Journal of Engineering and Applied Sciences 11 (Special Issue 2): 3254-3264, 2016
ISSN: 1816-949X

Optimize Machine Learning Based Intrusion Detection for Cloud Computing: Review Paper

[1]Mohammed Hasan Ali, [2]Mohamad Fadli Zolkipli and [1]Mohammed Abdulameer Mohammed
[1]Faculty of Computer Systems and Software Engineering, School of Computing,
University Utara Malaysia, Changlun, Kedah, Malaysia
[2]University Malaysia Pahang, Pahang, Malaysia

Abstract: Security is a rich research area and there are many solutions create to protect the information and make the systems safer, intrusion detection is one of the powerful solutions in security. Current day network Intrusion Detection Systems (IDS) has several flaws such as low detection rates and high rates of false positive alerts and the need for constant human intervention and tuning. This research shows some of the related researchers based on IDS, shows the advantages and limitations of these researches also this research focus on IDS based hybrid as powerful more than the single systems. By use two or more methods and algorithms in one system, to take advantages from each of them as they algorithms complement the other. This research tries to analysis the data set. KDD99 is the most popular data set in the IDS. It's facing some disadvantages even the new version NSL-KDD still facing some problems.

Key words: IDS, powerful solutions, advantages and limitations, algorithms complement, Malaysia

INTRODUCTION

Cloud computing moves the application software and databases to the large data centers where the management of the data and services are not trustworthy for the user. This unique attribute, however, poses many new security challenges (Wang *et al*., 2009). The fully distributed and open structure of cloud computing and services becomes an even more attractive target for potential intruders. It involves multi-mesh distributed and service oriented paradigms, multi domains and multi user autonomous administrative infrastructures which are more vulnerable and prone to security risks (Patel *et al*., 2013). Figure 1 it illustrates some of the recent statistics about attack distributions for 2012.

All network and computer systems need to defend against a number of threats. Amateur hackers, rival corporations, terrorists have the motive and capability to carry out sophisticated attacks against computer systems (Choo, 2011). Therefore, the field of information security has become vitally important to the safety and economic well-being of society as a whole. The rapid growth and widespread use of electronic data processing and electronic business conducted through the massive use of the wired and wireless communication networks, the Internet, web application, cloud computing along with numerous occurrences of international terrorism, raises

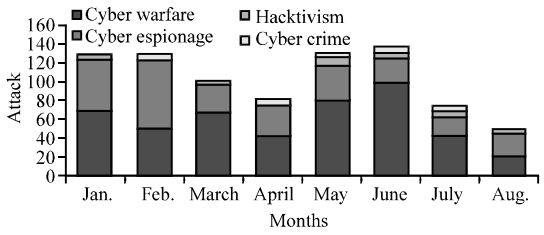

Fig. 1: Attack distribution for 2012 (Alharkan and Martin, 2012)

the need for providing secure and safe information security systems through the use of firewalls intrusion detection and prevention systems, encryption, authentication and other hard-ware and software solutions. The cloud computing had been changing the way computer resources have been used up to now. By make it available for everyone and this is the reason why cloud computing become attract attention for the attacker. Traditional network security mechanisms face new challenges in the cloud such as virtual machine intrusion attacks and malicious user activities. New security methods are therefore needed to increase user's level of trust in clouds. Currently, cloud providers enforce data encryption for the storage containers, virtual fire walls and access control lists. However, cloud consumers need to develop secure and customizable solutions to comply

Corresponding Author: Mohammed Hasan Ali, Faculty of Computer Systems and Software Engineering, School of Computing, University Utara Malaysia, Changlun, Kedah, Malaysia

with their application requirements. There are many tools and developed to provide solve to security problem in cloud computing. Intrusion Detection System (IDS) is one of the most important components of security. This research proposes to develop a new model of IDS that has been used to solve complex optimization problem can be adapted to addressing detect an attack problem in cloud computing.

MATERIALS AND METHODS

Data set: The rapid development of technology over the internet makes computer security a critical issue. Now a days, artificial intelligence, data mining and machine learning algorithms have been subjected to extensive research in intrusion detection with emphasis on improving the accuracy of detection and make an immune model for Intrusion Detection System (IDS) to tackle zero day attacks or novel attacks. To make an IDS model faster with more accurate detection rates, selection of important features from the input dataset is highly essential. Feature selection in learning process while design the model leads to reduction in computational cost, over fitting, model size and improve accuracy (Tavallaee *et al*., 2009).

Create a collection of data include normal data and kinds of attack was a powerful point to find the detect rate and false alarm rate and other measurements for the models that proposed. During 1998 and 1999 group of MIT has collected and distributed the first standard corpora for evaluation of computer network intrusion-detection systems under Defense Advanced Research Projects Agency (DARPA) the initial formal, repeatable and statistically significant evaluations of intrusion-detection systems.

These evaluations contributed significantly to the intrusion-detection research field by providing direction for research efforts and an objective calibration of the technical state of the art. They are of interest to all researchers working on the general problem of workstation and network intrusion detection. The evaluation was designed to be simple to focus on core technology issues and to encourage the widest possible participation by eliminating security and privacy concerns and by providing data types that were used commonly by the majority of intrusion-detection systems. Lee and Stolfo, 2000) used data mining techniques to select KDD features from DARPA 98 dataset. Intrusion detection datasets contain huge amount of observations or records with high dimensional data. To handle such large datasets and implement a model for IDS is not an easy task for the novice researcher (Lee and Stolfo, 2000).

KDD CUP 99: Many researchers since 1999 evaluate their intrusion-detection models using KDD Cup 99 dataset. It is 15 year old benchmark data set which is openly available. The objective of the KDD 99 intrusion detection contest is to create a standard data set for survey and evaluate research in intrusion detection which is prepared and managed by MIT.

