Active Flows in Diagnostic of Troubleshooting on Backbone Links [1]

A.M. Sukhov[1),2] D.I. Sidelnikov[2)] A.P. Platonov[3)]

M.V. Strizhov[1), 3] A.A. Galtsev[1)]

August 8, 2011

[1]The research is supported partially by grant N06-07-89074 of RFBR

[2]corresponding author, *e-mail: amskh@yandex.ru*

[3]M. Strizhov is a PhD Student at the Computer Science Department, Colorado State University since May 2009

[1)] *Samara State Aerospace University, Moskovskoe sh., 34, Samara, 443086, Russia; e-mails: amskh@yandex.ru, strizhov@cs.colostate.edu, galaleksey@gmail.com*

[2)] *Institute of Organic Chemistry of RAS, Leninsky pros., 47, Moscow, 119991, Russia; e-mail: sid@free.net*

[3)] *Russian Institute for Public Networks, Kurchatova sq. 1, Moscow, 123182, Russia; e-mail: plat@ripn.net*

Abstract

In this paper, we propose a novel approach to finding and predicting anomalous network states based on a flow monitoring mechanism. We assume that number of active flows can show a real network state. Moreover, the dependence between flow number and link utilisation allows us to derive an equation for the confidence interval on high-loaded network links. Experiments have been conducted that confirmed the basic position of the model and identified the anomaly network states. A software package based on this model has been created that allows the prevention of DDoS attacks. For successful operation of this software the number of active flows that single IP address can generate has been analysed.

1 Introduction

With the growth of the global network and number of tasks, the risk of malicious intrusions is snowballing in the network. Hackers are constantly improving network attack technologies. Intrusion prevention is one of the most important challenges facing those who research the networks.

In order to resist an intrusion successfully as well as to detect the IP address source, enough time to detect the start of the intrusion is needed. There are many papers dedicated to detection and prevention of network intrusions, but it has only recently been discovered how to detect intrusions at the flow level [19, 20]. In this paper, we attempt to describe a flow-based application traffic model, which is the basis for diagnostic of troubleshooting on backbone links.

Our contribution lies in the fact that the real network state of the backbone link is described by two parameters: the number of active flows and link utilisation.

For management of the connection quality in IP networks, several recent proposals [5] refer to use of flow oriented architecture.

A network flow has been defined in many ways [4]. The traditional Cisco definition is to use a 7-tuple key, where a flow is defined as a unidirectional sequence of packets sharing all of the following 7 values:

- Source IP address
- Destination IP address
- Source port for UDP or TCP, 0 for other protocols
- Destination port for UDP or TCP, type and code for ICMP, or 0 for other protocols
- IP protocol
- Ingress interface (SNMP ifIndex)
- IP Type of Service

The basic principle of flow-aware networking relies on the fact that flows are elementary entities associated with user behavior [11, 13]. In order to provide an acceptable network quality to end users, it is important to take flows into account [3, 16] as key parameters of sharing network bandwidth. To efficiently implement flow aware resource management techniques, it is essential to estimate the number of flows which are active on a link in an operational network [1, 7].

In this paper, the analytical traffic model is constructed to diagnose network errors on the flow level. The reasons for the occurrence of network defects can vary; for example: hardware problems, insufficiency of available bandwidth, network attacks such as distributed denial-of-service (DDoS) and port scanning [14, 18]. It should be noted that the network state of a link is described by two parameters: network utilisation and number of active flows.

Barakat et al [2] propose a model that relies on flow-level statistics to compute the total (aggregate) rate of data observed on an IP backbone link. For modelling purposes, the traffic is viewed as the superposition (i.e., multiplexing) of a large number of flows which arrive at random times and stay active for random periods.

Our paper presents a technique for estimation of the network behaviour based on the utilisation curve, which represents the correlation between link utilisation and the number of active flows in it. It is supposed that the number of active flows may be considered as the *real* network state [1, 18].

The utilisation curve defines key parameters of the network: length of operational region, mean flow performance, confidence interval, points of overload, etc. Based on data received from large enterprise networks [10, 15], we will try to discover possible bottlenecks [8] and find a threshold point, where the number of arriving flows doesn't increase link utilisation.

In order to prove our hypothesis, measurements from the border gateway routers of Russian internet service provider (ISP) FREEnet were made. FREEnet (The Network For Research, Education and Engineering) is an academic and research network. The total capacity of the upstream links is approximately *2.1 Gbps*, *1.3 Gbps*. These links interconnect FREEnet with its peers (excluding FREEnet members).

We also used data from the border gateway routers of Russian ISP SamaraTelecom (ST), from HEAnet - Ireland National Research and Education Network, and from Samara State Aerospace University (SSAU).

