Fraud Detection – Analytical Report

Overview

This analysis investigates potentially fraudulent activity by examining relationships between user identities, devices, IP addresses, and behavioral fingerprints. The goal is to uncover hidden connections that might indicate coordinated abuse or fraud across a digital banking environment.

Technologies and Libraries Used

- Pandas: For data loading, manipulation, and filtering.
- GeoIP2: To extract geographic and subnet information from IP addresses.
- User-Agents: For parsing browser and device information.
- NetworkX: For visualizing relationships in a graph structure.

These tools were chosen because they are reliable and widely used in data analysis and fraud detection.

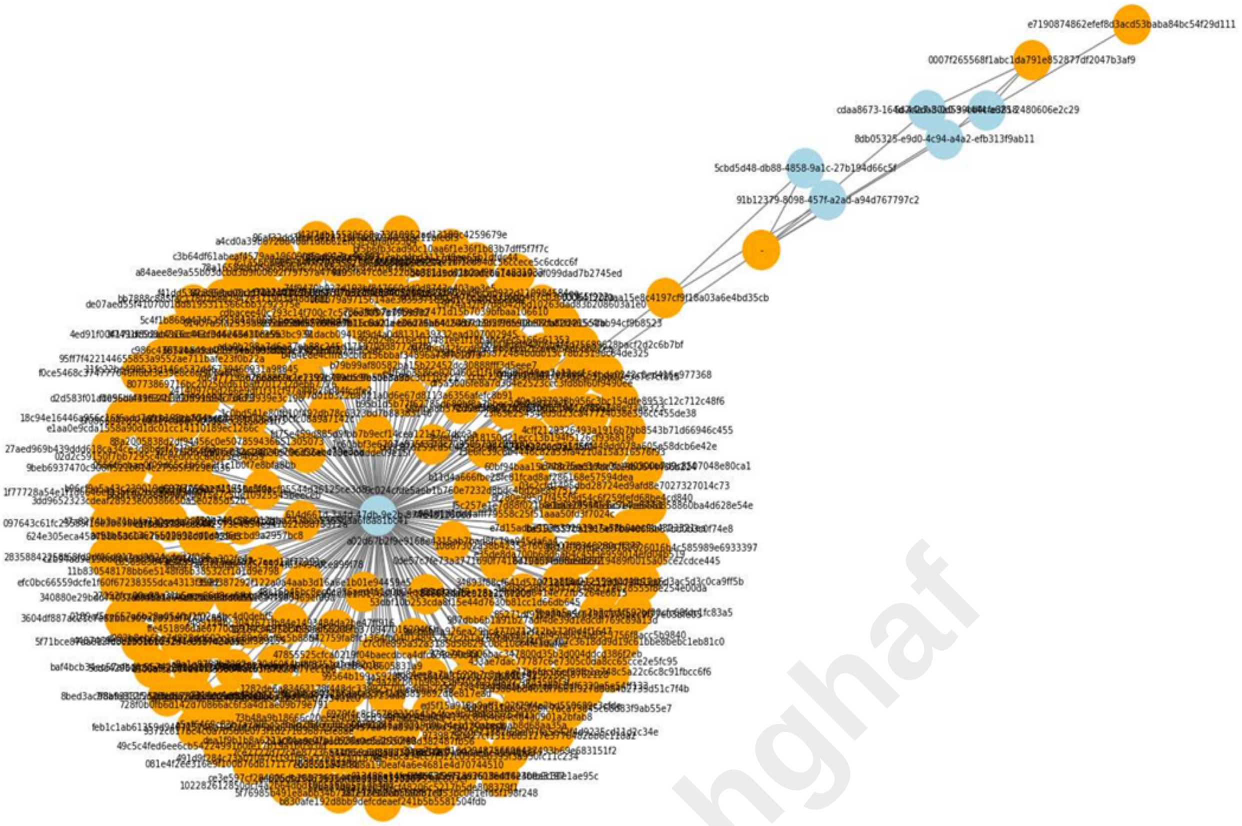
Analytical Process

1. Data Enrichment:
IP addresses and user-agent strings were enriched with geographic and device data.

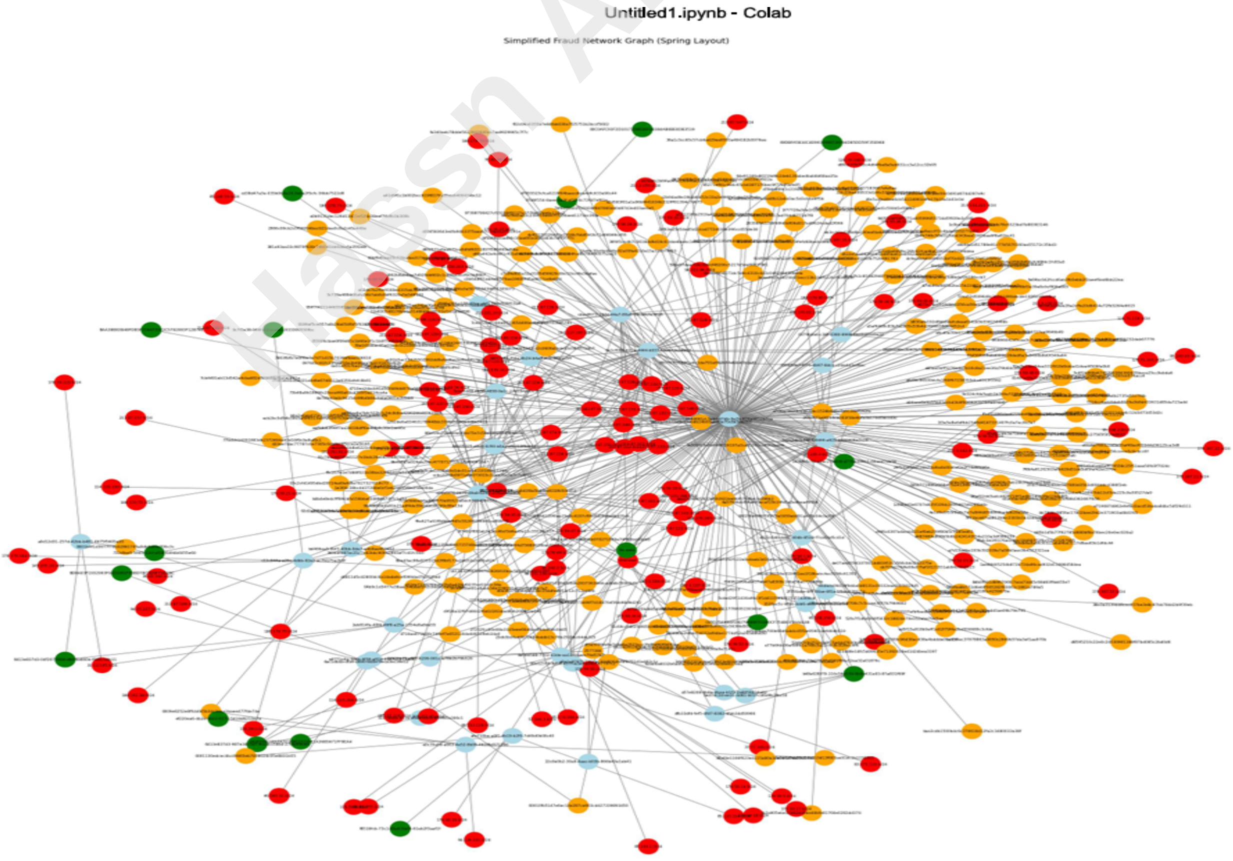
2. Finding Relationships:
Devices and identities sharing the same device fingerprint, subnet, screen resolution, or GPU renderer were linked.

3. Network Mapping:
We created two distinct network graphs:
- MAC Address-Identity Graph: Started from a compromised identity, then retrieved all MAC addresses linked to it, followed by all other identities sharing those MAC addresses.
- Full Device-Network Graph: Included nodes for devices, identities, fingerprints, and subnets. Edges represented shared attributes. This helped visualize the underlying structure of suspicious connections.

