Detection of Abrupt Changes in Network Traffic Dynamics

V.V. Ivanov[1], **Val.V. Ivanov**[1], **Yu.A. Kryukov**[2], **P.V. Zrelov**[1]
[1]*Laboratory of Information Technologies, JINR*
[2]*University "Dubna", Dubna*

1. Introduction. We propose new methods for detection of abrupt changes in the network traffic dynamics. In our study we use the measurements obtained at the input of the Dubna University local area network (LAN) with approximately 200-250 interconnected computers. We analyze data flows, i.e. values of packet sizes transmitted within n subsequent time intervals (preliminary aggregation of the data has been performed). The methods use a self-nonself discrimination scheme based on a comparison of samples composed from the successive values of the analyzed time-series. We demonstrate that these methods can be successfully applied to reveal differences in the network dynamics for various periods of observation.

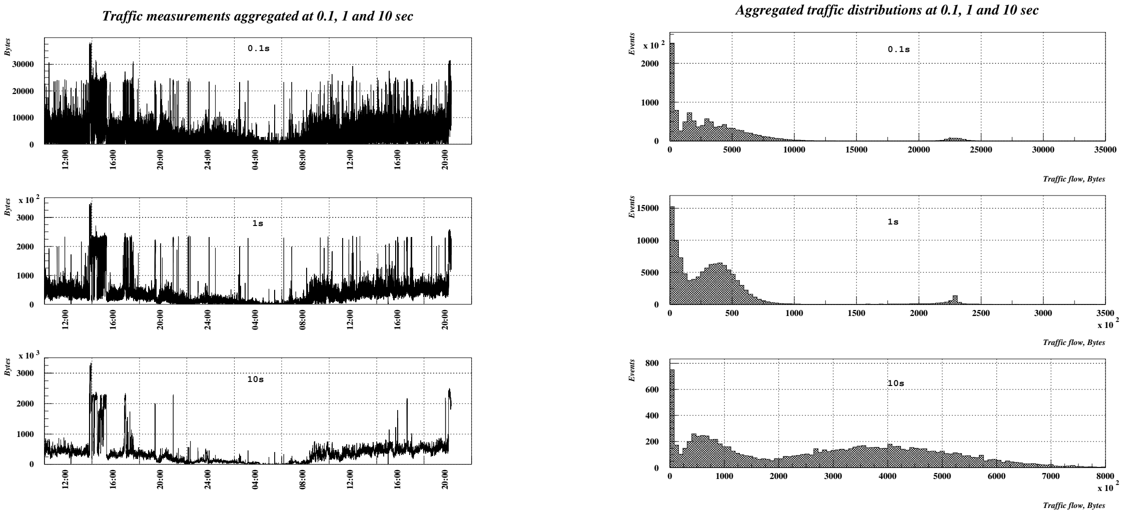

Figure 1: Left figure: Aggregated (at 0.1 sec, 1 sec and 10 sec) time-series of the traffic flow. $5 \cdot 10^6$ points, 32 hours (including day- and night-time). Right figure: Distribution of aggregated traffic flow at 0.1 sec, 1 sec and 10 sec

2. Data. The duration of the data set taken at the input of the Dubna University local area network is around 32 hours, including day- and night-time (the data contain measurements which belong to different states of the system). The data were preliminary aggregated at $0.1, 1$ and 10 seconds levels. The data sets demonstrate quite an unstable character. This means definite changes in the properties of internal dynamics of the system. The regions corresponding to different time series regions can be considered as "before" and "after" some change-point on time axes.

3. Log-normal distribution. Following our previous investigations [1, 2], one should expect that the statistical distribution of aggregated traffic measurements should have (in general) a character of the log-normal law. The correspondence of distributions to the log-normal form is different for various data sets. In some cases, a uniform (or close to uniform) "background" in the analyzed statistical distribution may exist. In other cases there are noticeable deviations from the log-normal form in the region of small values of the analyzed distribution. In order to overcome such problems, additional parameters b and c are introduced:

$$f(x) = b + \frac{A}{\sqrt{2\pi}\sigma(x-c)} \exp^{-\frac{1}{2\sigma^2}(\ln(x-c)-\mu)^2} . \qquad (1)$$

4. Change-point detection in the network traffic dynamics. We present different change-point detection algorithms which are able to identify changes in the network traffic dynamics, including hacker's attacks.