This paper describes a simple flow-based model [1, 2] which can be used for networking monitoring and estimating quality of backbone links. We present our findings under the following headings:

- Section 2 - the review of related work
- Section 3 - operating region of network
- Section 4 - a test for network quality
- Section 5 - diagnostic of network states
- Section 6 - results from experiments

- Section 7 - traffic parameters for the model of DDoS attacks prevention

2 Previous study

In this paper traffic is considered as a stationary process, using the results from the papers of Barakat et al [2] and Ben Fredj et al [3]. They proposed a traffic model for uncongested backbone links that is simple enough to be used in network operations and engineering. The model used by Barakat et al relies on Poisson shot-noise. With only three parameters (λ, arrival rate of flows, $\mathbb{E}[S_n]$, average size of a flow, and $\mathbb{E}[S_n^2/D_n]$, average value for the ratio of the square of a flow size and its duration), the model provides approximations for the average of the total rate (the throughput) on a backbone link and for its variations at short timescales. The model is designed to be general so that it can be easily used without any constraints from the definition of flows, or on the application or the transport protocol.

Define $B(t)$ as the total rate of data (e.g., in *bits/s*) on the modelled link at time t. It is determined by adding the rates of the different flows. We can then write

$$B(t) = \sum_{n \in \mathbb{Z}} X_n(t - T_n) \tag{1}$$

The process from Eq. (1) can describe the number of active flows N found at time t in an $M/G/\infty$ queue [12], if $X_n(t - T_n) = 1$ at $t \in [T_n, T_n + D_n]$.

The model presented by Barakat et al [2] can compute the average and the variation of traffic on the backbone. In summary:

- The average total rate of the traffic is given by the two parameters λ and $\mathbb{E}[S_n]$:
$$\mathbb{E}[B(t)] = \lambda \mathbb{E}[S_n] \tag{2}$$
- The variance of the total rate $\mathbb{V}[B(t)]$ (i.e., burstiness of the traffic) is given by the two parameters λ and $\mathbb{E}[S_n^2/D_n]$:
$$\mathbb{V}[B(t)] = \lambda \mathbb{E}[S_n^2/D_n] \tag{3}$$

It should be mentioned that Eq. (2) is true only for the ideal case of a backbone link of unrestricted capacity, which can be applied to underloaded links. The main drawback of the ratio (2) is its lack of definite usage limits, due to the fact that the variables $\lambda, \mathbb{E}[S_n]$ describing the system are in no way connected with its current state. The average flow size $\mathbb{E}[S_n]$ does not depend on a specific system; it is a universal value determined by the current distribution of file sizes found in the Internet.

The arrival rate of flows λ describes the user' s behaviour, and doesn' t depend on the network state and utilisation. The cumulative number of flows that arrive at a link will remain linear even if the network has problems and doesn' t satisfy all the incoming demands.

In order to describe the real network state with an arbitrary load we should use Little' s law:

$$N = \lambda \mathbb{E}[D_n] \tag{4}$$

Here, $\mathbb{E}[D_n]$ is the mean duration of flow and N is the mean number of active flows. Formula (4) is true for any flow duration [12] and thus for an arbitrary flow size distribution and rate limit. This formula describes the network state more precisely than Eq. (2), as the average number of active flows on the bandwidth unit increases with the utilisation. In other words, the average duration of flow $\mathbb{E}[D_n]$ enables us to judge the real network state in contrast to its average value $\mathbb{E}[S_n]$.

3 Operating region of the network

In order to analyse the connection quality at the backbone area or the link to the provider we are going to investigate a graphical dependence between the link utilisation U and the number of active flows N in it [1]. These variables are easily measurable quantities in spite of the average values $\mathbb{E}[D_n]$ and $\mathbb{E}[S_n]$. The separate network state is pictured by single point on a coordinate plane with axes N and U. The curve depicting average values has been shown in Fig. 1.

On the curve shown in Fig. 1, three parts can be identified, corresponding to the different network states [17]. The first part of the curve describes the network state close to the ideal. If the investigated link has unrestricted capacity, there should be a stable linear relationship between the number of flows N and link utilisation U. The curve describing the network behavior beyond a certain point will be convex. The linear part of the curve corresponding to the ideal network behaviour is defined as the operational region. The operational region ends at the threshold point which should be found experimentally. The dislocation of this point depends on many factors, such as transport layer protocol, network topology, the amount of buffering at the link, etc.

The second part of the curve corresponds to the moderately loaded network, when the diversion from the ideal network state becomes obvious. There is an increase in the average duration of a flow compared to the working area, and therefore, a larger number of active flows on the bandwidth unit characteristic of this network state.

