UNIVERSITY CORPORATIVE NETWORK TRAFFIC ANALYSIS BASED ON THE METHODS OF NONLINEAR DYNAMICS

M.A. Basarab, I.P. Ivanov, A.V. Kolesnikov, L.I. Kolobaev
Bauman MSTU
Moscow
e-mail: bmic@mail.ru

Abstract

As the number of hosts connected to the corporate network, increasing not only the amount of information transmitted, but also the load on the hardware of the server. Identification of statistical characteristics, as well as the correlations between network activity and dedicated computing resources is an important task in the field of optimization of computer networks.

The paper shows the results of some statistical characteristics of incoming and outgoing traffic, as well as associated distribution apparatus resources of one server on the corporate network of university.

1 Introduction

In the computer processing and storage of information it is necessary to exchange data between the actors. So since late 70s begins the rapid development of computer networks and related equipment. Local and wide area networks evolve every year. The number of connected subscribers increases, as well as the total amount of transmitted information.

This fast development trend raises a number of problems. Intensive exchange of data requires increase in computing power and can lead to a reduction in the quality of service. Network topology, transportation protocol, different type of web services and many other reasons can influence on the characteristics of data traffic. So the actual mathematical models of traffic which could help to optimize the network load are needed. It is also will be useful in developing software and hardware ways of increasing network reliability.

It is obviously, that macro parameters of network traffic determine by man. But parameters on timescales equals microseconds determines by hardware and transportation environment. So research of servers hardware and operating system behavior could be another useful instrument in the task of network load optimization. Investigation of memory layout, CPU, operating system state and other characteristics of the host may reveal hidden correlations, cycles and phase transitions in the network traffic distribution.

2 Overview of traffic models

Stochastic models of traffic incoming, which was used in the past, most is a Markov process, or in a general sense - processes with short-term dependence. Such models described by Poisson distribution with an exponential message length and were based on terms of queuing theory. It was a time of early ARPANET networks. Then results of data traffic modeling by queuing theory was almost equal to traditional telephony call hold time distribution.

Later appear many studies that indicate increasing overall amount of network traffic. Furthermore new data transport protocols have a great impact on characteristics of traffic. As a result of such studies was developed concept of packet trains in 1986. This model assumes that messages in network moving together as a train. When in Poisson assumes that each packet processed individually.

One of the popular contemporary data traffic model consider a long term dependencies in data volume distributions. Also it was concluded, that network traffic is self-similar with heavy tailed distribution.

2.1 Classic model

First traffic models consider that data sources in networks often has a pulse character. Periods with intense activity are followed by long idles. So it was concluded that interarrival times of messages are exponentially distributed and data source behaves as a Poisson process. The length of messages is exponentially distributed too. And all process is stationary and independent.

Poisson model does not capture periods of traffic bursties, which describe network traffic. Any renewal traffic process has an autocorrelation function vanished identically in Poisson process. But in real networks traffic has bursty periods which conclude to positive autocorrelation. Thus, Poisson model does not work in case when burstiness happens on multiple time scales.

2.2 Packet train model

This popular network traffic model was formalized in 80s. It assumes that packets in traffic flow travel together and may be proceed as a unit. And each point of the network should make decision about train destination considering only first packet of the chain. Such algorithm would save network hardware from useless operations of analyze equals frames. But it should be noticed that packet train model is a source model. And it could be applied only if all packets in chain have similar destination point. It is obviously that network hardware and transport protocols would be different from one based on classic model.

2.3 Self-similar model

Some current researches indicate that union of lots highly variable data sources make network traffic strongly auto-correlated with long-range dependency[4]. It means that persistence in correlation structures does not disappear on even large lags. In other words, the imposition of a set ON / OFF-sources (data sources with variable activity periods) exhibit infinite variance syndrome, resulting in united self-similar network traffic, striving for fractional Brownian motion. In other works was mentioned that client/server architecture may be a reason of high frequently variation of ON/OFF periods.

There are many factors could be a reason of self similarity of network traffic:

- User behavior;
- Searching, storing and processing of data;
- Union of traffic streams;
- Network management systems;
- Network optimization based on feedback;
- Network development;

Usually self-similar processes described by Hurst coefficient. It accepts values from 0 to 1. For white noise Hurst coefficient equal 0.5 it means that there is no any long term dependence. When $H > 0.5$, the data array has long-range dependence. If $H < 0.5$, this is the case of independence process.[1]

3 Data acquisition

In research was monitored one of the servers of corporate network in STDU of Bauman. Host used for the needs of web portal named Electronic University. It has few virtual machines with Linux operating systems. And used for its own purposes like several data bases, web servers, file servers and other. Mostly every traffic packets belong to HTTP. Ethernet 100Mbps was used.