4. Tabular Summary:
A matrix was created showing each node (device, identity, fingerprint, subnet) and whether it was connected to known fraud nodes. We used the flag 't' to indicate a confirmed connection, while unconnected nodes were left as NaN to distinguish them from valid matches.


e7190874862efef8d3acd53baba84bc54f29d111
0007f265568f1abc1da791e852877df2047b3af9
2480606e2c29
8db05325-e9d0-4c94-a4a2-efb313f9ab11
5cbd5d48-db88-4858-9a1c-27b194d66c5f
91b12379-8098-457f-a2ad-a94d767797c2
a15e8c4197cf918a03a6e4bd35cb

MAC Address-Identity Graph



Full Device-Network Graph

- Blue nodes: Devices
- Orange nodes: Identities
- Red nodes: Subnets
- Green nodes: Fingerprints

Results & Key Insights

- Several user identities were connected via shared subnets and fingerprints.
- A single device was used across multiple accounts - a strong indicator of fraud.
- The network graphs revealed central devices/subnets acting as infrastructure for multiple suspicious accounts.

Conclusion

This analysis successfully uncovered a fraud network using both enriched data and graph-based exploration. It provides clear, actionable insight that can guide further investigation by security teams.

Recommendations

- Review identities connected through reused infrastructure.
- Monitor new accounts appearing with the same fingerprint or subnet.
- Automate this process to enable real-time fraud detection and early intervention.