion detection in computer networks. Performance of the proposed method is evaluated using detection rate and false positive rate parameters. An experimental result shows that the proposed method would be more reliable than other approaches. The false alarm rate in GFS based intrusion detection systems is considerably lower than other approaches. **BOULAICHE A et al. [59]** has proposed a quantitative approach for intrusions detection and prevention based on statistical n-gram models. An experimental result shows that the proposed approach is very promising to provide systems security in a fully automatic way without any human intervention. **Mohammed M N et al. [60]** has proposed an Intrusion Detection System Based on SVM for WLAN. Proposed system produced better result in terms of the detection efficiency and false alarm rate, which may give better coverage, and make the detection more effective.

### 3.3 Anomaly and Misuse Detection

**Lee W et al. [61]** has described a framework MADAM ID for mining audit data for automated models for intrusion detection. Proposed framework uses data mining algorithm to compute activity pattern from system audit data and extracts predictive features from patterns, then it uses machine learning algorithm to audit data rules. Proposed system gave best result on DARAPA dataset against all of the participating system. **Liao et al. [62]** has proposed a new algorithm based on the k-Nearest Neighbor classifier method for modeling program behavior in intrusion detection. Proposed method gave better result as compare to other methods using short system call sequence. **Joo et al. [63]** has proposed a neural network model based on asymmetric costs of false positive errors and false negative errors. System performance of the proposed method is based on asymmetric costs of errors. The results of the empirical experiment indicate that the neural network model provides very high performance for the accuracy of intrusion detection. **Qin et al. [64]** has proposed intrusion detection systems (IDS) in a network environment for anomaly detection. Frequent episode rules are generated for anomaly detection. The performance of the proposed method is effective in detecting unknown network attacks embedded in traffic connections often requested in many Internet services such as TELNET, HTTP, FTP, SMTP, Email, authentication, and authorization. **Siraj A et al. [65]** has described the workings of a decision engine for intelligent intrusion detection system that utilizes causal knowledge inference based on Fuzzy Cognitive Maps (FCMs). **Dasgupta et al. [66]** has proposed a novel recursive data mining method for Masquerade Detection and author identification, its result are better as compared to other. **Aydın M A et al. [67]** has proposed a hybrid IDS by combining the two approaches in one system. The hybrid IDS is obtained by combining packet header anomaly detection (PHAD) and network traffic anomaly detection (NETAD) which is anomaly based IDSs with the misuse based IDS Snort. **Chou T et al. [68]** has proposed hybrid classifier system for intrusion detection. The performance of the proposed method is evaluated on the basis of FPR and DR. The experimental results of the proposed method show that hybrid model has a better detection performance with low FPR on normal computer usages and high DR on malicious activities. It also shows that the overall performance of this hybrid architecture is better than that of each individual base feature selecting classifier. **Shanmugam B et al. [69]** has proposed improved intrusion detection system using Fuzzy Logic for Detecting Anomaly and Misuse type of Attacks. Parameters detection rate and false positive rate are used for evaluation. The performance of the proposed method shows that the detection rate is comparatively higher than all other systems. The false positive rate is also low in comparison to the values obtain from other models. **Mabu S et al. [70]** an intrusion detection model based on Fuzzy Class association rule mining

 

using genetic network programming. Experimental results on the proposed method using KDD Cup and DARPA98 databases shows that it provides competitively high detection rates compared with other machine learning techniques. **Soleimani et al. [71]** has proposed a multi layer episode filtering for the multi step attack detection. The experimental result on the proposed method shows that proposed method effectively discover known and unknown attacks with high accuracy. **Lei et al. [72]** has proposed two new clustering algorithms, the improved competitive learning network (ICLN) and the supervised improved competitive learning network (SICLN), for fraud detection and network intrusion detection. The result demonstrates that both the ICLN and the SICLN achieve high performance, and the SICLN outperforms traditional unsupervised clustering algorithms. **Anming Z [73]** has proposed an intrusion detection algorithm based on NFPA. Performance of the proposed method has better results as compared to the other classical algorithm. **Pandaa M et al. [74]** has proposed a hybrid intelligent approach for network intrusion detection. the performance of the proposed system was evaluated on the detection rate and false alarm rate parameter, proposed system gave high detection rate and low false alarm rate. **Mukherjee S et al. [75]** has proposed a feature vitality based reduction method, to identify important reduced input features. Empirical results shows that selected reduced attributes give better performance to design IDS that is efficient and effective for network intrusion detection. **Prasenna P et al. [76]** has proposed network programming and mining classifier for intrusion detection using probability classification. This method can sufficient to evaluate misuse and anomaly detection. Experiments on KDD cup and DARPA dataset.it gave the high detection rate and accuracy as compared to other conventional methods. **Hussein S M et al. [77]** hybrid IDS by integrated signature based (Snort) with anomaly based (Naive Bayes) to enhance system security to detect attacks. Accuracy, detection rate, time to build model and false alarm rate were used as parameters to evaluate performance between hybrid Snort with Naïve Bayes, Snort with J48graft and Snort with Bayes Net. The result shows that it provides better performance using hybrid Snort with Naive Bayes algorithm.

