International Journal of Computer Science Engineering & Technology
APPROVED BY
NATIONAL SCIENCE LIBRARY (NSL)
NATIONAL INSTITUTE OF SCIENCE COMMUNICATION AND INFORMATION RESOURCES (NISCAIR)
COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH (CSIR)– NEW DELHI INDIA .
ISSN :2455-9091

SUBSCRIPTION FRAUD DETECTION USING PRIVATE SET INTERSECTION PROTOCOL

[1] S. Subhashini, [2] P. Shanmugaraja, M.E., Ph.D.
[1] PG Scholar, [2] Associate Professor Department of Information Technology,
[1, 2] Sona College of Technology, india.

ABSTRACT: Subscription fraud remains one of the top types of fraud, and is widespread across all operations. Fraudsters subscribe to services without intending to pay, employing schemes such as identity theft and the use of false, stolen or fictitious details. Data sharing between telecoms would increase fraud detection rates, but phone records are protected bylaw. Several protocols have to be proposed to enable fraud detection across multiple databases without revealing additional information. A model has to be proposed to generate phone records, with which we evaluate how the choice of parameters affects detection performance. The technique to several protocols that allow Telco's to match their customers to the other Telco's' fraudster database without revealing any additional information about the data than is necessary to classify the fraudster are also extended. Private Set Intersection Protocol is used to detect the problem of computing the intersection of private datasets.

KEYWORDS: [Self-Organizing Maps (SOM), Garbled Circuit (GC), Internet Protocol (IP), Local Area Network (LAN).]

1. INTRODUCTION

Telecommunications fraud continues to be a big problem in the industry today. Advancements in technology have made life easier and more convenient for most people today, but not without a price.

Telecommunications fraud can be broken down into several generic classes which describe the mode operators are defrauded but for the purpose of this paper the study is focused on "Subscription Fraud". Subscription fraud is a contractual fraud. In these kinds of fraud revenue is generated through the normal use of a service without having to pay. In this scenario, the fraudster operates at level of phone numbers where all transactions from this number is fraudulent and all activities in such cases are further abnormal throughout the active period of the account. Subscription fraud can be divided into two categories. These are: Subscription fraud for the purpose of personal usage by the fraudster and Subscription fraud for profit. In the second category, the fraudster opens a small outfit where he starts up a call center. The fraudster has no intensions of paying his bills but he sells the airtime to people who intend to make cheap long distance calls for cash. Fraud detection problems are found in many sectors of lives endeavor and the telecoms sector is not an exception. Hence fraud detection is referred to as the attempt engaged in discovering illegitimate usage of a communication network by identifying fraud

as quickly as possible once it has been perpetrated.

Background

There are many different types of telecommunications fraud and these can occur at various levels. The two most common types of fraud are subscription fraud and superimposed fraud. Others are intrusion fraud, fraud based on loopholes in technology, social engineering, fraud based on new technology, fraud based on new regulation and masquerading as another user. Fraud management systems have proved to be a suitable tool to detect fraud in different networks with diverse techniques such as self-organizing maps (SOM), general data mining, rule based systems profiling through Artificial intelligence techniques like neural networks or decision trees based on the hierarchical regime switching models, Bayesian networks, fuzzy rules or other data mining techniques.

Fraud detection can also be done at 2 levels call or behavior and with two different approaches user profile or signature based. Most of the techniques use the CDR data to create a user profile and to detect anomalies based on these profiles. The mined large amounts of CDR have in order to find patterns and scenarios of normal usage and typical fraud situations.

The most common and best succeeded methods for fraud analysis are signature based. These methods detect the fraud based on deviation detection by comparing the recent activity with the user behavior data which is expressed through the user signature.

Private Set Intersection

This is a two-party protocol between a client C and a server S. Each holds a set of inputs drawn from the same domain. At the conclusion of the protocol, C learns which elements of the sets are shared by both C and S. But neither party learns any element of the opposite party's set that is not an element of the intersection.

Freedman et al. present different versions of the protocol which are secure against semi-honest or malicious adversaries. We implement the semi-honest case and leave the malicious case to the interested reader.

