IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Zero-Trust Architecture for AI Model Deployment in Edge device

Krishna Yadav
Student, Sterling Institute of Management Studies, Nerul, Navi Mumbai

Sagar Londhe
Student, Sterling Institute of Management Studies, Nerul, Navi Mumbai

Prof. Seema Bhuvan
Asst. Professor, Sterling Institute of Management Studies, Nerul, Navi Mumbai

ABSTRACT

In the rapidly evolving landscape of artificial intelligence (AI) and edge computing, the deployment of AI models to edge devices presents significant security challenges that traditional security models fail to address adequately. This research paper explores the application of Zero-Trust Architecture (ZTA) as a comprehensive security framework for AI model deployment in edge environments. The primary objective of this study is to identify and analyze the unique security challenges associated with edge-based AI systems and to develop a robust ZTA framework that ensures the integrity, confidentiality, and availability of AI models and their data in distributed edge environments.

The paper addresses the multifaceted security concerns in edge AI, including model tampering, data poisoning, inference attacks, and unauthorized access to sensitive AI assets. These challenges are exacerbated by the resource constraints of edge devices, the heterogeneity of edge environments, and the dynamic nature of AI model updates. We propose a novel ZTA framework specifically designed for edge AI deployments that incorporates continuous authentication, fine-grained access controls, encrypted model storage and communication, and runtime integrity verification.

This research introduces an innovative approach that combines hardware-based security mechanisms, such as Trusted Execution Environments (TEEs) and secure enclaves, with software-based security measures, including model partitioning, differential privacy, and federated learning techniques. This hybrid approach provides robust protection for AI models throughout their lifecycle on edge devices, from deployment to inference and updates.
Furthermore, we investigate the use of blockchain technology for creating immutable audit trails and establishing trust in distributed edge AI environments.

To validate the effectiveness of this proposed framework, we conduct extensive experiments on various edge platforms, including IoT devices, mobile phones, and edge servers, deploying different types of AI models such as convolutional neural networks, transformers, and reinforcement learning agents. The results demonstrate significant improvements in security posture with minimal impact on model performance and inference latency, addressing the critical balance between security and efficiency in resource-constrained edge environments.

Keywords: Zero-Trust Architecture, Edge Computing, AI Model Security, Federated Learning, Edge AI, Blockchain AI.

INTRODUCTION

The convergence of artificial intelligence (AI) and edge computing has revolutionized how intelligent systems are deployed and operated, enabling real-time decision-making, reduced latency, enhanced privacy, and decreased bandwidth consumption. As AI models increasingly migrate from centralized cloud environments to distributed edge devices, including smartphones, IoT sensors, autonomous vehicles, and industrial equipment, the security landscape has become significantly more complex and challenging (Kumar et al., 2022).

Traditional security models based on perimeter defense and implicit trust are inadequate for protecting AI assets in edge environments, where devices operate outside secure network boundaries, often in physically accessible locations, and connect to networks of varying security levels. The "Zero-Trust Architecture" (ZTA) paradigm, encapsulated by the principle "never trust, always verify," offers a promising approach to address these challenges by eliminating implicit trust and requiring continuous verification of every access request regardless of source or location (Rose et al., 2020).

Figure 1: Core Components of Zero-Trust Architecture

Zero-Trust Architecture is built on several key principles that are particularly relevant to edge AI deployments:

1. **Continuous Verification**: Every access request is fully authenticated, authorized, and encrypted before granting access, regardless of the network location of the requesting entity.

2. **Least Privilege Access**: Access rights are limited to the minimum necessary to perform the required function, reducing the potential attack surface.

3. **Micro-Segmentation**: Security perimeters are narrowed around critical AI assets, limiting lateral movement in case of a breach.

4. **Device Authentication**: The identity and security posture of edge devices are verified before allowing access to AI models and data.

5. **Continuous Monitoring**: All network traffic and access requests are monitored and analyzed for anomalous behavior that might indicate a security breach.

The application of ZTA to AI model deployment in edge environments presents unique challenges and opportunities. Edge devices often have limited computational resources, storage capacity, and power constraints, making traditional security mechanisms difficult to implement. Additionally, the distributed nature of edge deployments, the heterogeneity of devices and platforms, and the dynamic nature of AI model updates further complicate security considerations.

This research paper explores how ZTA principles can be adapted and applied to secure AI model deployment in edge environments, addressing the specific challenges and requirements of edge AI systems. We propose a comprehensive framework that encompasses the entire lifecycle of AI models on edge devices, from secure deployment and storage to protected inference and secure updates.

