

МИНИСТЕРСТВО НАУКИ
И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ







СПб
ФИЦ
РАН

Санкт-Петербургский
Федеральный исследовательский центр
Российской академии наук



Zhernova Ksenia, Maxim Gorda

«Detecting Anomalies in Network Traffic Data and Its Graphical Representation Using Transformers»

SpacSec-2026

Relevance

- Most computer attacks occur over a network. Consequently, the analysis of network traffic is a critical security task.
- Traffic can be analyzed either as individual packets or as statistical data extracted from those packets. Visual analysis can simplify this process.
- A promising approach to visual analysis involves the use of computer vision, which enables the discovery of hidden patterns.
- Vision Transformers have proven highly effective in image analysis, as they can identify relationships between distant pixels and are capable of analyzing different parts of a dataset in parallel.

Contradiction:

- The vast majority of research relies on publicly available, preprocessed statistical datasets.
- Such datasets may contain errors, redundancies, and incorrect labeling.

→ There is a need to develop an approach that enables the rapid conversion of raw data into images, facilitating highly efficient visual analysis.

This would make it possible to combine the advantages of visual analysis with transformer architectures.

Novelty

Scientific novelty:

- the use of RGB matrices to represent packets of traffic by **byte-by-byte rendering without pre-processing** PCAP files,
- the subsequent comparison of the training performance on RGB matrices and on geometric images such as donut charts.

Practical significance:

- This will allow developers to use the proposed approach in information security systems close to real time.

State of the Art

- The subject is traffic anomaly detection using Transformers (TF) and Vision Transformers (ViT).
- Relevant papers in the subject area were divided into the following categories:

- TF in computer security:

Hartono, B., Silalahi, F. D., Muthohir, M.: Transformers in cybersecurity: Advancing threat detection and response through machine learning architectures. Journal of Technology Informatics and Engineering 3(3), 382–396 (2024)

Wu, Z., Zhang, H., Wang, P., Sun, Z.: RTIDS: A robust transformer-based approach for intrusion detection system. IEEE access 10, 64375–64387 (2022)

Long, Z., Yan, H., Shen, G., Zhang, X., He, H., Cheng, L.: A Transformer-based net-work intrusion detection approach for cloud security. Journal of Cloud Computing13(1), 5 (2024)

Karlsen, E., Luo, X., Zincir-Heywood, N., Heywood, M.: Benchmarking large language models for log analysis, security, and interpretation. Journal of Network and Systems Management 32(3), 59 (2024)

Bahmanisangesari, S.: A novel transformer-based multi-step approach for predicting common vulnerability severity score (2024)

Yin, K., Yang, Y., Yao, C., Yang, J.: Long-term prediction of network security situation through the use of the transformer-based model. Ieee Access 10, 56145–56157 (2022)

- Alphanumeric data analysis using ViT.

Sana L. et al.: Securing the IoT cyber environment: Enhancing intrusion anomaly detection with vision transformers. IEEE Access 12, 82443–82468 (2024)

Shaikh, J., Syed, T.A., Shah, S.A., Jan, S., Ain, Q.U., Singh, P.K.: Advancing DDoS attack detection with hybrid deep learning: integrating convolutional neural networks, PCA, and vision transformers. International Journal on Smart Sensing and Intelligent Systems 17(1), 20240040 (2024)

- Combining ViT with another model:

Du, C., Guo, Y., Zhang, Y.: A deep learning-based intrusion detection model integrating convolutional neural network and vision transformer for network traffic attack in the internet of things. Electronics 13(14), 2685 (2024)

State of the Art

- Converting traffic data into CSV and subsequent RGB-matrix analysis

a) Zhou, H., Zou, H., Li, W., Li, D., Kuang, Y.: HiViT-IDS: an efficient network intrusion detection method based on vision transformer. Sensors 25(6), 1752 (2025)

b) Ho, C. M. K., Yow, K. C., Zhu, Z., Aravamuthan, S.: Network intrusion detection via flow-to-image conversion and vision transformer classification. IEEE Access 10,97780–97793 (2022)

c) He, K., Zhang, W., Zong, X., Lian, L.: Network intrusion detection based on feature image and deformable vision transformer classification. IEEE Access 12, 44335–44350 (2024)

d) Kim, T., Pak, W.: Deep learning-based network intrusion detection using multiple image transformers. Applied Sciences 13(5), 2754 (2023)