The third part of the curve corresponds to the totally disabled network, with considerable packet loss evident. We propose some simple preliminary models for an overloaded link, accounting for user impatience and reattempt behaviour. In a real network, if demand exceeds capacity, the number of flows in progress does not increase indefinitely. As flow throughput decreases, some flows or sessions will be interrupted, due either to user impatience or to aborts by TCP or higher layer protocols.

In this section, estimation of the confidence interval is given for the operational region of our curve. Since the total rate is the result of the multiplexing of $N(t)$ flows of independent rates, the Central Limit Theorem [12] tells us that the distribution of $B(t)$ tends to normal (Gaussian) at high loads, which is typical of backbone links.

It should be noted that the statement about normal distribution is reasonable for networks in which there are no significant deviations. For example, such deviations include disabling external or internal links, a DDoS attack to the network, or port scanning. For such anomalous situations, the distribution of B (t) no longer tends to Gaussian. In this case, consistent points of the network state will go beyond the confidence intervals. A method of detecting network anomalies could be based on this principle [9].

As is mentioned in Section 2, the variance of the total rate requires two parameters: the arrival rate of flows λ and the expectation of the ratio between the square of the size of a flow duration $\mathbb{E}[S_n^2/D_n]$ (see Eq. 3). This tells us that the total rate should lie between $\mathbb{E}[B] - A(\varepsilon)\sqrt{\mathbb{V}(B)}$ and $\mathbb{E}[B] + A(\varepsilon)\sqrt{\mathbb{V}(B)}$ in order to provide the required quality of service. When we talked about the required quality of service we implied the accordance of network behavior to Eq. (2) - here $A(\varepsilon)$ is the ε-quantile of the centered and normalised total rate $B(t)$.

Taking into account Eqs. (2-4) and theorems about average values, the confidence interval of the bandwidth B on an operational region of our curve is

$$B = b(N \pm \alpha A(\varepsilon)\sqrt{N}). \tag{5}$$

Here, $b = E[S_n/D_n]$ is the average flow performance which characterises the speed of user communication. Coefficient α could be found experimentally.

In the event of a network anomaly, two or more sequential measurements will exceed the confidence intervals, which are easy to find using diagnostic tools.

In our recent work, emulation of DDoS attacks has been carried out using the network tool LOIC [9]. The test shows that the number of active flows increases greatly when network utilisation is low. We have developed a utility that allows us to determine the start of the attack and to allocate

the IP addresses from which it takes place. In this case, intrusion detection technique is based on the measuring of active flows from single IP address which is under test (See Section 7).

4 Testbed Setup

In order to prove our hypothesis measurements were made on border gateway routers from FREEnet, HEAnet - Ireland' s National Research and Education Network, from the Russian ISP SamaraTelecom (ST), and also from Samara State Aerospace University (SSAU) network. All networks have several internal and external links. ST' s basic load lies on one channel to the Internet, whereas HEAnet and FREEnet relies on a number of connections. Measurements from Gigabit links were taken for FREEnet and *155 Mbps*, *622 Mbps* are for HEAnet and ST. The utilisation of these links varies widely from 5% to 60% with a clearly identifiable busy period.

A passive monitoring system based on Cisco' s NetFlow [6] technology was used to collect link utilisation values and active flow numbers in real-time. In Moscow and Samara we measured on a Cisco 7206 router with NetFlow switched on. At HEAnet a Cisco 12008 was utilised. A detailed description of Cisco NetFlow can be found in the Cisco documentation [6].

This is achieved using the following commands on the Cisco 7206:

- sh ip cache flow - gives information about the number of active and inactive flows; about the parameters of the flows in the real time.
- show interface summary - gives information about the current link utilization.

On a GSR Router these commands look like:

- enable
- attach slot-number
- show ip cache flow

The FREEnet data was obtained using scripts running every 30 minutes from the middle of January to the end of March 2008. The data sets from two routers of FREEnet have been collected for further analysis. The full loading of the routers varies in limits of hundreds of megabits per second (*100-220 Mbps*) for the first router and tens of megabits for the second router (see Fig. 5). During the tests we fixed information about any network events that could have an influence on connection quality. The ST data was recorded at

30-minute intervals, twenty-four hours a day for a week, to discover network behaviour with different loading levels. The HEAnet data was obtained using scripts running every 5 minutes for a period of 72 hours. It is quite easy to write a script which will collect the data from the router to the management server.

5 Diagnostic of network states

In order to perform experimental testing of our model, the data set needed to be divided into several intervals depending on number of active flows. Inside each interval, the average values and their variance were calculated for flow performances as well as for other parameters characterising network states.

The earlier tests have been conducted on the boundary routers of ST and HEAnet, and they didn' t allow verification of the model with high precision. The new data from FREEnet contains thousands of points describing network states.

