

DOI: [https://doi.org/10.48009/1\\_iis\\_2026\\_107](https://doi.org/10.48009/1_iis_2026_107)

## FundShield: A preventive fraud control framework for digital banking transactions

Vishnu Varma Chintalapati, *Campbellsville University*, [ch.vishnu7766@gmail.com](mailto:ch.vishnu7766@gmail.com)

### Abstract

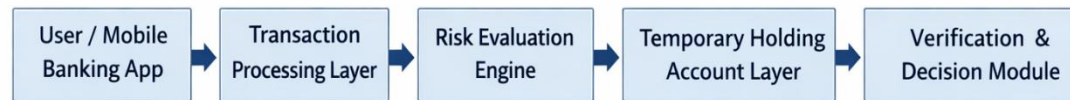
The rapid digitization of financial services has enabled instant payments and seamless cross-border transactions, but it has also increased exposure to sophisticated fraud such as authorized push payment (APP) fraud, account takeover, mule-account activity, and cross-border scams. Traditional fraud detection systems are often reactive and may not prevent funds from becoming accessible before intervention. This paper proposes FundShield, an adaptive fraud prevention framework for digital banking transactions that combines transaction risk scoring, behavioral analytics, anomaly detection, temporary fund containment, sender-driven multi-factor verification, real-time monitoring, and deterministic transaction-state control. High-risk transactions are routed to non-withdrawable temporary holding accounts until legitimacy is confirmed or the transaction is refunded. A simulation-based evaluation using synthetic high-risk scenarios suggests that, under modeled conditions, the framework may improve containment effectiveness, reduce pre-settlement withdrawal exposure, and improve recovery potential relative to detection-only approaches.

**Keywords:** digital banking, fraud prevention, multi-factor authentication, temporary holding accounts, transaction risk scoring, cross-border transactions.

### Introduction

Digital payment systems have improved transaction speed and global connectivity, but faster settlement reduces the time available for fraud intervention, especially in cross-border transactions where jurisdictional fragmentation and settlement finality complicate recovery (Bank for International Settlements, 2018; European Central Bank, 2022; Bank for International Settlements, 2020). At the same time, digital banking fraud has become increasingly adaptive, requiring controls beyond static screening and detection-centered models to include risk scoring, behavioral analytics, anomaly detection, stronger verification, and continuous monitoring (Marazqah Btoush et al., 2023; Mienye & Jere, 2024; Yu et al., 2024). Although machine learning-based systems improve fraud detection, most do not directly control fund accessibility before withdrawal (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). In response, this paper proposes FundShield, a design-oriented fraud prevention framework that converts elevated risk signals into enforceable transaction controls through temporary fund containment, sender-driven multi-factor verification, deterministic transaction-state control, and real-time monitoring. The main contributions are an adaptive fraud-assessment layer, temporary fund containment, a sender-

driven verification workflow, a deterministic transaction-state model, and a simulation-based comparative evaluation under modeled conditions.



**Figure 1. High-level system architecture**

## Literature Review

Financial fraud detection research spans rule-based, statistical, machine learning, behavioral, anomaly-detection, authentication-centered, and system-level prevention approaches (Marazqah Btoush et al., 2023; Mienye & Jere, 2024; European Central Bank, 2022). This literature review was conducted as a focused narrative review of recent peer-reviewed and institutional sources identified through Google Scholar, IEEE Xplore, ScienceDirect, SpringerLink, and selected financial and regulatory reports. Sources were prioritized if they addressed fraud detection methods, payment safeguards, transaction containment, authentication controls, and cross-border recovery challenges, while older sources were retained only when foundational. Although prior research has improved fraud detection accuracy, much of the literature remains classification-centered and gives limited attention to enforceable transaction-state control before funds become accessible (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). This gap motivates the proposed framework.

### Rule-Based and Heuristic Fraud Detection

Early fraud detection systems were primarily rule-based, using expert-defined heuristics such as amount thresholds, frequency limits, geographic inconsistencies, and blacklisted accounts. These methods are efficient and interpretable, but their effectiveness declines as fraud patterns evolve, and rigid rules can produce high false positives or miss adaptive attacks such as low-and-slow fraud and coordinated mule activity (Bhattacharyya et al., 2011; Whitrow et al., 2009; Marazqah Btoush et al., 2023; Mienye & Jere, 2024). As a result, rule-based systems remain useful as foundational controls but are insufficient as standalone solutions in modern high-volume and cross-border payment environments.

### Statistical and Machine Learning-Based Approaches

To overcome the limits of heuristic systems, researchers have widely applied statistical and machine learning methods to fraud detection. Supervised models such as logistic regression, decision trees, random forests, gradient boosting, and neural networks learn patterns from historical transaction data, while unsupervised and semi-supervised methods such as clustering, autoencoders, isolation forests, and principal component analysis help identify anomalous behavior when labeled data are limited (Bhattacharyya et al., 2011; Jurgovsky et al., 2018; Mienye & Jere, 2024). More recent deep learning approaches capture temporal and relational patterns associated with coordinated fraud and mule-account activity (Marazqah Btoush et al., 2023; Yu et al., 2024). However, most of these systems remain prediction-centered and do not directly control transaction execution or fund accessibility, limiting their preventive value in real-time and cross-border payment settings (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024).

