https://doi.org/10.59849/aidd.2024.41

Research Useful and Service Traffic Based on Machine Learning in Multiservice Software Defined Network

Bayram Ibrahimov[1], Farhad Valiyev[2]
[1]Department of RadioElectronics and Aerospace Systems,
Azerbaijan Technical University, Baku, Azerbaijan,
E-mail: i.bayram@mail.ru
[2]Institute of Control Systems, Baku, Azerbaijan,
E-mail: farhad.valiyev89@gmail.com

Abstract. Traffic classifications based on digital end-to-end technologies such as Software Defined Networking, Network Functions Virtualization and Machine Learning have been studied with an increase in the volume of transmitted traffic and in the provision infocommunication services. The analyzes showed that critical information and communication services require proactive monitoring of the Service Level Agreement in real time and guaranteed quality of service, which must be tested and monitored during the transmission, processing and reception traffic packet flows. This paper investigates the problems dynamic classification packet flows useful and service traffic using machine learning in a multiservice software-defined network. Methods for calculating dynamic traffic classification based on machine learning using intelligent technology are proposed and analytical expressions are used to evaluate the quality metrics of the classifier's work, such as overall accuracy, accuracy and completeness each class, and F1-measure using the Random Forest and XGBoost methods.

Keywords: F1-measure, prediction, machine learning, classical learning, accuracy, classification, protocol, SDN.

1.Introduction.

Currently, the development multiservice communication networks based on SDN technology (SDN, Software Defined Networking), subject to intensive growth in the volume transmitted useful and service traffic, requires the analysis of effective methods for dynamic classification traffic flows based on machine learning (ML, Machine Learning) to ensure quality of service in multiservice communication networks in real time [1, 2].

It should be noted that recently software-defined SDN networks have become increasingly widespread in multiservice communication networks, the architecture of which involves separating the control plane from the data transmission plane and forming a unit for a single centralized network management.

This concept allows you to obtain certain advantages in the organization and management useful and service flows, as well as achieve greater flexibility in accessing resources networks and devices, and more dynamic development of the existing network infrastructure, in contrast to the traditional approach to building multiservice telecommunication networks [2].

However, studies have shown [2, 3, 4] that multiservice SDN communication networks offer the user a wide range different multimedia services, including voice communications, video communications, video conferencing, and data transmission. These features make it easy to add new services and applications and change existing ones.

In a multi-service communication network using SDN technology, each of the created applications requires ensuring a certain level QoS (Quality of Service) and QoE (Quality of Experience), which becomes more and more difficult with the increase in the number consumers and multimedia services.

The work [1, 3] shows that new approaches to classifying flows in traditional communication networks were based on the well-known list of TCP and UDP ports, but with the advent dynamically changing ports, the use of such a method became impossible.

At the same time, the well-known DPI (Deep Packet Inspection) technology, which allows for "deep" analysis of packet headers at the upper levels of the OSI model. But with the help of the DPI system, it is also not always possible to identify the nature of the data flow, for example, in cases encrypted or tunneled traffic [1, 2, 3, 4].

Recently, in multiservice telecommunication networks, data mining methods, especially machine learning methods, are increasingly being effectively used to solve a wide range problems, including classifying useful and service traffic.

Taking into account the above, this paper examines the tasks research and analysis methods for classifying flows useful and service traffic packets using machine learning in multi-service SDN communication networks based on machine learning technology to ensure quality of service.

2. General statement of the problem.

For sources heterogeneous traffic known to the communication network operator, automatic or preliminary static marking of packets is possible, indicating the affiliation of the service, which is used for differentiated traffic management, requiring solutions using complex data processing algorithms, in particular, forecasting, classification and testing tasks.

In this case, a special place is occupied by intelligent technologies - machine learning, which is becoming increasingly in demand for application, but due to the diversity its algorithms and the impossibility of their universal use, there is a need to determine the scope application for each them.

One of the trends that arises with the current complication of a multiservice software-defined network and the use of big data are trends in artificial intelligence, machine learning, and neural networks. At this stage of the research, it was decided to consider the possibilities of machine learning traffic flows in a multiservice SDN network.

Machine learning is a technique for analyzing traffic flows in a multiservice SDN network that allows an analytical system to learn while solving many similar problems, extract information from source data, identify patterns and make decisions with minimal human intervention.

