# Abnormal Pattern Analysis in Online Transaction

Parikshith Nayaka S K
Prof. Dept. of Computer Science and Engineering
Alva's Institute of Engineering & Technology
Mangalore, India

Jeffin Boban
Dept. of Computer Science and Engineering
Alva's Institute of Engineering & Technology
Mangalore, India

Aishwarya K
Dept. of Computer Science and Engineering
Alva's Institute of Engineering & Technology
Mangalore, India

Arjun V
Dept. of Computer Science and Engineering
Alva's Institute of Engineering & Technology
Mangalore, India

Bhargavi Ravi
Dept. of Computer Science and Engineering
Alva's Institute of Engineering & Technology
Mangalore, India

***Abstract*— The Abnormal pattern analysis in online transaction features uses the behavior of the user and verifies the location to check for unusual patterns. If any unusual pattern is detected, the system re-verification. The System analyzes user credit card data for various characteristics. The drawback of the current E-bank log system was that it was not able to achieve or was not very secure for the users. There were security issues faced in the current E-bank log system. Thus, we implement the Sequential Pattern Matching Algorithm to overcome the security issue being faced in the current system.**

**Keywords—** ***Unusual Pattern, Sequential Pattern Matching Algorithm***

## I. INTRODUCTION

Internet Banking is being the current trend in India for money transactions. The number of users are increasing daily in India. The more people do transactions on Internet Banking, there are chances for hacking, which is a security issue. Security is most important in such kind of systems. The user is given with a username and password and the OTP is sent to the user's mobile phone at the time of transactions in the current system. The internet banking can be done in Smartphones as well. The disadvantage of the current system is that anyone who is having the username and password of the user can access the account. Also if the unauthorized person uses the user's mobile, he can get to know the OTP which is sent to the mobile during transaction.

Sequential pattern mining was first introduced by R. Agrawal and R. Srikanth in [1]. It aims at discovering subsequences as patterns in a sequence database. Since then, sequential pattern mining has become an important data mining task. There are several applications of sequential pattern mining including mining customer shopping sequences, DNA sequences and Web click streams.

Online banking requires strong user authentication. User authentication is achieved by two-factor authentication based on the user i.e a static password and an OTP. The advantage of involving a mobile phone is that most users have mobile phones, therefore no extra hardware token is needed to be bought, or supported or deployed. The traditional system sends an OTP over an SMS to user who wants to do an online transaction.

## II. LITERATURE SURVEY

### A. *Sequential Pattern Analysis*

Sequential pattern mining algorithm such as Generalized Sequential Patterns [2], Prefix Span [3], SPADE [3] and SPAM [3] are more popular in data mining research. Generalized Sequential Patterns and SPADE are based on breadth-first search, whereas Prefix Span and SPAM are based on depth-first search. There are two algorithms ,ColSpan[3] and BIDE [3] in closed sequential pattern mining .ColSpan produces a candidate set for closed sequential pattern performs post pruning on it. ColSpan requires more storage to store the sequence when mining long patterns or when the support threshold is low and it offers poor scalability. BIDE adopts PrefixSpan and uses BackSpan pruning method to stop growing redundant patterns. BIDE is computational intensive approach.

### B. *One Time Password*

One Time Password, also known as dynamic password. It is a unpredictable serial of numbers generated by a specified

algorithm. One password only works one time, so it's a effective way to avoid account being theft.

Technically speaking, OTP has three formats:(1)Time Synchronization; based on the time compare of dynamic token and dynamic password, the token based on time synchronization usually generate a new password every 60 seconds.(2)Event Synchronization; the server and the token use as specified sequence of events and the same seed value to compute a password using a hash algorithm.(3)Challenge/Respond; server generate a challenge number and send it to token, they both compute password using the challenge number and some parameters which are prearranged.

## III. IMPLEMENTATION

The unit proposed, the unauthorized user when logged in is blocked by analyzing their pattern using Sequential Pattern Mining Algorithm or system sends an OTP to the user's mobile phone. The 4 digit OTP along with a 2 digit code which user sets while registration is to be entered in order to access the account.

The current System was not secure enough for the users. The fraud users could easily access the genuine users account by having their user name and password. The current system could not trace out whether the user was a genuine user or a fraud. The security level in terms of identifying the fraud was not found in the current system.

## IV. METHODOLOGY

The working of the various components is explained below.

