Journal of Theoretical and Applied Information Technology
28th February 2015. Vol.72 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MACHINE LEARNING TECHNIQUES FOR INTRUSION DETECTION SYSTEM: A REVIEW

[1]SUNDUS JUMA, [1]ZAITON MUDA, [1]M.A. MOHAMED, [2]WARUSIA YASSIN
[1]Faculty of Computer Science and Information Technology, University Putra Malaysia,
43400 UPM Serdang, Selangor, Malaysia
[2]Faculty of Information and Communication Technology, University Technical Malaysia Melaka,
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia
E-mail: [1]p.sundus@hotmail.com, [1]zaitonm@upm.edu.my, [2]warusia@gmail.com

ABSTRACT

Intrusion detection is considered as one of the foremost research areas in network security, the challenge is to recognize unusual access that could lead to compromising the interconnected nodes. Anomaly-based intrusion detection system, that utilizes machine learning techniques such as single classifier and hybrid classifier have the capability to recognize unpredicted malevolent. In this paper, we examine different machine learning techniques that have been proposed for detecting intrusion by focusing on the hybrid classifier algorithms. The objective is to determine their strengths and weaknesses. From the comparison, we hope to identify the gap for developing an efficient intrusion detection system that is yet to be researched.

Keywords: *Intrusion Detection, Anomaly Detection, Machine Learning, Hybrid Classifier, Single Classifier*

1. INTRODUCTION

Securing information is becoming one of the biggest challenges in an interconnected world. The highly valued information, coupled with the vulnerabilities of the system lead to significant growth of cyber-attacks. There are many tools designed to prohibit the internet-based attacks such as firewall, intrusion prevention system and intrusion detection systems (IDS). Particularly, IDS was developed as a tool for detecting attacks mounted over the network. As the sophistication of attacks increases, the proficient intrusion detection technique is required to prevent these vexatious activities. Consequently, IDS has been a centre for the research and become one of the fundamental components in computer security. Its primary responsibility is to identify malicious attacks in order to protect computer systems from possible damages.

Via detection techniques, we can categorize IDS into signature based detection (SBD) and anomaly based detection (ABD) [1]. SDB is capable of detecting attacks, of which its operational pattern has been identified, similar to that of anti-virus applications in detecting viruses. Unfortunately, the drawback of this technique is the incapacity to detect novel attacks. Moreover, there is a need for continuous updating of attack signatures for newly identified attacks. Alternatively, ABD works by determining any deviation from the regular usage patterns as an intrusion. For this, ABD needs to be familiar with normal usage via training. Machine learning techniques have been widely used to serve this purpose. Until recently, many techniques have been developed and each one has their own strengths. In general, the effectiveness of IDS is a measure of its proficiency to detect intrusion, to the least those that could potentially cause devastating damages. Some common parameters for measurement are detection rate, false positive, false negative, true positive and false alarm. Most of the existing techniques concentrated on improving the detection rate and therefore, to some degree, the field has been tremendously well researched. However, the current state of research shows that these techniques having a high rate of false alarms [2],[3]. This new problem has widened up the window for further investigation.

Two classes of machine learning techniques, namely single classifier and hybrid classifier have been applied to promote numbers of anomaly intrusion systems. Some researchers applied single classifier techniques such as Genetic Algorithms (GA), One-R (1R) and Random Forest (RF) while others applied a combination of classifiers such as

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

K-Means and One-R (KM+1R), Self-Organizing Maps and Support Vector Machine (SOM+SVM) and K-Means Clustering and Naïve Bayes Classifier (KMC+NBC), and so forth.

In this paper, we review a number of ABD techniques that have been developed for IDS. Each technique was tested using different available datasets targeting a different set of attacks. Our main concern is to determine the key advantages of each technique as well as their weaknesses. In the future, this paper can serve as a reference point and provide opportunities to improve the existing approach to further research in this field. However, in this paper, we bound the topic to clustering and the classification method as a hybrid classifier.

