Network Security Based On Pattern Matching: An Overview

MR. V. K. PACHGHARE[1] , DR. PARAG KULKARNI[2]
[1] Assistant Professor, Computer Engineering & IT Department,
College of Engineering, Pune (An autonomous Institute of Government of Maharashtra, India)

[2]Chief Scientist and Research Head, Capsilon Research Labs, Capsilon India, Pune, India

Abstract
The security of computer networks plays a strategic role in modern computer systems. In order to enforce high protection levels against malicious attack, a number of software tools have been currently developed. Intrusion Detection System has recently become a heated research topic due to its capability of detecting and preventing the attacks from malicious network users. A pattern matching IDS for network security has been proposed in this paper. Many network security applications rely on pattern matching to extract the threat from network traffic. The increase in network speed and traffic may make existing algorithms to become a performance bottleneck. Therefore it is very necessary to develop faster and more efficient pattern matching algorithm in order to overcome the troubles on performance.
Keywords*: - Intrusion detection, Pattern Matching, Network Security*

I. Introduction

The Internet as well as local networks is expanding at a tremendous speed. This one way helps to improve the quality and convenience of the human life but on the other hand provides a platform for network criminals and hackers. The number of intrusions into computer systems is growing and raising concerns about computer security. So computer networks are usually protected against intrusions by the means of access restriction policies. Despite the effort devoted to carefully designing a system to protect such attacks, network security is very difficult to guarantee, since attacks exploit unknown weaknesses or bugs, which are usually contained in system and application software (McHugh et al, 200; Proctor, 2001). Intrusion detection which refers to a certain class of system attack detection problems is a relatively new research area in computer and information security [7]. In general IDS can be categorized into misuse and anomaly detection approaches. Misuse detection system can reliably identify intrusion attacks in relation to the known patterns of discovered vulnerabilities. However, emergent intervention of security experts is required to define accurate rules or patterns, which limit the applications of misuse detection systems, identify deviations from normal network behaviors and alert for potential unseen attacks [2]. It is able to detect novel attacks without a priori knowledge about them if the classification model has the generalization capability to extract intrusion pattern and knowledge during training. Unfortunately, anomaly detection approach suffers from high false positive rate on classifying normal network traffic. So, machine learning techniques have been used to capture the normal usable patterns and classify the behavior as either normal or abnormal [26].

While the pattern matching algorithms are applied to network security, the speed of pattern matching usually becomes a bottleneck.

This paper proposes a pattern matching algorithm which overcomes the shortcomings of traditional algorithms. The paper is organized as follows: section –II related work, section – III Pattern Recognition Methods, section – IV Pattern Recognition System, section – V conclusion.

II. Related Works

Pattern Recognition is one of the most important areas which have been studied in computer science. In a standard formulation of the problem, we are given a pattern and a data and it is required to find all occurrences of the pattern in the data [3].

Since the publication of the Bayer-Moore and Knuth-Morris-Pratt algorithm, several hundreds of papers have been published dealing with pattern recognition.

As seen from the literature survey carried above, the researchers worldwide have been working in various Intrusion Detection techniques in general and Pattern Matching techniques in particular. The above overview of related works indicates that pattern matching techniques are suitable to provide a solution to some open issues in IDS development. The various algorithms are available for Pattern Matching technique with varying degree of accuracy. It should be stated that, for the deployment of IDSs using pattern matching technique in operational environments, one of the main difficulties is the high production of false alarms. There is a need of an algorithm which gives low false alarm rate.

Manuscript received October 5, 2008
Manuscript revised October 20, 2008

III. The Pattern Recognition Methods

Pattern recognition undergoes an important developing for many years. Pattern recognition includes a lot of methods which are impelling the development of numerous applications in different field.

1. Statistical Pattern Recognition

The Statistical methods have been commonly used for pattern recognition. Statistical approaches have a number of advantages. It can provide accurate notification of malicious activities that typically occur over extended periods of time and are good indicators of impending denial-of-service attacks [23]. However, it also has drawbacks. It can be difficult to determine thresholds that balance the likelihood of false positive alarms with the likelihood of false negative alarms. In addition, this method need accurate statistical distributions, but, not all behaviors can be modeled using purely statistical methods. The statistical pattern recognition deals with features only without consider the relations between features.

2. Data Clustering

Data clustering is a technique for finding patterns in unlabeled data with many dimensions. It is an unsupervised method. The main advantage of data clustering is the ability to learn from and detect intrusions in the audit data, while not requiring the explicit descriptions of various attack classes. The method of data clustering can be partitioned into two classes, one is hierarchical clustering and the other is partition clustering.

