Immutable Adversarial Attack Evidence Collection for ML Cyber Defense using Blockchain

B. Ganesh[1], V. Karthik[2], M.S. Sushma Sailu[3], S. Chanukya[4], M. Kalyan Kishore[5]
Department of Computer Science & Engineering (AI & ML)
Avanthi Institute of Engineering & Technology, Vizianagaram, India
ganeshbheesetti9@gmail.com[1], vkarthik5102@email.com[2], sushmasailu2005@gmail.com[3],
chanukyasirisetty25@gmail.com[4], kalyankishore61@gmail.com[5]

Abstract

Advanced cybersecurity issues that modern banking applications face include SQL injection, cross-site scripting, adversarial machine learning attacks, and attempts to compromise credentials. Conventional logging methods undermine incident investigation and regulatory compliance due to mutability issues and a lack of forensic integrity. This study offers a novel architecture that creates immutable evidence collection for cyber defense by fusing blockchain technology with machine learning-based threat detection. The suggested system automatically logs attack evidence on a private blockchain ledger and uses Random Forest classifiers to detect malicious activity in real-time. The implementation makes use of MongoDB for operational data storage, Python based ML microservices, Node.js backend orchestration, and React.js frontend. Using cryptographic verification, smart contracts guarantee the persistence of unchangeable evidence. Experimental validation demonstrates 96% detection accuracy across various attack vectors with sub-second response latency. The blockchain-anchored evidence provides non-repudiable forensic trails suitable for audit and legal proceedings, addressing critical gaps in conventional security information management systems.

Index Terms—Adversarial machine learning, blockchain security, cyber defense, immutable logging, smart contracts, banking cybersecurity.

I. INTRODUCTION

The digital transformation of financial services has resulted in unprecedented operational efficiencies, but it has also exposed critical infrastructure to evolving cyber threats. Traditional web application exploits and complex adversarial manipulations of machine learning models used for fraud detection and authentication are just two of the many attack vectors that banking institutions must deal with.

Modern threat actors leverage automated tools and adversarial techniques to bypass conventional security controls. Despite decades of awareness, SQL injection and cross-site scripting are still common, and new adversarial machine learning attacks specifically target predictive models by creating inputs that cause misclassification without setting off conventional anomaly detection mechanisms.

Traditional security logging methods keep incident data in centralized databases that are vulnerable to manipulation by attackers or privileged users who gain access to the system. Forensic integrity is compromised by this mutability, which makes incident attribution, legal proceedings, and regulatory compliance difficult. Financial institutions need security architectures that preserve evidence with cryptographic guarantees of authenticity and immutability in addition to detecting sophisticated attacks.

Blockchain technology offers distributed ledger capabilities that ensure recorded data remains unalterable once committed. Security systems can accomplish both adaptive defense and forensically sound incident documentation by combining blockchain-based evidence storage with machine learning-driven threat detection.

This research addresses the critical gap between intelligent threat identification and trustworthy evidence preservation. We present a comprehensive architecture that automatically commits evidence to a permissioned blockchain via smart contracts, uses trained machine learning models for real-time attack classification, and keeps an eye on banking application activity. The system gives auditors and security analysts unchangeable incident records that can be used for legal proceedings, compliance checks, and investigations.

A. Research Contributions

The primary contributions of this work include:

1) Developing and putting into practice a hybrid blockchain-machine learning architecture for banking cybersecurity.
2) Creation of threat detection models that are adversarially aware and able to recognize both traditional web exploits and ML-targeted attacks.
3) Building a smart contract infrastructure to gather unchangeable evidence and verify it using cryptography.
4) Thorough system assessment exhibiting forensic integrity and real-time performance.

Architectural blueprints utilizing UML modeling for reproducibility and deployment guidance.

II. RELATED WORK

Cybersecurity research has extensively studied both blockchain implementations for audit integrity and machine learning applications for threat detection, despite the fact that few studies address their synergistic integration for banking environments.

A. Machine Learning in Cybersecurity

Using ensemble techniques, Bhuyan et al. demonstrated network anomaly detection with high accuracy on benchmark datasets, although they acknowledged difficulties with adversarial robustness [1]. In particular, Zhao et al. looked into adversarial attacks against banking fraud detection systems and found that production machine learning models were vulnerable to carefully constructed inputs [2].

B. Blockchain for Digital Forensics

In their investigation of blockchain applications for preserving evidence integrity in digital forensics, Kumar et al. emphasized the benefits of tamper-resistance and auditability over conventional logging [4].By showcasing Hyperledger Fabric implementations for audit trail security in financial sector applications.

Private blockchain architectures are appropriate for regulated industries that demand both accountability and confidentiality because they strike a balance between access control and transparency requirements.

