ISSN 2347 - 3983

Volume 8. No. 6, June 2020

International Journal of Emerging Trends in Engineering Research

Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter87862020.pdf

https://doi.org/10.30534/ijeter/2020/87862020

# Analysis of Methods for Measuring Available Bandwidth and Classification of Network Traffic

**Gulomov Sherzod Rajaboevich[1],Xoshimova Charos Saidaminovna[2], Ganiyeva Toxira Irkinovna[3], Djurayeva Shoxista Tagirovna[4]**

[1]Tashkent University of Information Technologies named after Muhammad al-Khwarizmi, Uzbekistan, sherhisor30@gmail.com

[2]Tashkent State Technical University named after Islam Karimov, Uzbekistan, charos.khoshimova@mail.ru

[3]Tashkent State Technical University named after Islam Karimov, Uzbekistan, tganiyeva@mail.ru

[4]Tashkent State Technical University named after Islam Karimov, Uzbekistan, 1965d@gmail.com

## ABSTRACT

This paper explores methods for measuring the available bandwidth of network traffic to assess the quality of routing in info communication networks as well presents the advantages and disadvantages of one-way and two-way network delays. A comparative analysis of the effectiveness of network traffic classification methods are carried out: characteristics used to solve it, existing approaches and areas of their applicability.

**Key words:** Round Trip Time,One Way Delay, Jitter, Jitter, DPI, machine learning.

## 1. INTRODUCTION

The active development of information technology has made them an integral part of life, production, and the service sector. Information systems are currently operated in both commercial and government organizations. The interaction between such systems is carried out through a global network. There is a rapid increase in network traffic, its structure is becoming more complicated. Traffic analysis is becoming more and more popular in the areas of control and management, optimization, and protection from harmful influences.Threats caused by attacks on distributed info communication systems require effective methods for their identification and response. The biggest problem is caused by attacks that have anomalous behavior in the characteristics of the selected packet of network traffic attributes. One of the directions of the development of traffic filtering systems is the classification of network traffic, which allows monitoring the quality of service and effectively managing channel bandwidth.

## 2. METHODS FOR MEASURING AVAILABLE NETWORK BANDWIDTH

General approaches in evaluating network connection performance; separate standards have been proposed that describe the following metrics:

- two-way network delay (Round Trip Time, RTT);
- one-way network delay (One Way Delay, OWD);
- variation of packet delay (Jitter);
- packet loss (Packet loss);
- the values of the full bandwidth (FB);
- the value of the available bandwidth (AB).

### 2.1.Two-ways network delay

Two-ways network delay is defined as the time it takes to transmit a packet between network nodes plus the time it takes to receive confirmation of packet delivery by the remote node. In other words, this is the time interval between sending the first bit of the packet from the source to the receiver and receiving the last bit of the response packet from the receiver to the source.

### 2.2. One-way network delay

One-way network delay is located as the transmission time of a packet of a certain type between two network nodes. A specific type is understood to mean a packet that has a set of predefined features; the standard does not rigidly stipulate these signs, but indicates that they can be, for example, the size of the packet, the type of application that generated the packet, the type of protocol of the transport layer that delivered the packet and some others. The meaning of the used set of features is to distinguish from the general packet stream arriving at the destination node those packets whose characteristics are of interest to the specialist conducting the measurements [1-2-3].Table 1 summarizes the advantages and disadvantages of one-way and two-ways network delays.

**Table 1:** Advantages and disadvantages of one-way and two-way network delays

| Metrics | Advantages | Disadvantages |
|---|---|---|
| One way network delay | It can be used to evaluate the performance of asymmetric channels, provides the user with reliable information about the efficiency of the network route | 1. The performance of a network application may depend on the efficiency of the communication channel in only one direction, which cannot be estimated when measuring two-way network delay. |

| | | |
|---|---|---|
| | in both directions. | 2. More difficult to measure, this requires synchronization of time samples for the data source and receiver with an accuracy of at least 1 ms. |
| Two ways network delay | 1. Simple of implementation of the measurement process - not only the installation of measuring equipment is required, but also special software on the receiver side of the network packets.<br>2.Simple interpretation: often of practical interest. | 1. Network data transmission channels in many cases are asymmetric, and the amount of two-way delay becomes an uninformative metric to assess network performance.<br>2. Even if the network channel is symmetric, the performance in the forward and reverse directions can dramatically vary due to the use of asymmetric mass service mechanisms in network routers along the IP packet path.<br>3. Network application performance can depend significantly on data transfer in only one of two directions.<br>in networks using QoS (Quality-of-Service) mechanisms, the allocation of resources in one of the directions can significantly differ from the allocation of resources in the opposite direction, which also makes two-way network delay an uninformative metric. |