KDD Cup 99 it remains the standard research dataset best suited to benchmark performance and compares the effectiveness of various approaches to network intrusion detection (Fossaceca *et al*., 2011). KDD CUP 99 built based on the data captured in DARPA'98 IDS program.

DARPA'98 is about four gigabytes of compressed raw (binary) tcpdump data of seven weeks of network traffic which can be processed into about 5 mln. connection records, each with about 100 bytes. The two weeks of test data have around 2 mln. connection records. KDD training data set consists of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or an attack with exactly one specific attack type. The simulated attacks fall into one of the following four categories (Tavallaee *et al*., 2009).

Denial Service of attack (DoS): Is an attack in which the attacker makes some computing or memory resource too busy or full of handle legitimate requests or denies legal users access to a machine.

User to Root attack (U2R): is a class of exploit in which the attacker starts out with access to a normal user account on the system (perhaps gained by sniffing passwords, a dictionary attack or social engineering) and can exploit some vulnerability to gain root access to the system.

Remote to Local attack (R2L): It occurs when an attacker who can send packets to a machine over a network but who does not have an account on that machine exploits some vulnerability to gain local access to a user of that machine.

Probing attack: Is an attempt to gather information about a network of computers for the apparent purpose of circumventing its security control.

Researchers found some disadvantage with KDD99, Olusol's analysis the KDD 99 data set for selecting a relevant feature. They proposed that some features or attributes were not related to any attack they have taken 10% of the whole data set and performed their analysis.

Tavallaee *et al*. (2009) analysis the KDD CUP 99 they employed 21 learned machines to label the records of

entire KDD train and test sets which provides with 21 predicted labels for each record. They found about 98% of the records in the train set and 86% of the records in the test set were correctly classification with all the learners, so they achieve about 98-86% classification rate applying every simple machine learning methods because a minimum classification rate of 86% which makes the comparison of IDS quite difficult it since they all vary in the range of 86-100%. The important deficiency in the KDD data set is the huge number of redundant records, when analyzing KDD train and test sets they found that about 78 and 75% of the records are duplicated in the both of sets. This deficiency will cause to important disadvantages, learning algorithms to be biased towards the more frequent records and thus prevent it from learning unfrequent records which are usually more harmful to networks such as U2R attacks. In addition, cause evaluation results to be biased by the methods which have better detection rates on the frequent records talk about same problems in another way, especially pattern recognition and machine learning algorithms trained with the KDD training data subset and tested on the KDD testing data subset failed to detect the majority of U2R and R2L attacks within the context of misuse detection. Researchers provided a solution to solve the mentioned issues, resulting in new train and test sets which consist of selected records of the complete KDD data set. The provided data set does not suffer from any of the mentioned problems (Tavallaee *et al*., 2009). The number of records on the train and test sets is reasonable. With this, advantage makes it affordable to run the experiments on the complete set without the need randomly select a small portion. And the evaluation results of different research will be consistent and comparable. Even wth this advantage may not be a perfect representative of existing real network but they still believe it, however, can be applied as an effective benchmark data set to help researchers compare different intrusion-detection methods. The new version of KDD data set, NSL-KDD is publicly available for researchers through the website. NSL-KDD version which has been streamlined, removing many duplicate records and adjusting the distribution of records based on difficulty (Fossaceca *et al*., 2011). Our research is based on apply the new model on both and measure effectives and different.

IDS based artificial intelligence: Intrusion-Detection System (IDS) is the process of monitoring the events occurring in a computer system or network and analyzing them for signs of intrusion. It is useful not only in detecting successful intrusions but also in monitoring attempts to break security which provides important information for timely counter-measures. Basically, IDS can be classified into two types: Misuse intrusion detection and anomaly intrusion detection. Traditional protection techniques such as user authentication, data encryption, avoiding programming errors and firewalls are used as first lines of defense for computer security. If a weak password is compromised, user authentication cannot prevent the unauthorized use. Moreover, firewalls are vulnerable to errors in configuration and susceptible to ambiguous or undefined security policies.

Recently, the use of Artificial Intelligence (AI) techniques has been employed in different data mining and machine learning classification and prediction modeling schemes. In addition to these, hybrid data mining schemes, hierarchical hybrid intelligent system models and ensemble learning approaches that combine the base models with other hybrid machine learning paradigms to maximize the accuracy and minimize both root mean squared errors and computational complexity have also gained popularity in the literature AI techniques were used in many IDSs to perform these important tasks.The ability of soft computing techniques to deal with uncertain and partially true data makes them attractive to be applied in intrusion detection. There are many techniques such as Artificial Neural Network (ANN), Fuzzy logic association rules mining, Support Vector Machine (SVM), Genetic Algorithm (GA) etc. used to improve detection accuracy and efficiency of signature based IDS or anomaly detection based IDS (Patel *et al*., 2013) as shown in Table 1, some related reearches about IDS based AI. Another part of this research will propose the hybrid approaches based on these techniques and machine learning methods.

Table 1 shown some of the works, some of them provide IDS based hybrid of AI algorithms and another depends on the single algorithm. These research presentation IDS still suffering from some problems and challenges like the high rate in false alarms and showed how this IDS subject still needs to develop.

