Adaptive method of detecting traffic anomalies in high-speed multi-service communication networks

Sergey Ageev[1,*] , *Vladimir* Karetnikov[2], *Evgeny* Ol'khovik[2], and *Andrey* Privalov[3]

[1]Radioavionica JSC, 4 lit. B Troitsky pr., St. Petersburg 190005, Russia
[2]Admiral Makarov State University of Maritime and Inland Shipping, 5/7 Dvinskaya srt., St. Petersburg, 198035, Russia
[3]Emperor Alexander I St. Petersburg State Transport University, 9 Moskovsky pr., Saint Petersburg, 190031, Russia

Abstract. In the paper, an adaptive hybrid heuristic (behavioral) method for detecting small traffic anomalies in high-speed multiservice communication networks, which operates in real time, is proposed and investigated. The relevance of this study is determined by the fact that network security management processes in high-speed multiservice communication networks need to be implemented in a mode close to real-time mode, as well as identifying possible network security threats in the early stages of the implementation of possible network attacks. The proposed method and algorithm belong to the class of adaptive methods and algorithms with preliminary training. The average relative error in estimating the evaluated traffic parameters does not exceed 10%, which is sufficient for the implementation of operational network management tasks. Anomalies of the expectation of traffic intensity and its dispersion are identified if their valuesexceed the normal values by 15% or more, which makes it possible to detect possible network attacks in the early phases of their implementation, for example, at the stage of scanning ports and interfaces of the attacked system. The procedure for detecting anomalous traffic behavior is implemented based on the Mamdani's method of hierarchical fuzzy logical inference. A study of the proposed method for detecting anomalous behavior of network traffic showed its high efficiency.

1 Introduction

The successes achieved in the development of telecommunication and communication technologies have led to the creation and implementation of the concept of a multiservice communication network (MCN), the basis of which is packet IP networks that integrate various voice, data and multimedia transmission services [1, 2]. However, the emergence of a large number of additional services at the MCN makes the actual problem of reliable provision of its network and information security (NIS) [3].

* Corresponding author: serg123_61@mail.ru

Traffic in the MCN is very diverse [4 - 6]. It consists, among other things, of multimedia traffic, which is very sensitive to delays, data transmission traffic, signaling information traffic, email traffic. At the same time, the specified requirements for the quality of services must be fully implemented. There are objective difficulties in building the NIS of MCN. These difficulties are caused by the complexity of the MCN structure, the large spatial scope of the network infrastructure, the need for quick and high-quality analysis of a large number of various dynamically changing network, information characteristics and parameters.

Therefore, the operational continuous assessment and detection of the anomalous behavior of high-speed network traffic with a priori unknown, dynamically changing characteristics is one of the key tasks of managing the MCN network, as well as its NIS, is an urgent scientific problem.

2 Analysis of methods for assessing the characteristics and parameters of traffic in high-speed MCN

As is known [5, 6], traffic in the MCN can be approximated using probability distributions of Poisson, Pareto, Weibull, log-normal distribution and exponential distribution. Traffic in the MCN is non-stationary in its nature, and the mathematical models that adequately describe its behavior are non-linear stochastic models [6]. This fact makes it difficult to implement procedures for evaluating the parameters and characteristics of network multiservice traffic with the required quality under conditions of a priori uncertainty both with respect to its current probabilistic distribution law and its parameters.

One of the constructive approaches to solving the problem of estimating the vector parameters of random processes with nonlinear observation models is the conditional non-linear Pareto - optimal filtering method [7, 8]. The essence of this approach is that the estimation of a vector unknown parameter is carried out in two stages. At the first stage, the function of the current forecast of estimates of the values of the vector parameter is calculated. At the second stage, with the help of corrective functions and the obtained additional posterior information on the values of these estimates, they are corrected. The choice of the class and type of assessment functions for the current forecast, the class and type of correction functions is quite free and is determined by the specific formulation of the problem being solved.

