適用於入侵偵測之高準確度階層式分群演算法設計與實作

游錦昌、曹偉駿

E-mail: 9422463@mail.dyu.edu.tw

摘 要

隨網際網路的發展迅速，駭客人數也不斷的快速成長，要如何維護資訊安全，避免駭客的入侵，成了當下相當重要的議題。為了防護網路入侵行為，許多的防護軟體或方法相繼被提出，在這些方法中最受人注目的便是「入侵偵測系統」。然而，入侵偵測系統發展至今約二十餘年，卻還是無法非常有效的應用在現今網路環境上，探討其原因，無非是因為現有的入侵偵測系統偵測率過低且誤報率過高，其中誤報率過高的問題更是讓管理人員拒絕使用入侵偵測系統的主因。為了提升入侵偵測系統之偵測率及降低其誤報率，本研究之具體作法為設計一適用於入侵偵測之高準確度階層式分群演算法，該演算法可適用於分群網路封包，而結果將比現有之分群方法擁有更高的準確度。本論文從入侵偵測系統之分群演算法著手改進，讓群集分析的結果可以更準確的分析出正、異常群集，藉此提昇入侵偵測系統之判斷能力。

關鍵詞 : 入侵偵測系統 ; 分群演算法 ; 偵測率 ; 誤報率

目錄

參考文獻

[1] CERT/CC Statistics 1988-2004, http://www.cert.org/stats/cert_ stats.html.

[2] M.R. Anderberg, Cluster Analysis for Applications, Academic Press, New York, 1973.

[3] R.A. Baezda-Yates, Introduction to Data Structures and Algorithms Related to Information Retrieval, W. B. Frakes and R. Baeza-Yates, Eds., Prentice-Hall, New Jersey, 1992.

[4] M. Berry and G. Linoff, Data Mining Techniques: for Marketing, Sales, and customer support, Published by Arrangement with Wei Keg Publishing Co., 1997.

[5] M. Dash, H. Liu. and X. Xu.,“`1+1>2': Merging Distance and Density Based Clustering,” Database Systems for Advanced Applications (DASFAA'01), pp. 18-20, 2001.

[6] R. Durst, T. Champion, B. Witten, E. Miller and L. Spagnuolo,“ Testing and Evaluating Computer Intrusion Detection System,” Communications of the ACM, Vol. 47, No. 7, pp. 53-61, 1999.

[7] M. Ester, H.P. Kriegel, J. Sander and X. Xu ,“ Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications,” Data Mining and Knowledge Discovery, Vol. 2, No. 2, pp. 169-194, 1998.

[8] K.C. Gowda and E. Diday,“ Symbolic Clustering Using a New Dissimilarity Measure,” Pattern Recognition, Vol. 24, No. 6, pp. 567-578. 1991.