Genetic algorithm: Genetic Algorithms (GA) are inspired by Darwin's theory of biological evaluation as an optimization algorithm. The main idea of this technique is derived from natural evolution, so there are biologic operators such as crossover, mutation and selection. Genetic algorithm has three randomly generated phases: Initial population of chromosomes, crossover operator and mutation operator. Each chromosome represents some "solution" of the problem and its quality is done by the value of so-called object/fitness function. Genetic algorithm starts by generating of some random solutions which are marked as an initial population. In the second step, random crossovers lead to produce new offspring and in the third step with random value of mutation, a few

Table 1: Some related work on AI based IDS

Researchers	Description
Senthilnayaki and Venkatalakshmi	IDS proposed on genetic algorithm based feature selection approach and a support vector machine based classification algorithm
Richa Shivhare and Sushil Chaturved	Proposed framework consists of three parts, feature reduction by (ANN), clustering and classification and tested the proposed by KDD 99
Osama Alomari and Zulaliha Ali Othman	Proposed Bees algorithm as feature selection and using (SVM) as classifier and applied the model on KDD 99
Anand Kannan and Ayush Sharma	Intrusion detection model in which combine a proposed genetic based feature selection algorithm and an existing Fuzzy Support Vector Machines (SVM) for effective classification as a solution and applied on KDD 99
Laheeb Mohammad Ibrahim	Proposed IDS based on distributed time-delay artificial neural network
Wei Li	This research applied Genetic Algorithm (GA) to network Intrusion Detection Systems and focused on the network protocols
TCP/IP	
Jonatan Gomez	This research representation scheme for evolving fuzzy rules using the concept of complete binary tree structures. also genetic operators such as gene addition
Mehdi Moradi and Mohammad Zulkernine	This research presents a neural network approach to intrusion detection. A Multi-Layer Perception (MLP), While most of the previous studies tried to classified between normal and attack this work tried to give the type of attack also
Piyakul Tillapart and Pratit Santiprabhob James Cannady	In this research, Fuzzy rule-based system has been introduced to implement IDS in this framework This research proposed misuse detection based artificial neural networks provide the potential to identify and classify network activity based on limited, incomplete and nonlinear data sources

of genes in chromosome are changed or replaced. The new generation of candidate solutions is then used in the next iteration of the algorithm. GA is a computing technique used as an exhaustive search paradigm to find exact or approximate solutions to optimization problems. GA is categorized as global search heuristics. Its paradigm is based on a particular class of evolutionary algorithms that uses techniques inspired by evolutionary biology such as inheritance, mutation, selection and crossover. GA is implemented in a computer simulation framework in which a population of abstract representations (representing chromosomes) of candidate solutions (representing biological creatures or phenotypes) to an optimization problem produces better solutions. Traditionally, solutions are represented in bits (a set of 0s and 1s) but other encoding are also possible.

GA is the most commonly known evolutionary algorithm for optimization which uses techniques inspired by natural evolution to allow a population of solutions (also called, individuals) candidates to solve the optimization problem to evolve i.e., move toward the best solution. The evolution usually starts from a population of randomly generated individuals which change at each iteration, called a generation. In each generation, the fitness of every individual in the population is evaluated and the most fitting individuals are selected. Everyone is modified by mating and a new, more evolved generation of candidate solutions is formed. Many researchers use GA as the automatic generation algorithm for optimization to find the best solution as shown in Table 2 some related research about this as following.

Traditional algorithms facing some limitations and problems like random select of parameters and generate randomly, it is still an open problem (Castano *et al*., 2016). Table 2 shown some works that is depended on hybrid models based GA is a global optimal algorithm. In fact, GA is an adaptive and exploratory search and research algorithm based on evolutionary ideas of natural genetics. GA generates the population of primary individuals in high quality of individuals. On the other hand, each one of these individuals represents a solution for the problem (Aslahi-Shahri *et al*., 2016).

An issue with ELM is that as the hidden node learning parameters in ELM are randomly assigned and they remain unchanged during the training procedure, the classification boundary may not be an optimal one (Silva *et al*., 2011; Cao *et al*., 2012a, b). The original ELM does not provide an effective solution for architecture selection of SLFNs, although in the framework of model selection of neural networks, there are many approaches have been proposed there is no a universal method so far to determine the optimal architecture of SLFNs, i.e., the optimal number of hidden nodes. The number of hidden nodes of SLFNs is usually pre-determined by users who use their experience to determine this number which may be far away from the optimal number of hidden nodes of SLFNs, in other words, it may not be a suitable number of hidden nodes of SLFNs (Zhai *et al*., 2016). So this research proposes GA to optimize ELM by choose the best classifier that ELM can produce.

