Advances in Science and Technology Research Journal

Volume 13, Issue 2, June 2019, pages 51–55
https://doi.org/10.12913/22998624/106162

Received: 2019.03.19
Revised: 2019.04.09
Accepted: 2019.04.26
Available online: 2019.05.17

Research of Typical Information Objects Traffic

Vadim Goykhman[1], Luiza Korganbaeva[2], Alexey Ermakov[3,4], Maria Nikolaeva[4*]

[1] NTC SOTSBI, Pestel str. 7, 191028, Saint-Petersburg, Russia
[2] Gumilyov Eurasian National University, Satpaeva 2, 010008, Astana, Kazakhstan
[3] LO ZNIIS, Varshavskaya 11, 196128, Saint-Petersburg, Russia
[4] North-Eastern Federal University, Belinskogo 58, 677000, Yakutsk, Russia
* Corresponding author's e-mail: mv.nikolaeva@s-vfu.ru

ABSTRACT
The article considers network traffic as an aggregation of information flows from typical information objects, which can include apartment buildings, student campuses, office centers, sports and entertainment facilities of general use, etc. The study showed that the amount of information transferred by the TCP protocol significantly exceeds (by orders of magnitude) the amount of information transferred by the UDP protocol, which illustrates the average daily traffic graph of the amount of information transferred by the TCP and UDP protocols. The methods of traffic analysis were considered, the most significant characteristics of traffic were determined.

Keywords: traffic, TCP session, Netflow, packets, flows.

INTRODUCTION

Traffic analysis is a direction of scientific research, the results of which are essential for improving the efficiency of decisions made in various aspects of infocomm industry, such as development and management of networks, protection, maintaining a given level of quality.

Telecommunications traffic represents a large number of parameters and therefore, depending on the purpose of the analysis, used method, network points, where traffic data can be received, resources and capabilities of software and hardware, the components that characterize the traffic in the proposed circumstances most fully were selected.

Traffic circulating in a telecommunications network is a set of information flows, each generated by a typical information object. For each of these typical information objects, one can assume the similarity of the traffic structure. The examples of such typical information objects are: apartment buildings, student campuses, office centers, sports and entertainment facilities for general use, etc. Since the traffic of the objects of the same type is, in a probabilistic sense, stable in nature, its characteristic features can be described by certain models. The presence of such mathematical models can be used in simulation modeling and will allow building and developing communication networks focused on specific sources of traffic, which in turn will, with due quality, increase the economic efficiency of projects.

METHODS

Two main methods of traffic analysis are known [1]: packet analysis (packet-based), which, in turn, is subdivided into surface, medium and deep analysis (SPI, MPI, DPI), and flow analysis (flow-based), selectable or aggregated by specified criteria.

It should be emphasized that flow analysis is currently used more often than packet analysis, since this method imposes fewer requirements on the resources used, due to a significant reduction in the amount of computation for processing. In addition, in recent years, there have been more and more statements about the inadmissibility and illegality of deep packet analysis, since it

requires an analysis of the content and thus may violate the secrecy of correspondence. Therefore, the traffic analyzed with the packet analysis method must first undergo a special "anonymization" procedure. This also causes the spread of flow-based analysis methods [2].

Netflow is a protocol that was developed by Cisco Systems. It is commonly used for analyzing network traffic using a flow-based method. Currently, de facto, it is a standard supported not only by Cisco Systems, but also by other leading manufacturers of network equipment.

The architecture of the Netflow traffic accounting system is based on three components: a sensor, a collector, and an analyzer. In order to collect information about traffic, Netflow uses one or more sensors that collect the statistics about the traffic passing through routers, and a collector that receives the information from sensors and saves it in storage. The data received by the collector is recorded in the form of flows. The analyzer reads these files and generates reports in a user-friendly form.

In NetFlow, flows are sets of packets that are transmitted from a specific IP address and port to a specific IP address and port, i.e., going in the same direction and having the same parameters. Thus, the NetFlow flow is determined by means of five parameters [3]:

- source IP address;
- destination IP Address;
- source port;
- destination port;
- IP protocol code (TCP, UDP, ICMP, etc.).