A semi-honest adversary is a participant in the protocol that correctly follows the protocol but tries to gain more information about the other party's input than can be inferred by his own private input and the output of the protocol. A protocol is secure in the semi-honest model if there is no efficient semi-honest adversary, i.e., an adversary that runs in polynomial time.

Related Work

Improved Garbled Circuit Building Blocks and Applications to Auctions and Computing Minima use rely on recently proposed "free XOR" GC technique. It is having Input/Output Conversion Protocols and Efficient OT protocol. Building an Effective Representation for Dynamic Networks it develops a generic framework for evaluating and tuning any dynamic graph. In addition, it presents a preliminary analysis on Web logs and e-mail networks. It is used for Approximation technique, Evaluation technique and Application of technique

Faster Secure Two-Party Computation with Less Memory present an improved implementation of Yao's garbled circuit protocol in the semi-honest adversaries setting which is up to 10 times faster than previous implementations. It is secure two-party computation. Secure Distributed Subgroup Discovery in Horizontally Partitioned Data present new protocols which allow distributed subgroup discovery while avoiding the disclosure of the individual databases. It is having a Multi-Party Computation protocol. Those are Top-1 Subgroup Discovery, Top-k Subgroup Discovery, Weighted Covering and Greedy Top-1 Subgroup Discovery. Security Analysis and Enhancements of 3GPP Authentication and Key Agreement Protocol shows to address such security problems in the current 3GPP AKA then present a new

authentication and key agreement protocol which defeats redirection attack and drastically lowers the impact of network corruption.

The protocol, called AP-AKA, also eliminates the need of synchronization between a mobile station and its home network. AP-AKA specifies a sequence of six flows. Universal Mobile Telecommunication, 3GPP AKA protocol and APAKA protocol are used. The main The main disadvantage of this one of this one an authentication and key agreement protocol which can defeat the redirection attack and may drastically lower the impact of network corruption.

The protocol, called AP-AKA, also eliminates the need of synchronization between the mobile station and the home network. In AP-AKA, the home network does not maintain dynamic states for each individual subscriber. The advantages are Provide efficient authentication, Improve network performance and effectively maintained home network and foreign network. It first gives the background, a brief overview of the overall EPS architecture. It goes on to list the various requirements to be met for EPS security.

A description of the EPS security architecture and detailed security procedures are given subsequently. The innovations that have been introduced in EPS, on top of UMTS, are highlighted all through the article. The article concludes by listing some open security issues at the moment.

Network Access Security in Next-Generation 3GPP Systems: A Tutorial says gives the background, a brief overview of the overall EPS architecture. It goes on to list the various requirements to be met for EPS security. A description of the EPS security architecture and detailed security procedures are given subsequently. The innovations that have been introduced in EPS, on top of UMTS, are highlighted all through the article. Integrity algorithm is used this paper. The main disadvantage of this one of this one in this heterogeneous framework, there is a greater risk of unlawful accessing and tampering with information that travels between the various entities. The advantages are KASME is that it is bound to the MS identity and the identity of the SN. Another advantage is that KASME is returned to the SN only after the UE authentication response is validated by the HSS. The NAS security context has a longer lifetime than the AS security context. It can also stay alive when the UE goes to idle.

2. SYSTEM MODEL

New Training Algorithm Due to increase in number of users on internet, many people want to attack other system resources. Competitors also want to make their web site more popular than others. So they want to attack the service of other's web site. They keep on logon to a particular web site more times, and then service provided by the web server performance keeps degraded.

To avoid that one, this application maintains a status table. In that it keeps the IP addresses of current users and their status. If the particular IP address has been signed on for a first time, it makes the status as genuine user. For it marks as Normal user. For the fifth time it makes the particular IP address status as Attacker. In the time calculations there are only consider times.

User wish to server increase the time depends up on the application. After that, the user cannot allow get the service of that particular web site. The service is denied to that particular IP address.

To prevent the server from accessing the server and interruption of the performance in server is distribute successfully in this system. This is very useful for the users to determine the efficiency of our proposed algorithm named as New Training Algorithm.