FREEnet Data Set 1 has been divided into seven intervals according to the number of active flows (15000 - 20000, 20000 - 25000, 25000 - 30000, 30000 - 40000, 50000 - 60000, $\geq$60000). Inside each interval, basic parameters characterising active flows have been calculated, and the result is represented in Table 1. Here, N is the average number of active flows for the interval denoted by n. B describes the average router loading in *Megabits per second (Mbps)*, and b is average flow performance measured in *bits per second (bps)*. $\sigma(b) and \sigma(B)$ are the standard deviations for flow performance b and router loading B respectively.

The results of the measurement for FREEnet Data Set 1 are pictured in Fig. 2. A basic curve is constructed as the line of average values; it describes network states on flow level. The error bar restricts the confidential interval for network states. Comparison with the theoretical prediction from Section 2 leads to the conclusion that the area of network exploitation lies inside the operational region.

In order to restrict the operational region, the straight portion of the curve from Fig. 2 should be marked as shown in Fig. 3. The sloped angle of this straight line is found as the average flow performance b. Only three initial points from the investigated data set may be placed in the limits of the operational region. The angle of inclination gives an average flow performance equal to 5700 *bps* for FREEnet. If the number of flows exceeds 30000, then the network gets moderately loading which leads to a reduction in the flow performance. The router loading does not increase uniformly with the number of requests, and the connection quality becomes almost twice as

bad (see Table 1).

In the conclusion of this section it should be noted that expression (5) allows formulation of the easy rule of how to display the network defects. If two consistent measurements running every 5(30) minutes give the deviation of the real network states B_i, N_i from the confidence interval with $A(0.05)$, then a network problem has been detected. The confidence interval is described by flow performance b and the values $\alpha, \sigma(B)$ which may be found only as result of data processing.

This rule received an apt illustration during FREEnet network testing. A network incident has been detected: a wide number of links to a large FTP server have been temporarily turned down. These anomaly network states B_i, N_i depart from the confidence interval corresponding to standard behaviour of the investigated network, and form a separate cluster as shown in Fig. 6.

So our model receives the experimental confirmation and the diagnostic method based on introduction of the confidence interval may be applied to network monitoring.

6 Statistical tests

The network state at every instant can be described by a point on a two-dimensional plot where abscise shows the number of active flows N and ordinate shows router loading B. Basic tasks of experimental validation of our model consist of

- construction of curve of average values and comparison with the theoretical prediction shown in Fig. 1
- calculation of variance for flow performances and verification of the parabolic form of the confidence interval from Eq. (5)
- examination on normal distribution for flow performances in the absence of network state deviation.

In Table 2, the Data Set 2 from the second router of FREEnet with lower loading is presented. The investigated region divides into six intervals according to the number of active flows (5000 - 6000, 6000 - 7000, 7000 - 8000, 9000 - 10000, $\geq$10000). The operational region for the second router of FREEnet is restricted by 10 000 active flows, as shown in Fig. 4. Only the last intervals should be excluded from the straight portion of utilisation curve.

Fig. 5 illustrates the network states and the form of the confidence interval for an operation region with normal quantile function $A(\varepsilon), \varepsilon = 0.05$.

A correlation coefficient indicates the strength and direction of a linear relationship between two random variables. In order to verify the parabolic form of the confidence interval the correlation coefficient between variables $\sigma_i(B)$ and $\sqrt{N_i}$ should be calculated for both data sets of FREEnet (see Tables 1 and 2). Comparing the second and fourth columns of the aforementioned tables, the correlation coefficient values are equal to 0.70 for Data Set 1 and 0.93 for Data Set 2. These magnitudes allow us to discuss high correlation between the theoretical model and its experimental examination.

A significant question concerns the numerical value for the numerical coefficient α_1, α_2 from Eq. (5). The function $A(\varepsilon)$ can be computed using the Gaussian approximation, which gives, for example, $A(0.05) = 1.96$. Our data from Tables 1 and 2 allow calculation of their magnitudes:

$$\alpha_1 \approx 13, \alpha_2 \approx 4.5 \tag{6}$$

Significant assumption underlies a theoretical model that distribution of flow performances b may be considered as a Gaussian distribution. The number of testing network states inside many intervals from Tables 1 and 2 allow us to check a given set for similarity to the normal distribution. Here, we use the Pearson χ^2 test, the results of which can be found in Tables 3 and 4. Column 2 shows the value of χ^2 for $\alpha = 0.95$; table values for χ^2 are shown in round brackets. It should be noted that all investigated intervals with a sufficient number of states discover the normal type of distribution.

7 Traffic parameters for the model of DDoS attacks prevention

This section discusses the number of completed flows, which IP addresses generate during normal activity, at a time when there is no outside intrusion. During the attack, the number of completed flows amounts to several thousand per minute from a single attacker' s address. Samara State Aerospace University developed and launched a set of utilities for protection against DDoS attacks. This method is based on a flow traffic model, presented in this paper.