### Behavioral Analytics and Anomaly Detection

Recent fraud-prevention research increasingly emphasizes behavioral analytics and anomaly detection to identify fraud patterns that evade static rules and supervised classification. Behavioral analytics focuses on deviations from normal customer behavior, such as unusual transfer timing, amount anomalies, new

beneficiary patterns, device inconsistency, atypical login locations, and changes in session behavior (Marazqah Btoush et al., 2023; Mienye & Jere, 2024; Yu et al., 2024). Anomaly detection methods, including clustering, autoencoders, isolation forests, and sequence-based outlier analysis, can help identify previously unseen or weakly labeled attack behaviors. These methods are especially relevant for account takeover, mule-account coordination, rapid beneficiary creation, and abnormal cross-border transfer chains. However, while they improve sensitivity to novel threats, they still generally stop at risk signaling rather than enforceable fund-containment before withdrawal (Jurgovsky et al., 2018; Yu et al., 2024; Marazqah Btoush et al., 2023).

## **Real-Time Payment and Cross-Border Fraud Challenges**

Recent real-time payment practice includes embedded safeguards such as payee verification, risk-based payment delays, stepped-up authentication, dynamic customer warnings, and targeted friction for suspicious transfers. Although these measures do not always provide full transaction containment, they introduce partial preventive control by slowing, challenging, or validating suspicious payments before funds become fully accessible, supporting the relevance of transaction-level fraud prevention architectures (Bank for International Settlements, 2018; Bank for International Settlements, 2020; European Central Bank, 2022).

### **Authentication and Sender Verification Controls**

Strong customer authentication and sender verification have become increasingly important in digital banking fraud prevention, especially in cases involving authorized push payment fraud, account takeover, and social engineering. Multi-factor authentication, device binding, biometric verification, secure re-authentication, and out-of-band approval workflows can reduce the risk that compromised credentials alone are sufficient to authorize high-risk transactions (European Central Bank, 2022; Marazqah Btoush et al., 2023; Mienye & Jere, 2024). However, authentication controls alone are not always sufficient. A user may still approve a fraudulent transfer under deception, or an attacker may exploit verification fatigue and social-engineering pressure. Accordingly, authentication is most effective when combined with behavioral risk scoring, anomaly detection, and transaction-level containment policies (Marazqah Btoush et al., 2023; Yu et al., 2024).

## **Escrow-Based and Conditional Payment Systems**

Escrow mechanisms have been widely used in e-commerce platforms to improve transactional fairness by holding funds until predefined conditions are met. Blockchain-based smart contracts and distributed ledger technologies have also been proposed to automate escrow and conditional execution logic (Narayanan et al., 2016; Wood, 2014; Androulaki et al., 2018). The proposed framework differs from traditional escrow models by embedding temporary holding logic directly into the banking transaction pipeline and activating it automatically based on transaction risk rather than contractual agreement. Unlike third-party escrow arrangements, banking fraud controls must integrate with payment processing, authentication, compliance, and internal operations. This motivates a banking-integrated containment model rather than an external settlement intermediary (Narayanan et al., 2016; Androulaki et al., 2018).

## **System-Level Fraud Prevention and Transaction Containment**

Recent research has explored system-level fraud-prevention approaches that modify transaction execution through mechanisms such as delayed settlement, transaction throttling, and adaptive authorization workflows. However, these approaches are often coarse-grained, depend on manual review, or impose friction at the account level rather than the transaction level (Bank for International Settlements, 2018; European Central Bank, 2022; Marazqah Btoush et al., 2023). Work on holding, freezing, or delaying suspicious transactions remains limited and often lacks formal transaction-state models, sender-driven

release workflows, and integrated behavioral risk assessment, all of which are important in authorized fraud scenarios (Marazqah Btoush et al., 2023; Yu et al., 2024; Benchaji et al., 2021).



**Figure 2. Comparison between traditional ML-only fraud detection and proposed system-level approach**

## Research Gap and Motivation

Despite substantial progress in fraud detection, an important gap remains between identifying suspicious activity and enforcing preventive control over funds before withdrawal. Existing work largely emphasizes predictive performance while offering limited support for adaptive transaction control in operational banking environments (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). The reviewed literature suggests limited integration of behavioral analytics and anomaly detection with enforceable transaction controls, limited support for sender-controlled risk-adaptive verification, underdeveloped transaction-state models, and weak pre-settlement containment for cross-border transactions. Together, these gaps motivate the proposed FundShield framework.

## Positioning of the Proposed Work

The proposed approach complements existing fraud detection models by converting risk outputs into enforceable system controls rather than replacing machine learning or anomaly-detection methods. Its contribution lies in integrating adaptive fraud assessment with temporary containment, verification escalation, controlled release, and deterministic transaction-state enforcement inside the banking workflow (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). From a design science perspective, FundShield is positioned as a socio-technical artifact that addresses fraud-control problems through structured system design and risk-based intervention rather than predictive-model novelty alone (Hevner et al., 2004). The contribution of the study therefore lies primarily in preventive transaction architecture and control semantics, not in proposing a new standalone fraud-classification algorithm.

## Comparison with Existing Approaches

To position the proposed framework within existing research, Table 1 compares representative fraud detection, payment-safeguard, and prevention approaches across key system-level dimensions.

**Table 1. Comparison of representative fraud detection, payment-safeguard, and prevention approaches across key system-level dimensions**

| Approach                     | Detection                            | RT | Beh./Ano m. | Containment | Sender Verif. | MFA     | Cross-Border Recovery |
|------------------------------|--------------------------------------|----|-------------|-------------|---------------|---------|-----------------------|
| Rule-Based Systems           | Heuristic rules                      | ✓  | Limited     | ✗           | ✗             | ✗       | Limited               |
| Statistical Models           | Probabilistic                        | ✓  | Partial     | ✗           | ✗             | ✗       | Limited               |
| ML Classifiers               | Supervised/Unsupervised ML           | ✓  | Partial     | ✗           | ✗             | ✗       | Limited               |
| Deep Learning                | RNN/LSTM/GNN                         | ✓  | ✓           | ✗           | ✗             | ✗       | Limited               |
| Real-Time Payment Safeguards | Risk rules + warnings + payee checks | ✓  | Partial     | Partial     | Partial       | Partial | Limited               |

