JCSE International Journal of Computer Sciences and Engineering Open Access
Research Volume-6, Issue-6 E-ISSN: 2347-2693

Fraud Pattern Recognition In Banking Sector Using Graph Database

Sonali Sen[1], Trishita Mukherjee[2], Sunanda Pal[3], Sumana Ghosh[4]

[1,2,3,4]Department of Computer Science, St. Xavier's College [Autonomous] Kolkata, Calcutta University, Kolkata, India

**Corresponding Author: sonalisen@sxccal.edu, Tel.: 9830616421*

Available online at: www.ijcseonline.org

Accepted: 13/Jun/2018, Published: 30/Jun/2018

***Abstract*—** Bank sector gives the proper economic structure and support in a country. Frauds in the banking sector have become a major issue in the banking arena. Therefore, it has become a necessity in implementing fraud pattern detection mechanisms to unmask the fraudsters. Ideal mechanism of recognizing such fraud patterns can be implemented using a graphical structure. Graph database provides such a graphical structure with node-relationship analysis. Typical pattern fraudulent methods like bust-out fraud (BOF) and credit-card fraud (CRF) can be recognized via such a graphical structure analysis. The motive behind such a proposal is to detect fraudulent patterns and implement transaction analysis in a bust-out fraud and credit-card fraud. We are trying to observe possible fraud rings in the bust-out fraud and in the credit-card fraud we are trying to identify the origin of the scam. This proposal provides the ideal solution for the investigation of large amounts of heterogeneous data that is required to recognize the fraudulent patterns in the bust-out and credit-card fraud.

***Keywords*—** Graph database, Bust-out fraud, Credit card fraud.

I. Introduction

The fraudsters have spread their tentacles in the banking sector. Banks have incurred a number of losses in the recent years. The traditional fraud detection methods have become almost obsolete as too many fraud cases occurred in recent years. Intelligent fraudsters easily find out innovative ways to trick such system by masking their identities. The traditional fraud detection methods use behavior profiling and anomaly detection in the banking sector. RDBMS is used currently in the banking sector in India. There is currently no data visualization database present in the current database management system. Therefore, recognizing such ring patterns becomes difficult. Introduction of graph database leads to the conception of a graphical structure that is much more flexible than the current database that is used in the banking sector. In such a case data visualization plays an important role in detecting such fraud patterns. We have taken certain references from some papers to get a concept of existent fraud patterns in India. In this proposal we have merged both the bust-out fraud and credit card fraud detection with transaction analysis.

In case of Bust-out fraud, the fraudsters forge the first identity by combining real contact information such as phone number, home address, Aadhaar number or pan number etc., with some fake documents such as name, address, date-of-birth, age etc. That is they fraudulently map those real information [1]. They create the synthetic identities which seem like an existent one. The criminals are smart enough to hide their identity so that they can bypass the security checks of their targets. A verifiable piece of identity is useful in such a case. A smart fraudster becomes effective at reusing information and creating synthetic identities with less effort. In this way, the criminals form a "ring". Then they open accounts in the bank and initially over a certain period of time they use those accounts as legitimate ones with regular purchases and timely payments. Bank issues them credit cards and personal loans (which don't have the mortgage scheme) when they ask for it as they previously exhibited responsible behavior. Then one day they bust out that is they disappear without paying any debts. This causes bank a huge financial loss as they have to write off the uncollectible debt [2]. Through graph database we can expose such hackers who use synthetic identities.

In case of credit card fraud, credit card data can be stolen by a fraudster using a myriad of methods. A fraudster copies the credit card information (name, expiration date, cvv no, etc.) using a skimmer and a hidden spy cam can capture the pin (when the user is entering it), of that legitimate credit card holder and then uses it for his/her own benefits. Fraud pattern detection in credit card fraud can be implemented by unearthing the point of origin of the scam. We identify the fraudulent transactions and then try to discover the suspicious store from where the hacker had stolen the victim's credit card information. Our point in this case is to find the vulnerable merchants from where most of the credit information is hacked. These fraudsters can either be the clerks, the salesman or the bank employees [3]. Graph

© 2018, IJCSE All Rights Reserved

database is extremely efficient in handling such fraud patterns.

