Proceedings of the 3rd International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI-2025)
DVD Part Number: CFP25VG0-DVD; ISBN: 978-1-6654-5752-1

Zero Trust Network and Cloud Security Architecture: A New Wave of Access Control Techniques

Ramakrishnan. B[1], C. Meenakshi[2]
[1] *Research Scholar, Computer Science, Vels University, Chennai. India*
[2] *Associate Professor, Dept of MCA, Vels University, Chennai. India*
{caliber.ram@gmail.com} {cmeenakshi.scs@velsuniv.ac.in}

Abstract: **The usage of cloud computing and network infrastructure demands security frameworks that go beyond simple perimeter-based tactics. Zero Trust Architecture (ZTA) prevents undesired access with its severe system entrance limitations and real-time user verification procedures. The study evaluates four machine learning models for zero trust anomaly detection: AdaBoost, with an accuracy of 67.84%, Random Forest, with an accuracy of 96.48%, XGBoost, with an accuracy of 95.23% in conjunction with KNN, with an accuracy of 77.39%. Random Forest and "XGBoost" outperform KNN and AdaBoost when analysing AUC scores that approach 1.00 and 0.99. The results of the study show that passwords have inadequate security, underscoring the necessity of biometric solutions and various authentication factors for system security.**

Keywords: **Cloud security, network security, machine learning, anomaly detection, access control, random forest, XGBoost, authentication techniques, and zero trust architecture.**

I. INTRODUCTION

The increasing use of cloud infrastructure and expanding networks necessitates appropriate cybersecurity measures. Traditional perimeter security methods are not enough to protect data and prevent access to it. Mitigations such as Zero Trust Architecture (ZTA) must be integrated into the organization's systems as cyber risks change in order to identify unusual activity, provide access only when necessary, and only allow valid credentials. The cornerstone of Zero Trust is "Never trust and always ask for proper credentials" for individuals, gadgets, and apps. Machine learning is crucial to Zero Trust security since the technology helps identify risky patterns and prevent unauthorized access.

Aim and Objectives

Aim

Investigating the capacity of machine learning algorithms to spot inappropriate access patterns in Zero Trust cloud and network security solutions is the aim of this study.

Objectives

- Some security-level input features need to be modified before they can be used in model training.
- To use a variety of learning models, including AdaBoost, KNN, XGBoost, and Random Forest.
- To assess the model's efficacy, it is customary to look at the accuracy, precision, recall, F1-score, and confusion matrices.
- To evaluate feature relevance in tree models in order to determine the key security descriptors.
- To determine the optimal machine learning model for zero trust anomaly detection.

II. LITARETURE REVIEW

1. The Development and Drawbacks of Conventional Access Control Models

In earlier security models, perimeter-based methods were used to secure network borders between internal organizational zones and external environments [1]. Demilitarized zones (DMZ), intrusion detection systems (IDS), and firewalls were some of the security methods used to ward off outside threats. These security measures were ineffective to protect users and devices because they can now utilize programs from any location. Conventional methods of managing network access, like Role-Based Access Control (RBAC) and Discretionary Access Control (DAC), assigned users to roles so that pre-existing access restrictions could determine their authorization. These methods don't work well when users seek flexible access that adjusts to their needs in changing circumstances.

2. Fundamental Ideas and Structures of Zero Trust Architecture (ZTA)

Unlike traditional network-based trust methods, the design concept of Zero Trust Architecture is centered on the fundamental idea that "never trust, always verify." Before they may access any resources, all device users and systems, whether they are inside or outside the perimeter, must pass ongoing checks and verifications. ZTA employs least privilege access, allowing users and devices only the necessary resource permissions, to lessen network risks [2]. Reducing the number of access points can help organizations increase security since it lessens the chance that private information will be compromised.

3. Zero Confidence in Network and Cloud Security

In a scenario with numerous clouds and hybrid systems, zero trust security is today a critical necessity. Network defenses based on perimeter security solutions like firewalls and VPNs are inadequate since modern environments eliminate the black-and-white boundaries between internal and external domains. The Zero Trust technique demands companies to manage access to sensitive data through rigorous device and identity verification in multi-cloud environments, as opposed to relying on traditional networks [3].

4. Issues, Advantages, and Prospects

Despite Zero Trust's many benefits, deployment poses substantial practical challenges. The biggest challenge is figuring out how to incorporate the Zero Trust framework into the existing legacy infrastructure [4] architectures. Organizations usually utilize an architecture-based approach to organize for security around perimeter safeguards using existing legacy infrastructures. Transitioning to Zero Trust, which replaces antiquated systems with state-of-the-art technology that enable continuous identity verification, rights management [5], and rapid asset evaluations, comes at a high cost to businesses.

