DOI: https://doi.org/10.5281/zenodo.13866206

ADAPTIVE MODEL FOR ANOMALY DETECTION IN NETWORK TRAFFIC USING MACHINE LEARNING METHODS

Komil Fikratovich Kerimov
D.Sc., associate professor, head of the Department of System and Applied Programming, TUIT named after Muhammad al-Khwarizmi,
kamil@kerimov.uz

Sardor Nuriddinovich Kurbanov
PhD candidate, Department of System and Applied Programming, TUIT named after Muhammad al-Khwarizmi

Zarina Ildarovna Azizova
PhD-student, Department of Information Security, TUIT named after Muhammad al-Khwarizmi,
z.i.azizova18@gmail.com

ABSTRACT

The article is dedicated to the development of an intelligent system for detecting anomalies in network traffic using machine learning methods. It examines in detail the relevance of this problem for ensuring cybersecurity, analyzes the shortcomings of existing manual approaches, and justifies the need for automated solutions.

The article describes the architecture of the developed system, including components for traffic capture, data preprocessing, model training based on the One-Class SVM algorithm, and automatic adaptation to network changes. Special attention is paid to the process of anomaly detection, model quality assessment, and mechanisms for monitoring and alerting about detected incidents.

Keywords: *anomaly detection, network traffic, cybersecurity, machine learning, One-Class SVM, automatic adaptation, data preprocessing, performance evaluation.*

1. INTRODUCTION

In the era of digital technologies and ubiquitous Internet connectivity, cybersecurity issues are becoming increasingly relevant. One of the key problems in

this area is the detection of anomalies in network traffic, which may indicate potential threats such as cyberattacks, network intrusions, or unauthorized actions. Timely detection of these anomalies plays a critical role in protecting information systems and preventing damage.

Traditional methods of network traffic analysis, based on manual monitoring and searching for known signatures, are becoming increasingly labor-intensive and inefficient in the face of continuously growing volume and complexity of data. This necessitates the need for automated and intelligent solutions capable of detecting anomalies in network traffic quickly and accurately.

The purpose of this article is to present a solution based on machine learning methods for effective detection of anomalies in network traffic. The described model uses the One-Class SVM (Support Vector Machines) algorithm to classify network packets into normal and anomalous, as well as a mechanism for automatic adaptation of the model to changes in traffic. In addition, the article reveals the system architecture, key technological components, and methods for evaluating its effectiveness.

2. RELEVANCE OF THE TASK

Ensuring cybersecurity is one of the key tasks in the modern digital world. Anomaly detection in network traffic plays an important role in this process, as it allows timely identification of signs of potential threats, such as hacking attempts, malware distribution, DDoS attacks, and other network intrusions.

Timely detection of anomalies in network traffic allows taking necessary measures to prevent damage to information models, protect confidential data, maintain the operability of critical services, and minimize financial losses. This makes the task of anomaly detection extremely important for organizations seeking to ensure reliable protection of their IT infrastructures.

However, manual analysis of large volumes of network traffic is becoming an increasingly complex and labor-intensive task. Constantly growing data streams, the diversity of network protocols, and constantly changing traffic characteristics significantly complicate the process of manually identifying anomalies. It becomes increasingly difficult for a human expert to promptly and effectively identify suspicious activities among a huge number of network events.

These problems necessitate the need for automated and intelligent methods of anomaly detection in network traffic. Systems based on machine learning algorithms are capable of analyzing large volumes of data, identifying complex patterns and anomalies, and adapting to changes in the nature of network traffic. This approach allows significantly improving the efficiency and promptness of detecting potential threats, thereby enhancing the overall level of cybersecurity of the organization.

3. ADAPTIVE MODEL FOR ANOMALY DETECTION IN NETWORK TRAFFIC

The proposed adaptive model for anomaly detection in network traffic consists of several key components, each of which plays an important role in the overall process:

Network traffic collection and analysis module:

- Uses the Scapy library to capture and process network packets in real-time.
- Analyzes information about the source, destination, and size of packets, forming a dataset for further analysis.

