Investigation of network communications by using statistical processing of monitored data

RADI ROMANSKY
TECHNICAL UNIVERSITY OF SOFIA (BULGARIA)

Contents

I. Introduction

II. Organization of Program Monitoring

III. Statistical Processing of Monitored Data

- *A. Application for initial evaluation*
- *B. Calculation of basic statistics*
- <u>Comments:</u>

IV. Conclusion

References

Abstract

The communications in a given network structure are related to the generation of network traffic and one possibility for its investigation is to conduct program monitoring with subsequent statistical processing of the registered data. The article presents a study of communication parameters of traffic based on measurement experiments in a universal network segment and analysis of the formed sample to determine statistical estimates to detect relationships between parameters.

Keywords

communication parameters, network traffic, measuring experiments, program monitoring, statistical assessments.

I. Introduction

The main feature of the contemporary digital age is the mass informatization of society based on the penetration of network technologies in all areas of public life, social interactions, communications and traffic management, as well as monitoring governance processes. All this is related to the organization of effective management of network communications in traditional networks, setting the goal of optimizing the network infrastructure and the transition to new generations and green communications.

The paper discusses the problem of organizing program monitoring in a universal network environment and statistical processing of registered data to obtain estimates for selected parameters. The measurement experiments were performed using a selected software, and the registrations were saved in a file in a suitable format. The program monitor is installed in a workstation within a local area network and runs on Windows to "listening" the network traffic. This is possible for "broadcast" type Ethernet segments, and the study itself depends on the platform used. The formed initial sample of registrations (levels of monitored parameters) was used to conduct statistical analysis, and for this purpose an application was developed that allows automated determination of basic statistics such as sample power and range, descriptive estimates, correlation coefficients and regression analysis to detect interdependence between parameters.

II. Organization of Program Monitoring

OmniPeek network analyser was chosen to organize the measurement experiments, which provides detailed information on latency, performance, possible causes of solution-finding errors, as well as visualization of its communication nodes and the traffic exchanged between them.

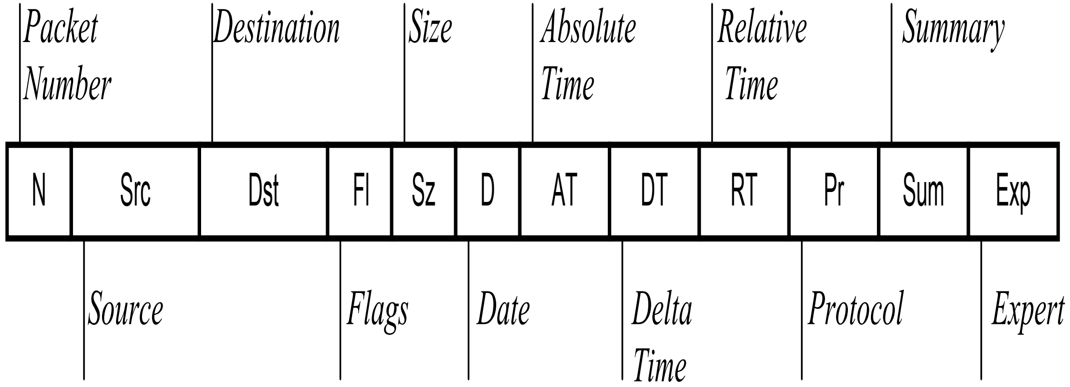
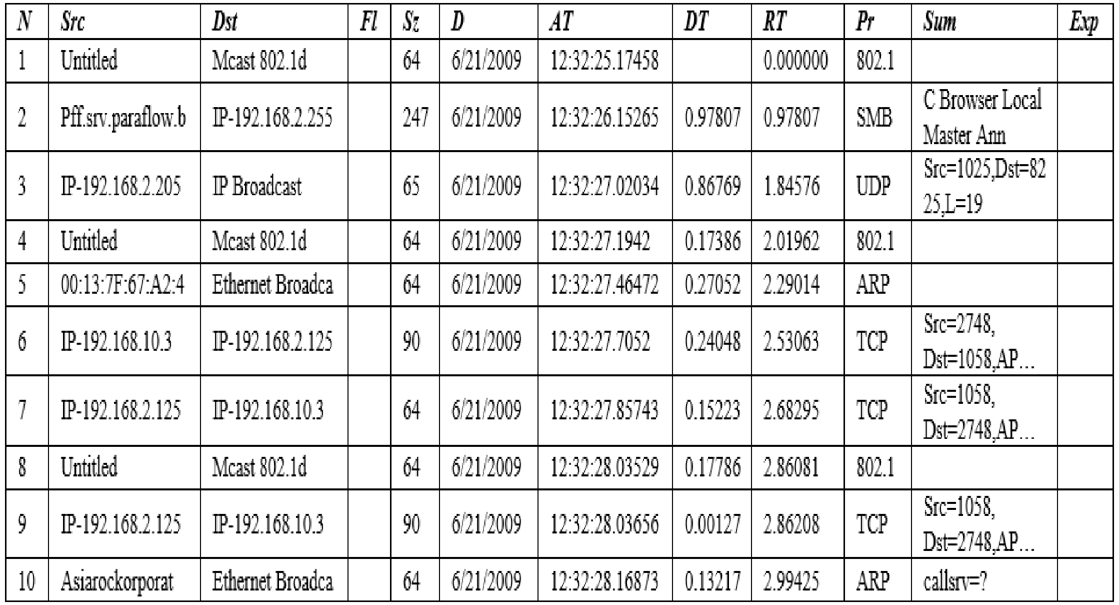


