Design and Implementation of Highly Accurate Hierarchical Clustering Algorithm for Intrusion Detection

游錦昌、曹偉駿

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

ABSTRACT

With the growth of Internet, the number of hackers is increasing. Therefore, how to protect information security and avoid intrusions is an important issue. In order to prevent the behavior of intrusion to Internet, many software tools or methods such as intrusion detection systems have been proposing. However, in the past twenty years, the operation of intrusion detection systems still cannot be efficient. The reason is that existing intrusion detection systems are still with low detection rate and high false positive. Especially, high false positive lets system managers refuse to use intrusion detection systems. Therefore, in order to increase the effectiveness of intrusion detection and reduce the false positive, we propose a hierarchical clustering algorithm for intrusion detection. Our proposed method is the highly accurate hierarchical clustering algorithm, which is suitable for clustering network packets. The proposed clustering algorithm can accurately generate normal and abnormal clusters, and is more efficient and accurate than existing clustering methods.

Keywords : Intrusion Detection System ; Clustering Algorithm ; Detection Rate ; False Positive

Table of Contents

REFERENCES

[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.