Research on Principle Techniques for Network Intrusion Detection based on Data Mining and Analysis Approach

JIANG Shan[1, a], CHEN Changai[2]

[1,2] Institute of information technology, Henan University of Traditional Chinese Medicine, Zhengzhou 450000,China

[a] j_xueshan@126.com

Keywords: Network Intrusion Detection; Data Mining; Data Analysis Technique; Network Structure Optimization and Security; Theoretical Background.

Abstract. The security of software applications, from web-based applications to mobile services, is always at risk because of the open society of internet. With the increase in the number of network throughput and security threats, intrusion detection system has attracted much attention in recent years. In this paper, we undertake the research on the principle techniques for network intrusion detection based on data mining and analysis approach. We adopt the prior knowledge on Bayesian network which is a directed acyclic graph, each node represents a random variable and an edge said direct probabilistic dependencies between two connected nodes. Then, we use the traditional risk assessment model to measure the possibility of being hearted. The numeric analysis and experimental illustration indicates the effectiveness of our method compared with other popular adopted state-of-the-art methodologies. In the future, we plan to introduce the graph and complex network theory into our prototype system to enhance the performance.

Introduction

The security of software applications, from web-based applications to mobile services, is always at risk because of the open society of internet. With the increase in the number of network throughput and security threats, intrusion detection system has attracted much attention in recent years. IDS mechanism for monitoring system and network case, collect useful data, such as suspicious activity and environmental background, and analyzes the data to detect malicious intent. In general, intrusion detection method is divided into signature-based intrusion detection or anomaly-based intrusion detection system (ads). SD is a known process comparison signature pattern attack or threat to capture events to identify possible invasion [1-2]. Found in the process of advertising from a known behavior, behavior and construct summary on behalf of the normal or expected from monitoring routine activities, network connection, the host or the user for a period of time. The current industrial NIDS misuse-based method and practical solutions, using the signature against intrusion detection model each of these types of attacks [3-5]. As a typical misuse detection method, search package attack pattern matching methods and use agreement rules and string matching. Pattern matching method can effectively detect the invasion of the famous. But they rely on timely generate attack signatures, and to detect the novel and unknown attacks [6-7]. In the spread of the novel and unknown attack and defense based on signature of any known attacks are possible. In addition, increase the diversity of attack signature block modeling [8]. Machine learning process will automatically from the data dependence, inference and generalization to invisible data extrapolation of dependencies. Machine learning method of the intrusion detection model and attack data of normal network data, and allow the network characteristics were used to detect the unknown attacks.

To address the problems raised above, we built a Bayesian classifier for intrusion detection by Bayesian Model Averaging (BMA) over the k-best BN classifiers. When future data points are classified, the decision is made by averaging over the prediction results of the k-best BN classifiers. The motivation of doing this is that multiple BNs are better than one BN in representing the probability distribution of the model space, thus they offer better predictive power than one network, particularly in the domain where only small training datasets are available.

In this paper, we conduct research on principle techniques for network intrusion detection based on data mining and analysis approach. We will adopt the state-of-the-art machine learning and data mining tools to help detecting the dangerous elements in the network. The systematic description of our approach and the algorithm analysis are shown in the following sections.

Our Proposed Approach

Bayesian Network based Theory. This is a directed acyclic graph, each node represents a random variable and an edge said direct probabilistic dependencies between two connected nodes. For each node, contains the node has a conditional probability distribution probability of the different values in the value of his parents. Formal, lattice structure assertions, each node is conditionally independent of all non-descendants to its parent node. The probability is shown in the formula 1.

$$p(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} p\left(X_i \mid Pa_G(X_i)\right) \tag{1}$$

Figure 1 gives a simple example of BN that portrays the probabilistic relationships. Model resolution is the choice of training data sparseness of lack of scoring index, namely the size of the data set, small relative to the number of variables. In this case, there can be many different BN the same training data fitting. Therefore, using a single BN can lead to bad data to predict the future.

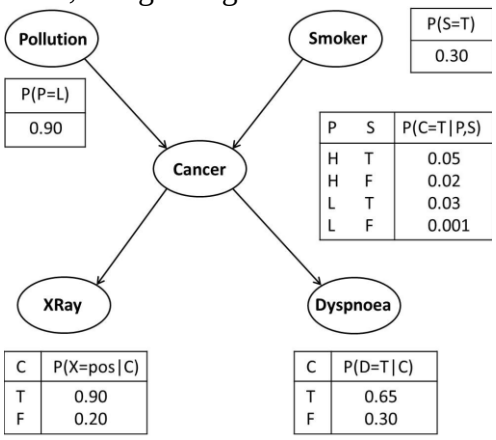

Fig.1The Illustrative Example of BN

A promising solution to alleviate this problem is to employ BMA, which provides a principled approach to the model uncertainty problem by integrating all possible models weighted by their respective posterior probabilities. The prior probability is shown in the formula 2.