[9] K.C. Gowda and G. Krishna, “ Agglomerative Clustering Using The Concept of Mutual Nearest Neighborhood, ” Pattern Recognition, Vol. 10, No. 2, pp. 105-112, 1978.
[10] Y. Guan, A. Ghorbani and N. Belacel, “ Y-Means: A Clustering Method for Intrusion Detection, ” Canadian Conference on Electrical and Computer Engineering, pp. 1083-1086, 2003.
[11] H. Han, X.L. Lu and L.Y. Ren, “ Using Data Mining to Discover Signatures in Network-Based Intrusion Detection, ” Machine Learning and Cybernetics, Vol. 1, pp. 13-17, 2002.
[12] P. Hansen and N. Mladenovi?, “ J-Means: A New Local Search Heuristic for Minimum Sum of Squares Clustering, ” Pattern Recognition,Vol. 34, No. 2, pp. 405-413, 2001.
[13] R. Heady, G. Luger, A. Maccabe and M. Servilla, “ The Architecture of A Network Level Intrusion Detection System, ” Technical report CS90-20, 1990.
[14] S. Hirano, X. Sun and S. Tsumoto, “ Comparison of Clustering Methods for Clinical Databases, ” Information Sciences, Vol. 159, No. 3-4, pp. 155-165, 2004.
[15] A.K. Jain and R.C. Dubes, Algorithms for Clustering Data, Prentice-Hall, New Jersey, 1988.
[16] A.K. Jain, M.N. Murty and P.J. Flynn, “ Data Clustering: A Review, ” ACM Computing Surveys, Vol. 31, No. 3, pp. 264-323, 1999.
[17] J.S. Jang, C.T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing, Prentice-Hall, New Jersey, 1997.
[18] M.F. Jiang, S.S. Tseng, and C.M. Su, “ Two-Phase Clustering Process for Outliers Detection, ” Pattern Recognition Letters, Vol. 22, No. 6, pp. 691-700, 2001.
[19] R.A. Johnson, and D.W. Wichern, Applied Multivariate Statistical Analysis, Prentice-Hall, New Jersey, 1998.
[20] 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.
[21] R.J. Kuo, L.M. Ho and C.M. Hu, “ Integration of self-organizing feature map and K-means algorithm for market segmentation, ” Computers and Operations Research, Vol. 29, No. 11, pp. 1475-1493, 2002.
[22] M.J. Laan and K.S. Pollard, “ A New Algorithm for Hybrid Hierarchical Clustering with Visualization and The Bootstrap, ” Journal of Statistical Planning and Inference, Vol. 117, No. 2, pp. 275-303, 2003.
[23] R. Lippmann, R.K. Cunningham, “ Guide to Creating Stealthy Attacks for The 1999 DARPA Off-Line Intrusion Detection Evaluation, ” MIT Lincoln Laboratory Project Report IDDE-1, 1999.
[24] R. Lippmann, D. Fried, I. Graf, J. Haines, K. Kendall, D. McClung, D. Weber, S. Webster, D. Wyschogrod, R. Cunningham, M. Zissman, “ Evaluation Intrusion Detection Systems: The 1998 DARPA Off-Line Intrusion Detection Evaluation, ” Proceedings of the 2000 DARPA Information Survivability Conference and Exposition (DISCEX), IEEE Computer Society, Vol. 2, pp. 12-26, 2000.
[25] R. Lippmann, J.W. Haines, D.J. Fried, J. Korba and K. Das, “ The 1999 DARPA Off-Line Intrusion Detection Evaluation, ” Computer Networks, Vol. 34, No. 4, pp. 579-595, 2000.
[26] Y. Liu, K. Chen, X. Liao and W. Zhang, “ A Genetic Clustering Method for Intrusion Detection, ” Pattern Recognition, Vol. 37, No. 5, pp. 927-942, 2004.
[27] J. Mao and A. Jain, “ A Self-Organizing Network for HyperEllipsoidal Clustering, ” IEEE Transactions on Neural Networks, Vol. 7 No. 1, pp. 16-29, 1996.
[28] R. Michalski, R.E. Stepp and E. Diday, “ Automated Construction of Classifications: Conceptual Clustering Versus Numerical Taxonomy, ” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 5, No. 4, pp. 396-409, 1983.
[29] S.J Oh and J.Y. Kim, “ A Hierarchical Clustering Algorithm for Categorical Sequence Data, ” Information Processing Letters, Vol. 91, No. 3, pp. 135-140, 2004.
[30] 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, pp. 370-379, 1997.
[31] V. Paxson and S. Floyd, “ Wide-Area Traffic: The Failure of Poisson Modeling, ” IEEE/ACM Transactions on Networking, Vol. 3, No. 3, pp. 226-244, 1995.
[32] W.J. Tsaur and I.M. Fan, “ Anomaly Detection Mechanisms for Web Severs in Linux Environments, ” Communications of the CCISA, Vol. 8, No.4, pp. 59-78, 2002.
[33] W.J. Tsaur and D.S. Lin, “ Applying Effective Cluster Analysis Schemes to Intrusion Detection Systems, ” Industry Forum, Vol. 7, No. 2, pp. 213-236, 2005.
[34] W.J. Tsaur and Y.C. Shieh, “ Fuzzy Association Rules Mechanism for Intrusion Detection Systems, ” The 2003 National Computer Symposium (NCS ’ 03), pp. 1256-1263, 2003.
[35] X. L. Xie, and G. Beni, “ A Validity Measure for Fuzzy Clustering, ” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 8, pp. 841-847, 1991.
[36] Y. Zhang and V. Paxson, “ Detecting Backdoors, ” Proceedings of the 9th USENIX Security Symposium, pp. 157-170, 2000.
[37] T. Zhang, R. Ramakrishnan and M. Livny, “ BIRCH: An Efficient Data Clustering Method for Very Large Databases, ” SIGMOD

Conference, pp. 103-114, 1996.
[38] M. Zissman, DARPA Intrusion Detection Evaluation, http://ww w.ll.mit.edu/IST/ideval/index.html, 2000.