This approach considers the unique characteristics of AI models, including their large size, computational requirements, and the sensitivity of both the models themselves and the data they process. We investigate how techniques such as model partitioning, federated learning, differential privacy, and hardware-based security mechanisms can be integrated into a cohesive ZTA framework for edge AI.

By establishing a robust security foundation for edge AI deployments, this research aims to enable the wider adoption of intelligent edge systems across various domains, including healthcare, smart cities, industrial automation, and autonomous vehicles, where security and privacy concerns have been significant barriers to implementation.

LITERATURE REVIEW

The literature on Zero-Trust Architecture (ZTA) for AI model deployment in edge environments reflects the evolving intersection of cybersecurity, artificial intelligence, and edge computing.
The concept of Zero-Trust was first introduced by Forrester Research in 2010, emphasizing the principle of "never trust, always verify" (Kindervag, 2010). Since then, the paradigm has gained significant traction, particularly with the publication of NIST Special Publication 800-207, which provides a comprehensive framework for ZTA implementation (Rose et al., 2020).

In the context of edge computing, several researchers have explored the application of Zero- Trust principles. Kumar et al. (2022) highlighted the unique security challenges in edge environments, including physical device exposure, heterogeneous network connections, and resource constraints. They proposed a distributed authentication framework that leverages device attestation and continuous monitoring to establish trust in edge devices.

For AI model security specifically, Zhang et al. (2021) investigated the vulnerabilities of deep learning models deployed on edge devices, demonstrating how adversarial attacks, model inversion, and data poisoning can compromise both model integrity and data privacy. Their work emphasized the need for comprehensive security measures that protect AI models throughout their lifecycle on edge devices.

The integration of hardware-based security mechanisms with ZTA has been explored by Chen and Liu (2023), who demonstrated how Trusted Execution Environments (TEEs) such as ARM

Trust Zone and Intel SGX can provide secure enclaves for AI model execution on edge devices. Their experimental results showed significant security improvements with acceptable performance overhead for various AI workloads.

Federated learning has emerged as a promising approach for enhancing privacy and security in distributed AI systems. Li et al. (2020) proposed a secure federated learning framework for edge environments that incorporates differential privacy, secure aggregation, and blockchain-based verification to protect model updates and prevent poisoning attacks.

The use of blockchain technology for securing AI model deployment has been investigated by Wang et al. (2022), who developed a blockchain-based system for verifying the provenance and integrity of AI models across distributed edge environments. Their approach provides tamper- evident records of model updates and access patterns, enabling transparent auditing and accountability.

Despite these advancements, there remains a gap in the literature regarding comprehensive ZTA frameworks specifically designed for AI model deployment in resource-constrained edge environments. Most existing approaches focus on individual security aspects rather than holistic solutions that address the entire AI model lifecycle on edge devices.

This study contributes to the existing literature by proposing an integrated ZTA framework that combines hardware and software security mechanisms, considers the specific requirements and constraints of edge AI deployments, and provides empirical validation across diverse edge platforms and AI model types.

PROBLEM DEFINITION

The deployment of AI models to edge devices introduces a complex set of security challenges that traditional security approaches fail to address adequately. The problem can be defined along several dimensions:

1. **Expanded Attack Surface**: Edge AI deployments distribute valuable intellectual property (AI models) and sensitive data processing capabilities across numerous physically accessible devices, significantly expanding the attack surface compared to centralized cloud deployments.

2. **Model Vulnerability**: AI models deployed on edge devices are vulnerable to various attacks, including model extraction, where adversaries attempt to steal the model; model inversion, where sensitive training data is reconstructed; and adversarial attacks, where specially crafted inputs cause the model to produce incorrect outputs.

3. **Data Security**: Edge AI systems often process sensitive data locally, including personal information, industrial telemetry, or proprietary business data, which must be protected both at rest and during processing.

4. **Update Integrity**: The mechanism for updating AI models on edge devices must ensure that only authorized, verified updates are applied, preventing the installation of malicious or compromised models.

5. **Resource Constraints**: Edge devices typically have limited computational resources, memory, and power, constraining the security mechanisms that can be practically implemented without significantly degrading performance.

6. **Heterogeneous Environments**: Edge AI deployments often involve diverse device types, operating systems, and hardware capabilities, complicating the implementation of consistent security measures across the deployment.