3. SYSTEM MODULES

1. Data Collection
2. Privacy Preserving Data Analysis
3. Data Preprocessing
4. Privacy Preserving Data Publishing
5. Hacker

4. MODULE DESCRIPTION

Data Collection

Collecting and preparing sample data is the first step in designing models. The data is gotten from historical data of fraudulent and non-fraudulent subscriptions.

Privacy Preserving Data Analysis

The privacy preserving data analysis protocols assume that participating parties are truthful about their private input data. The techniques developed in assume that each party has an internal device that can verify whether they are telling the truth or not.

Data Preprocessing

Three data preprocessing procedures are conducted to train the system more efficiently. These procedures are: (1) solve the problem of missing data, (2) normalize data and (3) randomize data. The missing data are replaced by the average of neighboring values during the same week.

Privacy Preserving Data Publishing

The incentive compatible model is only concentrating on the secure data sharing process and does not consider the data storage publishing. All the users' private information's are stored in the particular database that is more securable one. Using a symmetric encryption algorithm, Triple Des algorithm is used to encrypt the all the users sensitive information in the secure database.

Hacker

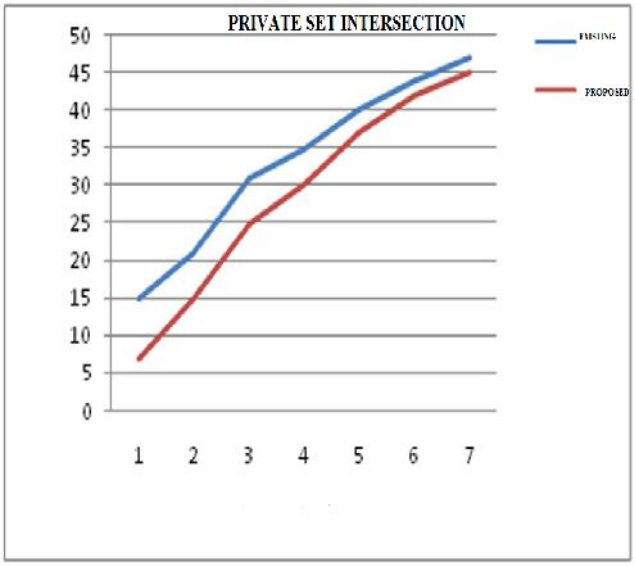
The occurrence of fraud is detected by using the naïve Bayesian algorithm with the help of the historical data of hackers.

5. SYSTEM IMPLEMENTATION

We implemented the private set intersection protocol using the Java programming language. We chose Java because of its suitability for rapid prototyping and its vast collection of libraries. The core of the program is Twisted, an event-driven networking engine. It simplifies the complexity of networking by providing a suitable set of primitives.

The implementation consists of two programs, one implementing the client's, and one implementing the server's functionality. They communicate over the network using a TCPsocket,thus they can be run on different computers. However, although we implemented all parts necessary to run the protocols in a secure fashion we do not want to omit the limitations of our implementation.

There is neither a graphical user interface nor a connector to a database. We also assume that the program is run through a secure communication channel. But since this are all fairly standard components we believe that they can be easily added to our implementation and their absence does not affect the measurements.



6. ALGORITHMIC STEPS

The protocol follows the following basic structure. C defines a polynomial P whose roots are her inputs:

$P(y) = (x1 - y)(x2 - y). . .(x_{kC} - y) = \Sigma {}^{k}c_u=_0\alpha_u y^u.$

She sends to S homomorphic encryptions of the coefficients of this polynomial. S uses the homomorphic properties of the encryption system to evaluate the polynomial at each of his inputs. He then multiplies each result by a fresh random number r to get an intermediate

result, and he adds to it an encryption of the value of his input, i.e., S computes Enc(r · P(y) + y). Therefore, for each of the elements in the intersection of the two parties' inputs, the result of this computation is the value of the corresponding element, whereas for all other values the result is random

7. TRAINING ALGORITHM

Input: C's input is a set X = {x_1, . . . , x_{kc} }, S's input is a set Y = {y_1, . . . , y_{ks} }. The elements in the input sets are taken from a domain of size N.

1. C performs the following:

(a) She chooses the secret-key parameters for a semantically-secure homomorphic encryption scheme, and publishes its public keys and parameters. The plaintexts are in a field that contains representations of the N elements of the input domain, but is exponentially larger.