It is the minimal human involvement and the ability to efficiently process data that makes machine learning methods so relevant with the exponential growth of data volumes in our time.

In [1, 3, 4], classical learning was analyzed, in turn, divided into supervised and unsupervised learning. In the first case, the data streams of traffic packets must be pre-labeled in some way, and in the second, they may not be categorized in any way.

Three methods were chosen for testing: regression - supervised learning, clustering and association - unsupervised.

Currently, due to the increase in the volume processed traffic in communication networks with the introduction SDN technology, it requires classification transmitted traffic in order to ensure QoS and QoE, through an ML machine learning mechanism using intelligent technology.

To create a feature matrix for real-time traffic classification in order to maintain QoS and QoE, new approaches are proposed, which are based on calculating statistical characteristics in a multiservice SDN network.

To create a feature matrix for real-time traffic classification in order to maintain QoS and QoE, new approaches are proposed, which are based on the calculation of statistical characteristics.

To create a matrix of features for useful and service traffic, we used parameters such as average packet size $E[\mathrm{L}_{\mathrm{P}}(\lambda_{\mathrm{i}})]$, average arrival time between packets $E[\tau_{ji}(\lambda_i)]$, average of the entire flow $E[N_p(\lambda_i)]$ and packet speed $V_p(\lambda_i)$, which are described by the following functional dependence:

$$\beta[\mathrm{X}(\lambda_{\mathrm{i}}),\mathrm{Y}(\lambda_{\mathrm{i}})]=\mathrm{W}\{\mathrm{E}[\mathrm{L}_{\mathrm{p}}(\lambda_i), E[\tau_{ji}(\lambda_i)], V_p(\lambda_i), E[N_p(\lambda_i)]\}\ ,\ \ i=\overline{1,k}\ . \quad (1)$$

Expression (2) is a feature matrix for classifying active flows in real time, which requires a feature matrix available over a small number of initial packets.

3. Description of the method calculation and analysis communication network indicators based on ML.

To accurately classify useful and service traffic based on intelligent ML technology, an analysis of their statistical characteristics QoS and QoE, packet parameters and heavy-weight monitoring algorithms in dynamically changing networks using SDN is required.

Consider the problem of traffic classification can be formulated as follows.

Let there be a system for servicing useful and service traffic, the input of which can receive two types of packets [3, 4]:

1. $$Y(\lambda_{i.\mathrm{u}}) = \{y(\lambda_1), y(\lambda_2),..., y(\lambda_i)\},\ t_{1i} < t_{2i} <,..., < t_{n.i}\ ,\ L_{ji} \in [\mathrm{L}_{\mathrm{i.min}}, \mathrm{L}_{\mathrm{i.max}}],\ i = \overline{1,k}, \quad (2)$$

where $\mathrm{L}_{\mathrm{i.min}}, \mathrm{L}_{\mathrm{i.max}}$ – the minimum and maximum length of the payload and service load packet, depending on the settings of the SDN network and MTU (Maximum Transmission Unit), the maximum transmission unit and equal to $L_{ji} \le (40,...,279)$ bute .

2. $$X(\lambda_{i.\mathrm{s}}) = \{x(\lambda_1), x(\lambda_2),..., x(\lambda_i)\},\ t_{1i} < t_{2i} <,..., < t_{n.i},\ L_{ji} \in [\mathrm{L}_{\mathrm{i.min}}, \mathrm{L}_{\mathrm{i.max}}],\ i = \overline{1,k}, \quad (3)$$

where $t_{n.i}$ – time of arrival of the i – th packet flow at the switch interface.

We assume that, based on (2) and (3), the packet flow $X(\lambda_{i.u})$ and $Y(\lambda_{i.u})$ is ordinary and represents sequential, unidirectionally arriving packets.

In this case, the maximum inter-interval time between two successively arriving packets is determined from the dynamics of the SDN network as follows:

3. $$E[\tau_{ji}(\lambda_i)] = t_{(j+1)i}(\lambda_i) - t_{ji}(\lambda_i)\quad ,\ i = \overline{1,k}\quad . \quad (4)$$

4. Based on machine learning methods using intelligent technology and a method for calculating dynamic traffic classification, we will consider metrics for assessing quality indicators - the completeness of each class, which is expressed as follows:

$$T_{j.\mathrm{Re}c.} = \frac{TP_j}{TP_j + \sum_{i=1}^{n} F_{i\,j}} \le 1\ ,\ i, j = 1, 2,..., \mathrm{n}. \quad (5)$$

where TP_i – number 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.

