**Andrzej PASZKIEWICZ**, Marek BOLANOWSKI
DEPARTMENT OF COMPLEX SYSTEMS, THE FACULTY OF ELECTRICAL AND COMPUTER ENGINEERING, RZESZÓW UNIVERSITY OF TECHNOLOGY
Rzeszów, Poland

# Long-range dependence in DataCenter networks transmission

**Abstract**

The paper presents the mechanisms of long-range dependence measurement in the context of data transmission in Data Center networks. The research involved mainly analyzing network traffic generated by protocols such as CIFS and iSCSI, which are commonly used in such infrastructures. The purpose of the paper was to determine whether the network traffic of above mentioned protocols encapsulated in TCP/IP protocol will have persistent, anti-persist, or random walk character. By indicating long-range dependencies for this type of network traffic, it will be possible to develop effective mechanisms for detecting anomaly in its transmission as well as flow control, including QoS mechanisms, load balancing, etc.

**Keywords**: Hurst exponent, long-term memory, network convergence, data center protocols.

## 1. Introduction

Computer networks are a very important component of modern IT systems. Taking into account rapid development of computer networks and accompanying rapid increase in the amount of data collected and processed, it is necessary to develop new effective mechanisms supporting the management of network traffic in such systems. So far all mechanisms of so-called traffic engineering were based solely on the assumptions of the accepted paradigm of simple systems [1, 2]. Therefore, they have taken into account the features of complex systems such as self-similarity, emergence, power laws, and long-range dependencies to a small extent [3]. Of course, given the complexity of possible models based on the theory of complex systems, it is not possible to account for all of their features in one model or network mechanism.

Until now, we focused in our studies primarily on the use of complex system properties in anomaly detection. The results of these works were presented, among others, in the following papers [4] and [5]. For this purpose, we based on the selection of statistical parameters, which describe the traffic transmitted in the computer networks. This made it possible to determine the vector of parameters that triggered a specific response to the threat. This approach is a part of an active computer network management model that can be implemented in an SDN network environment [6]. The high bandwidths used in this class of networks require the use of two-step traffic analysis, the idea of which is illustrated in Fig. 1.

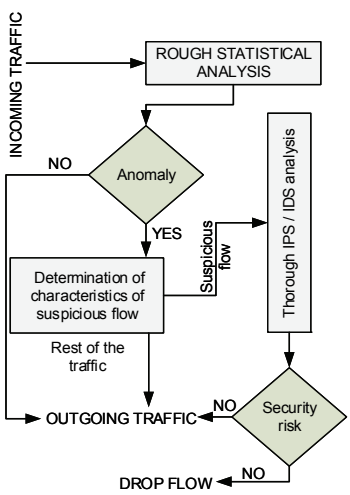


Fig. 1. Diagram of active response to threats using coarse anomaly detection

During a rough statistical analysis, the flow is represented in just a few selected time series (e.g. UDP/s, TCP/s, Utilization [%]/s, etc.). Statistical analysis of these series in predetermined intervals allows to detect suspicious traffic (flow) and redirect it to a thorough analysis. The high computational complexity of the processes of detailed traffic analysis prevents it from being used for all traffic. For the described method to work effectively it is necessary to use simple to determine statistical parameters that will create the vector value for normal traffic. An example of value vector is described in the paper [7]. In the next step, the authors wanted to test whether the methods developed for TCP/IP in particular for core or local networks will also be validated for SANs [8] used in Data Center environments, where network traffic is tunneled in TCP/IP channels. At this stage of the study, the analysis of the traffic of this class was carried in terms of its characteristics in complex systems [9], in particular with respect to long-range dependence. For this purpose, the Hurst exponent described in Chapter 2 will be used.

## 2. Hurst exponent as an element of coarse anomaly detection vector

In computer networks, analysis of network traffic both in terms of security and improving its performance, and planning for infrastructure development and upgrades play an important role. Therefore, apart from analyzing individual events, defining and confirming trends is important, as well as predicting possible changes. Of course, the basic statistical analysis of network traffic can be useful in this regard, but it does not allow to accurately determine whether the current events are random walk or for example a long-term trend. Thus, the Hurst exponent can be used to determine the long-range dependence of the process under investigation [10]. This makes it possible to distinguish random series from non-random ones, even if the first ones are non-Gauss series. There are three classes of Hurst exponent size:

- $H < 0.5$ when the processes are antipersistent, also referred to as 'returning to the average'. So there is a high probability of change.
- $H = 0.5$ when the processes are independent of one another on a time scale and have a random walk.
- $H > 0.5$, when the process is said to be persistent, there is a correlation between the data and long-range dependencies in the time series studied. It can then be said that the current trend will be reinforced.