2. MACHINE LEARNING TECHNIQUES

In this section, we examine different machine learning techniques that include a single classifier and hybrid classifier. For each technique, we identify its strengths and weaknesses.

2.1 Single Classifiers

Single machine learning classifier can be used to address the problem of intrusion detection. In literatures, researchers have used machine learning techniques such as Support Vector Machine (SVM), Self-Organizing Maps (SOM) and K-Nearest Neighbour (KNN) to resolve this problem and the results have shown some significant achievements.

Author [4] proposed an approach for classification of attacks in Session Initiation Protocol (SIP)-based Voice over Internet Protocol (IP) environments using support vector machine (SVM). Initially, the analyser assesses an arrangement of pre-specified features for building up vectors. All vectors are defined as either attack or normal vector. The classifier determines if a vector corresponds to a several anomaly and issue an alarm event if needed. Experimental results show a performance of real time and better accuracy of identifying Spam over Internet Telephony (SPIT) attacks as well as flooding particularly when these combined with proficient event correlation rules. Unfortunately, their study did not cover other type of detection because these were considered to be their future work.

Author [5] proposed a simple and well-known learning algorithm known as 1R or One Rule. On particular tested attributes, 1R generates classification based the value of a single attribute. 1R selects the less error rate attribute for its "one rule". The error rate increases when there are instances that do not belong to the corresponding attribute class. Furthermore, the accuracy of 1R increases when applying more complex rules that have various implications for machine learning applications and researches. 1R has the ability to predict the accuracy of the rules that were created by more advanced machine learning systems.

RF is a popular regression technique and part of an ensemble classification approach. Author [6] proposed RF based on growing multiple randomized trees without pruning. A random Forest algorithm is a classification method for multifunction data using random sampling and attribute selection. Thus, RF is a natural choice for a variety of tasks such as probability estimation [7], prediction [8], and pattern analysis in Bioinformatics [9]. Author [10] proposed a framework for anomaly detection that utilizes a RF algorithm in order to detect uninhabitable intrusion. The implementation has been tested over different datasets acquired from the KDD Cup '99 datasets. Compared to other reported unsupervised anomaly detection approaches, this approach increases the detection rate and lowering down false positive rate.

Author [11] utilized a conditional probability for every relationship by analyzing the relationship between independent variables and dependent variables. Naïve Bayes (NB) is based on a very strong independence assumption with a quite simple structure. In [12] the author stated that Naïve Bayes classifiers (NBC) offer more reasonable result, even with an easy structure. Experiments showed that NB is very competent in classification task, but lack in classifying User-to-Root (U2R) and Remote-to-Local (R2L) based attacks correctly.

Another use of NBC was proposed for anomaly based network intrusion detection in [13]. It was demonstrated that the NBC is extra effective in identifying network intrusion as a contrast to the neural network technique. However, they found that although their technique generates reasonable detection rate of up to 95%, the generation of false positive remains high. Due to this, the research on single classifier has been slowly replaced by the hybrid classifier which offers much promising results.

2.2 Hybrid Classifiers

The essential objective of the development of IDS is to acquire superior probable accuracy for a task [14]. This goal leads to the invention of hybrid classifiers to conquer the problems related to IDS.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Hybrid classifier combines many machine learning techniques, with one intention in mind, to further enhance the performance of the detection system [15].

Particularly, there are two kinds of hybrid classifiers. The first type can be based on combining diverse classifiers such as Nero-fuzzy technique [16]. The second type of hybrid classifier can be based on combining clustering technique and classification technique together. In this paper, we bound our concentration of the second type of hybrid classifier.

In hybrid machine-learning model that combines clustering and classification techniques, the clustering is employed as the initial module for "pre-classification" assignment and the classification technique as the subsequent module is employed for the ultimate classification task [17],[18]. In particular, the clustering technique performs data reduction task by filtering out unrepresentative data. Specifically, the data which will not be able to cluster correctly can be considered as outlier's data. The typical data devoid of the outlier's data is applied to train the classifier in an aim to maximize the classification result.

