Conf42 Cloud Native

Scalable Cloud-Native Friendly Fraud Detection using Temporal Graph Attention and Transformers

Presented By:

Gautham Paspala

Staff Software Engineer,

ServiceNow

The Growing Threat of Friendly Fraud

Friendly fraud has become one of the most significant challenges facing the payments industry today. Unlike traditional fraud, these are legitimate transactions later disputed through chargebacks, making detection exceptionally difficult.

The Scale of the Problem

60–75%

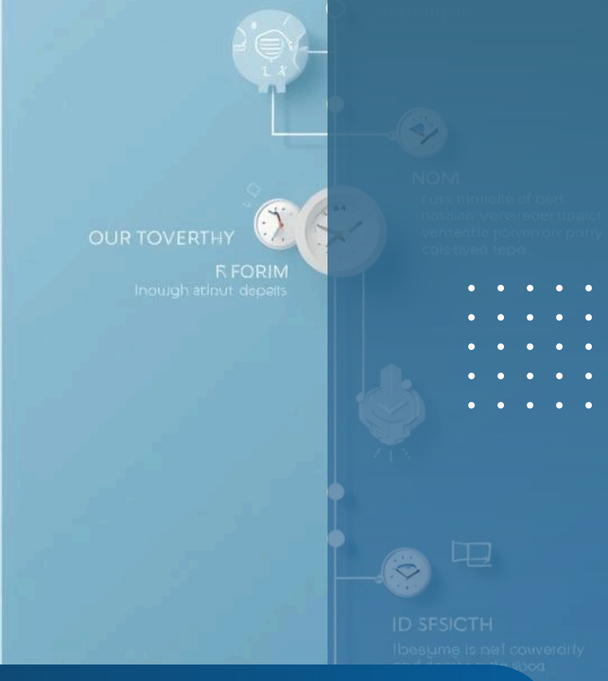
An alarming 60–75% of all chargeback volume is now attributed to friendly fraud, representing the majority of disputes that financial institutions must handle and investigate.

$48B+ Losses

The financial impact is staggering, with global losses exceeding $48 billion annually. This figure continues to grow as e-commerce expands and consumers become more aware of chargeback processes.

Understanding Friendly Fraud

Friendly fraud occurs when legitimate cardholders make valid purchases, then later dispute the charges through their bank. This creates a temporal disconnect that makes point-of-sale prevention nearly impossible.



- Legitimate Purchase Made by Cardholder
- Goods or Services Delivered
- Chargeback Filed Weeks Later

The delay between transaction and dispute can span weeks or months, giving fraudsters time to receive goods while planning chargebacks. Traditional controls cannot anticipate future disputes.

Limitations of Current Detection Approaches

Rule-Based Systems

Detection Rate: 23–31%
Relies on rigid, predefined rules that cannot adapt to evolving fraud patterns. Limited ability to capture complex behavioral signals or network relationships.

Traditional Machine Learning

Detection Rate: 38–47%
Constrained by manual feature engineering. Lacks capability to model network structures and temporal relationships inherent in fraud patterns.

Introducing Our Attention-Driven Framework

Our novel approach combines heterogeneous temporal transaction graphs with transformer-based sequence modeling to detect friendly fraud patterns that traditional systems miss.

Core Framework Components

Graph Component

Temporal Graph Attention Networks model complex relationships between cardholders, merchants, devices, and transactions. Graph attention highlights suspicious structural patterns like device-sharing and cross-merchant behaviors.

Transformer Component

Transformer architecture captures sequential behavioral signals across time windows. Self-attention mechanisms identify dispute targeting patterns and temporal anomalies in transaction sequences.

Representing Payment Ecosystems as Large-Scale Graphs

Modern fraud detection requires modeling the complex web of relationships between entities in payment ecosystems. By representing these connections as heterogeneous graphs, we capture structural patterns invisible to traditional approaches.

Graph entities extend beyond simple account-merchant pairs to include devices, IP addresses, physical addresses, and behavioral fingerprints. This rich representation enables detection of sophisticated fraud rings and shared-identity schemes.

- **Cardholders: 50–100M accounts**
- **Merchants: 15–25M entities**
- **Graph edges: 10–50B total**
- **Daily additions: 100–500M new connections**

Temporal Transaction Graph Structure

Our heterogeneous graph captures the complex relationships within payment ecosystems. Each node type represents a distinct entity, while edges encode behavioral and transactional connections that evolve over time.

- Accounts, Devices, and Merchants
- Transactions and IPs
- Physical and Digital Addresses

Edges represent temporal relationships: purchase transactions, device usage patterns, merchant visits, and address associations. This time-aware structure enables detection of evolving fraud patterns.