Table 2: Some related work applied GA to optimization to find the best solution

Researchers	Brief description
Compare and Martini	This study proposes and compares different techniques for maintenance optimization based on Genetic Algorithms
SaisoKarakatic and Vili Podgorelec	This study presents a survey of genetic algorithms that are designed for solving multi depot vehicle routing problem. Genetic algorithm based solutions are compared with other existing approaches, both exact and heuristic, for solving this same problem
Ruben Aguilar-Rivera and Manuel Valenzuela-Rendon	This study presents a review of the application of evolutionary computation methods to solving financial problems. Genetic algorithms, Although, genetic algorithms have remained the most popular approach in the literature and The schemata theorem shows how the best solutions pass
Saulo Oliveira	Seam carving is a method which aims at retargeting images, adjusting input images into arbitrary dimensions based on GAs are optimization tools Due to the underlying features of GAs, some optimization problems can be solved
Thiago Nogueira and Iverson Farias Costa	A method to efficiently perform the selection of multiple contingencies is presented. The issue is modeled as a combinatorial optimization problem and solved by genetic algorithms
Mohammad Ammar and Amr Elhadidy	Markov-chain models are used for predicting the performance of road pavement and to calculate the expected decline at different periods of time. A genetic-algorithm-based procedure is developed for solving the multi-objective optimization problem
Santhanam and Padmavathi	Medical data mining is one major research area where evolutionary algorithms and clustering algorithms play a vital role. In this research work, K-Means is used for removing the noisy data and genetic algorithms for finding the optimal set of features with Support Vector Machine (SVM) as classifier
Spranger and Capelli	In this study, they perform a comparative analysis between two computational methods for virtual stent deployment: a novel fast virtual stinting method, which is based on a spring–mass model, is compared with detailed finite element analysis in a sequence of in silicon experiments. They present a way to optimize the fast method by calibrating a set of parameters with the help of a genetic algorithm
Hung *et al.* (2013)	Proposed in this study. Calculations of explicit equations were performed so as to assisting in measuring the aesthetic characteristics; next, a fuzzy judgment was invoked to calculate the perceptual aesthetic measures of a product style so as to establish the overall aesthetic standard for the product. Aesthetic measurement principles were combined with the Genetic Algorithm (GA) and applied to the optimization of the product's shape
Maryam Sarkhosh and Riccardo Leardi	In this study, they investigate the effect of pixel selection by application of Genetic Algorithms (GAs) for PLS model. Gas is very useful in the variable selection in modeling and calibration because of the strong effect of the relationship between presence/absence of variables in a calibration model and the prediction ability of the model itself

RESULTS AND DISCUSSION

IDS based machine learning: Many programs make it easy for an attacker to create a plethora of malware instances that cannot be matched by classical signature-based anti-virus products. Thus, it is important to find other means of classifying an unknown malware sample. Rieck presents a system that uses the behavioral information contained in the analysis reports. First, behavioral profiles of each know malware family is extracted. Second, machine-learning techniques are applied to derive classifiers from these profiles for grouping malware instances that share similar behavior.

To support this, the others run a commercial anti-virus scanner on large body of collected malware samples to obtain labels for the individual samples. For the samples that could be identified these labels correspond to the families the samples belong to. Based on these labels and the behavioral profiles extracted from the reports, the authors trained the machine learning algorithms like Support Vector Machines (SVM) to build classifiers for the individual families. A given SVM only states the probability at the analyzed sample belonging to a family. The system may be confronted with samples that do not belong to any known families or exhibit behaviors that are characteristic for multiple families (Egele *et al.*, 2012). Depending on this concept of machine learning, there are numerous recent articles that test the application of machine learning techniques for intrusion detection as mentioned them in Table 3.

The researches shown in Table 3 include hybrid and single algorithms to work based IDS but they still suffering from some limitations. Machine learning methods like SVM, ANN have widely used for IDS but these methods generally suffer from some points very important for IDS, like the long training time, require parameter training or do not perform well in multi-class classification.

Jaiganesh and Sumathi (2012) another overview of potential techniques used with a NIDS is provided. The study not only covers Neural Networks and SVMs but also suggests that ELMs would be useful for a NIDS based on the fact that they are easy to implement have fast learning speed, high generalization ability and research well with on-linear activation functions and kernels. Although, the researchers suggest that ELMs may be useful in overcoming many of the challenges discussed earlier in Patel *et al.* (2013), they do not go into details of prior research on ELMs with NIDS nor do they discus show an ELM could be applied the NID problem. They do suggest that disadvantages of individual algorithms can be overcome by combining different learning approaches in a hybrid manner to take the advantages of algorithms.

Table 3: Some related work on IDS based on machine learning

Researchers	Brief description
Salma Elhag and Fernandez	Consider the use of genetic fuzzy systems within a pairwise learning framework for the development of such a Alberto system
Datta H. Deshmukh	IDS based a hybrid of decision tree classifiers and Naïve Bayes classifiers and applied on NSL-KDD
Varuna and Natesan Jojn	IDS Based Hybrid Learning method(K-means clustering and Naïve Bayes classifiers)
M.Fossaceca and Thomas A. Mazzuchi	Design of framework to improve the efficacy of Anomaly IDS based on ELM
Bahareh Gholipour	IDS based hybrid SVM and ABC algorithms and data analysis was undertaken using KDD CUP 99
Datta H.Deshmukh	IDS based naive bayes classifier is used in supervised learning method which classifies various network events for the KDD cup 99 dataset
Chi Cheng and Wee Peng Tay	Used ELM to classify and detect the intrusions
Shi-Jinn Horng	This study proposed an SVM-based intrusion detection system which combines a hierarchical clustering, KDD CUP 99 used to evaluate the proposed system
Yinhui Li	Proposed IDS based on combination of clustering method, ant Colony algorithm and SVM to classification Data normal or not
Phurivit Sangkatsnee	Developed a real-time IDS using the Decision Tree technique to classify On-line network data as normal attack data
Latifur Khan and Mamoun Awad	They used (SVM) for classification and Dynamically Growing Self Organizing Tree (DGSOT) for clustering based IDS
TaeshikShon and Jongsub Moon	IDS based on used Enhanced SVM that combines unsupervised And supervised) SVM and used GA for feature selection
Mahesh Kumar Sabhna and Gursel Serpen	Proposed IDS based on subset of machine learning algorithm(multilayer perceptron, K-means, Gaussian) and applied on KDD 99
Yihua Liao and Rao	Design IDS approach based on the K-Nearest Neighbor(KNN) Classifier, it is used to classify programs behavior as normal or intrusive

Extreme learning machine: IDS need to be adaptable with all a new nature of attacks and ELM has a version Online Sequential Extreme Learning Machine (OSELM) enables the system to learn in a progressive way and this fit the incremental nature of data in cloud environment (Liang *et al*., 2006). IDS strategy depends on the speed and accuracy on detect, ELM if compare with the SVM, ELM runs faster than the SVM while accuracy is similar.