The principle of attack detection is that the number of flows increases drastically at constant link utilisation during the attack. The purpose of this section is to clarify the traffic parameters that are used in software to prevent attacks. Two parameters are most important; namely the maximum number

of active flows that a single IP address can generate and type of distribution of the number of flows from single IP address, ranked by popularity. It should be noted that the software collects statistics once per minute and all the analysed data is of this time interval. Netflow statistics was used to find the parameters from the boundary router of SSAU. The results are shown in Figs. 7-9.

A typical graph of the distribution of active flows from a single IP address, ranked by popularity, is shown in Fig. 7. The number of active flows N_i and the sequence number i of the addresses in the ranked list are plotted on a logarithmic scale. Points on the graph lie on a straight line, suggesting that the distribution conforms with Zipf's law [21]:

$$N_i = N_1/i \tag{7}$$

To find the maximum number of active flows per single IP address, statistics for the week have been analysed. The resulting graph is shown in Figs. 8 and 9. The x-axis displayed the time in minutes since the beginning of observation; the vertical axis represents the maximum number of flows N_1.

Graphs in Figs. 8 and 9. show that the maximum number of flows from single IP address does not exceed 100 for the server 91.222.129.201 and not more than 20 for 91.222.128.200 during normal network activity.

For the future we are planning to make the intrusion prevention system (IPS) to defend against DDoS for public data centers. The major parameters of this software include:

- *sFlow* protocol instead of *NetFlow*
- *JAVA* application (for use in the Linux operating system), which will also handle the statistics obtained directly from the router without any intermediate agents.
- firewall *iptables*

The foundation of the software under development will be based on the traffic model, which describes the state of the network by two parameters: the number of active flows and link utilisation. When two sequential measurements go beyond the confidence interval an error message will display.

8 Conclusion

In this paper we showed that simple modelling methods can estimate a quality of links and predict an appearance of bottlenecks in large networks. At

the moment we are working on developing utilities, which will make it possible to construct the dependence of link loading on the number of active flows automatically, define attacks and calculate the length of the operation region as well as coefficients for the confidence interval.

This paper has demonstrated that it is possible to easily determine the confidence interval, operation region and the overload point of a network connection, utilising low cost hardware and simple software. This allows us to identify the anomaly network states including network attacks and the moment when a backbone upgrade is required. Further experiments are necessary in order to develop software utilities for this purpose. Thus, providing analytical generalisations, we established common terminology for processes, taking place in networks.

References

[1] F. Afanasiev, A. Petrov, and A. Sukhov, A Flow-based analysis of Internet traffic, Russian Edition of Network Computing, 5(98) (2003) 92-95 (arXiv:cs/0306037)

[2] C. Barakat, P. Thiran, G. Iannaccone, C. Diot, P. Owezarski P., A flow-based model for Internet backbone traffic, IEEE Transactions on Signal Processing - Special Issue on Signal Processing in Networking, vol. 51, no. 8 (2003) 2111-2124

[3] S. Ben Fredj, T. Bonald T., A. Proutiere, G. Regnie, J. Roberts, Statistical Bandwidth Sharing: A Study of Congestion at Flow Level, ACM SIGCOMM, August 2001

[4] N. Brownlee, C. Mills, G. Ruth, Traffic Flow Measurement: Architecture (RFC 2722), October 1999

[5] Y. Chabchoub, C. Fricker, F. Guillemin, and P. Robert, A Study of Flow Statistics of IP Traffic with Application to Sampling, Lecture Notes in Computer Science, 4516 (2007) 678-689

[6] Cisco IOS NetFlow site, Cisco Systems, http://www.cisco.com/go/netflow/

[7] L. Deri, nProbe: an Open Source NetFlow Probe for Gigabit Networks, TERENA 2003

[8] C. Fraleigh, F. Tobagi, C. Diot, Provisioning IP Backbone Networks to Support Latency Sensitive Traffic, INFOCOM, Volume: 1 (2003) 375-385

[9] A. A. Galtsev and A. M. Sukhov, Network attack detection at flow level, NEW2AN/ruSMART 2011, Lecture Notes of Computer Science, 6869 (2011) 326-334

[10] R. Lippmann, J. Haines, D. Fried, J. Korba and K. Das, The 1999 DARPA off-line intrusion detection evaluation, Computer Networks, 34(4) (2000) 579-595

[11] Y. Jiang, P. Emstad, A. Nevin, V. Nicola, and M. Fidler, Measurement based admission control for a flow-aware network, in Next Generation Internet Networks, (2005) 318-325.