Graph Attention Mechanism

Graph attention networks dynamically weight relationships between entities, automatically highlighting suspicious structural patterns. By learning which connections matter most, the model identifies fraud signals that traditional approaches miss.

Relationship Weighting

Attention scores automatically highlight suspicious structural relationships between accounts, devices, and merchants in the transaction graph.

Device Sharing Detection

Identifies multiple accounts sharing devices or IP addresses a strong indicator of coordinated friendly fraud rings operating across the network.

Cross-Merchant Patterns

Detects repeated dispute patterns across different merchants, revealing systematic friendly fraud behavior that spans the payment ecosystem.

Transformer-Based Behavioral Sequence Modeling

Our transformer architecture processes transaction sequences to identify subtle behavioral patterns indicative of friendly fraud. Self-attention mechanisms weigh relationships between transactions across variable time windows, capturing dispute targeting behaviors.

Sequential Pattern Analysis

Time Windows

Captures sequential signals such as dispute targeting patterns over configurable time windows. The model identifies cardholders who systematically dispute transactions following specific purchase behaviors or merchant interactions.

Self-Attention

Self-attention layers model complex behavioral patterns within transaction sequences, learning which past transactions are most relevant for predicting future dispute likelihood without manual feature engineering.

Evaluation Metrics for Fraud Detection

Measuring model performance requires carefully selected metrics that balance detection accuracy with operational costs. Our framework employs industry-standard evaluation approaches benchmarked against rule-based and traditional ML systems.

Precision & Recall

Precision measures the accuracy of fraud predictions, while recall captures the proportion of actual fraud cases detected. Balancing these metrics is critical to minimize both false positives and missed fraud.

AUC Score

Area Under the ROC Curve provides a threshold-independent measure of model discrimination ability. Higher AUC indicates better separation between legitimate transactions and friendly fraud cases.

Cost Optimization

Cost-sensitive threshold optimization aligns detection thresholds with business impact, weighing the cost of false positives against missed fraud losses to maximize operational value.

Benchmark Results Summary

Our attention-driven framework demonstrates significant performance improvements over traditional detection methods, with measurable gains in precision, recall, and overall detection accuracy across multiple fraud categories.

Performance improvements over baselines

Detection Rate Gains

Detection rate increased from 38-47% (traditional ML) to 67-78% with our graph attention approach a 40-65% relative improvement in identifying friendly fraud cases.

False Positive Reduction

False positive reduction of 35-45% compared to rule-based systems, significantly lowering operational costs and improving customer experience during dispute resolution.

Cloud-Native Deployment Challenges

Deploying attention-driven fraud detection at enterprise scale introduces significant infrastructure challenges. Building and maintaining temporal transaction graphs with billions of edges requires distributed processing capabilities that can handle continuous data streams while preserving graph integrity and query performance.

The complexity of managing heterogeneous graph databases alongside transformer inference pipelines demands sophisticated orchestration. Systems must balance computational costs, latency requirements, and data freshness while maintaining high availability across distributed cloud environments.

- Scalable graph construction at issuer scale
- Massive data ingestion (10-50B+ edges)
- Real-time processing needs
- Infrastructure complexity and orchestration

Cloud-Native Solutions and Architecture

Our fraud detection framework leverages modern cloud-native principles to achieve scalability, reliability, and real-time processing capabilities required for enterprise-grade deployment at issuer scale.

Distributed Processing

Distributed graph databases enable parallel processing of billion-edge transaction networks. Partitioned storage and compute clusters handle massive data volumes efficiently.

Segment Calibration

Segment-level model calibration improves accuracy across diverse merchant categories and transaction patterns. Localized thresholds reduce false positives significantly.