Table 1: Comparison of two types of intrusion detection system based on detection approach and data sets used

| S. No. | Authors | Type | | Data set | | |
|---|---|---|---|---|---|---|
| | | Anomaly Detection | Misuse detection | DARPA | KDD Cup | Others |
| 1 | Lippmann et al.(2000) | Y | | Y | | Y |
| 2 | Gomez et al.(2001) | Y | | | Y | |
| 3 | Balajinath et al.(2001) | Y | | | | Y |
| 4 | Barbara et al.(2001) | | Y | | | Y |
| 5 | Lee W. et al.(2001) | Y | Y | Y | | Y |
| 6 | Ye N (2001) | Y | | | | Y |
| 7 | Zhang et al.(2001) | Y | | | | Y |
| 8 | Liao et al.(2002) | Y | Y | Y | | |
| 9 | Hoang et al.(2002) | Y | | | | Y |
| 10 | Jha et al.(2002) | Y | | | | Y |
| 11 | Ning et al. (2002) | | Y | | | |
| 12 | Florez et al.(2002) | Y | | Y | | |
| 13 | Helmer et al.(2002) | | Y | | | Y |
| 14 | Ye et al. (2002) | Y | | | | Y |
| 15 | Yeung et al.(2003) | Y | | | | Y |
| 16 | Jun et al.(2003) | Y | | | | Y |
| 17 | Joo et al.(2003) | Y | Y | | | Y |
| 18 | Guan et al. (2003) | | Y | | Y | |
| 19 | Feng et al.(2004) | Y | | | | Y |
| 20 | Estevez Tapiador et al. (2004) | Y | | Y | | |
| 21 | Wang et al.(2004) | Y | | | | Y |
| 22 | Xiang et al.(2004) | Y | | | Y | |
| 23 | Qin et al.(2004) | Y | Y | Y | | |
| 24 | Siraj et al.(2004) | Y | Y | | | Y |
| 25 | Dasgupta et al.(2005) | Y | Y | | | Y |
| 26 | Casas et al. (2012) | Y | | | Y | |

 

| | | | | | | |
|---|---|---|---|---|---|---|
| 27 | Soleimani et al. (2012) | Y | Y | Y | | |
| 28 | Devarakonda et al. (2012) | Y | | | Y | |
| 29 | Lei et al. (2012) | Y | Y | | Y | |
| 30 | Kavitha et al. (2012) | Y | | | Y | |
| 31 | Gong et al. (2012) | Y | | | | Y |
| 32 | Pastrana S et al. (2012) | Y | | | | Y |
| 33 | Pereira C R et al. (2012) | Y | | | Y | Y |
| 34 | Sindhu S S S et al. (2012) | Y | | | Y | |
| 35 | Kang I et al. (2012) | Y | | Y | | |
| 36 | Li Y et al. (2012) | Y | | | Y | |
| 37 | Koc L et al. (2012) | Y | | | Y | |
| 38 | Altwaijry H et al. (2012) | Y | | | Y | |
| 39 | Jamdagni A et al. (2012) | Y | | Y | | |
| 40 | Chung Y Y et al. (2012) | Y | | | Y | |
| 41 | Lin S et al. (2012) | Y | | | Y | |
| 42 | Anming Z et al. (2012) | Y | Y | | | Y |
| 43 | BOULAICHE A et al. (2012) | | Y | | | Y |
| 44 | Zheng L et al. (2012) | Y | | | | Y |
| 45 | Pandaa M et al. (2012) | Y | Y | | Y | |
| 46 | Mohammed M N et al. (2012) | Y | | | | Y |
| 47 | Mukherjee D S et al. (2012) | Y | Y | | Y | |
| 48 | Chetan R et al. (2012) | Y | | | Y | |
| 49 | Brauckhoff D et al. (2012) | Y | | | | Y |
| 50 | Om H et al. (2012) | Y | | | Y | |
| 51 | Sharma M et al. (2012) | Y | | | Y | |
| 52 | Prasenna P et al. (2012) | Y | Y | Y | Y | |
| 53 | Hussein S M et al. (2012) | Y | Y | | Y | |
| 54 | Sharma S K et al. (2012) | Y | | | Y | |
| 55 | Sangkatsanee P et al.(2011) | | Y | | Y | Y |
| 56 | Mabu, S et al. (2011) | Y | Y | Y | Y | |
| 57 | Abadeh M S et al.(2011) | | Y | | Y | |
| 58 | Lee S et al.(2011) | Y | | | Y | |
| 59 | Tsai C et al. (2010) | Y | | | Y | |
| 60 | Chen C et al. (2010) | Y | | Y | | |
| 61 | Wang G et al(2010) | Y | | | Y | |
| 62 | Mok M S et al. (2010) | Y | | | Y | |
| 63 | Aydın M A et al. (2009) | Y | Y | | | Y |
| 64 | Hoang X D et al.(2009) | Y | | | | Y |
| 65 | Chou T et al.(2009) | Y | Y | | Y | |
| 66 | Shanmugam B et al.(2009) | Y | Y | Y | | |
| 67 | Powers S T et al.(2008) | Y | | | Y | |
| 68 | Ramachandran C et al. (2008) | Y | | | | Y |
| 69 | Xiang C et al. (2008) | Y | | | Y | |
| 70 | Tsang C et al. (2007) | Y | | | Y | |
| 71 | Kayacik H G et al.( 2007) | Y | | | Y | |
| 72 | Abadeh M S et al.(2007) | | Y | Y | | |
| 73 | Depren O et al. (2005) | Y | | | Y | |
| 74 | Jiang S Y et al.(2006) | Y | | Y | Y | |
| 75 | Perdisci R et al. (2006) | Y | | | Y | |