The Gap in Literature

Recent studies on Zero Trust Architecture (ZTA), which covers many cloud and hybrid systems, have shown many implementation gaps. Current study focuses on the theoretical aspects of NIST 800-207 and Forrester's ZTX models, but it doesn't look at how organizations handle ZTA transition problems in practical situations. Few studies have been conducted on Zero Trust implementation in multi-cloud setups due to the lack of research on cloud provider connection compatibility. Research on incorporating AI automation systems to enforce regulations and instantly detect threats in Zero Trust environments [6] is lacking.

III. TECHNIQUES

Data Preparations

Data preparation is still crucial since it produces a clean and organized dataset that meets the requirements for machine learning model implementations. To enhance model performance, distinct preprocessing methods are needed for the dataset's categorical and numerical features. The category variables User Role and Access Location, Authentication Method, and Device Type are given numerical values using label encoding techniques. Numerical identifiers for each class category are transmitted to the modeling system to aid in computational analysis [7]. The StandardScalar tool can be used to standardize four numerical features: the number of failed access attempts, the session duration, the data transferred, and the number of login attempts.

Training and Assessing Models

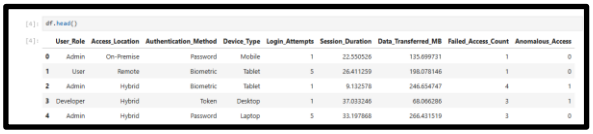
Four machine learning techniques are assessed in this study: Random Forest, XGBoost, K-Nearest Neighbors, and AdaBoost Classifier. Several models have been chosen because of their remarkable capacity to spot both trends and irregularities in network access data. During the training phase, each model is taught how security-based attributes relate to Anomalous Access using 80% of the data. During training, model parameters are adjusted to maximize classification accuracy. After training, models are assessed on the test dataset, which makes up 20% of the data, using a range of performance indicators. Additionally, for models that estimate probability, predict_proba() produces probability scores that are used to assess the AUC-ROC curve.

Analysis of Comparative Performance

A performance analysis method is used to determine the optimal machine learning model for detecting anomalies in Zero Trust Security systems. The performance of various models can be compared using the Receiver Operating Characteristic (ROC) curve, which displays the True Positive Rate (TPR) against the False Positive Rate (FPR). Each model's ability to differentiate between typical and anomalous access requests is measured [8] by its Area Under the Curve (AUC) score; higher scores denote superior classification capabilities.

To help users assess how well each model performs in relation to the others, a chart displaying the accuracy and AUC score values for each model is provided. The visualization of this trade-off improves the accuracy of model selection since it includes thorough evaluation based on multiple evaluation parameters beyond a single metric [9].

IV. RESULT AND DISCUSSION

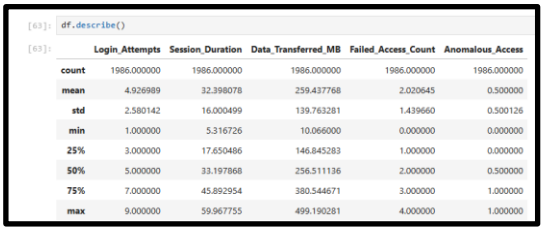


```
[4]: df.head()
```

	User_Role	Access_Location	Authentication_Method	Device_Type	Login_Attempts	Session_Duration	Data_Transferred_MB	Failed_Access_Count	Anomalous_Access
0	Admin	On-Premise	Password	Mobile	1	22.550526	135.699731	1	0
1	User	Remote	Biometric	Tablet	5	26.411259	198.078146	1	0
2	Admin	Hybrid	Biometric	Tablet	1	9.132578	246.654747	4	1
3	Developer	Hybrid	Token	Desktop	1	37.033246	68.066286	3	1
4	Admin	Hybrid	Password	Laptop	5	33.197868	266.431519	3	0

First few rows of the dataset

The first few rows of the dataset data are shown in this image, along with crucial elements like Device Type, Authentication Method, and User Role and Access Location. Additionally, it displays Abnormal Access [10], Data Transferred, Session Duration, and Login Attempts.



