Framework for Analysis of Anomalies in the Network Traffic

L.S. Silva[1], T.D. Mancilha[2], J.D.S. Silva[3], A.C.F. Santos[4], e A. Montes[5]

[1,2]Ground Segment Development Division - DSS
[3,4]Laboratory for Computing and Applied Mathematics – LAC
Brazilian National Institute for Space Research - INPE
C. Postal 515 – 12245-970 – São José dos Campos – SP - BRASIL
[5]Research Center Renato Archer – CenPRA

Rodovia Dom Pedro I, km 143,6 - 13069-901- Campinas – SP - BRASIL
E-mail: lilia@dss.inpe.br, thiago@dss.inpe.br, demisio@lac.inpe.br, adriana.ferrari@lac.inpe.br, antonio.montes@cenpra.gov.br

Keywords: anomaly detection, anomaly analysis, network traffic analysis, network traffic visualization, traffic anomalies.

Abstract

In this paper, a framework in development at INPE for analysis of anomalies in the network traffic is presented. The stages of this work, including the preparation of the environment for tests and development, data selection, data capture, data reduction and data presentation are described. Also, techniques and tools used in each stage will be approached.

1. Introduction

Tools for network traffic analysis allow the detection of anomalies in the environment, including attacks and not usual actions in the network, and enable fast execution of actions to avoid that detected threats propagate through the network.

The network traffic should be monitored in regular intervals to obtain data that will be analyzed by statistical or intelligent techniques for identifying anomalies. The idea is to store data from normal traffic for future comparison with current data searching eventual anomalies.

In this work, network traffic data from session TCP/IP packets, more specifically, deriving from HTTP communication between client and server machines. These data belong to large sets, are found in several types, and are stored in different scale and measure units. Considering this context, the use of techniques to reduce the very large dataset and tools to present data on the computer screen are necessary.

The techniques used to reduce the traffic data volume for analysis and the tools applied to visualize the traffic will be presented in the next section.

2. Framework for Anomaly Analysis

The implemented framework comprises the following stages: preparation of the development and test environment, data selection, data capture, data reduction and data presentation.

The traffic behavior from two networks, test network and production network, is analyzed and the data collection is performed through mechanisms of traffic monitoring installed. Test network corresponds to the network used for application of simulated attacks, traffic monitoring and tests. Production network considered in this work is an internal network at INPE whose data are captured by means of a network sensor installed outside its boundary and observed [4].

Both network traffics are monitored using tools like *tcpdump*, *ethereal* and monitoring scripts. Network packet data are remounted in sessions and stored in a database through the RECON system [2].

Packet header data are used for mapping the network traffic and they are classified as primitive and derived attributes or features [5]. Primitive attributes are directly obtained from packet fields, such as: source and destination IP addresses, source and destination ports, and service type in use. Derived attributes are constructed with basis on the primitive ones. Derived attributes carry stronger semantically information for the traffic mapping. The derived attributes [3] used in this work are: medium size of network packets received by the client (in bytes); medium size of network packets received by the server (in bytes); number of packets received by the client; number of packets received by the server; small packet rate or packet with size less than 130 bytes (%); traffic direction; data bytes received by the client; data bytes received by the server; and session duration.

A set of these attributes is obtained from each database register and represent only one network session or connection record. A TCP/IP network session can be defined as any sequence of packets, that characterizes a information exchange between two IP addresses, and that has information of beginning, middle and end (even so all communication be resident in a unique packet). A set of sessions represents the network traffic observed in a given instant.

In order to reduce the traffic data, enabling a visual analysis, clustering techniques based on neural networks are being applied. Traffic sessions with similar behavior participate of the same cluster. Further, reduced data are exhibited in a graphical way on the computer screen using the Matlab tool.

The graphical representation of the network traffic permits to obtain better insight and understanding of its behavior, facilitating the identification of anomalies, because pictures can convey an overall message much better than a list of numbers. Some work have been conducted in this research line [6][7][8][9].

A network traffic visualization tool named RGCom [1] is being developed at INPE. This application performs data reading from a database, data normalization, and data plotting in parallel coordinates on the computer screen. Graphical and database communication resources from Java programming environment are used in its implementation.

The graph produced by RGCom contains nine parallel coordinates with values of a determined attribute each one. Nine attribute points of a same session are plotted in that axis and they are interconnected, shaping a session line. All lines of one happened session in a selected by the user date and time interval are drawn and the resultant graph represents the network traffic behavior in that time.

4. Conclusion

The mentioned framework for analysis of anomalies in the network traffic has concluded the following stages preparation of development and test environment, data selection and data capture. Clustering techniques based on neural networks used for data reduction are being tested and analyzed. In despite of these one perform a significant data reduction, they need to be improved, because the formed clusters do not satisfactorily represent all traffic sessions with similar characteristics.

RGCom does not perform data reduction. But, at this moment, the more important objective of this tool is to permit that combinations of attributes better describing the network traffic behavior be selected, making the anomaly identification task easier.

REFERENCES

[1] Mancilha, T.D, Silva, L.S, Salgado, A.E.M, Montes, A. and Paula, A. R. (2006), *Desenvolvimento em Java de uma Ferramenta de Visualização Gráfica do Tráfego de Rede*, Paper a ser apresentado no X Encontro Latino Americano de Iniciação Científica – Universidade do Vale do Paraíba, São Jose dos Campos, SP.

[2] Chaves, M.H.P (2002), *Analise de Estado do Trafego de Redes TCP/IP para Aplicação em Detecção de Intrusão*, Dissertação de mestrado em Computação Aplicada, INPE, São Jose dos Campos, SP.

[3] Chaves, C.H.P.C and Montes, A. (2005), *Detecção de Backdoors e Canais Dissimulados*, V Workshop dos cursos de Computação Aplicada (Worcap'2005), INPE, São José dos Campos, SP.

[4] Silva, L.S., Montes, A. and Silva, J.D.S (2005), *Evolução dos Trabalhos em Detecção de Anomalias na Rede*, V Workshop dos cursos de Computação Aplicada (Worcap'2005), INPE, São José dos Campos, SP.

[5] Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Dokas, P., Srivastava, J. and Kumar, V. (2004), *Detection and Summarization of Novel Network Attacks Using Data Mining*, disponível em: http://www.cs.umn.edu/research/minds/papers/raid03.pdf, acessado em julho 2004.

[6] Kim, S.S. and Reddy, A. L. N. (2005), *A Study of Analyzing Network traffic as Images in Real-Time*, Department of Electrical Engineering ,Texas A&M University.

[7] Kim, S.S.; Reddy A. L. N. and Vannucci M. (2004), *Detecting Traffic Anomalies through Aggregate Analysis of Packet Header Data*, Proceedings of Networking 2004, LNCS 3042, pp 1047-1059, Athens, Greece.

[8] Kim, S.S. and Reddy, A. L. N. (2005), *Modeling Network traffic as Images*, Proceedings of IEEE International Conference on Communications, Seoul Korea.

[9] Barford P., Kline J.; Plonka D. and Ron A. (2002), *A Signal Analysis of Network Traffic Anomalies*, Proceedings of ACM SIGCOMM Internet Measurement Workshop, Marseille, France.