Final Project Proposal

(CS5800 Fall 2023 Semester)

Real-Time Fraud Detection in Financial Transactions Using Cuckoo Filters

Arya Dhorajiya (NUID: 002780002)
Spandan Maaheshwari (NUID: 001525706)
Sudarshan Paranjape (NUID: 002817988)
Sumit Hawal (NUID: 002979933)

Project Context

Project Context in view of Arya:

With swindles and fraudulent activities causing havoc, we are diving into a project which involves detecting shady transactions. The rise in sneaky moves, from credit card shenanigans to identity theft, has left people and businesses counting losses and feeling skeptical about the whole financial scene. Unauthorized transactions and data breaches are like the new normal, and we need a solid defense. So, the motivation for this project is to figure out how these tricksters operate by analyzing transaction data and produce ways to stop them in their tracks. It is all about giving folks and businesses a shield against these digital troublemakers. By digging into the patterns these scammers use, the ultimate goal is to beef up security measures and make our online world a safer place. Concisely, it is about putting a dent in fraud, bringing back trust in online transactions, and giving cybercriminals a run for their money.

Project Context in view of Spandan:

My journey in the field of electronics has always been marked by a fascination with how complex systems manage and process data with remarkable efficiency. This curiosity has now led me to the exploration of Cuckoo Filters, a sophisticated data structure known for its dynamic capabilities in handling data sets. Unlike the static nature of Bloom Filters, Cuckoo Filters offer the unique advantage of allowing additions and deletions, making them particularly suited for environments where data is continually changing. This feature of Cuckoo Filters aligns perfectly with my growing interest in hashing and its myriad applications across various computer systems.

Our project aims to leverage the dynamic capabilities of Cuckoo filters for real-time fraud detection in financial transactions. We focus on their ability to quickly adapt to changing fraud patterns, due to their efficient and precise set membership testing. The goal is to develop a prototype that promptly identifies and responds to potential fraud. We will explore the working and application of Cuckoo filters, inspired by their use in tech giants like Meta. Ultimately, this project seeks to innovate in financial security, enhancing the reliability of such systems.

Project Context in view of Sudarshan:

Fraud detection has become increasingly crucial in various industries, including finance, insurance, and e-commerce. The prevalence of internet banking and online transactions necessitates heightened vigilance to safeguard against fraudulent activities. In the realm of real-time fraud detection, the primary objective is to swiftly identify and prevent fraudulent transactions as they occur. By harnessing transaction data, the goal is to develop an algorithm that effectively detects and mitigates fraudulent activities. Similarly, we try to implement a Fraud/Unauthorised transaction detection system using Cuckoo Filter. Cuckoo filters can be used for fraud detection by storing fingerprints of known fraudulent identifiers in the hash table. When a new transaction is received, its fingerprint is computed and compared to the fingerprints in the hash table. If the fingerprint is found in the hash table, then the transaction is flagged as potentially fraudulent.

Project Context in view of Sumit:

Recently while discussing HashMap’s and their different applications with my group members, one of our group members suggested Cuckoo filters. Cuckoo filters aim to improve the performance of filters used for high-speed set membership tests. They do

this by allowing alterations to the membership sets. We are thinking of applying this to financial transactions to understand the set memberships of the particular transaction to the whole set of transactions. The reason why this problem is complex is because of the rarity of the frauds that happen within millions of transactions, therefore simple machine learning algorithms fail at such tasks. My interest lies in Finance and Data Science, which is why I am excited to solve this problem at the intersection of anomaly detection and financial leveraging data structures.

Question

In the realm of real-time fraud detection within financial transactions, the primary question centers on optimizing the identification and prevention of fraudulent activities occurring in the moment. Given the availability of transaction data, the objective is to devise an algorithm that efficiently detects and mitigates fraudulent transactions. The algorithm must leverage advanced techniques such as the cuckoo filter and other machine learning models to rapidly identify anomalous patterns indicative of fraud. Considerations will extend beyond mere identification, incorporating the optimization of resources by prioritizing transactions based on the risk level, thereby minimizing the impact on legitimate users and ensuring the algorithm's scalability. Using the concepts, we learnt in CS5800 we want to devise an optimized solution that swiftly and accurately identifies real-time fraudulent transactions while efficiently utilizing resources and minimizing false positives.

Scope

1. The scope of this project is focused on developing and implementing a prototype based on algorithm of Cuckoo filter for real-time fraud detection in financial transactions.
2. Specifically, the project aims to examine the efficiency of leveraging advanced techniques, including the Cuckoo filter and other machine learning models, to swiftly identify and prevent fraudulent activities as they occur.
3. The algorithm will consider transaction data, aiming to optimize the identification

process and minimize false positives. Additionally, the project will explore the prioritization of transactions based on risk levels to enhance resource utilization and scalability.

4. The project does not intend to delve into the utilization of State-Of-The-Art Ribbon filters, or other complex machine learning models or explore additional features beyond user locations and transaction data. The focus remains on delivering a practical, efficient, and accurate real-time fraud detection solution.