### 2.3. Packet delay variation

The packet delay variation is defined as the difference between the values of the one-way network delays in two consecutive measurements. According to him, a network packet is considered received if the transmitter sent the first bit of the packet at time $\boldsymbol{T}$, and the receiver received this network packet [4]. In this case, the packet loss metric takes the value of a logical unit. If the packet was not received by the receiver during the packet lifetime (according to the IP protocol standard, this value does not exceed the theoretical maximum of 255 seconds), then the metric is assumed to be logical zero.

### 2.4. Packet loss

Packet loss necessarily happens from time to time. Due to constant use and high demand, packets are confused or lost along the way, and these are some of the most common reasons.

*Crowded networks*

Networks that achieve maximum throughput are called congested networks and are more likely to experience packet loss due to increased traffic. Since the packet transfer process follows certain steps, connection failures can lead to the loss of some packets so that the network can handle the incoming load [5-6]. However, as modern technologies evolve, many applications and programs are now able to process rejected data using another method, which involves slowing down the transfer rate or automatically forwarding lost data packets.

*Mistakes*

Software errors are also another cause of packet loss on the network. Applications that are accessed without proper software testing are likely to cause network problems and, in turn, will affect packet transmission. Software reboots often solve this problem, however, a software update or a full application fix may be required.

*Network hardware and software issues*

There are several possible hardware or software problems that can significantly affect incoming traffic to the network. When legacy hardware devices are used to start the system, packets may be lost due to slow data transfer. Companies and individuals are encouraged to constantly upgrade or upgrade their hardware to optimize network process performance. This is necessary to avoid network delays, packet loss, or even a complete loss of connection to the system.

*Threats and attacks*

Security leaks and network threats can also cause packet loss. Recently, cyber-attacks, known as packet drop attacks, have become popular with cybercriminals. Some people send commands that send data packets to the data stream. These malicious users can do this by gaining access to a network router. These types of attacks can be identified by monitoring the packet loss rate on the network. A sudden jump in these statistics can be a sign of an online attack.

### 2.5. The values of the full bandwidth (FB) and available bandwidth (AB)

The FB and AB of the network channel are not precisely defined in the RFC standards, which is why several different print sources may have slightly different definitions. FB - the maximum bandwidth of the network IP channel, which can be available on the route in the absence of a competing stream. AB - the maximum bandwidth of the network IP channel, which can be provided to the stream in a specific situation of congestion of the route by competing traffic.

And so, the four most used methods for measuring the bandwidth of a network channel are highlighted:

1. The method of variable packet size.
2. The method of sequence pairs of packets.
3. The method of periodic flow.

The variable packet size method is used to measure the full throughput of an end-to-end connection [7-8]. The packet pair sequence method and the periodic flow method are designed to measure the available channel capacity.

*The method of variable packet size*

The main experimental dependence that was proposed to be investigated is the two-way network delay function, where the size of the test packet is used as an argument. The router discards the expired packets, passing the error message back to the sender, using the standard ICMP protocol for internetwork control messages. The collected ICMP packets are processed and the two-way network delay values for the network connection are calculated depending on the size of the packet.

The following network delay components can be distinguished:

- processing delay $\boldsymbol{D_p}$ – the time required to create a packet with data, prepare it for transmission, as well as process the packet at the destination to retrieve data;
- transmission delay $\boldsymbol{D_t}$–packet transmission time from the first to the last bit in the telecommunication channel. This value depends on the width of the communication channel;
- queuing delay $\boldsymbol{D_q}$ – time spent by the packet on waiting and processing on network devices. It is described by queuing theory.

The network delay is determined by the $\boldsymbol{W/C}$ratio, where the total bandwidth is $\boldsymbol{C}$(byte/s), the network packet size $\boldsymbol{W}$is measured in bytes.

