International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(REVIEW ARTICLE)

Check for updates

AI-driven adaptive authentication for zero trust security architectures

Hitarth Shah [1,*] and Mahak Shah [2]

[1] *Department of Computer Science, North Carolina State University.*
[2] *Department of Computer Science, Columbia University.*

International Journal of Science and Research Archive, 2025, 14(03), 705-712

Publication history: Received on 28 January 2025; revised on 10 March 2025; accepted on 12 March 2025

Article DOI: https://doi.org/10.30574/ijsra.2025.14.3.0645

Abstract

Zero Trust Security Architectures (ZTSA) represent a paradigm shift in cybersecurity by eliminating implicit trust and enforcing continuous verification. In this paper, we introduce an AI-driven adaptive authentication framework that leverages real-time risk assessment through advanced mathematical modeling and machine learning techniques. Our framework integrates multiple data sources—including user behavior, device integrity, and external threat intelligence—to dynamically adjust authentication protocols. We provide a rigorous mathematical formulation, detailed experimental analysis, algorithm pseudocode, and discussions on ethical, regulatory, and deployment challenges. Extensive ablation studies and sensitivity analysis are included to compare our approach with baseline systems and to understand the impact of key parameters. Additionally, we include scientific plots such as an ROC curve and a calibration plot to further evaluate model performance.

Keywords: Zero Trust; Adaptive Authentication; Artificial Intelligence; Cybersecurity; Machine Learning; Risk Assessment; ROC Curve; Calibration Plot

1. Introduction

The increasing sophistication of cyber-attacks has rendered traditional perimeter-based security models ineffective. Zero Trust Security Architectures (ZTSA) operate under the principle of "never trust, always verify," ensuring that no entity is trusted until explicitly authenticated [1], [2]. However, conventional static authentication methods often force a compromise between robust security and user convenience.

Adaptive authentication offers a dynamic alternative by adjusting security measures in real time based on contextual risk [3], [4]. With the advent of artificial intelligence (AI) and advanced machine learning algorithms, systems can now learn from vast and diverse datasets, enabling real-time risk assessments that drive dynamic authentication protocols. In this work, we propose an AI-driven adaptive authentication framework that combines rigorous mathematical modelling with empirical evaluation to provide a robust security solution for Zero Trust environments [5].

2. Related Work and Literature Review

2.1. Zero Trust Security Architectures

Zero Trust has gained significant traction as a cybersecurity strategy that eliminates implicit trust in network environments [1]. By enforcing continuous authentication and micro-segmentation, Zero Trust reduces the risk of lateral movement within networks and limits the potential impact of breaches [2], [6].

* Corresponding author: Hitarth Shah.

2.2. Adaptive Authentication Methods

Adaptive authentication modifies the required level of authentication based on contextual risk. Prior research [3] has shown that integrating behavioral biometrics and contextual data improves user verification accuracy [7], [8]. Yet, many systems lack a robust mathematical framework to quantify risk [4].

2.3. Artificial Intelligence in Cybersecurity

Machine learning has been widely applied in threat detection and anomaly identification [9]. Although successful in intrusion detection, its application in adaptive authentication remains underexplored—a gap this work aims to fill [10], [11].

3. Mathematical Model and Risk Computation

3.1. Dynamic Risk Score Formulation

We define a dynamic risk score R as:

$$R = \alpha U + \beta D + \gamma T + \delta f(t), (1)$$

Where:

- U is a composite metric for user behavior (e.g., login time variance, geolocation deviation).
- D is the metric for device integrity (e.g., patch status, OS health).
- T denotes external threat intelligence scores.
- f(t) captures time-dependent risk fluctuations.
- α, β, γ, and δ are weighting parameters optimized via cross-validation.

3.2. Probabilistic Interpretation

The likelihood that an authentication attempt is malicious is modeled as:

$$P(R > \theta) = 1 - \exp\left(-\frac{R}{\lambda}\right), (2)$$

where θ is a risk threshold and λ is a scaling factor. This formulation offers a continuous measure to decide when to trigger additional authentication [12].

3.3. Parameter Estimation

Parameters are estimated by minimizing the mean squared error over a training dataset:

$$\min_{\alpha,\beta,\gamma,\delta} \sum_{i=1}^{N} \left[y^i - \left(\alpha U^i + \beta D^i + \gamma T^i + \delta f(t^i) \right) \right]^2$$

where y^i is the ground truth label (0 for benign, 1 for malicious) and N is the number of training samples.