### *A. HOTPAlgorithm*

HMAC-based One-time Password algorithm (HOTP) is a one-time password (OTP) algorithm based on HMAC (hash-based message authentication code). HOTP algorithm provides method of authentication by symmetric generation of human readable password, each used for only one authentication attempt. One time password leads directly from the single use of counter value.

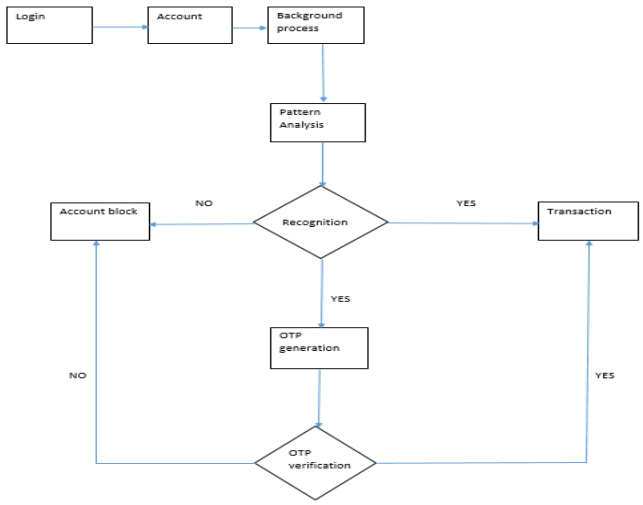


Figure 1. System Model

Users willing to use HOTP must establish parameters; typically these are specified by the authenticator, and either accepted or not by authenticatee:

i) A cryptographic hash method is represented, H (default SHA-1)

ii) A secret key, K, which is an arbitrary byte string, and must remain private

iii) A HOTP value length, d (default and minimum is 6, and the recommendation is 6–8)

Both parties compute the HOTP value, then the authenticator checks locally-generated value against the value supplied by the authenticate.

The HOTP value is the human readable design output,d-digit decimal number (without omission of leading 0s):

$$\text{HOTP value} = \text{HTOP}(K,C) \bmod 10^d$$

That is ,the value is d least significant base-10 digits of HTOP.

HTOP is a hashed message authentication code of the counter C (under the key K and hash function,H)..

$$\text{HOTP}(K, C) = \text{truncate}(\text{HMAC}_H(K, C))$$

Abridge takes four least significant bits of MAC and uses it as an offset, i.

abridge(MAC) = extract(MAC, MAC[156:159] *8)
Index i is used to select 31 bits from MAC, starting at bit i+1.
extract(MAC, i) = MAC[i + 1:i + (4 × 8) − 1]

iv) Note that 31 bits is single bit short of a 4-byte word. Thus, value can be placed inside such a word without using the sign bit(the most significant bit). This is done to avoid doing modular arithmetic on negative numbers, as this has many differing definitions and implementations.

### *B. TOTP Algorithm*

Time based One Time Password algorithm(TOTP) is an extension of the HMAC based one time password algorithm (HOTP) generating one time password by instead taking uniqueness from the current time. It has been adopted as internet engineering Task Force[1] standard RFC 6238,[1] is the initiative for open authentication (OATH), and is used in a number of two factor authentication systems.

To establish TOTP authentication, both parties must agree on both HOTP parameters and the additional TOTP parameters:

i) $T_0$, Unix time from which to start counting time steps(default is 0).

ii) $T_X$, interval will be used to calculate the value of the counter $C_T$ (default is 30 seconds).

iii) Both the server and client compute the TOTP value, then the server checks if the TOTP value supplied by the client matches the locally-generated TOTP value. Some servers allow values that should been generated before or after the current time in order to account for slight clock skews, network latency and user delays.

iv) TOTP uses rhe HOTP algorithm, substituting the counter with non-decreasing value based on the current time.

v) TOTP value(K) = HOTP value(K, $C_T$)

vi) The time counter, $C_T$, is an integer counting number of durations, $T_X$, is the difference between the current unix time,T, and some ($T_0$; cf. Unix epoch); the latter values all being in integer seconds.

vii) $$C_T = \left\lfloor \frac{T - T_0}{T_X} \right\rfloor$$

viii) Note that Unix time is not monotonic; specifically, when leap seconds are inserted into UTC.

### C. *PrefixSpan Algorithm*

This algorithm is another form projection based algorithm. The idea to check only prefix subsequences and only their corresponding postfix subsequences are projected into databases, rather than projecting sequence database. PrefixSpan uses a direct application of apriori property to reduce the candidate sequences alongside the databases..