When a new packet appears, key information is extracted from it, which enables to identify the flow to which the packet belongs. Next, a search is performed on the current set of flows. If a flow is found, the corresponding counters are incremented in its data. These counters include: flow lifetime, number of bytes and packets. In the cases where the required flow is not found, a new flow entry is created.

When a sensor determines that a flow is complete, the data about it is sent to a collector. The flow can be considered complete in the following cases [4]:

1) The end of the flow was detected. For example, the TCP flag FIN or RST was received;
2) No activity was observed in the flow for more than 15 seconds (or another time specified in the NetFlow configuration);
3) If flow is very long, then it must be exported at a certain time interval (depending on the protocol configuration);
4) If there are internal restrictions in protocol or software and hardware. For example, the available device memory is running out.

The collected information is stored in files of its own format; however, many analyzers enable exporting data in the form of tables in the .csv format. All fields of interest are available for export using the appropriate utilities.

After exporting the data, all information is presented in the form of a table. Each row in the table corresponds to one NetFlow flow. The list of fields available for export to NetFlow is presented in Table 1.

Exporting the set of fields, shown in the table, depends on the NetFlow settings. Additionally, with the help of various utilities, the set of displayed fields can vary.

For further processing and research of traffic received from NetFlow, an application software package was developed that performs the following functions:

- splitting the aggregated flow into incoming and outgoing flows;
- allocation of transport protocols;
- traffic intensity calculation;
- qualitative analysis of aggregated traffic - classification of applications used;
- flows transform into a TCP session.

The software was written in Python programming language using a data analysis library – pandas. With the help of these tools, the studies on the aggregated traffic of multiservice networks of a university dormitories and small offices, the traffic of individual apartments of the apartment building, and the superimposed network of the environmental monitoring system were carried out [5].

RESULTS

During the analysis, it was found that the traffic in modern multiservice networks is basically incoming, and it determines the intensity distribution of the load, as its average intensity is greater than 1-2 orders of intensity of the original traffic. This illustrates the average daily graph. It shows the intensity of incoming and outgoing traffic of the university dormitories – see Figure 1.

The graph of the intensity of incoming traffic is shown in blue, and the graph of the intensity of outgoing traffic is shown in red.

Table 1. The list of fields available for export to NetFlow

Field name	Content
Unix_secs	Record export time in seconds in Unix format.
Unix_nsecs	Residual nanoseconds of data export time.
Sysuptime	The export device running time since loading it in milliseconds.
Exaddr	Address of the exporting system.
Dpkts	The number of packets transmitted in a flow.
Doktets	The number of bytes transmitted in flow at the IP level.
First	Sysuptime system time at the time of the start of the flow.
Last	Sysuptime system time at the moment of arrival of the last packet of a flow.
Engine_type	Fields identifying sensor that collected and sent flow to collector, and its settings.
Engine_id	
Src_addr	IP address of the source of packets flow.
Dst_addr	IP address of the destination of packets flow.
Next_hop	IP address of the next router.
Input	SNMP-label of incoming interface входящего интерфейса.
Output	SNMP-label of outgoing interface.
Src_port	Packets source port.
Dst_port	Packets destination port.
Prot	IP protocol code (for example, TCP=6, UDP=17).
ToS	Type of Service label.
Tcp_flags	The sum of the flags of all TCP packets received in this flow.
Src_mask	Source address mask.
Dst_mask	Destination address mask.
Src_as	The number of autonomous system of the source.
Dst_as	The number of autonomous system of the destination.

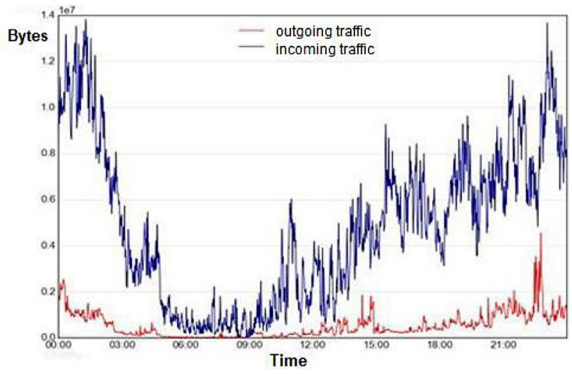

Fig. 1. The graph of the intensity of outgoing and incoming traffic

Similar results were obtained while analyzing apartment traffic. The outgoing traffic is 3-5% of incoming traffic. These correlations are illustrated by the graph shown in Figure 2, which shows the dynamics of incoming (blue) and outgoing (orange) traffic, during the month.