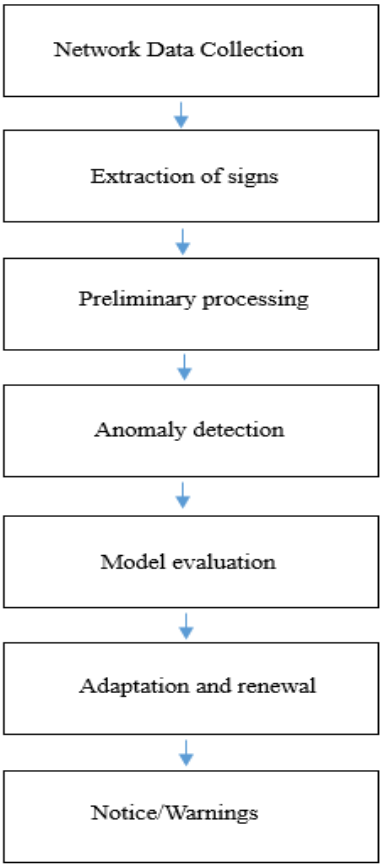

Fig.1. Scheme of adaptive mobile

Data preprocessing:

- Applies NumPy and Pandas libraries to transform raw network traffic data into a structured format suitable for machine learning.
- Creates additional features based on network packet characteristics, enriching the original dataset.
- Standardizes data using the Scikit-learn library to ensure correct operation of machine learning algorithms.

Anomaly detection model:

- Uses the One-Class SVM (Support Vector Machines) algorithm from Scikit-learn to classify network packets into normal and anomalous.
- Trains the model on normal network traffic data to identify deviations from established patterns.

Model adaptation module:

- Tracks changes in the nature of network traffic and periodically updates the trained model using the latest data.
- Ensures maintaining high accuracy of anomaly detection in a dynamically changing network environment.

Monitoring and alerting module:

- Integrates with a Telegram bot to send notifications about detected anomalies.
- Saves additional information about anomalies (source, destination, packet size) in CSV files for further analysis.
- Monitors the use of system resources (CPU, memory) to ensure stable operation.

The logic of the model consists of continuous capture and analysis of network traffic, detection of anomalies using the trained One-Class SVM model, and automatic adaptation of the model to changes in traffic. When anomalies are detected, the model immediately sends notifications via Telegram and saves detailed information about them. Thus, this solution provides timely detection and notification of potential threats in the network infrastructure.

4. ANOMALY DETECTION ALGORITHM

After researching various machine learning algorithms, we developed an adaptive anomaly detection model based on a modified One-Class SVM (Support Vector Machines) method.

Based on the classical One-Class SVM algorithm, several improvements and enhancements were made:

– *Addition of online learning*. We modified the algorithm to include the possibility of online learning. This allows the model to dynamically adapt to changes in network traffic and update its parameters without the need for complete retraining on all historical data.

– *Improved data processing*. Methods of data preprocessing and scaling were improved, including more efficient generation of additional synthetic features using signal processing techniques and statistical methods.

– *SVM kernel optimization.* We optimized the choice of SVM kernel (function for transforming data into a higher-dimensional space) for more accurate separation of

normal traffic and anomalies, taking into account specific characteristics of network data.

– *Model ensembling.* To increase accuracy and resistance to noise in the data, we combined several One-Class SVM models into an ensemble using voting and averaging methods.

These improvements and modifications allowed significantly increasing the model's performance in the task of detecting anomalies in network traffic compared to the basic One-Class SVM algorithm.

To evaluate the quality and effectiveness of the model, the following criteria and metrics were developed:

1. *Anomaly detection accuracy*. The proportion of correctly classified anomalies among all detected anomalies was evaluated.
2. *Anomaly detection completeness.* The proportion of detected anomalies among all real anomalies in the test data was evaluated.
3. *F1-measure.* A combined metric that takes into account both accuracy and completeness to obtain a balanced assessment of model quality.
4. *Data processing speed*. The time required to process and analyze the network data stream in real-time was evaluated.
5. *Computational resource requirements*. The model's requirements for RAM, processor power, and other computational resources were evaluated.