(b) She uses interpolation to compute the coefficients of the polynomial P(y) = $\Sigma^{k} c_u=_0 \alpha_u y^u$ of degree k_C with roots {x1, . . . , x_{kC}}.

(c) She encrypts each of the (k_C + 1) coefficients by the semantically-secure homomorphic encryption scheme and sends to S the resulting set of ciphertexts, {Enc(α_0), . . . , Enc(α_{kC})}.

2. S performs the following for every y ∈ Y ,

(a) He uses the homomorphic properties to evaluate the encrypted polynomial at y. That is, he computes Enc(P(y)) = Enc($\Sigma^{k} c_u=_0 \alpha_u y^u$).

(b) He chooses a random value r and computes Enc(rP(y) + y). (One can also encrypt some additional payload data py by computing Enc(rP(y) + (y|p_y)). C obtains p_y iff y is in the intersection.)

He randomly permutes this set of k_S ciphertexts and sends the result back to the client C.

3. C decrypts all k_S ciphertexts received. She locally outputs all values x ∈ X for which there is a corresponding decrypted value .

CONCLUSION

Fraud is a multi-billions problem around the globe. The problem with telecommunication fraud is the huge loss of revenue and it can affect the credibility and performance of telecommunication companies. The most difficult problem that faces the industry is the fact that fraud is dynamic. This means that whenever fraudster's feel that they will be detected they find other ways to circumvent security measures.

Telecommunication fraud also involves the theft of services and deliberate abuse of voice and data networks. In such cases the perpetrator's intention is to completely avoid or at least reduce the scharges for using the services. Over the years, fraud has increased to the extent that losses to telephone companies are measured in terms of billions of American dollars.

Private set intersection is a useful building block for many privacy-preserving applications. Our results show that protocols based on generic secure computation can offer performance that is competitive with the best known custom protocols, without the need to rely on application specific techniques.

Since our protocols are built using generic garbled circuits they can be easily incorporated into larger secure-computation protocols, or combined with auditing mechanisms. Our work provides evidence that many secure computation problems can be solved without resorting to the design of custom protocols.

REFERENCES

[1] M. Freedman, K. Nissim, and B. Pinkas, "Efficient private matching and set intersection," in Proc. Adv. Cryptology - EUROCRYPT, 2004, pp. 1–19.

[2] Bolton, R. J. and Hand, D. J. (2002): Statistical Fraud Detection. A Review, Institute of Mathematical Statistics, 17(3), 235–255.

[3] Barson, P., Field, S., Davey, N., McAskie, G. and Frank, R. (1996): The Detection of Fraud in Mobile Phone Networks. Neural Network World, 6(4), 477–484.
[4] H. Grosskreutz, B. Lemmen, and S. R€uping, “Secure distributed subgroup discovery in horizontally partitioned data,”Trans. Data Privacy, vol. 4, no. 3, pp. 147–165, 2011.
[5] V. Kolesnikov, A. Sadeghi, and T. Schneider, “Improved garbled circuit building blocks and applications to auctions and computing minima,” in Proc. 8th Int. Conf. Cryptol. Netw. Secur., 2009, pp. 1–2.
[6] 2013 global fraud loss survey by the CFCA. (2013). [Online]. Available: http://cfca.org/pdf/survey/CFCA2013GlobalFraudLossSurvey-pressrelease.pdf
[7] Telecom fraud survey. [Online]. Available: http://www.prweb. com/releases/neuraltechnologies/fraudandrisk management/ prweb8472098.htm, 2011.
[8] Y. Huang, D. Evans, J. Katz, and L. Malka, “Faster secure twoparty computation using garbled circuits,” in Proc. USENIX Secur. Symp., 2011, pp. 35–35.
[9] W. Henecka and T. Schneider, “Faster secure two-party computation with less memory,” in Proc. 8th ACM SIGSAC Symp. Inform., Computer Commun. Secur., 2013, pp. 437–446.
[10] R. A. Becker, C. Volinsky, and A. R. Wilks, “Fraud detection in telecommunications: History and lessons learned,” Technometrics, vol. 52, no. 1, pp. 20–33, 2010