In this study, based on the concept of conditional non-linear Pareto - optimal filtering, a method and algorithm for detecting anomalous traffic behavior are developed using joint estimates of the current value of the mathematical expectation and dispersion (standard deviation (SD)) of the MCN traffic intensity. The adaptation of correction functions to unknown characteristics of the MCN traffic intensity is proposed to be performed using pseudo-gradient procedures, the general theory of which was laid down in [9–11]. In this case, the adjustment of the parameters of the correcting functions depending on the parameters of the random sequence is carried out using Takagi-Sugeno fuzzy logic inference [12, 13], taking into account the dynamics of changes in their values. Using the fast Fourier transform method [14], the spectral power densities of the obtained estimates of the values of the mathematical expectation and dispersion of the MCN traffic are located in a sliding window. By their increments, which are determined in two sliding windows, the anomalous behavior of the traffic of the multiservice communication network is determined.

3 Method and algorithm for detecting traffic anomalies in high-speed multiservice communication networks

Let the MCN traffic observations be presented in the form of a random sequence (RS) *x(i)* having finite mathematical expectation and dispersion given at discrete time instants $t = i = \{1, 2, ..., n,..\}$ and described additively – multiplicatively a model having the form:

$$x(i) = \Theta(i) \cdot w(x(i-1)) + \xi(i), \quad (1)$$

where *w(*)* is a random function of the observations, $\Theta(i)$ is a random variable, and $\xi(i)$ is the interference of observations with zero mathematical expectation and finite dispersion.

It is necessary to construct a vector recursive procedure for estimating the mean values of the mathematical expectation of a random sequence *x(i)* and its standard deviation by the criterion of the minimum mean square error, of the form:

$$J(i) = M\{\bar{\varepsilon}\} = \{M\,(m(i) - \hat{m}(i))^2 \rightarrow min,\ M(\sigma(i) - \hat{\sigma}(i))^2 \rightarrow min\ \}, \quad (2)$$

where $\hat{m}(i)$, $\hat{\sigma}(i)$ are the estimates of the mathematical expectation and standard deviation of the random sequence *x(i)* at step *i*, and $m(i)$, $\sigma(i)$ are their true values at this step.

The forecast function for the current value of the mathematical expectation of the random sequence is defined as:

$$\hat{m}(i) = \frac{1}{N}\sum_{k=1}^{N} x(i-k), \quad i = 1, 2, .., n, .., \quad (3)$$

where *N* is the size of the sliding window, which is selected relatively small size [6].

Further, the forecast of the estimation of the standard deviation of the random sequence in step *i* is made in the same sliding window:

$$\hat{\sigma}(i) = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N} x^2(i-k) - (\frac{1}{N}\sum_{k=1}^{N} x(i-k))^2} \quad (4)$$

Further consideration of the construction of the corrective procedure will be carried out for the component of the value of the mathematical expectation of functional (2), with a generalization to the vector case.

The value of the functional $J(\hat{m}(i))$ may not be observable, and only the implementation of its gradient with a random error is available:

$$\nabla Q\big(\xi, \hat{m}(i)\big) = \nabla J\big(\hat{m}(i)\big) + \xi, \xi \in R^n, \quad (5)$$

where ξ is the error of observing the gradient. Let ξ be the centered, uncorrelated errors in estimating the gradient of the quality functional. Functionality (5) will be minimized using a recurrence algorithm of the form:

$$\hat{\hat{m}}\,(i+1) = \hat{\hat{m}}\,(i) - \lambda_m(i+1)\nabla Q\big(\xi, \hat{m}(i+1)\big) \quad (6)$$

where $\nabla Q\big(\xi, \hat{m}(i+1)\big)$ is a random direction in the phase space at the point $\hat{m}(i+1)$, $\hat{\hat{m}}\,(i)$ is the adjusted estimate of the mathematical expectation at the previous step, $\{\lambda_m(i)\}$ is a sequence of positive numbers, which for a stationary random sequence, must satisfy the conditions of Dvoretsky [9 - 11].

$$\sum_{i=1}^{\infty} \lambda_m(i) = \infty, \quad \sum_{i=1}^{\infty} \lambda_m^2(i) < \infty \quad (7)$$

