Abstract

The collective behavior of network data segments corresponding to port 110 (E-mail) was visualized and analyzed with a set of nonlinear and computational parameters such as: Lyapunov Exponent, Fractal Dimension, Recurrence, Determinism, Grammatical Rules, and Lempel-Ziv Complexity. Common methods for network data analysis are based mainly in descriptive statistics, although useful and widely accepted they provide us with a global and partial view of the network dynamics. In this work we applied a series of analytical techniques from nonlinear and computational origin in order to elaborate a more complete view of the network dynamics for data segments from port 110. These techniques preserve information of local and global nature and their combination allows a more complete view about the collective dynamics of this network traffic. The visual representation with Visual Recurrence Analysis and Grammatical Rules is useful for the detection of changes and events in the network dynamics.

1. Introduction

The explosion of network traffic generates bottlenecks and security problems, due to these problems, is necessary a permanent monitoring of the network traffic, many tools for monitoring and analysis of different aspects of networks are available for example: tcpdump, pload, ipac, nmap, ethereal, etc. [1]. These tools use mainly descriptive statistics (mean, variance, standard deviation, frequency) in order to generate a global view of the data behavior from the network traffic, but the local information about the network traffic dynamics is missed with the statistical analysis. In the fields of time series analysis a diversity of analytical techniques have been developed: Visual Recurrence Analysis [2], Lyapunov Exponent, Phase Space Representation, Fractal Dimension [3]. These techniques based in nonlinear dynamical systems allow the analysis of complex behavior such as in network traffic. In the scientific literature there are reported the application of specific nonlinear analysis techniques such as the fractal analysis for the study of network traffic [4, 5]. Also computational metrics of complexity can be applied to sequence of data, for example: Grammatical Rules and Lempel-Ziv Complexity [6], with these techniques a characterization of the level of complexity in network traffic data can be done. The nature of the analytical techniques used in this work allows the preservation of local and global information about the network dynamics; also the integration of different attributes for the network data from different kinds of analysis allows a holistic view of their dynamic behavior. In section 2 the non linear and computational techniques are described: Visual Recurrence Analysis, Fractal Dimension, Grammatical Rules, Lempel-Ziv Complexity, etc. In section 3, the experimental method and the data under study are described. In section 4, the results of the analysis are presented. Finally, in section 5 the discussion of this work is developed.

2. Nonlinear and computational analytical techniques

2.1 Nonlinear analytical techniques

In the study of time series different behaviors can arise for example: periodical, chaotic or random dynamics; when a statistical analysis is applied to a set of data a linear behavior is assumed, but in general complex systems such as networks of computers that communicated via Internet or Intranet have a dynamics that is nonlinear, then analytical techniques such as Visual Recurrence Analysis, Lyapunov Exponent,
Fractal Dimension, etc.; can be useful for a better characterization of such systems. Next, a brief description of the nonlinear techniques applied in this work is presented.

The Visual Recurrence Analysis (VRA) is a technique that uses a graphical representation of a temporal and spatial correlation matrix (recurrence plot) derived from the differences between each datum of a data sequence compared with the rest of them. The recurrence plots are useful for the detection of hidden patterns and structural changes into the data, also allowing the identification of similarities between datasets [2]. The geometrical analysis of the recurrence plot allows the quantification of a set of attributes: Determinism, Recurrence, Shannon Entropy.

The Lyapunov Exponent is a measure of the sensibility of a system to perturbations on their initial conditions this is associated with a chaotic dynamics of the system [3, 6]. A greater Lyapunov Exponent corresponds with a greater chaotic behavior.

The Fractal Dimension measures the degree of self-similarity of a system, how a system is invariant under changes in spatial scale, allows to quantify the degree of heterogeneity of a signal for different scales [3, 6].

2.2 Computational analytical techniques

The theory of computation is the origin of other class of analytical techniques; two of these techniques were applied in the present work: Grammatical Rules and Lempel-Ziv Complexity.

The Grammatical Rules Analysis is based on the generation of a free context grammar for a sequence of data, this technique applies the Sequitur algorithm as grammar generator, the number of rules generated from a sequence of data is a measure of its computational complexity [6].

The Lempel-Ziv Complexity measures the number of different chain patterns that must be used in order to reproduce a data sequence. A sequence is analyzed from left to right and each time a new subsequence is founded a complexity counter is incremented. The complexity is the ratio between the total number of subsequences and the total length of the full sequence. This analysis measures the algorithmic complexity of a data sequence [6].

3. Description of data and methodology for their analysis.

A collection of datasets from network traffic corresponding to port 110 (E-mail) were generated by monitoring the network of the DES-Ciencias de la Información from the Universidad Autónoma del Carmen (UNACAR), this network has a extended star topology.

