UNIVERSITI PUTRA MALAYSIA

DEVELOPMENT OF A METHOD FOR FRAUD SEVERITY MEASUREMENT BASED ON USAGE PROFILING

MOHD SHAFRI BIN KAMARUDDIN

FSKTM 2006 14

DEVELOPMENT OF A METHOD FOR FRAUD SEVERITY MEASUREMENT BASED ON USAGE PROFILING

By

MOHD SHAFRI BIN KAMARUDDIN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirement for the Degree of Master of Science

July 2006

Dedicated to my mother; Roosnia Samsodin,

my wife; Rashidah Abidin,

my daughters; Aina Shahidah, Aina Shahirah, Aina Shafiqah

and my family

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

DEVELOPMENT OF A METHOD FOR FRAUD SEVERITY MEASUREMENT BASED ON USAGE PROFILING

By

MOHD SHAFRI BIN KAMARUDDIN

July 2006

Chairman : Associate Professor Ramlan Mahmod, PhD

Faculty : Computer Science and Information Technology

The nature of fraud has changed from cloning fraud to subscription fraud, which makes specialized detection methodologies inadequate. Instead, the focus is on the detection methodologies that based on the subscriber's calling activity or calling pattern, which can be roughly divided into two main categories: absolute analysis and differential analysis.

Absolute analysis is capable at detecting the extremes of fraudulent activity. However, absolute analysis cannot trap all types of fraud especially usage behavior fraud related. An alternative approach to this problem is to perform a differential analysis against subscriber's behavioral patterns. Certain behavioral patterns may be considered anomalous or abnormal for certain subscriber and potentially indicative of fraud but would be considered acceptable for another. In order to overcome the uncertainty in

behavioral patterns, in this research, we propose to conduct the usage profiling at individual subscriber level. Usage profiling is a process of generating calling statistic based on predefined categories, which involve some form of aggregation from subscriber's calling activity or CDR.

Usage profiling process will generate two forms of usage profile : usage profile history (UPH) and current usage profile (CUP). In fraud detection system, comparison of these two types of usage profile will generate a measure known as fraud severity measurement. Implementation of the Hellinger distance for measuring a fraud severity, lack of detection accuracy as this method does not properly define the measurement scale as the Hellinger distance method will generate variation of values for fraud severity measurement. Therefore, it is very difficult to define the actual severity level of detected fraud.

In this research, we propose a new method for measuring fraud severity. The advantages of the method are detection accuracy and detection speed. With the new method, the severity measurement scale is properly defined and the detection speed is faster than the Hellinger distance.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Master Sains

PEMBANGUNAN KAEDAH BAGI PENGUKURAN TAHAP KEPARAHAN FRAUD MENGGUNAKAN USAGE PROFILING

Oleh

MOHD SHAFRI BIN KAMARUDDIN

Julai 2006

Pengerusi : Profesor Madya Ramlan Mahmod, PhD

Fakulti : Sains Komputer dan Teknologi Maklumat

Keadaan fraud telah berubah dari fraud pengklonan kepada fraud langganan, yang menyebabkan kaedah pengesanan yang terhad tidak lagi sesuai digunakan. Malahan, fokus kaedah pengesanan adalah tertumpu kepada kaedah pengesanan berdasarkan kepada aktiviti panggilan atau ragam panggilan pelanggan yang boleh dibahagikan kepada dua kaedah utama: "absolute analysis" atau "differential analysis".

"Absolute analysis" berupaya untuk mengesan aktiviti fraud yang keterlaluan. Bagaimanapun, "absolute analysis" tidak berupaya untuk mengesan kesemua jenis fraud terutama sekali fraud yang berkaitan dengan ragam penggunaan. Penggunaan kaedah "differential analysis" adalah dilihat sebagai kaedah alternatif kepada masalah ini. Sebahagian daripada corak penggunaan mungkin dianggap ganjil dan berkemungkinan menyebabkan kes fraud bagi sebahagian pelanggan, tetapi tidak bagi sesetengah yang

lain. Bagi mengatasi masalah ketidaktentuan corak penggunaan ini, di dalam penyelidikan ini, telah dicadangkan perlaksanaan kaedah “usage profiling” di peringkat setiap individu pelanggan. “Usage profiling” adalah satu kaedah pengumpulan statistik panggilan mengikut kategori yang telah ditetapkan berdasarkan kepada aktivti panggilan pelanggan atau CDR.