Fig. 1. Monitoring file format

<u>Legend:</u> **N** – serial number of the package in the registration file; **Src** – packet sender; **Dst** – recipient of the packet; **Fl** – information about the conditions to which the packet meets (a packet received with an error in the data or package according to a pre-set protocol); **Sz** – length (size) of the package; **D** – date of receipt of the moment; **AT** – time of packet receipt according to monitor system timer computer; **DT** – time elapsed since the previously registered package; **RT** – time from the beginning of the measurement; **Pr** – used protocol for the package; **Sum** / **Exp** – additional information about the package

Table 1. Monitoring Results



N	*Src*	*Dst*	*Fl*	*Sz*	*D*	*AT*	*DT*	*RT*	*Pr*	*Sum*	*Exp*
1	Untitled	Mcast 802.1d		64	6/21/2009	12:32:25.17458		0.000000	802.1		
2	Pff.srv.paraflow.b	IP-192.168.2.255		247	6/21/2009	12:32:26.15265	0.97807	0.97807	SMB	C Browser Local Master Ann	
3	IP-192.168.2.205	IP Broadcast		65	6/21/2009	12:32:27.02034	0.86769	1.84576	UDP	Src=1025,Dst=8225,L=19	
4	Untitled	Mcast 802.1d		64	6/21/2009	12:32:27.1942	0.17386	2.01962	802.1		
5	00:13:7F:67:A2:4	Ethernet Broadca		64	6/21/2009	12:32:27.46472	0.27052	2.29014	ARP		
6	IP-192.168.10.3	IP-192.168.2.125		90	6/21/2009	12:32:27.7052	0.24048	2.53063	TCP	Src=2748, Dst=1058,AP…	
7	IP-192.168.2.125	IP-192.168.10.3		64	6/21/2009	12:32:27.85743	0.15223	2.68295	TCP	Src=1058, Dst=2748,AP…	
8	Untitled	Mcast 802.1d		64	6/21/2009	12:32:28.03529	0.17786	2.86081	802.1		
9	IP-192.168.2.125	IP-192.168.10.3		90	6/21/2009	12:32:28.03656	0.00127	2.86208	TCP	Src=1058, Dst=2748,AP…	
10	Asiarockorporat	Ethernet Broadca		64	6/21/2009	12:32:28.16873	0.13217	2.99425	ARP	callsrv=?	

