Development of the Concept of a System for Collecting and Visualizing Statistical Network Traffic Data in the ASNET-AM Network

Gurgen Petrosyan
Institute for Informatics and
Automation problems of NAS RA
Yerevan, Armenia
e-mail: gurgen@sci.am

Eugene Prokhorenko
Institute for Informatics and
Automation problems of NAS RA
Yerevan, Armenia
e-mail: eugene@sci.am

Arthur Petrosyan
Institute for Informatics and
Automation problems of NAS RA
Yerevan, Armenia
e-mail: arthur@sci.am

***Abstract* - The ability to collect, visualize and statistically process network traffic allows to keep records, optimize loads, detect anomalous activities, predict Internet channels bandwidth needed, as well as plan the expansion and upgrade of network equipment.**

***Keywords* – Network, traffic data, collecting data, visualizing, netflow**

I. Introduction

The ability to collect, visualize and statistically process network traffic allows to keep records, optimize loads, detect anomalous activities, predict Internet channels bandwidth needed, as well as plan the expansion and upgrade of network equipment. ASNET-AM network (Academic Research Computer Network of Armenia), which connects more than 65 scientific and educational institutions, has a complex distributed infrastructure in Yerevan and other several large cities of Armenia. Because of this there are several output channels to the Internet from different locations and through different providers. Therefore we see that in our case it is necessary to have a possibility to keep traffic data records for important parameters and visualize them.

II. Concept of the System

There are two Ubuntu servers in ASNET-AM running netflow [1] collectors. Two Ubuntu servers are installed in the ASNET AM network [3], on which, using the netflow protocol, statistics is collecting. Statistics from all border routers (external traffic) is on one server, and the core routers (internal traffic) - on the other. Traffic from core routers provides greater capabilities to trace flows within the network. Tests have shown that the presence of netflow traffic does not overload the CPU of routers (MikroTik) and server monitoring (Ubuntu), since netflow traffic in ASNET-AM rarely rises above 10 mbps.

In the general case, we have traffic from our AS, traffic from ASs that have peering with us, and other parts of the Internet. We split traffic into:

1. entering our AS through border routers to the Internet
2. leaving our AS
3. internal traffic having an outgoing IP and a destination IP in our AS.

Further separation of traffic between our institutions is not needed. The main system for collecting and visualizing statistical data is a web interface, created on the basis of the PHP/MySQL interaction. Let us briefly describe the mechanism of its operation using external traffic as an example.

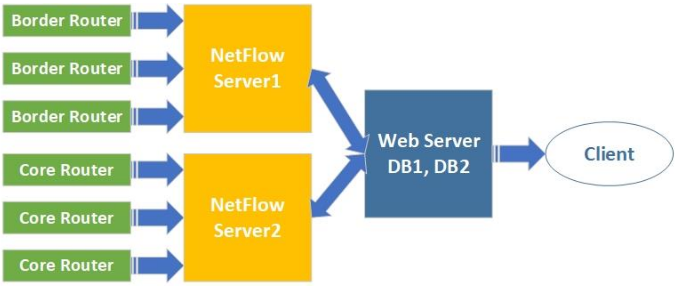


Fig. 1. Scheme of the concept of the data collection system

In the first step, the names of all ASNET-AM organizations and their real IP addresses; they can be as many as needed: separated IP addresses and subnets (one IP address is /32 subnet).

On the web server, using the “except” tool [2], which can be used to program a dialogue with interactive programs, a special script is run with a certain frequency, in our case at the beginning of every hour, using “cron”. Script establishes an ssh connection to the nfsen server (using predefined credentials (login/password or ssh key)) and executes for each organization from the database the commands of the following type:

nfdump -M [PATH_TO_DATA_FOLDER_CHANEL1] -T -R 2023/01/01/nfcapd202301011400:2023/01/01/nfcapd 202301011455 -s dstip/bytes -o csv DST/SRC NET [SUBNET1] or DST/SRC NET [SUBNET2] or …

https://doi.org/10.51408/csit2023_58

nfdump -M [PATH_TO_DATA_FOLDER_CHANEL2] -T -R 2023/01/01/nfcapd202301011400:2023/01/01/nfcapd 202301011455 -s dstip/bytes -o csv DST/SRC NET [SUBNET1] or DST/SRC NET [SUBNET2] or …