[12] L. Kleinrock, Queueing Systems, Wiley, NY, 1975, Vol. I: Theory

[13] A. Kortebi, L. Muscariello, S. Oueslati, and J. Roberts, Minimizing the overhead in implementing flow-aware networking, in ANCS 05: Proceedings of the 2005 symposium on Architecture for networking and communications systems. New York, NY, USA: ACM Press (2005) 153-162

[14] John McGlone, Alan Marshall, Roger Woods, "An Attack-Resilent Sampling Mechanism for Integrated IP Flow Monitors," icdcsw, pp.233-238, 2009 29th IEEE International Conference on Distributed Computing Systems Workshops, 2009

[15] NSS Group, Intrusion Detection Systems Group Test (Edition 4), NSS Group, 2004

[16] S. Oueslati and J. Roberts, A new direction for quality of service: Flow aware networking, in Next Generation Internet Networks (2005) 226-232

[17] K. Papagiannaki, N. Taft, Z.-L Zhang, C. Diot, Long-Term Forecasting of Internet Backbone Traffic: Observations and Initial Models, INFOCOM 2003

[18] W. Yang W., J. Gong, W. Ding, X. Wu, Network Traffic Emulation for IDS Evaluation, IFIP International Conference on Network and Parallel Computing, ISBN: 978-0-7695-2943-1 (2007) 608-612

[19] Daniel Reichle, Analysis and Detection of DDoS Attacks in the Internet Backbone using Netflow Logs, Diploma Thesis DA-2005.06, TIK, ETH Zurich, 2005

[20] G. Munz, G. Carle, Real-time Analysis of Flow Data for Network Attack Detection, in Proc. pf 10th IFIP/IEEE International Symposium on Integrated Network Management (IM'07), 2007, pp. 100-108

[21] George Kingsley Zipf. Relative frequency as a determinant of phonetic change. Reprinted from the Harvard Studies in Classical Philiology, Volume XL, 1929.

Table 1: Parameters for active flows of FREEnet, Data Set 1, 2008

n	N	B, *Mbps*	$\sigma(B)$, *Mbps*	b, *bps*	$\sigma(b)$, *bps*
1	17489	113.1	23.1	6784	1386
2	23260	126.0	21.4	5682	965
3	27007	152.0	39.2	5628	1452
4	34902	156.7	26.9	4990	770
5	45104	163.9	33.9	3634	752
6	55019	176.3	33.2	3205	604
7	64778	215.4	42.2	3325	652

Table 2: Parameters for active flows of FREEnet, Data Set 2, 2008

n	N	B, *Mbps*	$\sigma(B)$, *Mbps*	b, *bps*	$\sigma(b)$, *bps*
1	5446	15.42	2.25	2843	413
2	6531	17.11	2.45	2364	377
3	7508	17.74	2.35	2364	313
4	8370	18.92	2.24	2261	268
5	9443	20.67	3.81	2190	404
6	15495	28.05	5.40	1811	349

Table 3: Statistical tests, Data Set 1, 2008

n	χ^2 for $\alpha = 0.95$	Gaussian test	Correlation coefficient
1	not enough data		
2	not enough data		
3	not enough data		
4	3.49 (9.49)	+	0.70
5	not enough data		
6	3.45 (7.81)	+	
7	0.50 (9.49)	+	

Table 4: Statistical tests, Data Set 2, 2008

n	χ^2 for $\alpha = 0.95$	Gaussian test	Correlation coefficient
1	9.15 (12.6)	+	
2	10.0 (11.1)	+	
3	3.24 (14.1)	+	0.93
4	10.0 (14.1)	+	
5	not enough data		
6	1.94 (14.1)	+	