 

Table 2: Comparison of data mining methods in intrusion detection

| S. NO. | Author | A | AL | GA | C | SP | KNN | HMM | RB | ML | SVM | NN | DBTN | C 4.5 | NB | FL | OTH | Remarks |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Lippmann et al.(2000) | | | | | | | | | | | | | | | | Y | test bed |
| 2 | Gomez et al. (2001) | | | Y | | | | | | | | | | | | | | generate fuzzy classifier |
| 3 | Balajinath et al.(2001) | | | | | | | | | | | | | | | | Y | Genetic algorithm based on intrusion detection (GBID). |
| 4 | Barbara et al.(2001) | Y | | | Y | | | | | | | | | | | | | ADAM ( Audit Data Analysis and Mining) system |
| 5 | Lee W et al.(2001) | | | | Y | | | | | | | | | | | | | Mining Audit Data for Automated Models for Intrusion Detection(MADAMID) |
| 6 | Ye N (2001) | | | | | | | | | | | | | | | | Y | Clustering and Classification Algorithm Supervised (CCA-S) |
| 7 | Zhang et al.(2001) | | | | | Y | | | | | | Y | | | | | | Hierarchical Intrusion Detection (HIDE) system |
| 8 | Liao et al.(2002) | | | | | | Y | | | | | | | | | | | the k-Nearest Neighbor (kNN) classifier, is used to classify program behavior as normal or intrusive |
| 9 | Hoang et al.(2002) | | | | | | | Y | | | | | | | | | | A hidden Markov model (HMM) detection engine and a normal database detection engine have been combined to utilise their respective advantages. |
| 10 | Jha et al. (2002) | | | | | | | | Y | | | | | | | | | Pre-emptory, Reactionary |
| 11 | Ning et al. | | | | | | | | | | | | | | | | Y | coordinate attacks response and detection system |
| 12 | Florez et al.(2002) | Y | Y | Y | | | | | | | | | | | | | | improved algorithm for fuzzy data mining |
| 13 | Helmer et al.(2002) | | | Y | | | | | | Y | | | | | | | | distributed intrusion detection architecture |
| 14 | Ye et al. (2002) | | | | | | | | | | | | | | | | Y | multivariate quality control technique to detect intrusions |
| 15 | Yeung et al.(2003) | | | | | | | Y | | | | | | | | | Y | dynamic and static behavioral models |
| 16 | Jun et al.(2003) | | | | | Y | | Y | Y | | | | | | | | | Multiple measures intrusion detection model |
| 17 | Joo et al.(2003) | | | | | | Y | | | | | | | | | | | neural network model based on asymmetric costs |
| 18 | Guan et al. (2003) | | | | | | | | | | | | | | | | | Y means algorithm. |
| 19 | Feng et al.(2004) | | | | | | | | | | | | Y | | | | | intrusion detection systems (IDS) in a network environment |
| 20 | Estevez-Tapiador et al. (2004) | | | | | | | | | | | | | | | | Y | Markov chains |
| 21 | Wang et al.(2004) | | | | | | | | | | | | | | | | Y | Principle Component Analysis (PCA) |
| 22 | Xiang et al.(2004) | | | | | | | | | | | | | Y | | | | multiple-level tree classifiers |
| 23 | Qin et al.(2004) | | | | | | | | | | | | | | | | Y | |
| 24 | Siraj et al.(2004) | | | | | | | | | | | | | | | | Y | Decision Engine |
| 25 | Dasgupta et al.(2005) | | | | | | | | | | | | | | | | Y | CIDS |
| 26 | Casas et al. (2012) | | | | | | | | | | | | | | | | Y | UNIDS |

 

