https://doi.org/10.59849/aidd.2024.37

Research Classification Traffic Flows Based on Machine Learning to Ensure Quality of Service in Radio Systems and Mobile Cellular Networks

Beytulla Huseynov
Department of Radio Engineering and Telecommunications,
Azerbaijan Technical University, Baku, Azerbaijan,
E-mail: h.beytulla63@gmail.com

Abstract. Methods for the efficiency classifying traffic packets based on the technology distributed communication networks such as Software Defined Networking, Internet Protocol Multimedia Subsystem, and Machine Learning are analyzed in order to ensure a certain level quality of service (QoS) and quality of experience (QoE) and reliability of stream transmission packets with an increase in the volume of transmitted useful and service traffic in radio systems and mobile cellular networks. This paper examines the problem researching methods for classifying flows of useful and service traffic packets using machine learning. To create a feature matrix for real-time traffic classification in order to maintain QoS, QoE and the reliability of the transmission packet flows, a new approach is proposed, which is based on calculating the statistical characteristics of packet flows.

Keywords: classification, accuracy, flow, machine learning, IMS, package, F1-measure.

1. Introduction.

Currently, it remains an urgent task to increase the throughput of modern radio systems and mobile cellular networks based on the architectural concepts of the following NGN (Next Generation Networks) and future FN (Future Networks) networks and the results of the Network 2030 study (Network 2030), carried out by the ITU-T FG NET-2030 focus group to study the capabilities and principles building fixed communication networks for the period up to 2030 and beyond [1, 2].

It is worth noting that such NGN and FN concepts make it possible to obtain certain advantages in the organization and management information flows in packets, as well as to achieve greater flexibility in accessing the resources wireless networks and hardware and software complexes of radio systems. Here, a more dynamic development of the existing network infrastructure, in contrast to the traditional approach to building multi-service cellular networks, requires the creation of a method for dynamic classification of traffic flows based on machine learning to ensure quality of service and reliability of message transmission in real time.

Consequently, in this direction, modern technologies for fixed and wireless networks of the next (ITU-T, Y.2000) and future generation (ITU-T, Y.3000) include machine learning (Machine Learning, ML), SDN (Software Defined Networking), NFV (Network Functions Virtualization), IMS (Internet Protocol Multimedia Subsystem), artificial intelligence, mobile 4G-LTE (Long Term Evolution), IoT (Internet of Think), 5G–NR-U (New Radio-Unilence) so and 6G [2, 3].

The conducted research shows [2, 3] that among the above, a special place is occupied by SDN, IMS and ML technologies in order to ensure quality of service QoS (Quality of Service), quality of perception QoE (Quality of Experience) and reliability of transmission traffic messages that became one of the most important tasks.

In works [1, 4] devoted to classification for the purposes of QoS, QoE and transmission reliability, the main problems are real-time operation and the impossibility of adding new classes to the existing model of mobile cellular networks when introducing new information technologies.

This paper examines the tasks research and analysis methods for classifying flows useful and service traffic packets using machine learning traffic in radio systems and mobile cellular networks based on SDN, IMS and ML technologies.

2. General statement of the research problem.

Conducted research and analysis show [1, 3, 4, 5] that radio engineering systems and mobile cellular networks based on SDN and IMS technologies have significantly greater capabilities for providing QoS, QoE and classification of useful and service traffic flows based on machine learning.

Therefore, managing heterogeneous traffic within this facility is most effective when studying the quality of functioning radio systems and mobile cellular networks using methods for classifying traffic flows based on machine learning to ensure quality of service and quality of perception, as well as the reliability of transmission traffic messages in real time .
These tasks, such as the efficiency of the functioning of a mobile network with packet switching, are described by the following functional dependence:

$$Q_{EF}(\lambda_i) = E[\mathrm{U}_{\mathrm{QoS}}(\lambda_i), A_{am}(\lambda_{i.u}, \lambda_{i.s}), U_{QoE}(\lambda_i),\ D_{BER}(E_b, \lambda_i)]\ ,\ i = \overline{1,k}\ , \quad (1)$$

