NETFLOW-BASED TRAFFIC CLASSIFICATION APPROACHES IN ESG NETWORKS

Maruf Juraev
Tashkent Institute of Technology, Management and Communication
E-mail: juraevmaruff@gmail.com
Dilmurod Normurodov
The University of World Economy and Diplomacy,
Shohista Akhmadalieva
Tashkent Institute of Technology, Management and Communication
Shakhriyor Khudoyberdiev
Tashkent Institute of Technology, Management and Communication

Abstract. ESG (Environmental, Social, and Governance) network traffic classification is a very important mathematical and statistical tool in communications and computer networks, which is used to determine the average and statistical data of the traffic passing through a particular line or node. The correct use of traffic analysis methods results for ESG network in valuable information such as: how busy the link is, the average point-to-point delay, and the average packet size. Such important information helps engineers design reliable networks, prevent potential congestion, and predict future growth. This article is designed to provide an overview of traffic classification techniques and to consider them in packet-based, flow-based, and application-based contexts.

Keywords: Mice flow, elephant flow, netflow, network traffic, classification, packet

Аннотация. Классификация сетевого трафика по критериям ESG (Environmental, Social, and Governance) является очень важным математическим и статистическим инструментом в сетях связи и компьютерных сетях, используемым для определения средних и статистических данных трафика, проходящего через конкретную линию или узел. Правильное использование методов анализа трафика позволяет получить для сети ESG ценную информацию, такую как: загруженность канала, средняя задержка между точками и средний размер пакета. Такая важная информация помогает инженерам проектировать надежные сети, предотвращать потенциальные перегрузки и прогнозировать будущий рост. Данная статья призвана дать обзор методов классификации трафика и рассмотреть их в контексте пакетной, потоковой и прикладной классификации.

Ключевые слова: поток мышей, поток слонов, сетевой поток, сетевой трафик, классификация, пакет

ISSN 3030-3907

Annotatsiya. ESG (Environmental, Social, and Governance) tarmoq trafik tasnifi aloqa va kompyuter tarmoqlarida juda muhim matematik va statistik vosita bo‘lib, u ma’lum bir liniya yoki tugun orqali o‘tadigan trafikning o‘rtacha va statistik ma'lumotlarini aniqlash uchun ishlatiladi. Trafikni tahlil qilish usullaridan to‘g‘ri foydalanish ESG tarmog‘i uchun quyidagilar kabi qimmatli ma’lumotlarga olib keladi: aloqa qanchalik band, o‘rtacha nuqtadan nuqtaga kechikish va o‘rtacha paket hajmi. Bunday muhim ma’lumotlar muhandislarga ishonchli tarmoqlarni loyihalashda, potentsial tirbandlikning oldini olishda va kelajakdagi o‘sishni bashorat qilishda yordam beradi. Ushbu maqola trafikni tasniflash usullariga umumiy nuqtai nazar berish va ularni paketga asoslangan, oqimga asoslangan va dasturga asoslangan kontekstlarda ko‘rib chiqish uchun mo‘ljallangan.

Kalit so‘zlar: Sichqon oqimi, fil oqimi, tarmoq oqimi, tarmoq trafigi, tasniflash, paket

INTRODUCTION.

Today, ESG networks, digital infrastructure based on environmental, social and governance criteria – play an important role in ensuring sustainable and reliable network design. Such systems should ensure not only technical stability, but also transparency and security of information flow. From this point of view, the problem of network traffic detection and classification is one of the urgent issues for ESG networks [1-6]. Network traffic classification is the process of observing data passing through a network at the packet, flow, or application level and classifying them into specific categories. This process identifies network loads, user activity, security conditions, and other important indicators. In the research work, flow features were used to classify traffic flows. Flow feature extraction was performed in a laboratory setting and 89 different features were obtained. Traffic flows were divided into mice flow and elephant flow types, taking into account the number of packets in the flow duration and the time it took to record the flow [9, 10]. This new approach statistically analyzes traffic flow data (e.g., number of packets, bit rate, session duration) and determines the type of traffic based on behavioral patterns. This approach allows for increased security, reliability, and efficiency in ESG networks. In the research work, network status monitoring was performed as a result of traffic flow classification. As a result of monitoring, it is possible to identify and analyze unexpected network conditions and provide effective work for system users [18].