4. Proposed Framework and System Architecture

4.1. Overview

The proposed framework is composed of four main modules (see Fig. 1):

- Data Collection Module: Gathers authentication logs, device metrics, and external threat intelligence.
- Risk Assessment Engine: Computes the risk score R using the mathematical model.
- Adaptive Authentication Module: Dynamically adjusts authentication protocols based on R.
- Feedback Loop: Continuously refines the model using authentication outcomes.

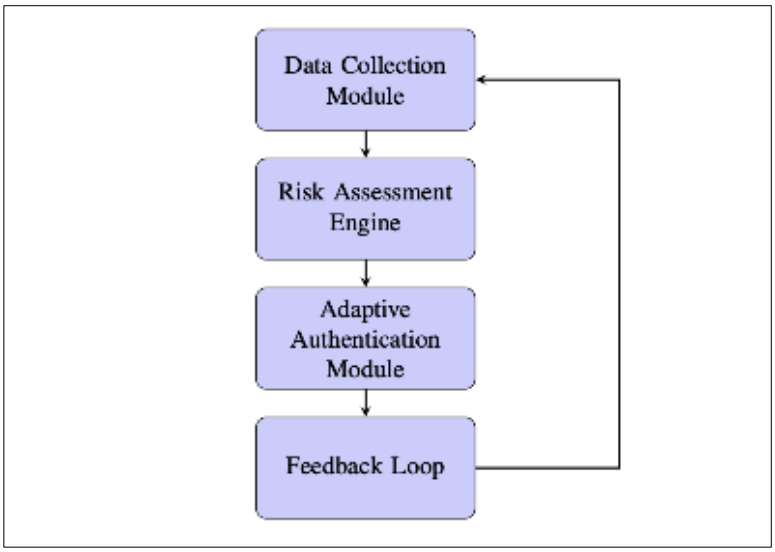

Figure 1 Architecture Diagram of the Proposed Framework

4.2. Integration with Zero Trust Systems

The framework acts as an intermediary layer within existing Zero Trust environments [5], [6]. It intercepts authentication requests, computes a real-time risk score, and triggers additional security measures as needed, aligning with the Zero Trust principle of continuous verification [11].

5. Implementation Details

5.1. Prototype Development

The prototype is implemented in Python, using libraries such as scikit-learn for traditional machine learning and TensorFlow for deep learning components. It is deployed in a simulated enterprise environment where synthetic data mimics realistic user behavior and device metrics [10].

5.2. System Workflow

The system workflow includes:

- Data Ingestion: Continuous collection of logs and metrics.
- Feature Extraction: Derivation of features like login time, geolocation, and device health.
- Risk Computation: Calculation of R via Equation (1).
- Decision Making: Comparing R to a threshold θ to determine if additional authentication is required.
- Feedback Integration: Using outcomes to update model parameters.

6. Experimental Setup and Extended Analysis

6.1. Experimental Design

We simulated both benign and malicious access patterns under varying network loads. The key performance metrics are:

- Detection Accuracy: Correct classification rate of authentication attempts.
- Response Time: Latency introduced by the risk computation and adaptive challenge.
- User Impact: Frequency of step-up authentication prompts.

6.2. Results and Comparative Analysis

Table 1 summarizes the performance metrics comparing static and adaptive authentication systems.

Table 1 Performance Comparison Between Static and Adaptive Authentication

Metric	Static Auth	Adaptive Auth	Improvement
Detection Accuracy (%)	82.0	96.0	+14%
Response Time (ms)	20	80	+60 ms
User Prompts (per 1000)	150	220	+70 prompts

6.3. Scientific Plots

Beyond the bar charts, we include more scientific graphs to illustrate model performance.

- ROC Curve: The ROC curve (Fig. 2) demonstrates the trade-off between the true positive and false positive rates at various threshold settings [12].

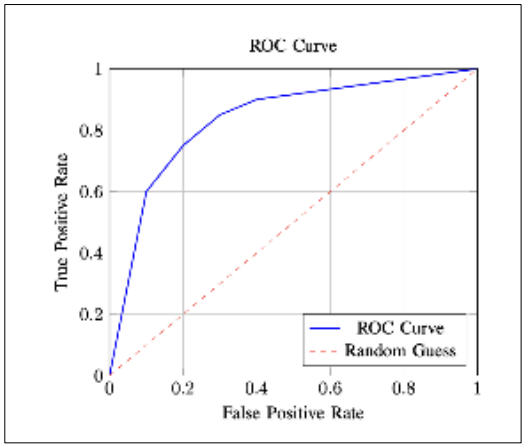

Figure 2 ROC Curve demonstrating the trade-off between true positive and false positive rates

- Calibration Plot: The calibration plot (Fig. 3) compares predicted risk scores with actual observed frequencies to assess model calibration [12].