“Usage profiling” akan menghasilkan dua jenis profail: profail penggunaan lepas dan profail penggunaan terkini. Di dalam sistem pengesanan fraud, perbandingan kedua-dua profail ini akan menghasilkan satu tahap pengukuran yang boleh menunjukkan tahap keparahan sesuatu kes fraud tersebut. Perlaksanaan pengukuran keparahan kes fraud menggunakan kaedah “Hellinger distance” mempunyai sedikit kelemahan dari aspek ketepatan tahap keparahan sesuatu kes fraud itu, oleh kerana kaedah ini tidak menentukan dengan baik skala sebenar tahap pengukuran yang digunakan memandangkan ia menghasilkan nilai pengukuran keparahan fraud yang pelbagai. Ini menyebabkab kesukaran untuk menentukan tahap keparahan sebenar kes fraud yang dikesan.

Satu kaedah baru telah dicadangkan bagi mengukur tahap keparahan kes fraud di dalam penyelidikan ini. Kelebihan kaedah baru ini adalah dari aspek ketepatan pengesanan dan kepantasan pengesanan. Kaedah yang dicadangkan ini dapat menetapkan skala pengukuran keparahan dengan baik disamping kepantasan pengesanan yang lebih baik daripada kaedah “Hellinger distance” .

ACKNOWLEDGEMENTS

Assalamualaikum.

Alhamdulillah, all praise to ALLAH S.W.T for giving me strength, patience and motivation to complete this thesis as a fulfillment for the degree of Master of Science.

I would like to express my deepest appreciation and gratitude to the research supervisory committee, lead by *Prof. Madya Dr. Ramlan Mahmod, Prof. Madya Dr. Hj. Md. Nasir Sulaiman* and *En. Mohd Taufik Abdullah* for their guidance, support, encouragement and valuable advice throughout my study in this faculty.

My deepest thanks to my mother, wife, daughters and family, for their support and understanding during my study. I also like to thank all staffs at the Faculty of Computer Science and Information Technology, UPM for their administrative support.

Finally, I would like to extend my gratitude to management of TM Research And Development Sdn. Bhd. (TMR&D) for sponsoring and giving me opportunity to further study in UPM.

Thank you all and may ALLAH S.W.T bless all individuals for their kindness.

Wassalam.

Mohd Shafri Kamaruddin

July 2006

I certify that an Examination Committee has met on 6th July 2006 to conduct the final examination of Mohd Shafri bin Kamaruddin on his Master of Science thesis entitled "Development of a Method for Fraud Severity Measurement Based on Usage Profiling" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Hamidah Ibrahim, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Hj. Mohd Hassan Selamat
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Abdul Azim Abd. Azim, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Abdullah Mohd Zin, PhD
Professor
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
(External Examiner)

_____

HASANAH MOHD GHAZALI, PhD
Professor/Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee are as follows:

Ramlan Mahmod, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Md. Nasir Sulaiman, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Mohd Taufik Bin Abdullah
Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor/ Dean
School of Graduate Studies
Universiti Putra Malaysia

Date :

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not previously or concurrently submitted for any other degree at UPM or other institutions.