Therefore, if Hurst's exponent is approaching 1 then the process is more strongly self-similar.

The analysis of rescaled R/S range can be used to identify cyclic fluctuations that allow to determine the existence of long-term memory and causality in time series [11]. In order to obtain a unified time-independent measure, Hurst created a dimensionless estimator that divides the range of fluctuations by the standard deviation of observation. The range of the process variability can be described by the equation:

$$(R/S)_n = cn^H, \tag{1}$$

where $(R/S)_n$ denotes (rescaled) range of variance of n observations, $n$ refers to the number of observations, $H$ is Hurst exponent, and $c$ is positive constant. Of course, there are many methods of determining the value of $H$ exponent. They were well presented in the paper [12] and include i.a.

- *Absolute Value method* – in which the aggregated series $X(m)$ is determined, dividing a series of observations of $N$ length into blocks of $m$ length and averaging each block. By presenting the

aggregation level in relation to the absolute first moment of the aggregated series $X(m)$ in the log-log graph we obtain a line at $H$-1 angle in relation to the axis X.

- *Variance of Residuals* – in this method data is divided into blocks of $m$ size in which later sums are calculated. The log-log graph for average variance should be similar to the straight line at the angle of $2H$.
- *Periodogram method* – makes use of a spectral density estimator to determine the self-similarity index for time series.

## 3. Traffic analysis of data exchange protocols

Data Center is designed to collect and process large amounts of data. For this purpose, the designed and implemented network infrastructure must be characterized by very high reliability, redundancy and efficient network connections and nodes. However, in addition to the hardware, the protocols and network mechanisms involved play an important role. This paper focuses on the analysis of two network-relevant data exchange protocols:

- CIFS (Common Internet File System) – is a protocol that allows sharing resources within a computer network. Originally, it was developed by Microsoft as SMB - Server Massage Block. This protocol allows applications and users to access files and other remote resources. Its performance is based on the client-server model. Because of its simplicity, it is widely used and implemented.
- iSCSI (Internet Small Computer Systems Interface) – it is a protocol used in SANs (Storage Area Networks) to transfer data between storage devices. This protocol sends SCSI commands using TCP/IP protocol. The greatest advantage of iSCSI is the ability to create large SANs using common network components, making it easier to build system and less expensive than traditional Fiber Channel solutions.

A dedicated environment was prepared for the test (Figure 2). In this environment, a software traffic generator played an important role, its task was to simulate the action of both protocols. The IxChariot [13] software used for this purpose by means of available scripts allows to test this type of traffic in any size network environment, including Data Center networks. Thanks to its functionality and prepared input data, it was possible to make a real assessment of network performance taking into account both the user's actions and the system itself. In addition, properly prepared scripts have allowed us to evaluate the performance of SAN devices and determine their behavior during the load.

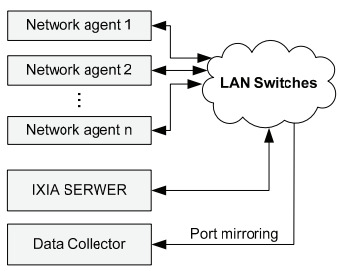


Fig. 2. Scheme of test environment

Laboratory tests were carried out at various time intervals (from a dozen minutes to several hours) to observe, compare and analyze the characteristics of network traffic. IXIA Server was responsible for initiating and controlling the flow of data between the various agents representing servers and computer units in SAN communicating with the above-mentioned protocols. The number of agents for different simulations ranged from 10 to 100. In order to analyze the traffic appearing on intermediate network nodes independently from IxChariot the data mirroring functionality was launched for the dedicated Data Controller. This mechanism did not affect the efficiency of the device itself, nor the quantity and quality of the flows tested.

As a result of the simulations, Hurst exponent values were estimated for the following parameters: response time, efficiency (throughput), and transaction frequency. In order to verify the results obtained, the calculation was performed using three methods: absolute value method, variance of residuals, periodogram method. The results are presented in Tables 1 and 2.