where $Q_{EF}(\lambda_i)$ - a function that takes into account the criteria for the effectiveness of the functioning multiservice telecommunication networks based on packet switching for transmitting $i-$th stream packets, taking into account the speed incoming useful and service traffic λ_i $i = \overline{1,k}$, $\lambda_i = \beta(\lambda_{i.u},\ \lambda_{i.s})$, where $\lambda_{i.u}$, $\lambda_{i.s}$ $-$intensity of the incoming flow useful and service traffic packets, respectively, $i = \overline{1,k}$;
$\mathrm{U}_{\mathrm{QoS}}(\lambda_i),\ U_{QoE}(\lambda_i)-$ a function that takes into account criteria for quality of service and quality of perception of the flow useful and service traffic packets in real time, respectively, $i = \overline{1,k}$; $A_{am}(\lambda_{i.u}, \lambda_{i.s})-$metric indicators accuracy assessment classification useful and service traffic flows based on machine learning, taking into account the intensity incoming useful $\lambda_{i.u}$ and service traffic $\lambda_{i.s}$, $i = \overline{1,k}$; $D_{BER}(E_b, \lambda_i)-$ a function that takes into account the reliability criteria for transmitting traffic messages radio engineering systems and mobile cellular networks in real time, taking into account the energy of the transmitted bit signal E_b when transmitting the $i-$th stream of packets, respectively, $i = \overline{1,k}$.
Expression (1) takes into account the new approach for the real-time feature matrix and the characteristics of the classification of traffic flows in order to ensure QoS, QoE and message transmission reliability, as well as quality of assessment metrics and accuracy indicators for different machine learning methods in radio engineering systems and mobile cellular networks using SDN, IMS and ML technologies [1, 5].
Thus, taking into account the above, the task arises - to create an effective calculation method for the dynamic classification of service and useful traffic based on the proposed model, which has the ability to detect new classes, reliability and accuracy, working in real time[5].

3. Research on methods for calculating dynamic traffic classification based on machine learning.

It is worth noting that the class traffic packet flows means a certain type of traffic, determined by its characteristics as an infocommunication application, a type of multimedia service, and enlarged categories. In this case, in the traffic model under study there is a certain known database traffic classes – space:

$$Z(\lambda_i) = \{z_1(\lambda_i), z_2(\lambda_i), ..., z_n(\lambda_i)\},\quad i = \overline{1,k}\quad , \quad (2)$$

where $n-$number of known classes of traffic under study.
Based on (1) and (2), for the space $Y(\lambda_i)$ there is a function of traffic classes, both True traffic and Predicted classes, which are described as follows [1, 3, 4]:

$$S_{tc}(\lambda_i) = f[\mathrm{B}(\lambda_{\mathrm{i}})] = Z(\lambda_i)\ ,\ S_{pc}(\lambda_i) = \varphi[\mathrm{B}(\lambda_{\mathrm{i}})] = Z(\lambda_i)\ ,\ i = \overline{1,k}\quad , \quad (3)$$

Expressions (3) characterize the true traffic class, which uniquely determine the correspondence between the feature vectors of traffic flows from space $Y(\lambda_i)$ and classes from space $Z(\lambda_i)$. However, for the machine learning mode, data on the marking of service traffic $S_{tc}(\lambda_i)$ flows is used as data.
Research shows [3, 4] that each of the space $X(\lambda_{i.s})$ and $Y(\lambda_{i.u})$ flows can be described by its own feature vector using the feature extraction function β:

$$\beta[\mathrm{X}(\lambda_{\mathrm{i.s}})] = \mathrm{A} = \{\mathrm{A}_1(\lambda_{\mathrm{i.s}}), \mathrm{A}_2(\lambda_{\mathrm{i..s}}), ..., \mathrm{A}_{\mathrm{m}}(\lambda_{\mathrm{i.s}})\}\quad ,\ i = \overline{1,k}\quad , \quad (4)$$

$$\beta[\mathrm{Y}(\lambda_{\mathrm{i.u}})] = \mathrm{B} = \{\mathrm{B}_1(\lambda_{\mathrm{i.s}}), \mathrm{B}_2(\lambda_{\mathrm{i..s}}), ..., \mathrm{B}_\mathrm{m}(\lambda_{\mathrm{i.s}})\} \quad , \; i = \overline{1, k} \; , \tag{5}$$

where $m -$ total number of features for spaces $X(\lambda_i)$ and $Y(\lambda_i)$, which use statistical characteristics traffic packet flows as features.
The last expressions (1),…,(5) take into account the characteristics, true and predicted traffic classes before the queuing system is put into operation, to assess the performance of the algorithm, part of the traffic packet flows, which in machine learning terms, this stage is called testing.
Let's consider machine learning methods using intelligent technology SDN, IMS, ML and a method for calculating dynamic traffic classification, as well as quality assessment metrics and accuracy indicators.
1. Overall Accuracy - is the proportion correctly classified threads out all and is found by the following expression:

$$T_{i.\,Acc.} = \frac{\sum_{i=1}^{n} TP_{ii}}{\sum_{i=1}^{n} TP_{ii} + \sum_{i=1}^{n}\sum_{j=1}^{n} F_{ij}} \le 1, \quad i, j = 1, 2, ..., \mathrm{n}\,. \tag{6}$$

where $TP_i -$ number of correctly classified traffic packet flows;
$F_{i\,j} -$ the number of incorrectly classified traffic packet flows, and i – predicted class number; j – true class number.
2. Reliability of transmission traffic flows - accuracy of message reception

$$D_{BER}(E_b, \lambda_i) = \sum_{i=0}^{t} C_n^{n-i} \cdot (1 - P_{sd})^{n-i} \cdot P_{sd}^{i} \quad , \tag{7}$$

where $C_n^{n-i} -$ binomial coefficient of n by $n-i$ and equals $C_n^i = \frac{n!}{i! \cdot (n-i)!}$, $n = k + r$, k, r - the number information and verification code elements messages, respectively.
3. F1-measure – the harmonic average between the completeness and accuracy of each class of traffic flows:

$$F1 = [2 \cdot \frac{T_{\mathrm{Pre.}} \cdot T_{\mathrm{Rec}}}{T_{\mathrm{Pre.}} + T_{\mathrm{Rec.}}}] \to \max \quad , \tag{8}$$