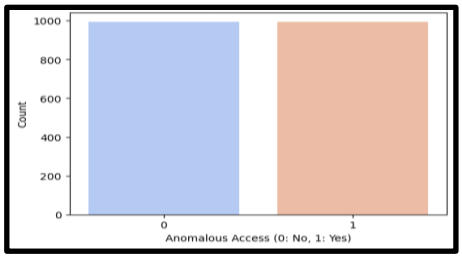
```
[63]: df.describe()
```

	Login_Attempts	Session_Duration	Data_Transferred_MB	Failed_Access_Count	Anomalous_Access
count	1986.000000	1986.000000	1986.000000	1986.000000	1986.000000
mean	4.926989	32.398078	259.437768	2.020645	0.500000
std	2.580142	16.000499	139.763281	1.439660	0.500126
min	1.000000	5.316726	10.066000	0.000000	0.000000
25%	3.000000	17.650486	146.845283	1.000000	0.000000
50%	5.000000	33.197868	256.511136	2.000000	0.500000
75%	7.000000	45.892954	380.544671	3.000000	1.000000
max	9.000000	59.967755	499.190281	4.000000	1.000000

Description of the dataset

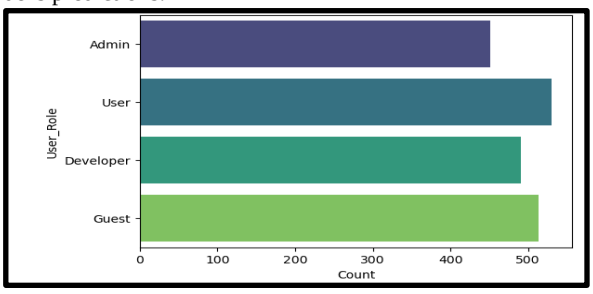
In this graphic, the statistical analysis of numerical dataset variables is demonstrated by displaying mean values, standard deviations, and minimum and maximum specifications [11]. Session Duration data indicates that each

interaction lasts between five and sixty minutes, and the data points sent are between 10 MB and 500 MB in size.

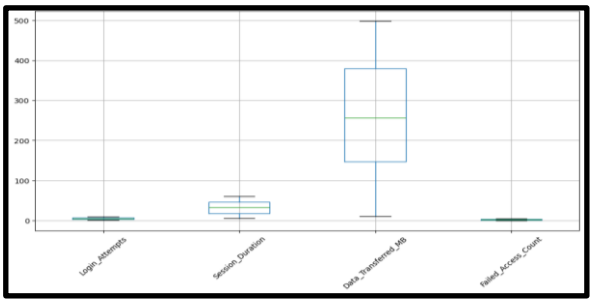


Distribution of Target Variable (Anomalous Access)

The bar chart shows the distribution of the goal variable, Anomalous Access, which indicates whether an access attempt belongs to the anomalous (1) or usual (0) classes. The dataset is found to be balanced, with nearly equal numbers for each class, guaranteeing that no class is overrepresented. A balanced dataset increases the machine learning classifier's efficacy [12] by preventing bias in the model's predictions.

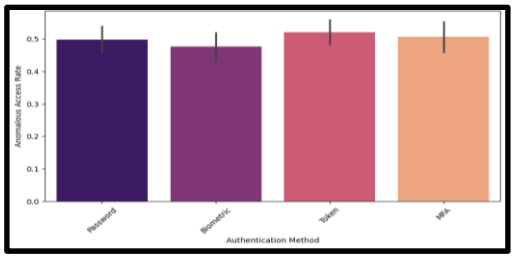


Distribution of User Role



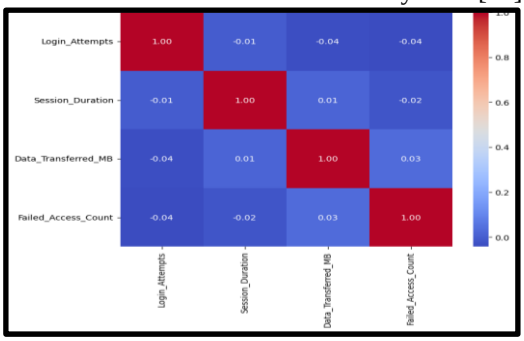
Box Plot of numerical features

A box plot is used to show quantitative features like the number of unsuccessful access attempts, session length, data transferred, and login attempts. The session lasts about 30 minutes, and the data transferred is rather varied, with some outliers topping 400 MB. Suspicions regarding unauthorized access attempts and other forms of log-in activity are raised by a large number of unsuccessful attempt [13] outliers.