In both the cases we see the use of credit cards. In bust out fraud the fraudsters use credit cards as regular ones and repays on-time initially. But suddenly they purchase with large amounts and without any intention of paying back they disappear. In credit card fraud several observations state that for fraudulent transactions there are difference in behavioural pattern between the legitimate ones and the fraudulent ones. Which is basically determined based upon the customer's previous transaction history (the amount they purchased and the time they took to repay) and recent purchases. But a customer may have different transaction patterns which are actually the legitimate ones. A legitimate customer may have purchase history of small amounts and as well as large amounts. Also a criminal who is in bust out fraud, may create transaction histories with large amounts with paid those back. In such cases behavioral pattern recognition will not give a high success rate in fraud detection. Rather we can just focus on the amount because a criminal normally goes for higher amount to commit fraud and not for a small amount. He can defraud by doing several small amount transactions or one or two big amount transactions. So the question arises that how to determine the large amount transaction. It is basically the credit limit ratio that one individual is spending through his card. According to the experts, using 50% of the credit limit is definitely considered as large purchase and some people have the opinion that it is 20%.

This model implements transaction analysis and detects fraudulent patterns in a bust-out and credit card fraud. The proposed model will further be discussed in the proposed work, results and conclusion.

Rest of the paper is organized as follows; Section I contains the introduction of fraud pattern in banking sector. Section II contains proposed work of detecting different fraud patterns, where section 2.1 describes the steps to detect bust-out fraud, section 2.2 explains the essential steps in detecting the origin of credit card fraud, and section 2.3 contains the transaction analysis for both of the fraud cases and section 2.4 explains the working methodology of the total system in proper steps. Section III of the paper gives the results where section 3.1 describes the results related to Bust-out fraud section 3.2 gives the results of credit card fraud detection and section 3.3 explains the results about transaction analysis. Section IV of the paper concludes the research work.

II. RELATED WORK

As the frauds have different patterns we use graph database to recognize those patterns by analyzing the links between different entities.

Graph database is a database management system which is used to create, read, update and delete operations. It gives a graphical representation of the whole database in the form of nodes and relationships. It uses only nodes and relationships to represent the data in the database. Nodes and relationships may have some properties which describes themselves. The nodes have specific labels and properties associated with them. Cypher query language (CQL) is used to write the queries for the subsequent graph database [4].

In the proposed system, the following conceptual model is used to create the graph database. As, you can see from the following conceptual mode, the following node-relationships are used to describe the graph database. A database is created and then a certain number of queries are designed to run to get the desired results [5].

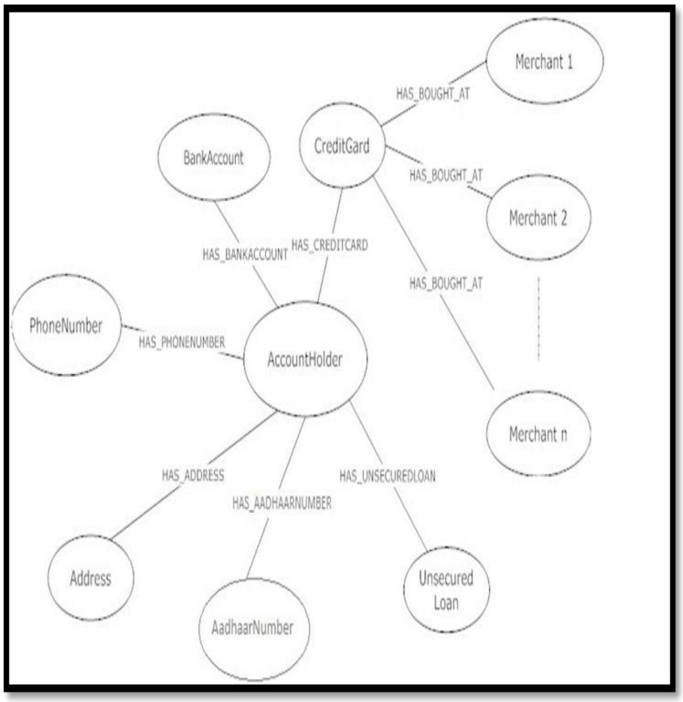


Fig.1. CONCEPTUAL MODEL OF THE DATABASE

2.1 BUST-OUT FRAUD DETECTION

For the ring fraud, we are calculating the number of rings that are formed by a criminal or a group of criminals and then we are also calculating the financial risk that the bank suffers from this fraud when the ring busts-out. The financial risk is the summation of the accountholder's credit card's limit and the loan amount minus the balance of the account of corresponding loan. The accountholders that are participating in a ring is also displayed. As discussed before, the accountholders in a ring shares a subset of legitimate contact information (like phone numbers, addresses, and Aadhaar or pan number) including some of the synthetic information to create a fraudulent identity. The fraudster maps the fraudulent data with the original documents. They then open accounts and ask for credit cards and unsecured loans from the bank. The bank issues it to them. Then, suddenly one day the ring "busts-out", that is the fraudster vanishes off and the bank incurs a huge financial loss and the uncollectible debt is written off.