III. Statistical Processing of Monitoring Data

A. Application for initial evaluation

For additional analysis of the observed traffic, an application has been developed that offers the following functionalities:

✓ selection of a file with monitoring registrations;

✓ review of the registered information for each of the packages;

✓ selection of a specific source and recipient of the package;

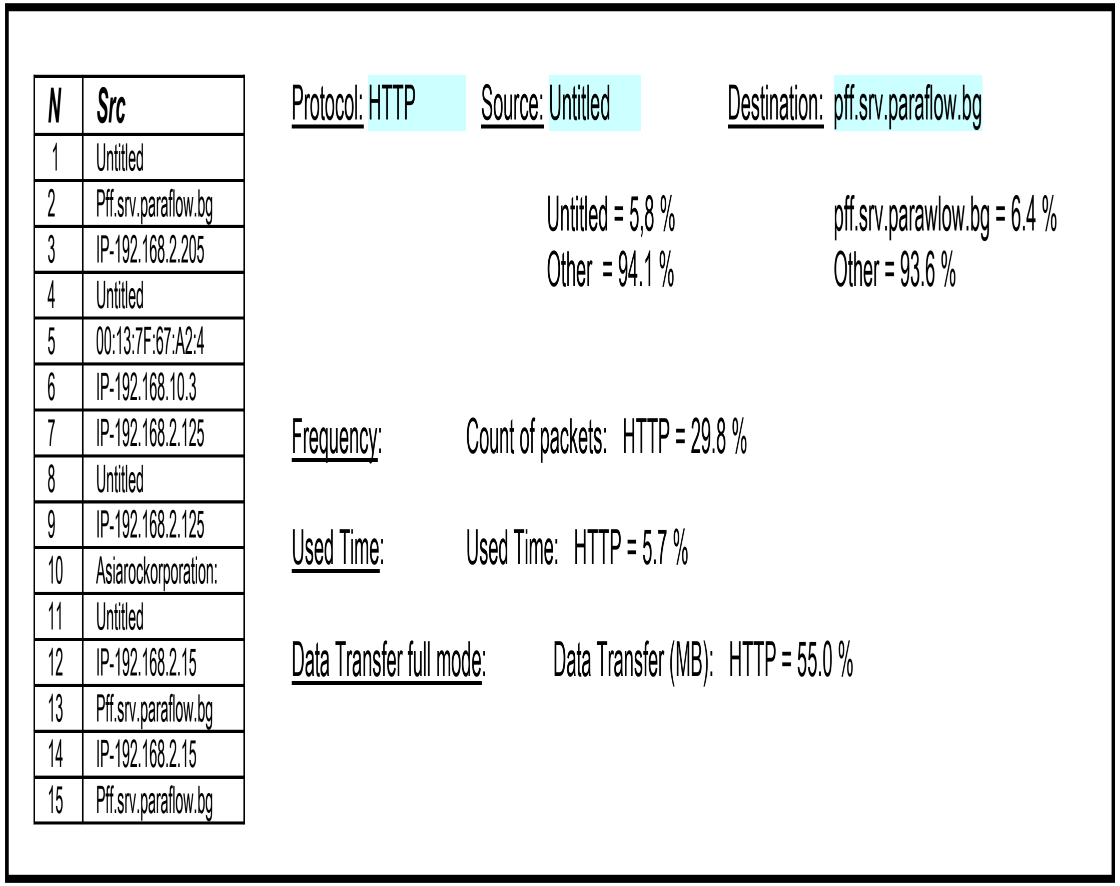
✓ calculation of the number of packets transmitted per unit time for a selected protocol;

✓ determination of statistical estimates - minimum, maximum and mean value, standard deviation, variance, correlation matrix, frequency of use for a given protocol, mean time of use;

✓ calculation of the generated traffic. The application requires before starting to be determined a working file for registration monitored data, which permits to calculate automatically calculates the statistics for the specified protocol.

An example is presented in figure 2.

As can be seen from Figure 2, the determined statistics for a given protocol are: ✓ relative frequency of using the protocol when transmitting packets; ✓ relative time to use the protocol; ✓ relative pat of data transferred (volume in [MB]).