5. Research: Examine current methodologies in fraud detection and the use of probabilistic data structures in similar applications.
6. Algorithm Design: Architect a system that integrates Cuckoo filters for real-time analysis of financial transactions.
7. Implementation: Develop a functional prototype of the username validation system using Bloom filters.
8. Performance Evaluation: Evaluate the efficiency and accuracy of the Cuckoo filter focusing on detection accuracy and processing speed.

Description

Cuckoo Filters, an advanced and efficient data structure, are particularly well-suited for Real-Time Fraud Detection in Financial Transactions due to their unique characteristics and operational advantages. In the fast-paced and dynamic environment of financial transactions, the ability to determine set membership quickly and accurately is crucial. Cuckoo Filters excel in this regard, offering a robust mechanism to ascertain if a transaction or entity is part of a set of known fraudulent patterns or not. Unlike Bloom Filters, Cuckoo Filters have the added advantage of supporting the dynamic deletion of items. This feature is vital in the financial sector, where fraud patterns constantly evolve, and data sets require frequent updates without incurring significant overhead.

Operationally, Cuckoo Filters manage data through compact representations called "fingerprints," efficiently stored in a hash table-like structure. These fingerprints are small bit strings, efficiently derived from transactions using hash functions. The dense packing of these fingerprints in the filter contributes to high space efficiency, a critical factor when dealing with large volumes of financial data. Cuckoo Filters perform

dynamic insertions, lookups, and deletions effectively, making them adaptable to the rapidly changing landscape of financial fraud. While they do have a certain rate of false positives, Cuckoo Filters ensure no false negatives, a feature crucial in fraud detection where missing a fraudulent transaction can have significant repercussions.

In the context of Real-Time Fraud Detection in Financial Transactions, Cuckoo Filters are incredibly versatile. They can quickly adapt to new types of fraudulent activities by adding or removing transaction patterns. This adaptability makes them superior to other data structures that are static and cannot evolve as quickly with changing fraud trends. Financial institutions can leverage Cuckoo Filters to maintain and update a dynamic database of transactional fingerprints, which can be used to cross-reference incoming transactions for potential fraud. This cross-referencing is done with minimal latency, an essential requirement for real-time processing in finance.with large volumes of financial data. Cuckoo Filters perform dynamic insertions, lookups, and deletions effectively, making them adaptable to the rapidly changing landscape of financial fraud. While they do have a certain rate of false positives, Cuckoo Filters ensure no false negatives, a feature crucial in fraud detection where missing a fraudulent transaction can have significant repercussions.

Implementing Cuckoo Filters for fraud detection in financial transactions offers a system that is not only efficient and scalable but also highly responsive to new fraud patterns. The design and deployment of such a system necessitates careful consideration of the dynamic nature of financial fraud, ensuring that the system remains effective over time. By adopting Cuckoo Filters, financial institutions stand to enhance their fraud detection mechanisms significantly, making them more adept at identifying and responding to fraudulent activities swiftly. This approach marks a significant step forward in the realm of financial security, providing a sophisticated tool to combat the ever-evolving challenge of financial fraud.

Once we have gone through understanding of the Cuckoo Filter's internal mechanics, the project will shift towards developing a specialized algorithm tailored for real-time fraud detection in financial systems. This algorithm will be designed to strike a delicate balance between rapid detection of fraudulent transactions and minimizing false positives, which are particularly critical in the financial domain. The ability of Cuckoo Filters to dynamically update their dataset makes them ideally suited for this application, where fraudulent patterns can evolve rapidly. Rigorous testing will be integral to this

phase, ensuring the algorithm not only detects fraud effectively but also adapts to new patterns of fraudulent behavior swiftly. This adaptability is crucial for maintaining the relevance and efficacy of the fraud detection system over time.

Our end goal is to deliver a well-documented, scalable, and highly efficient system that leverages the unique capabilities of Cuckoo Filters for real-time fraud detection in financial transactions. This system is expected to significantly enhance the accuracy and responsiveness of fraud detection mechanisms in financial institutions. By successfully completing this project, we aim to contribute to the advancement of financial security technologies, providing a robust tool against the dynamic and sophisticated nature of financial fraud. Additionally, this project will enrich our practical experience and understanding of complex data structures like Cuckoo Filters, furthering our knowledge in the realms of algorithm design, system performance optimization, and efficient data management.

References

1. https://en.wikipedia.org/wiki/Cuckoo_filter
2. https://en.wikipedia.org/wiki/Bloom_filter
3. https://www.cs.cmu.edu/~dga/papers/cuckoo-conext2014.pdf
4. Bin Fan, Dave G. Andersen, Michael Kaminsky, and Michael D. Mitzenmacher. 2014. Cuckoo Filter: Practically Better Than Bloom. In Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies (CoNEXT '14). Association for Computing Machinery, New York, NY, USA, 75–88. https://doi.org/10.1145/2674005.2674994
5. Peir, Jih-Kwon & Lai, Shih-Chang & Lu, Shih-Lien & Stark, Jared & Lai, Konrad. (2002). Bloom filtering cache misses for accurate data speculation and prefetching. Proceedings of the International Conference on Supercomputing. 189-198. 10.1145/514191.514219.
6. P. K. Sadineni, "Detection of Fraudulent Transactions in Credit Card using Machine Learning Algorithms," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2020, pp. 659-660, doi: 10.1109/I-SMAC49090.2020.9243545.