For using the method of varying the packet size, the sender transfers a chain of measuring packets of constant size to all intermediate network devices of the channel under study in the global network.

To apply this measurement method, it is necessary to take into account the accepted simplification that the delay in the queue $\boldsymbol{D_q}$, that is, the time spent by the packet on waiting and processing on network devices, is assumed to be zero. In this case, the magnitude of the one-way network delay is determined by two components: the processing delay $\boldsymbol{D_p}$ and the transmission delay $\boldsymbol{D_t}$.

*The method of sequence pairs of packets*

The implementation of the packet pair method is currently possible using two similar methods.

In the first implementation, a set of packet pairs is transmitted over a network channel and the variation in the time interval between packets of each pair is analyzed [9]. Under the variation of the packet pair for the network route under study, we consider the difference in the time intervals between the reception of the last bits of each of the two packets for the input and output streams. Figure 1 shows a graphical representation of the variation of a pair of packets in a channel as a given pair travels along a route with full bandwidth.

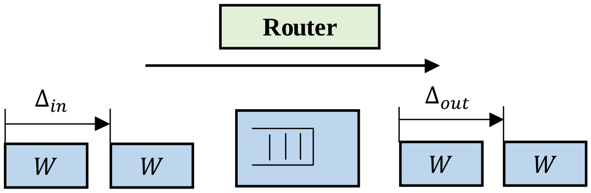

**Figure 1:** Variation of a pair of packets

Router with capacity equal to $C_i$. The method is applied on the assumption that there are no competing traffic flows on this network section, as a result, we will measure the available bandwidth. Assuming that the total throughput of the first communication channel on this route is equal to $C_0$, and the size of the test packets is constant and equal to $W$, the variation of the pair of packets for the investigated communication channel will be proportional to the total throughput: $\Delta_0 = W/C_0$. This is true for any route if the variation of the previous section of the route is $\Delta_{ds}$and the total throughput of this channel is $C_i$. Then the variation at the output of this section of the route will be determined as:

$$\Delta_{out} = \max(\Delta_{in})\frac{W}{C_i}.$$

After a pair of packets passes through the entire network route, the variation at the time of arrival of packets to the recipient ($\Delta_{field}$) can be measured:

$$\Delta_{field} = max_{i=0,\dots,k}\left(\frac{W}{C_i}\right) = \frac{W}{min_{i=0,\dots,k}(C_i)} = \frac{W}{C},$$

where $C$ – is the total bandwidth for the entire network route. As a result, the recipient can measure the total throughput of the network route as $C = W/\Delta_{field}$

In modern public networks, the situation when there are no competing traffic flows throughout the network route is too simplified. The presence of third-party traffic in the network channel only increases the value of the $\Delta_{field}$variation, which entails the erroneous calculation of the available bandwidth on the route under study. This is due to the ever-changing value of the $D_q$ component for successive packets [10]. The use of mathematical statistics algorithms and the re-sending of additional packet pairs in order to discard misses in measuring the full throughput reduces the influence of external traffic in the network channel and increases the accuracy of the measurements.

The second implementation method uses sequences of equal-sized packets at equal intervals between packets. In this case, the variation of the sequence of packets is calculated as the difference between the timestamps of the first and last packet of the sequence. After the recipient finds the variation value $\Delta_{field}(M)$for the sequence consisting of $M$ packets, the frequency of variation is calculated:

$$R = \frac{(M-1)W}{\Delta_{field}(M)}$$

In this method, the accuracy of calculating the full throughput is much less affected by third-party traffic in the network channel.

*The method of periodic flow*

The following method is designed only to measure the available bandwidth of a network channel and cannot be used to measure the full bandwidth [11]. This method involves measuring the variation in the magnitude of the one-way delays of test packets. If the bit stream speed exceeds the available channel capacity, then the $D_q$ latency in the bottleneck of the route will increase. In a situation where the bit rate does not exceed the available channel capacity, the packet chain will not experience additional time delays.

## 3. TRAFFIC CLASSIFICATION METHODS

With the growing number of approaches, a need arose for their classification. One of the options for the analysis of the classification of methods is shown in Figure 2.