Table 2 presents the ratio of incoming and outgoing traffic in one of the flats of an apartment building, daily/monthly.

In addition, the tendency of preferential use of TCP protocol at the transport level was identified and confirmed. It provides information transmission with acknowledge and, in fact, establishing a virtual connection between participants of the data transfer session for the entire session, which in turn, suggests the analogies between TCP data transmission sessions and telephone connections established on traditional telephone networks – the traffic of which, for a period of more than one hundred years, has been well studied.

The study showed that the amount of information transferred by the TCP protocol significantly exceeds (by orders of magnitude) the amount of information transferred by the UDP protocol, which illustrates the average daily traffic graph of the amount of information transferred by the TCP and UDP protocols, respectively, shown in Figure 3.

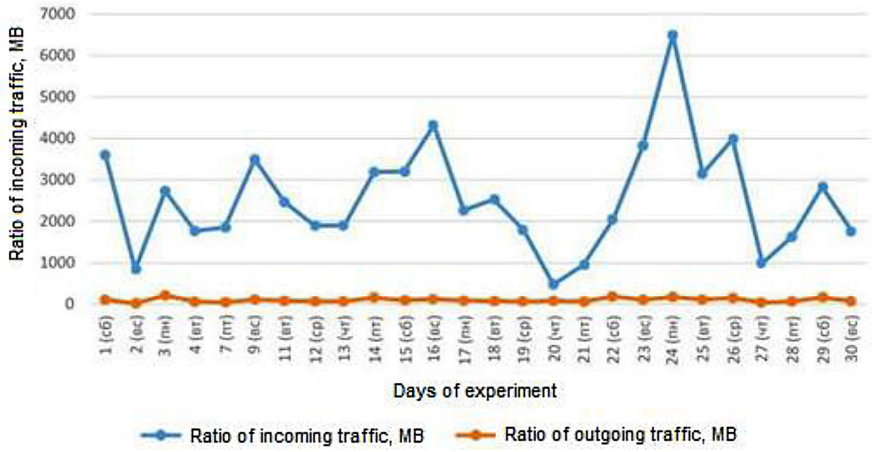

Fig. 2. Ratio of incoming and outgoing traffic

The graph of the amount of information transferred by the TCP protocol is shown in blue, and the graph of the amount of information transferred by the UDP protocol is shown in red.

Traffic transmission (using TCP protocol) is performed by TCP sessions. A single TCP session refers to establishing a connection, transferring data, and terminating a connection.

Table 2. The ratio of incoming and outgoing traffic

Day	Daily traffic volume, MB	Volume of incoming traffic, MB	The ratio of incoming traffic to all daily traffic, %	Volume of outgoing traffic, MB	The ratio of outgoing traffic to all daily traffic, %
1 (Sat)	3702.098	3594.638	97.1%	107.4601	2.9%
2 (Sun)	866.7493	845.0093	97.5%	21.73997	2.5%
3 (Mon)	2940.903	2728.565	92.8%	212.3375	7.2%
4 (Tue)	1828.379	1766.663	96.6%	61.71643	3.4%
7 (Fri)	1902.613	1852.78	97.4%	49.8336	2.6%
9 (Sun)	3606.584	3492.397	96.8%	114.1866	3.2%
11 (Tue)	2552.628	2466.499	96.6%	86.1292	3.4%
12 (Wed)	1964.989	1897.08	96.5%	67.90944	3.5%
13 (Thu)	1969.584	1900.314	96.5%	69.26988	3.5%
14 (Fri)	3347.499	3186.78	95.2%	160.7195	4.8%
15 (Sat)	3295.428	3203.312	97.2%	92.11607	2.8%
16 (Sun)	4437.191	4314.262	97.2%	122.9288	2.8%
17 (Mon)	2352.612	2264.873	96.3%	87.7396	3.7%
18 (Tue)	2604.883	2526.428	97.0%	78.45482	3.0%
19 (Wed)	1856.511	1791.635	96.5%	64.87591	3.5%
20 (Thu)	560.0585	478.0615	85.4%	81.99696	14.6%
21 (Fri)	1008.62	946.261	93.8%	62.35879	6.2%
22 (Sat)	2232.613	2046.163	91.6%	186.4503	8.4%
23 (Sun)	3931.255	3824.078	97.3%	107.1762	2.7%
24 (Mon)	6667.245	6492.768	97.4%	174.4779	2.6%
25 (Tue)	3266.534	3152.483	96.5%	114.0509	3.5%
26 (Wed)	4133.146	3980.907	96.3%	152.2397	3.7%
27 (Thu)	1040.007	997.3101	95.9%	42.69673	4.1%
28 (Fri)	1696.887	1628.743	96.0%	68.14435	4.0%
29 (Sat)	2992.466	2827.724	94.5%	164.7419	5.5%
30 (Sun)	1829.275	1757.339	96.1%	71.93542	3.9%