3. Fuzzy Set

As the quantitative features in the intrusion data are partitioned into the interval with crisp boundary, there might exists a sharp boundary problem for pattern classification. The fuzzy logic provides the partial membership in set theory to integrate with the association rules and frequent episodes which solved the above problem. Fuzzy rule-based systems inspired by the fuzzy set theory have been successfully applied to solve many complex and non-linear problems. The application of fuzzy sets in pattern recognition started in 1966, where two basic operations – abstraction and generalization. Pattern Recognition system based on fuzzy sets theory can imitate thinking process of human being widely and deeply.

4. Artificial Neural Networks

The Artificial Neural Network methodology enables us to design useful nonlinear systems accepting large numbers of inputs, with the design based solely on instances of input-output relationship. The first Artificial Neural Network model was proposed in 1943. Today it is developing very fast. Basically it is a data clustering method based on distance measurement. This approach applies biological concepts to machines to recognize patterns. Pattern Recognition using Artificial Neural Network is a very attractive since it requires minimum priory knowledge, and with enough layers and neurons, an Artificial Neural Network can create any complex decision region.

5. Structural Pattern Recognition

Structural Pattern Recognition emphases on the description of the structure, namely explain how some simple sub-patterns compose one pattern. The syntax analysis and structure matching are the two main methods in structural pattern recognition. The basis of syntax analysis is the theory of formal language, the basis of structure matching is some of special technique of mathematics based on sub-patterns. The structural pattern recognition handles with symbol information. This method can be used in applications with higher level, such as image interpretation. Pattern Recognition of multidimensional objects can be done by structural pattern recognition with static classification or artificial neural networks.

6. Support Vector Machine (SVM)

Support Vector Machine based on the statistical theory and method of SVM is an effective tool that can solve the problems of Pattern Recognition.

7. Approximate reasoning approach to Pattern Recognition -

This method used two concepts- one is fuzzy applications and the other is compositional rule of inference can cope with the problem for rule based pattern recognition [31].

8. A logical combinatorial approach to Pattern Recognition

This approach can apply for both supervised pattern recognition and unsupervised pattern recognition.

IV. Pattern Recognition System

The anomaly intrusion detection system suffers from high false alarm rates while the misuse intrusion detection system lacks generalization capabilities and cannot detect new attack types. Pattern Recognition techniques have been found to strike a fine balance in this trade off. The use of pattern recognition and classification has grown in the past few years. The complexity of the classification systems and their increased availability has made them more accessible. They are able to filter noise and extract features from traffic to facilitate classification. Pattern classification is a series of steps, starting with the input, moving to segmentation, data extraction and translation and finally classification. After the classification, cost factors can be added to increase the power of the decision to act. In order for pattern recognition to be useful in network security, two large problems must be addressed; Data extraction and classification. Information from a single packet is inadequate for feature extraction.

Collating multiple packets might provide a basis for describing features but how many packets are enough and how do we synthesize the data from multiple packets to make it useful for data extraction.

The aim of pattern classification is to utilize the information acquired from pattern analysis to discipline the computer in order to accomplish the classification. The step of classification is the kernel of the pattern recognition system.

The general pattern recognition system is given in fig. 1 below:

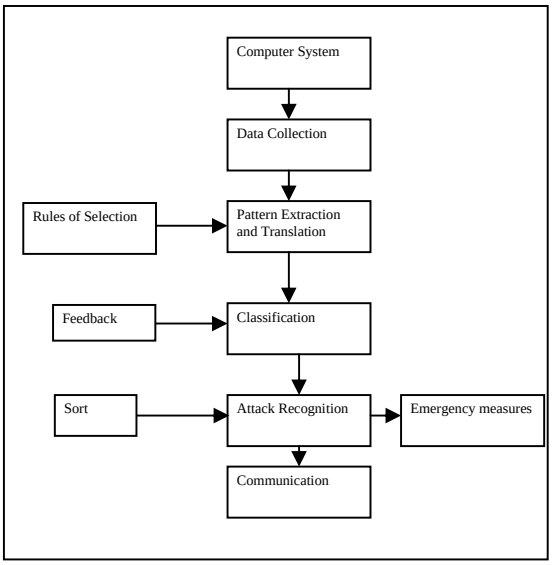

Fig. 1 The general pattern recognition system

The next step following Data extraction is classification. It is the process of using the data set to classify the traffic as normal or illegitimate traffic. The classifications can be divided into three categories: normal, Denial of Service and Scan. Numerical values were assigned the three categories based on their probability. Four types of classifiers are used: Bayesian, Feed Forward with Backpropagation, ART2 and Kohonen neural networks.