In accordance with [9, 10], the vector $\nabla Q\big(\xi, \hat{m}(i)\big)$ is called a pseudogradient at the point $\hat{m}(i)$, if the condition is satisfied at this point:

$$\nabla J\big(\hat{m}(i-1)\big) \cdot M\{\nabla Q\big(\xi, \hat{m}(i)\big) \geq 0\ \}, \quad (8)$$

where $M\{*\}$ is the operation of mathematical expectation, that is, the vector $\nabla Q\big(\xi, \hat{m}(i)\big)$ on average is an acute angle with the gradient vector of the quality functional $\nabla J\big(\hat{m}(i-1)\big)$. The implementation of the quality functional at the point $\hat{m}(i+1)$, in accordance with [6, 9 - 11], can be represented as follows:

$$\nabla Q\big(\xi, \hat{m}(i+1)\big) = (\hat{m}(i+1) - \hat{\hat{m}}\,(i))^2. \quad (9)$$

After simple algebraic transformations, the recursive pseudo-gradient algorithm (PGA) for estimating the current value of the mathematical expectation will look like:

$$\hat{\hat{m}}\,(i+1) = \hat{\hat{m}}\,(i) + \lambda_m(i+1)\left(\hat{m}(i+1) - \hat{\hat{m}}\,(i)\right). \quad (10)$$

If the distribution density of the values of the random sequence $\hat{m}(i)$ $p(\hat{m})$ is symmetric with respect to the mathematical expectation, then it is possible to use a pseudo-gradient algorithm of the form:

$$\hat{\hat{m}}\,(i+1) = \hat{\hat{m}}\,(i) + \lambda_m(i+1)\varphi\left(\hat{m}(i+1) - \hat{\hat{m}}\,(i)\right), \quad (11)$$

where a non-decreasing monotonic function can be used as the function $\varphi(*)$, for example, the sign function $\varphi(*) = sign\,(*)$. It was noted in [9–11] that the use of this function makes it possible to increase the stability of pseudo-gradient algorithm to errors in estimating the gradient of the quality functional.

A generalization of algorithm (11) is a vector pseudo-gradient algorithm for estimating the parameters of a random sequence, having the form [6]:

$$\hat{\hat{G}}\,(i+1) = \hat{\hat{G}}\,(i) + R(i+1) \times \nabla Q(i+1) \quad (12)$$

where $\hat{\hat{G}}\,(i+1)$ is the vector of estimates of the parameters of the random sequence at step *i* + 1, which can be represented as:

$$\hat{\hat{G}}\,(i+1) = \left[\hat{\hat{m}}\,(i+1), \hat{\hat{\sigma}}(i+1)\right]^T. \quad (13)$$

The matrix $R(i+1)$ is the diagonal matrix of step coefficients of the estimated parameters.

Regarding algorithms (11) and (12), one can formulate statements that:

1. These algorithms are pseudo-gradient algorithms. The proof of this statement is based on the correct verification of condition (8). A consequence of this statement is the fact that these procedures have all the properties of pseudo-gradient algorithms [9 - 11].

2. The structure of algorithms (11) and (12) is invariant with respect to the statistical characteristics of random sequence *x(i)*, with an accuracy determined by the accuracy of identification of its parameters. The proof of this statement is based on the application of the central limit theorem [15]. The consequence of this statement is that for any probabilistic properties of traffic, the structure of the algorithm for estimating its parameters is constant, only its settings can change.

To evaluate the parameters of non-stationary random sequences, condition (8) restricts the use of pseudo-gradient algorithm, since the pseudo-gradient algorithm must monitor changes in the value of traffic parameters, and not converge to their specific values. Therefore, it is proposed to restrict the sequence $R(i+1)$ from below to a constant value. As a consequence of choosing a limited step coefficient, the dispersion of the estimation of the parameters of the random sequence will also be limited from below. Therefore, it is necessary to find a compromise solution between the speed and accuracy of estimating the parameters of the random sequence [6, 9 - 11].