Tab. 1. Hurst exponent values for the CIFS protocol

| CIFS | Response Time | Throughput | Transaction |
|---|---|---|---|
| Absolute Value method | 0,575 | 0,577 | 0,570 |
| Variance of Residuals | 0,706 | 0,724 | 0,710 |
| Periodogram method | 0,654 | 0,648 | 0,655 |

Tab. 2. Hurst exponent values for the iSCSI protocol

| iSCSI | Response Time | Throughput | Transaction |
|---|---|---|---|
| Absolute Value method | 0,634 | 0,630 | 0,638 |
| Variance of Residuals | 0,862 | 0,865 | 0,859 |
| Periodogram method | 0,722 | 0,764 | 0,705 |

Obtained results clearly indicate that the processes connected with the implementation of data transmission connections using two protocols have persistent character, i.e. $0.5<H<1$. Thus, this proves long-range temporal dependencies with respect to each other. For example, Figure 3 shows the results obtained for the Response Time parameter for the CIFS protocol and Figure 4 for the iSCSI protocol based on the variance of residuals. The existence of long-range dependencies in tunneled traffic indicates the possibility of effective use of statistical analysis methods for detecting anomalies. It is possible while analyzing the value of the $H$ parameter itself. Its significant changes in a given time window can indicate anomaly, such as attack, disk crash, system breakage, etc. However, it should be emphasized that the use of coarse anomaly detection requires, especially initially, tuning of threshold values of statistical parameters to avoid reporting a large number of false positive attacks.

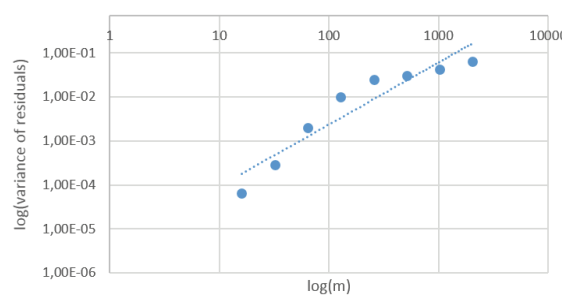


Fig. 3. Results for variance of residuals for CIFS Response Time

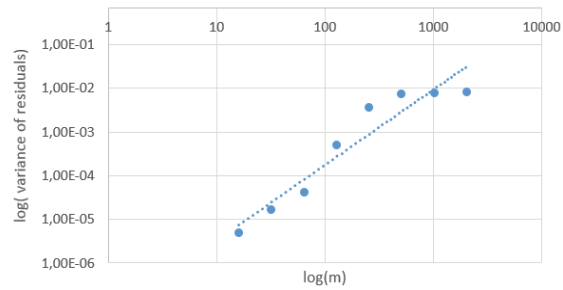


Fig. 4. Results for variance of residuals for iSCSI Response Time

Similar results were obtained for CIFS Transaction (Fig. 5) and iSCSI Transaction (Fig. 6). Not only the used bandwidth and the response time, but also the Transaction Rate show long-term dependence. Therefore, Hurst exponents for these parameters can form the basis of the vector of so-called coarse detection. Of course, as a result of possible further studies, it is possible to extend this vector with additional parameters relevant to the type of data transmission characteristic of the Data Center.

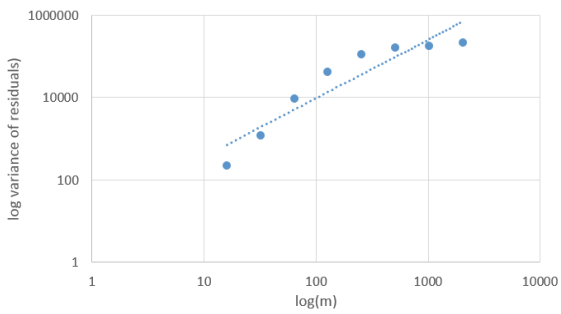


Fig. 5. Results for variance of residuals for CIFS Transaction parameter

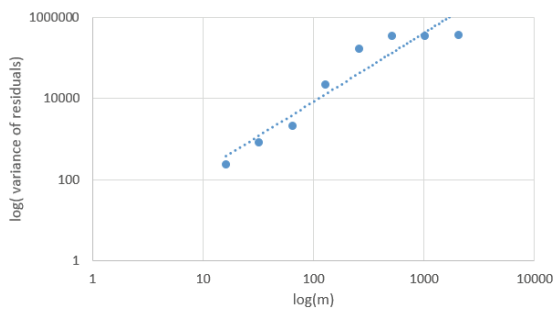


Fig. 6. Results for the variance of residuals for the iSCSI Transaction parameter