In contrast, the classification method can be applied as the initial module and the clustering as the subsequent module as the clustering method cannot differentiate the data more precisely. Hence, a classifier can be trained a priori, and its production is later utilized as the input for the cluster to increase the clustering result.

Author [19] proposed a weighted polynomial equation as the objective method of an evolutionary procedure. The factor values of the proposed equation are anomalous scores of every single Transmission Control protocol/Internet Protocol (TCP/IP) fields in compliance to numerous anomaly attacks. Genetic Algorithm (GA) was used with the proposed objective function for the optimized field selection because GA has fast processing time and good detection rate. The result has shown an increasing in detection rates of up to 97.56% when the number of selected fields used is 15.

Author [20] introduced a Flexible Neural Tree Model (FNT) based on a combination of genetic algorithms and neural networks. Moreover, the IDS was modelled using a hierarchical hybrid intelligent system that combines support vector machine (DT-SVM) and decision tree. While the DT-SVM produced high detection rate, it's insufficient in the proficiency to segregate an attack from normal behaviour [21].

Author [22] introduced a parallel genetic local search algorithm. In this algorithm the universal population is separated into several sub-populations, every one appointed to a discrete processor. Every single sub-population comprises of the similar class fuzzy rules. Results disclosed that the proposed algorithm was able to maximize the detection rate to 96.3% and minimize the false alarm rate to 0.29% concurrently. The training duration of the presented learning algorithm is reduced extensively using the recommended parallel learning framework. This improvement might be utilized to create high-performance classifiers which able to solve the complex classification problems in a significant short computation duration.

Author [23] proposed a novel supervised network intrusion detection method based on Transductive Confidence Machines for K-Nearest Neighbours (TCM-KNN) machine learning algorithm and active learning based training data selection algorithm. This approach can proficiently identify anomalies with higher detection rate, lower false positives under the condition of using much less chosen data and chosen features for training in contrast to the traditional supervised intrusion detection methods. A number of experiments conducted against the well-known KDD Cup '99 dataset demonstrate that this method is more robust and efficient than the state-of-the-art intrusion detection methods.

Author [3] proposed a hybrid learning model based on the Triangle Area-based Nearest Neighbours (TANN) which consists of K-Means clustering and K-Nearest Neighbours (K-NN) classifier in order to detect attacks efficiently. Initially, K-Means clustering is used to extract a number of cluster centres that represents one particular category of attacks before the K-NN classifier is applied. This approach offered a high detection rate at 98.95%, but unfortunately came with a high false alarm rate at 3.83%.

According to [24], many works have been done to enhance detection capabilities of IDS. Artificial Neural Network (ANN) has been widely used and successfully applied to solve many complex practical problems. However, ANN-based IDS failed in detecting low-frequent attacks such as R2L and U2R based attacks. ANN meets difficulties to learn the behaviour of this kind of attack because the learning sample size is too small compared to the high-frequent attacks. Therefore, the author proposed a novel approach for ANN-based IDS using ANN and Fuzzy Clustering called FC-ANN

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to overcome these drawbacks. The fuzzy clustering approach is applied to generate different training subsets before a different ANN models are trained to formulate different models. Then, fuzzy aggregation module is employed to aggregate the result. Each subset of the training set has a lower complexity by employing fuzzy clustering and this directly enable the ANN to learn each subset more precisely in detecting low-frequent attacks. As a result, FC-ANN approach achieved a higher detection rate for R2L and U2R attacks, respectively at 83.33% and 93.18%, compared to other well-known methods. Nonetheless, a potential drawback of this approach is in its weakness to detect probe based attacks with just at 48.12%.