LITERATURE REVIEW: Wassie et al. [1] developed a traffic prediction model that can detect elephant flows. The performance of the Elephant flow prediction models used XGBoost, GBM, and GDF algorithms, respectively. The packet size and byte size attributes were considered as the main factors. Shi et al.

ISSN 3030-3907

[2] tried to find a new load balancing solution to ensure elephant flow detection and to alleviate network congestion and utilize high bandwidth. They proposed an OpenFlow-based load balancing strategy for data center networks that accurately detects elephant flows and applies separate routing schemes to different flow types to achieve high network capacity utilization. Data centers analyze many mice flows and elephant flows that are sensitive to latency [3], which account for more than 80% of the total load. The persistent congestion of long-term elephant flows is one of the main causes of network congestion. They considered a method to analyze the network traffic separately when elephant flow and mouse flow are mixed. Huang et al. [4] used the AdaBoost algorithm to analyze heavy and active flows using the Q-learning algorithm. As a result, efficient flow rules for incoming flows can be dynamically implemented without controller intervention. Riadi et al. [5] ANN can be used as an effective tool for classifying network packets. They found that the best classification accuracy, 99.6%, was obtained by ANN with the number of hidden layer neurons half the number of input neurons and twice the number of input neurons. Abdulhassan et al. [6] proposed that network switches classify all packets be-longing to the same flow by applying the same actions defined in the corresponding rule. Packet classification is to facilitate the support of new services by filtering, routing packets based on packet header information, and prioritizing certain flows. A new method for implementing multi-field packet classification of SDN flow charts was presented. Zhang et al. [7] proposed a network load prediction method called IoTGuard. It can predict the payload status in IoT networks based on real-time IoT network traffic. The IoTGuard method consists of two stages, namely, the application layer and the network flow payload status classification analysis. There are various network packet generators, but there is no set of rules that specify the requirements for packet generators. Therefore, when analyzing packets, they need to be recorded. During the recording process, the conditions are given and the analysis results are performed [8].

TRAFFIC CLASSIFICATION PARAMETERS: In this section, we will look at a set of packet-level features that are important for deep network traffic analysis. These features play a key role in the classification of network traffic, especially in the analysis of data received in the NetFlow format. During our research, the traffic data collected in the lab in the form of NetFlow was separated into 89 distinct features. These features represent specific parameters within the packet and are divided into the following categories [12, 13]:

• Source and destination attributes: IP addresses, port numbers, protocol types;

ISSN 3030-3907

• Time-related metrics: flow start and end times, time differences between packets;

• Volume-related metrics: number of packets sent and received, total bytes;

• TCP flag statistics: frequency of SYN, ACK, FIN, RST, PSH, and other flags;

• Directional flows: separate statistics for incoming and outgoing flows;

• Additional network metrics: TTL values, DSCP fields, flow-related entropy or dispersion.

89 features are separated at each packet level, and their combination creates the possibility of identifying application or user activity taking place on the network, monitoring security, and detecting anomalies at an early stage. Also, based on these features, automatic classification and prediction can be performed using various machine learning algorithms [11]. Thus, NetFlow-based packet feature analysis is recognized as one of the effective approaches to ensure reliable traffic monitoring and security in modern ESG networks. Packet size is an important measurement in classifying network traffic, which represents the number of bytes in a packet. In Internet traffic, packet sizes are often observed to have very small or very large values, which indicates the presence of an unbalanced distribution in the network. The distribution of packet size often takes the form of a heavy-tailed distribution, in which small packets are more common, but large packets carry more information in the network. The following formula can be used to express this situation mathematically:

$$P(X > x \sim x^{a}, \quad a > 0 \qquad (1)$$

where $P(X > x)$ is the probability that the packet size is greater than x, and a is the heavy packet parameter, the value of which determines how stretched the distribution is. The distribution of packet sizes in Internet traffic also uses Zipf's law, which is expressed as follows:

$$f(r) \propto \frac{1}{r^{s}} \qquad (2)$$

where f(r) is the frequency of the numerical (ranked) packet size r, and s is a distribution parameter. This law shows that there is a strong inverse relationship between the ranking of packet sizes in the frequency table and their occurrence rate. For example, if the packet sizes in a network are ordered as $x_1, x_2, \ldots, x_n$, the largest size will have the highest frequency, and subsequent sizes will decrease in speed. Understanding and modeling this imbalance in network traffic volume helps to achieve high efficiency in traffic classification and control systems. Especially since large-sized packets (elephant flows) constitute the main load of the network, monitoring and control mechanisms for them require special attention. Traffic