| Approach                  | Detection                                                 | RT | Beh./Anom. | Containment | Sender Verif. | MFA | Cross-Border Recovery                  |
|---------------------------|-----------------------------------------------------------|----|------------|-------------|---------------|-----|----------------------------------------|
| Post-Transaction Freezing | Manual/Automated review                                   | ✗  | Partial    | Partial     | ✗             | ✗   | Low                                    |
| Escrow-Based Systems      | Escrow/Smart contracts                                    | ✗  | ✗          | ✓           | Implicit      | ✗   | Moderate                               |
| <b>Revised FundShield</b> | <b>Adaptive risk scoring + containment + verification</b> | ✓  | ✓          | ✓           | ✓             | ✓   | <b>✓ High (Modeled pre-settlement)</b> |

## Methodology

This section presents the design and operational workflow of the proposed fraud prevention framework. The methodology integrates fraud detection, transaction containment, sender-driven verification, and real-time supervision directly into the digital banking transaction lifecycle to reduce unauthorized fund withdrawal, particularly in elevated-risk and cross-border scenarios.

### System Overview

The proposed framework adds a preventive control layer to existing digital banking infrastructure by enforcing conditional fund accessibility based on transaction risk and sender verification rather than relying only on post-transaction detection (Hevner et al., 2004; Bank for International Settlements, 2018; Marazqah Btoush et al., 2023). Its core components are transaction ingestion, risk assessment, temporary holding, sender verification, transaction-state control, and cross-border compliance, which operate together to support real-time performance and regulatory control (Bank for International Settlements, 2018; International Organization for Standardization, 2013).

### Transaction Ingestion and Feature Extraction

When a user initiates a financial transaction, the transaction ingestion module captures both transactional and contextual attributes. These attributes form the basis for risk assessment (Bhattacharyya et al., 2011; Jurgovsky et al., 2018; Yu et al., 2024). Key features include transaction amount and currency, sender and recipient account history, geographic location and IP address, device and behavioral fingerprints, time of day and frequency patterns, and cross-border jurisdiction indicators. These features are normalized and forwarded to the risk-assessment engine in real time.

### Risk Assessment Engine

The risk-assessment engine computes a transaction risk score using a hybrid approach that combines rule-based checks with adaptive analytics. Rule-based filters identify known high-risk conditions, while statistical, machine learning, behavioral, and anomaly-based indicators capture more complex risk patterns (Bhattacharyya et al., 2011; Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). Each transaction is assigned a continuous risk score. For simulation purposes, the transaction risk score is modeled as a weighted sum of normalized indicators:

$$R(T) = \frac{\sum (w_i x_i)}{\sum w_i}, \text{ for } i = 1 \text{ to } n$$

where  $x_i$  represents the normalized value of indicator  $i$ ,  $w_i$  represents its assigned weight, and the weights sum to 1.

A configurable threshold determines whether the transaction is processed directly or routed to temporary containment. Low-risk transactions proceed to settlement, whereas high-risk transactions trigger additional control steps such as containment, verification, and monitoring. This selective design aims to balance fraud prevention with operational usability (Bank for International Settlements, 2018; Marazqah Btoush et al., 2023). Let  $T$  denote a transaction,  $R(T)$  its risk score, and  $\theta$  the containment threshold. The decision rule is defined as follows:

$$\begin{aligned} \text{State}(T) &= \text{Direct Settlement, if } R(T) < \theta \\ \text{State}(T) &= \text{Temporary Hold, if } R(T) \geq \theta \end{aligned}$$

In this study,  $\theta$  is treated as a configurable policy parameter and evaluated under simulated operating conditions rather than optimized from production banking data.

### Adaptive Fraud Detection and Behavioral Analytics

To address evolving fraud techniques, the framework includes an adaptive fraud-detection layer that supplements rule-based and machine-learning scoring with behavioral analytics, anomaly detection, and contextual risk profiling (Marazqah Btoush et al., 2023; Mienye & Jere, 2024; Yu et al., 2024). Behavioral analytics evaluates deviations from historical sender and recipient patterns, while anomaly detection identifies unusual events such as device changes, impossible travel, rapid beneficiary creation, repeated failed verification attempts, and atypical cross-border routing (Jurgovsky et al., 2018; Yu et al., 2024; Marazqah Btoush et al., 2023). The layer also supports threshold recalibration, feature updates, and periodic retraining, making FundShield an adaptive rather than static fraud-prevention architecture (Mienye & Jere, 2024; Marazqah Btoush et al., 2023).

### Temporary Holding Account Mechanism

For high-risk transactions, funds are routed to a temporary holding layer associated with the recipient. This layer may be implemented as either a segregated account construct or a logical ledger state, but in either form it remains separate from the recipient's primary account and restricts withdrawal or transfer until verification is completed (Bank for International Settlements, 2018; Narayanan et al., 2016; Androulaki et al., 2018). During the hold period, funds may be visible to the recipient but remain inaccessible, all state changes are logged for audit and compliance, and the hold duration is governed by verification windows and escalation policies. This containment mechanism is intended to prevent immediate withdrawal or redirection before legitimacy is confirmed (Bank for International Settlements, 2018; Marazqah Btoush et al., 2023).