Expressions (5) characterize the completeness each Recall class - the proportion correctly predicted traffic packet flows out all those belonging to this class in the multiservice SDN network.

5. The accuracy each Precision - class is the proportion correctly identified threads of a class out of all threads that were assigned to this class:

$$T_{i.\mathrm{Pr}e.} = \frac{TP_{ii}}{TP_{ii} + \sum_{j=1}^{n} F_{i\,j}} \le 1\ ,\quad i, j = 1, 2,..., \mathrm{n}. \quad (6)$$

The resulting analytical expressions (1),..., (6) describe the method for calculating the dynamic classification traffic flows based on machine learning in order to maintain QoS and QoE in a in a multiservice SDN network in real time and characterize the quality assessment metrics, accuracy indicators, feature matrix, true and predicted traffic packet flow classes.

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.

Figure 1 shows a graphical dependence of the overall accuracy on the classification method for UDP traffic with different machine learning methods.

Analysis of the graphical dependence $T_{i.Acc.} = \mathrm{W}[\mathrm{A}_{\mathrm{am}}(\lambda_{i.UDP})]$ shows that among the important methods for training are XGB and RF, with the help of which high overall accuracy is achieved.

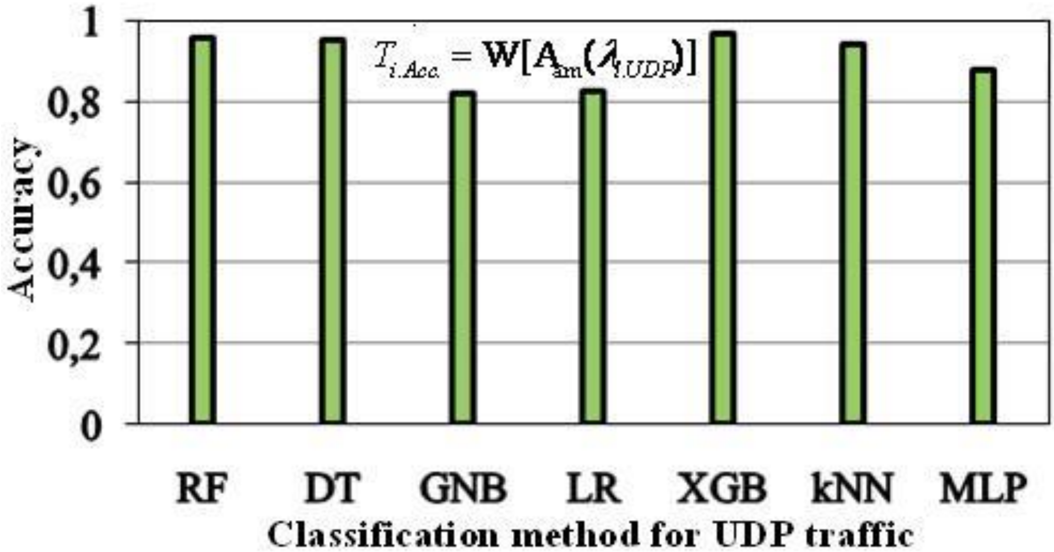

Figure 1. Graphical dependence of the overall accuracy on the classification method for UDP traffic with different machine learning methods

This graphical dependence also uses the XGBoost and Random Forest traffic classification algorithms.

Overall Accuracy is the proportion correctly classified flows out of all UDP traffic. In this case, critical and important values are obtained only at XGB = 0.964 and RF = 0.956, which meet the requirement of the efficiency metric for traffic classification by methods based on machine learning.

References

1. Patterson, J., Gibson, A. Deep learning from a practitioner's perspective. M.: DMK Press, 418 (2018).
2. Ibrahimov, B.G., Alieva, A.A. Research and Analysis Indicators of the Quality of Service Multimedia Traffic Using Fuzzy Logic. 14th International Conference on Theory and Application of Fuzzy Systems and Soft Computing – ICAFS-2020. 1306, 773-780 (2021).
3. Belozor, A. M., Goldshtein, A. B. Analysis of the application of machine learning methods for the tasks of telecom operators. Information technologies and telecommunications. 7, 3. 26–30 (2019).
4. Brink, H., Richards, J., Feverolf M. Machine learning. St. Petersburg: Peter. 336, (2017).