Each physical server in university network use Zabbix monitoring system. Zabbix is client-server application for real time collection, storing and processing information about servers operating system, hardware and network traffic parameters. Also it allows notifying network administrator about high loads or deny of service situations. Zabbix is a freeware software suite. For later processing was collected data about next servers parameters:

- Cached memory;
- Buffers memory;
- CPU input/output wait time
- CPU idle time;
- CPU system time;
- CPU user time;
- Free memory;

- Incoming/outgoing traffic (bit/s);
- Number of OSs running processes;
- Number of apaches running processes;
- Summary processor load.

Data collected in real time with different timescales from 1 to 60 seconds 24/7 and stored as time series.

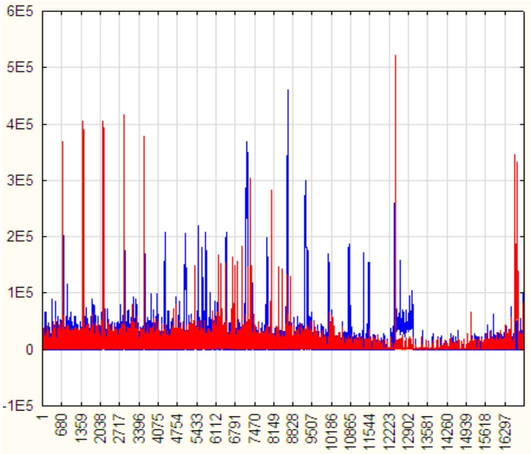

Figure 1: Daily traffic

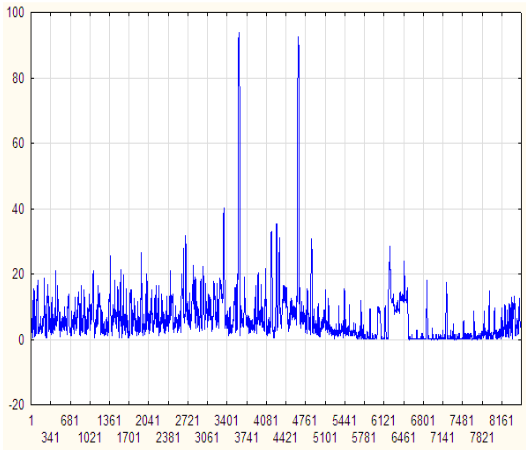

Figure 2: Daily CPU user time

4 Autocorrelation function analysis

First of all autocorrelation function was analyzed to assume order of increasing dependence. For self similar processes autocorrelation function of aggregated process $X^{(m)}$ at $m \to \infty$ does not vanish.

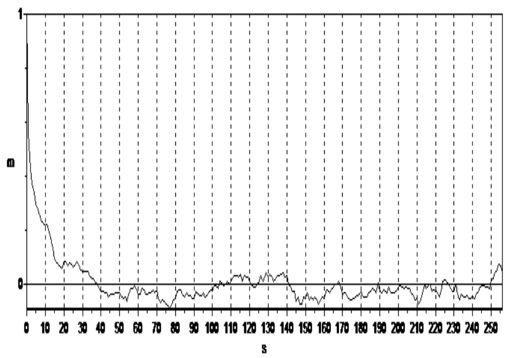

Figure 3: ACF of incoming traffic

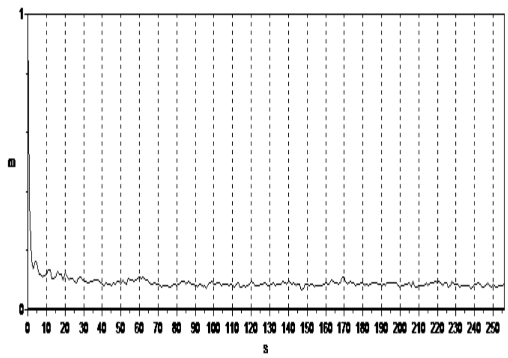

Figure 4: ACF of outgoing traffic

It is important to note that ACF of such memory parameters of server as amount buffered or free memory shows periodical behavior. But other hardware and operating system parameters still match ACF for self similar process. It probably appears because memory controls first of all by operating system.

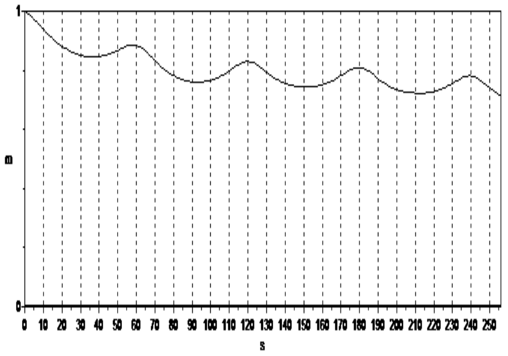

Figure 5: ACF of cached memory

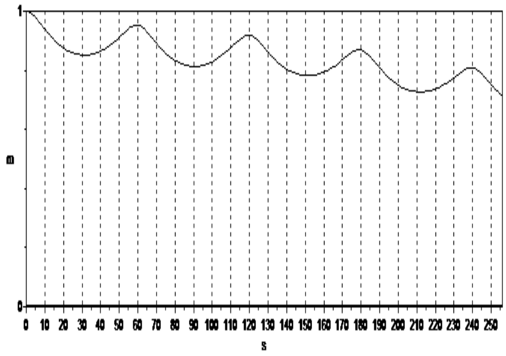

Figure 6: ACF of free disc space

So analysis of autocorrelation function show us that network traffic and some server monitored parameters could be seen as self similar process.

