Volume 3, Issue 7, July 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Synthetic Identity of Crime Detection

T.P.Latchoumi
Assistant Professor,
Department of Computer Science and Engineering
CCET-Pondicherry University
Pondicherry- India

V.M.Vijay Kannan
Assistant Professor,
Department of Computer Science and Engineering
Erode Sengunthar Engineering College,
Perundurai-India.

Abstract - Synthetic identity of crime detection is a specific case of identity crime. Now-a-days online transaction is a tedious task because of online frauds. Online transaction is suited for both credit card and debit card application is a specific case of identity crime. The existing non-data mining detection system of business rules and scorecards, and known fraud matching have limitations. To address these limitations and combat identity crime in real time, in this paper proposes three levels of security methods. First level is generating short term password, second level is an arising security questions and the third level is to prevent the SQL injection attack (i.e.) Incidents of this nature put the identity of customers at risk as hackers are able to access their personal data, sometimes including credit card details, email addresses and passwords. If an incoming credit card transaction is not accepted by the trained value with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. Each incoming call sent into the bank database, matching the all attributes of the user is done to generate the security code on the user mobile. If the user enters the security code on the application, then it provides the way for transaction. Suppose, if the mobile is in disconnection state or tower less state the cardholder must answer all the security questions and do the transaction. In case of cardholder affected by online frauds by using request and response time to prevent the SQL injection attacks.

Keywords- Data mining based fraud detection, security, anomaly detection, data stream

I. Introduction

Identity crime is defined as broadly as possible in this paper. At one extreme, synthetic identity fraud refers to the use of plausible but fictitious identities. These are effortless to create but more difficult to apply successfully. At the other extreme, real identity theft refers to illegal use of innocent people's complete identity details. In reality, identity crime can be committed with a mix of both synthetic and real identity details. Identity crime has become prominent because there is so much real identity data available on the Web, and confidential data accessible through unsecured mailboxes. Credit applications are Internet or paper-based forms with written requests by potential customers for credit cards, mortgage loans, and personal loans. Credit application fraud is a specific case of identity crime, involving synthetic identity fraud and real identity theft. As in identity crime, credit application fraud has reached a critical mass of fraudsters who are highly experienced, organized, and sophisticated. Their visible patterns can be different to each other and constantly change. Based on anecdotal observations of experienced credit application investigators, fraudsters can use software automation to manipulate particular values within an application and increase frequency of successful values. To detect the crime in credit card applications. To address these limitations and combat identity crime in real time, this paper proposes a new multilayered detection system complemented with two additional layers: communal detection (CD) and spike detection (SD). CD finds real social relationships to reduce the suspicion score, and is tamper resistant to synthetic social relationships. It is the white list-oriented approach on a fixed set of attributes. SD finds spikes in duplicates to increase the suspicion score, and is probe-resistant for attributes. It is the attribute-oriented approach on a variable-size set of attributes. To address these limitations and combat identity crime in real time, in this paper proposes three levels of security methods. First level is generating short term password, second level is an arising security questions and the third level is to prevent the SQL injection attack (i.e.) Incidents of this nature put the identity of customers at risk as hackers are able to access their personal data, sometimes including credit card details, email addresses and passwords. If an incoming credit card transaction is not accepted by the trained value with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. Each incoming call sent into the bank database, matching the all attributes of the user is done to generate the security code on the user mobile. If the user enters the security code on the application, then it provides the way for transaction. Suppose, if the mobile is in disconnection state or tower less state the cardholder must answer all the security questions and do the transaction. In case of cardholder affected by online frauds by using request and response time to prevent the SQL injection attacks. Credit -card -based purchases can be categorized into two types: 1) physical card and 2) virtual card. In a physical-card based purchase, the cardholder presents his card physically to a merchant for making a payment. To carry out fraudulent transactions in this kind of purchase, an attacker has to steal the credit card. In the second kind of purchase, only some important information about a card (card number, expiration date, secure code) is required to make the payment. Such purchases are normally done on the Internet or over the telephone. To commit fraud

in these types of purchases, a fraudster simply needs to know the card details. Most of the time, the genuine cardholder is not aware that someone else has seen or stolen his card information. In this paper proposes three levels of security methods. First level is generating short term password, second level is an arising security questions and the third level is to prevent the SQL injection attack (i.e.) Incidents of this nature put the identity of customers at risk as hackers are able to access their personal data, sometimes including credit card details, email addresses and passwords. If an incoming credit card transaction is not accepted by the trained value with sufficiently high probability, it is considered to be fraudulent. Each incoming call sent into the bank database, matching the all attributes of the user is done to generate the security code on the user mobile. If the user enters the security code on the application, then it provides the way for transaction. Suppose, if the mobile is in disconnection state or tower less state the cardholder must answer all the security questions and do the transaction. In case of cardholder affected by online frauds by using request and response time to prevent the SQL injection attacks.

