The Application of Adaptive Fuzzy Data Mining Techniques on Intrusion Detection

鍾居璋、陳鴻文

E-mail: 9607617@mail.dyu.edu.tw

ABSTRACT

The chances of invasion through Internet are increasing for the popularity of computer networks. Thus, intrusion detection systems (IDS) are eagerly in demand. However, for traditional IDS, it is vital that their accuracy rates are too low and their false alarm rates are too high due to the high complexity of invasion behaviors. The issue of promoting the effectiveness of IDS has become unavoidable today. This thesis focused on analyzing the packets of DARPA set with technologies of data mining and fuzzy theory. First, misuse rules of intrusion detection were constructed by using the C5.0 algorithm to build decision trees of intrusion training packets. Following that, all training packets were classified by ISODATA algorithm into 67 clusters; each cluster will form a decision tree by C5.0 algorithm, and each route from the root node to a leaf node within the decision tree becomes a near-anomaly rule of intrusion detection. For each newly incoming packet, the misuse rules of intrusion detection will be applied first to identify its attack pattern. If not identifiable, the packet will be examined by those near-anomaly rules of intrusion detection. The distances between the packet and each center of 67 clusters (only one near-anomaly rule will be fired for each cluster) were calculated. Finally, the packet will be regarded as belonging to a particular behavior (normal or intrusion) which corresponds to the largest weighted sum of memberships relative to the calculated distances. On the other hand, three kinds of sequential invasion patterns were also proposed by examining the packets of the same source and destination IP ’ s. An alarm would be issued early when a user was trying to invade a server under one of the proposed sequential intrusion patterns. Thus, damages of the three invasions could be reduced. Experimental results of testing DARPA dataset illustrated the effectiveness of proposed methods, including both the fuzzy classification for a single packet and sequential invasion patterns for multiple packets. The average accuracy rates of fuzzy classification and sequential invasion patterns are 86% and 60%, respectively. In addition, the rate of false alarm of the proposed fuzzy classification is also reduced to only 2%. Thus, the proposed techniques can improve the problems of traditional techniques of intrusion detection in both accuracy and false alarm rates.

Keywords : Data Mining, Fuzzy Theory, Intrusion Detection System, Decision Tree, Sequential Pattern

Table of Contents

REFERENCES

一、中文部份 台灣網路資訊中心(2007), [線上資料], 來源: http://www.myhome.net.tw/2007_03/main03_1.htm [2007, September]. 邱美珍(1996), 決策樹學習法中連續屬性之分類研究, 中原大學資訊工程研究所未出版碩士論文。 柯文元(2005), 模擬關聯與情境法則探勘於

入侵偵測之研究，大葉大學資訊管理學系碩士班未出版碩士論文。 二、英文部份 Adler, P. S. (1993). Time-and-motion regained. Harvard Business Review, 71 (1), 97-108. Agrawal, R. & Srikant, R. (1994). Mining Sequential Patterns. Proceedings of the International Conference on Data Engineering. Berry, M. J. A. & Linoff, G.. (1997). Data Mining Technique for Marketing Sale and Customer Support. Wiley Computer, New York, NY. Berry, M., & Linoff, G.. (1997). Data Mining Techniques for marketing, sales, and Customer Support. New York. Wiley Computer Publishing. Bezdek, J. C. (1981). Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum, New York. C5.0. Available: http://www.rulequest.com/ [No date]. Debar, H., Becker, M., & Siboni, D. (1992). A Neural Network Component for an Intrusion Detection System. IEEE Security and Privacy, 10(2), 155-169. Dickerson, J. E., Juslin, J., Koukousoula, O., & Dickerson, J. A. (2001). Fuzzy intrusion detection. IFSA World Congress and 20th NAFIPS International Conference, 3, 1506-1510. Fayyad, U., Shapiro, G. P. , & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 31(7), 27-34. Han, J. (1999). Data Mining. Encyclopedia of Distributed Computing, Kluwer Academic Publishers, 1-7. Helman, P., & Liepins, G. (1993). Statistical Foundations of Audit Trail Analysis for the Detection of Computer Misuse. IEEE Software Engineering. 14(5), September – October. Kumar, S., & Spafford, E. (1994). A Pattern Matching Model for Misuse Intrusion Detection. the 17th National Computer Security Conference. Marin, J. A., Ragsdale, D. J., & Surdu, J. R. (2001). A Hybrid Approach to Profile Creation and Intrusion Detection. Proceedings of the DARPA Information Survivability Conference and Exposition - DISCEX, 69-76. MIT Lincoln Laboratory - DARPA Intrusion Detection Evaluation. Available: http://www.ll.mit.edu/IST/ideval/index.html [1997]. Piatetsky-Shapiro, G., & Frawley, W. J. (1991). Knowledge Discovery in Databases. AAAI/MIT Press. Portnoy, L., Eskin, E., & Stolfo, S. J.(2001). Intrusion Detection with Unlabeled Data Using Clustering. Proceedings of the ACM CCS Workshop on Data Mining for Security Applications. Smith, R., Bivens, A., & Embrechts, M.(2002). Clustering Approaches for Anomaly Based Intrusion Detection. Proceedings of the Walter Lincoln Hawkins Graduate Research Conference. Sundaram, A. An Introduction to Intrusion Detection, ACM Crossroads Student Magazine. Available: http://www.acm.org/ crossroads/xrds2-4/intrus.html [No date]. Symantec. Available: http://www.symantec.com/index.htm [2007]. Tsaur ,W. J., & Fan, I M. (2002). Anomaly Detection Mechanisms for Web Severs in Linux Environments. Communications of the CCISA, 8(4). Qin, M, & Hwang, K, (2004). Frequent Episode Rules for Internet Anomaly Detection. Proceedings of The Third IEEE International Symposium on Network Computing and Applications, 161-168. Quinlan, J.R., (1993). C4.5 Programs for machine learning. Morgan Kaufmann Publishers, San Mateo, California. Zadeh, L. A., (1965). Fuzzy sets. Information and Control , 8, 338-353. Zadeh, L. A., (1975). The concept of a linguistic variable and its application to approximate reasoning I, II, III. Information Science , 8 , 199-251 , 301-357 ; 9, 43-80. Zaiane, O. R., Xin, M., & Han, J. (1998). Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs. Proceedings of Advances in Digital Libraries Conference (ADL- 98), 19-29.