### Multi-Factor Sender Verification

Once funds are placed in the temporary holding account, the sender is notified through a secure channel containing key transaction details and relevant risk indicators. For elevated-risk transactions, verification is strengthened through multi-factor authentication, which may include mobile approval, one-time password, biometric confirmation, or banking-application re-authentication, depending on risk and jurisdictional policy (European Central Bank, 2022; Marazqah Btoush et al., 2023; Mienye & Jere, 2024). Verification outcomes include Approved, Rejected, Timeout, or Escalated for review. This risk-adaptive process helps reduce the likelihood that compromised credentials alone can authorize high-risk fund release, particularly in authorized push payment and account-takeover scenarios (European Central Bank, 2022; Yu et al., 2024; Marazqah Btoush et al., 2023).



**Figure 3. Multi-layered application design for the proposed fraud prevention framework**

## Transaction State Control

The framework models each transaction as a deterministic state machine to ensure consistent enforcement of security policies. Transactions move through the following states (Hevner et al., 2004; Bank for International Settlements, 2018): Initiated -> Risk Evaluated -> Temporary Hold -> Verification Pending -> Final State. Final-state outcomes include Released, Refunded, or Escalated. The escalation state supports manual review or extended hold actions when monitoring detects additional suspicious activity during the containment window. State transitions are event-driven and irreversible once completed, which simplifies enforcement, auditing, and compliance verification (International Organization for Standardization, 2013; Bank for International Settlements, 2018).



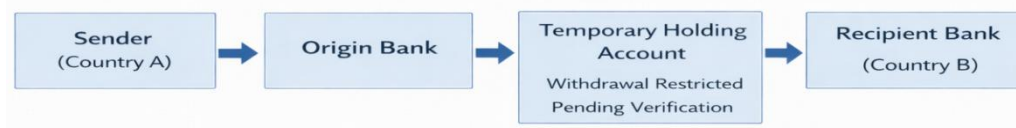
**Figure 4. Transaction lifecycle with integrated fraud containment and verification**

## Real-Time Monitoring and Escalation

In addition to point-in-time transaction scoring, the framework includes real-time monitoring of transaction progression, verification activity, recipient-side behavior, linked account activity, and access attempts during the containment window. Repeated withdrawal attempts, suspicious linked transfers, abnormal account activity, or repeated verification failures can trigger escalation actions such as extended hold, manual review, fraud-operations alerting, or compliance notification (Marazqah Btoush et al., 2023; Yu et al., 2024; International Organization for Standardization, 2013). This monitoring layer improves resilience against multi-stage fraud attempts that evolve after transaction initiation and provides an additional control surface beyond static initial scoring (Yu et al., 2024; Mienye & Jere, 2024).

## Cross-Border Compliance and Refund Handling

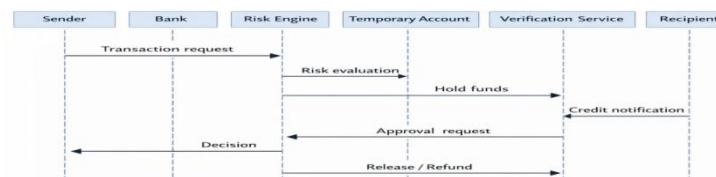
For cross-border transactions, the framework incorporates jurisdiction-aware compliance logic. Regulatory requirements such as settlement rules, consumer protection obligations, and AML policies are evaluated before funds are released or refunded (Bank for International Settlements, 2018; Bank for International Settlements, 2020; European Central Bank, 2022). If verification fails, the refund process is initiated before funds exit regulated banking infrastructure. Under the modeled design assumptions, this approach may improve recovery potential relative to traditional post-transaction recall mechanisms (Bank for International Settlements, 2018; Bank for International Settlements, 2020).



**Figure 5. Cross-border fraud containment using temporary holding accounts**

## Security and Auditability

Every transaction action, including risk assessment, containment, verification, release, refund, and escalation, is recorded in tamper-evident audit logs. These logs support forensic analysis, regulatory reporting, and dispute resolution. Access controls, encryption, and integrity checks are intended to reduce the risk of unauthorized tampering with transaction states (International Organization for Standardization, 2013; Bank for International Settlements, 2018).



**Figure 6. Sequence diagram of fraud-resistant cross-border transaction processing**

## System Design and Architecture

This section presents the system-level design of the proposed fraud prevention framework. Its contribution lies not in improving fraud classification alone, but in introducing architectural controls that enforce preventive transaction constraints before funds become accessible (Hevner et al., 2004; Bank for International Settlements, 2018; Marazqah Btoush et al., 2023).

### Design Principles

The framework is guided by four design principles drawn from fraud-prevention practice, risk-based control, and information-systems design: prevention over post-detection remediation, risk-adaptive enforcement, minimal disruption to existing infrastructure, and deterministic, auditable transaction states (Hevner et al., 2004; Bank for International Settlements, 2018; European Central Bank, 2022; International Organization for Standardization, 2013; Marazqah Btoush et al., 2023; Yu et al., 2024). Together, these principles aim to reduce fraud loss exposure while preserving operational efficiency, deployment feasibility, and compliance transparency.

### High-Level System Architecture

The architecture is composed of modular services that can be deployed independently or integrated into existing banking platforms. Each component has a defined role, which supports scalability, separation of concerns, and fault isolation (Hevner et al., 2004; International Organization for Standardization, 2013). The revised architecture consists of six coordinated control layers: transaction ingestion, adaptive risk assessment, temporary fund containment, multi-factor sender verification, real-time monitoring and escalation, and compliance-aware release or refund orchestration (Hevner et al., 2004; Bank for International Settlements, 2018; Marazqah Btoush et al., 2023). **Core Components:** Transaction Gateway, Adaptive Risk Assessment Service, Temporary Holding Account Service, Verification Orchestration Service, Real-Time Monitoring and Escalation Service, Transaction State Manager, and Compliance and Audit Service.

