A Study on Applying Effective Cluster Analysis Schemes to Intrusion Detection Systems

林東森、曹偉駿

E-mail: 9222744@mail.dyu.edu.tw

ABSTRACT

In recent years, along with the popularity of Internet, the pervasion of the computer usage and the maturity of the network technique, it makes Internet become the media of commercial transaction, and attract the attention from society. On the contrary, it also causes hackers' attacks and a variety of network crimes. The potential damage that caused by the intrusion is not only difficult to estimate, but expose the security problem existing in the system. The host-based intrusion detection system is to defend and detect attack behaviors from hacker via log files in the server. If we can analyze log files precisely and efficiently, it will be useful to establish an intrusion detection system. Therefore, in this thesis we mainly analyze log files by using clustering and classification methods to distinguish normal and abnormal behaviors. Although the clustering method can analyze huge log files, it generally needs to set up a known cluster number in advance before starting clustering; however, it is difficult for us to decide the known number. If we employ a cluster validity index to evaluate automatically the number of clusters, we can acquire better results. When the amount of log files are gradually added, clustering procedure needs to be run again. If we can adopt the advantage of classification, we can handle cumulative log files efficiently. Based on the requirements stated above, we integrate a faster clustering algorithm into the cluster validity index to solve the cluster analysis problems, and further use the classification rule of modified K-nearest neighbor to handle the increasingly cumulative log files. The results derived in this thesis can be used to construct the preceding part of an intrusion detection system. Furthermore, we also develop a practical system. Through this implemented system, we can cluster and classify log files, and validate the feasibility of the proposed methods in this thesis.

Keywords : Intrusion detection system, Clustering, Cluster validity index, Classification.

Table of Contents

REFERENCES

[1] 李志明，影像最佳類別數目之研究，國立中央大學碩士論文，民國90年6月。

[2] 范義明，Linux作業系統下異常偵測之研究，大葉大學碩士論文，民國91年6月。

[3] 許文豪，圖形辨識概述與實作，國立清華大學碩士論文，民國89年7月。

[4] 黃景彰，資訊安全-電子商務之基礎，華泰文化事業公司，民國90年6月。

[5] 賴溪松、葉育斌，資訊安全入門，全華科技圖書，民國90年6月。

[6] 譚嘉慧，模糊分類適切性分析，中原大學碩士論文，民國89年6月。

[7] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.

[8] R. H. Charles, Cluster analysis for researchers, 茂昌圖書有限公司, 1985.

[9] T. W. Cheng, D. B. Goldgof, and L. O. Hall, Fast Fuzzy Clustering, Fuzzy Sets and Systems, Vol. 93, pp. 49-56, 1998.

[10] V. Cherkassky, and F. Mulier, Learning from Data : Concepts, theory, and methods. New York : Wiley, 1998.

[11] Y. Fukayama and M. Sugeno, A New Method of Choosing the Number of Cluster for the Fuzzy c-mean Method, Proceedings of 5th Fuzzy Systems Symposium, pp. 247-250, 1989.

[12] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press, 1990.

[13] J. S. Jang, C. T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing, Prentice Hall, New Jersey, 1997.

[14] M. F. Jiang, S. S. Tseng, and C. M. Su, Two-Phase Clustering Process for Outliers Detection, Pattern Recognition Letters, Vol. 22, pp.

691-700, 2001.

[15] R. A. Johnson, and D. W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, New Jersey, 1998.

[16] D. J. Kim, Y. W. Park, and D. J. Park, A Novel Validity Index for Determination of the Optimal Number of Clusters, IEICE Transactions on Information and Systems Society, Vol. E84-D, No. 2, pp. 281-285, 2001.

[17] NetCraft http://www.netcraft.com/ [18] N. R. Pal and J. C. Bezdek, On Cluster Validity for the Fuzzy c-means Model, IEEE Transactions on Fuzzy Systems, Vol. 3, No. 3, 1995.

[19] N. R. Pal and J. C. Bezdek, Correspondence to "On Cluster Validity for the Fuzzy c-means Model", IEEE Transactions on Fuzzy Systems, Vol. 5, No. 1, 1997.

[20] Superscan http://www.foundstone.com/ [21] A. Sundaram, An Introduction to Intrusion Detection, ACM Crossroads Student Magazine. http://www.acm.org/crossroads/xrds2-4/intrus.html [22] X. L. Xie, and G. Beni, A Validity Measure for Fuzzy Clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 841-847, 1991.

[23] L. A. Zadeh, Fuzzy Sets, Information Control, Vol. 8, pp. 338-353, 1965.