International Journal of Enterprise Computing and Business Systems

ISSN (Online) : 2230-8849

Volume 2 Issue 1 January 2013

International Manuscript ID : ISSN22308849-V2I1M6-012013

A STUDY OF ANOMALY INTRUSION DETECTION USING MACHINE LEARNING TECHNIQUES

Zakiya Malek, Dr. Bhushan Trivedi
GLS Institute of Technology

Abstract-In the era of information systems and internet there is more concern rising towards information security in daya to day life, along with the availability of the vulnerability assessment mechanisms to identifying the electronic attacks.Anomaly detection is the process of attempting to identify instances of attacks by comparing current activity against the expected actions of intruder. Machine learning based intrusion detection has the ability to change its execution plan as it obtains new information. The goal of this paper is to provide a comprehensive review of some machine learning based techniques have been applied to AIDS with identifying their main pros and cons.

Keywords-anomaly detection; machine learning; intrusion detection

I. INTRODUCATION

Intrusions and misuse of computer systems are becoming a major concern of our time [3], [2] Traditionally intrusion detection systems (IDS) are classified based on the style of detection they are using: systems relying on misuse detection monitor activity with precise descriptions of known malicious behavior, while anomaly detection systems have a notion of normal activity and flag deviations from that profile [1]. Misuse detectors most commonly used in the form of signature systems that scan network traffic for characteristic byte sequence where as machine-learning is frequently forms the basis for anomaly detection.

When a computer needs to perform a certain task, a programmer's solution is to write a computer program that performs the task. A computer program is a piece of code that instructs the computer which actions to take in order to perform the task. The field of machine learning is concerned with the higher-level question of how to construct computer programs that automatically learn with experience. As machine learning algorithm helps to build detection model from training data automatically, this will save human labor from writing signature of attacks or specifying the normal behavior of a user/system. In many cases, the applicability of machine learning principles coincides with that for the statistical techniques, although the former is focused on building a model that improves its performance on the basis of previous results.

Hence, A-IDS using machine learning can detect new kinds of attacks from the typical characteristics of system users and identify significant variation from the user's established behavior. Although to identify new attack could make it desirable to use such schemes for all situations but the major shortcoming is their resource expensive nature.

In this paper we set out to examine the intrusion detection domain where machine learning is used to identify intruder. We concentrate some of the machine learning based techniques as Bayesian Networks, Markov models, Neural network, fuzzy logic, Genetic algorithm and clustering. Following section presents the various proposed algorithm.

II. BAYESIAN NETWORKS

A Bayesian network is a model that encodes probabilistic relationships among variables of interest. This technique is generally used for intrusion detection in combination with Statistical schemes, a procedure that yields several advantages [4] including the capability of encoding interdependencies between variables and of predicting events, as well as the ability to incorporate both prior knowledge and data.

However, as pointed out in [5], a serious disadvantage of using Bayesian networks is that their results are similar to those derived from threshold-based systems, while considerably higher computational effort is required.

Study says that the f Bayesian networks could be effective in certain situations, because the results obtained from it are highly dependent on the assumptions about the behavior of the target system, therefore a deviation in these hypotheses leads to detection errors.

III. MARKOV MODELS

Two main approaches are their Markov chains and hidden Markov models. A Markov chain is a set of states that are interconnected through certain transition probabilities, which determine the topology and the ability of the model. During a first training phase, the probabilities associated to the transitions are estimated from the normal behavior of the given system. The anomaly detection is done by comparing the anomaly score (associated probability) obtained for the observed sequences with a fixed threshold. Where as in the case of a hidden Markov model, the system of interest is assumed to be a Markov process in which states and transitions are hidden. Only the productions are observable.

Markov-based techniques mostly used in host IDS, normally applied to system calls[6].Also some of its approaches used in network IDS[7]. In general, the model derived for the given system gives a good approach for the claimed profile, while, as in Bayesian networks, the results are highly dependent on the assumptions about the behavior accepted for the system.