Figure 1: Link utilisation vs. the number of active flows

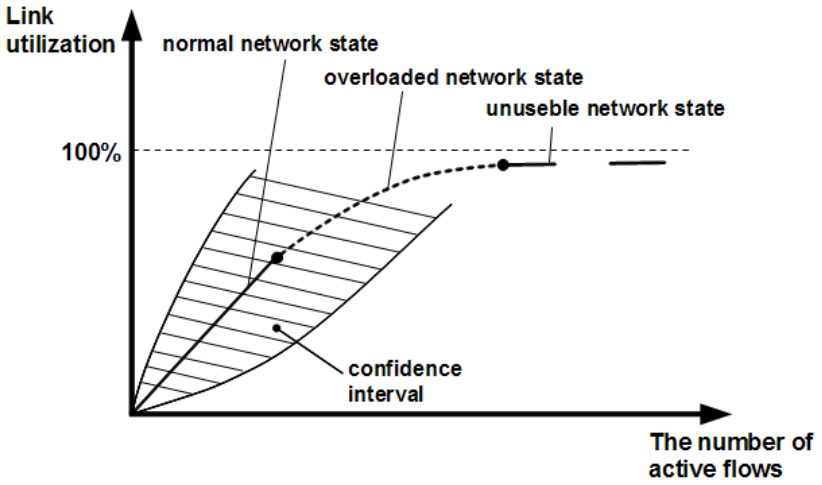

Figure 2: FREEnet router loading vs the number of active flows, Data Set 1

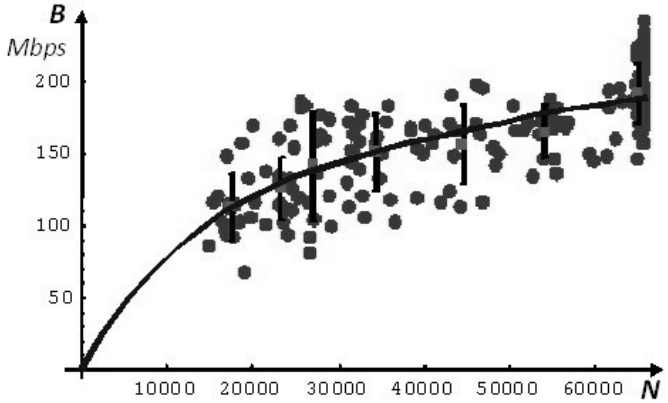

Figure 3: The operational region of network, Data Set 1

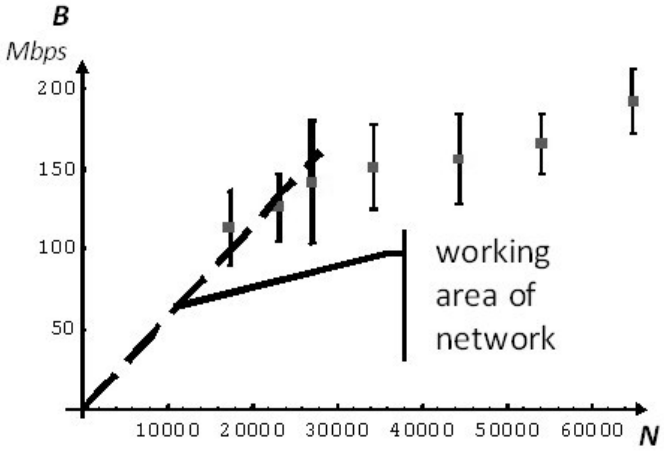

Figure 4: The operational region of network, Data Set 2

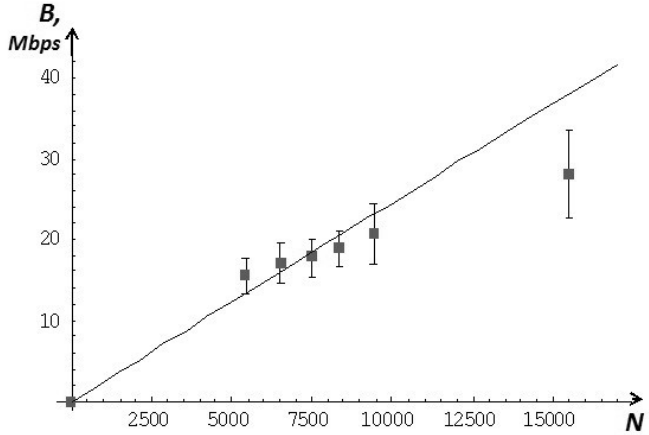

Figure 5: Confidence interval for operation region, Data Set 2 from FREEnet

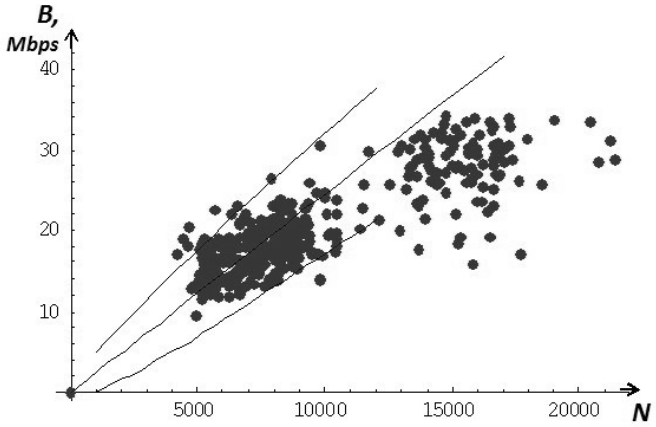

Figure 6: Detection of anomaly network state

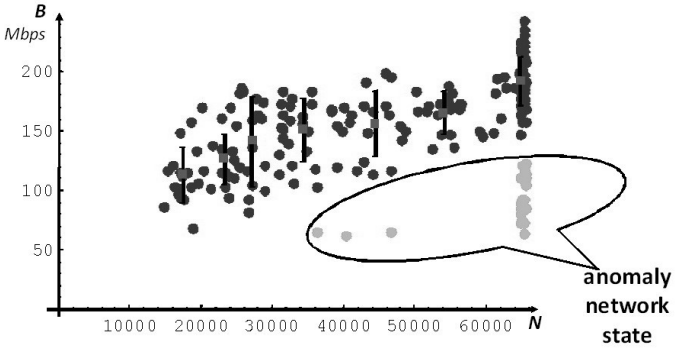

Figure 7: The number of active flows from a single IP address in descending order