Checking for bust out fraud is totally dependent on the entity link analysis. We are finding out which accountholders share same contact information and we are considering them as ring. Also we are calculating the maximum possible financial risk [6].

1) PSEUDO ALGORITHM.

```
Step1: Match account holders with the contact information which gives the accountholders who have any contact information in common.
Step2: Count the account holders as the size of ring.
Step3: If the size is greater than 1
        Then
                Check if account holder has credit card
                        Then take credit limit
                Check if account holders has loan
                        Then take (loan amount-account balance)
                Sum up the credit limit with (loan amount-account balance) as the possible risk
Step4: Return the rings with size and possible risk
```

2.2 CREDIT-CARD FRAUD DETECTION

In case of the credit card fraud, the transactions are represented as a graph. Then, we can look for the common denominator/merchant who is involved in the fraud case and also find the point of origin of the scam. The credit card and the merchant node are linked by the transaction itself. The transaction contains the following properties-a date, amount and status. The status can either be fraudulent or legitimate. Before a fraudulent transaction takes place, a legitimate transaction need to be done from where the hacker has copied the credit card information of that particular legitimate credit card user. We calculate the number of customers from a store from where their credit card information has been stolen by the hacker that is we are trying to find out the origin of the fraud.

2) PSEUDO ALGORITHM.

```
Step1: match the credit cards which did transaction at merchants (say merchants1) where the transactions are "fraudulent"
Step2: match those credit cards which did transaction at merchants (say merchants2) where the transactions are "legitimate" and are done before the "fraudulent" ones
Step3: now the common credit cards are taken
Step4: return the merchants2 as the probable vulnerable merchants
Step5: return the cardholders of those common cards as the persons
```

2.3 TRANSACTION ANALYSIS FOR BOTH THE CASES

As discussed before in both of the cases usage of credit card is noticeable. Traditional behavioural pattern recognition system for credit card fraud detection may falter when a customer has history of various small and as well as large amount of legitimate purchasing with different timings of paying the debt and also if a fraudster in bust out fraud intentionally creates transaction histories of large amount and repaying those before making another large amount of purchase and getting vanished. A fraudster will always tend to make up large amount of money by cheating. Hence we are mainly aiming at the amount of money for each transaction. As said earlier spending 50% of credit limit is definitely a huge debt. And even 20% of credit line is also considered to be a big amount. Sudden buying of big amount or not paying a huge amount after a certain period of time is definitely considered to be a bad move. [7]

Hence to check for the credit behaviour for each individual customer, we propose a scoring system where the set Score has three values. 1 for positive score, -1 for negative score, 0 for suspicious score.

Score = {-1, 0, 1}

Credit balance is the amount of money a customer needs to pay to the bank. And the balance is modified every time the customer purchases anything with the card or he is making payment in the bank or any interest is added with that balance.

Now if the credit balance is greater than or equals to 50% of credit limit then we will be assigning negative score. If the credit balance is greater than or equals to 20% of credit limit and less than 50% of credit limit then we will be assigning suspicious score, otherwise a positive score will be assigned. A negative score means a high chance of fraud case whereas a positive score indicates not a fraud case. A suspicious score indicates that it has a tendency to be a fraud case but it may not be one.

We will include a property ‘previous_credit_score’ in the node credit card in order to keep track of the previous balance modification score which will help in later cases. The property ‘previous_credit_score’ indicates the score assigned to the credit card when the balance was last updated. We will now determine the current score of the credit card status based upon the credit balance and credit limit.

3) CURRENT SCORE DETERMINATION PSEUDO ALGORITHM.

```
Step 1: Begin
Step 2: Check if credit balance >= 50% of credit limit
                If true then return negative score i.e. -1
Step 3: Else check if credit balance >= 20% of credit limit
                If true then return suspicious score i.e. 0
Step 4: Else return positive score i.e. 1
Step 5: End
```

We can merge this concept of score of credit card with bust out fraud detection and credit card fraud detection for better result.