N	*Src*
1	Untitled
2	Pff.srv.paraflow.bg
3	IP-192.168.2.205
4	Untitled
5	00:13:7F:67:A2:4
6	IP-192.168.10.3
7	IP-192.168.2.125
8	Untitled
9	IP-192.168.2.125
10	Asiarockorporation:
11	Untitled
12	IP-192.168.2.15
13	Pff.srv.paraflow.bg
14	IP-192.168.2.15
15	Pff.srv.paraflow.bg

Protocol: HTTP Source: Untitled Destination: pff.srv.paraflow.bg

Untitled = 5,8 %
Other = 94.1 %

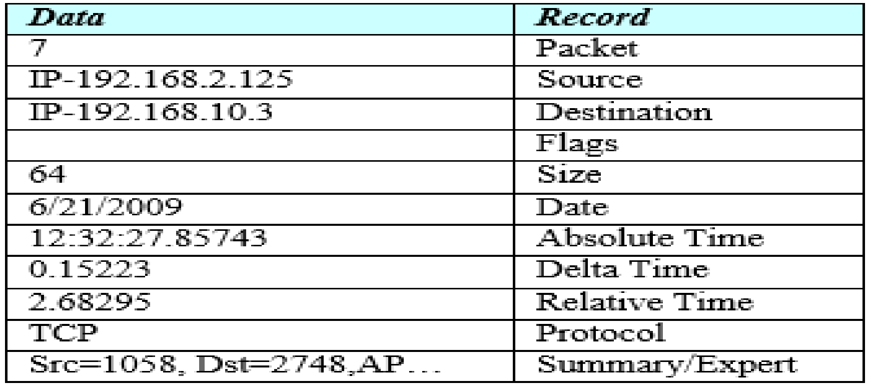
pff.srv.parawlow.bg = 6.4 %
Other = 93.6 %

Frequency: Count of packets: HTTP = 29.8 %

Used Time: Used Time: HTTP = 5.7 %

Data Transfer full mode: Data Transfer (MB): HTTP = 55.0 %

For each selected monitoring data file, the application provides a visualization of the registrations made and by selecting a line (registration) a window appears containing the complete information about the package (Fig. 3), consistent with the determined format.



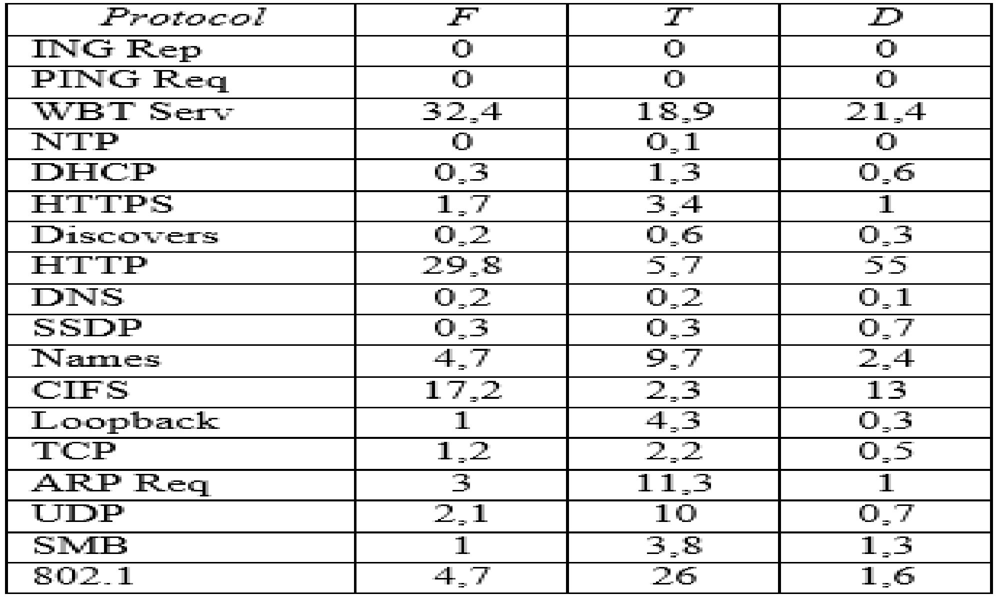
Data	*Record*
7	Packet
IP-192.168.2.125	Source
IP-192.168.10.3	Destination
	Flags
64	Size
6/21/2009	Date
12:32:27.85743	Absolute Time
0.15223	Delta Time
2.68295	Relative Time
TCP	Protocol
Src=1058, Dst=2748,AP…	Summary/Expert

Table 2 summarizes the calculated relative values of:

F – frequency of use of packet transfer protocols during monitoring;