Huang *et al*. (2012) it has been investigated that the performance of ELM is not sensitive to the number of hidden neurons as long as it is large enough and 1000 hidden neurons are good enough for many real-world applications. Some researchers talk briefly about the disadvantages of other machine learning methods (Patel *et al*., 2013; Fossaceca *et al*., 2011):

Bayes network classifier: Utilizes feature probability distributions to estimate conditional probabilities that observations belong to specific classes. Although, this technique works well with large datasets, the classes are assumed to be independent and it is difficult to estimate the actual probabilities of the network traffic

K Nearest Neighbors (KNN): Each sample (i.e., observation) is classified based on a majority vote of its "nearest neighbors" determined by measure of similarity. Although this technique is fairly easy to implement, KNN requires a significant amount of storage, tends to perform poorly on datasets with many attributes (highly dimensional) and researches slowly on test data.

Decision tree: Uses a recursive approach to learning to divide data into specific classes however decision tree algorithms can be unstable or very complex

Neural networks: Based on modeling how neurons operate in the human brain, single or multiple layer "perceptrons" are trained with a gradient descent algorithm such as Back Propagation to optimize neuron weights in order to minimize error between actual and predicted training samples. However, Neural networks have a large computational burden are prone to over-fitting and can have long processing times

Support Vector Machines (SVMs): Employ statistical optimization techniques to construct a set of "hyper planes" in high dimensional space to classify traffic into proper categories. SVMs require careful selection of kernel type and adjustment of several parameters. SVMs are very accurate and are capable of modeling complex decision boundaries with fewer propensities for over-fitting data than other approaches. However, the SVM algorithm is highly complex with large memory requirements. Kernel selection can be difficult and the algorithm tends to run slowly on larger datasets. ELMs have recently been applied to the network intrusion detection problem in several studies as Table 4 shown some of these studies:

Table 4 shows many of the studies tried to apply single or hybrid ELM as IDS because the advantages of ELM but even with all the ELM advantages it's still, ELM random determination of the input weights and hidden

Table 4: ELMs based IDS

Resaechers	Brief description
Fossaceca *et al.* (2011)	Chosen the Extreme Learning Machine (ELM) as the core learning algorithm based IDS, due to recent research suggests that ELMs are straightforward to implement, computationally efficient and have excellent learning performance characteristics
Singh *et al.* (2015)	In this study used a technique based on the online Sequential (ELM) is presented for intrusion detection. The Harish proposed technique user alpha profiling to reduce the time complexity while irrelevant features are discarded using an ensemble of filtered, correlation and consistency based feature selection techniques.
Xue *et al.* (2014)	They proposed the use of Extreme Learning Machine (ELM) for detecting network intrusion attempts
Jaiganesh and Sumathi (2012)	Overview of intrusion detection systems and introduces the reader to some fundamental concepts of IDS methodology, Extreme Learning Machine (ELM) is a new emergent technology which provides good generalization performance for both classification and regression problems at highly fast learning speed
Chi Cheng and Wee PengTay	The risk of information systems to external attacks or intrusions has increased tremendously. They consider the problem of intrusion detection in computer network, and investigate the use of Extreme Learning Machines (ELMs) to classify and detect the intrusions. With increasing connectivity between networks
Jaiganesh and Sumathi (2012)	Extreme Learning Machine (ELM) has been extensively applied to provide potential solutions forthe IDS problem But, the practicability of ELM is affected because of the complexity in choosing the suitable ELM Parameters and to provide the parameters by using Levenberg-Marquardt (LM)
Gilles Paiva	This research focuses on the use of ELM (Extreme Learning Machine) and OS-ELM (Online Sequential ELM) techniques applied to IDSs. Some features of these methods that motivate their use for building IDSs are: easy assignment of parameters; good generalization; and fast and online training.

biases may lead to non-optimal performance and it might suffer from the overfitting as the learning model will approximate all training samples well (Silva *et al*., 2011; Jaiganesh and Sumathi, 2012). This problem will cause a high rate of false positives which is a big challenge for network operators (Fossaceca *et al*., 2011). It also lowers the predicting accuracy (Wang *et al*., 2011, 2009). Bartlett (1998) pointed out that for feedforward neural networks the smaller the norm of weights and training error are the better generalization performance the networks tend to have. Therefore, it is necessary to develop a more effective learning method that can overcome this shortcoming and approximate as fast as the ELM algorithm. So, the determination of the optimal number of basic functions to be included in the hidden layer is still an open problem (Castano *et al*., 2016).

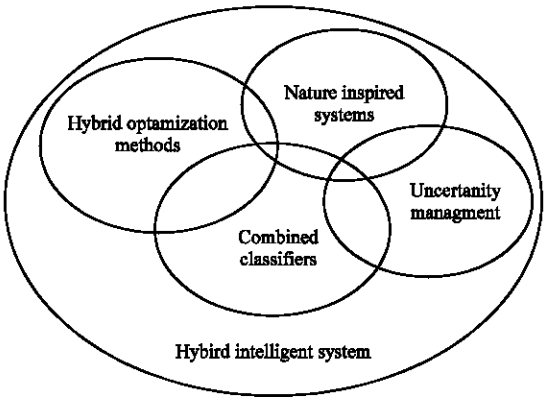

Fig. 2: Domains of hybrid intelligent systems (Wozniak *et al*., 2014)