ISSN 3030-3907

Engineering (TE) is a systematic process that aims to group traffic flows across a network, efficiently use network resources, and reduce congestion [14]. TE is primarily designed to manage large volumes of traffic, and it improves system efficiency by identifying and optimizing flows in the network. However, Traffic Engineering is not very effective when working with small flows (mice flows). The reason for this is the variability of traffic flows in a random and unstable environment such as the Internet, in particular, the non-constant size of flows and the bandwidth allocated to them reduce the effectiveness of Traffic Engineering in accurately classifying traffic [15]. The variability of bandwidth observed in traffic flows is a natural phenomenon in many cases, which complicates the criteria for classifying traffic and creates the need for frequent reclassification of flows. This reduces the efficiency of overall network management. In network traffic, two main types of flows are distinguished according to their size and duration: heavy flows ("elephant" packets) and small flows ("mice" packets). This classification plays an important role in network management and traffic engineering [16-18]. Typical values for packet size in wired networks are as follows:

• The lower limit of the packet size is a few tens of bytes (for example, for control packets);

• The upper limit is the MTU (Maximum Transmission Unit) value, usually 1514 bytes.

When large files are transmitted, they are divided into fragments, and each fragment has a maximum size. In Table 1, S denotes the packet size and they are divided into groups.

Table 1.

Packet grouping based on real traffic data.

Packet type	**Size (bytes)**	**Description**
Control packets	$S < 200$	Control commands, no data upload
Data packets	$S > 200$	Data payload packets
Heavy flow packets	$300 < S < 1514$	Packets with a large payload

The duration of a network flow is the difference between the start and end times of a flow. This criterion is used to determine the nature of a flow and to distinguish between short-term and long-term flows. The duration of a flow can vary significantly depending on the application, protocol, and user activity.

ISSN 3030-3907

Table 2.

Classification by flow duration.

Flow type	Duration (Δt)	Description
Short flow	$1\ ms \leq \Delta t < 60\ sec$	For example, ICMP, DNS, HTTP requests, ACK packets
Average flow	$1\ min \leq \Delta t < 5\ min$	Normal user session or file transfer
Long flow	$\Delta t \geq 5\ min$	Video streaming, VPN, chat apps

Results: A flow is a sequence of IP packets transmitted over a network, uniquely identified by the following attributes:

- Source IP address
- Destination IP address
- Source port number
- Destination port number
- Transport protocol (Protocol ID) — usually TCP or UDP

If the above attributes are the same, these packets are considered to belong to the same flow. Typically, such flows are monitored in a unidirectional sequence, that is, packets are transmitted in only one direction. Flow-based classification is the process of dividing network traffic into flows and classifying them based on the parameters of each flow (for example, duration, average packet size, protocol type, etc.). This approach is widely used in processes such as real-time network activity monitoring, threat detection, quality of service (QoS), and traffic balancing. It classifies flows according to their size, duration and inter-arrival time. These parameters are defined as follows: The stream size is the total number of bytes transferred between the server and the wireless client during the connection. From the client's perspective, the measurement is not affected if a new server starts serving during the connection. However, this measurement is usually taken separately for the server/client pair. The mice flow and elephant flow were separated using the features Total Fwd Packets and Flow Duration extracted from the data recorded in the laboratory.

