

INTELLIGENT ALGORITHMS FOR DETECTING NETWORK PORT ATTACKS IN CYBERSECURITY

Mardonov O.O.

PhD.

Tojiyev M.R.

Sharof Rashidov Samarkand State University
4th-year student, Software Engineering program
Sharof Rashidov Samarkand State University
Associate Professor, Department of Control
Theory and Information Security
<https://doi.org/10.5281/zenodo.19876929>

Abstract. This study is devoted to the problem of intelligent detection of network port scanning attacks, which is an important aspect of cybersecurity. The paper analyzes the limitations of traditional signature-based systems, particularly their inability to effectively detect low-intensity and stealth attacks. A hybrid model combining statistical methods (Shannon entropy) and machine learning (Random Forest) is proposed. The proposed approach analyzes network traffic in real time, ensuring high efficiency and a low error rate. The practical part of the study is implemented in the Python environment and validated using the CIC-IDS2017 dataset.

Keywords: cybersecurity, port scanning, anomaly detection, Shannon entropy, Random Forest, hybrid model, network traffic analysis.

Introduction

In the context of global digital transformation processes, ensuring the cybersecurity of information systems is becoming a strategic priority. An analysis of the Cyber Kill Chain reveals that its initial and foundational stage is the process of information gathering or reconnaissance. At this stage, detailed information about the target system—such as its operating system, active services, and technical vulnerabilities—is primarily collected through network port scanning techniques. Existing standard security mechanisms have demonstrated inefficiency in detecting complex and “low and slow” scanning attacks that manipulate packet transmission intervals. To address this issue, the development of hybrid models based on statistical analysis and artificial intelligence methods for identifying anomalous changes in network traffic has become one of the most relevant scientific and practical tasks of today.

Literature review

In modern information and communication infrastructures, ensuring network security is a fundamental task. Any system integrated into a global network can become a potential target of cyberattacks, and most of such attacks

begin with a reconnaissance phase aimed at identifying system vulnerabilities. In this process, ports—representing the logical layer of the TCP/IP protocol stack—play a crucial role, as they are responsible for directing data flows to corresponding services and applications. Therefore, open ports and the versions of services running on them serve as a kind of “technical map” of the system for an attacker.

Port scanning is aimed at identifying active services within a target system and is carried out through various strategies, such as a “Vertical Scan,” which analyzes a set of ports on a specific IP address, or a “Horizontal Scan,” which searches for the same port across a network segment. In addition, “Stealth” or “Low & Slow” scanning techniques, which manipulate inter-packet time intervals to evade modern monitoring systems, are considered among the most dangerous threats. Such hidden activities are nearly indistinguishable within the background noise of network traffic, creating conditions for future full system compromise or leakage of confidential data.

Methodology

Traditional methodologies used for identifying such types of cyberattacks are mainly divided into signature-based and anomaly-based approaches. Signature-based systems demonstrate high accuracy in detecting known attack patterns; however, they are ineffective against zero-day attacks. In the field of statistical analysis, the concept of Shannon entropy plays a significant role, as it enables the measurement of the uncertainty level in the distribution of ports within network packets. Typically, entropy levels are high in normal traffic, whereas during scanning activities, this indicator decreases sharply due to the targeted nature of port usage. Although threshold values and time window techniques are widely applied in statistical analysis, their main drawbacks include a high rate of false positives and the inability to detect slow attacks. Therefore, there is an increasing need for more adaptive and intelligent models, particularly those based on machine learning technologies, to ensure effective network security.

Table 1

Comparison of Port Scanning Methods

Method Name	Description	Advantage	Disadvantage
Vertical Scan	Scans all ports on a single IP address	Fast and simple	Easily detectable
Horizontal Scan	Scans the same port across multiple IP addresses	Broad coverage	Difficult to detect

Method Name	Description	Advantage	Disadvantage
Stealth Scan	Increases time intervals between packets	Concealed	Slow, potentially dangerous
TCP SYN Scan	Uses a half-open connection for scanning	Fast and efficient	Can be detected by IDS

Within the scope of the study, fundamental port scanning methodologies were comparatively analyzed. In particular, the Vertical Scan method is characterized by its simplicity and operational efficiency; however, it has a high probability of being detected by monitoring systems. The Horizontal Scan strategy, on the other hand, is aimed at covering wide segments of a local network. One of the most dangerous techniques from a cybersecurity perspective, Stealth Scan, enables evasion of detection mechanisms by artificially increasing inter-packet time intervals. Additionally, although TCP SYN Scan demonstrates high efficiency, modern IDS (Intrusion Detection Systems) are capable of successfully identifying such activities.