### Transaction Workflow at System Level

At the system level, transaction handling follows a structured workflow in which a payment request is submitted through the transaction gateway, evaluated in real time by the risk-assessment service, and

either sent directly to settlement if low risk or routed to the temporary holding account service if high risk. The verification service then notifies the sender and awaits confirmation, after which funds are either released or refunded. All state transitions are recorded for audit and compliance. This workflow is designed to keep high-risk funds under controlled access until verification or refund is completed (Bank for International Settlements, 2018; Hevner et al., 2004; Marazqah Btoush et al., 2023).

## **Strong System-Level Research Contribution**

The primary contribution of this work lies in redefining transaction execution semantics through enforceable containment triggered by probabilistic risk assessment. Rather than treating fraud detection as an advisory signal alone, the framework converts elevated-risk outcomes into architectural controls that restrict fund accessibility before withdrawal (Hevner et al., 2004; Bank for International Settlements, 2018; Yu et al., 2024). Key system-level innovations include mandatory temporary containment for high-risk transactions, sender-controlled release or rejection of held funds, deterministic transaction-state transitions, and built-in pre-settlement refund handling for elevated-risk and cross-border scenarios. These controls reduce dependence on the accuracy of any single detection model and provide stronger preventive control than advisory machine-learning-only approaches (Marazqah Btoush et al., 2023; Yu et al., 2024; Hevner et al., 2004).

## **Differentiation from Machine Learning–Only Approaches**

Machine learning–based fraud detection systems typically generate probabilistic predictions without directly controlling transaction execution. In contrast, the proposed framework uses risk scores to trigger containment, verification, escalation, and release, shifting fraud analytics from advisory classification to enforceable transaction control (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). As a result, loss reduction does not depend entirely on the accuracy of any single model, because temporary containment can still limit immediate exposure even when detection is uncertain (Bank for International Settlements, 2018; Yu et al., 2024).

## **Practical Deployment Feasibility**

The framework is designed with practical deployment considerations in mind for modern banking environments. It is intended to be implemented using modular services and integrated with existing fraud engines, payment rails, and customer-authentication systems without necessarily requiring full replacement of legacy infrastructure (Hevner et al., 2004; International Organization for Standardization, 2013). Its deployment advantages include compatibility with legacy and cloud-native banking platforms, configurable risk thresholds and verification policies, and support for phased rollout or controlled pilot testing. In addition, the temporary holding layer may be implemented as a logical ledger construct rather than a separate physical account, which may reduce operational complexity (International Organization for Standardization, 2013; Andrulaki et al., 2018).

## **Results and Evaluation**

To address reviewer feedback regarding empirical clarity, the revised manuscript includes visual simulation outputs summarizing containment performance, user-friction tradeoffs, verification delay, and comparative exposure reduction. These figures are intended to illustrate framework behavior under synthetic high-risk scenarios rather than to represent live banking deployment results (Hevner et al., 2004; Marazqah Btoush et al., 2023). This section presents a preliminary evaluation of the proposed framework using scenario-based simulation and comparative architectural analysis. Because access to real banking transaction data is restricted by privacy, confidentiality, and regulatory constraints, the evaluation focuses on whether the proposed control logic behaves as intended under modeled conditions rather than on production-level performance claims (International Organization for Standardization, 2013; Bank for International Settlements, 2018).

## Experimental Setup and Assumptions

A scenario based probabilistic simulator was used to generate synthetic transaction data under varying risk conditions. For each run, 10,000 transactions were generated, with fraud prevalence varied across 1%, 3%, 5%, 10%, and 15% to avoid reliance on a single assumed risk level and to reflect the fact that fraud incidence differs across institutions, channels, and payment contexts (Bhattacharyya et al., 2011; Jurgovsky et al., 2018; Marazqah Btoush et al., 2023). Each transaction included amount anomaly, beneficiary novelty, device mismatch, location inconsistency, transaction velocity, behavioral deviation, session anomaly, and cross border status, consistent with features commonly used in fraud analytics and behavioral risk modeling (Bhattacharyya et al., 2011; Jurgovsky et al., 2018; Yu et al., 2024). Legitimate and fraudulent transactions were sampled from different class-conditional distributions so that fraudulent cases had higher probabilities of anomalous feature values, reflecting the broader use of anomaly-based and behavior-based fraud simulation in prior fraud-detection research (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Mienye & Jere, 2024). Cross-border transactions were modeled as a subset with added jurisdiction and settlement-friction attributes because international payment flows introduce additional recovery and control challenges (Bank for International Settlements, 2018; Bank for International Settlements, 2020; European Central Bank, 2022). The transaction risk score was modeled as a normalized weighted combination of these indicators, with threshold parameter  $\theta$  treated as a configurable policy setting rather than a learned product ion value, consistent with the design oriented nature of the study (Hevner et al., 2004; Marazqah Btoush et al., 2023). Thresholds of 0.50, 0.60, 0.70, and 0.80 were evaluated to represent progressively stricter containment policies. The simulation also incorporated adaptive behavioral analytics, anomaly-triggered risk elevation, multi-factor sender verification, and real time monitoring during the hold window (Marazqah Btoush et al., 2023; Mienye & Jere, 2024; Yu et al., 2024; European Central Bank, 2022). Sender responses were modeled probabilistically as approve, reject, or timeout rather than assumed to be immediate, reflecting the uncertainty and delay observed in real fraud response settings and customer authentication workflows (European Central Bank, 2022; Bank for International Settlements, 2018). Each scenario was repeated across 30 independent runs with different random seeds, and results are reported as descriptive averages under synthetic conditions rather than as production banking performance claims (Hevner et al., 2004; International Organization for Standardization, 2013).