IDS based hybrid system: Hybrid intelligent systems offer many alternatives for unorthodox handling of realistic increasingly complex problems, involving ambiguity, uncertainty and high-dimensionality of data. They allow using both a priori knowledge and raw data to compose innovative solutions. Therefore, there is growing attention to this multidisciplinary research field in the computer engineering research community (Wozniak *et al*., 2014). Figuer 2 is a rough representation of the computational domains covered by the hybrid intelligent system approaches

In the development of IDS, the ultimate goal is to achieve the best possible accuracy for the task at hand. This objective naturally leads to the design of hybrid approaches for the problem to be solved. The idea behind a hybrid classifier is to combine several machine learning and Artificial intelligence techniques so that the system performance can be significantly improved. More specifically, a hybrid approach typically consists of two functional components. The first one takes raw data as input and generates intermediate results. The second one will then take the intermediate results as the input and produce the final results (Jang and Mizutani, 1996). Finally, hybrid classifiers can also be based on the integration of two different techniques in which the first one aims at optimizing the learning performance (i.e., parameter tuning) of the second model for prediction (Wozniak *et al*., 2014). Some researchers divided the hybrid systems. As there are three strategies to design hybrid classifiers, used in the intrusion detection domain for the review including single, hybrid and ensemble classifiers as Fig. 3.

Table 5 shown that the researchers focusing on the hybrid system as a new and active idea by integrate two algorithms or more but the IDS using machine learning and artificial intelligence techniques still need to be

Table 5: Shown some related work about Hybrid systems based IDS

Researchers	Brief description
Lorena Cazorla and Lopez	Carry out an extensive study to determine the requirements imposed by the Control System (CS) on the IDS Javier solutions using the Non-Functional Requirements (NFR) Framework
Rahmani and Chizari	A hybrid method of support vector machine and Genetic Algorithm (GA) is proposed. And its implementation in intrusion detection problem
Rajpal and Rohini Kaur	Proposed an algorithm based on combination of misuse detection and genetic algorithm approach based IDS
Swati Sharma and Santosh	GA and fuzzy logic based approaches are used for detecting network intrusions
Kumar Parveen Kumar and Nitin Gupta	Presented an intrusion detection model based on genetic algorithm and neural network
Enache and Patriciu	Propose an IDS model based on Information Gain for feature selection combined with the SVM classifier. The parameters for SVM will be selected by a swarm intelligence algorithm (Particle Swarm Optimization or Artificial Bee Colony)
Changning Ca and Guojian Cheng	This study proposes a new network anomaly detection method in order to deal with the low detection rate and high false alarm rate problem. Based Ball Vector Machine (BVM) and Extreme Learning Machine (ELM)
Gilles Paiva	ELM (Extreme Learning Machine) and OS-ELM (Online Sequential ELM) techniques applied to IDSs
Jaiganesh and Sumathi (2012)	Proposed IDS based sigma () of the radial basis kernel function is tuned using Levenberg-Marquardt (LM) learning and proposed kernelized Extreme Learning Machine with LM
Gang Wang and Jinxing Hao	Based on ANN and fuzzy clustering, to solve the problem and help IDS achieve higher detection rate, less false positive rate and stronger stability
Tanselozyer and RedaAlhajj	Propose a method based on iterative rule learning using a fuzzy rule-based genetic classifier, to work as an intelligent Intrusion Detection System (IIDS)
Dong Seong Kim and Ha-Nam Nguyen	Proposed Genetic Algorithm (GA) to improve Support Vector Machines (SVM) based intrusion detection system (IDS)
Sandhya Peddabachigari and Ajith Abraham	Intrusion detection with Decision trees and SVM were tested with benchmark 1998 DARPA Intrusion Detection data set

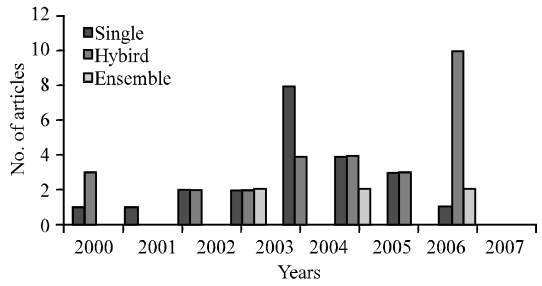

Fig. 3: Yearwise distribution of studies for the types of classifier design (Tsai *et al*., 2009)

researched, even that if many researchers proposed it and in Table 5 shown some related research about hybrid systems based IDS.

Table 5 shown how researchers focusing on how hybrid approaches had been proposed to enhance the capabilities of current IDS by combining the two methods or algorithms and shown how the different of traditional algorithm's results and the enhance of the results after the hybrid them. In (Xue *et al*., 2014) ELM is a tuning-free algorithm with an extreme fast learning speed by randomly generating the input weights and the hidden bias. However, due to the random determination of learning parameters, some optimal input weights and hidden biases may be generated which may have negative impact on its generalization ability and performance stability. To alleviate such Weakness, some modifications have been proposed, a hybrid model is proposed to alleviate such weaknesses of ELM. The model adopted Genetic Algorithms (GAs) to produce a group of candidate networks first and according to specific ranking strategy, some of the networks are selected to ensemble a new network. To verify the performance of our method, called the Genetic Ensemble of extreme Learning Machine (GE-ELM). In the researchers mention (GE-ELM) can make networks more robust and generate better generalization performance. Alexandre *et al*. (2015) created a hybrid model to improving intelligent transportation systems. The hybridized was between the GA and ELM. The ELM is greats generalization performance at a very fast learning speed. The ELM plays the key role of providing the fitness of candidate solutions in each generation of the GA. The proposed method helps the ELM based classifier to increase its performance from a mean probability of correct classification of 74.83% (with no feature selection) up to 93.74% (when using the optimum subset of selected features). Since that the huge difference in the results between classification with feature selection or without, this research proposed the hybrid model to work based IDS.