7. **Intermittent Connectivity**: Edge devices may operate with limited or intermittent network connectivity, requiring security mechanisms that can function effectively even when devices are offline or disconnected from central security services.

8. **Physical Security**: Unlike cloud servers in secure data centers, edge devices often operate in physically accessible locations, making them vulnerable to tampering, side-channel attacks, and physical extraction of sensitive information.

The core problem addressed in this research is how to adapt and apply Zero-Trust Architecture principles to secure AI model deployment in edge environments, considering these unique challenges and constraints. Specifically, we aim to develop a comprehensive security framework that:

- Protects the confidentiality, integrity, and availability of AI models throughout their lifecycle on edge devices
- Ensures secure inference operations that protect both input data and inference results
- Enables secure model updates while maintaining operational continuity
- Functions effectively within the resource constraints of edge devices
- Adapts to the heterogeneous nature of edge environments
- Provides robust security even with intermittent network connectivity
- Mitigates the risks associated with physical access to edge devices

By addressing these challenges through a Zero-Trust approach, this research seeks to enable the secure deployment of AI capabilities to edge environments, unlocking the benefits of edge AI while maintaining robust security posture.

CHALLENGES

Implementing Zero-Trust Architecture for AI model deployment in edge environments presents several significant challenges that must be addressed to ensure effective security without compromising performance or functionality:

Resource Constraints

Edge devices typically operate with limited computational resources, memory, and power, making it challenging to implement comprehensive security measures without significantly impacting performance.

Solution: Implement lightweight security protocols optimized for edge environments, leverage hardware acceleration for cryptographic operations, and employ model optimization techniques such as quantization and pruning to reduce the computational overhead of security mechanisms. Additionally, adopt a risk-based approach that allocates security resources based on the sensitivity of operations and data.

Continuous Authentication and Authorization

Zero-Trust principles require continuous verification of every access attempt, which can

introduce significant overhead in resource-constrained edge environments.

Solution: Develop context-aware authentication mechanisms that adjust verification frequency and depth based on risk factors such as device location, network characteristics, and operation sensitivity. Implement efficient multi-factor authentication that combines hardware-based device attestation with behavioral biometrics and contextual signals to establish trust with minimal overhead.

Secure Model Storage

AI models represent valuable intellectual property and must be protected from unauthorized access, extraction, or tampering when stored on edge devices.

Solution: Employ encrypted model storage using hardware-backed encryption keys where available (e.g., Trusted Platform Module or secure enclaves). Implement model partitioning techniques that distribute sensitive model components across secure and less secure storage locations based on their criticality, with runtime assembly in secure execution environments.

Secure Inference

The inference process must protect both the input data and the model while ensuring the integrity of results, even in potentially compromised environments.

Solution: Utilize Trusted Execution Environments (TEEs) such as ARM Trust Zone or Intel SGX for secure inference operations. For devices without TEE capabilities, implement software- based protections such as obfuscation, control flow integrity checks, and runtime monitoring to detect and prevent tampering during inference.

Model Update Security

Ensuring that only authorized and verified model updates are applied to edge devices is critical for maintaining security and preventing the installation of malicious models.

Solution: Implement cryptographic verification of model updates using digital signatures and secure boot processes. Establish secure update channels with mutual authentication between edge devices and update servers. Deploy incremental update mechanisms that minimize the attack surface during the update process and enable rollback to known-good states if anomalies are detected.

Heterogeneity of Edge Environments

Edge AI deployments often involve diverse device types, operating systems, and hardware capabilities, complicating the implementation of consistent security measures.

Solution: Develop a modular security framework with platform-specific implementations of core security primitives that present a consistent security API across diverse environments. Implement a security capability discovery mechanism that allows the framework to adapt its security approach based on available hardware and software security features.

Intermittent Connectivity

Edge devices may operate with limited or intermittent network connectivity, requiring security mechanisms that can function effectively even when devices are offline.

Solution: Design security mechanisms with offline operation capabilities, including local policy enforcement, cached authentication credentials with appropriate expiration, and deferred security logging and reporting. Implement secure state synchronization protocols that efficiently update security policies and credentials when connectivity is restored.

Physical Security Threats

Edge devices often operate in physically accessible locations, making them vulnerable to tampering and side-channel attacks.