## 4. Conclusions

The studies conducted using three methods indicated that the network traffic implementing connections with the CIFS and iSCSI protocols has persistent character, i.e. the probability that the current trend will be maintained rather than the change will occur is very big. This makes it possible to estimate the stability of transmission parameters describing a given network traffic. In the case of Data Center networks, which are characterized by the transmission of large amounts of traffic, this may allow to develop new effective method of balancing network infrastructure load. In addition, it seems possible to use up-to-date information to create QoS mechanisms dynamically adjusting priorities and handling queues in intermediate nodes for current network conditions. On the other hand, the persistent character of traffic in the Data Center networks can allow for detection of any deviations from the norm. The causes of such changes can be attacks on data center infrastructure, failures and malfunctioning of network interfaces, the spread of viruses, etc. Of course, in such case, it is also necessary to develop mechanisms determining acceptable values of deviations individually for particular phenomena. However, such activities should be the basis for further research and subsequent publications. In further papers, we will explore the possibility of using neural networks [14] to improve the detection level of anomaly or introduce the next (third) level of their detection.

## 5. References

[1] Nicolis G., Nicolis C.: Foundations of Complex Systems: Emergence, Infor-mation and Predicition. World Scientific Publishing Co., Singapure, 2012.

[2] Grabowski F., Paszkiewicz A., Bolanowski M.: Computer Networks as Complex Systems in Nonextensive Approach. Journal of Applied Computer Science, vol. 21, No. 2, pp. 31-44, 2013.

[3] Domańska J., Domańska A., Czachórski T.: A Few Investigations of Long-Range De-pendence in Network Traffic. In: Proceedings of the 29th International Symposium on Computer and Information Sciences, Springer International Publishing, pp. 137-144, 2014.

[4] Bolanowski M., Paszkiewicz A., Wroński M., Żegleń R.: Representativeness analysis and possible applications of partial network data flows. Measurement Automation Monitoring, t.62, z.1, s.29-32, 2016.

[5] Bolanowski M., Paszkiewicz A., Zapała P., Żak R.: Stress test of network devices with maximum traffic load for second and third layer of ISO/OSI model. Pomiary Automatyka Kontrola, t.60, z.10, s.854-857, 2014.

[6] Nunes B.A.A., Mendonca M., Nguyen X.N., Obraczka K., Turletti T.: A Survey of Software-Defined Networking: Past, Present, and Future of Programmable Networks. IEEE Communications Surveys & Tutorials, vol. 16, issue 3, pp. 1617-1634, 2016.

[7] Bolanowski M., Paszkiewicz A.: Nowy model detekcji zagrożeń w sieci komputerowej, Przegląd Elektrotechniczny, t.89, z.11, s.308-311, 2013.

[8] Tate J., Beck P., Ibarra H.H., Kumaravel S., Miklas L.: Introduction to Storage Area Networks. An IBM Redbooks publication, 2016.

[9] Grabowski F.: Nonextensive model of self-organizing systems. Complexity, vol. 18, issue 5, pp. 28-36, 2013.

[10] Ciftlikli C., Gezer A.: Comparison of Daubechies wavelets for Hurst parameter estimation. Turk J Elec Eng & Comp Sci, vol. 18, pp. 117-128. 2010.

[11] Sheng H., Chen Y.Q., Qiu T.S.: Fractional Processes and Fractional-Order Signal Processing: Techniques and applications. Signals and Communication Technology, Springer, 2011.

[12] Karagiannis T., Faloutsos M., Riedi R.H.: Long-Range Dependence: Now you see it, now you don't! Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE.

[13] https://www.ixiacom.com/products/ixchariot. Access: 10.07.2017.

[14] Gomółka Z., Twaróg B., Bartman J., Kwiatkowski B.: Improvement of Image Processing by Using Homogeneous Neural Networks with Fractional Derivatives Theorem, AIMS, Discrete and Continuous Dynamical Systems, Journal of American Institute of Mathematical Sciences, pp. 505-514, 2011.

*Received: 18.05.2017* *Paper reviewed* *Accepted: 02.07.2017*

**Andrzej PASZKIEWICZ, PhD, eng.**

He received a Ph.D. degree in Computer Science from Lodz University of Technology in 2009. His current research interests focus on widely under-stood processes in computer networks. He works at the Department of Power Electronics, Power Engineering and Complex Systems Rzeszow University of Technology as an assistant professor.

*e-mail: andrzejp@prz.edu.pl*

**Marek BOLANOWSKI, PhD, eng.**

He received a Ph.D. degree in Computer Science from Lodz University of Technology in 2009. His current research interests focus on computer network design and interconnection network performance. He works at the Department of Distributed Systems Rzeszow University of Technology as an assistant professor.

*e-mail: marekb@prz.edu.pl*