2.4 WORKING OF THE TOTAL SYSTEM

```
Step 1:  a.  Start bust out fraud checking
         b. Start parallel checking of credit card fraud with the transaction status at hand i.e. which is fraudulent and which is legitimate
         c. Start calculating current score for every credit card parallely
Step 2: Give corresponding results of bust out fraud and credit card fraud detection
Step 3: If the current score is -1 then
         Step 3.1: Investigate to check whether fraudulent transaction or legitimate transaction
         Step 3.2: If fraudulent transaction
                  Then go to credit card fraud checking
               Else
                  Check for bust out fraud
                        If ring formation is caught
                              Then Potential fraud
                        Else
                              Investigate the contact details provided
                              If not correct then fraud
                              Else not fraud
                              End
                        End
               End
Step 4: If the current score is 0 then
         Step 4.1: If prev_credit_score is 1
                  Then wait for the payment
         Step 4.2: Else
                  Go to Step 3.1 and do the similar checking and follow Step 3.2
                  End
Step 5: If the current score is 1
           Then no fraud check
                        Step 6: End
```

III. Results

3.1 RESULT FOR BUST-OUT FRAUD

In the bust-out fraud detection, a graphical representation is given below in fig2. From the graphical representation we can see that the green nodes represent the account holders, the red ones represent the address of a particular accountholder, the yellow ones represent the phone number of an accountholder and the blue ones represent the AadhaarNumber of an accountholder. The nodes are connected with the following relationships: HAS_ADDRESS, HAS_PHONENUMBER and HAS_AadhaarNumber.

The result of the following will be as follows:

Ring: [Bivash, Sourav, Rajiv] size: [3] Contact Information: [AadhaarNumber, Phone Number]

Ring: [Bivash, Sourav] size: [2] Contact Information: [Address]

The result with the contact information AadhaarNumber will be a potential fraud.

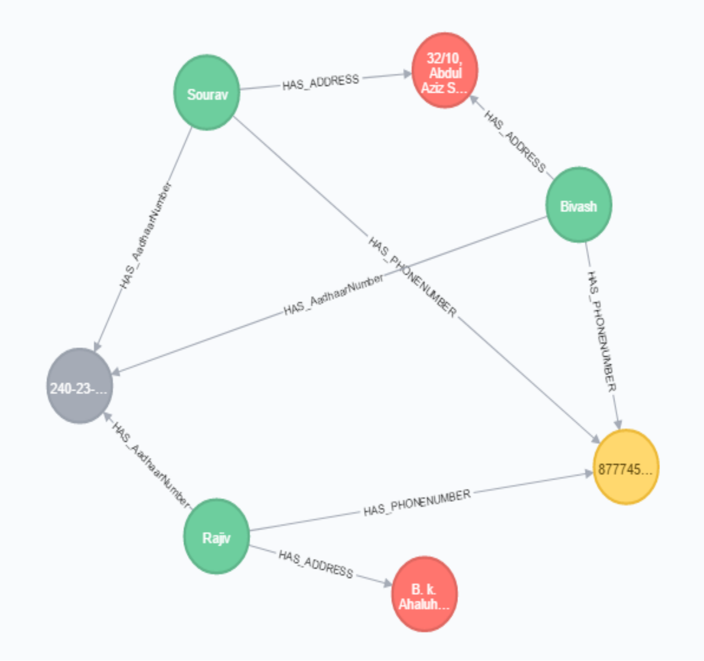
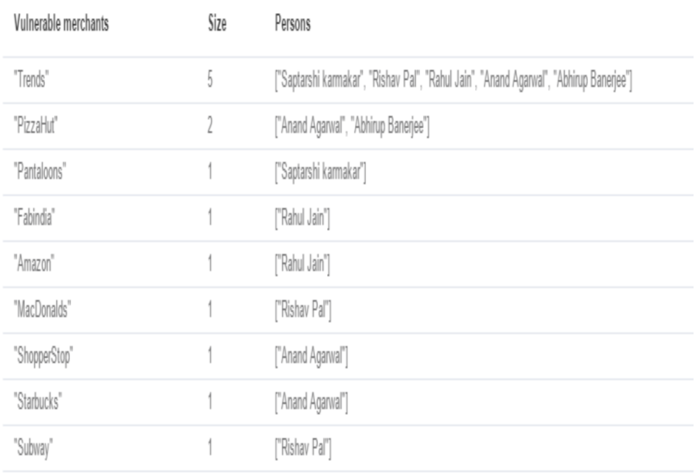


Fig2. BUST-OUT FRAUD GRAPHICAL REPRESENTATION

3.2 RESULT FOR CREDIT-CARD FRAUD

In the credit-card fraud detection, a graphical representation is given below in fig3. From the graphical representation we can see that the green nodes represent the account holders, the red ones represent the stores/merchants, and the pink ones represent the credit card of an accountholder. The nodes are connected with the following relationship: Did_transaction_at.