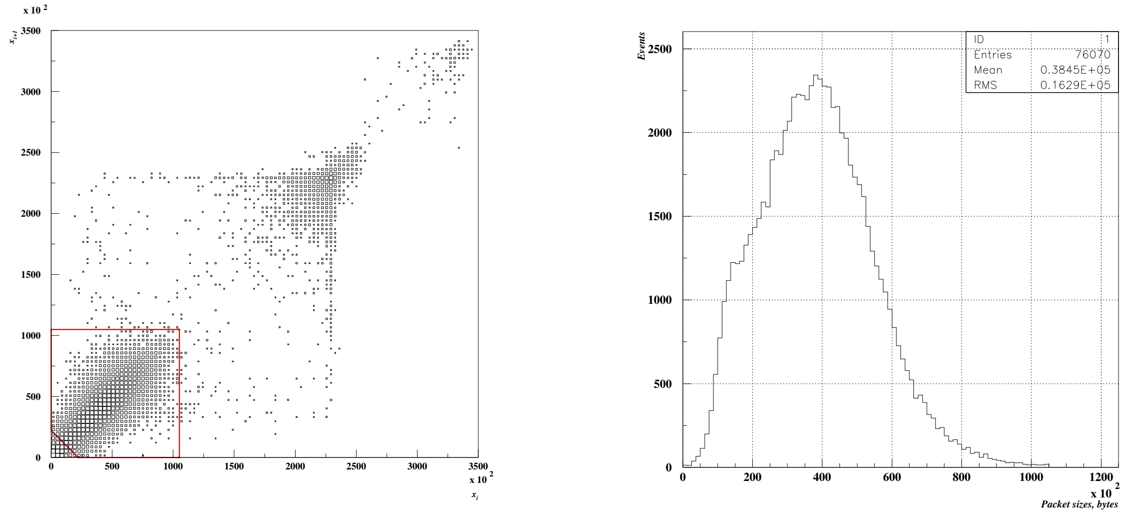

Figure 2: Left figure: 2D delay plot ($\tau = 1$) for aggregated at 1 sec time-series of the traffic flow (32 hours). Right figure: Distribution of the traffic flow (after the cut)

We propose a new approach for change point detection in the time-series, based on immunocomputing principles. This approach can be considered as an extension of the negative selection algorithm. The modified algorithm could be formalized as follows:

- Define a set consisting of the samples[1] (of predefined volume n) obtained from successive values of the time-series.
- Construct a vector of some general parameters, which characterize the sample.
- Define "self" as a collection S of the vectors which concern with the system behavior (considered as "normal").
- Define a matching rule (which can be either logical or probabilistic) for distinguishing between "self" and "nonself" vectors.

In this study we consider (for simplicity) only two-dimensional vectors. The method is well illustrated in this case by using two-dimensional plots.

6. Time delay method. The method consists in construction of the plot in delayed coordinates $(x_i, x_{i+\tau})$ (in our case $\tau = 1$) and consequent classification of different attractors a "normal" or "abnormal". The parameters are: sample volume $n = 2$, sample type – "moving" ($\tau = 1$), vector – $(x_i, x_{i+\tau})$, rule – appropriate region in 2D - plane.

[1] We consider two types of samples: $(x_1, ..., x_n), (x_{n+1}, ..., x_{2n}), \ldots$ and "moving" samples $(x_1, ..., x_n), (x_{1+\tau}, ..., x_{n+\tau}), \ldots$, where $\tau = 1, 2, \ldots$