$$p(x \mid D) \approx \frac{\sum_{G \in g} p(x \mid G, D)\, p(G \mid D)}{\sum_{G \in g} p(G \mid D)} \tag{2}$$

In spite of the difficulties and costs related to estimate probability, risk assessment model is arguably the most useful tools, the existing investment decisions. Common alternative is to make decisions based on intuition, in this case, assuming that there is no clear discussion or analysis. Using the risk assessment framework, we can analyze the decision-making is based on the hypothesis that clear: assuming that can discuss, review, and improve the future researchers. Therefore, this article takes the first step to create a more systematic approach to evaluate ids in the AMI network deployment options [9-10]. Its theoretical basis is reasonable, its implementation is simple. In addition, the learning algorithm to correct the error classification by weak classifier, it is better than most of the learning algorithm is not so easy to learn. Adaboost based classifier to identify performance usually is encouraging. In our framework, a mixed network of weak classifier and an online algorithm results in a parameterized place on each node intrusion detection model. All nodes in the parameters of the model is combined into a global intrusion detectors on each node using a small amount of samples, and the combination is by using a based on particle swarm optimization (PSO) algorithm and vector machine (SVM). The figure 2 shows the revised framework.

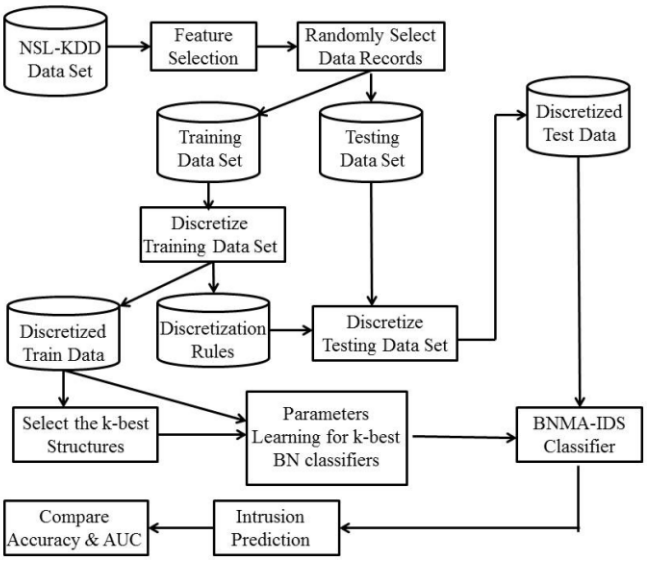

Fig.2The revised framework

The Risk Assessment Model. The detection rate and false alarm rate we use to evaluate the performance of the algorithm to detect network intrusion. It is necessary to pay more attention to the false alarm rate, because in a real application, most of the network behavior is normal. High false alarm rate waste of resources, because each alarm must be checked. Adaboost-based learning algorithm of detection rate and false alarm rate depends on the initial weights of the training sample. So we propose to adjust the initial sample weights in order to balance the detection rate and the false alarm rate. The formula 3 calculate the solution.

$$R_c = C_c + \sum_i p_i \left(d_c^i A_c^i + \left(1 - d_c^i\right) A_m^i \right) \tag{3}$$

Even if a utility investment in the infrastructure of a distribute IDS it still needs to be a center for receiving and alarm management. We noticed that distributed infrastructure refers to the intrusion detection sensor deployment in AMI. Alarm reports those sensors, however, will be in a centralized way of management, therefore, in our framework, we assume that the investment in a distributed sensor will also need a basic investment in a centralized solution. The measurement of the risk is shown in the figure 3.

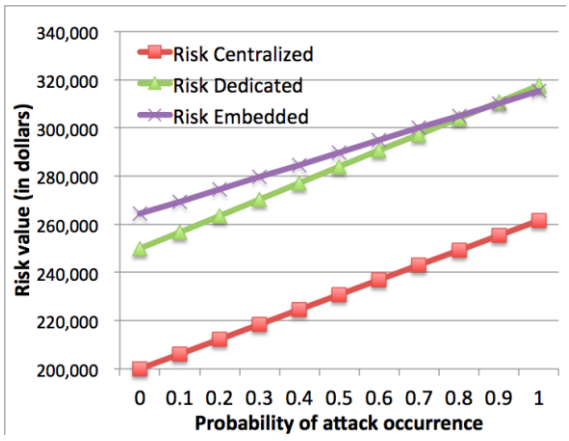

Fig.3The Probability Attack Statistical Figure

Simulation Result and Experiment

We proposed a Bayesian classifier using BMA of k-best BN classifiers, called BNMA classifier, for intrusion detection. Previous IDS using BN classifier has two problems. The training set used for constructing the global model only contains 4000 randomly chosen samples, and the test set for the global model contains 284672 samples of the four types of attacks and all the normal samples. We from the traditional decision tree stumps and online learning algorithms and online GMM and our

online learning algorithm, the detection rate and the local model in each node of the rate of false positives and based on global model and algorithm, as well as those by combining the local model using the sum rule and support vector machine (SVM) algorithm. See, our algorithm and based on the algorithm greatly improves the detection precision and PSO algorithm and sum rule based on the results of each node to obtain more accurate and support vector machine (SVM) algorithm. The figure 4-5 shows our result.