Author [25] proposed SVM-based IDS with BIRCH hierarchical clustering as a pre-processing phase and simple feature selection procedures to reduce some unimportant features. The performance of SVM is further improved by using a hierarchical clustering algorithm. Moreover, the simple feature selection procedure allows SVM to categorize data much more accurately. The author showed that this technique can achieve high detection rates at 99.5%, 97.5%, and 99.3% for Denial of Service (DoS), probe and normal data respectively. However, the predicted percentage of this system to make distinction between R2L and normal data was decreased dramatically to 28.8%.

Author [26] used the Feature Selection (FS) method in order to improve the performance of existing classifiers by excluding non-consequential features. Furthermore, an improved Partial Least Squares Support Vector Machine called PLSSVM has been introduced. A linear and non-linear measure for the feature selection within pre-processing phase has been considered in this work. PLSSVM performed well in classifying normal and probe attacks records, respectively at 95.69% and 86.46%. In contrast, PLSSVM missed a big number of dynamic attacks such as DoS and U2R attacks that behave quite similar to the normal behaviour, which were recorded at 78.76% and 30.7% respectively.

Author [27] used K-Means clustering and One-R classifier (KM+1R) to produce a hybrid machine learning technique. The essential solution is to break up instances between the possible attacks and the normal instances through a first stage in different cluster. Afterwards, the clusters are assorted into Probe, R2L, U2R, DoS and Normal attacks. The performance of KM+1R was measured using KDD Cup '99 dataset. This hybrid approach achieved a low false alarm rate at 2.73%, while accuracy and detection rate at 99.26 and 99.33 respectively.

Author [28] used the hybrid of Support Vector Machine (SVM) and genetic algorithms (GA) to evaluate IDS's performance by examining the most representative parameters which are accuracy, false and true alarms. The result reported high accuracy rate to detect intrusion at 98.33% and good quality percentage of false positive, true positive, false negative and true negative at 1.78%, 99.49%, 0.5%, 98.21%, respectively, compared to technique based on SVM alone which was 2.14, 98.75%, 1.25% and 97.85% respectively.

Author [29] introduced the hybrid approach that contains RF classifier and Synthetic Minority Oversampling Technique (SMOTE). RF was used to develop proficiency and effectual IDS while SMOTE was applied to enhance the detection rate of R2L and U2R classes in the imponderables training dataset. After that single out the whole of the fundamental features of the minority classes using R2L and U2R classes attack mode. Using this approach it was shown that the time required to build the model was decreased and the detection rates for R2L and U2R were increased to 0.963% and 0.962% respectively when the number of features is 22.

Author [30] proposed hybrid approach that combines K-Means clustering and Naïve Bayes Classifier (KMC+NBC). The performance of the proposed approach was evaluated using the ISCX 2012 dataset. The group of authors concluded that the KMC+NBC has considerably enhanced the accuracy and detection rate up to 99% and 98.8%, respectively, whereas decreased the false alarm rate to 2.2%.

Author [31] proposed a novel hybrid learning model for intrusion detection based on k-means clustering, naïve bayes and decision table majority approaches to improve the detection rate, accuracy and false alarm. K-means clustering employed as a pre-classification component to gather a similar behaviour of data in a single group. Next, the clustered grouped by data classified into normal and abnormal classes through the naïve Bayes classifier in order to reduce the amount of missclassified results during the clustering stage. Subsequently, the classified data pass to decision table majority for conclusive progression. These methods evaluated with KDD Cup '99 dataset. However, no result reported on this work.

A proficient integrated data mining methods consisting of k-means clustering, neural network as

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

well as support vector machine has been applied as a competent intrusion detection model in [32]. The clustering algorithm used to group the entire input data into k clusters sets, and subsequently these clustered data trained with neural network. In the final stage, the data classified with support vector machine for intrusion detection. The approach has been assessed using KDD Cup '99 dataset, the result has compared with other existing methods, and the performance estimated in term of accuracies. The result of the experiments has shown an improvement with 97.5%, 98.7%, 98.8%, 98.8% as the averaged accuracy rate for the type of attacks of DoS, Probe, R2L and U2R.