II. Related Works

In this paper to detect the crime in credit card application Resilient Identity Crime Detection; in other words, the real-time search for patterns in a multilayered and principled fashion, to safeguard credit applications at the first stage of the credit life cycle. This paper describes an important domain that has many problems relevant to other data mining research. It has documented the development and evaluation in the data mining layers of defence for a real-time credit application fraud detection system. In doing so, this research produced three concepts (or "force multipliers") which dramatically increase the detection system's effectiveness (at the expense of some efficiency) [7]. These concepts are resilience (multilayer defence), adaptively (accounts for changing fraud and legal behaviour), and quality data (real-time removal of data errors). Resilience is the ability to degrade gracefully when under most real attacks. The basic question asked by all detection systems is whether they can achieve resilience. To do so, the detection system trades off a small degree of efficiency (degrades processing speed) for a much larger degree of effectiveness (improves security by detecting most real attacks). In fact, any form of security involves tradeoffs [20]. The detection system needs "defence-in-depth" with multiple, sequential, and independent layers of defence [25] to cover different types of attacks. These layers are needed to reduce false negatives. In other words, any successful attack has to pass every layer of defence without being detected.

The two greatest challenges for the data mining-based layers of defence are adaptively and use of quality data. These challenges need to be addressed in order to reduce false positives. Adaptivity accounts for morphing fraud behavior, as the attempt to observe fraud changes its behavior. But what is not obvious, yet equally important, is the need to also account for changing legal (or legitimate) behavior within a changing environment. In the credit application domain, changing legal behavior is exhibited by communal relationships (such as rising/falling numbers of siblings) and can be caused by external events (such as introduction of organizational marketing campaigns) [19]. This means legal behavior can be hard to distinguish from fraud behavior, but it will be shown later in this paper that they are indeed distinguishable from each other. The detection system needs to exercise caution with applications which reflect communal relationships. It also needs to make allowance for certain external events. Quality data are highly desirable for data mining and data quality can be improved through the real time removal of data errors (or noise) [24]. The detection system has to filter duplicates which have been reinterred due to human error or for other reasons. It also needs to ignore redundant attributes which have many missing values, and other issues.

Automated adversarial detection systems can fail when under attack by adversaries. As part of a resilient data stream mining system to reduce the possibility of such failure, adaptive spike detection is attribute ranking and selection without class-labels. The first part of adaptive spike detection requires weighing all attributes for spiky-ness to rank them [18].The second part involves filtering some attributes with extreme weights to choose the best ones for computing each example's suspicion score. Within an identity crime detection domain, adaptive spike detection is validated on a few million real credit applications with adversarial activity. The results are F-measure curves on eleven experiments and relative weights discussion on the best experiment [8]. The results reinforce adaptive spike detection's effectiveness for class-label-free attribute ranking and selection.

As one of the most pervasive methods of individual identification and document authentication, signatures present convincing evidence and provide an important form of indexing for effective document image processing and retrieval in a broad range of applications [17]. However, detection and segmentation of free-form objects such as signatures from clustered background is currently an open document analysis problem. In this paper, we focus on two fundamental problems in signature-based document image retrieval [16]. First, we propose a novel multi-scale approach to jointly detecting and segmenting signatures from document images. Rather than focusing on local features that typically have large variations, our approach captures the structural saliency using a signature production model and computes the dynamic curvature of 2-D contour fragments over multiple scales [20]. This detection framework is general and computationally tractable. Second, we treat the problem of signature retrieval in the unconstrained setting of translation, scale, and rotation invariant non-rigid shape matching. We propose two novel measures of shape dissimilarity based on anisotropic scaling and registration residual error, and present a supervised learning framework for combining complementary shape information from different dissimilarity metrics using LDA [6]. We quantitatively study state-of-the-art shape representations, shape matching algorithms, measures of dissimilarity, and the use of multiple instances as query in document image retrieval. We further demonstrate our matching techniques in off-line signature verification. Extensive experiments using large real world collections of English and Arabic machine printed and handwritten documents demonstrate the excellent performance of our approaches.

In computing it may be required to generate permutations of a given sequence of values. The methods best adapted to do this depend on whether one wants some randomly chosen permutations, or all permutations, and in the

latter case if a specific ordering is required [6]. Another question is whether possible equality among entries in the given sequence is to be taken into account; if so, one should only generate distinct multiset permutations of the sequence. An obvious way to generate permutations of n is to generate values for the Lehmer code (possibly using the factorial number system representation of integers up to $n!$), [14] and convert those into the corresponding permutations. However, the latter step, while straightforward, is hard to implement efficiently, because it requires n operations each of selection from a sequence and deletion from it, at an arbitrary position; of the obvious representations of the sequence as an array or a linked list, both require (for different reasons) about $n^2/4$ operations to perform the conversion [11]. With $n$likely to be rather small (especially if generation of all permutations is needed) that is not too much of a problem, but it turns out that both for random and for systematic generation there are simple alternatives that do considerably better [12]. For this reason it does not seem useful, although certainly possible, to employ a special data structure that would allow performing the conversion from Lehmer code to permutation in $O(n \log n)$ time.