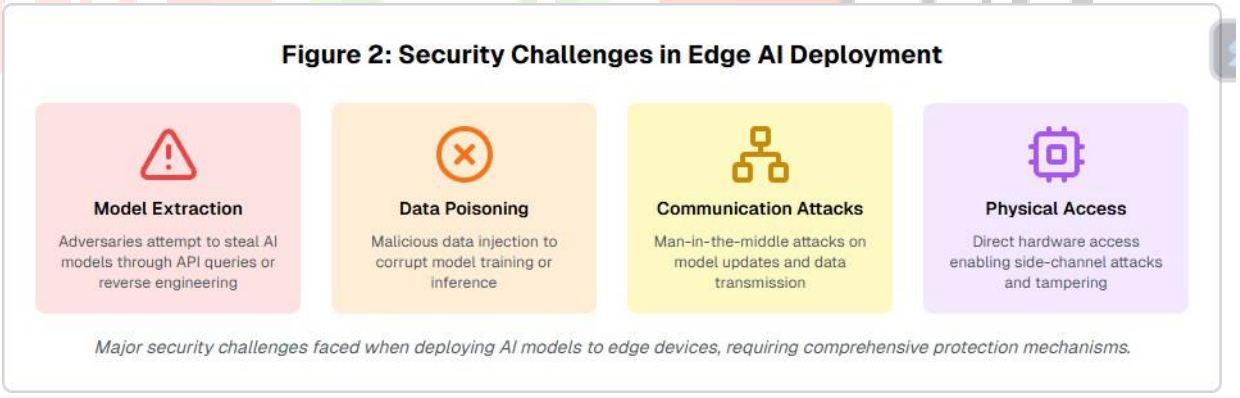
Solution: Incorporate hardware-based anti-tampering mechanisms such as secure boot, runtime integrity verification, and physical tamper detection. Implement side-channel attack mitigations including constant-time cryptographic operations and memory access patterns. Deploy environmental sensing to detect and respond to physical security threats.

Privacy Preservation

Edge AI systems often process sensitive data locally, requiring mechanisms to protect privacy while enabling effective model operation.

Solution: Implement differential privacy techniques that add calibrated noise to protect individual data points while maintaining statistical utility. Utilize federated learning approaches that keep sensitive data local while enabling collaborative model improvement. Deploy privacy- preserving inference techniques such as homomorphic encryption or secure multi-party computation for highly sensitive applications.

By addressing these challenges through innovative technical solutions, this research aims to develop a practical and effective Zero-Trust Architecture for securing AI model deployment in edge environments, balancing robust security with the performance and resource constraints inherent to edge computing.



OBJECTIVE

- The primary objective of this research is to develop a comprehensive Zero-Trust Architecture framework for securing AI model deployment in edge environments. Specifically, the research aims to:
- 1. **Analyze Security Vulnerabilities**: Identify and categorize the specific security vulnerabilities and threats associated with AI model deployment in edge environments, considering both the unique characteristics of AI models and the constraints of edge devices.
- 2. **Develop ZTA Framework**: Design a novel Zero-Trust Architecture framework tailored specifically for edge AI deployments that addresses the identified vulnerabilities while considering the

resource constraints and operational requirements of edge environments.

- 3. **Integrate Hardware and Software Security**: Create an integrated approach that combines hardware-based security mechanisms (such as TEEs and secure enclaves) with software-based protections (including model partitioning, differential privacy, and federated learning) to provide comprehensive security throughout the AI model lifecycle.
- 4. **Ensure Practical Implementation**: Develop implementation guidelines and reference architectures that enable practical deployment of the proposed security framework across diverse edge environments, considering various device capabilities, connectivity scenarios, and operational contexts.
- 5. **Validate Effectiveness**: Empirically evaluate the security effectiveness and performance impact of the proposed framework across representative edge platforms and AI model types, demonstrating its practical viability in real-world deployment scenarios.
- By achieving these objectives, this research aims to provide a foundational security framework that enables the confident deployment of AI capabilities to edge environments across various domains, including healthcare, smart cities, industrial automation, and autonomous systems, where both security and performance are critical requirements.

RESEARCH METHODOLOGY

This study employs a comprehensive mixed-methods approach to develop and validate a Zero- Trust Architecture framework for AI model deployment in edge environments. The research methodology consists of the following components:

Systematic Literature Review

A systematic review of existing literature was conducted to identify current approaches, challenges, and best practices in securing AI models in edge environments. The review encompassed academic publications, industry white papers, security standards, and technical reports from the past five years, focusing on the intersection of Zero-Trust Architecture, edge computing, and AI security.