5 Hurst parameter analyze

H.E. Hurst experimentally observed that for many time series valid the next expression:

$$\frac{R}{S} = (\frac{N}{2})^H,$$

where H – Hurst parameter; R – range series; S – standard deviation series.

As it mentioned above Hurst parameter could be a measure of long term dependence of time series. Estimation of it would not only make a conclusion about self similarity but allows using fractal forecast models. Short term forecast could help in development systems of network and hardware overload balances.

Parameter	Hurst coefficient
Buffers memory	0.9656
Cached memory	0.9868
CPU idle time	0.9575
CPU system time	0.9903
Free memory	0.9336
Incoming traffic	0.9775
Number of OSs running processes	0.8835
Number of Apaches running processes	0.8343
Outgoing traffic	0.9712

Table 1: Values of Hurst coefficient for monitored data

Hurst coefficient was estimated by R/S analysis of time series. And as expected all monitored parameters have show self similar nature with long range dependence.

	Hurst value parameter			
	00:00 to 6:00	06:00 to 12:00	12:00 to 18:00	18:00 to 24:00
Incoming traffic	0.9516	0.9836	0.6154	0.7142
Outgoing traffic	0.9112	0.9766	0.7385	0.8180

Table 2: Values of Hurst coefficient for monitored data

Also it was confirmed that Hurst parameter depends from traffic intense. For that data series was divided by time periods and H was calculated.

Such big value of Hurst parameter in morning time can be explained by the service of document flow in university. Automatic system makes document distribution in morning hours increasing amount of transmitted traffic.

6 Traffic as deterministic chaos

Some recent studies made an interesting try to consider network traffic in equations of chaos theory. Chaos theory studies the behavior of dynamical systems that are highly sensitive to initial conditions. Plotting daily traffic data series in phase space it always has some kind of attractors.

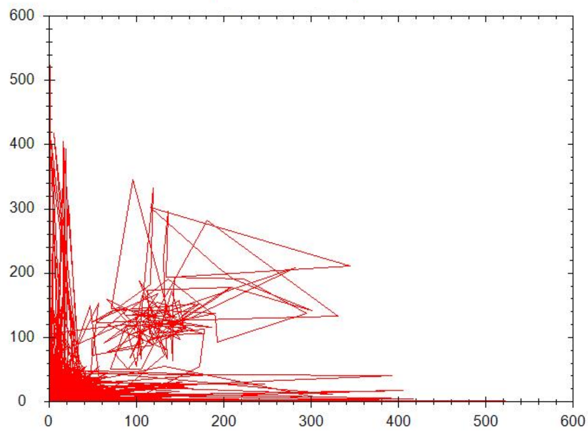

Figure 7: Attractor of daily traffic in phase space

Was made an estimation of Lyapunov exponent as a parameter of chaotic systems.

$$\lambda_1(i) = \frac{1}{i\Delta t}\frac{1}{(M-i)}\sum_{j=1}^{M-k}\ln\frac{d_j(i)}{d_j(0)},$$

where Δt is the sampling period of the time series, and $d_j(i)$ is the distance between the jth pair of nearest neighbors after i discrete-time steps and M is the number of reconstructed points.

First Lyapunov exponent for all types of traffic takes values from 1.2 to 2.5. It let us work with traffic as a chaos, use attractors and nonlinear dynamics methods. Also in is worth to notice that Lyapunov exponent for the most of other servers parameter takes value above zero that means of periodic properties of them.

7 Conclusion

During research was performed analyse of server incoming/outgoing traffic and important corresponding hardware and operating system parameters. Self similar and long range dependence was confirmed not only for network traffic but other monitored parameters also. This fact will allow us using fractal methods in forecasting, searching hidden correlations and other.

Also for network traffic was calculated first Lyapunov exponent which shows that traffic could be considered as deterministic chaos and non linear dynamics methods applicable in future works.

References

[1] Hurst H. E. Long-term storage capacity ofres ervoirs. *Trans. Amer. Soc. Civ. Engrs.*. Vol. **116**, pp. 770 – 808.

[2] Wolf A.,Swift J. B.,Swinney H. L.,Vastano J. A. Determining Lyapunov exponents from a time series. *Physica*. Vol. **116**, pp. 285.

[3] Benettin G., Galgani L., Strelcyn J.M. Kolmogorov entropy and numerical experiments. *Phys. Rev.*. Vol. **A14**, pp. 23382345.

[4] Leland W.E., Taqqu M.S., Willinger W., Wilson D.V. On The Self-Similar Nature Of Ethernet Traffic. *ACM SIGCOMM93.* **San-Fransisco, 1993**, pp. 183 – 193.