Current work: Extreme Learning Machines (ELMs), first introduced by Huang have been widely studied by researchers and applied to a variety of applications (Ding *et al*., 2014). In anticipation of achieving our goal to apply a machine learning approach to the large scale problem of network intrusion detection, we selected a computationally efficient and scalable base algorithm that we adapted to handle large datasets. Due to its reduced computational complexity, ease of implementation, scalability and good learning performance with respect to other learning algorithms we chose the ELM as the basis

for this research (Huang *et al*., 2012).As mentioned earlier, one of the motivations for using Elms instead of SVMs is that ELMs deliver comparable classification accuracy rates but have lower computational requirements than SVMs (Huang *et al*., 2012). Most journal study comparing SVMs to ELMs conclude that the ELM significantly outperforms the SVM in terms of computational speed and with learning (classification) performance that is on par with SVMs (Chorowski *et al*., 2014).

The basic ELM randomizes the values of the parameters in the hidden layer of an SLFM prior to calculation ofthe Moore-Penrose inverse to obtain a suitable solution for assigning each input data sample to the proper target class (Man *et al*., 2012). Since the parameters are not adjusted during the training process the overall computed classification boundary may not be optimal. However, since each time an ELM is executed it generates random parameters for the SLFN hidden layer, the classification decision boundary will vary for each trial. Cao *et al*. (2012a, b) recently introduced the Voting ELM or VELM that takes advantage of this diversity by applying a voting rule to an accumulator function that tracks the number of times a particular sample is assigned to a particular class over K trials.

The VELM uses a majority voting technique to assign each sample to a particular class. Results of Cao's experiments with VELM demonstrate that theVELM outperforms several machine learning approaches including the basic ELM, SVM, BPNN and several other techniques. Later research resulted in several variants of VELM with improved performance including a Minimum Square Error (MSE) weighted majority voting approach. Although the VELM results are promising, the improved performance comes at a cost as the basic ELM must be executed for several successive decision rounds and this can be very time-consuming.

Huang *et al*. (2012) and Fu *et al*. (2016), they used kernel-based ELM and in some cases yields very good results, kernel selection is critical for achieving good learning performance. But the Kernel-based ELM algorithm computes a kernel over the entire set of input samples and can consume a lot of memory (Fossaceca *et al*., 2011). For the large datasets computation of a full kernel is often not feasible due to memory constraints and smaller datasets that do lend themselves to full kernel computation an approach for combining multiple kernels or multiple classifiers is sometimes required in order to achieve good results.

Fossaceca *et al*. (2011) tried to explore the efficacy of combining the learning decisions of multiple classifiers to formulate a single decision that is more accurate than any of the individual classifiers. The motivation for using an ensemble of classifiers is that prior research demonstrated that individual classifiers have varied ability to detect specific classes in a multi-class learning problem. By introduced the novel Multiple Adaptive Reduced Kernel Extreme Learning Machine (MARK-ELM) to research based IDS,MARK-ELM suitable for processing multi-class NIDS. Approaches have displayed good detection performance for some classes of attack but poor performance for others because the it's depends on unbalance dataset (KDD99). The proposed approach showed good detection performance with a high rate of false positives which is huge challenge for network operators. Also the author during testing mode didn't depend on the data set testing mode to evaluate the results.

Singh *et al*. (2015) research explained that the common challenges for IDSs are large amounts of data to process, low detection rates and high rates of false alarms. They used the online sequential extreme learning machine to design the IDS based anomaly by analyzing the network traffic. For performance evaluation proposed technique the standard NSL-KDD and Kyoto university benchmark datasets are used to test the proposed IDS. Both datasets features that used in this work extracted from KDD cup 99 data set. So the algorithm has not been validated on large data set such as KDD 99 and further validation has to be performed.Concept of data profiling that used in this work is heuristic and not model based, heuristic is vague and general because it's created by the designer based on observed patterns and for example, used the heuristics in detection when you might not be sure, a virus is there but you can look for specific key attributes of a virus. Heuristic is an approach to problem solving, learning or discovery that employs a practical method not guaranteed to be optimal or perfect (Aslahi-Shahri *et al*., 2015).

CONCLUSION

This research is aimed at presenting an anomaly detection technique based on GA and SVM they used GA and SVM in order to improve the performance of classification for SVM and evaluation Proposed technique just based KDD CUP 99. As mentioned the limitation in SVM as it provides the binary classification as normal data or attack and also the ELM is faster than it. In additionally, evaluated the system in this research only by KDD99 data set and with disadvantage mentioned in the KDD99 the result was not a good accurate.

REFERENCES

Alexandre, E., L. Cuadra, S. Salcedo-Sanz, A. Pastor-Sanchez and C. Casanova-Mateo, 2015. Hybridizing extreme learning machines and genetic algorithms to select acoustic features in vehicle classification applications. Neurocomputing, 152: 58-68.

Alharkan, T. and P. Martin, 2012. IDSaaS: Intrusion detection system as a service in public clouds. Proceedings of the 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (ccgrid 2012), May 13-16, 2012, IEEE, New York, USA., ISBN:978-0-7695-4691-9, pp: 686-687.

Aslahi-Shahri, B.M., R. Rahmani, M. Chizari, A. Maralani and M. Eslami *et al*., 2016. A hybrid method consisting of GA and SVM for intrusion detection system. Neural Comput. Appl., 27: 1669-1676.

Bartlett, P.L., 1998. The sample complexity of pattern classification with neural networks: The size of the weights is more important than the size of the network. IEEE. Transac. Inf. Theor., 44: 525-536.