### 3.1. Traffic classification based on payload

DPI systems are primarily designed to identify applications involved in network interactions. Therefore, the "in-depth" analysis involves the analysis of the contents of network packets at all levels. Each network packet consists of control information and payload. Here, the term “packet” is used as a universal term for generalizing such concepts as a frame, datagram, and a segment of the corresponding network protocols. In the process of parsing the packet, the protocol headers are highlighted, the field values in them are analyzed.

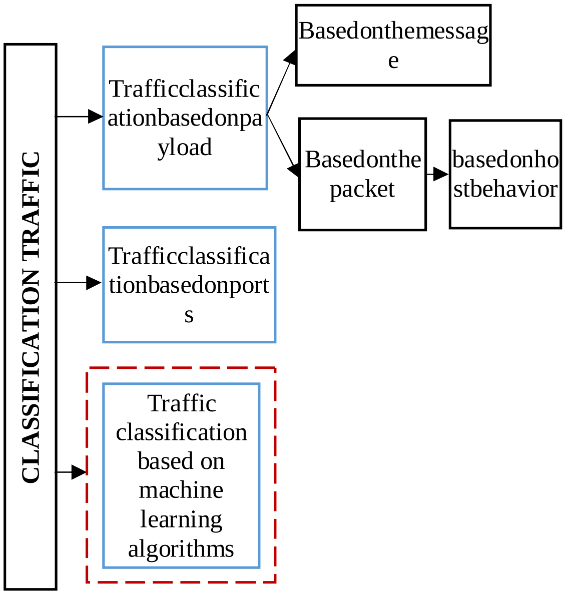
CLASSIFICATION TRAFFIC
Trafficclassificationbasedonpayload
Basedonthemessage
Basedonthepacket
basedonhostbehavior
Trafficclassificationbasedonports
Traffic classification based on machine learning algorithms

**Figure 2:** Traffic Classification Methods

The structure of the header is determined by the specification, while the payload may contain randomly organized data, although it is usually a protocol packet of the next higher level: to continue the analysis, it is necessary to determine which protocol it is.Figure 3 shows the allocation and analysis of protocol headers in a packet.

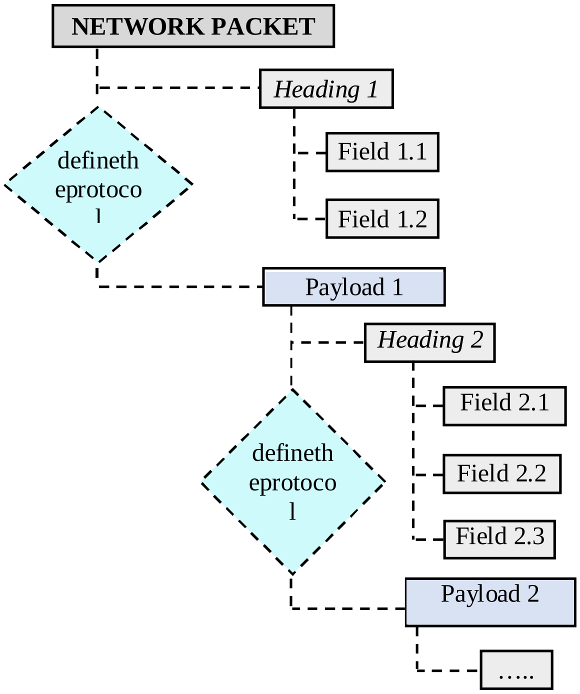

**Figure 3:** Highlighting and parsing protocol headers in a packet

In accordance with the OSI model, the headers of the network protocols of the packet form a stack and, as a rule, follow each other in a natural order - from low to high. However, when organizing tunnel connections, this order may be violated - for example, when transmitting IPv4 packets (network layer) within UDP protocol packets (transport layer). Tunneling protocols are now widespread: in particular, they are used in organizing virtual private networks. In the general case, a tunnel of arbitrary configuration is possible: in particular, one tunnel can be nested in another [12]. Tunnel traffic parsing must be supported by a network analyzer. Figure 4 shows the taxonomy of payload-based approaches depending on processing methods.