III. System Analysis

People nowadays rely heavily on the Internet since conventional activities or collaborations can be achieved with network services (e.g., web service). Widely deployed web services facilitate and enrich several applications such as online banking, social networks, cloud computing and ecommerce.

Existing Crime Detection

In case of the existing system the fraud is detected after the fraud is done that is, the fraud is detected after the complaint of the card holder. And so the card holder faced a lot of trouble before the investigation finish. And also as all the transaction is maintained in a log, we need to maintain a huge data. And also now a day's lot of online purchase are made so we don't know the person how is using the card online, we just capture the IP address for verification purpose. So there need a help from the cyber crime to investigate the fraud. To avoid the entire above disadvantage we propose the system to detect the fraud in a best and easy way. The first existing defence is made up of business rules and scorecards. The second existing defence is known fraud matching. The existing non data mining detection system of business rules and scorecards, and known fraud matching have limitations. To detect the crime in credit card application by applying Communal detection and Spike detection algorithm.

Communal Detection

CD is derived from white-list oriented approach. In these approach makes a link to the multivalve attributes. Suppose there were two credit card applications that provided the same postal address, home phone number, and date of birth, but one stated the applicant's name to be John Smith, and the other stated the applicant's name to be Joan Smith. These applications could be interpreted in three ways:

1. Either it is a fraudster attempting to obtain multiple credit cards using near duplicated data.
2. Possibly there are twins living in the same house who both are applying for a credit card.
3. Or it can be the same person applying twice, and there is a typographical error of one character in the first name.

With the CD layer, any two similar applications could be easily interpreted as (1) because this paper's detection methods use the similarity of the current application to all prior applications (not just known frauds) as the suspicion score. However, for this particular scenario, CD would also recognize these two applications as either (2) or (3) by lowering the suspicion score due to the higher possibility that they are legitimate. To account for legal behavior and data errors, CD is the white list-oriented approach on a fixed set of attributes.

Table.1: Sample of Six Credit Applications with Six Attributes

i or j	Given name	Family name	Unit no.	Street name	Home phone no.	Date of birth
1	John	Smith	1	Circular road	91234567	1/1/1982
2	Joan	Smith	1	Circular road	91234567	1/1/1982
3	Jack	Jones	3	Square drive	93535353	3/2/1955
4	Ella	Jones	3	Square drive	93535353	6/8/1957
5	Riley	Lee	2	Circular road	91235678	5/3/1983
6	Liam	Smyth	2	Circular road	91235678	1/1/1982

The white list, a list of communal and self-relationships between applications, is crucial because it reduces the scores of these legal behaviours and false positives. The white list is constructed by ranking link-types between applicants by volume.

Spike Detection

SD complements CD. Before proceeding with a description of SD, it is necessary to reinforce that CD finds real social relationships to reduce the suspicion score, and is tamper resistant to synthetic social relationships. It is the white list-oriented approach on a fixed set of attributes. In contrast, SD finds spikes to increase the suspicion score, and is probe resistant for attributes. Probe resistance reduces the chances a fraudster will discover attributes used in the SD score calculation. It is the attribute-oriented approach on a variable-size set of attributes. A side note: SD cannot use a white list-oriented approach because it was not designed to create larger the volume for a link-type, the higher the probability of a communal relationship.

However, there are two problems with the white list. First, there can be focused attacks on the white list by fraudsters when they submit applications with synthetic communal relationships. Although it is difficult to make definitive statements that fraudsters will attempt this, it is also wrong to assume that this will not happen. The solution proposed in this paper is to make the contents of the white list become less predictable. The values of some parameters (different from an application's identity value) are automatically changed such that it also changes the white list's link types. In general, tampering is not limited to hardware, but can also refer to manipulating software such as code. For our domain, tamper resistance refers to making it more difficult for fraudsters to manipulate or circumvent data mining by providing false data. Second, the volume and ranks of the white list's real communal relationships change over time. To make the white list exercise caution with (or more adaptive to) changing legal behavior, the white list is continually being reconstructed.

New Crime Detection and Prevention

Synthetic identity of crime detection is a specific case of identity crime. Now-a-days online transaction is a tedious task because of online frauds. Online transaction is suited for both credit card and debit card application is a specific case of identity crime. The existing non-data mining detection system of business rules and scorecards, and known fraud matching have limitations. To address these limitations and combat identity crime in real time, in this paper proposes three levels of security methods. First level is generating short term password, second level is an arising security questions and the third level is to prevent the SQL injection attack (i.e.) Incidents of this nature put the identity of customers at risk as hackers are able to access their personal data, sometimes including credit card details, email addresses and passwords. If an incoming credit card transaction is not accepted by the trained value with sufficiently high probability, it is considered to be fraudulent. At the same time, we try to ensure that genuine transactions are not rejected. Each incoming call sent into the bank database, matching the all attributes of the user is done to generate the security code on the user mobile. If the user enters the security code on the application, then it provides the way for transaction. Suppose, if the mobile is in disconnection state or tower less state the cardholder must answer all the security questions and do the transaction. In case of cardholder affected by online frauds by using request and response time to prevent the SQL injection attacks. To increase the security levels by using three methods 1. To secure short time password generation, 2. Security questions generation, 3.Prevent SQL injection attacks.