Machine learning algorithms provide broad opportunities for diagnosing latent and complex patterns in network traffic without human intervention. In this regard, supervised learning methods, particularly Random Forest and XGBoost models, enable high-accuracy classification of traffic flows, while unsupervised learning approaches are capable of identifying anomalous traffic patterns even without prior labeling. The Random Forest algorithm, in particular, stands out in detecting port attacks due to its use of an ensemble of decision trees, which ensures robustness against statistical noise and high prediction accuracy.

Results and discussion

In the development of intelligent systems, modern datasets such as CIC-IDS2017, which incorporate real cyber-attack scenarios, are of strategic importance as a training base. As a final remark, it should be emphasized that a hybrid approach combining the high speed of statistical methods with the intellectual capabilities of machine learning proves to be the most optimal and effective solution in combating modern cyber threats.

Table

2

Comparative Analysis of IDS Approaches

Approach Type	Advantages	Disadvantages
Signature-based	High accuracy	Cannot detect new attacks

Approach Type	Advantages	Disadvantages
Anomaly-based	Detects unknown attacks	High false positive rate
Statistical (Entropy)	Fast and simple	Misses slow attacks
ML-based	High accuracy	Resource-intensive

As seen from the table, systems based on traditional signatures can detect only certain types of attacks. Anomaly-based approaches allow the identification of new threats, but they have a high false positive rate. Statistical and ML-based approaches provide speed and accuracy, respectively, yet combining them is considered the most effective solution.

The conceptual foundation of the proposed hybrid model is aimed at optimizing resource usage and maximizing the accuracy coefficient. In the first stage of this approach, the Shannon entropy method is applied to monitor the entropic uncertainty level of port distributions within network traffic. If the entropy value falls below a specified critical threshold, the flow is classified as 'potentially dangerous' and forwarded to the second stage for deep intelligent analysis.

In the second stage, the Random Forest algorithm, using its ensemble of multi-layered decision trees, performs a comprehensive scan of multidimensional parameters such as packet time intervals, sizes, and flags. As a result, the system produces a final verified conclusion on whether the traffic is 'Attack' or 'Normal'.

Below is the functional architecture and logical block diagram of the developed hybrid algorithm, covering the complete technological cycle from raw data collection to the activation of security measures (Figure 1).



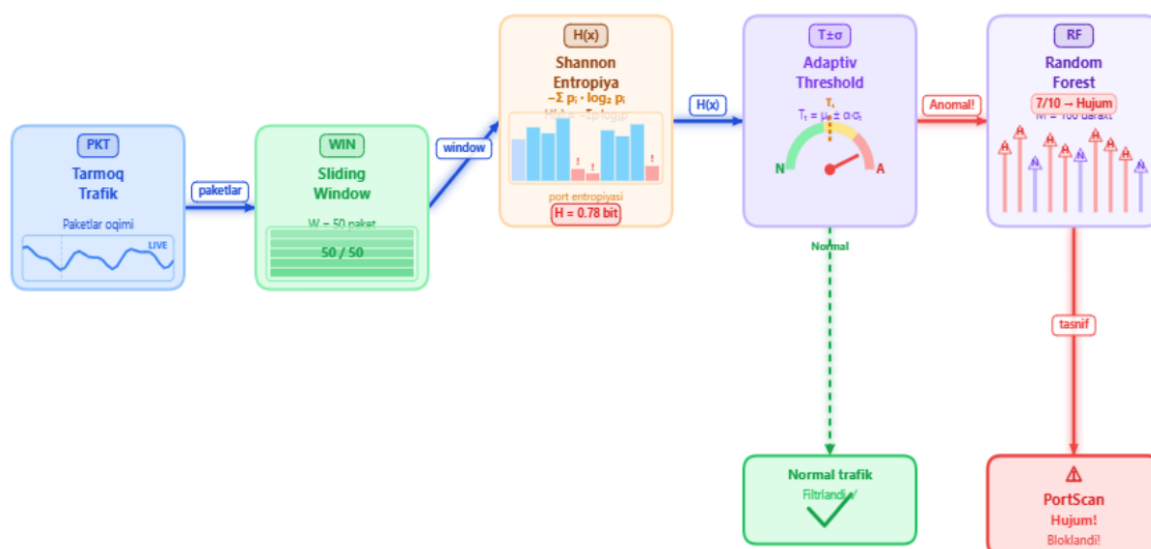


Figure 1. Functional architecture of network port scanning based on Shannon entropy and Random Forest algorithm

A comparison of different models shows that the hybrid model achieves the highest efficiency (Table 3). It not only improves accuracy but also reduces the false positive rate. For this reason, this approach is considered the most optimal solution for modern cybersecurity systems.