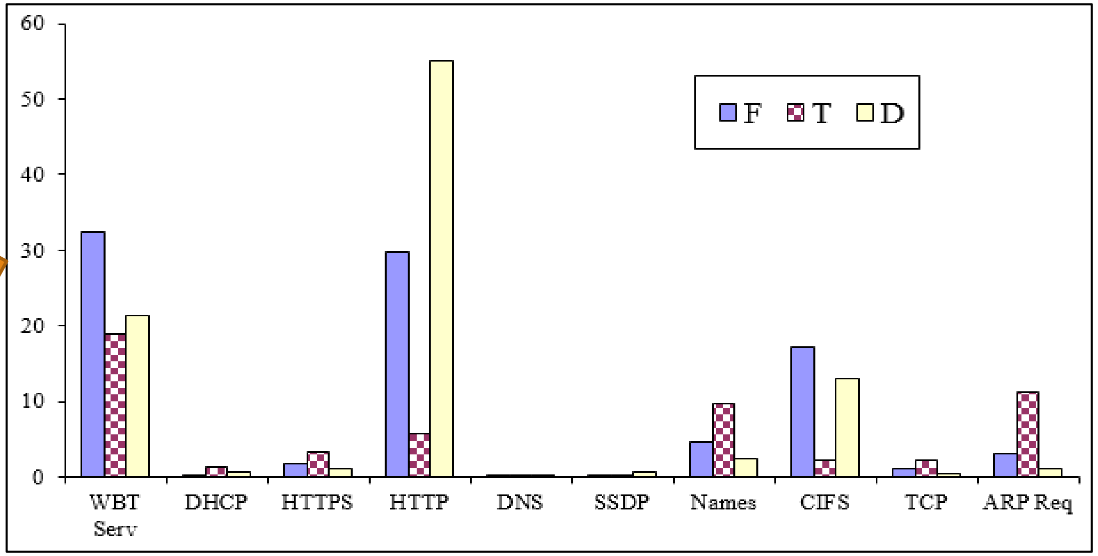
T – relative time to use the protocols;

D – estimate of the transferred data as a relative share to all transferred data



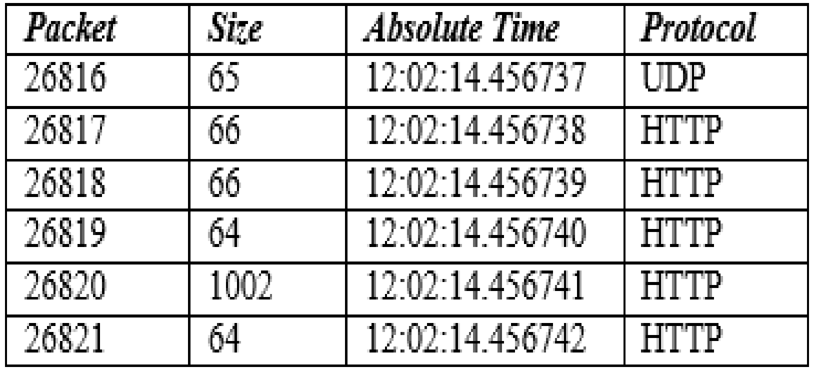
Protocol	*F*	*T*	*D*
ING Rep	0	0	0
PING Req	0	0	0
WBT Serv	32,4	18,9	21,4
NTP	0	0,1	0
DHCP	0,3	1,3	0,6
HTTPS	1,7	3,4	1
Discovers	0,2	0,6	0,3
HTTP	29,8	5,7	55
DNS	0,2	0,2	0,1
SSDP	0,3	0,3	0,7
Names	4,7	9,7	2,4
CIFS	17,2	2,3	13
Loopback	1	4,3	0,3
TCP	1,2	2,2	0,5
ARP Req	3	11,3	1
UDP	2,1	10	0,7
SMB	1	3,8	1,3
802.1	4,7	26	1,6

Based on the parameter F most used protocol is WBT Server, followed by HTTP and CIFS. At the relative time (parameter T) for use with the highest value is protocol 802.1, followed by WBT Server, and most data (parameter D) is transmitted via HTTP (55%).