| S. NO. | Author | A | AL | GA | C | SP | KNN | HMM | RB | ML | SVM | NN | DBTN | C 4.5 | NB | FL | OTH | Remarks |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 27 | Soleimani et al. (2012) | | | | | | | | | | | | | | | | Y | Multi-layer episode filtering for the multi-step attack detection |
| 28 | Devarakonda et al. (2012) | | | | | | | Y | | | | | | | | | Y | Intrusion Detection System using Bayesian Network and Hidden Markov Model |
| 29 | Lei et al. (2012) | | | | | | | | | | | | | | | | Y | the improved competitive learning network (ICLN) and the supervised improved competitive learning network (SICLN) |
| 30 | Kavitha et al. (2012) | | | | | | | | | | | | | | | Y | Y | intrusion detection system using Neutrosophic Logic classifier |
| 31 | Gong et al. (2012) | | | | | | | | | | | | | | | | Y | FTNSA |
| 32 | Pastrana S et al. (2012) | | | | | | | | | | Y | | | | Y | | Y | detect malicious activities in MANETs |
| 33 | Pereira C R et al. (2012) | | | | | | | | | | Y | | | | Y | | Y | optimum-path forest (OPF) |
| 34 | Sindhu SSS et al. (2012) | | | | | | | | | | | | | Y | Y | | Y | lightweight Intrusion Detection System (IDS) |
| 35 | Kang I et al.. (2012) | | | | | | | | | | Y | | | | | | | Support vector data description (SVDD) |
| 36 | Li Y et al. (2012) | | | | | | | | | | Y | | | | | | Y | GFR |
| 37 | Koc L et al. (2012) | | | | | | | | | | | | | | Y | | | network intrusion detection system |
| 38 | Altwaijry H et al. (2012) | | | | | | | | | | | | | | Y | | | Bayesian based intrusion detection system |
| 39 | Jamdagni A et al. (2012) | | | | | | | | | | | | | | | | Y | RePIDS |
| 40 | Chung Y Y et al. (2012) | | | | | | | | | | | | | | | | Y | hybrid network intrusion detection system |
| 41 | Lin S et al. (2012) | | | | | | | | | | Y | | | | | | Y | intelligent algorithm with feature selection and decision rules |
| 42 | Anming Z (2012) | | | | | | | | | | | | | | | | Y | NFPA |
| 43 | BOULAICHE A et al. (2012) | | | | | | | Y | | | | | | | | | Y | quantitative-based approach for the detection and the prevention of intrusions |
| 44 | Zheng L et al. (2012) | | | | | | | | | | | | | | | | Y | Filter-ary-Sketch |
| 45 | Pandaa M et al. (2012) | | | | | | | | | | | | | | | | Y | A Hybrid Intelligent Approach for Network Intrusion Detection |
| 46 | Mohammed M N et al. (2012) | | | | | | | | | | Y | | | | | | | Intrusion Detection System Based on SVM for WLAN |
| 47 | Mukherjee D S et al. (2012) | | | | | | | | | | | | | | Y | | | Feature Vitality Based Reduction Method |
| 48 | Chetan R et al. (2012) | | | | | | | | | | | | | | | | Y | DAID |
| 49 | Brauckhoff D et al. (2012) | Y | | | | | | | | | | | | | | | | Anomaly Extraction in Backbone Networks |
| 50 | Om H et al. (2012) | | | | | | Y | | | | | | | | Y | | Y | A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System |
| 51 | Sharma M et al. (2012) | | | | | | | | | | | | | | | | Y | Pre-Clustering Algorithm |

 

| S. NO. | Author | A | AL | GA | C | SP | KNN | HMM | RB | ML | SVM | NN | DBTN | C 4.5 | NB | FL | OTH | Remarks |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 52 | Prasenna P et al.. (2012) | | | | | | | | | | | | | | | | Y | Network Programming And Mining Classifier For Intrusion Detection Using Probability Classification |
| 53 | Hussein S M et al. (2012) | | | | | | | | | | | | | | Y | | Y | Hybrid IDS Using Snort with Naïve Bayes to Detect Attacks |
| 54 | Sharma S K et al. (2012) | | | | | | | | | | | | | | Y | | Y | NIDS |
| 55 | Sangkatsanee P et al.(2011) | | | | | | | | | Y | | | | | | | | (RT-IDS) |
| 56 | Mabu s et al.(2011) | | | | | | | | | | | | | | | | Y | novel fuzzy class-association rule mining method based on genetic network programming (GNP) |
| 57 | Abadeh M S et al.(2011) | | | | | | | | | | | | | | | | Y | genetic fuzzy systems |
| 58 | Lee S et al.(2011) | | | | | | | | | | | | | | | | Y | a novel adaptive and dynamic clustering framework |
| 59 | Tsai C et al. (2010) | | | | | | Y | | | | | | | | | | Y | TANN |
| 60 | Chen C et al.(2010) | | | | | | | | | | | | | | | | Y | Lightweight Network Intrusion Detection system (LNID) |
| 61 | Wang G et al(2010) | | | | | | | | | | | | | | | | Y | FC-ANN |
| 62 | Mok M S et al.(2010) | | | | | | | | | | | | | | | | Y | Random effects logistic regression model |
| 63 | Aydın M A et al. (2009) | | | | | | | | | | | | | | | | Y | neural network model based on asymmetric costs of false positive errors and false negative errors |
| 64 | Hoang X D et al. (2009) | | | | | | | Y | | | | | | | | | Y | A program-based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference |
| 65 | Chou T et al.(2009) | | | | | | | | | Y | | | | | | | Y | hybrid intrusion detection model |
| 66 | Shanmugam B et al.(2009) | | | | | | | | | | | | | | | Y | Y | Improved Intrusion Detection System |
| 67 | Powers S T et al. (2009) | | | | | | | | | | | | | | | | Y | hybrid IDS |
| 68 | Ramachandran C et al. (2008) | | | | | | | | | | | | | | | | Y | FORK |
| 69 | Xiang C et al. (2008) | | | | | | | | | | | | | | | | Y | multiple-level hybrid classifier |
| 70 | Tsang C et al. (2007) | | | | | | | | | | | | | | | | Y | fuzzy rule-based system |
| 71 | Kayacik H G et al.( 2007) | | | | | | | | | | | | | | | | Y | A hierarchical SOM-based intrusion detection system |
| 72 | Abadeh M S et al.(2007) | | | | | | | | | | | | | | | | Y | Parallel genetic local search algorithm (PAGELS) |
| 73 | Depren O et al. (2005) | | | | | | | | | | | | | | | | Y | An intelligent intrusion detection system |
| 74 | Jiang S Y et al.(2006) | | | | | | | | | | | | | | | | Y | A clustering-based method for unsupervised intrusion detections |
| 75 | Perdisci R et al. (2006) | | | | | | | | | | | | | | | | Y | Alarm clustering for intrusion detection systems |