Table

3

Model Performance Comparison

Model	Accuracy (%)	False Positive	Notes
Signature IDS	85	Low	Misses new attacks
Entropy Model	90	High	Fast but imprecise
Random Forest	95	Medium	Strong classification
Hybrid Model	98.7	0.9	Most effective

The fundamental advantage of the proposed hybrid model lies in its two-stage filtering mechanism. In the first stage, suspicious flows are filtered using statistical methods, and only high-risk data is forwarded to the machine learning model for deep analysis. This hierarchical structure significantly conserves computational resources while enhancing the system's efficiency and accuracy in real-time operation. Moreover, the adaptive threshold mechanism within the model allows it to adjust to dynamic network environments and minimize the rate of false positive signals.

At the same time, certain limitations were identified during the study: the difficulty of detecting low-intensity covert attacks and the dependence of model performance on the quality of the training dataset. As a prospective direction, integrating the model with recurrent neural networks such as LSTM (Long Short-Term Memory) is planned to improve its ability to detect complex cyber threats with temporal dependencies and to adapt the algorithm for IoT (Internet of Things) ecosystems.

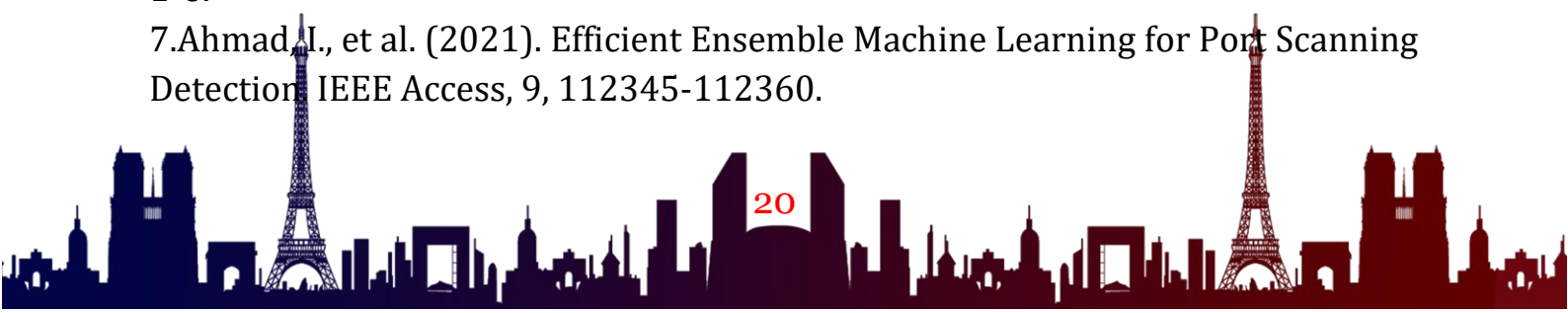
Conclusion

The conducted research has shown that traditional network security tools fail to achieve the expected effectiveness in detecting covert and low-intensity port scanning attacks. This underscores the need for dynamic and adaptive intelligent approaches to counter modern cyber threats. The hybrid model, based on the synthesis of Shannon entropy and the Random Forest algorithm, serves to reduce system load and provide high-accuracy monitoring by progressively filtering network traffic.

Experimental results confirmed the model's stable performance and minimal error rate, providing a solid foundation for its integration into practical information security systems. This solution enables early detection of cyberattacks and enhances the overall stability of network infrastructure. In the future, it is advisable to further improve the model's intellectual capabilities using Deep Learning methods.

References:

- 1.Shannon, C. E. (1948). A Mathematical Theory of Communication.
- 2.Bell System Technical Journal, 27(3), 379-423. (Entropiya asoschisi).
- 3.Lashkari, A. H., et al. (2017). Toward Generating a New Dataset for Intrusion Detection System Framework. Proceedings of the 15th Annual Conference on Privacy, Security and Trust (PST), IEEE, 1-10. (CIC-IDS2017 dataseti mualliflari).
- 4.Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. (Algoritm asoschisi).
- 5.Scarfone, K., & Mell, P. (2018). Guide to Intrusion Detection and Prevention Systems (IDPS). NIST Special Publication 800-94, National Institute of Standards and Technology.
- 6.Tavallae, M., et al. (2009). A Detailed Analysis of the KDD CUP 99 Data Set. IEEE Symposium on Computational Intelligence for Security and Defense Applications, 1-6.
- 7.Ahmad, I., et al. (2021). Efficient Ensemble Machine Learning for Port Scanning Detection. IEEE Access, 9, 112345-112360.



- 8.Yusupov, Sh. R. (2022). Kiberxavfsizlikda anomaliyalarni aniqlashning intellektual usullari. O'zbekiston Milliy universiteti xabarлари, 2(1), 145-152.
- 9.Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.
- 10.V. Paxson. (1999). Bro: A System for Detecting Network Intruders in Real-Time. Computer Networks, 31(23-24), 2435-2463.
- 11.Zheng, D., et al. (2020). Port Scan Detection Based on Evidence Theory and Entropy. Future Internet, 12(10), 164