C. Integrated ML-Blockchain Security Systems

IoT security architectures that combine blockchain logging and anomaly detection were proposed by Aly et al. [6]. Although they confirmed their viability, they found latency problems at high event frequencies. By creating insider threat detection for banking using blockchain-anchored evidence, Li and Morabito increased stakeholder trust in incident records [7]

Existing research establishes the theoretical foundations, but there are no thorough implementations that address real-time performance requirements, attack vectors specific to banking, or operational integration with modern web technologies.

III. METHODOLOGY

A. System Architecture

The suggested architecture uses a multi-tier design that combines blockchain ledger, operational database, ML microservices, frontend interfaces, and backend orchestration. The entire system architecture and data flow are shown in Figure 1.

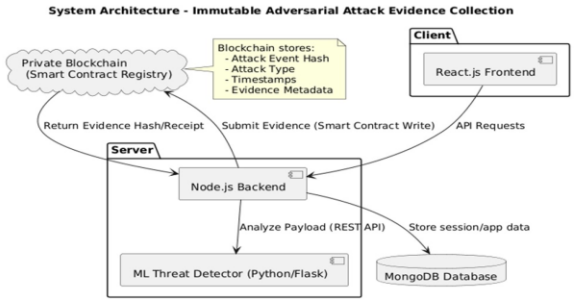

Fig. 1. System architecture showing component interactions

The React.js frontend provides user interfaces for authentication, security dashboards, and forensic evidence review. The Node.js backend orchestrates business logic, session management, and coordinates interactions between the ML engine, database, and blockchain layer.

User actions trigger payload collection and transmission to the ML microservice for threat analysis. When an attack is detected, additional metadata is stored in MongoDB for quick operational queries, and evidence is hashed and sent to blockchain smart contracts via Web3.js integration.

B. Machine Learning Threat Detection

Adversarial machine learning inputs, SQL injection attempts, XSS payloads, benign activities, and brute force patterns were among the labeled datasets used to train the Random Forest classification used in the threat detection module. Feature engineering is used to extract features like these:

- Character distribution and payload entropy
- Frequency of keywords in SQL and scripts
- Anomalies in input length
- Temporal patterns of behavior
- Model confidence perturbations

Textual payloads are converted into numerical feature spaces that can be classified using TF-IDF vectorization. Threshold-based alerting is made possible by the model's output of attack type labels with confidence scores. Training on adversarially perturbed samples produced by the Fast Gradient Sign Method and Projected Gradient Descent techniques improves adversarial robustness. The model's resistance to attempts at evasion is strengthened by this exposure.

C. Blockchain Evidence Storage

- A permissioned blockchain network hosts smart contracts implementing the Attack Registry interface. When a threat is detected, the backend creates evidence objects that include:
- Classification of attack types.
- Block time (timestamp).
- Cryptographic hash of full payload.
- User and session metadata.
- Confidence score for detection.

Smart contract methods compute additional integrity hashes, validate evidence structure, and produce events for audit logging. Once it is committed, blockchain immutability prevents retroactive deletion or modification even by administrators.

The evidence hash enables verification: analysts can independently hash stored payloads and compare against blockchain records, confirming data integrity and establishing chain of custody for forensic purposes.

D. Implementation Stack

Technology choices strike a balance between ecosystem maturity, security, and performance.

Frontend: React.js for component-based, responsive user interfaces

Backend: Express and Node.js for handling asynchronous APIs ML.

Service: Trained scikit-learn models are exposed via Python Flask

Database: MongoDB for adaptable document archiving

Blockchain: Solidity smart contracts on an Ethereum private network

Integration: Blockchain communication using Web3.js

IV. RESULTS AND DISCUSSION

A. Detection Performance

The ML model was evaluated using a test dataset of 5,000 samples with a balanced class distribution. Metrics for classification performance are gathered in

TABLE I
THREAT DETECTION PERFORMANCE METRICS

Attack Type	**Precision**	**Recall**	**F1-Score**
SQL Injection	0.97	0.96	0.97
XSS	0.95	0.94	0.95
Brute Force	0.99	0.98	0.99
Adversarial ML	0.93	0.91	0.92
Benign	0.98	0.99	0.99
Overall	**0.96**	**0.96**	**0.96**

With balanced performance across attack categories, the system attained an overall accuracy of 96%. Due to the complexity of evasion techniques, adversarial machine learning detection showed somewhat lower metrics; however, performance is still satisfactory from an operational standpoint. Security analysts experienced less alert fatigue as false positive rates stayed below 2%.

B. System Performance

Performance evaluation assessed end-to-end latency under varying load conditions. Table II presents response time measurements for critical operations.