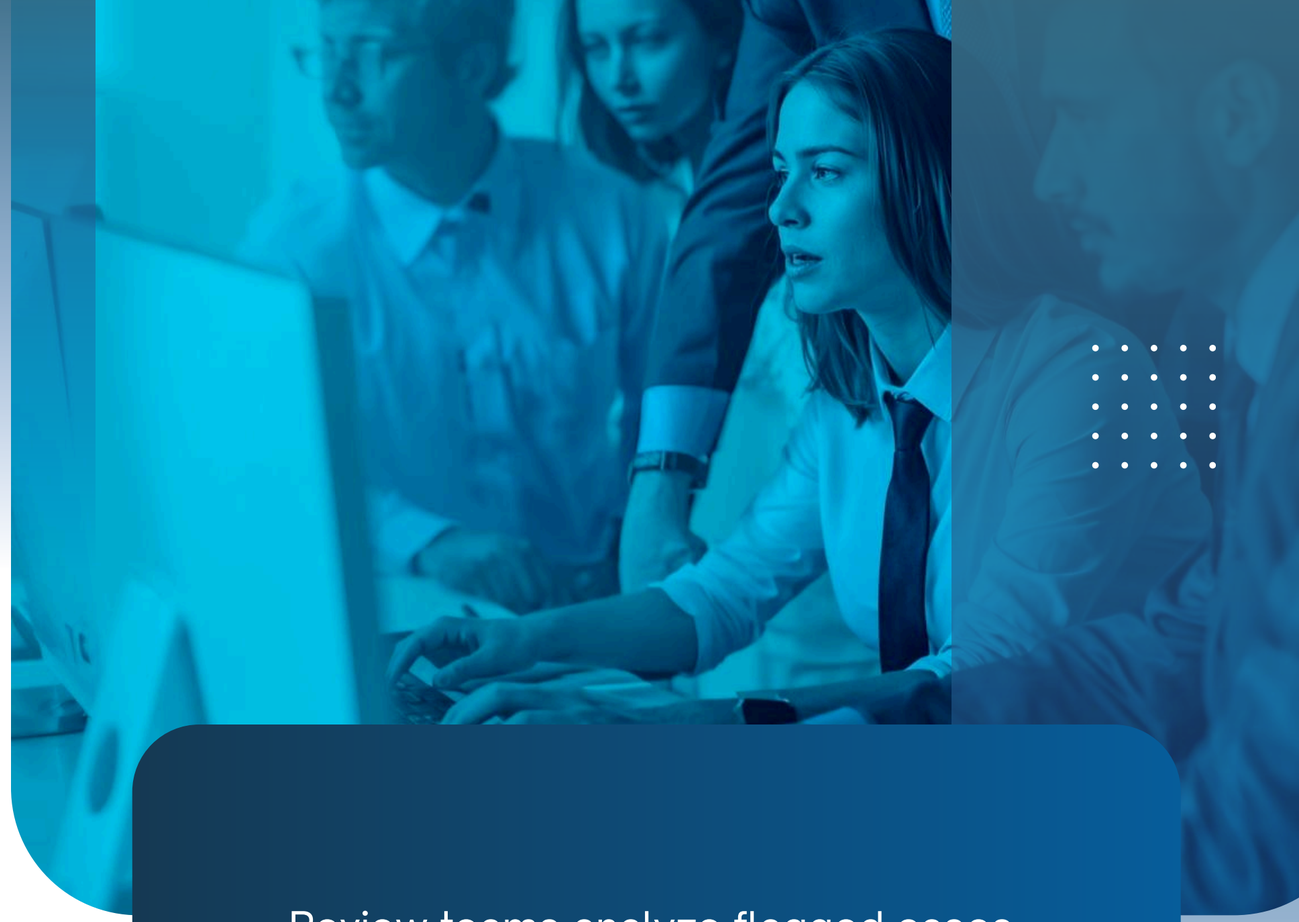
Auto-Scaling

Kubernetes-based autoscaling and fault tolerance ensure consistent performance during traffic spikes. Self-healing infrastructure maintains 99.9% uptime for critical fraud detection.

Human-in-the-Loop Review Workflows

Effective fraud detection requires balancing automation with human expertise. By incorporating expert review feedback loops, organizations can continuously refine detection models while ensuring legitimate customers aren't wrongly flagged.

- Reducing False Positives with Expert Review
- Minimizing Customer Friction
- Continuous Model Refinement via Feedback



Review teams analyze flagged cases, providing valuable feedback that improves model accuracy over time. This collaborative approach optimizes the balance between catching fraud and maintaining positive customer experiences.

Real-World Impact and Benefits

Our attention-driven fraud detection framework delivers measurable improvements across key business metrics, transforming how organizations combat friendly fraud while maintaining excellent customer relationships.

Increased detection accuracy and reduced financial losses

Enhanced customer experience by lowering false alerts

Scalable, maintainable cloud-native infrastructure

Summary and Key Takeaways

This presentation explored how combining temporal graph attention networks with transformer-based sequence modeling creates a powerful framework for detecting friendly fraud at scale. Cloud-native architecture enables the processing power needed for real-time detection across billions of transactions.

The attention-driven approach significantly outperforms traditional rule-based and ML systems by capturing complex relational patterns and behavioral sequences that indicate fraudulent dispute behavior.

- Friendly fraud is a growing, complex problem
- Graph + transformer approach improves detection
- Cloud-native enables scale
- Human-in-the-loop crucial for operational success

Thank You