VOL. 11, NO. 1, JANUARY 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

REVIEW ON HYBRID EXTREME LEARNING MACHINE AND GENETIC ALGORITHM TO WORK AS INTRUSION DETECTION SYSTEM IN CLOUD COMPUTING

Mohammed Hasan Ali and Mohamad Fadli Zolkipli
Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang, Pahang, Malaysia
E-Mail: Mh180250@gmail.com

ABSTRACT

Today security is a major challenge, many tools provided in this issue of instant firewall and intrusion-detection system (IDS) .IDS is one of the powerful tools in the security. IDS works depend on the fastest to detect and accuracy of detection. In other hand the IDS facing problem with high false alarm rate. This work proposes to solve this problem by hybrid between the Extreme Learning Machine (ELM) and Genetic Algorithm (GA). ELM work depends on two parameters weight (W) and biases (B) that will provide by GA. ELM has set of properties that make it attractive to be adopted for intrusion detection system in cloud environment. our work approach and integrate GA ELM work as IDS with high hopes detection rate and accuracy to the second problem and suggest dividing the training mode for virtual training and virtual testing to ensure selecting a best classifier.

Keywords: security, extreme learning machine, genetic algorithm, cloud computing.

INTRODUCTION

Background

The idea of providing a centralized computing service dates back to the 1960s, when computing services were provided over a network using mainframe time-sharing technology. In 1966, Canadian engineer Douglass Parkhill published his book The Challenge of the Computer Utility(Parkhill, 1966), in which he describes the idea of computing as a public utility with a centralized computing facility to which many remote users connect over networks.

In the 1960s, the mainframe time-sharing mechanism effectively utilized computing resources, and provided acceptable performance to users; however, mainframes were difficult to scale and provision up-front because of increasingly high hardware costs. Accordingly, users didn't have full control over the performance of mainframe applications because it depended on how many users utilized the mainframe at a given moment. As such, with the introduction of personal computers users loved the idea of having full control of their computing resources, even though these resources are not as effectively utilized.

For many enterprises, the long-standing dream has been to background information technology issues and concentrate on core business instead. Although the effect of the cloud computing adoption is yet to be seen, many companies believe that cloud computing may offer feasible alternative model that may reduce costs and complexity while increasing operational efficiency. Today Small and Medium Business (SMB) companies are increasingly realizing that simply by tapping into the cloud they can gain fast access to best business applications or drastically boost their infrastructure resources, all at negligible cost. Gartner (Hunter, De Lotto *et al*. 2009) defines cloud computing ((Stanojevi and Shorten, 2008); (Koehler, Anandasivam *et al*. 2010); (Weiss, 2007); (Hibbett, Binder *et al*. 2007); (Whyman, 2008); as ''a style of computing where massively scalable IT- enabled capabilities are delivered 'as a service' to external customers using Internet technologies''. Cloud providers currently enjoy a profound opportunity in the marketplace. The cloud offers several benefits (Patel, Taghavi *et al.* 2013) like fast deployment, pay-for- use, lower costs, scalability, rapid provisioning, rapid elasticity, ubiquitous network access, greater resiliency, hypervisor protection against network attacks, low-cost disaster recovery and data storage solutions, on demand security controls, real time detection of system tampering and rapid reconstitution of services. Figure-1 illustrates the cloud computing

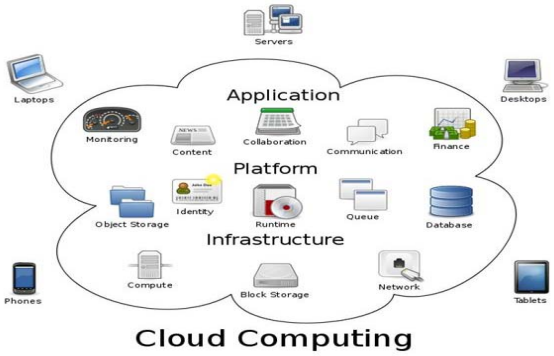

Figure-1. Cloud computing(Shelke, Sontakke *et al*. 2012).

Cloud computing moves the application software and databases to the large data centers, where the management of the data and services are not trustworthy. This unique attribute, however, poses many new security challenges (Wang, Wang *et al.* 2009). These challenges in clude but not limited to accessibility vulnerabilities, virtualization vulnerabilities, web application vulnerabilities such as SQL(Structured Query Language) injection and cross-site scripting, physical access issues, privacy and control issues arising from third parties having physical control of data, issues related to identity and credential management, issues related to data verification, tampering, integrity, confidentiality, data loss and theft, issues related to authentication of the respondent device or devices and IPspoofing.