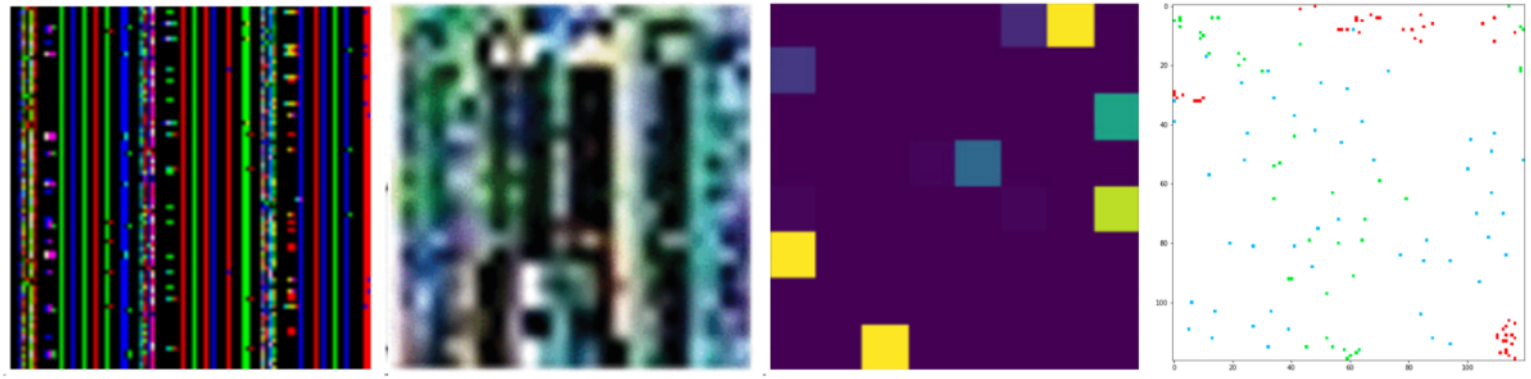


Fig. 1. Traffic data visualization in the other studies.

State of the Art

- Malware recognition:

Seneviratne, S., Shariffdeen, R., Rasnayaka, S., Kasthuriarachchi, N.: Self-supervised vision transformers for malware detection. IEEE access 10, 103121–103135 (2022)

Obidiagha, C.C., Rahouti, M., Hayajneh, T.: Deepimagedroid: A hybrid frameworkleveraging visual transformers and convolutional neural networks for robust androidmalware detection. IEEE Access 12, pp. 156285–156306 (2024)

Table 1. Applying datasets to network security research.

Study	Name of a data set	Data type	Data set	Study	Name of a data set	Data type	Data set
[4]	CICIDS 2017	CSV	[20]	[12]	CIC-DDoS2019	CSV	[22]
	UNSW-NB15	CSV	[21]	[13]	TON-IoT	CSV	[30]
[5]	CICIDS2017	CSV	[20]		Edge-IIoTset	CSV	[32]
	CIC-DDoS2019	CSV	[22]	[14]	CIC IDS2017	CSV	[20]
[6]	CIC-IDS 2018	CSV	[23]		UNSW-NB15	CSV	[21]
[7]	AA	CLF	[24]	[15]	CIC IDS2017	CSV	[20]
	CSIC	TXT	[25]		UNSW-NB15	CSV	[21]
	ECML/PKDD	TXT	[26]	[16]	ISCXIDS2012	CSV, PCAP	[33]
[8]	NVD API	JSON	[27]		CSE-CIC-IDS2018	CSV	[34]
[9]	UNSW-NB15	CSV	[21]	[17]	MalNet	IMG	[35]
	CSE-CIC-IDS2018	CSV	[28]	[18]	Drebin-215	IMG	[36]
[10]	NSL KDD train+	CSV	[29]		HealthCheck	IMG	[37]
[11]	TON-IoT	CSV	[30]		NATICUSdroid	IMG	[38]
	Bot-IoT	CSV	[31]				

Conclusion:
in the majority of studies, preprocessed statistical data in CSV format are used.
Some of these studies used RGB-matrices. However, the images were generated based on the statistical CSV data.