A. Security Short Time Password Generation

A physical random number generator can be based on an essentially random atomic or subatomic physical phenomenon whose unpredictability can be traced to the laws of quantum mechanics. However, physical phenomena and tools used to measure them generally feature asymmetries and systematic biases that make their outcomes not uniformly random. A randomness extractor, such as a cryptographic hash function, can be used to approach a uniform distribution of bits from a non-uniformly random source, though at a lower bit rate. Various imaginative ways of collecting this entropic information have been devised. One technique is to run a hash function against a frame of a video stream from an unpredictable source.

Pseudo-random number generators (PRNGs) are algorithms that can automatically create long runs of numbers with good random properties but eventually the sequence repeats (or the memory usage grows without bound). The string of values generated by such algorithms is generally determined by a fixed number called a **seed.** One of the most common PRNG is the linear congruential generator, which uses the recurrence $X_{n+1} = (aX_n + b) \bmod m$ to generate numbers. The maximum number of numbers the formula can produce is the modulus, *m*. To avoid certain non-random properties of a single linear congruential generator, several such random number generators with slightly different values of the multiplier coefficient *a* can be used in parallel, with a "master" random number generator that selects from among the several different generators. Most computer programming languages include functions or library routines that purport to be random number generators. They are often designed to provide a random byte or word, or a floating point number uniformly distributed between 0 and 1.

B. **Security Questions Generation**

Sensitive security questions are reasonably easy to reason about. The answers are generally infeasible to guess, and so attackers must somehow learn them. Users generally know who has access to their PIN number or bank account number, and can change them relatively easily if they suspect compromise. In contrast, Social Security numbers have only limited utility for authentication. They are frequently compromised in institutional data losses, and are frequently sold in bulk on the black market. The Social Security Administration has a policy of not assigning individuals a new number unless supplied with proof of fraud. As a result, Social Security numbers cannot be reset to a "secret" value between disclosure and attack. Therefore, the pool of Social Security numbers available to attackers is likely to grow over time.

However, Social Security numbers are by no means useless for security: they are likely able to defeat casual attackers, such as curious acquaintances, who might be able to guess or learn answers to personal security questions. Further, requiring them will raise the bar somewhat on identity theft, at fairly modest cost to the user and to the bank. In contrast,

personal security questions are comparatively difficult to analyze. The questions themselves are far more varied than sensitive security questions, and attackers can learn or guess the answers in a variety of ways. Such questions, though, are commonly used in practice, and thus demand careful consideration and analysis.

Most institutions that rely on personal security questions allow their users to choose the questions for which they will supply answers. However, at a few institutions, one or more security questions are mandated by the mechanism designer. In particular, four require users to specify their date of birth to authenticate, three require a ZIP code, and one requires the user's mother's maiden name. These questions have the benefit of having unambiguous answers for most users.

C. To Prevent Online Attacks

Injecting a Web application is the synonym of having illegal access to the data stored in the database. The data sometimes could be confidential and of high value like the financial secrets of a bank or list of financial transactions or secret information of some kind of information system. An unauthorized access to this data by a crafted user can pose threat to their confidentiality, integrity, and authority. As a result, the system could bear heavy loss in giving proper services to its users or it may even face complete destruction. Sometimes such type of collapse of a system can threaten the existence of a company or a bank or an industry. If it happens against the information system of a hospital, the private information of the patients may be leaked out which could threaten their reputation or may become a case of defamation. Attackers may even use such type of attack to get confidential information that is related to the national security of a country.

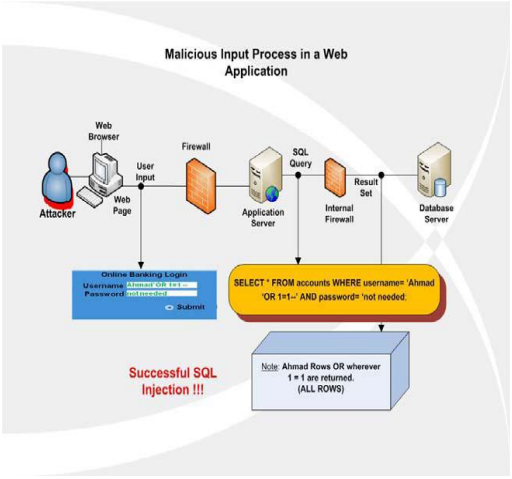


Fig.1: Malicious input process in a Web application

In case of SQL Injection, those schemes which work for preventing SQL Injection also do the curing of the system (or application) in early stage. Hence, in plain term, we could call the schemes 'countermeasures'. A strong countermeasure can remove or at least block all the available vulnerabilities in a system and thus it could protect it against various types of attacks that take advantage of the vulnerabilities.