All these systems need to be defended against a number of threats. Amateur hackers, rival corporations, terrorists and even foreign governments 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, Internet, Web application, cloud computing along with numerous occurrences of international terrorism, raises 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.

Cloud computing is increasingly popular even though unresolved security and privacy issues are slowing down their adoption and success ,indeed, integrity, confidentiality, and availability concerns are still open problems that call for effective and efficient solutions (Lombardi and Di Pietro 2011). There are various different measures to improve the security of network of either distributed orstand alone systems are already known to be researched, proposed, and deployed. Some of them measure include firewalls, network intrusion detection, and prevention systems.

The concept of intrusion detection was known since past and was first proposed by a well-known researcher named Anderson in 1980s (Anderson 1980). Since the concept was very effective in increasing the security measures of end systems, it gets popularized and now is used and applied in various domains of security.

ID is the identification of attempted or ongoing attacks on network or computer system. ID issues in cloud data collection, data reduction, reporting and response, behavior classification. Researchers used many techniques and modify them to work as IDS like artificial intelligence (AI), Data mining, Machine learning, Signature analysis, etc.

This research focus on hybrid between the classifier, Extreme Learning Machine (ELM) that was created by (Huang, Zhu *et al.* 2004) is one of the faster algorithms of classification. And Genetic Algorithm (GA) is one of the famous algorithms of (AI). (GA) That was John Holland who invented it in the early 1970s. This work tries to integrate between the advantages of (ELM) and (GA) to make our model to works as IDS.

STATEMENT OF PROBLEM

Extreme Learning Machine has set of properties that make it (Cheng, Tay *et al.*, 2012; Jaiganesh, Mangayarkarasi *et al.* 2013) attractive to be adopted for intrusion detection and prevention system in cloud environment.

Firstly, this classification algorithm (Uçar, Demir *et al.* 2014) (Huang, Zhu *et al.* 2006) has an online version named online sequential extreme learning machine OSELM. This online version enables the system to learn in a progressive way. This fits the incremental nature of data in cloud environment. In addition, this fits the needs of intrusion detection and prevention to be adaptable with all new nature of attacks that the attacker finds.

Secondly, (Fossaceca, Mazzuchi *et al.* 2015) this approach has proven itself to be faster than the common support vector machine SVM. Thirdly, this approach behaves well in multi-class problems in which intrusion detection and prevention system is regarded as one of them (Wei, Li *et al.* 2006). This is why many researchers have recommended ELM based approaches for this application.

Despite the variety of approaches that were built on ELM, researchers have agreed that ELM does not provide the optimal solution in the process of making the classification decision (Cao, Lin *et al.* 2012, Fossaceca, Mazzuchi *et al.* 2015). More specifically, Extreme learning machine has a random nature in the step of initiating the hidden layer neurons (Huang, Zhu *et al*. 2006).This creates infinite number of degrees of freedom, and as a result. It causes range of diversity in the decision making if two different chance seeds were initiated in two Extreme Learning Classifiers with the same exact structure. Some researchers have tried to exploit this fact by creating an ensemble of classifiers and developing a voting strategy under the concept decision fusion (Fossaceca, Mazzuchi *et al*. 2015). In our perspective, ensemble of totally arbitrary ELM classifiers will not perform well unless an assumption of optimal consensus is made. This research asserts that this assumption is not proven theoretically nor met practically.

In addition, many researchers use arbitrary strategies in the way of data partitioning into training and testing. Data unbalance is also a problematic nature through the application of intrusion detection and prevention system (Portnoy, Eskin *et al.* 2001) due to the fact of low frequency of some attacks and high frequency

of others. Thus, reported accuracy of performance is not meaningful unless data is partitioned fairly not arbitrarily as in most approaches (Fossaceca, Mazzuchi *et al*. 2015). Moreover, classifier training has to ensure its adequate learning before enabling it in testing mode.

Any intrusion detection and prevention system in cloud environment has to be aware of the fact that most attackers can develop new types of attacks continuously. Thus, the detection system has to be built based on learning capability in online mode similar to OSELM core systems. Therefore, improving ELM from the above declared perspectives will result in a significant improve in the overall intrusion detection and prevention system?