Dataset (PCAP)

Table 2. Attacking profiles: MITRE ATT&CK techniques, targets and tools.

Profile	ATT&CK technique	Target and settings	Tool
nmap_scan	T1046	laboratory network, SYN type, timing 4	nmap -sS
nmap_vuln	T1595.002	web server, timing 4	nmap -sV –script vuln
nmap_os_detect	T1592.001	laboratory network, timing 4	nmap -O
hydra_ssh	T1110	server node, SSH credential selection	hydra
ping_flood	T1499	web server, 30 sec., 0.01 sec. interval	ping -i
syn_flood	T1498.001	web server (port 80), 30 sec., high intensity	hping3 -S
udp_flood	T1498.001	DNS server (port 53), 30 sec., high intensity	hping3 –udp
slowloris	T1499.002	web server (port 80), 200 connections, 60 sec.	slowhttptest -H
arp_spoof	T1557.002	user node, gateway - server node	arpspoof
smb_enum	T1135	file server	smbclient -L
gobuster_dir	T1595.003	web server, 10 threads	gobuster dir

The collected dataset simulates the traffic of a real network:

- an isolated laboratory network was deployed,
- user profiles were granted partial access to the Internet at authorized addresses.

The server-side background:

- a web server, DNS, SMTP, a Postgres database management system, the Samba file service, NTP, a Git daemon, a syslog receiver, and an SSH service used for credential harvesting.