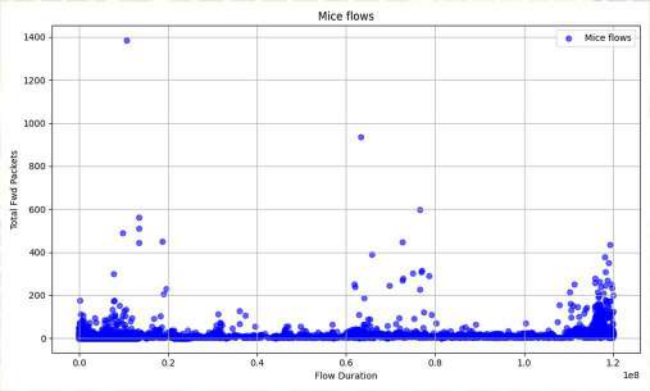


Figure 1. Mice flow chart.

ISSN 3030-3907

Mice flows are relatively small streams that are transmitted over short periods of time. Their duration is less than the time it takes to collect 10 KB of data, and each packet size is typically less than 500 bytes. Elephant flows, on the other hand, are streams that typically last more than an hour and have relatively large packet sizes. Therefore, an average elephant flow transmits over 3 MB of data, while a mice flow transmits only 10 KB.

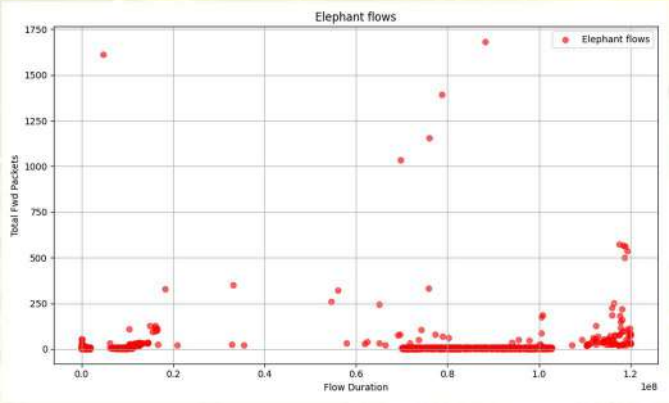


Figure 2. Elephant flow chart.

You can better understand the mice flow and elephant flow graphs by visualizing them together. As you can see in Figure 3, the mice flow and elephant flow graphs have similar visual representations, but one of the main differences between them is the duration of the flow. That is why they are depicted in this way in the graphical representation.

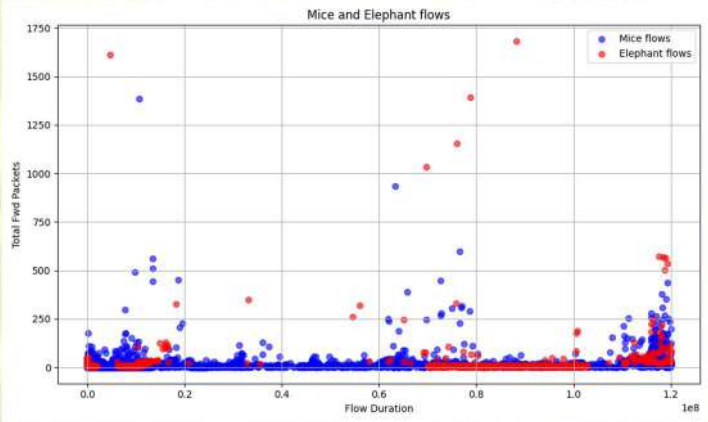


Figure 3. Mice flow and elephant flow graph combined view.

ISSN 3030-3907

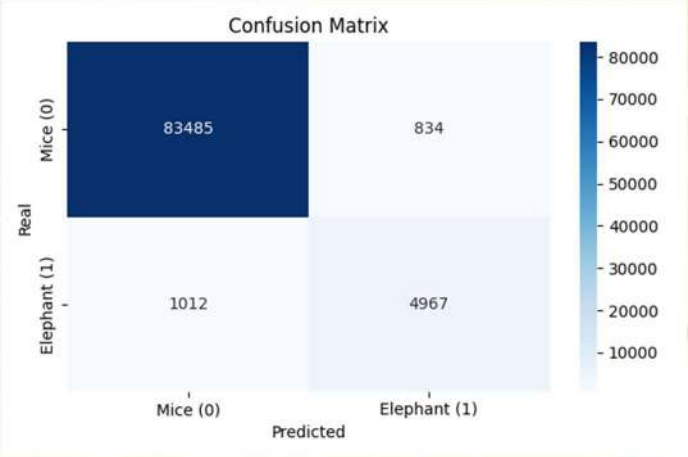


Figure 4. Confusion matrix of mice and elephant.