The prevention in these stored procedures is implemented by a combination of static analysis and runtime analysis. The static analysis used for commands identification is achieved through stored procedure parser and the runtime analysis by using a SQL Checker for input identification.

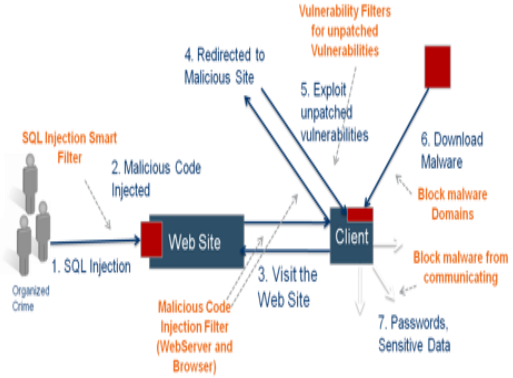


Fig.2: Detecting and Preventing SQL Injection Attack

IV.SYSTEM DESIGN

SQL injection detection and random password generation for prevent the attacks. This chapter includes the mechanism of Multifactor Authentication where the username and the long-term password is entered into the un-trusted computers by the clients. The long-term password entered generates the Short-term passwords for that particular session alone and it expires once the client logs out.

Moreover for improved security against the long-term password, an image is also given as a password. The client can enter into the website only after entering both the long-term password and the image. This Multifactor authentication consists of separate modules for each and every operation. In this section, we introduce a method in which the user can authenticate over the data to prevent from attacks.

This paper consists of four modules

1. User Authentication phase
2. Attack detection phase
3. Attack prevention phase
4. Security phase
5. Recovery phase

1 User Authentication Phase

In this phase we stock up the details of username and password in the database. Users also clearly identify their long-term password which will be used by user through the entire trusted browsing. It balances the given details of username and password with the details stored in the database. It will login if the details are correct. Then it originates the short term password which is used as a onetime password. Therefore the short-term should be altered for every successive trust browsing.

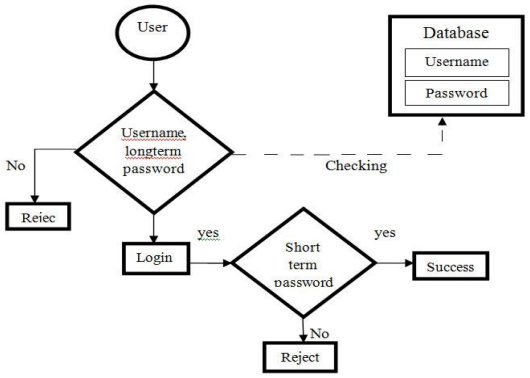


Fig.3 User Authentication Phase

2 Attack detection phase

In this phase by using SQL injection attacker queries put it on to the username and password. It result display all the user information in the database. So that to detect it by using encoding techniques. Using this technique attackers allow to the user account but the user details is hidden. It results will be unpredictable.

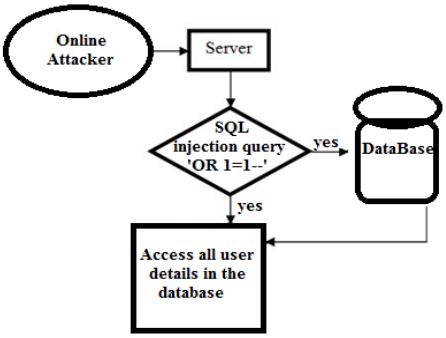


Fig.4 SQL injection attack

3 Attack Prevention Phase

This phase endow with defence to our data in many ways from the hacker from being slashed. If the hacker comes to know the long-term password of the user, then the website is prone to be attacked. Though the hacker knows the long-term password, the image which is also used as the password for login must be entered. Even the hacker manage to trace the image, the password code of the mobile is an integral part of the authentication process. So this is believed to be the security phase in this project in order to improve the security of the data.

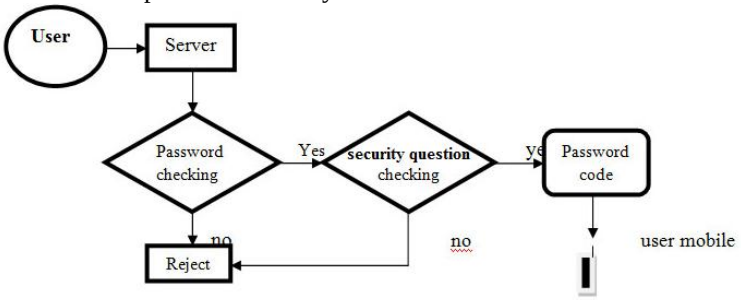


Fig.5 Attack prevention phase

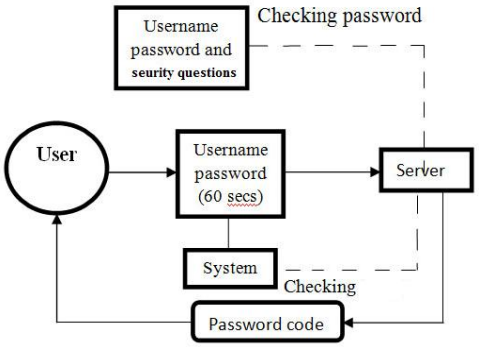


Fig.6 Security phase

4 Security Phase

In this phase security is enhanced by using the image retrieval method. Once the long-term password is entered it ensure for the image in the system by comparing with the image stored in the database. This phase is mainly used for the image checking and penetrates into the vital website by breeding the one time password. Then the password code is sent to the users' mobile from the server signifying that authentication is accomplished successfully.