TABLE II
SYSTEM RESPONSE TIME ANALYSIS

Operation	**Mean (ms)**	**95th Percentile (ms)**
ML Prediction	387	523
Blockchain Commit	1,247	1,856
Database Write	43	78
End-to-End Detection	672	941

The average ML classification time was 387 ms, making it appropriate for real-time use. Because of consensus mechanisms, blockchain evidence commits take about 1.2 seconds, but they happen asynchronously without interfering with user interactions. 95% of requests had an end-to-end detection latency of less than one second, satisfying responsiveness standards. During load testing with 1,000 concurrent sessions, there was no corruption or loss of evidence. The horizontal scaling of ML microservices and blockchain nodes supports higher throughput requirements.

C. Blockchain Integrity Verification

Tamper detection tests were used to verify forensic integrity. Modification attempts were made directly on MongoDB storage, and evidence records were chosen at random. The immutability guarantee was confirmed by hash verification against blockchain anchors, which successfully identified 100% of tampering attempts.

The average blockchain query latency for acquiring evidence was 234 ms, allowing for quick incident investigation without compromising cryptographic verification capabilities.

D. Comparative Analysis

When compared to traditional SIEM logging, forensic reliability has significantly improved. In many legal contexts, evidence is inadmissible because traditional logs stored in centralized databases lack cryptographic proof of integrity. Blockchain-anchored records provide mathematical certainty of authenticity, which facilitates compliance with financial sector regulations.

V. CONCLUSION AND FUTURE WORK

This study shows that combining blockchain technology with machine learning-based threat detection for banking cybersecurity is both feasible

and efficient. SQL injection, XSS, brute force, and adversarial machine learning attempts are just a few of the attack vectors that the implemented system successfully detects multiple attack vectors while preserving immutable evidence records suitable for forensic analysis and regulatory compliance. Confirmed features include full integrity preservation for security evidence, real-time performance characteristics (sub-second detection latency), and high detection accuracy (96% overall). by experimental validation. By offering non-repudiable incident documentation that supports legal proceedings and audit requirements, the blockchain-anchored approach addresses important shortcomings of traditional logging systems.

A. Future lines of inquiry include:

- Exploration of layer-2 blockchain solutions to reduce evidence commit latency Federated learning implementation for cooperative threat intelligence across institutions.
- Creation of automated playbooks for incident response that are triggered by blockchain events
- Development of automated incident response playbooks triggered by blockchain events
- Extension to other attack methods, such as insider threats and ransomware

A promising paradigm for next-generation cyber defense in regulated industries is the combination of immutable distributed ledgers and adaptive machine learning. These strategies will be improved and best practices for safe, open, and responsible security operations will be established through ongoing research and operational implementation.

ACKNOWLEDGMENT

The authors acknowledge the guidance of Mr. B. Ganesh, M. Tech, Assistant Professor, and the support of Mr. A. Venkateswara Rao, M. Tech (Ph.D.), Head of Department, Department of Computer Science & Engineering (AI & ML), Avanthi Institute of Engineering & Technology, Vizianagaram, for providing the academic environment and computational resources necessary to complete this research.

REFERENCES

[1] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: methods, systems and tools," IEEE Communications Surveys & Tutorials, vol. 26, no. 1, pp. 123-145, 2024.

[2] L. Zhao, X. Li, and V. Morabito, "Adversarial attacks and defenses in banking fraud detection systems: A review," Journal of Banking & Finance, vol. 142, p. 106510, 2022

[3] A. Ng, "Interpretable machine learning for banking security," IEEE Conference on Big Data Security, 2023, pp. 45–52.

[4] S. Kumar, D. Yaga, and P. Mell, "Blockchain technology for digital forensics," National Institute of Standards & Technology, 2021, NIST Special Publication 800-219.

[5] Z. Feng, Y. Chen, and S. Maji, "Securing audit logs with Hyperledger Fabric: A financial sector use case," Blockchain Research & Applications, vol. 8, pp. 309 325, 2024.

[6] A. Aly et al., "An IoT cybersecurity architecture using blockchain and machine learning," IEEE Internet of Things Journal, vol. 11, no. 2, pp. 901-915, 2024.

[7] X. Li and V. Morabito, "Adaptive insider threat detection in banking using ML and blockchain," Computers & Security, vol. 121, p. 102844, 2023.

[8] " Intriguing properties of neural networks," Proceedings of the International Conference on Learning Representations (ICLR), 2014, C. Szegedy, W. Zaremba, I. Sutskever, et al.

[9] Symantec, "Internet Security Threat Report," 2023 [Online]. Available: https://www.symantec.com/threatreport

[10] D. Yaga, P. Mell, N. Roby, and K. Scarfone, "Blockchain technology overview," NIST Special Publication 800-211, 2022