It is proposed, when choosing the step coefficient vector, to take into account the dynamics of changes in the estimated parameters and characteristics of the random sequence. Obviously, the moduli of the gradients of the components of the vector quality functional are proportional to the dynamic properties of the random sequence. Such dependencies are in the nature of hard-to-formulate tasks, therefore, it is proposed to automate the pseudo-gradient algorithm step coefficient adjustment using the Takagi-Sugeno fuzzy logic inference method or based on its particular form, the singleton method [12, 13], which has the form:

$$\begin{aligned} \boldsymbol{IF} < \hat{\hat{G}}\,(i) \in D1 > \boldsymbol{OR} < \nabla Q(i) \in D2 > \boldsymbol{OR} < \hat{\sigma}(i) \in D2 > \text{TO}\ R(i+1) \\ = R(z)\ \boldsymbol{AND}\ N = Nk \end{aligned} \quad (14)$$

To implement these rules, the fuzzy logical inference system is preliminarily trained according to the experimental data obtained at the stage of its design, in test random sequences with known statistical parameters [6]. An increase in the size of the sliding

window, if such a need arises, is carried out sequentially, with a step equal to one cell of the sliding window. This allows us to ensure the observability of the estimated parameters of the random sequence. The structure of the fuzzy inference system during operation remains constant.

For the obtained estimates of the mathematical expectation and dispersion of the traffic intensity, the power spectral density is determined:

$$\tilde{m}(i,\omega) = \sum_{n=0}^{N_F} \left\{ \frac{1}{N_F} \sum_{n=i}^{i+N_F} \hat{\hat{m}}\,(i+n) e^{j\omega n} \right\}^2, \qquad (15)$$

$$\tilde{\sigma}(i,\omega) = \sum_{n=0}^{N_F} \left\{ \frac{1}{N_F} \sum_{n=i}^{i+N_F} \hat{\hat{\sigma}}\,(i+n) e^{j\omega n} \right\}^2,$$

where N_F is the basis of the fast Fourier transform. After obtaining current estimates of spectral powers (15), their sum is determined in successive sliding windows W_1 and W_2 (Figure 1). In this figure, arrow 1 conditionally shows the direction of movement of the sliding windows W_1 and W_2. If the total power density of any random sequence (14) in window W_1 exceeds its value in window W_2 by a predetermined value, then a decision is made about the presence of anomalous traffic behavior.

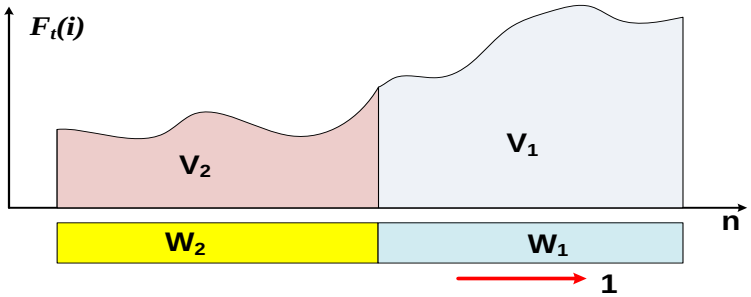

Fig. 1. W_1, W_2 - sliding windows, 1- direction of movement of the sliding windows; V_1 and V_2 - values of the power spectral densities in the corresponding sliding windows.

4 Analysis of the results of an experimental verification of the traffic anomaly detection algorithm

Mathematical modeling of the verification of the effectiveness of the developed algorithms for assessing the characteristics of the MCN traffic was carried out for bistochastic traffic with a log-normal distribution. In this case, the mathematical expectation and dispersion of the processes were modeled using first-order autoregression processes (AR-1).

Figure 2 shows a graph of traffic intensity for a changing amplitude of mathematical expectation.

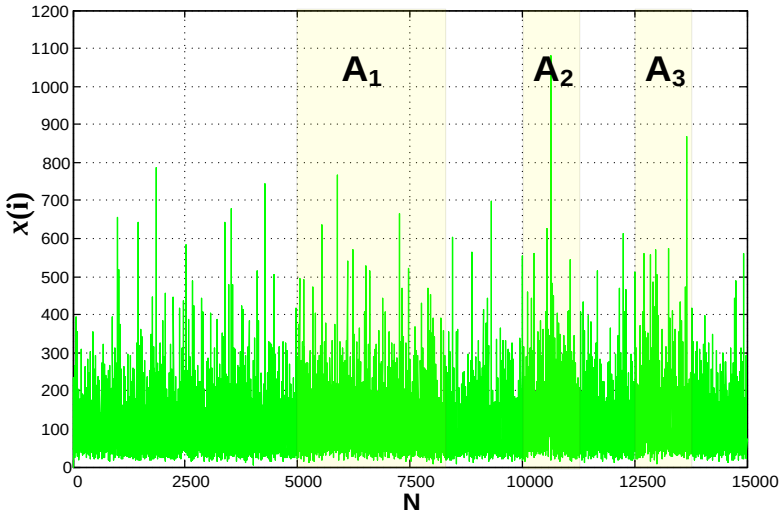

Fig. 2. The traffic intensity of the MCN with a log-normal distribution. Zones A_1, A_2 and A_3 – the presence of traffic anomalies