V. Conclusion

In the last twenty years, Intrusion Detection Systems have slowly evolved from host and operating system specific application to distributed systems that involve a wide array of operating system. The challenges that lie ahead for the next generation of Intrusion Detection Systems are many. Traditional Intrusion Systems have not adapted adequately to new networking paradigms like wireless and mobile networks. Factors like noise in the audit data, constantly changing traffic profiles and the large amount of network traffic make it difficult to build a normal traffic profile of a network for the purpose intrusion detection.

A perennial problem that prevents widespread deployment of IDS is their inability to suppress false alarms. Therefore, the primary and probably the most important challenge that needs to be met is the development of effective strategies to reduce the high rate of false alarms.

Pattern Recognition is the heart of all scientific inquiry, including understanding ourselves and the real-world around us and the developing of Pattern Recognition is increasing very fast, the related fields and the applications of pattern recognition became wider and wider. IDS based on pattern recognition have the specialties such as self adaptability, low consume, tolerances and self learning and so on. So it can wholly improve the all performance of security system. Of course, the building of the patterns is the precondition.

References

[1] Zhou Chunyue, Liu Yun, Zhang Hongke, “A Pattern Matching Based Network Intrusion Detection System”, 1-4244-0342-1/06/2006 IEEE

[2] Chi-Ho Tsang, Sam Kwong, Hanli Wang, Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. The journal of Pattern Recognition 40 (2007)

[3] Shai Rubin, Somesh Jha, Barton Miller, “Protomatching Network Traffic for High Throughput Network Intrusion Detection”, CCS’06, Oct 30-Nov 3, 2006, Alexandria, Virginia, USA

[4] Monther Aldwairi, conte, and Paul Franzon, “Configurable String Matching Hardware for Speeding up Intrusion Detection, ACM SIGARCH Computer Architecture News, Vol. 33, No. 1, March 2005

[5] Animesh Patcha, Jung-Min Park, “An Overview of Anomaly Detection Techniques: Existing Solutions and latest Technological Trends”, Computer Networks 51 (2007)

[6] Zachary K. Baker and Viktor K. Prasanna, High-throughput Linked-Pattern Matching for Intrusion Detection Systems, ANCS’05, Oct 26-28, 2005, Princeton, New Jersey, USA

[7] Dit-Yan Yeung, Yuxin Ding, Host-based Intrusion Detection using Dynamic and Static Behavioral Models, The journal of Pattern Recognition 36 (2003).

[8] Wu Yang, Bin-Xing Fang, Bo Liu, hong-li Zhang, Intrusion Detection System for High-Speed Network, Computer Communications 27 (2004)

[9] Zachary Baker, V.K. Prasanna, Time and Area Efficient Pattern Matching on FPGAs, FPGA’04, Feb 22-24, 2004, Monterey, California, USA