[**A:** Association rule, **C:** Classification, **SP:** Statistical Preprocessing, **NN:** Neural Network, **kNN:** k-Nearest Neighbor, **RB:** Rule based, **AA:** Apriori algorithm, **GA:** genetic algorithm, **SVM:** support vector machine, **HMM:** Hidden Markov model, **FL:** fuzzy Logic, **NB:** Naïve Bayes, **GMM:** Gaussian Mixture Model, **MLP:** Multilayer Perceptron (MLP), **LM:** Linear Model ]

 

## IV Discussion

This paper provides comprehensive study on the intrusion detection in the data mining domain. Our study concludes that there are two type of intrusion detection system that are anomaly detection and misuse detection. The figure 2 shows the percentage of papers based on the anomaly detection and misuse detection. This figure states that the ratio of papers on the anomaly detection is two times more than more than the misuse detection.

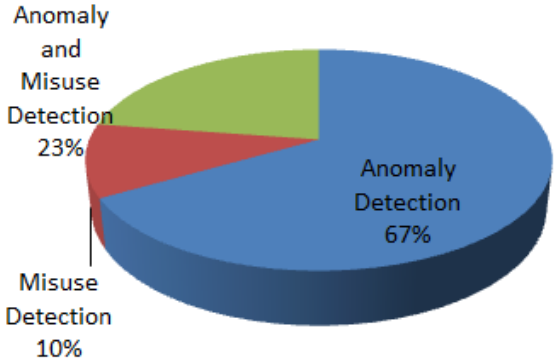


Fig. 2: Classification of intrusion detection based on detection approach

Existing IDS can be divided into two categories according to the detection approaches: anomaly detection and misuse detection. After reviewing these papers, we can see that almost 67 % researches focused on anomaly detection and 23 % researches focused on Anomaly and misuse detection and 10 % researches focused on misuse detection.

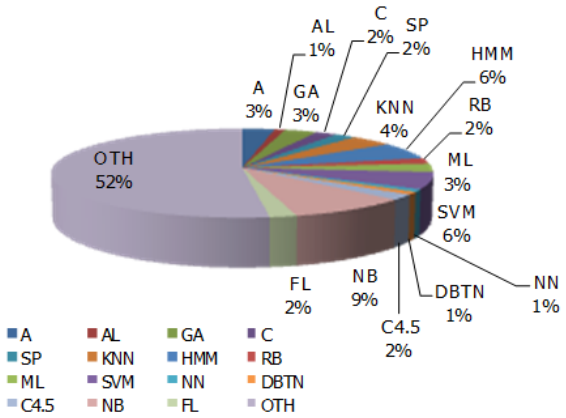


Fig. 3: Statistics of data mining model/ algorithm in intrusion detection

The figure 3 shows that the comparison of the different data mining techniques/algorithm for intrusion detection. The large numbers of data mining methods such as KNN, GA, HMM, SVM etc. are used for intrusion detection. There are fifteen main algorithms used in intrusion detection. Besides these fifteen algorithms, we put a category, other. Other category includes authors own proposed methods, most of the researcher proposed their own method but none of them are 100% correct. Most researches in intrusion detection uses NB (Naïve Bayes) main algorithm. Because NB is more stable than other models and algorithms. Besides, the second mostly used models are SVMs and HMM.

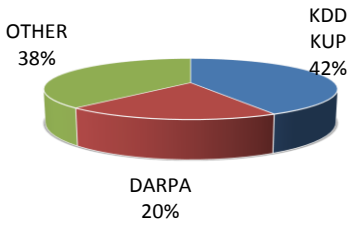


Fig. 4: Statistics of dataset that are used in reviewed papers

Figure 4 shows statistics of the datasets that are used for intrusion detection. Datasets are categorized in three categories DARPA, KDD Cup and some real world datasets used in these reviewed papers. The KDDCUP dataset is widely used by the researchers to test the effectiveness of the developed method for intrusion detection. 20 % of the studied papers used DARAPA dataset to check the effectiveness of the methods for intrusion detection, and the rest of the studied papers used other real world data sets.

## V Conclusion

The security is the primary concern in every field such as to prevent the computer networks from the intruders. This paper provides in depth study on the role of the data mining for intrusion detection. IDS can be divided into two categories according to the detection approaches: anomaly detection and misuse detection. Approaches for anomaly detection are neural network, statistics, predictive pattern generation, sequence matching and learning, approaches for misuse detection are expert system, pattern matching, state transition analysis. In this paper, we compared 75 different papers for finding out the situation of intrusion detection in current situation. After the comparison of these papers, we can see that most researches focused on anomaly detection. Most of the researchers uses DARPA and KDD Cup datasets for experiments.

## References

[1] http://en.wikipedia.org/wiki/Intrusion_detection_system.

[2] http://www.windowsecurity.com/articles/IDS-Part2-Classification-methods-techniques.html.

[3] Lippmann R. P., Fried D. J., Graf I., Haines J. W., Kendall K. R., McClung D., Weber D., Webster S.