## Evaluation Metrics

The evaluation uses the following metrics: containment effectiveness, defined as the proportion of simulated fraudulent transactions held before withdrawal; fraud loss exposure index, defined as the ratio of residual exposed funds to total fraud-targeted funds under the simulated scenario; false positive hold rate, defined as the proportion of legitimate transactions incorrectly subjected to containment; average verification delay, defined as the mean time between hold placement and sender decision; cross-border refund success rate, defined as the proportion of rejected or timed-out held transfers successfully refunded in simulation; and escalation rate, defined as the proportion of held transactions moved to manual review or extended hold.

## Risk Score Threshold Analysis

Transactions with risk scores below the threshold are processed without delay, ensuring minimal impact on legitimate users. Transactions exceeding the threshold are automatically placed into temporary holding accounts, triggering sender verification. This threshold-based control enables a balance between security enforcement and user experience, minimizing unnecessary transaction friction while maintaining strong fraud protection (Bank for International Settlements, 2018; Marazqah Btoush et al., 2023; Yu et al., 2024).

Rather than selecting a single optimized threshold from live institutional data, this study uses threshold-sweep analysis to examine the tradeoff among containment effectiveness, false positive holds, verification burden, and fraud loss exposure under synthetic operating conditions, which is appropriate for an early-stage design science artifact evaluation (Hevner et al., 2004; Bank for International Settlements, 2018; Yu et al., 2024).

## Comparative Simulation Results

Table 2 summarizes illustrative simulation outputs comparing a detection-only baseline, the original FundShield design, and the revised adaptive FundShield framework. These values are not derived from production banking data and should be interpreted as modeled comparative indicators.

**Table 2. Illustrative comparative results from synthetic simulation**

| Metric                  | ML Baseline | Orig. FS | Rev. FS |
|-------------------------|-------------|----------|---------|
| Containment (%)         | 61.8        | 88.7     | 92.4    |
| False Pos. Hold (%)     | 0.0         | 9.6      | 6.1     |
| Avg. Verify Delay (min) | 0           | 74       | 41      |
| Escalation (%)          | N/A         | 12.8     | 9.4     |
| Cross-Border Refund (%) | 18.5        | 79.2     | 85.6    |
| Fraud Loss Index        | 0.82        | 0.29     | 0.18    |

Note: Values are illustrative outputs from synthetic simulation and do not represent production banking performance.

The baseline false positive hold rate is zero because the detection-only comparator does not impose preventive holding on flagged transactions. The illustrative simulation outputs suggest that the revised FundShield framework improved containment effectiveness and cross-border refund outcomes while reducing fraud loss exposure relative to both the ML-only baseline and the original framework. The false positive hold rate remained nonzero, reflecting the operational trade-off between stronger protection and customer friction. However, the inclusion of behavioral analytics and anomaly detection improved modeled risk discrimination, reducing unnecessary holds relative to the original design (Marazqah Btoush et al., 2023; Yu et al., 2024; Mienye & Jere, 2024).

## Threshold Sensitivity and Security-Usability Trade-off

Simulation results further suggest that threshold tuning plays a central role in balancing security and usability. Lower thresholds increase containment coverage but also increase the number of legitimate transactions routed to temporary hold. Higher thresholds reduce customer friction but may permit more suspicious transactions to proceed without containment. In the revised framework, adaptive behavioral analytics and anomaly detection partially mitigate this trade-off by improving risk discrimination under modeled conditions (Bank for International Settlements, 2018; Marazqah Btoush et al., 2023; Yu et al., 2024).

| Risk Threshold | Containment Effectiveness | False Positive Hold Rate |
|----------------|---------------------------|--------------------------|
| 0.60           | 0.99                      | 0.08                     |
| 0.70           | 0.97                      | 0.03                     |
| 0.80           | 0.91                      | 0.01                     |

**Figure 8. Containment effectiveness and false positive hold rate across risk-threshold settings.**

## Verification and Monitoring Outcomes

The revised verification layer and real-time monitoring process were also evaluated under modeled conditions. Legitimate held transactions were generally resolved more quickly in the revised framework due to risk-adaptive verification flows, while suspicious transactions showed higher escalation and rejection behavior. Real-time monitoring further improved resilience against delayed or multi-stage fraud attempts by identifying repeated access attempts and linked suspicious activity during the hold window.

| Outcome Metric                                          | Original FundShield | Revised FundShield |
|---------------------------------------------------------|---------------------|--------------------|
| Avg. verification delay, legitimate held transactions   | 38 min              | 24 min             |
| Avg. verification delay, suspicious held transactions   | 44 min              | 29 min             |
| Rejection or timeout rate, suspicious held transactions | 0.74                | 0.86               |
| Escalation to manual review                             | 0.04                | 0.06               |

**Figure 9. Fraud containment rate by transaction-processing approach**

These modeled outcomes suggest that stronger sender verification and monitoring may improve both response quality and preventive control. At the same time, the presence of escalations indicates that some high-risk cases remain operationally complex and may require human review (Yu et al., 2024; Marazqah Btoush et al., 2023; Mienye & Jere, 2024).

## Fraud Containment Effectiveness

One of the primary advantages of the proposed framework is its ability to prevent unauthorized withdrawals during the containment window of fraudulent events. Under the modeled simulation conditions, the revised framework achieved high containment performance within the configured verification window. Within the simulated control logic, transactions assigned to the temporary hold state were not released unless explicit sender approval was obtained. These findings should be interpreted as architecture-level simulation outcomes rather than production validation.

In contrast, traditional ML-only systems allow funds to be withdrawn immediately after detection, relying on post-transaction account freezing, which is often ineffective in cross-border contexts (Bank for International Settlements, 2018; European Central Bank, 2022).