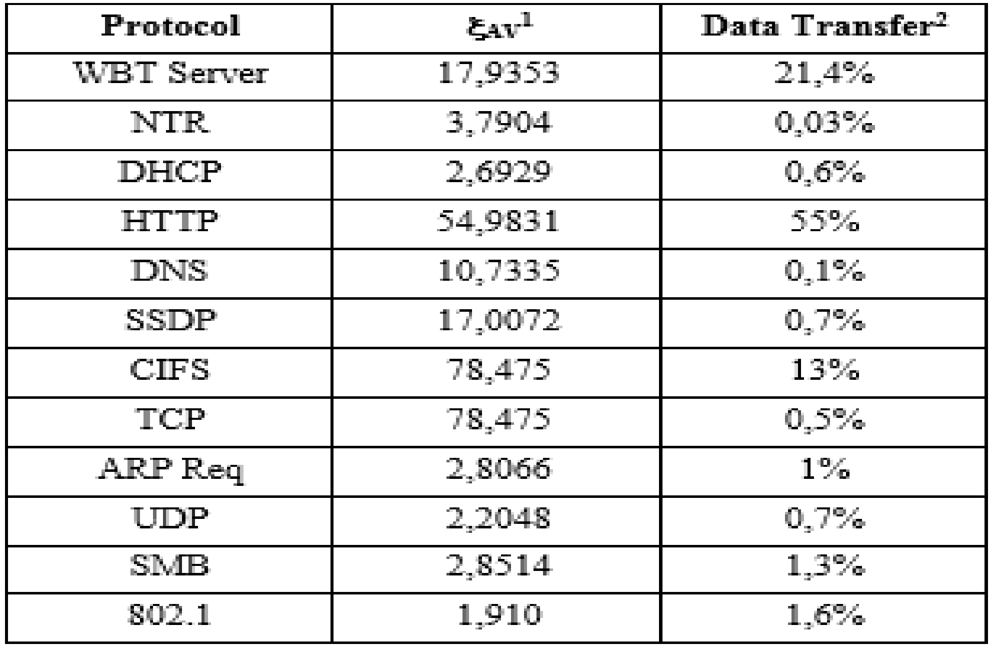


B. Calculation of basic statistics

Fig. 5. Segment of the registration file



Packet	*Size*	*Absolute Time*	*Protocol*
26816	65	12:02:14.456737	UDP
26817	66	12:02:14.456738	HTTP
26818	66	12:02:14.456739	HTTP
26819	64	12:02:14.456740	HTTP
26820	1002	12:02:14.456741	HTTP
26821	64	12:02:14.456742	HTTP



Protocol	ξ_{AV}[1]	**Data Transfer**[2]
WBT Server	17,9353	21,4%
NTR	3,7904	0,03%
DHCP	2,6929	0,6%
HTTP	54,9831	55%
DNS	10,7335	0,1%
SSDP	17,0072	0,7%
CIFS	78,475	13%
TCP	78,475	0,5%
ARP Req	2,8066	1%
UDP	2,2048	0,7%
SMB	2,8514	1,3%
802.1	1,910	1,6%

Basic statistical estimates for each individual protocol can be calculated on the basis of analysis of all measured information, by determining a random variable ξ – number of transmitted packets per unit time. For example, for the HTTP protocol, the calculated statistics are as follows:

- ✓ average value: ξ_{AV} = 554.9830648 [packets per second];
- ✓ packet transmission time: t_{MIN} = 0,000021 [s]; t_{MAX} = 1,882229 [s];
- ✓ mathematical expectation: E[ξ] = 0.1091066;
- ✓ variance: V[ξ] 0.1051385;
- ✓ square deviation: σ = 0.148704.