The classification results were analyzed using accuracy, precision, recall, and f1-score evaluations. Table 3 shows the support column, which indicates the number of streams recorded for each evaluation result.

Table 3.

Classification by flow duration.

Flow type	Precision	Recall	F1-score
mice	0.91	0.93	0.99
elephant	0.86	0.83	0.84

Table 4.

Classification according to the duration of the flow.

Indicator	Value	Explanation
accuracy	0.98	Overall percentage of correctly classified samples (98%). This is very good.
macro avg	0.94 (precision, recall, f1-score)	Simple average of the scores for each class.
weighted avg	0.98 (precision, recall, f1-score)	Weighted average of the classes by the number of samples (support).

If the number of elephant flows increases and their size increases, the load on the network increases, resulting in increased consumption of network resources. If the number of mice flows is large and of normal size, the network state can be considered stable. However, if the size of mice flows is larger than usual or the duration changes, a threat or anomaly may occur. When unusually large mice flows or too many elephant flows are detected, this indicates a potential attack or network failure.

$$N(t) = M(t) + E(t) \quad (3)$$

where $M(t)$ is the number of mice flows, $E(t)$ is the number of elephant flows, and $N(t)$ is the total number of traffic flows. Elephant flows are usually

ISSN 3030-3907

analyzed more often because they are one of the most effective ways to indicate network failures. By default, the system should issue a warning if elephant flows exceed 20% of the network load.

$$U(t) = \frac{E(t)}{N(t)} \qquad (4)$$

where $U(t)$ is the network load limit, which should be 20% as a rule. $M(t) = 83485, E(t) = 4967$. So $N(t) = 83485 + 4967 = 88452$ is the total traffic flow. From this, $U(t) \approx 0.06$ was obtained. The conclusion is that elephant flows accounted for 6% of the traffic flow, which shows that the traffic flow is stable. Because although the network has elephant flows, they are not at a level that affects the flow stability. The result of the traffic flow recorded in laboratory conditions leads to the conclusion that the system is operating stably. Because although mice flows account for 94% of the total traffic flow, there were no interruptions or larger mice flows that would affect the system stability.

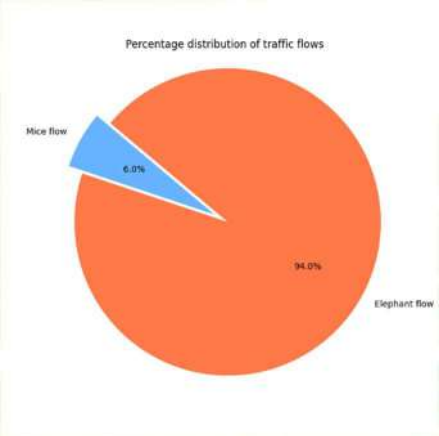


Figure 5. Mice flow and elephant evaluation diagram.

CONCLUSION: Which was used to determine the average and statistical data of traffic passing through a certain direction or node, ESG network traffic classification based on mice flow and elephant flow has been a very important mathematical and statis-tical formalization in cyberspace and computer networks. For ESG network in valuable information, such as: how loaded the link is, average point-to-point de-lay and average packet size, the correct application of traffic analysis results and synthesis methods. Preventing potential congestion and predicting future growth. Such important information supported engineers in designing reliable networks. In packet, flow and application contexts, this study was designed to provide an overview of traffic classification methods and consider them.