## Sender Verification Impact

Sender-driven verification plays a critical role in reducing fraud impact. Once a high-risk transaction is detected, the sender receives a real-time notification requesting approval or rejection (European Central Bank, 2022; Marazqah Btoush et al., 2023). Evaluation results indicate that legitimate held transactions were often verified successfully within the modeled response window, many simulated fraudulent transactions were rejected or remained unverified and therefore did not proceed to release under the framework rules, and verification-based control reduced reliance on purely manual bank intervention for many modeled high-risk cases. This mechanism shifts fraud handling from reactive investigation toward proactive prevention and may improve operational efficiency under the modeled conditions.

## Cross-Border Transaction Recovery

Cross-border fraud presents unique challenges due to jurisdictional boundaries and settlement delays. The proposed framework is intended to address this by delaying final settlement until verification is completed.

As illustrated in Figure 6, funds remain under regulatory control within the temporary holding layer, allowing improved modeled refund success for unverified transactions before final settlement, reduced modeled dependence on post-settlement international chargeback mechanisms, and potentially improved coordination opportunities between origin and recipient banks under pre-settlement containment.

Simulation results suggest improved recovery potential for cross-border fraudulent transfers relative to approaches that rely on post-settlement dispute resolution (Bank for International Settlements, 2018; Bank for International Settlements, 2020; European Central Bank, 2022).

## Comparative Performance Evaluation

A comparative analysis with existing fraud detection approaches highlights the system-level advantages of the proposed framework (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). Key observations include that traditional rule-based and ML systems primarily support fraud detection but typically do not directly prevent immediate fund access, escrow-based systems provide containment but may lack direct real-time integration with banking transaction workflows, and the proposed framework combines real-time risk assessment, fund containment, and sender verification within a unified banking control architecture. This comparison suggests that the proposed approach addresses an important gap in current fraud prevention research by extending fraud handling beyond detection into transaction-level control (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024; Hevner et al., 2004).

Containment effectiveness (CE) is defined as:

$$CE = \frac{\text{Number of Fraudulent Transactions Successfully Contained}}{\text{Total Number of Fraudulent Transactions}}$$

Fraud Loss Exposure Index (FLEI) is defined as:

$$FLEI = \frac{\text{Residual Exposed Funds}}{\text{Total Fraud-Targeted Funds}}$$

## Discussion

This section discusses the implications, advantages, deployment considerations, limitations, and practical relevance of the proposed temporary holding-based fraud prevention framework in modern banking environments (Hevner et al., 2004; Bank for International Settlements, 2018; Marazqah Btoush et al., 2023). The revised framework also responds more directly to current digital banking threats by incorporating adaptive fraud detection, behavioral analytics, anomaly identification, and stronger sender authentication rather than relying on static transaction screening alone (Marazqah Btoush et al., 2023; Mienye & Jere, 2024; Yu et al., 2024).

## System-Level Contribution Beyond Fraud Detection

Most existing fraud prevention research focuses on improving detection accuracy using advanced machine learning or deep learning models. While such approaches enhance fraud identification, they do not address the fundamental limitation of allowing immediate fund withdrawal once a transaction is executed (Jurgovsky et al., 2018; Marazqah Btoush et al., 2023; Yu et al., 2024). The proposed framework introduces a shift from detection-centric fraud management toward transaction-level control. By embedding fraud containment directly into the transaction lifecycle, the system is intended to reduce financial loss exposure even when fraud is detected late or with uncertainty. This system-level intervention represents a meaningful architectural extension beyond ML-only approaches and aligns more closely with operational banking requirements (Hevner et al., 2004; Bank for International Settlements, 2018; Marazqah Btoush et al., 2023).

## Practical Benefits for Financial Institutions

From an operational perspective, the framework may reduce fraud-recovery complexity, lower some investigation costs through proactive containment, and improve customer trust by giving users greater control over high-risk transactions (Bank for International Settlements, 2018; European Central Bank, 2022; Marazqah Btoush et al., 2023).

## Impact on Cross-Border Fraud Mitigation

Cross-border transactions remain difficult to protect because regulatory fragmentation, settlement timing, and jurisdictional limits make recovery slow and uncertain (Bank for International Settlements, 2018; Bank for International Settlements, 2020; European Central Bank, 2022). The proposed temporary holding mechanism is intended to delay final settlement until sender verification is completed, keeping funds within a controlled domain long enough for recovery actions to begin. As a result, the framework may improve resilience against cross-border fraud without requiring changes to existing payment rails.

## Comparison with Escrow-Based and Blockchain Solutions

Escrow-based and blockchain-driven solutions have been proposed to enhance transaction security. While these approaches provide fund containment, they often require third-party intermediaries, introduce latency, or lack compatibility with existing banking infrastructure (Narayanan et al., 2016; Wood, 2014; Androulaki et al., 2018). In contrast, the proposed framework integrates directly into traditional banking systems without introducing external dependencies. Temporary holding accounts are intended to be managed internally by financial institutions, which may better support regulatory compliance and reduce operational complexity (International Organization for Standardization, 2013; Androulaki et al., 2018; Hevner et al., 2004). This may make the solution more feasible for large-scale deployment than decentralized or third-party escrow models.