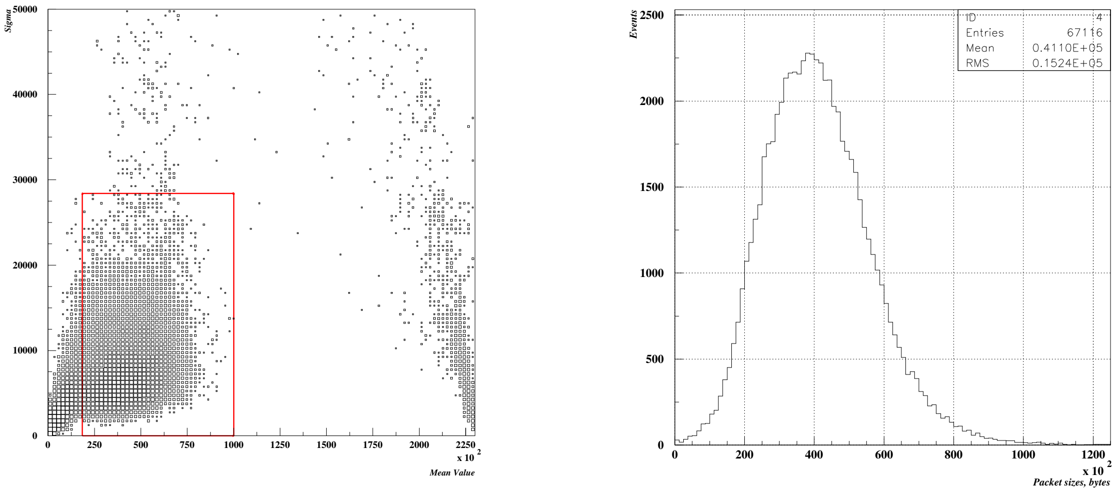

Figure 3: Results of application of the "moving sample" method to aggregated at 1 sec traffic (all 32 hours). Left figure: σ Vs "Mean value" plot. Right figure: Distribution of the traffic flow (after the cut)

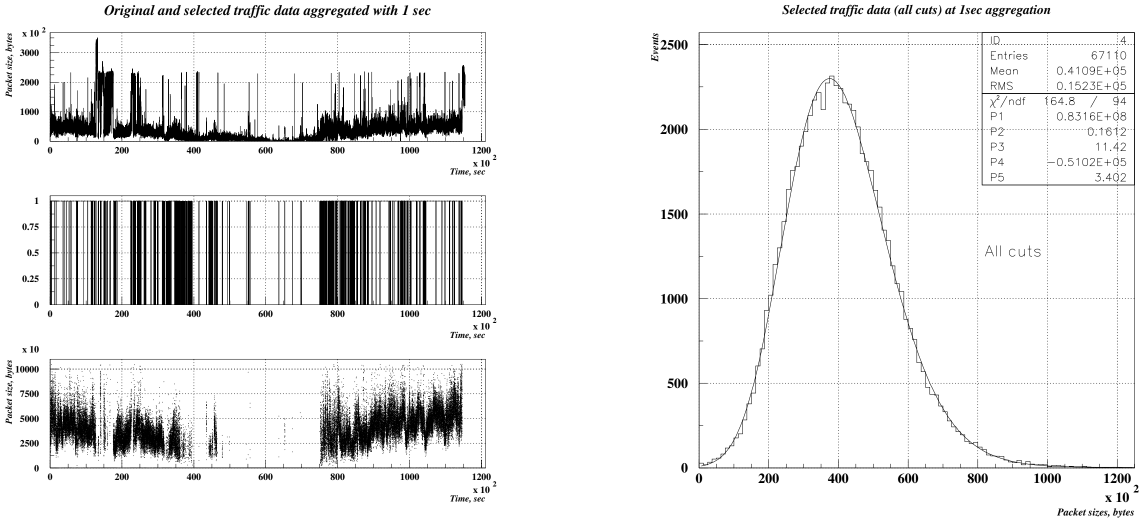

Figure 4: Aggregated (at 1 sec) time-series of the traffic flow (32 hours). Moments of the structural changes - 237 intervals with duration more than 10 sec, 63 - more than 1 min, 25 - more than 5 min, 6 - more than 30 min, 1 - more than 2 hours (at night). "Filtered" traffic - basic (daily) state (18.5 hours, i.e. 58% of the hole time)

7. A "moving sample" method. We consider mean values and dispersions for "moving" samples from original time-series of volume 10. Here the parameters of the method are: sample volume $n = 10$, sample type – "moving", vector – $(\mu_j, \sigma_j))$, matching rule – appropriate region in $(\mu, \sigma))$ - plane.