Moreover, author [33] has proposed a hybrid framework which is based on clustering and association for better intrusion detection. The clustering algorithm utilized to split the data in a basis of different classes for the reason the classes can be classified further in the next stage. Thus, the association classifier based on the FP growth algorithm used to classify the data correspondingly into a similar set of category. In order to evaluate the effectiveness of these algorithms, the KDD Cup '99 dataset considered. The proposed algorithm manages to obtain 100%, 100%, 97% and 99% as an accuracy rate for the attack type of DoS, U2R, R2L and Probe.

In addition, a novel method through data mining techniques for obtaining high detection rate has been carried on [34]. Out of 5 principal stages, two of them involving a process of clustering and classification using k-means, support vector machine and particle swarm optimization. K-means used to produce five different training subsets where 4 of them are for intrusion dataset while the remaining for normal dataset. Based on these sets, a vector formed for support vector and particle swarm optimization classification task. However, no result reported on this work.

Table 1 lists a number of recent works on IDS based on hybrid classifier, techniques are compared and dataset, detection techniques, related issues, and measurement. The detection rate (DR), true positive (TP), false alarm (FA), false positive (FP), false negative (FN) and accuracy AC) for every approach are also investigated. Each approach has unique advantages and disadvantages.

Briefly, many hybrid classifiers has been developed in intrusion detection fields such as feature selection with SVM, BIRCH Clustering with SVM, ANN with Fuzzy Clustering, AIN with NN, Decision Tree with SVM, Genetic Algorithm with SVM, SOM with ANN and so forth; but there is still room to improve the accuracy and detection rate and the false alarm rate.

The aforementioned literature discussed works were significant in using clustering as a method to pre-process datasets. In recent times, a lot of learning techniques have been investigated in selection to clustering and classification for the task of anomaly detection. The flexibility of hybrid approach makes it one of the popular among others.

3. OPEN RESEARCH ISSUES

Choosing a significant hybrid classifier is not an easy tack to be accomplished. The ineffective combinational could downgrade the detection performance. For example, the rate of false alarm of previous work is not much improved. The hybrid approaches should focus to maximize the rate of true positive and true negative as well as minimize the rate of false positive and false negative, which could directly increase the rate of accuracy with the better false alarm rate. Moreover, a number of previous works evaluated with KDD Cup 1999 dataset that contains the old type or version of the attacks. Nowadays, the nature of an attack has undergone significant changes and resembles the characteristic of a normal situation where it is difficult to be identified. This remains as an open challenge in the research community. In addition, constraints on clustering stage (i.e. errors of cluster arrangement or cannot differentiate and group the data more accurately) need to be improved on earlier to help the classification method get the best performance.

4. CONCLUSION

This paper reviews a series of studies on IDS that is based on machine learning techniques. Particularly, we reassess papers published between the year 1993 and 2014. Many machine learning techniques such as single classifier and hybrid classifier have been used in the domain of intrusion detection. Each technique has successfully shown significant improvement over the others. However, in general, there is no one technique that is the best among others, except under certain working conditions. We conclude that the research into IDS utilizing machine learning technique is way from complete and therefore, further improvements are yet to be undertaken.

REFERENCES:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[1] W. Lee, J.S. Stolfo and W.K. Mok, "A data mining framework for adaptive intrusion detection", *Proceedings of the lEEE Symposium on Security and Privacy*, New York, USA, 1998, pp. I20-132.

[2] S. Juma, Z. Muda and W. Yassin, "Reducing False Alarm Using Hybrid Intrusion Detection based on X-Means Clustering and Random Forest Classification", *Journal of Theoretical and Applied Information Technology*, Vol. 68, No. 2, 2014, 249-254.

[3] C.F. Tsai and C.Y. Lin, "A triangle area based nearest neighbours approach to intrusion detection", *Pattern Recognition*, Vol. 43, 2010, pp.222-229.