The result of the following will be as follows:



Vulnerable merchants	Size	Persons
"Trends"	5	["Saptarshi karmakar", "Rishav Pal", "Rahul Jain", "Anand Agarwal", "Abhirup Banerjee"]
"PizzaHut"	2	["Anand Agarwal", "Abhirup Banerjee"]
"Pantaloons"	1	["Saptarshi karmakar"]
"Fabindia"	1	["Rahul Jain"]
"Amazon"	1	["Rahul Jain"]
"MacDonalds"	1	["Rishav Pal"]
"ShopperStop"	1	["Anand Agarwal"]
"Starbucks"	1	["Anand Agarwal"]
"Subway"	1	["Rishav Pal"]

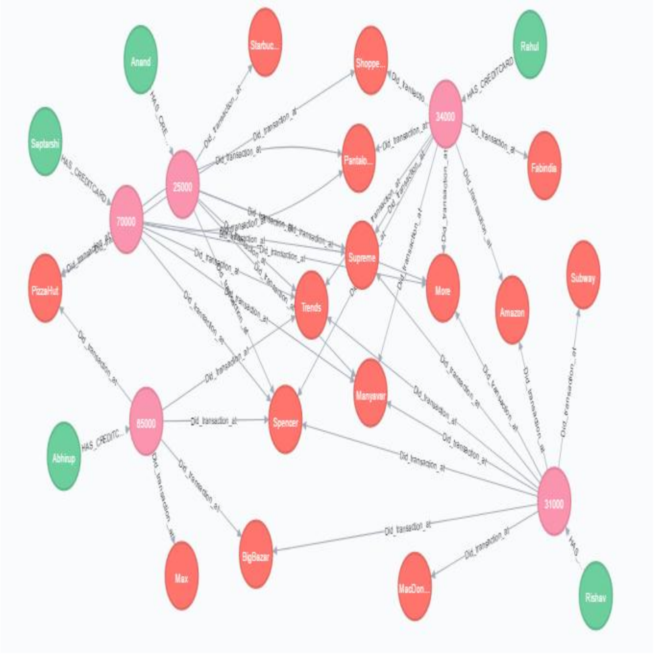


Fig3.CREDIT CARD FRAUD GRAPHICAL REPRESENTATION

3.3 RESULT FOR TRANSACTION ANALYSIS FOR BOTH THE CASES

For determining the score of a credit card, let a card limit is 100000 and card balance is 60000 then current score will be -1. For the same credit limit if the balances are 45000 and 5000 respectively it will give the score as 0 and 1 respectively. This analysis is applied both for the bust-out and the credit-card fraud detection.

Therefore the detection model gives the following result based on a given data set.

IV. CONCLUSION AND FUTURE SCOPE

Our proposal suggests identifying the recognition pattern of frauds related to the banking sector. The two kinds of fraud that we chose to discuss in our proposal are the Bust-out fraud and the credit card fraud. We have also implemented transaction analysis for both of the cases. The proposal covers the solutions to the problems of fraudulent cases in the banking sector. Also, in the current bank database system only behavior profiling is done, which does not gives a higher success rate. This approach is very simple and straightforward. Therefore, introducing this model in the database of banking system will give more efficient results in fraud detection.

REFERENCES

[1]. Harsha R. Vyawahare, Dr. P. P. Karde, An Overview on Graph Database Model, Vol. 3, Issue 8, August 2015, IJIRCCE.

[2]. Shefali Patil, Gaurav Vaswani, Anuradha Bhatia, Graph Databases- An Overview , IJCSIT, Vol. 5(1), 2014,657-660

[3]. Arnaud Castelltort, Anne Laurent, Rogue behavior detection in NoSQL graph databases. Journal of Innovation in Digital Ecosystems, Elsevier 2016, 3 (2), pp.70-82.

[4]. In: Saeed K., Homenda W ,Pokorný J. (2015) Graph Databases: Their Power and Limitations, (eds) Computer Information Systems and Industrial Management. CISIM 2015. Lecture Notes in Computer Science, vol. 9339. Springer, Cham

[5]. Graph Databases – Book by Emil Eifrem, Ian Robinson, and Jim Webber,-O'REILLY

[6]. Harsha R Vyavahare, Dr.P.P.Karde, SHORT SURVEY ON GRAPHICAL DATABASE, ISSN: 0976-5166 Vol. 6 No.4 Aug-Sep 2015-IJCSE.

[7]. R.Satraboyina, G.K Chakravarthi, Discovery of ranking fraud detection system for mobile apps-, vol.4, Issue 4, p.p.7-10, Aug-2016-IJSRCSE.