IV. NEURAL NETWROKS

Neural network consists of collection of processing elements that are highly interconnected and transform a set of inputs to a set of outputs. It conducts an analysis of information and provides the probability estimate that data match the characteristics which it has been trained to recognize. Neural network based misuse detection systems identify the probability that a particular event or series of events was indicative of an attack against the system. As the neural network gains experience it will improve ability to determine where these events are likely to occur in the attack process or not. The neural network has ability to learn the characteristics of misuse attacks and identify instances that are unlike any which have been observed before by the network

A neural network might be trained to recognize known suspicious events with a high degree of accuracy but for training routine require very large amount of data to ensure that results are statistically accurate. The training of neural network for misuse detection purposes may require thousands of individual attacks sequences and this quantity of sensitive information is difficult to obtain This detection approach has been employed to create user profiles[8], to predict the next command from a sequence of previous ones [9], to identify the intrusive behavior of traffic patterns [10]etc.

V. FUZZY LOGIC

Fuzzy logic is derived from fuzzy set theory under which reasoning is approximated. Thus it used in anomaly detection because the features to be considered can be seen as fuzzy variables [11]. This kind of processing scheme considers an observation as normal if it lies within a given interval[12].Fuzzy logic has proved to be effective; against port scans and probes, but its main disadvantage are the high resource consumption involved and fuzzy logic is always fuzzy.

VI. GENETIC ALGORITHMS

Genetic algorithms are categorized as global search heuristics, and are a particular class of evolutionary algorithms it use techniques inspired by evolutionary biology such as inheritance, mutation, selection and recombination. Thus, genetic algorithms constitute another type of machine learning-based technique, capable of deriving classification rules [13] and/or selecting appropriate features or optimal parameters for the detection process [14].

The main advantage of this subtype of machine learning IDS is the use of a flexible and robust global search method that converges to a solution from multiple directions, whilst no prior knowledge about the system behavior is assumed. Its main disadvantage is the high resource consumption involved.

VII. CLUSTERING DETECTION

Clustering techniques work by grouping the observed data into clusters, according to a given similarity or distance measure. The procedure most commonly used for this consists in selecting a representative point for each cluster. Then, each new data point is classified as belonging to a given cluster according to the proximity to the corresponding representative point [15]. Some points may not belong to any cluster; these are named outliers and represent the anomalies in the detection process. Clustering techniques determine the occurrence of intrusion events only from the raw audit data, and so the effort required to tune the IDS is reduced.

VIII. ISSUES & CHALLENGES

Anomaly Intrusion detection Techniques can give either positive or negative results. Positive results are designed to activate an alert when a pattern sufficiently matches known attack criteria and negative result provide an alert when the pattern does not match any known or expected pattern. In either of the two cases the techniques are essentially to identifying a pattern of user activity .This pattern of activity is indicate user’s behavior.

All available techniques are restricted in that they generate high false positive and false negative rates. High false positive rate usually [16] obtained from the lack of good studies on the nature of the intrusion events. Therefore it requires exploration and development of new, accurate processing schemes, as well as better structured approaches to modeling of system.

These false results must then be examined to determine their true nature. In essence, this becomes a question similar to that seen in the forensic sciences: Why did the user do that? Is this an attempt at malicious activity or is it innocuous? Much of this analysis can be done through the available data. Some of this analysis must be done in concert with the user through queries to identify their goals; as is done through typical forensic processes. Consequently, computer forensics [17], the process of analyzing the data, is critical to the effectiveness of the intrusion analysis process.

Other issue is the absence of appropriate metrics and assessment methodologies, as well as a general framework for evaluating and comparing alternative IDS techniques [18] [19] and despite that different mechanisms to elude IDS have been described in the literature [20] more significant efforts should be done to improve intrusion detection technique in the aspect where IDS systems perform poorly in protecting themselves from attacks [21].