[4] M. Nassar, R. State and O. Festor, "Monitoring SIP traffic using Support Vector Machines", *In Proc. of 11th International Symposium on Recent Advances in Intrusion Detection (RAID)*, 2008.

[5] R. Holte, "Very simple classification rules perform well on most commonly used datasets", *Machine Learning* , Vol. 11, 1993, pp.63-91.

[6] L. Breiman, "Random forests", *Machine Learning* , Vol. 45, 2001, pp.5–32.

[7] Wu, T.F., Lin, C.J., and Weng, R.C. (2004). Probability estimates for multi-class classification by pairwise coupling. Journal of Machine Learning Research, 5: 975-1005.

[8] B.E. Popescu, "Ensemble learning for prediction", Stanford, CA, USA Stanford University, 2004.

[9] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos S.Z. Li "Ensemble-based discriminant learning with boosting for face recognition" *Neural Networks, IEEE Transactions*, Vol. 17, 2006, pp.166-178.

[10] J. Zhang, M. Zulkernine and A. Haque, "Random-Forests-Based Network Intrusion", *IEEE Transactions on Systems, Man, And Cybernetics—Part C: Applications And Reviews*, Vol. 38, 2008, pp.649-659.

[11] H.J. George, "Estimating Continuous Distributions in Bayesian Classifiers", *In Proceedings of the Eleventh Conference on Uncertainly in Artificial Intelliegence, Morgan Kaufmann Publishers*, San Mateo, 1995.

[12] B.A. Nahla, B. Salem and and E. Zied, "Naive Bayes vs Decision Trees in Intrusion Detection Systems", *In Proceeding of the ACM Symposium on Applied Computing*, Nicosia, Cyprus, 2004.

[13] P. Mrutyunjaya and R.P. Manas, "Network Intrusion Detection Using Naïve Bayes", *International Journal of Computer Science and Network Security*, Vol. 7, No. 12, 2007, pp.258-263.

[14] W. Yassin, N.I. Udzir, A. Abdullah, M.T. Abdullah, Z. Muda and H. Zulzalil, "Packet Header Anomaly Detection Using Statistical Analysis", *International Joint Conference SOCO'14-CISIS'14-ICEUTE'14*, Vol. 47, No. 299, 2014, pp.473–482.

[15] M. Mohammadi, Z. Muda, W. Yassin and N.I. Udzir, "KM-NEU: An Efficient Hybrid Approach for Intrusion Detection System", *Research Journal of Information Technology*, Vol. 6, No.1, 2014, pp.46–57.

[16] J.S. Jang and C-T. Sun, "A Neuro-Fuzzy Classifier and its Applications", *Second IEEE International Conference on Fuzzy Systems,* 1993, pp 94-98.

[17] Y. Liu, K. Chen, X. Liao W. Zhang, "A genetic clustering method for intrusion detection", *Pattern Recognition*, Vol. 37, 2004, pp.927–942.

[18] K. Latifur, A. Mamoun and T. Bhavani, "A New Intrusion Detection System Using Support Vector Machines And Hierarchical Clustering", *The VLDB Journal*, Vol. 16, 2007, pp.507-521.

[19] T. Shona, X. Kovahb and J. Moon, "Applying genetic algorithm for classifying anomalous TCP/IP packets", *Neurocomputing* , Vol. 69, 2006, pp.2429–2433.

[20] Y. Chen, A. Abraham and B. Yang, "Hybrid Flexible Neural-Tree-Based Intrusion Detection Systems" *International Journal of Intelligent Systems*, Vol. 22, 2007, pp.37–352.

[21] S. Peddabachigaria, A. Abraham, C. Grosanc and J. Thomas, "Modeling intrusion detection system using hybrid intelligent systems", *Computer Applications*, Vol. 30, 2007, pp.114–132.