Areas A_1, A_2, and A_3 correspond to the presence of small anomalies in the mathematical expectation of traffic intensity. Figure 3 shows the results of evaluating the current value of the mathematical expectation of the random sequence. The average relative error in estimating the mathematical expectation and dispersion was less than 3.8%.

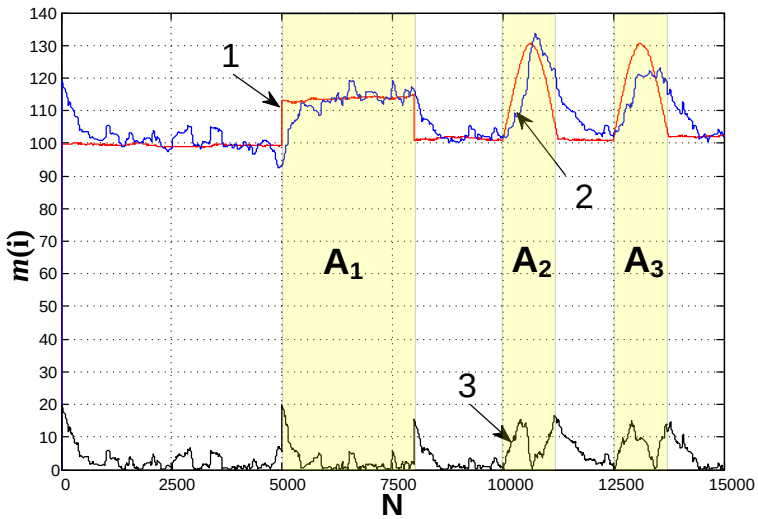

Fig. 3. Results of the assessment of the current value of the mathematical expectation of the random sequence *x(i)*. 1 – the true current value of the mathematical expectation of traffic intensity, 2 – the value of the estimate of the expectation of traffic intensity, 3 – the value of the absolute error module of the estimate of the mathematical expectation of traffic intensity. A_1, A_2, A_3 – zones of abnormal behavior of the current value of the mathematical expectation of traffic intensity

Figure 4 shows examples of assessing the current value of the power spectral density (PSD) of the sequence of estimates of the mathematical expectation of traffic intensity (1), as well as the values of the flags for detecting traffic intensity anomalies of the MCN (2).

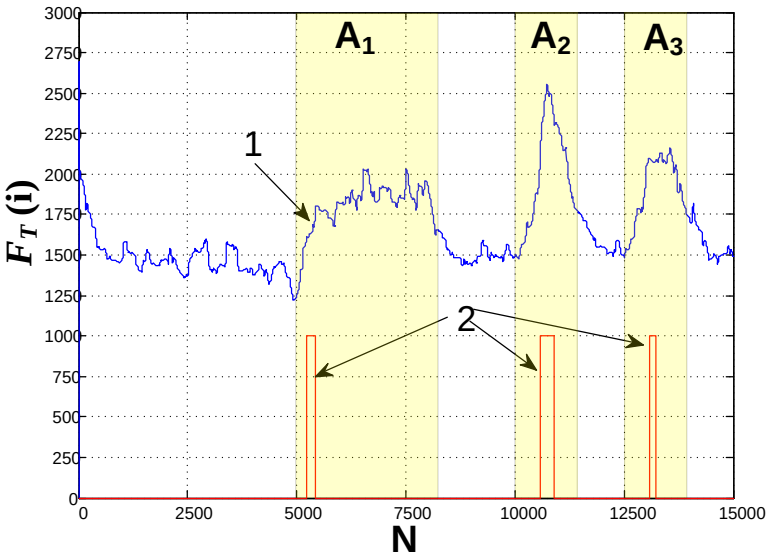

Fig. 4. Assessment of the current value of the power spectral density (PSD) of the sequence of estimates of the mathematical expectation of the traffic intensity of the MCN. 1 – power spectral density, 2 – flag values for detecting anomalies in traffic intensity of the MCN

In the course of computational experiments, the following results were achieved.