RESEARCH GOAL AND OBJECTIVE

The main goal of this study is to developing a new intrusion-detection system model to detect and analysis attacks with high speed and accuracy. Reach to this model by achieve the following objectives.

- Developing an optimization scheme to select the best set of ELM classifiers to be enabled for classification.
- To develop a strategy to ensure adequate training of classifier by considering virtual testing of classification in the training mode before enabling for testing.
- To evaluate the performance of proposed algorithm using standard and/or developed metrics.

RESEARCH SCOPE

Among the existing solutions for IDS, all-purpose systems which consider more aspects of the solution and components of the system are studied in this work. Some of the researchers focused on a particular IDS component or a specific type of attack or intrusion, targeted to decrease false alarm rates. For example, Carl et al. investigated different denial-of-service (DoS) detection techniques (Carl, Kesidis *et al.* 2006), given that cloud computing is the target environment; this research focuses more attention to apply ELM as IDS. Prevention Is Newly Acquired feature for intrusion detection systems thus, there are only few published research papers including this feature (Patel, Taghavi *et al.* 2013), this work focus to detection the attacks include the KDD 99 data set (DDos, Probe, R2L, U2R) with on line learning and function.

RESEARCH METHODOLOGY

Using an MATLAB to develop and evaluate the performance of the ELM and GA integrates, to find the best classifier for creating and implementing IDS. The introduced to ELM by Huang (Huang, Zhu et al. 2004), have been widely studied by researchers and applied to a variety of applications (Ding, Zhao et al. 2013). Apply a machine learning approach to the large scale problem of Network Intrusion Detection, Due to its reduced computational complexity, ease of implementation, scalability and good learning performance with respect to other learning algorithms(Cheng, Tay et al. 2012), ELM assigns randomly hidden node parameters, means that the weights between the input layer and the hidden layer, and the hidden layer biases are randomly generated(Huang, Zhu et al. 2006).This work try to use GA to improve the performance of parameters in ELM, because it represents an intelligent exploitation of random search to solve the optimization problems.GA is an iterative process and iterations are generation(Whitley 1994), As Figure-2:

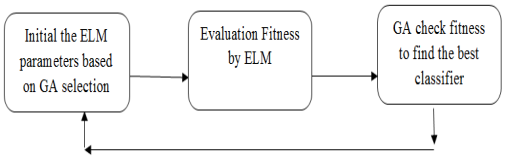

Figure-2 Summary of the methodology.

Detection rate(DR) and false positive rate (FPR) or false alarm rate are still needed to be improved (Kumar and Kumar 2012). So in this work well try to improve these rates based new propose, by divide the training mode to virtual training and virtual testing during the part of the ELM, to ensure adopt classifier enough training before the testing mode.

The intelligent intrusion detection system can only be built if there is availability of an effective data set(Dhanabal and Shantharajah). KDDCUP99 is the mostly widely used data set for evaluation of these systems(Tavallaee, Bagheri et al. 2009), The new version of KDD data set, NSL-KDD is publiclyavailable for researchers through our website. So this work will apply in both of them. Consequently, evaluation results of different research work will be consistent and comparable.

Finally, our approach to Evaluate the performance of this integrates will use many parameters, like the accuracy, recall, Precision and confusion matrix.

Research significant

Security is an important issue for all the networks of companies and institutions, because attackers are trying in many ways to successful access to the data network, despite the development of multiple ways to ensure that the infiltration to the infrastructure of the network via the Internet, firewalls, encryption, etc.

Intrusion detection system (IDS) is one of the techniques for intrusion detection methods, Intrusion detection is the process of monitoring computers or networks for unauthorized entry , activity or file modification(Whitman and Mattord, 2011).An IDS is a software that automates the detection process and detects possible intrusions (Patel, Taghavi *et al.* 2013).

Our goal is to have the ability to accurately classify traffic into normal and attack classes, scale to larger datasets and have a low rate of false alarms. We chose to leverage the state of the art machine learning algorithm called the Extreme Learning Machine (ELM) introduced by(Huang, Zhu *et al.* 2004).

