| | |
|---|---|
| P.O.Box 8718, Beijing 100080, China | Journal of Software，June 2007,18(7):1639-1651 |
| E-mail: jos@iscas.ac.cn | ISSN 1000-9825，CODEN RUXUEW，CN 11-2560/TP |
| http://www.jos.org.cn | Copyright © 2007 by *Journal of Software* |

# 基于特征选择的轻量级入侵检测系统

陈 友, 程学旗, 李 洋, 戴 磊

Full-Text PDF Submission Back

陈 友1,2, 程学旗1, 李 洋1,2, 戴 磊1,2
1(中国科学院 计算技术研究所,北京 100080)
2(中国科学院 研究生院,北京 100049)
作者简介: 陈友(1981－),男,安徽安庆人,博士,主要研究领域为网络安全,数据挖掘.程学旗(1971－),男,博士,研究员,博士生导师,CCF高级会员,主要研究领域为网络信息安全,大规模信息检索与信息挖掘,P2P计算.李洋(1978－),男,博士,主要研究领域为入侵检测,网络攻防对抗技术.戴磊(1979－),男,博士,主要研究领域为信息安全.
联系人: 陈 友 Phn: +86-10-62600949, E-mail: chenyou04@mails.gucas.ac.cn

Received 2006-10-08; Accepted 2007-03-26

## Abstract

The intrusion detection system based on feature selection deals with huge amount of data which contains redundant and noisy features causing slow training and testing process, high resource consumption as well as poor detection rate. Feature selection, therefore, is an important issue in intrusion detection and it can delete redundant and noisy features. In order to improve performances of intrusion detection system in terms of detection speed and detection rate, a survey of intrusion detection system based on feature selection is necessary. This paper introduces the concepts and algorithms of feature selection, surveys the existing lightweight intrusion detection systems based on feature selection algorithms, groups and compares different systems in three broad categories: filter, wrapper, and hybrid. This paper concludes the survey by identifying trends of feature selection research and development in intrusion detection system. Feature selection is not only useful for intrusion detection system, but also helpful to provide a new research direction for intrusion detection system.

Chen Y, Cheng XQ, Li Y, Dai L. Lightweight intrusion detection system based on feature selection. Journal of Software, 2007,18(7):1639-1651.
DOI: 10.1360/jos181639
http://www.jos.org.cn/1000-9825/18/1639.htm

## 摘要

基于特征选择的入侵检测系统处理的数据含有大量的冗余与噪音特征,使得系统耗用的计算资源很大,导致系统训练时间长、实时性差,检测效果不好.特征选择算法能够很好地消除冗余和噪音特征,为了提高入侵检测系统的检测速度和效果,对基于特征选择的入侵检测系统进行研究是必要的.综述了这一领域的研究进展,从过滤器、封装器、混合器3种模式对基于特征选择的轻量级入侵检测系统进行分类比较,分析和总结各种系统的优缺点以及它们各自适用的条件,最后指出入侵检测领域特征选择的发展趋势.特征选择不仅可以提升入侵检测系统的性能,而且使得对入侵检测的研究向特征提取算法的方向转移.

基金项目：Supported by the National Basic Research Program of China under Grant No.2004CB318109 (国家重点基础研究发展计划(973)); the National Information Security 242 Project of China under Grant No.2005C39 (国家242信息安全计划)

## References:

[1] Forres S, Perelson AS, Allen L, Cherukun R. Self-Nonself discrimination in a computer. In: Rushby J, Meadows C, eds. Proc. of the '94 IEEE Symp. on Research in Security and Privacy. Los Alamitos: IEEE Computer Society Press, 1994. 120-128.

[2] Narendra PM, Fukunaga K. A branch and bound algorithm for feature subset selection. IEEE Trans. on Computer, 1977,26(9): 917-922.

[3] Zhou R, Hansen E. Breadth-First heuristic search. Artificial Intelligence, 2006,170(4-5):385-408.

[4] Gheorghies O, Luchian H, Gheorghies A. A study of adaptation and random search in genetic algorithms. In: Proc. of the 2006 IEEE Congress on Evolutionary Computation Sheraton Vancouver Wall Centre Hotel. Vancouver: IEEE Computer Society Press, 2006. 2103-2110. http://ieeexplore.ieee.org/xpls/abs_all.jsp-arnumber=1688566

[5] Bhuyan J. A combination of genetic algorithm and simulated evolution techniques for clustering. In: Proc. of the ACM 23rd Annual Conf. on Computer Science. Nashville: ACM, 1995. 127-134. http://www.informatik.uni-trier.de/~ley/db/conf/acm/csc95.html#Bhuyan95

[6] Liu H, Setiono R. A probabilistic approach to feature selection: A filter solution. In: Proc. of the 13th Int'l Conf. on Machine Learning. 1996. 319-327. http://www.public.asu.edu/~huanliu/publications.html

[7] Das S. Filters, wrappers and a boosting based hybrid for feature selection. In: Brodley C, Danyluk A, eds. Proc of the 8th Int'l Conf. on Machine Learning. San Francisco: Morgan Kaufmann Publishers, 2001. 74-81.

[8] Yuan H, Tseng SS, Wu GS, Zhang FY. A two-phase feature selection method using both filter and wrapper. In: Proc of the '99 IEEE Int'l Conf. on Systems, Man, and Cybernetics. Piscataway: IEEE Computer Society Press, 1999. 132-136.