Approach: generating images

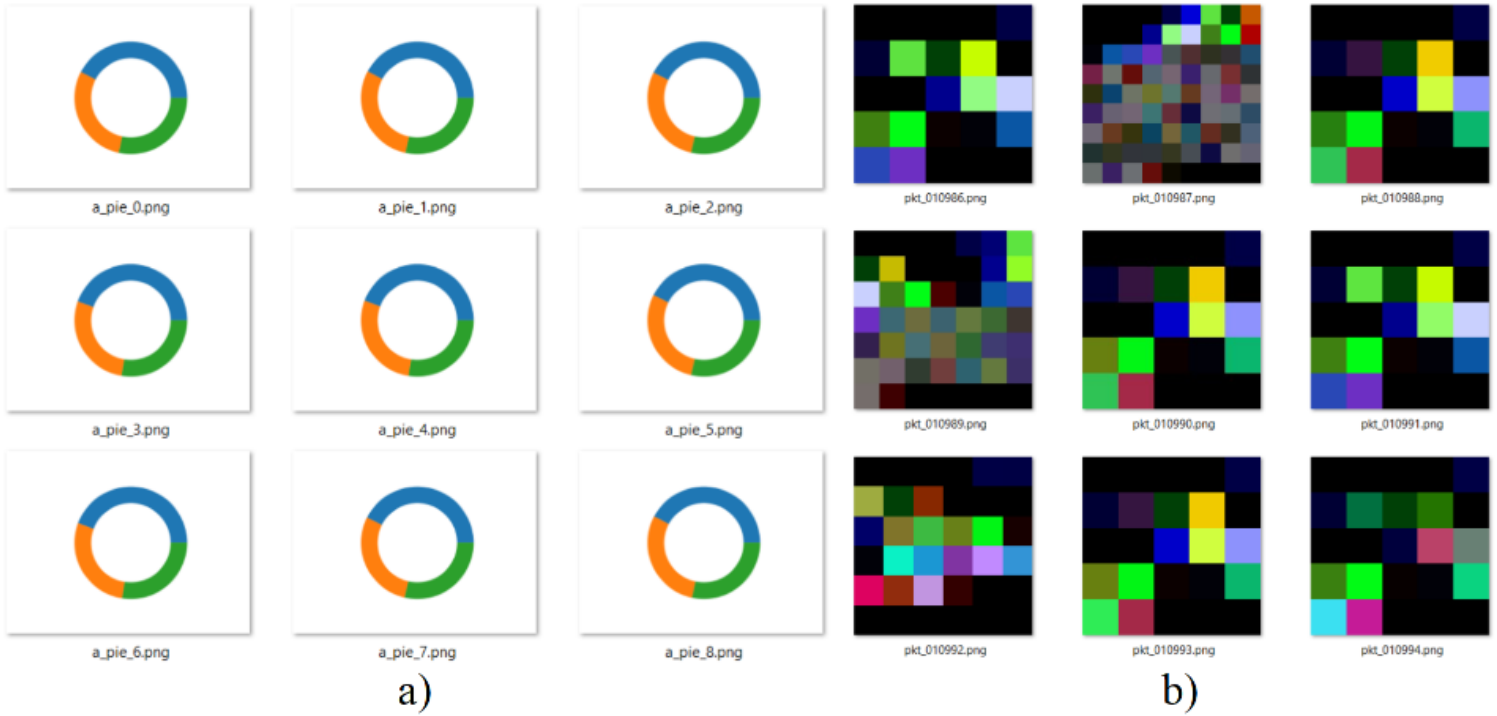


Fig. 2. Traffic visualization based on: a) geometry, b) byte-by-byte representation.

The general approach consists of:

(1) generating images from the collected traffic dumps;
(2) training models on the resulting datasets or images.

Images are divided into 2 groups:

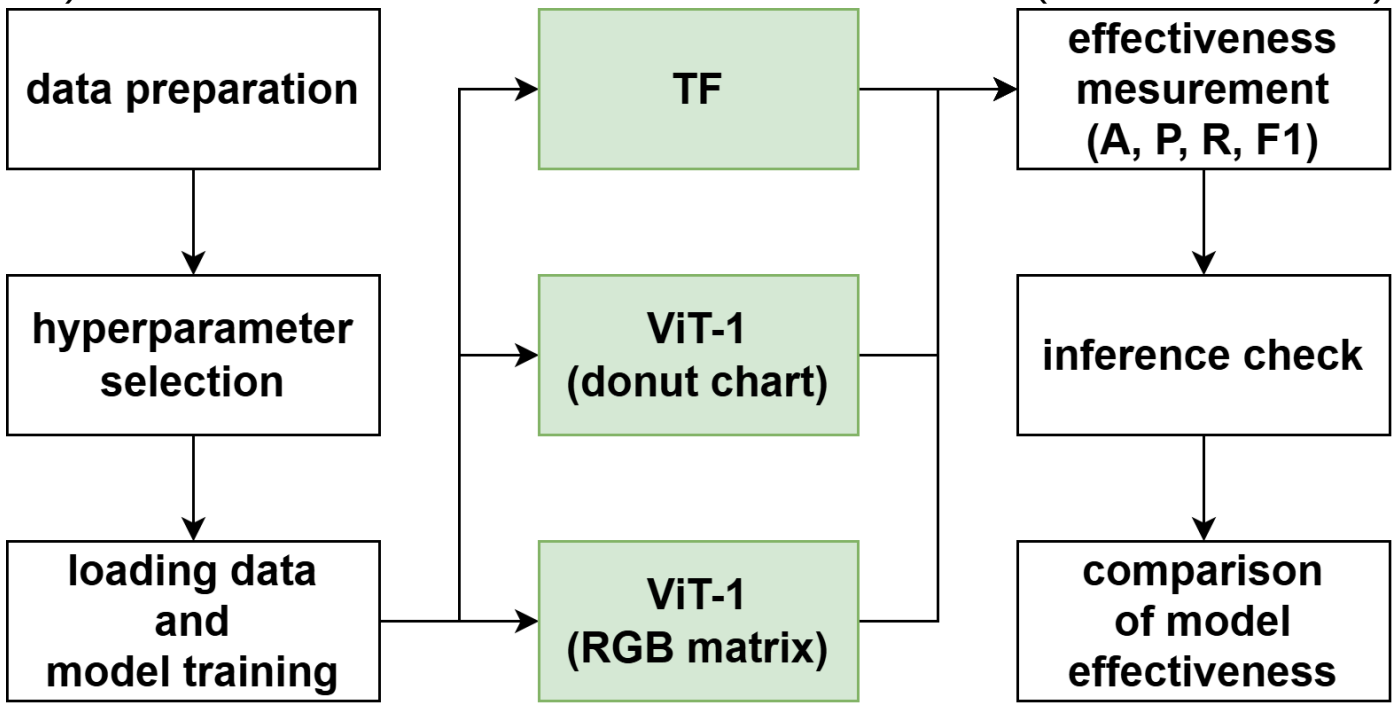
A) Donut chart. Selected metrics to be displayed: *window_size, packet_length, TTL*.
The PCAP files were analyzed using the *Scapy* library;
B) RGB matrix. Without preprocessing.
Packets of the PCAP files were converted into matrices byte-by-byte using *Pillow*.