[22] M.S. Abadeh, J. Habibi, Z. Barzegar and M. Sergi, "A parallel genetic local search algorithm for intrusion detection in computer networks", *Engineering Applications of Artificial Intelligence*, Vol. 20, 2007, pp. 1058–1069.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[23] Y. Liu and L. Guo, "An active learning based TCM-KNN algorithm for supervised network intrusion detection", *Computers & security* , Vol. 26, 2007, pp.459–467.

[24] W. Gang, H. Jinxing, and M. Jian, "A New Approach to Intrusion Detection Using Artificial Neural Networks and Fuzzy Clustering", *Expert Systems with Applications*, Vol. 37, 2011, pp.6225–6232.

[25] S.J. Horng, M.Y. Su, Y.H. Chen, T.W. Kao, R.J. Chen, J.L. Lai, and C.D. Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines", *Expert Systems with Applications*, Vol. 38, 2011, pp.306–313.

[26] F. Amiri, M. Yousefi, C. Lucas, A. Shakery and N. Yazdani, "Mutual Information-Based Feature Selection for Intrusion Detection Systems", *Journal of Network and Computer Applications*, Vol. 34, 2011, pp.1184–1199.

[27] Z. Muda, W. Yassin, M.N. Sulaiman and N.I. Udzir, "Intrusion Detection based on K-Means Clustering and OneR Classification", *In 7th International Conference on Information Assurance and Security (IAS),* Melaka, Malaysia, 2011, pp. 192-197.

[28] K. Atefi, S. Yahya, A.Y. Dak and A. Atefi, "A Hybrid Intrusion Detection System based on Different Machine Learning Algorithms", *Proceedings of the 4th International Conference on Computing and Informatics*, 2013, pp.312-320.

[29] A. Tesfahun and D.L. Bhaskari, "Intrusion Detection using Random Forests Classifier with SMOTE and Feature Reduction" *2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies*, 2013, pp. 127-132.

[30] W. Yassin, N.I. Udzir, Z. Muda and M.N. Sulaiman, "Anomaly-Based Intrusion Detection through K-Means Clustering and Naives Bayes Classification", *Proceedings of the 4th International Conference on Computing and Informatics (ICOCI)*, 2013, pp. 298-303.

[31] A. Purohit and H. Gupta, "Hybrid Intrusion Detection System Model Using Clustering, Classification and Decision Table", *Journal of Computer Engineering*, Vol. 9, No. 4, 2013, pp.103-107.

[32] A.M. Chandrashekhar and K. Raghuveer, "Amalgamation of K-means Clustering Algorithm with Standard MLP and SVM Based Neural Networks to Implement Network Intrusion Detection System", In M. Kumar Kundu, D. P. Mohapatra, A. Konar, and A. Chakraborty (Eds.), *Advanced Computing, Networking and Informatics,* Vol. 28, 2014, pp.273–283.

[33] M. Somani and R. Dubey, "Hybrid Intrusion Detection Model Based on Clustering and Association", *International Journal of Advanced Research in Electrical, Electronics and Instrumentations Engineering,* Vol. 3, No. 3, 2014, pp.8152-8160.

[34] H. Saxena and V. Richariya, "Intrusion Detection System Using K-Means, PSO with SVM", *International Journal of Emerging Technology and Advanced Engineering,* Vol. 4, No. 2, 2014, pp.653-657.

[35] H.G. Kayacik, Z.H. Nur and M.I. Heywood, "A hierarchical SOM-based intrusion detection system", *Engineering Applications of Artificial Intelligence*, Vol. 20, 2007, pp.439–451.

[36] T. Ozyer, R. Alhajj and K. Barker, "Intrusion detection by integrating boosting genetic fuzzy classifier and data mining criteria for rule pre-screening", *Journal of Network and Computer Applications*, Vol. 30, 2007, pp.99–113.

[37] W. Su-Yun and E. Yen, "Data mining-based intrusion detectors", *Expert System With Application*, Vol. 36, 2009, pp.5605-5612.