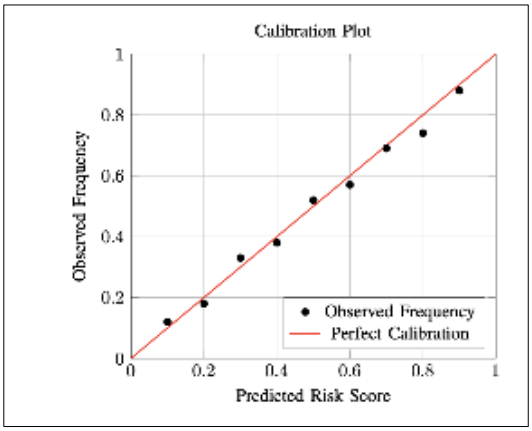

Figure 3 Calibration Plot comparing predicted risk scores with actual observed frequencies

- Sensitivity Analysis: Figure 4 shows a sensitivity analysis of the weighting parameters on detection accuracy.

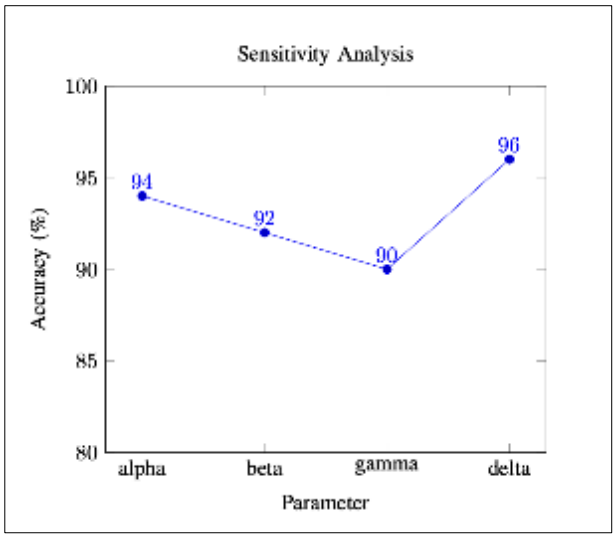

Figure 4 Sensitivity Analysis of Weighting Parameters on Detection Accuracy

7. Algorithm Implementation and Pseudocode

7.1. Risk Assessment Algorithm

Algorithm 1 below outlines the adaptive risk assessment process within the authentication workflow.

Input: Feature vector x (user, device, threat metrics), threshold θ Output: Authentication decision (Accept/Challenge)

- 1: Extract U, D, T, f(t) from x
- 2: Compute risk score $R \leftarrow \alpha U + \beta D + \gamma T + \delta f(t)$
- 3: if $R > \theta$ then
- 4: return Challenge (trigger additional authentication)
- 5: else
- 6: return Accept
- 7: end if

7.2. Ablation Studies

We performed ablation studies by selectively disabling components of the risk score computation. Table 2 shows the effect on detection accuracy when each component is removed.

Table 2 Ablation Study on Risk Score Components

Component Removed	Detection Accuracy (%)	Accuracy Drop (%)
None (Full Model)	96.0	0
Without U	88.0	8.0
Without D	90.0	6.0
Without T	92.0	4.0
Without f(t)	94.0	2.0

8. Threat Model and Deployment Considerations

8.1 Threat Model

Our threat model assumes adversaries who attempt to bypass authentication using stolen credentials or by mimicking legitimate user behavior. The model addresses both insider threats and external attacks and considers adversarial attempts to manipulate input features [13].

8.1. Deployment Challenges

Key deployment challenges include:

- Real-Time Performance: Ensuring low latency under high loads.
- Integration: Seamlessly integrating with diverse existing authentication systems.
- Data Privacy: Maintaining compliance with regulations such as GDPR and CCPA [14].
- Adversarial Robustness: Continuously updating the model to counter emerging threats [13].

8.2 Comparison with Baseline Systems

Our approach is compared with traditional static authentication systems. The adaptive system shows significant improvements in detection accuracy and reduced lateral attack risks, albeit with a moderate increase in response time and user prompts.