E., Wyschogrod D., Cunningham R. K., Zissman M. A. , Evaluating Intrusion Detection Systems: The 1998 DARPA Off-line Intrusion Detection Evaluation, Proceedings DARPA Information Survivability Conference and Exposition (DISCEX), IEEE Computer Society Press, Los Alamitos, CA (2000).

[4] Gomez J., Dasgupta D., Evolving Fuzzy Classifiers for Intrusion Detection, Proceedings of the 2002 IEEE Workshop on Information Assurance United States Military Academy, West Point, NY June (2001).

[5] Balajinath B., Raghvan S. V., Intrusion detection through learning behavior model, Computer communication 24 (2001) 1202-1212.

[6] Ye N., A Scalable Clustering Technique for Intrusion Signature Recognition, Proceedings of the 2001 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY.(2001).

[7] Zhang Z., Li J., Manikopoulos C. N., and Jorgenson J., Ucles J., HIDE: a Hierarchical Network Intrusion Detection System Using Statistical Preprocessing and Neural Network Classification, Proceedings of the 2001 IEEE Workshop on Information Assurance and Security United States Military Academy, West Point, NY(2001).

[8] Hoang X. D., Hu J., Bertok P., Program based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference, Journal of Network and Computer Applications.32 (2002) 1219–1228.

[9] Jha S., Hassan M., Building agents for rule-based intrusion detection system, Computer Communication. 25(2002) 1366-1367.

[10] Florez G., Bridges S. M., Vaughn R. B., An Improved Algorithm for Fuzzy Data Mining for Intrusion Detection. IEEE. (2002).

[11] Helmer G., Wong J. S. K., Honavar V., Miller L., Automated discovery of concise predictive rules for intrusion detection, The Journal of Systems and Software 60 (2002) 165–175.

[12] Ye N., Emran S. M., Chen Q., Vilbert S., Multivariate Statistical Analysis of Audit Trails for Host-Based Intrusion Detection, IEEE Transactions on computers.51(2002).

[13] Yeung D., Ding Y., Host-based intrusion detection using dynamic and static behavioral models, Pattern Recognition. 36 (2003) 229 – 243.

[14] Jun S., Cho S., Detecting intrusion with rule based integration of multiple models, Computers and security, 22 (2003) 613-623.

[15] Feng L., Guan X., Guo S., Gao Y., Liu P., Predicting the intrusion intentions by observing system call sequences, Computers & Security .23(2004)241-252

[16] Estevez-Tapiador J. M., Garcıa-Teodoro P., Dıaz-Verdejo J. E, Measuring normality in HTTP traffic for anomaly-based intrusion detection, Computer Networks 45 (2004) 175–193.

[17] Wang W., Guan X., Zhang X., A Novel Intrusion Detection Method Based on Principle Component Analysis in Computer Security, Springer-Verlag Berlin Heidelberg (2004) 657–662.

[18] Xiang C., Chong M. Y., Zhu H. L., Design of Multiple-Level Tree Classifiers for Intrusion Detection System, Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems Singapore, (2004).

[19] Depren O., Topallar M., Anarim E., Ciliz M. K., An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks, Signal Processing.85 (2005) 463–479.

[20] Jiang S. Y., Song X., Wang H., Han J. J., Li Q. H., A clustering-based method for unsupervised intrusion detections, Pattern Recognition Letters. 27 (2006) 802–810.

[21] Perdisci R., Giorgio G., Roli F., Alarm clustering for intrusion detection systems in computer networks, Engineering Applications of Artificial Intelligence. 19 (2006) 429–438.

[22] Tsang C., Kwong S., Wang H., Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection, Pattern Recognition. 40 (2007) 2373 – 2391.

[23] Kayacik H. G., Zincir-Heywood A. N., Heywood M. I., A hierarchical SOM-based intrusion detection system, Engineering Applications of Artificial Intelligence. 20 (2007) 439–451.

[24] Powers S. T., He J., A hybrid artificial immune system and Self Organizing Map for network intrusion detection, Information Sciences. 178 (2008) 3024–3042.

[25] Ramachandran C., Misra S., Obaidat M. S., FORK: A novel two-pronged strategy for an agent-based intrusion detection scheme in ad-hoc networks, Computer Communications. 31 (2008) 3855–3869.

[26] Xiang C., Yong P. C., Meng L. S., Design of multiple-level hybrid classifier for intrusion detection system using Bayesian clustering and decision trees, Pattern Recognition Letters .29 (2008) 918–924.

[27] Hoang X. D., Hu J.,Bertok P., A program-based anomaly intrusion detection scheme using multiple detection engines and fuzzy inference , Journal of

Network and Computer Applications. 32 (2009) 1219–1228.

[28] Tsai C., Lin C., A triangle area based nearest neighbors approach to intrusion detection, Pattern Recognition. 43 (2010) 222 – 229.

[29] Chen C., Chen Y., Lin H., An efficient network intrusion detection, Computer Communications. 33 (2010) 477–484.

[30] Wang G.,Hao J., Ma J.,Huang L., A new approach to intrusion detection using Artificial Neural Networks and fuzzy clustering, Expert Systems with Applications 37 (2010) 6225–6232.