The estimation of traffic intensity distribution parameters, namely, the current value of the mathematical expectation and dispersion, was implemented in hard real-time mode with an average relative error of less than 10%. The detection of a sharp, explosive change in traffic parameters, which is typical for the main phases of network attacks, was also detected in real time. The detection of small traffic anomalies, which is characteristic of the initial phases of network attacks, was detected with a delay of 120 - 180 μs (as shown in Figure 4), which is an acceptable result. The structure of the proposed method and algorithm for detecting the abnormal behavior of MCN traffic allows them to be implemented on parallel computing platforms. The developed method and algorithms have shown stable, with high accuracy, detection of anomalous behavior of MCN traffic under conditions of a dynamic change in its characteristics.

5 Conclusion

The obtained accuracy and dynamic characteristics of the developed method and algorithm ensure the detection of abnormal behavior of MCN traffic in high-speed multiservice communication networks with the required quality. The preliminary analysis performed in the work showed the possibility of hardware-software implementation of the developed algorithms on the existing hardware platform [16 - 18]. The most promising is the implementation of the algorithm as an intelligent agent for a multi-agent intelligent system of operational decision support. The hardware basis of such a system can be a system on a chip (SoC) and FPGA.

References

1. J. Song, M. Chang, S. Lee, J. Joung, IEEE Communications Magazine, **45(9)**, 116–123 (2007). doi:10.1109/mcom.2007.4342866

2. *Appendixes F: ITU-R Recommendation Titles. Hargrave's Communications Dictionary* (2009) doi:10.1109/9780470544822.app6
3. *ISO/IEC 27001:2005 «Information technology. Security techniques. Information security management Systems. Requirements»*
4. V*ertical QoS Mapping. QoS Over Heterogeneous Networks*, 201–223 (2007). doi:10.1002/9780470058763.ch8
5. T.M. Tatarnikova, A.V. Volskiy, Information and Control Systems **3(94)**, 54–60 (2018). doi:10.15217/issn1684-8853.2018.3.54
6. E.A. Rudenko, Journal of Computer and Systems Sciences International, 57(1), 43-62. (2018). doi: 10.7868/s0002338818010067
7. Yu.K. Ziatdinov, Problems of Informatization and Management, 1(29), 70–75 (2010). doi:10.18372/2073-4751.1.603
8. Ageev S. A., Gladkikh A. A., Kurnosov V. I., Privalov A. A. Adaptive method of detecting traffic anomalies in high – speed multiservice communication network. *H&ES* Research. 2019. Vol. 11. No. 5. Pp. 4 – 13. doi: 10.24411/2409-5419-2018-102-82 (in Russian).
9. I.N. Zaitseva, Vestnik of Astrakhan state technical university. Series: management, computer science and informatics, **3**, 90–96 (2019). doi: 10.24143/2072-9502-2019-3-90-96
10. Ya.I. Rabinovich, Doklady Akademii nauk, 462(2), 151–153 (2015). doi: 10.7868/s086956521514008x
11. M. Sugeno, T. Takagi, Readings in Fuzzy Sets for Intelligent Systems, 15(1), 387–403 (1993).
12. *Introduction to Fuzzy Control and Modeling. Fuzzy Control and Modeling*, (2009). doi:10.1109/9780470544730.ch2
13. H.J. Nussbaumer, *Fast Fourier Transform and Convolution Algorithms. Springer Series in Information Sciences*, **2** (Springer, Berlin, Heidelberg, 1982). doi:10.1007/978-3-642-81897-4_3
14. *Mathematical Statistics. Mathematical Statistics and Stochastic Processes.* Pp. 1–2 (2013). doi:10.1002/9781118562024.part1
15. URL: https://www.altera.com (date of access 10.10.2019).
16. URL: http:// www.xilinx.com (date of access 10.10.2019).
17. URL: http://opencores.org/projects (date of access 10.10.2019).
18. A. Butsanets, E. Ol'khovik, *2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)*. IEEE, 1-3 (2019). doi: 10.1109/fareastcon.2019.8934708