**PrefixSpan(α, i, S|α)**
**Begin**
1. Scan S|α once, find the set of frequent items b such that
   - b can be assembled to the last element of α to form a sequential pattern; or
   - <b> can be appended to α to form a sequential pattern.
2. For each frequent item b, appended it to α to form a sequential pattern α', and output α';
3. For each α', construct α'-projected database S|α', and call PrefixSpan(α', i+1,S|α').

**End**

Additionally, PrefixSpan is efficient beacause it mines the complete set of patterns and has a significantly faster running. Major cost of PrefixSpan,is the construction of projected databases. PrefixSpan needs to construct a projected database. After database projection is done, the use of bi-level projection represents FreeSpan and PrifixSpan by the S-Matrix is a faster way to mine. The main idea of PrefixSpan algorithm, is to use frequent prefixes to divide search soace and project sequence databases.It mains to search the relevant sequenses.

### D. *DFS-Prunning*

The algorithm adopting this feature show an ineffective pruning method and engender a great number of candidate sequences, which requires consuming a lot of memory in early stages of mining.

SPAM algorithm uses a depth-first traversal method combined with a vertical bitmap representing to store each sequence allowing a significant bitmap compression as well as efficient support counting.

### E. *Closed Sequential Pattern*

This algorithm is used to reduce the time and space cost when generating explosive numbers of frequent sequence patterns.

CloSpan mines only frequent closed subsequenses (the sequences containing no super sequence with the same support), instead of mining the complete set of frequent subsequences.

The mining process used by CloSpan is divided into two stages. A candidate set is generated in the first stage which is larger than the final closed sequence set. This set is called suspicious closed sequence set (the superset of the closed sequence set). A pruning method is called in the second stage to eliminate non-closed sequences.The main difference between ColSpan PrefixSpan is the implementation ColSpan which are an early termination mechanism that avoids the unnecessary traversing of search space.

DFS-Pruning (node $n = (s_1, \ldots, s_k)$, $S_n$ , $I_n$)
**Begin**
(1) $S_{temp} = \varphi$.
(2) $I_{temp} = \varphi$.
(3) **For each** ($i \in S_n$)
(4) **if** (($s_1, \ldots, s_k$ , {$i$}) is frequent)
(5) $S_{temp} = S_{temp} \cup \{i\}$
(6) **For each** ($i \in S_{temp}$)
(7) DFS-Pruning(($s_1, \ldots, s_k, \{i\}$),$S_{temp}$, all elements in $S_{temp}$ greater than $i$ )
(8) **For each** ($i \in I_n$)
(9) **if** (($s1, \ldots, sk$ » {$i$}) is frequent)
(10) $Itemp = Itemp \cup \{i\}$
(11) **For each** ($i \in I_{temp}$)
(12) DFS-Pruning (($s1, \ldots, sk \cup \{i\}$), $S_{temp}$, all elements in $I_{temp}$ greater than $i$)
**End**

The use of backward sub pattern and backward super pattern methods, some patterns will be absorbed or merged which, indeed reduce the search space growth.

## V. CONCLUSION

In this paper,abnormal pattern analysis in online transaction process was discussed. This can be used by the current banking system processes for added data security and transactions. This paper has overcome the drawbacks of security issues in the account log in and transaction by the OTP authentication process.One drawback that could be viewed in the paper is the data sets. More the data sets the system process is slow.

REFERENCES

[1] Mining Closed Sequential Patterns Using Genetic Algorithm V. Purushothama Raju\G.P. Saradhi Varma2 lDepartment of Computer Science & Engineering, Shri Vishnu Engineering College for Women,Bhimavaram, A.P., India 2Department of Information Technology, S.R.K.R. Engineering College,Bhimavaram, A.P., India.

[2] R. Srikant and R. Agrawal,"Mining sequential patterns : Generalizations and performance improvements," Proceedings of EOBT '96, pp. 3-17, Mar. 1996. Analysis and Design of E-banking Data Downloading.

[3] Sequential mining: patterns and algorithms analysis Thabet Slimani1, and Amor Lazzez2 Computer Science, Taif University & LARODEC Lab, Saudia Arabia, 2 Computer Science, Taif University , Saudia Arabia,thabet.slimani@gmail.com,a.lazzez@gmail.lazzez@gmail.com