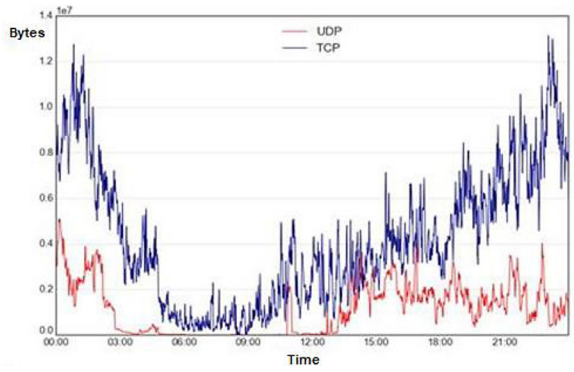

Fig. 3. Graph of the intensity of traffic transferred by the TCP and UDP protocols

A TCP session begins as soon as a packet has been sent with the SYN flag, and it has received confirmation from both sides (a triple handshake), and terminates after sending the RST or FIN flag and confirming it. Unlimited number of packets with unlimited packet length can be transmitted in one such session.

Since traffic is transmitted by TCP sessions, it is advisable to select those parameters that characterize TCP sessions as the most essential parameters characterizing the infocommunication traffic.

It should also be noted that in order to conduct research – not only among the objects of the "classical" Internet, but also to extend it to the objects of the Internet of things – it makes sense to expand the concept of "session" and treat it not only as a TCP session, but as a session meaning any process of unidirectional transfer of information, in which one can reliably determine its beginning and end. In addition, for the objects included in the Internet of Things, an important characteristic is such a value as the intervals between the arrival of messages to a server.

Thus, the statistical characteristics of traffic should include the following:

- duration of TCP sessions;
- volume (amount) of information transferred in TCP session;
- message between intervals;
- traffic intensity;

The values of all these parameters are probabilistic in nature, can change over time randomly and, accordingly, are subject to statistical analysis methods.

CONCLUSION

While analyzing the traffic of typical information objects we concluded that the study of incoming traffic transmitted at the transport level by the TCP protocol is the most significant. One of the urgent tasks involve determining the nature of the probability distributions of the above-mentioned traffic parameters and considering their possibility approximations by known "classical" distributions, which will allow, further, to simulate the traffic of these objects.

REFERENCES

1. Getman A.I., Evstropov E.F., Markin Yu.V. Real-Time Network Traffic Analysis: Overview of Applications, Approaches and Solutions. Preprint ISP RAN, 2015.
2. Goykhman V.Yu., Lushnikova T.Yu. Traffic analysis of a university dormitory. Telecommunication, 4, 2017, 51-55.
3. Goikhman V.Yu., Dremina A. Statistical analysis of network traffic, arising from downloading photos into social networks. Actual problems of infotelecommunications in science and education, 2016, 324-329.
4. Goykhman V.Yu., Sokolov N.A. Estimation of the growth of the intensity of incoming traffic" - Telecommunications, 3, 2018, 75-77.
5. Lushnikova T.Yu. Characteristic features of campus traffic. – Telecommunications, 4, 2017, 51-55.