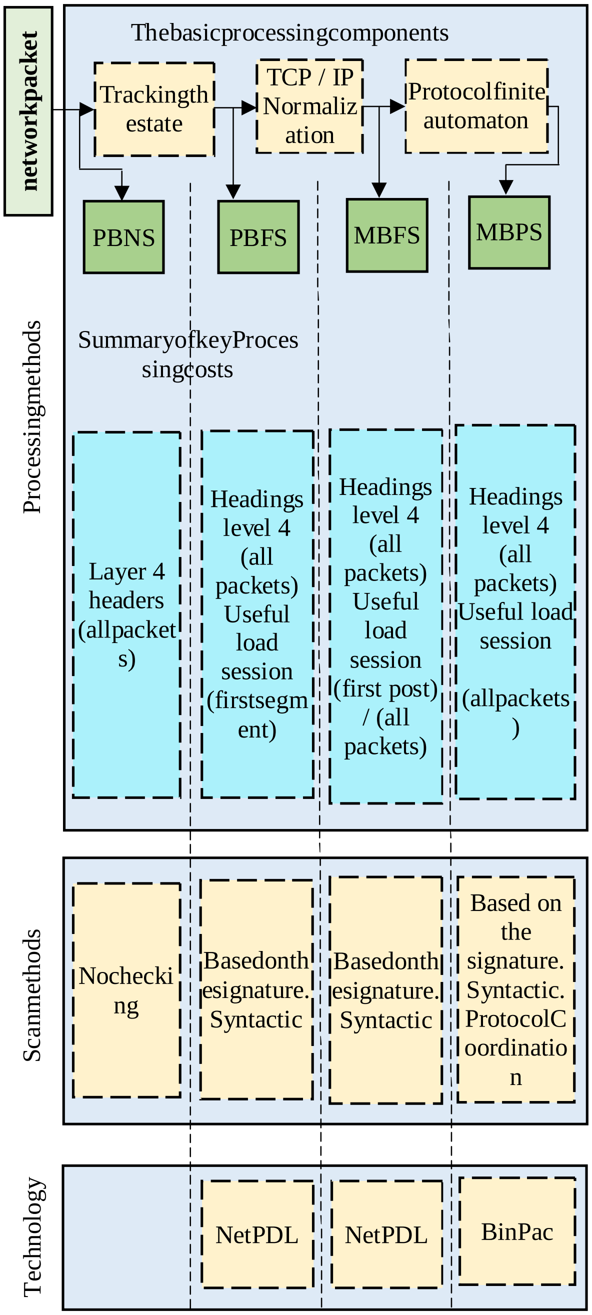
networkpacket
Thebasicprocessingcomponents
Trackingthestate
TCP / IP Normalization
Protocolfinite automaton
PBNS
PBFS
MBFS
MBPS
Processingmethods
SummaryofkeyProcessingcosts
Layer 4 headers (allpackets)
Headings level 4 (all packets) Useful load session (firstsegment)
Headings level 4 (all packets) Useful load session (first post) / (all packets)
Headings level 4 (all packets) Useful load session (allpackets)
Scanmethods
Nochecking
Basedonthesignature. Syntactic
Basedonthesignature. Syntactic
Based on the signature. Syntactic. ProtocolCoordination
Technology
NetPDL
NetPDL
BinPac

**Figure 4:** Traffic classification based on payload

The increasing complexity of these methods can be considered through an increase in the degree of processing from left to right, as well as in accordance with memory requirements.

*Four different levels of verification stand out*

*The first level* of verification is based on the signature; its purpose is to search for some signatures within the application layer payload. For example, an HTTP packet starts with a command following the URL and protocol version, while most Edonkey packets have fields containing the payload size. The signature-based method is based on matching the payload with the signature defined for this protocol. Signatures are usually regular expressions.

*The second level* of verification is syntactic. It can be considered as a more accurate version of signature verification, since it is aimed at verifying the correctness of the transmitted data from a syntactical point of view. In this case, it is necessary to decode all the fields contained in the message and ensure that the message is well formed.

*The third level*of control is related to the compliance protocol. It controls that the HTTP GET request from the client is really followed by a response from the server. This form of control is more accurate because it can verify the protocol’s actual behavior in accordance with the specification.