5 Recovery Phase

If the user tends to forget the password or if the image is lost, this recovery phase is used to recover the forgotten password. The recovery phase-1 performs the operation of displaying the queries given by the user at the time of registration. The user answers the queries and recovery phase-2 performs the operation of checking the answers with the stored answers in the database. After successfully completing the recovery phases, it helps to allow the user to change the password.

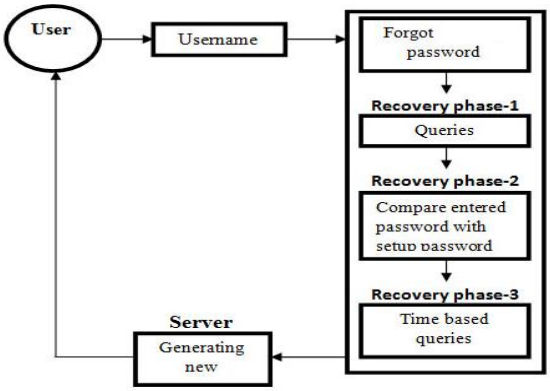
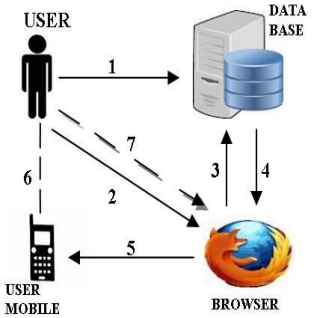


Fig.7 Recovery phase

V.ARCHITECTURE DESIGN

The overall description of online attacks detection and random password generation to prevent attacks is described below. Like standard web logins, it uses as a multi authenticator over the data. For multipurpose security, proposes generation of short term password and the image. Firstly, the user registers their specific aspects such as username, password, image, email Id and phone number. These specific aspects are reserved in the database of the server. The user access username and long-term password with the image in the web logins, then it establishes random alpha numeric that is used as a short term password for the user while entering into the browsing.

The short term password can compass the user through their mobile. The short term password is the one time password and this can be used only for the current browsing. This can be altering for successive web logins. So, the hacker doesn't have any adventitious to hack our short term password. Since it is only acknowledged to the user.



1. *Registration of user's specific aspects.*
2. *Detecting the attacks.*
3. *Checking process.*
4. *Derives specific aspects.*

5. *Receiving short term password.*
6. *Notification.*
7. *User browsing using short term password.*

Fig.8 Overall architecture for detect the online crime and prevent attacks by using short time security password.

VI. Experimental Results

The result of the project reveals that the password code is generated and is sent to the mobile of the user. This is one of the results exposed. The recovery phase produces a result to recover the password user must answer the security questions.