[10] Christopher Kruegel, Giovanni Vigna, William Robertson, A Multi-Model Approach to the Detection of Web-Based Attacks, Computer Networks 48 (2005)
[11] JIANG Bo, LIU Bin, High-Speed Discrete Content Sensitive Pattern Match Algorithm for Deep Packet Filtering, Proceedings of the 2003 International Conference on Computer Networks and Mobile Computing (ICCNMC'03), IEEE
[12] Ningning Wu, Jing Zhang, "Factor-analysis based anomaly detection and clustering", Decision support Systems 42 (2006)
[13] Qingbo Yin, Rubo Zhang, Xueyao Li, " An New Intrusion Detection Method Based On Linear Prediction", InfoSecu04, Nov 14-16, 2004, Pudong, Shanghai, China
[14] Robin Sommer, Vern Paxson, Enhancing Byte-Level Network Intrusion Detection Signatures with Context, CCS'03, Oct 27-31, 2003, Washington, DC, USA
[15] Steven L. Scott, A Bayesian pParadigm for Designing Intrusion Detection Systems, Computational Statistics & Data Analysis 45 (2004)
[16] R Sidhu and V. K. Prasanna, "Fast Regular Expression Matching using FPGAs", in IEEE Symposium on Field-Programmable Custom Computing Machine, Napa Valley, CA, April 2001, IEEE
[17] Lih-Chyau Wuu, Sout-Fong Chen, Building Intrusion Pattern Miner for Snort Network Intrusion Detection System, 0-7803-7882-2/03/2003 IEEE.
[18] Sourcefire.Snort, The Open Source Network Intrusion Detection System. http://www.snort.org, 2003
[19] Proctor, P.E., 2001. The Practical Intrusion Detection Handbook, Prentice-Hall, Englewood Cliffs, NJ
[18] Axelsson S., 2000, the Base-Rate Fallacy and the Difficulty of Intrusion Detection, ACM Trans, Inform. Syst. Security 3 (3)
[20] Fang Hao, Murali Kodialam, T.V.Lakshman, Hui Zhang, Fast payload-Based Flow Estimation for Traffic Monitoring and Network Security, ANCS'05, Oct 26-28, 2005, Princeton, New Jersey, USA
[21] Lothar Wendehals, Alessandro Orso, Recognizing Behavioral Patterns at Runtime using Finite Automata, WODA'06, May 2006, Shanghai, China
[22] Yang Wang and Hidetsune Kobayashi, High Performance Pattern Matching Algorithm for Network Security, IJCSNS, Vol.6 No.10, Oct 2006.
[23] Giacinto, Fabio Roli, Luca Didaci, Fusion of Multiple Classifier for Intrusion Detection in Computer Networks, Pattern Recognition Letters 24 (2003).
[24] Richard Lippmann, david Fried, Joshua, Kendall, McClung, Weber, Webster, Wyschogrod, Cunningham, Zissman, "Evaluation Intrusion Detection Systems:The 1998 DARPA Off-Line Intrusion Detection Evaluation", 0-7695-0490-6/99, IEEE
[25]Stephen Gossin, et al: Pattern Matching in Snort. http://www.sporksoft.com/~njones/notes/CSE202/project.pdf, 2002
[26] Mohammad Saniee Abadeh, Jafar Habibi, Zeynab Barzegar, Muna Sergi, A Parallel Genetic Local Search Algorithm for Intrusion Detection in Computer Networks, Engineering Applications of Artificial Intelligence 20 (2007).
[27] Shiuhpyng Winston Shieh and Virgil D. Gligor, "A Pattern Oriented Instrution Model and its Applications", in Proceedings of the 1991 IEEE Computer Society Symposium on Research in Security and Privacy, pages 327-342. IEEE, Service Center, Piscataway, NJ, May 1991
[28] Jian-Cai Huang, Jun-Feng Tian, Rui-Zhong Du, Jian-Qiang Zhai, " Research of Pattern Matching in Intrusion Detection", Proceedings of second international conference on matching learning and cybetnetics, Xi'an, 2-5 Nov. 2003, IEEE
[29] Young H. Cho and William H. Mangione-Smith, A Pattern Matching Co-processor for Network Security, DAC 2005, June 13-17, 2005, Anaheim, California, USA
[30] Lippmann, Fried, Graf, Haines, Kendall, McClung, Weber, Webster, Wyschogrod, Cunningham, Zissman, Evaluating Intrusion Detection Systems: The 1998 DARPA OFF-Line ID Evaluation, IEEE, 1999
[31] Liu, Sun, Wang, Pattern Recognition: an overview, IJCSNS, June 2006
[32] Zhou, Liu, Research on computer network security based on Pattern Recognition, IEEE, 2003

Mr. V. K. Pachghare has an experience of eighteen years in the teaching field. Presently, he is working as Assistant Professor in Information Technology, College of Engineering, Pune, India (An Autonomous Institute of Government Of Maharashtra). He worked as a member of Board of Studies, Computer Engineering, Pune University. Presently, he is a member in the Board of Computer Engineering, College of Engineering, Pune. He is author of a book "Computer Graphics". He has 5 International publications.

An alumnus of IIT and IIM, **Dr. Parag** completed his Ph.D. in Computer Engineering from IIT Kharagpur. He has been working in IT industry for last 17 years. He has worked as Research head, operations head, GM, Director and was instrumental in building world-class software product companies.. He is working as a Vice-President Strategic Development and Chief Scientist at Capsilon INDIA. His name and profile is selected for listing in "Marquis Who's Who in the world" (Science and Engineering) –2009.

He has written many business articles. He has more than 60 International publications and two patents pending in US PTO. He is member of IASTED technical committee, WSEAS working committee, board of studies of two institutes and is guiding 7 Ph.D. students. Parag has conducted more than 25 tutorials on research and business topics at various international conferences He is visiting faulty at IIM Indore. He is pioneer of new management program "Deliverance from Success" for Executives and author of books "Deliverance from Success" and "IT strategy". His areas of research and product development include M-maps, intelligent systems, text mining, image processing, Decision systems, Semi-constrained influence diagrams, forecasting, quantitative analysis, knowledge management, IT strategy, classification, distributed computing, AI and machine learning.