Minnesota Intrusion Detection System (MINDS)

Varun Chandola

1 Introduction

The *Minnesota INtrusion Detection System* (MINDS[1]) is an end-to-end data mining based intrusion detection solution that has been shown to be effective in detecting cyber intrusions on large scale networks [4, 7].

2 MINDS Architecture

The MINDS system consists of four key components:

Flow Converter This component converts the network data to internal flow format used by the MINDS system using specific converters developed for different types of network flow formats. Currently two converters are implemented that handle *CISCO netflow* format and *tcpdump* format.

Scan Detector This component labels each flow as a potential scan or not using a novel scan detection technique [8].

P2P Detector This component labels each flow as potential P2P or not using a set of heuristics and information regarding known P2P and good ports.

Anomaly Detector This component assigns an anomaly score to each flow using an unsupervised anomaly detection, known as *local outlier factor* (lof) [2]. Each flow is defined using a set of binary, categorical, and continuous features that are instrumental in differentiating between normal and intrusive flows [4]. The lof technique uses a distance measure between a pair of flows. To handle the categorical features, MINDS uses a novel data driven distance measure [1, 3].

Summarization This component provides a compact and informative summary of the top anomalous flows detected by the anomaly detector using a novel summarization technique [5, 6].

3 MINDS Software

The MINDS system is written in GNU C++ and Perl and tested extensively on Linux, Sun Solaris, FreeBSD, and Cygwin. The distribution is available as source as well as binaries which can run on ix86 architectures.

References

[1] S. Boriah, V. Chandola, and V. Kumar. Similarity measures for categorical data: A comparative evaluation. In *Proceedings of the eighth SIAM International Conference on Data Mining*, pages 243–254, 2008.

[1] http://www.cs.umn.edu/research/MINDS

[2] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. Lof: identifying density-based local outliers. In *Proceedings of 2000 ACM SIGMOD International Conference on Management of Data*, pages 93–104. ACM Press, 2000.

[3] V. Chandola, S. Boriah, and V. Kumar. A framework for exploring categorical data. In *Proceedings of the ninth SIAM International Conference on Data Mining*, 2009.

[4] V. Chandola, E. Eilertson, L. Ertoz, G. Simon, and V. Kumar. Data mining for cyber security. In A. Singhal, editor, *Data Warehousing and Data Mining Techniques for Computer Security*. Springer, 2006.

[5] V. Chandola and V. Kumar. Summarization – compressing data into an informative representation. In *Fifth IEEE International Conference on Data Mining*, pages 98–105, Houston, TX, November 2005.

[6] V. Chandola and V. Kumar. Summarization – compressing data into an informative representation. *Knowledge and Information Systems*, 12(3), August 2007.

[7] L. Ertoz, E. Eilertson, A. Lazarevic, P.-N. Tan, V. Kumar, J. Srivastava, and P. Dokas. MINDS - Minnesota Intrusion Detection System. In *Data Mining - Next Generation Challenges and Future Directions*. MIT Press, 2004.

[8] G. J. Simon, H. Xiong, E. Eilertson, and V. Kumar. Scan detection: A data mining approach. In *Sixth SIAM International Conference on Data Mining*, pages 118–129, 2006.