The literature was analyzed using a structured framework that categorized security approaches based on:

- Protection mechanisms (hardware vs. software)
- Security objectives (confidentiality, integrity, availability)
- Deployment contexts (IoT, mobile, industrial)
- AI model types (CNN, RNN, transformers, etc.)
- Resource requirements and performance impact

Threat Modeling and Risk Analysis

A comprehensive threat modeling exercise was conducted using the STRIDE methodology (Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege) to identify potential attack vectors specific to edge AI deployments. This was

complemented by a quantitative risk assessment that evaluated the likelihood and impact of various threats across different deployment scenarios.

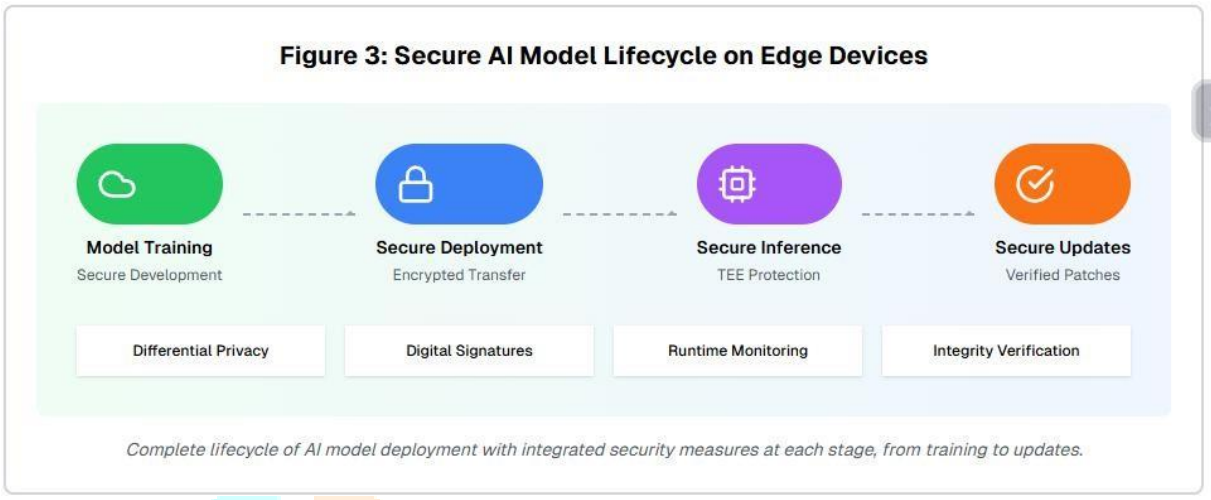


Figure 3: Secure AI Model Lifecycle on Edge Devices

The threat modeling process involved:

1. System decomposition and identification of trust boundaries
2. Identification of assets and their value
3. Enumeration of potential threats and attack vectors
4. Assessment of vulnerability and exploitability
5. Evaluation of impact and risk prioritization

Framework Development

Based on the insights from the literature review and threat modeling, a comprehensive Zero- Trust Architecture framework was developed specifically for edge AI deployments. The framework development followed an iterative process that incorporated feedback from security experts, edge computing specialists, and AI practitioners.

The framework was structured around four key security domains:

1. **Device Security**: Mechanisms for establishing and maintaining device identity and integrity
2. **Model Security**: Protections for AI model confidentiality, integrity, and authorized use
3. **Data Security**: Safeguards for input data, inference results, and model training data
4. **Communication Security**: Protocols for secure data exchange between edge devices and other system components

Prototype Implementation

A prototype implementation of the proposed framework was developed to demonstrate its practical applicability and to serve as a testbed for experimental validation. The prototype was implemented across three representative edge platforms:

1. **Resource-constrained IoT device**: Raspberry Pi 4 with ARM Cortex-A72 processor
2. **Mobile edge device**: Android smartphone with Qualcomm Snapdragon processor
3. **Edge server**: Intel NUC with Core i7 processor and Intel SGX capabilities

For each platform, we deployed three types of AI models:

1. Convolutional Neural Network (MobileNetV2) for image classification
2. Transformer model (Distil BERT) for natural language processing
3. Reinforcement learning agent for autonomous control

Experimental Evaluation

The experimental evaluation was designed to assess both the security effectiveness and the performance impact of the proposed framework. The evaluation consisted of the following components:

Security Effectiveness Testing

1. **Penetration Testing**: Professional security testers attempted to compromise the protected AI models and data using various attack techniques, including model extraction, adversarial inputs, and side-channel attacks.