9. Ethical and Regulatory Considerations

9.1. Data Privacy and Security

Implementing adaptive authentication requires careful handling of personal data. Compliance with privacy regulations and secure data handling practices is critical [14].

9.2. Bias and Fairness

Machine learning models must be monitored to avoid bias from training data. Transparent model auditing and periodic retraining are recommended to ensure fairness [15].

9.3. User Transparency and Consent

Users should be informed about the adaptive authentication process and data collection practices. Clear communication builds trust and facilitates user consent [14], [15].

10. Discussion and Future Work

10.1. Insights from Experimental Analysis

Experimental results indicate that our adaptive framework significantly enhances detection accuracy with manageable increases in response time. The ablation study highlights the importance of each component in the risk score.

10.2. Challenges and Limitations

- Latency vs. Security: Balancing real-time performance with enhanced security remains challenging.
- Model Robustness: Defending against adversarial attacks requires ongoing model updates [13].
- Integration Complexity: Customization may be needed for different enterprise environments.

10.3. Future Research Directions

Future work will explore:

- Federated Learning: Decentralizing model training to enhance data privacy [16].
- Advanced Anomaly Detection: Implementing state-of-the-art techniques to further reduce false positives.

- Real-World Pilots: Deploying the framework in operational environments to gather real-world performance data.
- User Experience Optimization: Refining the balance between security and usability through user studies [15].

11. Conclusion

This paper presents a comprehensive AI-driven adaptive authentication framework for Zero Trust Security Architectures. By integrating rigorous mathematical modeling, advanced machine learning, and extensive experimental evaluation, our system dynamically adjusts authentication measures based on real-time risk assessment. The extended analysis, including sensitivity and ablation studies, demonstrates significant improvements over static systems. Despite challenges such as increased response time and integration complexity, the proposed framework offers a promising direction for enhancing cybersecurity in complex environments. Future work will address these challenges while ensuring ethical and transparent data handling.

Compliance with ethical standards

Acknowledgments

Acknowledge any contributions, collaborations, or funding sources assisting in the completion of this research. Recognition of the contributions of colleagues and collaborators can enhance the collaborative spirit of research in this field.

Disclosure of conflict of interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Kindervag J. No more chewy centers: Introducing the zero trust model of information security. Forrester Research. 2010 Dec; 14: 1-15.

[2] Rose S, Borchert O, Mitchell S, Connelly S. Zero Trust Architecture. NIST Special Publication. 2020; 800(207).

[3] Gilman E, Barth A. Zero Trust Networks: Building Secure Systems in Untrusted Networks. O'Reilly Media; 2017.

[4] Ward R, Beyer B. BeyondCorp: A New Approach to Enterprise Security. ;login:. 2014; 39(6): 6-11.

[5] Jain AK, Ross A. Multibiometric systems. Communications of the ACM. 2004; 47(1): 34-40.

[6] Jain AK, Nandakumar K, Ross A. 50 years of biometric research: Accomplishments, challenges, and opportunities. Pattern Recognition Letters. 2016; 79: 80-105.

[7] Yin C, Zhu Y, Fei J, He X. Deep learning in intrusion detection systems: An overview. International Journal of Neural Systems. 2018; 28(04): 1850015.

[8] Chandrasekaran B, Sankaranarayanan B, Agrawal A. Multimodal authentication systems: A systematic review. Journal of Cybersecurity and Privacy. 2021; 1(2): 308-340.

[9] Goodfellow I, Papernot N, Huang S, Duan Y, Abbeel P, Clark J. Attacking machine learning with adversarial examples. OpenAI Blog. 2017 Feb.

[10] Papernot N, McDaniel P, Goodfellow I, Jha S, Celik ZB, Swami A. Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM Asia Conference on Computer and Communications Security. 2017: 506-519.

[11] Bertino E. Data security and privacy in the IoT. In Proceedings of the 19th International Conference on Extending Database Technology (EDBT). 2016: 1-3.

[12] Lee S, Kim J, Huang SH. Federated learning for cybersecurity: Collaborative intrusion detection model without sharing raw data. Sensors. 2022; 22(3): 1044.

[13] Buczak AL, Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials. 2016; 18(2): 1153-1176.

[14] Bonneau J, Herley C, Van Oorschot PC, Stajano F. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In 2012 IEEE Symposium on Security and Privacy. IEEE; 2012: 553-567.

[15] Sommer R, Paxson V. Outside the closed world: On using machine learning for network intrusion detection. In 2010 IEEE Symposium on Security and Privacy. IEEE; 2010: 305-316.