[31] Mok M. S., Sohn S. Y., Ju Y. H., Random effects logistic regression model for anomaly detection, Expert Systems with Applications. 37 (2010) 7162–7166.

[32] Lee S., Kim G., Kim S., Self-adaptive and dynamic clustering for online anomaly detection, Expert Systems with Applications. 38 (2011) 14891–14898.

[33] Casas P., Mazel J., Owezarski P., Unsupervised Network Intrusion Detection Systems: Detecting the Unknown without Knowledge, Computer Communications. 35 (2012) 772–783.

[34] Devarakonda N., Pamidi S., V. V. K., A. G. , Intrusion Detection System using Bayesian Network and Hidden Markov Model, Procedia Technology. 4 (2012) 506 – 514.

[35] Kavitha B.,Karthikeyan D. S., Maybell P. S. , An ensemble design of intrusion detection system for handling uncertainty using Neutrosophic Logic Classifier, Knowledge-Based Systems. 28 (2012) 88–96.

[36] Gong M., Zhang J., Ma J., Jiao L., An efficient negative selection algorithm with further training for anomaly detection, Knowledge-Based Systems. 30 (2012) 185–191.

[37] Pastrana S., Mitrokotsa A., Orfila A., Peris-Lopez P., Evaluation of classification algorithms for intrusion detection in MANETs, Knowledge Based Systems. 36 (2012) 217–225.

[38] Pereira C. R., Nakamura R. Y. M., Costa K. A. P., Papa J. P., An Optimum Path Forest framework for intrusion detection in computer networks, Engineering Applications of Artificial Intelligence. 25 (2012) 1226–1234.

[39] Sindhu S. S. S., Geetha S., Kannan A., Decision tree based light weight intrusion detection using a wrapper approach, Expert Systems with Applications. 39 (2012) 129–141.

[40] Kang I., Jeong M. K. , Kong D., A differentiated one-class classification method with applications to intrusion detection, Expert Systems with Applications. 39 (2012) 3899–3905.

[41] Li Y., Xia J., Zhang S., Yan J., Ai X., Dai K., An efficient intrusion detection system based on support vector machines and gradually feature removal method, Expert Systems with Applications .39 (2012) 424–430.

[42] Koc L., Mazzuchi T. A., Sarkani S., A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier, Expert Systems with Applications. 39 (2012) 13492–13500.

[43] Altwaijry H., Algarny S., Bayesian based intrusion detection system, Journal of King Saud University – Computer and Information Sciences .24 (2012), 1–6.

[44] Jamdagni A., Tan Z., He X., Nanda P., Liu R. P., RePIDS: A multi tier Real-time Payload-based Intrusion Detection System, Computer Networks (2012) .

[45] Chung Y. Y., Wahid N., A hybrid network intrusion detection system using simplified swarm optimization (SSO), Applied Soft Computing 12 (2012) 3014–3022.

[46] Lin S., Ying k., Lee C., Lee Z., An intelligent algorithm with feature selection and decision rules applied to anomaly intrusion detection, Applied Soft Computing. 12 (2012) 3285–3290.

[47] Zheng L., Zou P., Jia Y., Han W., Traffic Anomaly Detection and Containment Using Filter-Ary-Sketch. 2012 International Workshop on Information and Electronics Engineering (IWIEE), Procedia Engineering. (2011).

[48] Chetan R., Ashoka D.V., Data Mining Based Network Intrusion Detection System: A Database Centric Approach. 2012 International Conference on Computer Communication and Informatics (ICCCI -2012). (2012)

[49] Brauckhoff D., Dimitropoulos X., Wagner A. , Salamatian K. , Anomaly Extraction in Backbone Networks Using Association Rules, IEEE/ACM Transactions on networking.

[50] Om H., Kundu A., A Hybrid System for Reducing the False Alarm Rate of Anomaly Intrusion Detection System, 1st Int'l Conf. on Recent Advances in Information Technology. (2012).

[51] Sharma M., Toshniwal D., Pre-Clustering Algorithm for Anomaly Detection and clustering that uses variable size buckets, 1st Int'l Conf. on Recent Advances in Information Technology. (2012).

[52] Sharma S. K., Pande P., Tiwari S. K., Sisodiai M. S., An Improved Network Intrusion Detection Technique based on k-Means Clustering via Naive Bayes Classification. IEEE-International Conference On Advances In Engineering, Science And Management.(2012).

 

[53] Barbara B., Couto J., and Jajodia S., Wu N. , ADAM: A Testbed for Exploring the Use of Data Mining in Intrusion Detection,SIGMOD Record. 30 (2001) 15-24.

[54] Ning P., Jajodia S., Wang X. S., design and implementation of decentralized prototype system for detecting distributed attacks. Computer Communication. 25 (2002) 1374-1391.

[55] GuanY., Ghorbani A., Belacel N. , Y-Means: A Clustering Method for Intrusion Detection , Canadian Conference on Electrical and Computer Engineering. Montréal, Québec, Canada, (2003).

[56] Abadeh M. S., Habibi J.,Barzegar Z.,Sergi M., A parallel genetic local search algorithm for intrusion detection in computer networks, Engineering Applications of Artificial Intelligence. 20 (2007) 1058–1069.

[57] Sangkatsanee P., Wattanapongsakorn N., Charnsripinyo C., Practical real-time intrusion detection using machine learning approaches, Computer Communications. 34 (2011) 2227–2235.