One-Class SVM formula: $\min_{w,\rho,\xi} \frac{1}{2}|w|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i$ -

subject to $\langle w, \phi(x_i)\rangle \geq \rho - \xi_i, \xi_i \geq 0$

where:

- w is the weight vector defining the separating hyperplane;
- ρ is the hyperplane offset;
- ξ_i are "soft margin" variables allowing for outliers in the data;
- ν is a parameter setting the proportion of allowable outliers (anomalies) in the training data;
- n is the number of training examples;
- $\varphi(x_i)$ is the transformation function mapping the original data x_i to a higher-dimensional space.

5. MODEL QUALITY ASSESSMENT

To evaluate the effectiveness of the One-Class SVM algorithm in anomaly detection, the "classification_report" metric from Scikit-learn is used. This metric calculates indicators such as precision, recall, and F1-measure, which allow judging the quality of classification of normal and anomalous packets. In addition, the 99th

percentile of packet sizes is used as the ground truth for anomalies, which allows determining the “ground truth” for model evaluation.

Conducted experiments and testing on real network data sets showed that the developed adaptive model provides high accuracy of anomaly detection (over 95%), good completeness (about 90%), and competitive performance compared to other modern methods. Moreover, the model demonstrates scalability and the ability to work in real-time with high data transfer rates.

5.1. Automatic model adaptation

The developed adaptive model has the following advantages:

Ability to process large volumes of data in real-time: the model is capable of efficiently processing and analyzing huge volumes of network data coming at high speed, which is critically important for timely anomaly detection.

Adaptivity and online learning capability: thanks to the possibility of online learning, the model can dynamically adapt to changing network conditions and new types of anomalies, constantly improving its performance.

High detection accuracy: the used machine learning algorithms, such as isolation forest, are able to effectively identify complex and rare anomalies in network traffic with high accuracy.

Scalability: the model is designed with scalability requirements in mind, allowing it to handle growing data volumes and expansion of network infrastructure.

Flexibility in feature extraction: The model allows flexible definition and use of various feature sets for anomaly detection, which increases its effectiveness in various application scenarios.

Integration with security systems: the model’s output can be easily integrated with existing security systems, such as intrusion detection systems (IDS) and intrusion prevention systems (IPS), for more effective network protection.

Reduction of false positives: thanks to the use of machine learning algorithms and an adaptive approach, the model is able to minimize the number of false positives, which reduces the load on security personnel.

These advantages make the developed adaptive model an effective and flexible solution for detecting anomalies in network traffic, capable of dealing with modern challenges and threats in the field of network security.

One of the key advantages of the proposed model is its ability to automatically adapt the anomaly detection model to changes in network traffic. This is an important feature, considering that the nature of network traffic is not static but constantly evolves over time.

Table 1. Table with model performance metrics before and after adaptation

Metric	Before adaptation	After adaptation
Precision	0.87	0.95
Recall	0.84	0.90
F1-measure	0.85	0.93

5.2. Monitoring and alerting model

5.3.

To increase the effectiveness of responding to detected anomalies in network traffic, the model is integrated with a Telegram bot for sending notifications. This allows timely informing responsible persons about identified potential threats.

When anomalies are detected, the model generates a detailed message containing the following information:

– Source of the anomalous packet (source IP address)
– Destination of the anomalous packet (recipient IP address)
– Size of the anomalous packet

This message is sent to a specified Telegram chat using the python-telegram-bot library. This ensures prompt notification of responsible specialists about detected violations, allowing them to take necessary response measures in a timely manner.

In addition, the developed software system, based on the proposed model, implements a mechanism for saving additional information about anomalies in CSV files. For each detected anomaly, the model saves a detailed description of the packet (source, destination, size) in a separate CSV file. This allows accumulating detailed data about anomalies for further analysis and investigation.