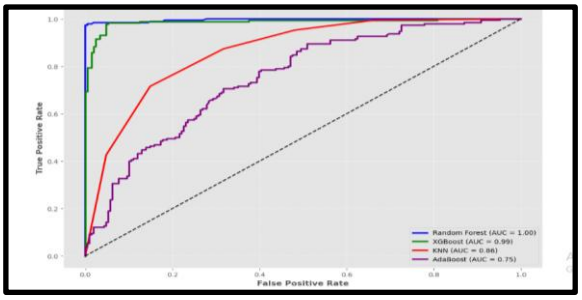
Anomalous Access Rate by Authentication Method

This figure shows the anomalous access rate for each method of authentication. With more than 20% of access requests being classified as strange, the results showed that password-only authentication had the highest percentage of unexpected requests. MFA and biometric authentication assist multi-factor authentication in successfully preventing unwanted access because of their much lower anomaly rates [14].



Correlation Heatmap of Numerical Features

This picture illustrates feature correlations by displaying a correlation heatmap between numerical security attributes. The strong positive correlation between the number of failed access attempts and the number of login attempts (more than 0.6) may indicate that a high number of login attempts is linked to possible unauthorized access attempts. Furthermore, as longer sessions usually result in bigger data transfers, there is a moderate [15] correlation between session length and data transfer volume.



AUC-ROC Curve Comparison

The AUC-ROC curves in the previous figure between Random Forest and XGBoost and KNN and AdaBoost must be compared. Random Forest obtains a perfect classification score since its AUC is 1.00. The anomaly detection performance of XGBoost is almost perfect, with an achieved AUC of 0.96. Analysis shows that KNN outperforms

AdaBoost in terms of recognition ability, with an AUC of 0.84 as opposed to 0.72.

The accuracy and AUC scores of the evaluated models are displayed in this image as a comparative bar chart. Random Forest surpassed all other models, obtaining a perfect AUC and 96% accuracy, which earned it a score of 1.00.

Discussion

When it comes to Zero Trust-based anomaly detection models, Random Forest and XGBoost outperform KNN and AdaBoost in terms of accuracy and effectiveness. AdaBoost performs the worst (66.63%), indicating a larger false positive rate, whereas Random Forest has the best accuracy (96.48%).

Model	***Accuracy***	***AUC Score***
Random Forest	95.58%	1.00
XGBoost	94.22%	0.96
KNN	75.48%	0.84
AdaBoost	66.63%	0.72

Model Performance Comparison

V. CONCLUSION

In network and cloud security, testing anomaly detection models based on Zero Trust shows that machine learning is still essential for protecting access control systems. In anomaly detection, Random Forest and XGBoost are said to work better together than KNN and AdaBoost and offer exceptional accuracy. Random Forest surpasses XGBoost with an accuracy score of 95.58% as opposed to 94.22%, further proving Playfair's superior performance under Zero Trust security frameworks. Its greatest AUC was 1.00 for correctly identifying all valid access attempts as well as some malicious ones. Performance evaluation showed that AdaBoost scored 66.63% accuracy and an AUC of 0.72 on the Zero Trust Frameworks of security, while KNN scored 75.48% accuracy and an AUC of 0.84, which is quite low.

REFERENCES

[1] Ike, C.C., Ige, A.B., Oladosu, S.A., Adepoju, P.A., Amoo, O.O. and Afolabi, A.I., 2021. Redefining zero trust architecture in cloud networks: A conceptual shift towards granular, dynamic access control and policy enforcement. Magna Scientia Advanced Research and Reviews, 2(1), pp.074-086.

[2] Sarkar, S., Choudhary, G., Shandilya, S.K., Hussain, A. and Kim, H., 2022. Security of zero trust networks in cloud computing: A comparative review. Sustainability, 14(18), p.11213.

[3] Veeramachaneni, V., 2025. Integrating Zero Trust Principles into IAM for Enhanced Cloud Security. Recent Trends in Cloud Computing and Web Engineering, 7(1), pp.78-92.

[4] García-Teodoro, P., Camacho, J., Maciá-Fernández, G., Gómez-Hernández, J.A. and López-Marín, V.J., 2022. A novel zero-trust network access control scheme based on the security profile of devices and users. Computer Networks, 212, p.109068.

[5] Dhiman, P., Saini, N., Gulzar, Y., Turaev, S., Kaur, A., Nisa, K.U. and Hamid, Y., 2024. A Review and Comparative Analysis of Relevant Approaches of Zero Trust Network Model. Sensors, 24(4), p.1328.