STATISTICAL EVALUATION OF NETWORK TRAFFIC VARIATIONS USING HYPOTHESIS TESTING METHODS

Orazdurdyyeva G. O., Bekiyeva M. B., Bekiyev A. R.
Department of Computer Sciences and Information Technologies,
Oguz han Engineering and Technology University of Turkmenistan
Department of Applied Mathematics and Informatics Oguz han Engineering and Technology University of Turkmenistan
Faculty of Engineering and Economics,
Belarusian State University of Informatics and Radioelectronics
Ashgabat, Turkmenistan; Minsk,Belarus
E-mail: gulshatorazdurdyyewa3@gmailcom, maral.bekiyeva@etut.edu.tm, successbmb@gmail.com

This paper presents a statistical approach for detecting and analyzing variations in network traffic using hypothesis testing methods such as the z-test, t-test, and chi-square test. The study aims to determine whether changes in traffic behavior are statistically significant and could indicate potential cyberattacks or anomalies. The proposed approach provides an interpretable, mathematically grounded framework for network monitoring without relying on complex machine learning algorithms. Experimental results demonstrate that statistical hypothesis testing can effectively differentiate normal traffic from abnormal or attack traffic, thereby contributing to improved cybersecurity analysis.

Introduction

The exponential increase in online services, cloud computing, and digital transactions has made network security one of the most critical challenges of the modern era. Network traffic often exhibits variations in volume, frequency, and packet distribution due to user activity or external interference. However, some of these variations may signify malicious activities such as Denial of Service (DoS) attacks, port scanning, or data exfiltration.

Traditional cybersecurity systems rely heavily on machine learning or signature-based detection, which may not always detect new or evolving threats. Therefore, there is a growing need for statistical methods capable of identifying anomalies in network behavior based on measurable, quantifiable differences.

This paper focuses on the use of statistical hypothesis testing methods—specifically, the z-test, t-test, and chi-square test—to evaluate whether observed changes in network traffic are statistically significant. Unlike heuristic or AI-based methods, this approach is grounded in mathematical rigor and can be easily implemented using standard statistical tools. The main objective is to provide a reliable and explainable framework for anomaly detection that enhances the overall cybersecurity infrastructure.[1-3]

I. Network Traffic Analysis

Network traffic represents the continuous flow of data packets transmitted across network devices such as routers, switches, and servers. Each packet contains information including its size, transmission rate, source and destination addresses, and protocol type. Under normal conditions, network traffic exhibits consistent and predictable statistical characteristics – for example, stable average packet sizes and steady transmission frequencies.

However, during malicious activities such as denial-of-service (DoS) attacks, port scans, or intrusion attempts, these characteristics deviate from their expected ranges. Detecting these variations through statistical evaluation allows for the identification of anomalies at an early stage, thereby enhancing cybersecurity defense mechanisms.[1,3]

II. Statistical Hypothesis Testing

Statistical hypothesis testing provides a formal framework to determine whether an observed pattern in data significantly differs from an expected norm. In the context of network security, it is used to evaluate whether current traffic behavior deviates from the baseline of normal activity. The process involves two competing hypotheses:

- Null hypothesis (H_0): There is no statistically significant difference between normal and current network traffic patterns.;
- Alternative hypothesis (H_1): There exists a statistically significant difference, suggesting potential anomalous or malicious activity.

A test statistic is calculated based on the traffic data, and its corresponding p-value is compared to a chosen significance level (commonly $\alpha = 0.05$). If the p-value is smaller than α, the null hypothesis is rejected, indicating that the difference between traffic samples is not due to random variation but rather a meaningful change – possibly related to a cyberattack.[5]

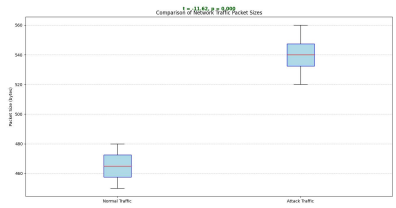

Figure 1 – Comparison on Network Traffic Packet Size

This example demonstrates how a t-test can identify statistically significant differences in average packet size between normal and attack traffic.