![Figure 1. Experimental method for analysis of network traffic data.](image)

The monitoring of the port was done with the software tool tcpdump in combination with a Perl script named gdumps.pl for setting up the initial parameters of each monitoring session, the variables that were monitoring are: IP server address, IP client address, number of packets processed by the server and by the client, size of the packets processed by the server and by the client, time length of connection acknowledge (ACK), time length of client request and answer by the server (SYN), initial send time for a packet to be processed (ITP), and time length of a packet (TLP).

The sniffer for the port 110 was active during 3 days in order to collect the data. The datasets generated were filtered with a custom Perl script named desplega2.pl in order to eliminate the registers corresponding to incomplete connections [7]. With the filtered datasets four variables were selected for analysis with nonlinear and computational techniques:
ACK, SYN, ITP, and TLP. The Figure 1 shows a flow diagram about the monitoring, processing and analysis of the data.

4. Analysis of results

The Figures 2 to 5, show the results of the application of Visual Recurrence Analysis, these figures shows the recurrence plots for the four variables analyzed: ACK, SYN, ITP, and TLP.

In the case of ACK (Figure 2) the pattern is almost uniform and have few changes this corresponds with a steady dynamics, in the case of SYN (Figure 3) its pattern show a repetitive but not periodic pattern of vertical lines (the information below and above the diagonal from left to right is the same for each recurrent plot), in the case of TLP the pattern shows a decreasing density of vertical lines corresponding with decreasing differences between their data. The recurrence plot for ITP (Figure 4) shows a richness of patterns where changes of dynamical regimes can be observed.

The Figures 6 to 9, show the behavior of the Determinism, this attribute characterize the predictability of a sequence of data, the minima observed in each graphic corresponds to data segments that show an increasing in the complexity of their dynamics that can be due to chaotic or random behavior for the variables under study, and in consequence a decreasing in their predictability, in the particular case of SYN (Figure 7) a complex behavior
can be observed, and in the others graphics changes from a stationary to a non stationary behavior in their determinism can be observed.

The Figures 10 to 13, show the behavior of the Recurrence, this attribute characterize the recurrent dynamic behavior along a sequence of data, sections of data with maximum recurrence corresponds to a dynamics that prevails in such sections. There is a correspondence between recurrence and determinism as can be observed in the figures for each variable.

The Figures 14 to 17, show the behavior of the Shannon Entropy, a decreasing of entropy corresponds
with a decreasing in the rate of information that is transmitted, the behaviors of determinism, recurrence and Shannon entropy are coupled as can be seen in the different figures for each variable.

Shannon Entropy or the Determinism corresponds to an increasing random dynamics associated to the network traffic of data segments.

The Table 1 shows the corresponding values of the maximum Lyapunov Exponent for the variables under study. In the case of SYN, it posses the greater value of the exponent this is indicative of the presence of a chaotic dynamics and in consequence a strong sensibility to small perturbations.

<table>
<thead>
<tr>
<th></th>
<th>Lyapunov Exponent values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACK</td>
<td>0.024</td>
</tr>
<tr>
<td>SYN</td>
<td>0.526</td>
</tr>
<tr>
<td>ITP</td>
<td>0.354</td>
</tr>
<tr>
<td>TLP</td>
<td>0.081</td>
</tr>
</tbody>
</table>

The Table 2 shows the Fractal Dimension for the variables under study. In the case of ITP it posses the greater value of the fractal dimension, this corresponds with the patterns observed in its recurrent plot where a self-similarity pattern can be observed, this mean that a different (time or spatial) scales the same pattern is reproduced.

<table>
<thead>
<tr>
<th></th>
<th>Fractal Dimension values.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACK</td>
<td>0.50</td>
</tr>
<tr>
<td>SYN</td>
<td>0.54</td>
</tr>
<tr>
<td>ITP</td>
<td>1.17</td>
</tr>
<tr>
<td>TLP</td>
<td>0.63</td>
</tr>
</tbody>
</table>

The Figures 19 to 22, show the evolution of the generation of grammatical rules for each variable under study. A continuous generation of new rules means a complex behavior, if the slope of the curve is high this means a more complex or random behavior, when the curve show a plateau as in Figures 19 and 21 it means a stationary dynamic for the corresponding data segments, also changes in the slope indicates changes in the computational complexity of the system as in Figure 22.
5. Discussion

A visual analysis with nonlinear and computational techniques was developed for four variables that characterized the network traffic of data segments corresponding to port 110 (E-mail), with these techniques global and local information was preserved and the integration of the different attributes allowed the construction of a holistic view of the dynamics for each variable under study. It was observed that the time length of client request and answer by the server (SYN) shows a chaotic behavior, and the initial send time for a packet to be processed (ITP) shows a self-similar behavior at different scales. In relation with the time length of connection acknowledge (ACK) and the time length of a packet (TLP) both showed a stationary and deterministic dynamics. Finally, this kind of visual analysis can be useful not only for understanding the dynamics of network traffic but also for the identification of particular events along the evolution of such dynamics.

6. References