Approach: training models (Table 3)

Step	Description
data preparation	image generation (donut chart or RGB matrix).
hyperparameter selection	searching for balance between memory consumption, learning rate, and model efficiency.
loading data into the model and training	the TF model was trained on two PCAP files: one containing abnormal traffic and one containing normal traffic. The ViT model, designed for analyzing graphical information, was trained on labeled image sets: RGB matrices of network traffic packets and donut charts based on selected metrics.
measurement of the model effectiveness	accuracy, precision, recall, and F1-score.
inference check	a new set of images not seen by the models was fed to the trained model to additionally check efficiency of the visualization techniques.
comparison of model effectiveness	comparing the metrics of models trained on textual and graphical data.

Experimental results

After model training three models were obtained:
1) **TF** - a standard Transformer for the alphanumeric data analysis;
2) **ViT-1** - a Vision Transformer for computer vision tasks (for donut charts);
3) **ViT-2** - a Vision Transformer identical to the **ViT-1** (for RGB matrices).



1) **TF**:
- trained on PCAP files with normal and attack traffic;
- raw traffic;
2) **ViT-1**:
- trained on donut charts;
- selected metrics with anomalies and normal behavior;
3) **ViT-2**:
- trained on RGB matrices of packets;
- raw traffic.

Fig. 3. The training approach scheme.

Experimental results

Table 4. The performance metrics of the trained models.

Model	Data type	Accuracy, %	Precision, %	Recall, %	F1-score, %
TF	PCAP	94.25	96.09	92.25	94.13
ViT-1	PNG (donut chart)	85.94	80.36	97.95	88.29
ViT-2	PNG (RGB matrix)	**98.98**	**98.69**	**99.53**	**99.11**

Table 5. The comparison of performance metrics with the other studies.

Model	Data type	Accuracy, %	Precision, %	Recall, %	F1-score, %
[Zhou, H. et al.]	CSV to RGB	100.00	100.00	100.00	100.00
[Ho, C.M.K. et al.]	CSV to RGB	98.50	98.23	-	-
[He, K. et al.]	CSV to RGB	99.50	99.59	-	-
[Kim, T. et al.]	CSV to RGB	95.50	94.22	93.63	92.52
Mean		98.375	98.01	96.815	96.26
Proposed	PCAP to RGB	98.98	98.69	99.53	99.11

Advantages and Limitations

The advantages of the presented approach:

1. high anomaly recognition accuracy;
2. no need for metric extraction and preprocessing;
3. the proximity of the trained model to real-time operation of security systems.

The limitations of the resulting models are shown in Table 6:

Disadvantage	Answer
Training time of the ViT on images for 10 training epochs took approximately 28 hours.	However, a trained model instantly classifies the images, allowing it to be used in real time.
Topological binding of features to the image coordinate grid, which could potentially allow attackers to bypass this security system.	Future work plans to explore possible modifications to the model to address this weakness.
The current method **does not consider image sequences**, although temporal dependencies are important for detecting traffic anomalies.	We also plan to address this shortcoming in future research.

Conclusion

1. The obtained TF-based models enable the detection of different network attacks, (port scanning, vulnerability detection, operating system fingerprinting, various types of DoS, and others) in raw network traffic.
2. The ViT model trained on RGB matrices generated byte-by-byte from traffic packets proved to be the most effective:

accuracy – 98.98%,precision – 98.69%, recall – 99.53%, and F1-score – 99.11%.

This study may be useful to researchers in the field of information and computer security, as well as to developers of automated information security systems and intrusion detection systems.

Future work will focus on:

- investigating the *sensitivity* of the ViT to *different intensities* of various network attacks;
- we also plan to address the limitations in future research.

Thank You for Your attention!

Zhernova Ksenia
zhernova@comsec.spb.ru