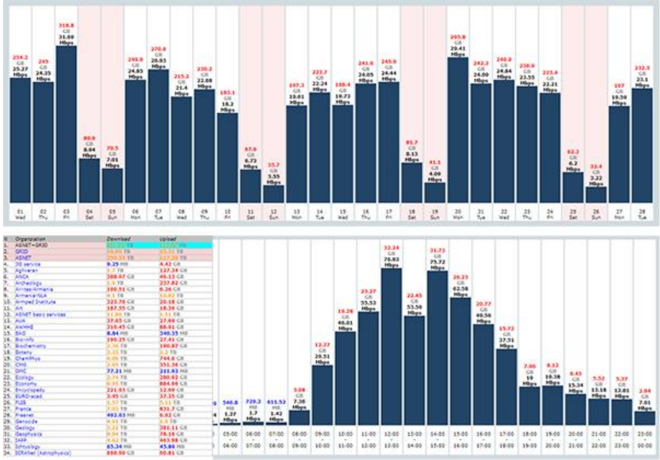
where *SUBNET1*, *SUBNET2*, etc, are the corresponding IP subnets of the given organization, and

PATH_TO_DATA_FOLDER_CHANEL1, *PATH_TO_DATA_FOLDER_CHANEL2* etc, - a path to data folders on nfsen server for this output Internet channel. For each subnet of the organization, each command is executed twice with DST NET and SRC NET, respectively, thus receiving in response both download (DST NET) and upload (SRC NET) traffic of the given subnet for the given channel. Depending on the number of output channels, the procedure is performed several times, with a different parameter PATH_TO_DATA_FOLDER_CHANEL.

The results of the above commands are summarized for each organization, and at the end we have two numbers: the total download traffic for this organization (all subnets of the organization) and, similarly, the total upload traffic. These two numbers are written in the indexed database indicating the time interval since the last retrieval of information, in our case, being 60 minutes.

In order to avoid overlaps when choosing an interval of 60 minutes, the following was done: after entering the IP addresses of organizations into the database, the duration of the system operation was estimated for one cycle with the current configuration of the web and netflow servers.

Monitoring shows that, depending on the network load, i.e., on the number of active users, which, in turn, also depends on the time of the working day, the duration of one cycle for a period of 60 minutes varies from 20 to 50 minutes. This interval strongly depends on the power of the hardware and the load of the netflow server, which performs the main function of calculating statistics and generating traffic for the web server.



N	Organization	Download	Upload
1.	ASNET+GRID	[illegible] TB	[illegible] TB
2.	GRID	18.88 TB	15.71 TB
3.	ASNET	203.33 TB	117.28 TB
4.	3D service	9.25 MB	4.42 GB
5.	Aghveran	1.7 TB	127.34 GB
6.	ANCA	388.07 GB	46.13 GB
7.	Archeology	1.9 TB	237.82 GB
8.	Arinpa-Armenia	100.51 GB	6.26 GB
9.	Armenia-NLA	4.5 TB	12.82 TB
10.	Armped Institute	323.76 GB	30.18 GB
11.	Art	187.55 GB	18.36 GB
12.	ASNET basic services	11.88 TB	1.51 TB
13.	AUA	37.65 GB	27.69 GB
14.	Awked	316.45 GB	88.91 GB
15.	BAO	8.84 MB	340.35 MB
16.	Bio-info	190.25 GB	27.41 GB
17.	Biochemistry	2.36 TB	190.87 GB
18.	Botany	3.22 TB	2.2 TB
19.	ChemPhys	4.06 TB	744.6 GB
20.	CMG	2.65 TB	351.36 GB
21.	DMC	77.21 MB	211.43 MB
22.	Ecology	2.74 TB	280.02 GB
23.	Economy	0.95 TB	884.06 GB
24.	Encyclopedy	221.93 GB	12.69 GB
25.	EURO-acad	3.45 GB	37.35 GB
26.	FLIB	1.57 TB	5.11 TB
27.	France	7.03 TB	831.7 GB
28.	Freenet	492.63 MB	6.92 GB
29.	Genocide	4.11 TB	2.6 TB
30.	Geology	3.22 TB	381.11 GB
31.	Geophysics	0.94 TB	78.16 GB
32.	IAPP	4.62 TB	463.98 GB
33.	Ichthyology	65.34 MB	45.89 MB
34.	ICRANet (Astrophysics)	660.99 GB	60.81 GB

Fig. 2. Some functionalities of visual interface

Thus, by running the script 24 times (once per hour), we, avoiding overlaps, collect all statistical data on the traffic of the specified subnets of organizations, taking into account several Internet channels. There were 66 organizations in the ASNET-AM database at the time of writing, with 1-7 subnets for each. In the presence of 4 output Internet channels, the duration of work script varies from 20 to 50 minutes depending on the workload of the channels.

After collecting statistical data, it is necessary to process and visualize them to simplify analysis. A visualization interface was created, on which, in particular, one can see statistical digital data (download / upload) for organizations by year, month, day and hour, as well as their total indicators. Some examples of the interface view are shown in Fig. 2.

References

[1] https://www.ietf.org/rfc/rfc3954.txt

[2] https://man7.org/linux/man-pages/man1/expect.1.html

[3] https://www.asnet.am/