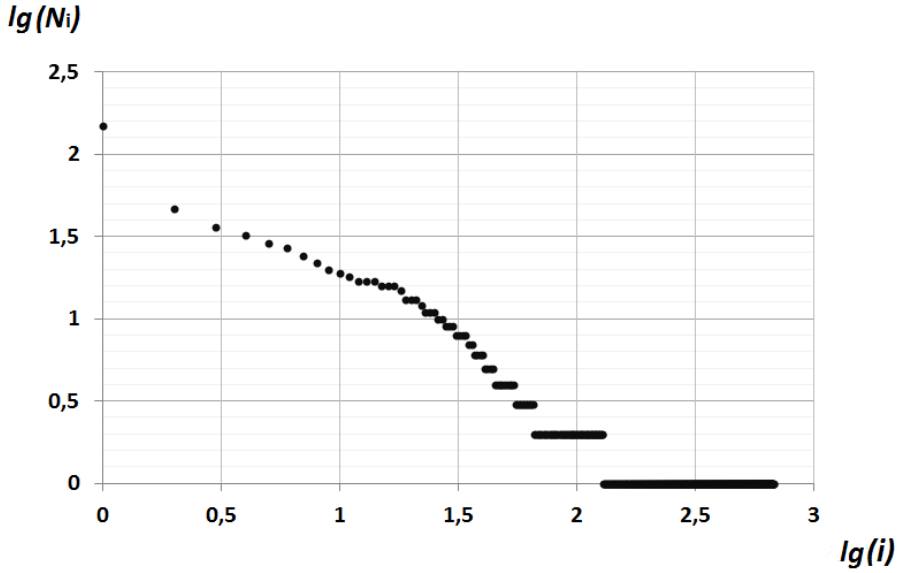

Figure 8: The maximum number of flows from single IP address, server 91.222.129.201

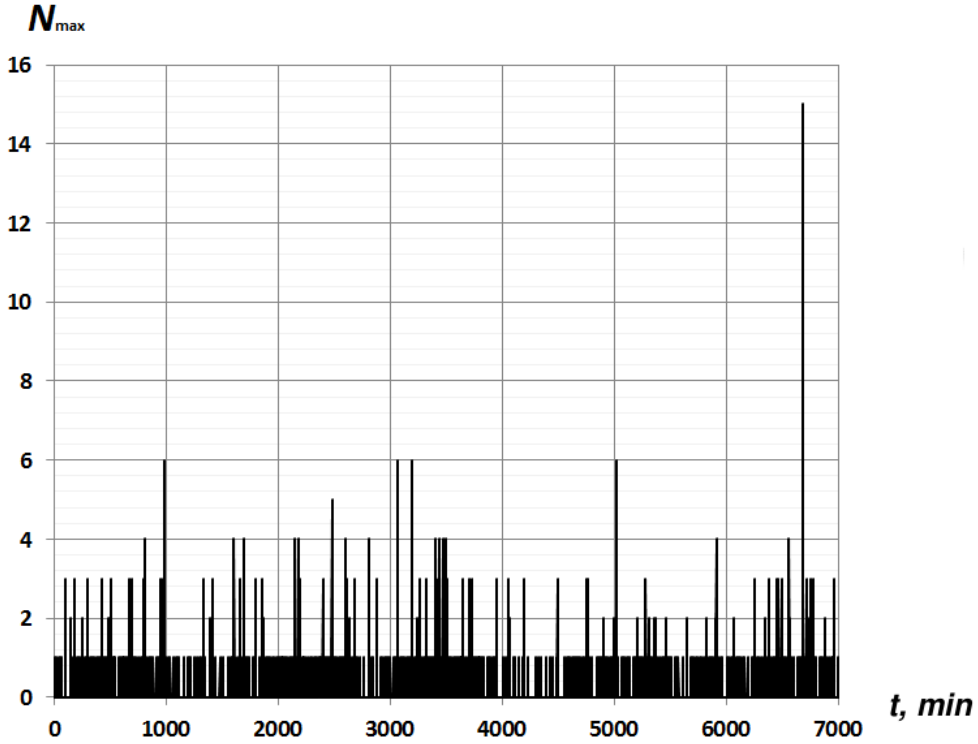

Figure 9: The maximum number of flows from single IP address, server 91.222.128.200