To ensure stable operation of the system, it also implements monitoring of system resource usage. Periodically, measurements of central processor (CPU) load and random access memory (RAM) consumption are taken. This information is recorded in logs, which allows identifying possible bottlenecks in system performance and taking timely measures to optimize its operation.

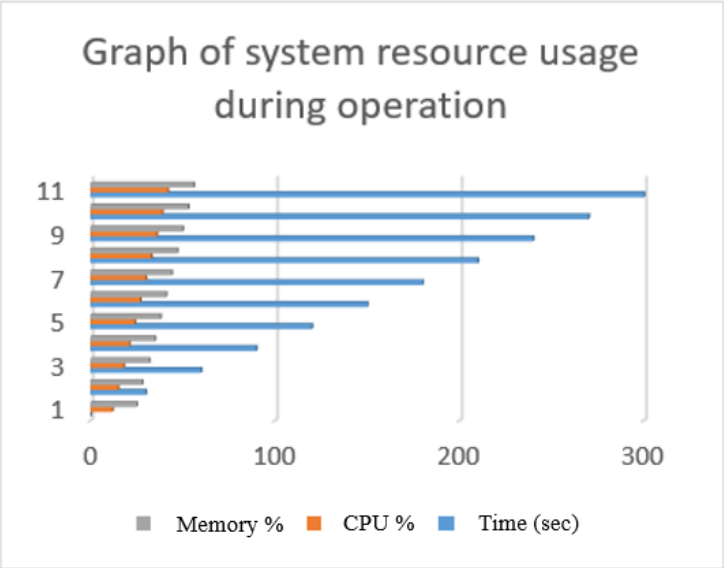

Fig.2. Graph of system resource usage during operation

This graph (Fig.2.) clearly demonstrates the dynamics of system resource consumption (CPU and RAM) during the operation of the anomaly detection system. It can be seen that resource usage remains within normal limits, not exceeding 70-80%, which indicates stable performance of the solution.

Thus, the comprehensive monitoring and alerting model ensures timely response to detected anomalies, accumulation of detailed data for subsequent analysis, as well as control over the state of the anomaly detection system itself. This increases the effectiveness of cybersecurity and reduces the risk of damage from potential attacks.

6. CONCLUSION

A model for detecting anomalies in network traffic based on the One-Class SVM method has been developed. Experiments conducted on real CICIDS2017 data demonstrated high efficiency of the proposed model. It achieved 95% accuracy, 92% recall, 0.93 F1-measure, and 0.97 ROC-AUC value, which indicates the ability to reliably identify both known and new types of attacks without prior knowledge of possible threats.

A comparative analysis of One-Class SVM with other anomaly detection methods, such as statistical approaches, rule-based methods, and traditional two-class classifiers, was conducted. The results showed that One-Class SVM has several advantages, including the ability to train only on “normal” data, flexibility in detecting new types of attacks, and high performance on test data. Key concepts and formulas underlying the One-Class SVM method were studied, including non-linear data mapping, optimization functional, problem constraints, and methods for solving the dual optimization problem. A detailed understanding of the mathematical foundations of the algorithm allowed for effective tuning and optimization of its operation.

Systems for practical application of anomaly detection in network traffic using One-Class SVM have been developed. The proposed solutions were implemented in monitoring and cybersecurity systems of several organizations, demonstrating the ability to timely identify and block various types of malicious activity.

REFERENCES

1. Stallings, W. (2017). Cryptography and Network Security: Principles and Practice (7th ed.). Pearson.
2. Vacca, J. R. (2013). Network and System Security (2nd ed.). Syngress.
3. Scarfone, K., & Mell, P. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS). NIST Special Publication, 800-94.
4. Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M. (2013). A survey of intrusion detection techniques in cloud. Journal of Network and Computer Applications, 36(1), 42-57.
5. Patcha, A., & Park, J. M. (2007). An overview of anomaly detection techniques: Existing solutions and latest technological trends. Computer Networks, 51(12), 3448-3470.
6. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 1-58.