[9] Kohavi R, John GH. Wrappers for feature subset selection. Artificial Intelligence Journal, 1997,97(1-2):273-324.

[10] Almuallim H, Dietterich TG. Learning boolean concepts in the presence of many irrelevant features. Artificial Intelligence, 1994, 69(1-2):279-305.

[11] Teodoro ML, Phillips GN, Jr Kavraki LE. A dimensionality reduction approach to modeling protein flexibility. In: Proc. of the 6th Annual Int'l Conf. on Computational Biology. Washington: ACM, 2002. 299-308. http://citeseer.ist.psu.edu/teodoro02dimensionality.html

[12] Hall MA. Correlation-Based feature selection for discrete and numeric class machine learning. In: Langley P, et al., eds. Proc. of the 17th Int'l Conf. Machine Learning. San Francisco: Morgan Kaufmann Publishers, 2000. 359-366.

[13] Liu H, Yu L. Towards integrating feature selection algorithms for classification and clustering. IEEE Trans. on Knowledge and Data Engineering, 2005,17(4):491-502.

[14] Yu L, Liu H. Efficient feature selection via analysis of relevance and redundancy. Journal of Machine Learning Research, 2004, 5 (10):1205-1224.

[15] Baglioni M, Furletti B, Turini F. DrC4.5: Improving C4.5 by means of prior knowledge. In: Proc. of the 2005 ACM Symp. on Applied Computing. Santa Fe: ACM, 2005. 474-481. http://www.informatik.uni-trier.de/~ley/db/conf/sac/sac2005.html#BaglioniFT05

[16] Fugate M, Gattiker JR. Anomaly detection enhanced classification in computer intrusion detection. LNCS 2388, Berlin, Heidelberg: Springer-Verlag, 2002. 186-197.

[17] Kim DS, Park JS. Network-Based intrusion detection with support vector machines. LNCS 2662, Berlin, Heidelberg: Springer- Verlag, 2003. 747-756.

[18] Beverly R, Sollins K, Berger A. SVM learning of IP address structure for latency prediction. In: Proc. of the 2006 SIGCOMM Workshop on Mining Network Data. Pisa: ACM, 2006. 1-6. http://www.sigcomm.org/sigcomm2006/papers/minenet-04.pdf

[19] Joachims T. Making large-scale SVM learning practical. In: Schlkopf B, Burges C, Smola A, eds. Advances in Kernel Methods—Support Vector Learning. Cambridge: MIT Press, 1999. 1-13.

[20] Sung AH, Mukkamala S. Identifying important features for intrusion detection using support vector machines and neural networks. In: Proc. of the 2003 Int'l Symp. on Applications and the Internet Technology. Orlando: IEEE Computer Society Press, 2003. 209-216. http://www.computer.org/portal/site/store/menuitem.41cf17dc879177c86ee948ce8bcd45f3/index.jsp-&pName=store_level1&path=store/catalog&file=pr01872.xml&xsl=generic.xsl&

[21] Reed R. Pruning algorithms—A survey. IEEE Trans. on Neural Network, 1993,4(3):740-662.

[22] Park JS, Shazzad KM, Kim DS. Toward modeling lightweight intrusion detection system through correlation-based hybrid feature selection. In: Feng D, Lin D, Yung M, eds. Proc. of the CISC. Heidelberg: Springer-Verlag, 2005. 279-289.

[23] Kim DS, Nguyen HN, Ohn SY, Park JS. Fusions of GA and SVM for anomaly detection in intrusion detection system. In: Advances in Neural Networks. LNCS 3498, New York, Berlin, Heidelberg: Springer-Verlag, 2005. 415-420.

[24] Kompella R, Singh S, Varghese G. On scalable attack detection in the network. In: Proc. of the 4th ACM SIGCOMM Conf. on Internet Measurement. Washington: ACM, 2004. 187-200. http://citeseer.ist.psu.edu/kompella04scalable.html

[25] Rao X, Dong CX, Yang SQ. An intrusion detection system based on support vector machine. Journal of Software, 2003,14(4): 798-803 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/14/798.htm

[26] KDD cup 1999 data. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[27] http://www.cs.waikato.ac.nz/ml/weka/index.html

[28] Chen Y, Li Y, Cheng XQ, Guo L. Survey and taxonomy of feature selection algorithms in intrusion detection system. In: Lipmaa H, Yung M, Lin D, eds. Proc. of the Conf. on Information Security and Cryptology. LNCS 4318, Berlin, Heidelberg: Springer-Verlag, 2006. 153-167.

[29] Chen Y, Dai L, Li Y, Cheng XQ. Building efficient intrusion detection model based on principal component analysis and C4.5 algorithm. In: Proc. of the 9th IEEE Int'l Conf. on Advanced Communication Technology. Korea: IEEE Computer Society Press, 2007. 2109-2112. http://www.icact.org/

[30] Taylor C, Alves-Foss J. NATE: Network analysis of anomalous traffic events, a low-cost approach. In: Proc. of the 2001 Workshop on New Security Paradigms. New Mexico: ACM, 2001. 89-96. http://www.csds.uidaho.edu/papers/Taylor01a.pdf

附中文参考文献:
[25] 饶鲜,董春曦,杨绍全.基于支持向量机的入侵检测系统.软件学报,2003,14(4):798-803. http://www.jos.org.cn/1000-9825/14/798.htm