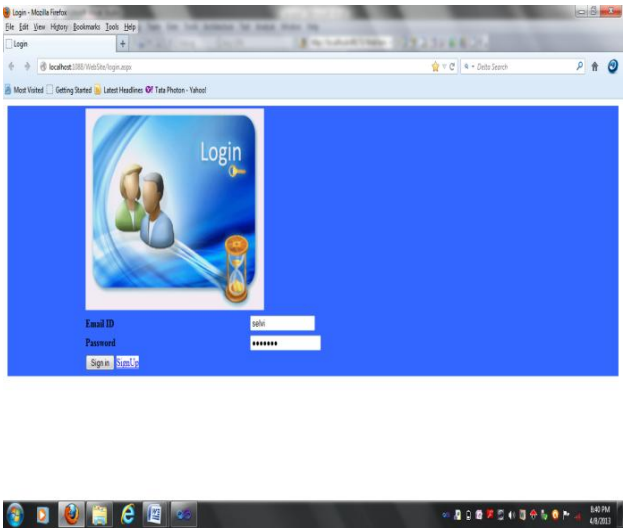


Fig.9 Login

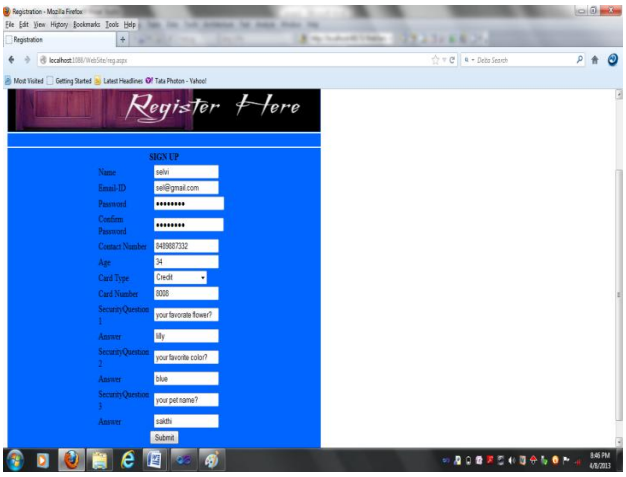


Fig.10 Registration

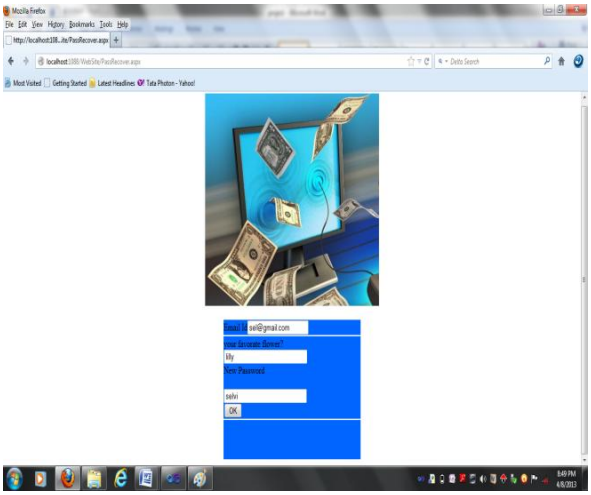


Fig.11 Password regeneration

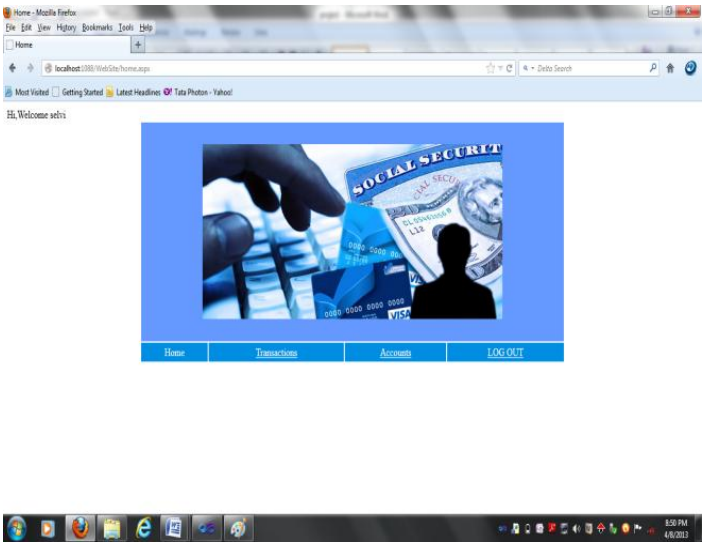


Fig.12 Home page

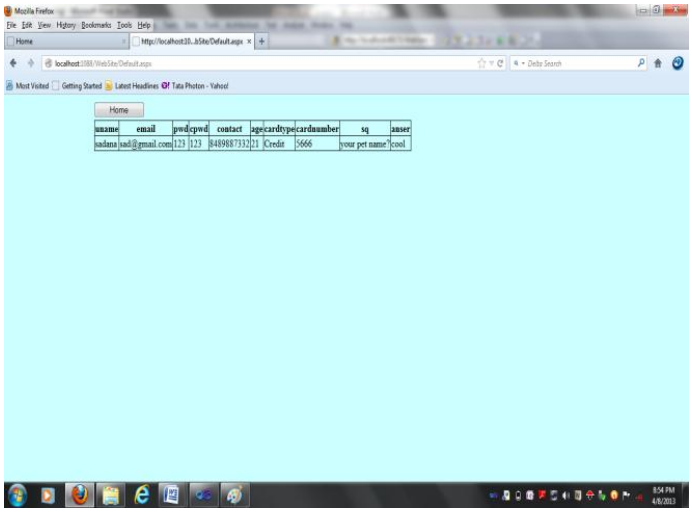


Fig.13 Verification

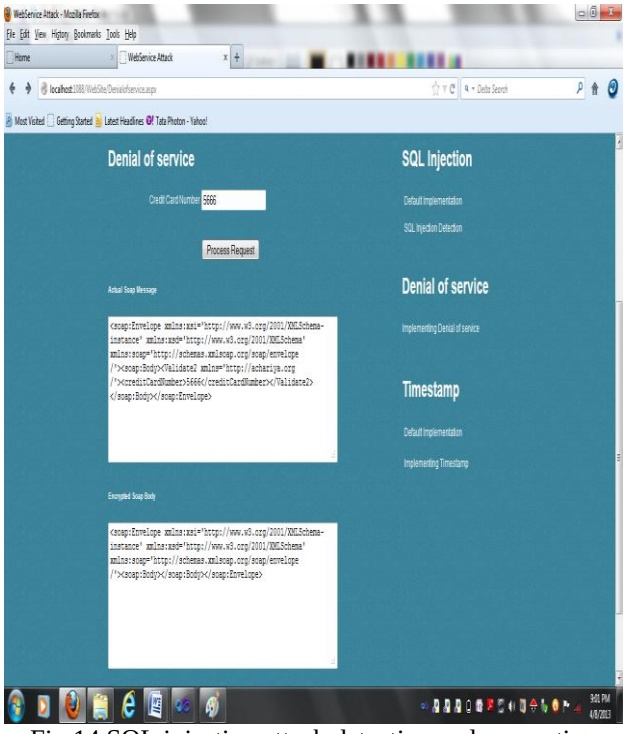


Fig.14 SQL injection attack detection and prevention

VII. Conclusion and Future Enhancement

In this report, we described in detail about Synthetic Identity of Crime Detection which is an efficient password management technique to remove password reuse attacks. While comparing with the other two factor authentication methods, this Multifactor Authentication and Random Password Generation is more efficient and economical .Firstly, the secure online money transactions are promoted using this multifactor authentication system .The Multifactor Authentication and Random Password Generation is resistant to Phishing attacks. Second, the weak password assumptions are avoided and in turn the password hacking is also reduced to a great extent. Third, the domino effect has been thoroughly flushed out since the passwords are different for each and every login. Hence, we propose that our new technique for detecting and Preventing by using Random Password Generation is more economical and reliable compared to other authentication factors. In this project, the authentication process needs a server and a system. The software should be installed onto the server and client does not work with all the system instead client works with the single system which is consistent for more security. It provides more security on online transaction by using biometrics recognitions.

References

[1] A. Bifet and R. Kirkby Massive Online Analysis, Technical Manual, Univ. of Waikato, 2009.

[2] A. Goldenberg, G. Shmueli, R. Caruana, and S. Fienberg, "Early Statistical Detection of Anthrax Outbreaks by Tracking Over-the-Counter Medication Sales," Proc. Nat'l Academy of Sciences USA (PNAS '02), 2002.

[3] B. Head, "Biometrics Gets in the Picture," Information Age, 2006.

[4] B. Schneier, Beyond Fear: Thinking Sensibly about Security in an Uncertain World. Copernicus, 2003.

[5] C. Phua, K. Smith-Miles and R. Gayler- Resilient Identity Crime Detection,2012.