The ELM algorithm has been shown in prior research to have performance on par with the Support Vector Machine (SVM) algorithm but runs much faster while maintaining good classification ability (Wei, Li et al. 2006). Other key advantages of ELMs include ease of implementation and ability to perform multi-class classification directly without using binary classification techniques in succession (Huang, Zhou *et al*. 2012). This work focus to design a new hybrid the ELM with GA to work as IDS in cloud computing environment with less rate of the false alarm

EXPECTED OUTCOMES

- Develop an optimization IDS, By select the best set of ELM classifier for high speed and accuracy for classification data.
- Reduce false alarm in IDS, By perform well in testing mode of classifier when ensure that adequate training of classifier, by considering virtual training and testing of classification in the training mode.
- Get the balance accuracy to evaluate the performance by developed metrics to solve unbalance of data set.

EXPECTED CONTRIBUTIONS

a) Increase the understanding of intrusion detects system with new mechanism.

b) Provide a new IDS framework to detect attacks and classification data with high accuracy and speed.

c) Provide a new strategy to ensure that adequate training of classifier before testing mode. By divided the training mode in virtual training and virtual testing.

d) Provide a new developing metrics to evaluate the performance of the proposed algorithm.

CONCLUSIONS

Nowadays the security is a big challenge in cloud computing; and many tools provided in this issue for instant firewall and intrusion IDS. Because the IDS works depend on the faster of detect and the accuracy of detect and IDS besides facing big problem with high rate in false alarm, our work focus on integrate the ELM and GA to work as IDS with expected high rate of detect and accuracy and for the second problem we suggested to divide the training mode to virtual training and virtual testing to be ensured of selecting a best classifier.

REFERENCES

Anderson J. P. 1980. Computer security threat monitoring and surveillance, Technical report, James P. Anderson Company, Fort Washington, Pennsylvania.

Cheng C., *et al.* 2012. Extreme learning machines for intrusion detection. Neural Networks (IJCNN), The 2012 International Joint Conference on, IEEE.

Choo K.-K. R. 2011. The cyber threat landscape: Challenges and future research directions. Computers and Security. 30(8): 719-731.

Ding S., *et al.* 2013. Extreme learning machine: algorithm, theory and applications. Artificial Intelligence Review. 44(1): 103-115.

Hibbett D. S., *et al*. 2007. A higher-level phylogenetic classification of the Fungi. Mycological research 111(5): 509-547.

Mohammed Hasan Ali, M. F. Z. (2015). Performance Comparison of Transmission Control Protocol and User Datagram Protocol Over Wireless Networks. The 2nd International Conference on Computational Science and Information Management.

Ali, M. H., & Othman, M. A. (2015). Towards a Exceptional Distributed Database Model for Multi DBMS. In Advanced Computer and Communication Engineering Technology (pp. 553-560). Springer International Publishing.

Huang G.-B., *et al*. 2012. Extreme learning machine for regression and multiclass classification. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on. 42(2): 513-529.

Huang G.-B., *et al.* 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on, IEEE.

Huang G.-B., *et al*. 2006. Extreme learning machine: theory and applications. Neurocomputing. 70(1): 489-501.

Hunter R., *et al.* 2009. What Does Google Know, Report G00158124. Stamford, CT: Gartner Research.

Koehler P., *et al*. 2010. Cloud Services from a Consumer Perspective. AMCIS.

Lombardi F. and R. Di Pietro. 2011. Secure virtualization for cloud computing. Journal of Network and Computer Applications. 34(4): 1113-1122.

Parkhill D. F. 1966. Challenge of the computer utility.

Patel A., *et al*. 2013. An intrusion detection and prevention system in cloud computing: A systematic review. Journal of Network and Computer Applications. 36(1): 25-41.

Shelke M. P. K., *et al*. 2012. Intrusion detection system for cloud computing. International Journal of Scientific and Technology Research. 1(4): 67-71.

Stanojevi R. and R. Shorten. 2008. Fully decentralized emulation of best-effort and processor sharing queues. ACM SIGMETRICS Performance Evaluation Review. 36(1): 383-394.

Wang Q., *et al.* 2009. Enabling public verifiability and data dynamics for storage security in cloud computing. Computer Security-ESORICS 2009, Springer: 355-370.

Wei X.-K., *et al*. 2006. Comparative study of extreme learning machine and support vector machine. Advances in Neural Networks-ISNN 2006, Springer: 1089-1095.

Weiss A. 2007. Computing in the clouds. Computing 16.

Whyman B. 2008. Cloud computing. Information Security and Privacy Advisory Board 3.

Zhu Q.-Y., *et al.* 2005. Evolutionary extreme learning machine. Pattern recognition. 38(10): 1759-1763.