The analytical expressions under study (2),..., (8) describe a method for calculating the dynamic classification of traffic flows based on machine learning in order to maintain QoS, QoE and reliability of message transmission in radio systems and mobile cellular networks.

4. Results analysis numerical calculations and experimental research.

To evaluate the effectiveness of the classification algorithms, traffic measurements were carried out in the data center from the AzerTelecom telecom operators using the “tcpdump” program in the edge router of the communication network. The bulk of the traffic was collected during the daytime hours. TCP (Transport Control Protocol) and UDP (Uzer Datagram Protocol) were used as transport layer protocols. For numerical analysis and experimental study of the effectiveness of the feature matrix in relation to the quality of network functioning, a feature matrix is proposed, in which the features are individual packet characteristics such as formula (6). In addition to formula (6), to classify active flows, the sizes and inter-interval time each of the first 15 packets were selected - a set of characteristics, as well as generally accepted characteristics. The performance of the proposed calculation method was analyzed in the Matlab, R 2019b (9.7; 64 bit) environment using the Communications Toolbox package, designed for calculating and modeling quality assessment metrics, accuracy indicators, feature matrix, true and predicted classes traffic packet flows when using the TCP and UDP protocol stack . Based on the numerical values in Figure 1, a graphical dependence of the overall accuracy on the classification method for TCP traffic with different machine learning methods was constructed.

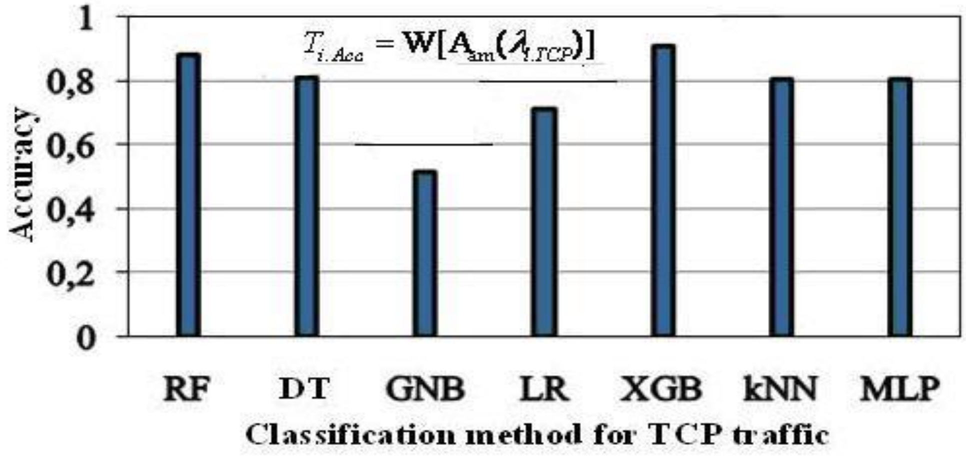

Fig.1. Graphical dependence of the overall accuracy on the classification method for TCP traffic with different machine learning methods

Analysis of the graphical dependence $T_{i.Acc.} = \mathrm{W}[\mathrm{A_{am}}(\lambda_{i.TCP})]$ shows that among the important methods for training are XGB and RF, with the help of which high overall accuracy is achieved. Overall Accuracy is the proportion of correctly classified flows out of all traffic, where XGB = 0.908 и RF = 0.879. These results improve the efficiency of traffic classification methods of machine learning. The graph shows that for the first test scenario, the boundary values obtained using the XGB and RF algorithms merge.

References

1. Krasnova, I.A., Deart, V.Yu., Mankov, V.A. Collection of Individual Packet Statistical Information in a Flow Based on P4-switch. Advances in Intelligent Systems, Computer Science and Digital Economics. CSDEIS 2019. 1127, 106-117 (2020).
2. Roslyakov, A.V. Future networks: a review of approaches to the new telecommunication Paradigm. Telecommunication. 9, 30-37 (2020).
3. Breiman L. Random Forests. Machine Learning. 45, 5-32 (2001).
4. Nguyen, T.T. and Armitage, G. A survey of techniques for internet traffic classification using machine learning. *IEEE Communication Surveys Tuts*. l, 10, 4, 56-76 (2008).
5. Belozor, A. M., Goldshtein, A. B. Analysis of the application machine learning methods for the tasks of telecom operators. Information technologies and telecommunications. 7, 3, 26 –30 (2019).