Cao, J., Z. Lin, G.B. Huang and N. Liu, 2012a. Voting based extreme learning machine. Inf. Sci., 185: 66-77.

Cao, J.J., S. Kwong, R. Wang and K. Li, 2012b. A weighted voting method using minimum square error based on Extreme Learning Machine. Proceedings of the International Conference on Machine Learning and Cybernetics (ICMLC), July 15-17, 2012, IEEE, New York, USA., ISBN:978-1-4673-1484-8, pp: 411-414.

Castano, A., F.F. Navarro, A. Riccardi and C.H. Martinez, 2016. Enforcement of the principal component analysis-extreme learning machine algorithm by linear discriminant analysis. Neural Comput. Appl., 27: 1749-1760.

Choo, K.K.R., 2011. The cyber threat landscape: Challenges and future research directions. Comput. Secur., 30: 719-731.

Chorowski, J., J. Wang and J.M. Zurada, 2014. Review and performance comparison of SVM-and ELM-based classifiers. Neurocomputing, 128: 507-516.

Ding, S., X. Xu and R. Nie, 2014. Extreme learning machine and its applications. Neural Comput. Appl., 25: 549-556.

Egele, M., T. Scholte, E. Kirda and C. Kruegel, 2012. A survey on automated dynamic malware-analysis techniques and tools. ACM Comput. Surveys, Vol. 44. 10.1145/2089125.2089126

Fossaceca, J.M., T.A. Mazzuchi and S. Sarkani, 2011. MARK-ELM: Application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection. Expert Syst. Appl., 42: 4062-4080.

Fu, H., C.M. Vong, P.K. Wong and Z. Yang, 2016. Fast detection of impact location using kernel extreme learning machine. Neural Comput. Appl., 27: 121-130.

Huang, G.B., H. Zhou, X. Ding and R. Zhang, 2012. Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B (Cybern.), 42: 513-529.

Huang, W., N. Li, Z. Lin, G.B. Huang and W. Zong *et al*., 2013. Liver tumor detection and segmentation using kernel-based extreme learning machine. Proceedings of the 35th Annual International Conference on Engineering in Medicine and Biology Society (EMBC), July 3-7, 2013, IEEE, New York, USA., ISBN:978-1-4577-0215-0, pp: 3662-3665.

Jaiganesh, V. and P. Sumathi, 2012. Kernelized extreme learning machine with levenberg-marquardt learning approach towards intrusion detection. Intl. J. Comput. Appl., 54: 38-44.

Jang, J.S.R. and E. Mizutani, 1996. Levenberg-Marquardt method for ANFIS learning. Proceedings of the Conference on North American Fuzzy Information Processing Society, June 19-22, 1996, IEEE, New York, USA., ISBN:0-7803-3225-3, pp: 87-91.

Lee, W. and S. Stolfo, 2000. A framework for constructing features and models for intrusion detection systems. ACM Trans. Inform. Syst. Secur. J., 3: 227-261.

Liang, N.Y., G.B. Huang, P. Saratchandran and N. Sundararajan, 2006. A fast and accurate online sequential learning algorithm for feedforward networks. IEEE. Trans. Neural Netw., 17: 1411-1423.

Man, Z., K. Lee, D. Wang, Z. Cao and S. Khoo, 2012. Robust single-hidden layer feedforward network-based pattern classifier. IEEE. Trans. Neural Netw. Learn. Syst., 23: 1974-1986.

Patel, A., M. Taghavi, K. Bakhtiyari and J.C. Junior, 2013. An intrusion detection and prevention system in cloud computing: A systematic review. J. Netw. Comput. Appl., 36: 25-41.

Silva, D.N., L.D. Pacifico and T.B. Ludermir, 2011. An evolutionary extreme learning machine based on group search optimization. Proceedings of the IEEE Congress on Evolutionary Computation (CEC), June 5-8, 2011, IEEE, New York, USA., ISBN:978-1-4244-7834-7, pp: 574-580.

Singh, R., H. Kumar and R.K. Singla, 2015. An intrusion detection system using network traffic profiling and online sequential extreme learning machine. Expert Syst. Appl., 42: 8609-8624.

Tavallaee, M., E. Bagheri, W. Lu and A.A. Ghorbani, 2009. A detailed analysis of the KDD CUP 99 data set. Proceedings of the 2nd IEEE Symposium on Computational Intelligence for Security and Defence Applications, July 8-10, 2009, IEEE, Ottawa, Ontario, pp: 1-7.

Tsai, C.F., Y.F. Hsu, C.Y. Lin and W.Y. Lin, 2009. Intrusion detection by machine learning: A review. Exp. Syst. Appl., 36: 11994-12000.

Wang, Q., C. Wang, J. Li, K. Ren and W. Lou, 2009. Enabling Public Verifiability and Data Dynamics for Storage Security in Cloud Computing. In: Computer Security. Backes, M. and P. Ning (Eds.). Springer Berlin Heidelberg, Berlin, Germany, ISBN: 978-3-642-04443-4, pp: 355-370.

Wang, Y., F. Cao and Y. Yuan, 2011. A study on effectiveness of extreme learning machine. Neurocomputing, 74: 2483-2490.

Wozniak, M., M. Grana and E. Corchado, 2014. A survey of multiple classifier systems as hybrid systems. Inf. Fusion, 16: 3-17.

Xue, X., M. Yao, Z. Wu and J. Yang, 2014. Genetic ensemble of extreme learning machine. Neurocomputing, 129: 175-184.

Zhai, J., Q. Shao and X. Wang, 2016. Architecture selection of ELM networks based on sensitivity of hidden nodes. Neural Process. Lett., 44: 471-489.