IX. CONCLUSION

In brief, the present paper discusses the some of the existing Machine learning based AIDS together with their general operational architecture. Another valuable aspect of this study is that it describes, open issue and challenges related with the current A-IDS which is helpful to design new approaches for intrusion detection system. It can be concluded from the study done by scanning various research papers and feed backs that still problem exist to identify new attacks which indicate that there is scope to research.

References

1. D. R. Ellis, J. G. Aiken, K. S. Attwood, and S. D. Tenaglia, “A Behavioral Approach to Worm Detection,” in *Proc. ACM CCS WORM Workshop*, 2004.
2. Richard A. Clarke, “Convergence and Transition, Privacy and Security,” Remarks at SafeNet 2000,Redmond, WA, December 2000.

3. Greg Farrell, “Police have few weapons against cyber-criminals. Problem stems from lack of funds, training,” *USA Today*, pp. 5B, December 6, 2000.
4. Heckerman D. A tutorial on learning with Bayesian networks.Microsoft Research; 1995. Technical Report MSRTR-95-
5. Kruegel C., Mutz D., Robertson W., Valeur F. Bayesian event classification for intrusion detection. In: Proceedings of the19th Annual Computer Security Applications Conference;2003.
6. Yeung DY, Ding Y. Host-based intrusion detection using dynamic and static behavioral models. Pattern Recognition 2003;36(1):229–43.
7. Mahoney M.V., Chan P.K. Learning nonstationary models of normal network traffic for detecting novel attacks.In: Proceedings of the Eighth ACM SIGKDD; 2002.p. 376–85.
8. Fox K., Henning R., Reed J., Simonian, R. A neural network approach towards intrusion detection. In: 13th National Computer Security Conference; 1990. p. 125–34.
9. Debar H., Becker M., Siboni, D. A neural network component for an intrusion detection system. In: IEEE Symposium on Research in Computer Security and Privacy; 1992.p. 240–50.
10. Cansian A.M., Moreira E., Carvalho A., Bonifacio J.M. Network intrusion detection using neural networks. In: InternationalConference on Computational Intelligence and Multimedia Applications (ICCMA’97); 1997. p. 276–80.
11. Bridges S.M., Vaughn R.B. Fuzzy data mining and genetic algorithms applied to intrusion detection. In: Proceedings of the National Information Systems Security Conference; 2000.p. 13–31.
12. Dickerson J.E. Fuzzy network profiling for intrusion detection. In:Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society (NAFIPS);2000. p. 301–6.
13. Li W. Using genetic algorithm for network intrusion detection. C. S.G. Department of Energy; 2004. p. 1–8.
14. Bridges S.M., Vaughn R.B. Fuzzy data mining and genetic algorithms applied to intrusion detection. In: Proceedings of the National Information Systems Security Conference; 2000. p. 13–31.
15. Portnoy L., Eskin E., Stolfo S.J. Intrusion detection with unlabeled data using clustering. In: Proceedings of The ACM Workshop on Data Mining Applied to Security; 2001.
16. Axelsson S. The Base-rate fallacy and its implications for the difficulty of intrusion detection. ACM Transactions on Information and System Security 2000;3:186–205
17. Kevin Mandia and Chris Prosise, *Incident Response:Investigation Computer Crime*, Osborne/McGraw-Hill, 2001.
18. Stolfo SJ, Fan W. Cost-based modeling for fraud and intrusion detection: results from the JAM project. DARPA Information Survivability Conference & Exposition 2000:130–44.
19. Gaffney J, Ulvila J. Evaluation of intrusion detectors: a decision vtheory approach. IEEE Symposium on Security and Privacy 2001:50–61
20. Ptacek T, Newsham T. Insertion, evasion and denial of service: eluding network intrusion detection. Secure Networks 2003

21. Axelsson S. Research in intrusion detection systems: a survey.Technical report. Chalmers University of Technology. Goteborg 1998.