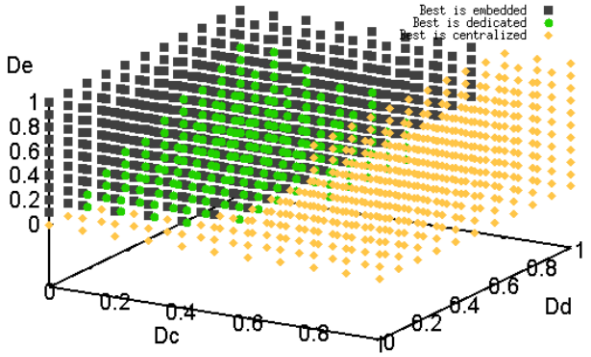

Fig.4The Selection of the Most Cost-efficient Architecture

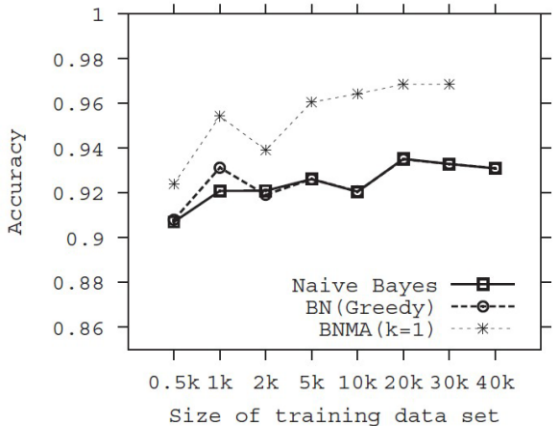

Fig.5The Comparison Result of the Technique

Conclusion and Summary

The security of software applications, from web-based applications to mobile services, is always at risk because of the open society of internet. With the increase in the number of network throughput and security threats, intrusion detection system has attracted much attention in recent years. IDS mechanism for monitoring system and network case, collect useful data, such as suspicious activity and environmental background, and analyzes the data to detect malicious intent. In this paper, we undertake the research on the principle techniques for network intrusion detection based on data mining and analysis approach. We adopt the prior knowledge on Bayesian network which is a directed acyclic graph, each node represents a random variable and an edge said direct probabilistic dependencies between two connected nodes. For each node, contains the node has a conditional probability distribution probability of the different values in the value of his parents. Through the combination of the modified risk assessment model, we finalize the proposed framework properly. The experimental analysis proves the feasibility and reliability of proposed approach.

References

[1] Agarwal R, Joshi M V. PNrule: A New Framework for Learning Classifier Models in Data Mining (A CaseStudy in Network Intrusion Detection)[J]. in Network Intrusion Detection), First SIAM Conference on Data Mining, 2000.

[2] Sommer R, Paxson V. Enhancing byte-level network intrusion detection signatures with context[J]. IN PROC. 10TH ACM CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY, 2003:262 - 271.

[3] Kruegel C, Valeur F, Vigna G, et al. Stateful Intrusion Detection for High-Speed Networks[J]. Security and Privacy, 2002. Proceedings. 2002 IEEE Symposium on, 2003:285 - 293.

[4] Kim J H, Im E G, Yoon J B, et al. Network intrusion model for analyzing intrusion patterns[J]. Advanced Communication Technology, 2004. The 6th International Conference on, 2004:303 - 308.

[5] Fang M L, Ai-xia W. Computer Network Intrusion Detection Technology[J]. Modern Computer, 2011.

[6] PETER L, RAINER L. Communications network intrusion monitor system uses filter rules and control lists: DE, DE10152010 A1[P]. 2001.

[7] Cho H, Kim D, Kim J, et al. Network Processor Based Network Intrusion Detection System[J]. Computer & Digital Engineering, 2009.

[8] Zhou J, Carlson A J, Bishop M. Verify results of network intrusion alerts using lightweight protocol analysis[C]. //Computer Security Applications Conference, 21st Annual. IEEE, 2005:10 pp. - 126.

[9] Xu, Biao, Xu-Huan Wang, Wei Wei, and Haoxiang Wang. "On reverse Hilbert-type inequalities." Journal of Inequalities and Applications 2014, no. 1 (2014): 1-11.

[10] Khoshgoftaar T M, Gao K, Lin H. Indirect classification approaches: a comparative study in network intrusion detection.[J]. International Journal of Computer Applications in Technology, 2007, (4):232-245.