8. Discussion of results. Fig. 1 shows that the distributions of the aggregated traffic flow (at different levels 0.1 sec, 1 sec and 10 sec) are not log-normal distributions. They have a composite structure of different distributions. Figures 2 and 3 show the results of applying different cuts in the correspondent 2D-planes of the "Time delay" and "Moving sample" methods. Right plot in Fig. 4 shows the distribution of the "filtered" traffic (after the all cuts), corresponding to a basic (daily) state (18.5 hours, i.e. 58% of the hole time). This distribution follows the log-normal law (1). Left plot in Fig. 4 shows the moments of the structural changes in the dynamics - 237 intervals with duration more than 10 sec, 63 - more than 1 min, 25 - more than 5 min, 6 - more than 30 min, 1 - more than 2 hours (at night). Fig. 5 shows only night traffic aggregated at different levels and right plot in Fig. 4 shows the corresponding distribution of traffic flow aggregated at 10 sec level, superimposed by the fitting curve of the log-normal function.

9. Conclusion. We propose the methods for change-point detection in the Internet traffic dynamics. These methods use a preliminary aggregation of the data at a certain level. The methods use a self-nonself discrimination based on comparison with the sets consisting of the samples of successive values of the time-series. In the case of the "moving sample" method the sample is presented by two values - dispersion and the mean value, while in the case of the time delay method the sample is formed by two successive time-series values.

We demonstrate that these methods can be successfully applied to reveal differences in dynamics for different periods of observation and that the statistical distributions of "cleaned" data are approximated with a good accuracy by the log-normal functions.

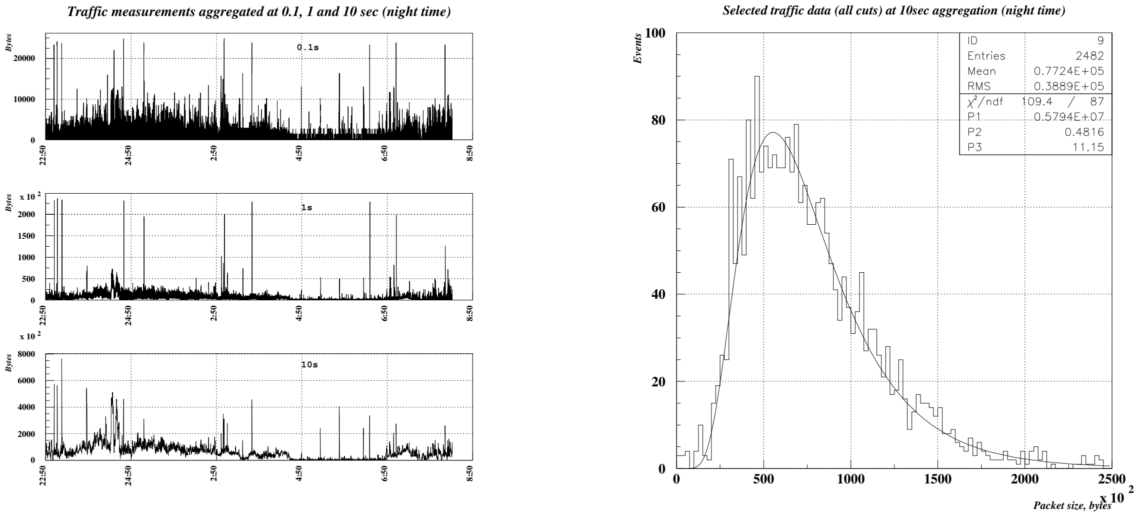

Figure 5: Left figure: Aggregated (at 0.1 sec, 1 sec and 10 sec) traffic (night-time). Right figure: Distribution of aggregated at 10 sec "night" traffic flow

References

[1] I. Antoniou, V.V. Ivanov, Val.V. Ivanov and P.V. Zrelov: *On the Log-Normal Distribution of Network Traffic*, Physica D, 167 (2002) 72-85.

[2] V.V. Ivanov, Val.V. Ivanov, Yu.A. Kryukov and P.V. Zrelov: *Detection of abrupt changes in network traffic dynamics*, In: Int. Conf. "Distributed computing and Grid-technologies in science and education", Dubna, June 29 - July 2, 2004, Book of abstracts, p.86.