[6] Clifton Phua, Kate Smith-Miles, Vincent Lee and Ross Gayler- Adaptive Spike Detection for Resilient Data Stream Mining, 2010.

[7] D. Hand, "Classifier Technology and the Illusion of Progress," Statistical Science, 2006.

[8] G. Gordon, D. Rebovich, K. Choo, and J. Gordon, "Identity Fraud Trends and Patterns: Building a Data-Based Foundation for Proactive Enforcement," Center for Identity Management and Information Protection, Utica College, 2007.

[9] J. Kleinberg, "Temporal Dynamics of On-Line Information Streams," Data Stream Management: Processing High-Speed Data Streams, M. Garofalakis, J. Gehrke, and R. Rastogi, eds., Springer, 2005.

[10] J. Neville, O. Simsek, D. Jensen, J. Komoroske, K. Palmer, and H. Goldberg, "Using Relational Knowledge Discovery to Prevent Securities Fraud," Proc. 11th ACM SIGKDD Int'l Conf. Knowledge Discovery in Data Mining (KDD '05), 2005.

[11] L. Hutwagner, W. Thompson, G. Seeman, and T. Treadwell, "The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS)," J. Urban Health, 2006.

[12] M. Jackson, A. Baer, I. Painter, and J. Duchin, "A Simulation Study Comparing Aberration Detection Algorithms for Syndromic Surveillance," BMC Medical Informatics and Decision Making, vol. 7, no. 6, 2007.

[12] M. Kantarcioglu, W. Jiang, and B. Malin, "A Privacy-Preserving Framework for Integrating Person-Specific Databases," Proc.UNESCO Chair in Data Privacy Int'l Conf. Privacy in Statistical Databases (PSD '08),2009.

[13] O. Kursun, A. Koufakou, B. Chen, M. Georgiopoulos, K. Reynolds, and R. Eaglin, "A Dictionary-Based Approach to Fast and Accurate Name Matching in Large Law Enforcement Databases," Proc. IEEE Int'l Conf. Intelligence and Security Informatics (ISI '06), 2006.

[14] P. Brockett, R. Derrig, L. Golden, A. Levine, and M. Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs, "The J. Risk and Insurance, 2002.

[15] P. Christen and K. Goiser, "Quality and Complexity Measures for Data Linkage and Deduplication," Quality Measures in Data Mining, F. Guillet and H. Hamilton, Springer, 2007.

[16] R. Bolton and D. Hand, "Unsupervised Profiling Methods for Fraud Detection," Statistical Science, 2001.

[17] R. Caruana and A. Niculescu-Mizil, "Data Mining in Metric Space: An Empirical Analysis of Supervised Learning Performance Criteria," Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '04), 2004.