Summary of the estimates for the average value ξ_{AV} of the parameter ξ by protocols and data transfer are presented in

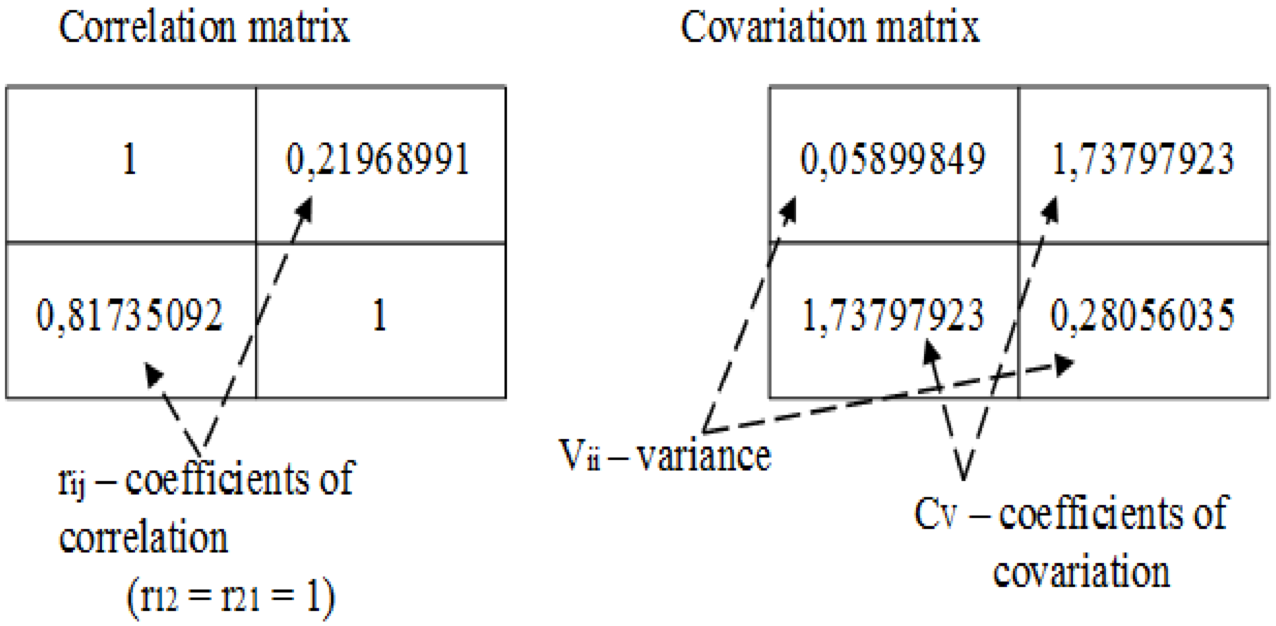
TABLE. 3. AVERAGE VALUES AND USABILITY IN DATA EXCHANGE

Comments:

(1) Average number of transmitted for 2 sec. packages. The most transmitted packet is the Common Internet File System (CIFS) protocol, which allows multiple subscribers to simultaneously access and modify the contents of the same file (through file sharing, synchronization and conflict prevention). In this case, the studied local area network is made up of many shared resources, to which all users have access, which explains the high usability.

(2) Relative part of transmitted data (in [%]) for protocols. HTTP has the highest value, which can be explained by the fact that most of the traffic consists of the transfer of information from external networks (Internet) and devices. At the same time, TCP is one of the least used protocols, which is normal because it is only used to transfer data between two nodes on the network. After the authentication, data transmission and use of protocols such as HTTP, CIFS, WBT Server, etc. begins.

For each individual protocol at a given time interval could be calculated the correlation and covariance matrix for the parameters "total number of packets" and "packets of the selected protocol", forming an simple from the registration file containing data only for the requested protocol – **Figure 6**.



Correlation matrix

1	0,21968991
0,81735092	1

r_{ij} – coefficients of correlation

$(r_{12} = r_{21} = 1)$

Covariation matrix

0,05899849	1,73797923
1,73797923	0,28056035

V_{ii} – variance

Cv – coefficients of covariation

IV. Conclusion

The results of the statistical processing of the experimental data obtained from the conducted program monitoring can be used for optimization of the information service processes in a distributed network environment. Additional possibilities provided by the presented software environment are calculation of correlation and covariance matrices for selected network traffic parameters, such as "total number of packets" and "packets of the selected protocol". For this purpose it is necessary to select a specific protocol and to indicate the time-interval, on the basis of which a sample is formed from the registration file for conducting the statistical analysis. The statistical analyzes offered by the application can be supplemented with variance and factor analyzes, which will lead to a variety of estimates and conclusions about the parameters of information services.

References

23 publications are included in the list of references

Thank you for your attention

Radi Romansky

rrom@tu-sofia.bg