III. Tests Used

Different statistical tests are employed depending on the nature and size of the data:

Table 1 – Summarizes the statistical tests

Test Type	Sample Size Condition	Data Type	Purpose in Traffic Analysis
Z-test	Large sample (n > 30), known variance	Continuous	Detects deviation in mean packet size or transmission rate
T-test	Small sample (n ≥ 30), unknown variance	Continuous	Compares means of normal and suspected traffic
Chi-square Test	Any sample size	Categorical	Compares frequency of packet types or IP categories

summarizes the statistical tests applied for evaluating variations in network traffic[3]

IV. Methodology

1. Data Collection. Network traffic data was collected from publicly available datasets, including NSL-KDD and UNSW-NB15, which provide both normal and attack traffic samples. These datasets contain a variety of attack types such as Denial of Service (DoS), Probe, and User to Root (U2R), allowing for

2. Feature Selection. The following parameters were selected as indicators of network behavior:

Table 2 – indicators of network

Feature	Description	Purpose
Packet Size	Number of bytes per packet	Detect anomalies in payload volume
Packet Rate	Number of packets per unit time	Identify traffic spikes
Protocol Type	TCP, UDP, ICMP, etc.	Observe changes in protocol distribution
Connection Duration	Time of each session	Capture unusual session behavior

3. Data Preprocessing

1. Normalization – Traffic parameters were scaled to ensure comparability and reduce bias from extreme values.

2. Grouping – Data was divided into two sets: Normal traffic Potential attack traffic

3. Statistical Testing. Different statistical tests were applied to identify anomalies in network traffic: Z-test: Compares average packet sizes between normal and abnormal traffic. Suitable for large sample sizes with known variance.T-test: Examines differences in packet transmission rates between groups with unknown variance or smaller sample sizes.Chi-square test: Evaluates deviations in categorical data, such as protocol type or source IP frequency.[3]

4. Decision Process. The results of the statistical tests are interpreted using ρ-values: ρ-value $<$ 0.05: The null hypothesis is rejected; traffic variation is statistically significant, indicating potential cyber threats and ρ-value $\geq$ 0.05: The null hypothesis is not rejected; observed variations are likely due to random fluctuation.[4]

V. Conclusion

This study demonstrates that statistical hypothesis testing methods can be effectively applied to evaluate variations in network traffic and detect anomalies that may indicate cyberattacks. The proposed framework is mathematically rigorous, transparent, and computationally efficient. Unlike machine learning models, it does not require extensive training data or complex tuning, making it suitable for lightweight intrusion detection systems. Future research may extend this approach by integrating Bayesian inference or combining statistical testing with AI-based classifiers to achieve even higher detection accuracy.

1. Lippmann, R., Haines, J., Fried, D., Korba, J., & Das, K. The 1999 DARPA off-line intrusion detection evaluation / R. Lippmann, J. Haines, D. Fried, J. Korba, K. Das // Computer Networks. – 2000. – Vol. 34, № 4. – P. 579–595. – DOI: https://doi.org/10.1016/S1389-1286(00)00075-X.
2. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. A detailed analysis of the KDD CUP 99 data set / M. Tavallaee, E. Bagheri, W. Lu, A. A. Ghorbani // Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. – 2009. – P. 1–6. – IEEE.
3. Moustafa, N., & Slay, J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) / N. Moustafa, J. Slay // 2015 Military Communications and Information Systems Conference (MilCIS). – 2015. – P. 1–6. – IEEE.
4. Chandola, V., Banerjee, A., & Kumar, V. Anomaly detection: A survey / V. Chandola, A. Banerjee, V. Kumar // ACM Computing Surveys. – 2009. – Vol. 41, № 3. – P. 1–58. – DOI: https://doi.org/10.1145/1541880.1541882.
5. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. Network anomaly detection: Methods, systems and tools / M. H. Bhuyan, D. K. Bhattacharyya, J. K. Kalita // IEEE Communications Surveys & Tutorials. – 2014. – Vol. 16, № 1. – P. 303–336. – DOI: https://doi.org/10.1109/SURV.2013.052213.00146.