[38] T. Shon and J. Moon, "A hybrid machine learning approach to network anomaly detection", *Information Sciences*, Vol. 177, 2007, pp.3799–3821.

[39] T. Arman, R. Mohammad and M. Abdolreza, "Intrusion detection using fuzzy association rules" *Applied Soft Computing*, Vol. 9, 2009, pp.462–469.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Comparison of Related Works

Author	Technique	Related Issues	Dataset	Measurement	Advantage	Disadvantage
Liu et al., 2004 [17]	SOM+ANN	Anomaly & Misuse Detection	DARPA 1998	DR, FA, FP	DR: 97.1%	FP: 2.8%
Shon et al., 2006 [19]	GA+ANN/ K-NN/SVM	Anomaly Detection	DARPA 1998	DR, FP, FN	DR: 98.6%	FP: 2.5% FN: 11%
Abadeh et al., 2007 [22]	GA+FL	Anomaly Detection	DARPA 1998	DR, FA	FA: 0.29% DR: 96.3%	-
Liu and Guo, 2007 [23]	TCM k-NN	Anomaly Detection	KDD-Cup '99	TP, FP	TP: 99.6% FP: 0.1%	-
Chen et al., 2007 [20]	GA +ANN	Anomaly Detection	DARPA 1998	FP, FN	FP: 0.61% FN: 0.37% DR: 97.98%	-
Kayacik et al., 2007 [35]	SOM	Anomaly Detection	KDD Cup '99	FP, DR	DR: 90.4%	FP: 1.38%
Latifur et al., 2007 [18]	SOM+SVM	Anomaly Detection	DARPA 1998	FP, FN, AC	-	AC: 69.8% FP: 37.8% FN: 29.8%
Ozyer et al., 2007 [36]	Genetic Fuzzy Classifier	Anomaly & Misuse Detection	KDD Cup '99	DR	-	DR: 53.02%
Shon et al., 2007 [38]	GA +SVM	Anomaly Detection	DARPA 1999	DR, FP, FN	-	DR: 87.4% FP: 10.20% FN: 27.27%
Su-Yun et al., 2009 [37]	DT, SVM	Anomaly Detection	KDD Cup '99	AC, FP, DR	FA: 1.0%	DR: 70.62% AC: 64.94%
Arman et al., 2009 [39]	Clustering, K-NN, SVM, H-SOM	Anomaly & Misuse Detection	KDD Cup '99	FP, DR	DR: 98%	FP: 10%
Tsai et al., 2010 [3]	KM+K-NN, SVM, K-NN and TANN	Anomaly Detection	KDD Cup '99	AC, DR, FA	DR: 98.95% AC: 96.91%	FA: 3.83%
Gang et al., 2010 [24]	ANN+Fuzzy Clustering	Anomaly Detection	KDD Cup '99	AC, FP, DR	AC: 96.71%	-
Horng et al., 2011 [25]	BIRCH Clustering+SVM	Anomaly Detection	KDD Cup '99	AC, FP, DR	AC: 95.7%	-
Muda et al., 2011 [2]	KM+1R	Anomaly Detection	KDD Cup '99	AC, DR, FA	AC:99.26% DR:99.33%	FA:2.73%
Tesfahun et al., 2013 [29]	SMOTE +FS+RF	Anomaly Detection	NSL-KDD	DR, FP	AC of R2L: 0.963, AC of U2R:0.962	-
Yassin et al., 2013 [30]	KMC+NBC	Anomaly Detection	ISCX 2012	AC, DR, FA	AC: 99% DR: 98.8%	FA: 2.2%
Chandrasekhar et al., 2014 [32]	k-means, NN, SVM	Anomaly Detection	KDD Cup '99	AC	AC: 98.45%	-
Somani et al., 2014 [33]	Clustering and Association	Anomaly Detection	KDD Cup '99	AC	AC: 99%	-