*The fourth level* of control relates to data semantics. It is able to verify whether an object transmitted over HTTP is an image or some other form of content. This control is very useful for detecting “tunnels” in which the application uses a different protocol to transport data.

*Various processing methods*

*The simplest method* is PBNS, which works by checking the values of some fields present in each packet. This method is very simple from the point of view of calculations (only packet headers up to L4 should be processed), it does not need to store states.

*The second method* - PBFS requires an implementation of a session table in which each record includes a session identifier and the corresponding application layer protocol. Each table occupies several tens of bytes.

*The third method*MBFS is message-based. This method requires a module for normalizing TCP/IP packets. MBFS-based technologies can perform the same checks as PBFS, but work on messages, therefore, their controls can be extended to the entire message in place of the first data segment [13-14]. In this case, the required memory sizes increase due to additional state information that must be stored for each session. All these parameters strongly depend on the nature of the traffic, that is, on the number of fragmented packets and “abnormal” TCP sessions. Depending on the implementation, some products may parse all messages.

*The fourth method* MBPS accurately interprets what each application transmits and receives. The MBPS handler understands not only the semantic part of the message, but also the various stages of messaging (HTTP GET, it must be accompanied), since this method fully understands the state machine of the protocol. The required memory volumes become even larger, because it is necessary to take into account not only the state of the transport session, but also the state of each application level session.
Performance is the highest among all methods - all application-level data must be processed to verify protocol compliance. PBFS-based implementations typically associate some additional state with each session in order to perform a more accurate classification.
Traffic classification based on host behavior
At the same time, the contents of the packets are not analyzed and, to classify network traffic, host behavior patterns are mapped to one or more applications.

### 3.2. Traffic classification based on ports

Historically, many applications use “well-known” ports on their local hosts. In this case, the classifier’s task is to search for TCP SYN packets in order to determine the server side of the new client-server TCP connection. Then, to make a conclusion about the application, the target port number of the packet is viewed in the list of registered IANA ports. UDP uses ports in a similar way, but without establishing a connection.

### 3.3. Traffic classification based on machine learning algorithms

Recently, the idea of classifying traffic using machine learning methods has been actively developed. The essence of this approach is to highlight certain attributes of network packets and form a training sample based on them, which is fed to the input by special algorithms for training. Trained algorithms will be able to determine the type of application with some accuracy without using port analysis or searching for labels in the payload [15-16]. Therefore, the undoubted advantage of this method can be considered the ability to classify encrypted traffic, which can be useful for detecting malicious applications that transmit encrypted network packets, as well as other network applications that violate the organization’s information security policy.
Taking into account the above methods, the parameters of the concrete implementation of the classification system are introduced, and the methods are evaluated for classification by these parameters.
*Accuracy* - is a general characteristic that reflects the share of correctly identified traffic from the total amount of analyzed traffic. The accuracy of the results is determined mainly by how well the features are selected by which the classification is carried out and the quality of the heuristic used.

*Reaction time* - the time from the moment of receiving the first packet of a network stream until its classification. It is critical for systems operating on the stream, in particular security and traffic management systems. This concept also includes the overall performance of the algorithm.

*Reliability* - reflects the area of applicability of the systemand resistance to effects arising during the transmission process, such as packet loss, asymmetry, etc.

Table2 shows the estimates of popular methods for classifying network traffic.

**Table2.** Estimates of popular network traffic classification methods

| № | Nameofmethods / Parameters | Accuracy | Reactiontime | Reliability | Advantage | Disadvantage |
|---|---|---|---|---|---|---|
| 1. | Traffic classification based on payload (DPI) | High | High | Average | - the full correspondence of the message structure to a certain format is checked;<br>- each well-known application contains certain signatures in the transmitted data, which make it possible to distinguish one application from another. | - the complexity of developing a complete message parser compared to relatively simple signatures;<br>- lower speed, which depends on the parsing algorithms used;<br>- the study of the payload of network packets violates user privacy, as the contents of the packets may contain confidential information, the use of which by third parties is prohibited by national laws;<br>- not applicable in case of encrypted traffic. |
| 2. | Traffic classification based on ports | Low | Average | Low | - simplicity, low computational cost. | - some applications may not have their own ports registered in IANA, such as Napster and Kazaa;<br>- in some cases, IP layer encryption can confuse TCP and UDP headers, making it impossible to determine the actual port number. |
| 3. | Traffic classification based on machine learning algorithms | High | High | High | - determining the type of application without using port analysis or searching for labels among the payload;<br>- definition of signs of traffic; useful for detecting malicious applications transmitting encrypted network packets. | - high computing costs and instability when traffic changes;<br>- the classification results depend on the features used, but there is no theory of choosing the optimal features. |