ISSN 3030-3907

REFERENCES

1. Wassie, G., Ding, J., & Wondie, Y. (2023). Traffic prediction in SDN for explainable QoS using deep learning approach. Scientific Reports (Nature Publisher Group), 13(1), 20607. doi:https://doi.org/10.1038/s41598-023-46471-8
2. Shi, X., Li, Y., Xie, H., Yang, T., Zhang, L., Liu, P., Liang, Z. (2020). An OpenFlow-based load balancing strategy in SDN. Computers, Materials, & Continua, 62(1), 385-398. doi:https://doi.org/10.32604/cmc.2020.06418
3. Cheng, W., Ren, F., Jiang, W., Qian, K., Zhang, T., & Shu, R. (2016). Isolating mice and elephant in data centers.
4. Huang, G., Ullah, I., Huang, H., & Kim, K. T. (2024). Predictive mobility and cost-aware flow placement in SDN-based IoT networks: A Q-learning approach. Journal of Cloud Computing, 13(1), 26. doi:https://doi.org/10.1186/s13677-024-00589-w
5. Riadi, I., Muhammad, A. W., & Sunardi. (2017). Network packet classification using neural network based on training function and hidden layer neuron number variation. International Journal of Advanced Computer Science and Applications, 8(6) doi:https://doi.org/10.14569/IJACSA.2017.080631
6. Abdulhassan, A., & Ahmadi, M. (2023). Many-fields packet classification using R-tree and field concatenation technique.
7. Zhang, H., Wang, J., Wang, X., Lu, K., Xu, T., & Zhou, Y. (2025). Payload state prediction based on real-time IoT network traffic using hierarchical clustering with iterative optimization. Sensors, 25(1), 73. doi:https://doi.org/10.3390/s25010073
8. Spiekermann, D., & Keller, J. (2022). Requirements for crafting virtual network packet captures. Journal of Cybersecurity and Privacy, 2(3), 516. doi:https://doi.org/10.3390/jcp2030026
9. Guan, S., Dong, T., & Cong, L. K. (2025). Method for EEG signal recognition based on multi-domain feature fusion and optimization of multi-kernel extreme learning machine. Scientific Reports, 15(1), 6601. doi:https://doi.org/10.1038/s41598-025-87569-5
10. Homaie, M., Mahboubi, A., Hartmann, D. J., Kadkhodaie, A., & Moussavi Harami, R. (2024). Flow unit classification and characterization with emphasis on the clustering methods: A case study in a highly heterogeneous carbonate reservoir, eastern margin of dezful embayment, SW iran. Journal of Petroleum Exploration and Production Technology, 14(10), 2703-2734. doi:https://doi.org/10.1007/s13202-024-01847-y
11. Kadish, S., Schmid, D., Son, J., & Boje, E. (2022). Computer vision-based classification of flow regime and vapor quality in vertical two-phase flow. Sensors, 22(3), 996. doi:https://doi.org/10.3390/s22030996

ISSN 3030-3907

12. Zhao, C., Liao, L. X., Chen, G., & Han-Chieh, C. (2025). Condensation of data and knowledge for network traffic classification: Techniques, applications, and open issues. Sensors, 25(8), 2368. doi:https://doi.org/10.3390/s25082368
13. Cui, S., Han, X., Han, D., Wang, Z., Wang, W., Li, Y., Lu, Z. (2024). FG-SAT: Efficient flow graph for encrypted traffic classification under environment shifts.
14. Sathish, K. M., & Praveena, T. (2020). Internet traffic analysis: mapreduce based traffic flow classification in hadoop environment. International Journal of Advanced Research in Computer Science, 11(2), 45-50. doi:https://doi.org/10.26483/ijarcs.v%25vi%25i.6507
15. Babadzhanov, A., Yarashov, I., Juraev, M., Otakhonov, A., Kudaybergenov, A., & Utemuratov, R. (2025). Mathematical Representation of Netflow Analysis Decision Making Based on Production Logic. International Journal of Advanced Computer Science & Applications, 16(9).
16. Juraev, M., Yarashov, I., Kudaybergenov, A., Babadzhanov, A., & Mamatova, Z. (2025). Research on Network Flow Based on Statistical Analysis Methods. International Journal of Advanced Computer Science & Applications, 16(6).
17. Yarashov, I., Kuvonchbek, R., Juraev, M., & Rahmonov, F. (2024, December). Processing and Analysis of Network Flow Data Using Intelligent Technologies. In Conference on Internet of Things and Smart Spaces (pp. 40-51). Cham: Springer Nature Switzerland.
18. Yarashov, I., Juraev, M., Ismatillayev, A., & Rakhimberdiev, K. (2024, December). Behavior Modeling in Computerized Production Systems. In Proceedings of the 8th International Conference on Future Networks & Distributed Systems (pp. 873-880).

ISSN 3030-3907