Applied Soft Computing 11 (2011) 2042–2056

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Neural visualization of network traffic data for intrusion detection

Emilio Corchado [a,*], Álvaro Herrero [b,1]

[a] Departamento de Informática y Automática, Universidad de Salamanca, Plaza de la Merced s/n, 37008, Salamanca, Spain
[b] Department of Civil Engineering, University of Burgos, C/Francisco de Vitoria s/n, 09006, Burgos, Spain

ARTICLE INFO

Article history:
Received 30 May 2009
Received in revised form 17 January 2010
Accepted 5 July 2010
Available online 2 August 2010

Keywords:
Neural and exploratory projection techniques
Connectionist unsupervised models
Computer network security
Intrusion detection
Network traffic visualization

ABSTRACT

This study introduces and describes a novel intrusion detection system (IDS) called MOVCIDS (mobile visualization connectionist IDS). This system applies neural projection architectures to detect anomalous situations taking place in a computer network. By its advanced visualization facilities, the proposed IDS allows providing an overview of the network traffic as well as identifying anomalous situations tackled by computer networks, responding to the challenges presented by volume, dynamics and diversity of the traffic, including novel (0-day) attacks. MOVCIDS provides a novel point of view in the field of IDSs by enabling the most interesting projections (based on the fourth order statistics; the kurtosis index) of a massive traffic dataset to be extracted. These projections are then depicted through a functional and mobile visualization interface, providing visual information of the internal structure of the traffic data. The interface makes MOVCIDS accessible from any mobile device to give more accessibility to network administrators, enabling continuous visualization, monitoring and supervision of computer networks. Additionally, a novel testing technique has been developed to evaluate MOVCIDS and other IDSs employing numerical datasets. To show the performance and validate the proposed IDS, it has been tested in different real domains containing several attacks and anomalous situations. In addition, the importance of the temporal dimension on intrusion detection, and the ability of this IDS to process it, are emphasized in this work.

1. Introduction

An attack or intrusion to a network would end up affecting any of the three computer security principles: availability, integrity and confidentiality, exploiting for example the Denial of Service, Modification and Destruction vulnerabilities [1]. One of the most harmful points of attacks and intrusions, increasing the difficulty of protecting computer systems, is the ever-changing nature of attack technologies and strategies.

For this reason among others, intrusion detection systems (IDSs) have become a required asset in addition to the computer security infrastructure of most organizations. In the context of computer networks, an IDS can roughly be defined as a tool designed to detect suspicious patterns that may be related to a network or system attack. Intrusion detection (ID) is then a field focused on the identification of attempted or ongoing attacks in a computer system (host IDS—HIDS) or network (network IDS—NIDS). The accurate detection of computer and network system intrusions in real-time has always been an interesting and intriguing problem for system administrators and information security researchers. It could mainly be attributed to the dynamic nature of systems and networks, the creativity of attackers, the wide range of computer hardware and operating systems and so on. Such complexity rises when dealing with distributed network-based systems and insecure networks such as the Internet.

This study introduces an NIDS characterized by the use of an unsupervised connectionist projection technique providing a novel approach based on the visual analysis of the internal structure of the flow of traffic data. Unsupervised learning meets the ID requirements as in a real-life situation there is no target reference with which to compare the response of the network. Additionally, this soft-computing approach is quite useful for identifying unknown or not previously faced attacks, known as 0-day attacks, based on the well-known generalization capability of the artificial neural networks (ANNs).

It is important to note that the authors propose MOVCIDS (mobile visualization connectionist intrusion detection system) also as a complementary tool to other network security ones, this is, MOVCIDS can work in unison with other defence mechanisms (even if they are IDSs), to provide an intuitive depiction of both normal and anomalous traffic.

The remaining five sections of this study are structured as follows: Section 2 contains a brief state of the art of IDSs (mainly visualization-based). Section 3 describes the neural projections

* Corresponding author. Tel.: +34 923294451; fax: +34 923294514.
E-mail addresses: escorchado@ubu.es, escorchado@usal.es (E. Corchado), ahcosio@ubu.es (Á. Herrero).
[1] Tel.: +34 947 259513; fax: +34 947 259395.

1568-4946/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.asoc.2010.07.002

techniques applied in this work, while Section 4 provides an overview of the proposed IDS, in which each step forming this system is described in detail. Some experimental results are presented and described in Section 5; the proposed IDS is tested in some different ways in Section 6; authors discuss the considered main advantages of MOVCIDS in Section 7 and finally, Section 8 puts forward a number of conclusions and pointers for future work.

2. Previous work

ID has been approached from several different points of view up to now; many different intelligent and soft-computing techniques (such as genetic programming [2,3], data mining [4–10], expert systems [11,12], fuzzy logic [13,14], or neural networks [15–20] among others) together with statistical [21] and signature verification [22] techniques have been applied mainly to perform a 2-class classification (normal/anomalous or intrusive/non-intrusive). Most of these systems can generate different alarms when an anomalous situation is detected, but they cannot provide a general overview of what is happening inside a computer network.