## 4. CONCLUSION

It should be noted that the methods for measuring available network bandwidth have limited applicability due to the asymmetric nature of modern network channels in the global network also have limited accuracy. It allows to find the basic performance metrics of IP networks with micro second accuracy. As well the popular network traffic classification methods have been researched and after receiving an analysis and exploration of the outcomes “the traffic classification based on machine learning algorithms” has been selected one the accurate and reliable method among them.

## REFERENCES

1. Demichelis C. **IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)**, RFC 3393// 25.06.2015.
2. Almes G., Kalidindi S., Zekauskas M., Morton A. **A one-way delay metric for IPPM**, STD 81, RFC 7679//20.06.2016. https://doi.org/10.1007/978-3-319-19608-4_12
3. Baranyi P., Csapo A., Sallai G. **Cognitive Capabilities in the Future Internet**//Cognitive Info communications– Springer International Publishing, 2015. – C. 173-185.
4. Mahdavi J., Paxson V. **IPPM metrics for measuring connectivity**, RFC 2678// 20.06.2015.
5. Vakili A., Gregoire J. C. **Accurate one-way delay estimation: Limitations and improvements** //IEEE Transactions on Instrumentation and Measurement. – 2012. – T. 61. – №. 9. – C. 2428-2435
6. HegazyZaher, H. A. Khalifa, Abeer Ahmed. **Fuzzy Max Plus Algebra Algorithm for Traffic Problems**. International Journal of Emerging Trends in Engineering Research. Volume 7, No. 11 November 2019. 530-535 PP. https://doi.org/10.30534/ijeter/2019/217112019
7. Jain M., Dovrolis C. **End-to-end available bandwidth: Measurement methodology, dynamics, and relation with TCP throughput**. – ACM, 2002. – T. 32. – №. 4. – C. 295-308.
8. Fettweis G. P. **The tactile Internet: Applications and challenges** //IEEE Vehicular Technology Magazine. – 2014. – T. 9. – №. 1. – C. 64-70.
9. Abdou A.R., Matrawy A., Van Oorschot P.C. **Accurate One-Way Delay Estimation with Reduced Client Trustworthiness** //IEEE Communications Letters. – 2015. – T. 19. – №. 5. – C.735-738.
10. Jain M., Dovrolis K. **End-to-end Estimation of the Available Bandwidth Variation Range** // SIGMETRICS05, Ban_Alberta,Canada. – 2005.
11. Kim J. C., Lee Y. **An end-to-end measurement and monitoring technique for the bottleneck link capacity and its available bandwidth** //Computer Networks. – 2014. – T. 58. – C. 158-179. https://doi.org/10.1016/j.comnet.2013.08.028
12. Imai M., Sugizaki Y., Asatani K. **A new estimation method using RTT for available bandwidth of a bottleneck link** //Information Networking (ICOIN), 2013 International Conference on. – IEEE, 2013. – C.529-534.
13. Hisamatsu H., Oda H. **Design, implementation and evaluation of ICMP-based available network bandwidth measurement based on IMTCP**//International Journal of Computer Networks & Communications. – 2014. – T. 6. – №. 3. – C. 1.
14. Sukhov A.M. **The distribution function of package delay in the global network for the control theory problems** //Telecommunications and Radio Engineering. – 2013. – T. 72. – №. 3.
15. Pavithra G, Abirami P, Bhuvaneshwari S, Dharani S, Haridharani B. **A Survey on Intrusion Detection Mechanism using Machine Learning Algorithms**. International Journal of Emerging Trends in Engineering Research. Volume 8. No. 4, April 2020. 945-949 PP. https://doi.org/10.30534/ijeter/2020/01842020
16. Bachman, P. and Precup, D. **Variational generative stochastic networks with collaborative shaping**. In Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6–11 July 2015, pages 1964–1972.