[58] Abadeh M. S., Mohamadi H., Habibi J., Design and analysis of genetic fuzzy systems for intrusion detection in computer networks, Expert Systems with Applications. 38 (2011) 7067–7075.

[59] Boulaiche A., Bouzayani H. , Adi K. , A quantitative approach for intrusions detection and prevention based on statistical n-gram models, The 3rd International Conference on Ambient Systems, Networks and Technologies (ANT), Procedia Computer Science .10 ( 2012 ) 450 – 457.

[60] Mohammed M. N., Sulaiman N., Intrusion Detection System Based on SVM for WLAN, Procedia Technology. 1 (2012) 313 – 317.

[61] Lee W., A Framework for Constructing Features and Models for Intrusion Detection Systems, ACM Transactions on Information and System Security. 3(2001) 227-261.

[62] Liao Y., Vemuri R. V., Use of K-Nearest Neighbor classifier for intrusion detection,Computers & Security. 21 (2002) 439-448.

[63] Joo D., Hong T., Han I., The neural network models for IDS based on the asymmetric costs of false negative errors and false positive errors, Expert Systems with Applications .25 (2003) 69–75 .

[64] Qin M., Hwang K., Frequent Episode Rules for Intrusive Anomaly Detection with Internet Data mining, USENIX Security Symposium.(2004).

[65] Siraj A., Vaughn R. B., Bridges S. M., Intrusion Sensor Data Fusion in an Intelligent Intrusion Detection System Architecture, Proceedings of the 37th Hawaii International Conference on System Sciences. (2004).

[66] Dasgupta D., Gonzalez F., Yallapu K., Gomez J.,Yarramsettii R., CIDS: An agent-based intrusion detection system, Computers & Security .24(2005) 387-398.

[67] Aydın M. A., Zaim A. H., Ceylan K. G., A hybrid intrusion detection system design for computer network security, Computers and Electrical Engineering. 35 (2009) 517–526.

[68] Chou T., Chou T., Hybrid Classifier Systems for Intrusion Detection, 2009 Seventh Annual Communications Networks and Services Research Conference.(2009).

[69] Shanmugam B.,Idris N. B., Improved Intrusion Detection System using Fuzzy Logic for Detecting Anamoly and Misuse type of Attacks, 2009 International Conference of Soft Computing and Pattern Recognition.(2009).

[70] Mabu S., Chen C., Lu N., Shimada K., Hirasawa K., An Intrusion-Detection Model Based on Fuzzy Class-Association-Rule Mining Using Genetic Network Programming, IEEE Transactions on systems, MAN, and Cybernetics—Part C: Applications and Reviews. 41(2011).

[71] Soleimani M., Ghorbani A. A., Multi-layer episode filtering for the multi-step attack detection, Computer Communications. 35 (2012) 1368–1379.

[72] Lei Z. J., Ghorbani A. A., Improved competitive learning neural networks for network intrusion and fraud detection, Neurocomputing. 75 (2012) 135–145.

[73] Anming Z., An Intrusion Detection Algorithm Based On NFPA, 2012 International Conference on Medical Physics and Biomedical Engineering, Physics Procedia. 33 (2012) 491 – 497.

[74] Pandaa M., Abrahamb A., Patrac M. R., A Hybrid Intelligent Approach for Network Intrusion Detection, International Conference on Communication Technology and System Design 2011. Procedia Engineering. 30 (2012) 1 – 9.

[75] Mukherjee D. S., Sharma N. , Intrusion Detection using Naive Bayes Classifier with Feature Reduction , Procedia Technology. 4 (2012) 119 – 128.

[76] Prasenna P., RaghavRamana A.V.T, Kumar R. K., Devanbu A., Network Programming And Mining Classifier For Intrusion Detection Using Probability Classification, Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering.( 2012).

[77] Hussein S. M., Ali F. H. M., Kasiran Z., Evaluation Effectiveness of Hybrid IDS Using Snort with Naive Bayes to Detect Attacks, IEEE.(2012).

 

**Authors' Profiles**

**Mr. Chandrashekhar Azad** is a Research scholar in the department of information technology, Birla institute of technology Mesra, Ranchi. He has completed Bachelor degree in 2007 and master degree in 2011 from Ranchi University, Ranchi. Ranchi University awarded him gold medal for best post graduate in professional courses in 2012. His research interests include data mining, web mining and network security.

**Dr. Vijay Kumar Jha** is working as an Associate Professor in the Department of Information Technology, Birla institute of technology Mesra, Ranchi. He has completed BE (Electronics) from SIT Tumkur in 1996, M.Sc. Engineering in Electronics from MIT Muzaffarpur in 2007 and PhD in Information Technology (Data Mining) from MIT Muzaffarpur in 2011. He has been associated with Birla Institute of Technology, Mesra, Ranchi since 2001. His research interests includes Data mining, ERP etc.

**How to cite this paper:** Chandrashekhar Azad, Vijay Kumar Jha,"Data Mining in Intrusion Detection: A Comparative Study of Methods, Types and Data Sets", International Journal of Information Technology and Computer Science(IJITCS), vol.5, no.8, pp.75-90, 2013. DOI: 10.5815/ijitcs.2013.08.08

 