From an opposite point of view, a great variety of visualization-based approaches to ID have been proposed as well [23–34]. In this case, the ID task is enabled by providing a visual depiction of the network or the traffic. Thus, the identification of attacks must be performed through visual features because no alarms are triggered. Visualization tools rely on the human ability to recognize different features and detect anomalies through graphical devices [35]. One of the main advantages is that apart from enabling the anomalies detection, this approach could provide a general snapshot of network traffic. As this study focuses on visualization of network traffic data rather than network structure or topology, previous work only on network data visualization is considered.

Network data are summarized in previous work by:

- *IP addresses*: that is the case of the Galaxy View of NVisionIP [36]. In [37], Border Gateway Protocol data are visualized by a diagram based on IP addresses. A matrix based on IP addresses is proposed as well in [30] to detect the propagation of the Welchia and Sasser. D worms. The Time-based Network Traffic Visualizer [31] combines a matrix display of host IP address and packets timestamp. IP segments are used in NIVA [38] to locate and colour the data.
- *Port numbers*: in [24] the main visualization proposed is based on port and time information. Stacked histograms of aggregate port activity are proposed in [25]. In the case of NVisionIP [36], the previously mentioned Galaxy View is completed by the Small Multiple View, that uses port numbers to visualize the data. By using port numbers and IP addresses, the system proposed in [25] is able to see the penetration and subsequent activity of the Sasser worm.
- *Different measurements of network traffic*: the Multi Router Traffic Grapher [26] shows the incoming/outgoing traffic in Bits per Second while IDGraphs [33] uses the number of unsuccessful connections [39].
- *Alarm data*: generated by different IDSs, such as Snort [40] or StealthWatch IDS [41].
- *Others*: additional kinds of data can be also processed by different visualization tools, such as VIAssist [42] or IDtk [28] that are applied to raw TCP packet data or alerts generated by IDS tools.

In contrast to other security tools, IDSs need to be monitored [43]. So, an IDS can be useless if nobody is looking at its outputs. In keeping with this idea, MOVCIDS goes beyond the state of the art in relation to previously mentioned visualization tools, combining features extracted from packet headers to depict each simple packet by using neural unsupervised methods based on exploratory projection pursuit (EPP) [44,45]. It provides the network administrator with a snapshot of network traffic, protocol interactions, and traffic volume generally in order to identify anomalous situations. To do so, an unsupervised neural model (see Section 3) is applied.

Most of the solutions described in this section use a glyph metaphor [28,38,46] to encode information by changing different features (colour, size, opacity, etc.) in addition to the spatial coordinates, while others use traditional representation techniques such as histograms [25,47,48], histographs [39] or other graphs [29,32]. The novel IDS proposed in this work employs the glyph metaphor as well, using different colours and shapes in addition to the spatial coordinates to offer information about the protocol each packet belongs to.

The connectionist visualization approach is not a new one; [34] proposes a visualization based on the information stored in event logs. These events are considered as multidimensional vectors, and a 2D representation of them is obtained by the self-organizing map (SOM) [49], where new (or anomalous) user activities are identified by visual comparison.

From a purely projection of packets standpoint, principal component analysis (PCA) [50,51], has been also proposed as a visualization tool for analyzing network data [23,27]. The PCA-based visualization provided in [23] does not enable to distinguish attacks from normal traffic. Furthermore, an explanation of the projection obtained by this technique is not yielded. In [27] PCA is proposed as a complementary tool to interpret the results obtained by a statistical analysis because the visualization does not allow the identification of attacks on its own. Previous work on this projection approach also includes the application of a visualization tool for intrusion detection [52]. Although some attacks are visually identified in that work by combining visualization and fuzzy feature extraction, explanations about the projection technique and the identification process are not provided.

The novel IDS presented in this study also employs scatterplot matrixes to visualize packet data and provides a proper explanation of the results obtained by projection methods such as PCA (based on the second order statistic, i.e., the variance) and also going further, applying connectionist models based on higher order statistics such as the kurtosis (which is a measure of how pointed a distribution is).

3. Unsupervised connectionist projection architectures

The identification of patterns that exist across dimensional boundaries in high-dimensional datasets is a fascinating task [44]. Such patterns may become visible if changes are made to the spatial coordinates. However, an a priori decision as to which parameters will reveal most patterns requires prior knowledge of unknown patterns.

Projection methods project high-dimensional data points onto a lower dimensional space in order to identify “interesting” directions in terms of any specific index or projection. Such indexes or projections are, for example, based on the identification of directions that account for the largest variance of a dataset – as is the case of PCA [50,51] – or the identification of higher order statistics such as the skew or kurtosis index – as is the case of exploratory projection pursuit (EPP) [44]. Having identified the most interesting projections, the data are then projected onto a lower dimensional subspace plotted in 2D or 3D, which makes it possible to examine its structure with the naked eye. The remaining dimensions are discarded as they mainly relate to a very small percentage of the information or the dataset structure. In that way, the structure identified through a multivariable dataset may be visually analyzed with greater ease. In this work, we take advantage of this dimensionality reduction ability to perform a 2D visualization of