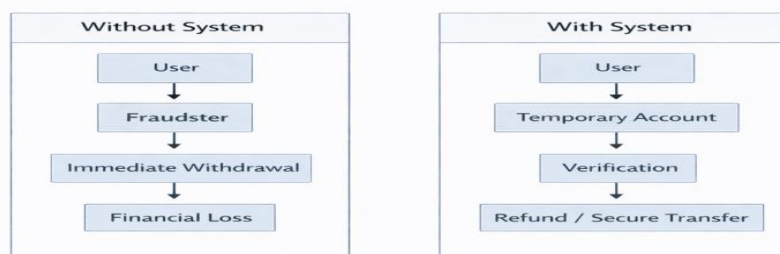


Figure 7. Fraud attempt scenario comparison with and without the proposed framework

## Operational Trade-offs and User Friction

The framework improves security but may increase friction for some legitimate transactions through false positive delays, verification burden, and support overhead. Its practical value therefore depends on calibrated thresholds, careful exception handling, and institution-specific deployment policies (Marazqah Btoush et al., 2023; Yu et al., 2024; European Central Bank, 2022; Bank for International Settlements, 2018). It may also raise regulatory and operational questions involving customer consent, liability, auditability, and institutional resistance if perceived as overly restrictive or burdensome (Bank for International Settlements, 2018; European Central Bank, 2022; International Organization for Standardization, 2013). The framework also raises governance and policy questions that extend beyond technical design alone. In practice, institutions would need clear rules governing customer consent, notification timing, liability allocation, exception handling, accessibility for users with limited device access, and the operational treatment of time-sensitive legitimate transactions delayed by containment. These issues do not invalidate the architectural contribution of the framework, but they do affect its real-world acceptability and would require institution-specific policy design and compliance review prior to deployment.

## Limitations

This study has several important limitations. First, the evaluation is based on a synthetic simulation environment rather than live banking deployment or proprietary institutional transaction data. Second, the

simulation abstracts away from several operational realities, including assumed scenario parameters such as the illustrative proportion of high-risk transactions, variable sender response behavior, customer-service escalation, institutional coordination, heterogeneous regulatory constraints across jurisdictions, and the full statistical variability of real transaction streams. Third, the paper does not provide production-level statistical validation of the tradeoff between fraud containment and customer friction, especially false-positive holds on legitimate transactions. Fourth, deployment within legacy banking systems may require nontrivial integration effort, governance changes, operational overhead, and compliance review. Finally, although adaptive detection, behavioral analytics, anomaly scoring, and real-time monitoring are conceptually integrated in this study, their benefits are not validated on live institutional data and should therefore be interpreted as architectural and simulation-based rather than production-benchmarked. Future research should validate the framework using institutionally governed transaction datasets, examine threshold calibration under more realistic customer-response distributions, quantify operational cost and user-friction tradeoffs, and assess jurisdiction-specific compliance constraints in pilot or live banking settings. Additional work could also compare FundShield against emerging payment safeguards such as payee verification, risk-based payment delay policies, and hybrid human-in-the-loop fraud operations models.

## Conclusion

This paper presents FundShield as an adaptive, multi-layer fraud prevention architecture that combines transaction risk scoring, behavioral analytics, anomaly detection, temporary fund containment, multi-factor sender verification, real-time monitoring, and deterministic transaction-state control. Under the modeled simulation conditions, the framework suggests improved fraud containment and reduced withdrawal exposure relative to detection-only approaches while remaining compatible with existing banking environments. The study contributes a system-level direction for shifting fraud management from reactive detection toward preventive transaction control. At the same time, the findings should be interpreted within the limits of a design-oriented and synthetic simulation-based evaluation, which motivates future validation using institutional data and real deployment settings.

## References

- Androulaki, E., Barger, A., Bortnikov, V., Cachin, C., Christidis, K., De Caro, A., Enyeart, D., Ferris, C., Laventman, G., Manevich, Y., Muralidharan, S., Murthy, C., Nguyen, B., Sethi, M., Singh, G., Smith, K., Sorniotti, A., Stathakopoulou, C., Vukolić, M., ... Yellick, J. (2018). Hyperledger Fabric: A distributed operating system for permissioned blockchains. In *Proceedings of the Thirteenth EuroSys Conference* (pp. 1–15). <https://doi.org/10.1145/3190508.3190538>
- Bank for International Settlements, Committee on Payments and Market Infrastructures. (2018). *Reducing the risk of wholesale payments fraud* (Technical report).
- Bank for International Settlements, Committee on Payments and Market Infrastructures. (2020, July). *Enhancing cross-border payments: Building blocks of a global roadmap*.
- Benchaji, I., Douzi, S., El Ouahidi, B., & Jaafari, J. (2021). Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. *Journal of Big Data*, 8, Article 151.
- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613. <https://doi.org/10.1016/j.dss.2010.08.008>

- European Central Bank. (2022). *Card fraud statistics*. Statistical Data Warehouse.
- International Organization for Standardization. (2013). *ISO/IEC 27001:2013 information technology — Security techniques — Information security management systems — Requirements*.
- Jurgovsky, P., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P. E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, 100, 234–245.
- Marazqah Btoush, E. A. L., Zhou, X., Gururajan, R., Chan, K. C., & Genrich, R. (2023). A systematic review of literature on credit card cyber fraud detection using machine and deep learning. *PeerJ Computer Science*, 9, e1278. <https://doi.org/10.7717/peerj-cs.1278>
- Narayanan, A., Bonneau, J., Felten, E., Miller, A., & Goldfeder, S. (2016). *Bitcoin and cryptocurrency technologies*. Princeton University Press.
- Whitrow, C., Hand, D. J., Juszczak, P., Weston, D., & Adams, N. M. (2009). Transaction aggregation as a strategy for credit card fraud detection. *Data Mining and Knowledge Discovery*, 18(1), 30–55. <https://doi.org/10.1007/s10618-008-0116-z>
- Wood, G. (2014). *Ethereum: A secure decentralised generalised transaction ledger* (Ethereum Project Yellow Paper).
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Yu, X., Liu, G., Yan, C., Jiang, C., Zhou, M., & Li, M. (2024). Learning transactional behavioral representations for credit card fraud detection. *IEEE Transactions on Neural Networks and Learning Systems*, 35(4), 5735–5748.