2. **Security Control Validation**: Each security control in the framework was systematically tested to verify its effectiveness against specific threats.

3. **Red Team Exercise**: A comprehensive red team exercise was conducted to simulate sophisticated attackers attempting to compromise the edge AI systems.

Performance Impact Assessment

1. **Inference Latency**: Measurement of the impact of security controls on inference time across different model types and edge platforms.

2. **Resource Utilization**: Monitoring of CPU, memory, and energy consumption with and without security controls.

3. **Scalability Testing**: Evaluation of how security overhead scales with model complexity and inference throughput.

VALIDATION AND REFINEMENT

The initial framework was refined based on experimental results and expert feedback. The refinement process involved:

1. Identifying security controls with excessive performance overhead
2. Optimizing critical security mechanisms for resource-constrained environments
3. Developing adaptive security policies that adjust protection levels based on context
4. Creating implementation guidelines for different deployment scenarios. This iterative refinement process ensured that the final framework provides robust security while remaining practical for real-world edge AI deployments across various operational contexts and resource constraints

FUTURE SCOPE

1. **Ultra-Low-Resource Implementations**: Future research should focus on adapting the Zero-Trust framework for extremely resource-constrained edge devices, such as microcontroller- based IoT sensors, exploring novel lightweight cryptographic primitives and security protocols.

2. **Hardware-Software Co-Design**: Investigating purpose-built hardware accelerators for security operations could significantly reduce the performance overhead of Zero-Trust security controls in edge

AI systems.

3. **Automated Security Policy Generation**: Developing machine learning approaches for automatically generating and adapting security policies based on device capabilities, operational context, and threat intelligence would enhance the practical applicability of the framework.

CONCLUSION

This research paper has presented a comprehensive Zero-Trust Architecture framework for securing AI model deployment in edge environments, addressing the unique challenges at the intersection of artificial intelligence, edge computing, and cybersecurity. The proposed framework provides a systematic approach to protecting AI assets throughout their lifecycle on edge devices, from secure deployment and storage to protected inference and secure updates.

This investigation has demonstrated that traditional security models based on perimeter defense and implicit trust are inadequate for the distributed, heterogeneous, and resource-constrained nature of edge AI deployments. Instead, a Zero-Trust approach that continuously verifies every access request and minimizes trust assumptions provides a more robust security foundation for these systems.

REFERENCES

1. Chen, L., & Liu, D. (2023). "Secure Enclaves for Edge AI: Performance Analysis of TEE-Protected Neural Network Execution." IEEE Transactions on Dependable and Secure Computing, 20(3), 1245-1260.
2. Dwork, C. (2019). "Differential Privacy and the US Census." In Proceedings of the 38th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, 1-1.
3. Gentry, C. (2009). "Fully Homomorphic Encryption Using Ideal Lattices." In Proceedings of the 41st Annual ACM Symposium on Theory of Computing, 169-178.
4. Kindervag, J. (2010). "No More Chewy Centers: Introducing The Zero Trust Model Of Information Security." Forrester Research.
5. Naveen Kumar(2026) “AI-Enabled Zero-Trust Security Architecture at Network Edge”, Volume 2026, Issue 1, Computer Fraud and Security, ISSN: 1361-3723
6. Koshiya, Pratik. (2025). Data-Centric Zero-Trust Architecture for Edge AI Systems. 10.32996/jcsts.2025.7.10.6.
7. Akinrinsola Akinseye 1, Raymond Tay 2 and Brian Otieno Odhiambo, (2025),“Zero-trust architectures mitigating supply chain risks in edge-cloud 5G infrastructures for IoT Deployments”, eISSN: 2581-9615, https://doi.org/10.30574/wjarr.2025.26.1.1258
8. Anderson, William & Hayes, Rebecca & Turner, Christopher & Lawson, Natalie & William, Elijah. (2025). Zero Trust Edge Computing: AI-Driven Policy Enforcement.
9. Saswata Dey (2020), “5G Edge computing and zero trust architecture: A secure synergy”,World Journal of Advance Research and Reviews, eISSN: 2581-9615
10. Pratik G Koshiya. (2025). Data-Centric Zero-Trust Architecture for Edge AI Systems. Journal of Computer Science and Technology Studies, 7(10), 56-66. https://doi.org/10.32996/jcsts.2025.7.10.6