_____

MOHD SHAFRI BIN KAMARUDDIN

Date :

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDR	Call Details Record
UPH	Usage Profile History
CUP	Current Usage Profile
CDMA	Code Division Multiple Access
TELCO	Telecommunication Company
MSC	Mobile Subscriber Center
ITU	International Telecommunication Union

CHAPTER 1

INTRODUCTION

1.1 Introduction

Fraud in telecommunication can be defined as any dishonest or illegal use of services where the intention of the sender is to avoid or reduce legitimate call charges (Johnson, 1996). Other definition of fraud is an attempt to obtain or gain illegitimate access to the network in order to enjoy unbillable services and undeserved fees (Davis and Goyal, 1993). The term "fraud" has a particular meaning in legal perspective; however the term is used broadly to mean misuse, dishonest intention or improper conduct without implying any legal consequences.

Historically, earlier types of fraud used technological means to acquire free access. Cloning of mobile phones by creating copies of mobile terminals with identification numbers from legitimate subscribers was used as a means of gaining free access (Davis and Goyal 1993). In the era of analog mobile terminals, identification numbers could be easily captured by eavesdropping with suitable receiver equipment in public places, where mobile phones were evidently used.

One specific type of fraud, tumbling, is quite prevalent in the United States (Davis and Goyal, 1993). It exploits deficiencies in the validation of subscriber identity when a mobile phone subscription is used outside of the subscriber's home area. The fraudster keep tumbling (switching between) captured identification numbers to gain access.

Davis and Goyal (1993) state that the tumbling and cloning fraud have been serious threats to operators' revenues.

The first fraud detection systems examine whether two instances of one subscription are used at the same time (overlapping calls detection mechanism) or at locations far apart in temporal proximity (velocity trap). Both the overlapping calls or calls collision (Patel , 1997), and the velocity trap try to detect the existence of two mobile phones with identical identification codes, which clearly evidencing cloning. As a countermeasure to these fraud types, technological improvements are introduced together with implementation of fraud detection system.

Fraud detection system can be considered as a basic tool to detect fraudulent activity where its implementation may reveal fraudulent activity in telecommunication network. This tool is very beneficial to network operator, who may lose some of their revenue due to fraudulent activity as service fees or charges are uncollected.

However, new forms of fraud come into existence. A few years later, Johnson (1996) and O'Shea (1997) reported that the so-called subscription fraud to be the trendiest and the fastest-growing type of fraud. In similar spirit, Hoath (1998) characterized subscription fraud as being probably the most significant and prevalent worldwide telecommunications fraud type. In subscription fraud, a fraudster obtains a subscription (possibly with false identification) and starts a fraudulent activity with no intention to pay the bill. It is indeed non-technical in nature and by call selling, the entrepreneur-

minded fraudster can generate significant revenues for a minimal investment in a very short period of time (Johnson, 1996).

As a countermeasure to subscription fraud, Barson et. al. (1996) conduct experiments for detecting fraud using simulated calls, which consists of six different user types. Supervised feed forward neural network was implemented to detect anomalous or deviations in user calling activity. Simulated calls were extracted into two types of features: one set describing the recent use and another set describing the long-term behavior. Both set are accumulated statistics of call data over a different length of time windows.

Burge and Shawe-Taylor, (1996) used the same concept of recent use and long term behavior. They reported that fraudulent activity can be easily monitored through the analysis on user behavior. User behaviors are reflected in calling detail or CDR, which consists of information about the call. Differential analysis and absolute analysis can be used to detect fraudulent activity. Burge and Shawe-Taylor,. (1997) implemented unsupervised learning techniques in computing user behavior profiles over sequences of call records. Hellinger distance method used to measure the changes between user behavior profiles as alarm indicator.

User behavior profile can be classified into three different types of categories: usage indicators, mobility indicators and deductive indicators (Burge, et. al 1997). Usage indicators will show how the service is used. Mobility indicators refer to user's mobility while using service, and deductive indicators reflect by-product of fraudulent behavior.

Mobility and deductive indicators are capable to trap call velocity and call overlapping, which are commonly related to phone cloning.

Moerau and Vandewalle, (1997) and Taniguchi, et. al (1998) used feed forward neural network with supervised learning to classify the subscribers between fraud and non fraud using call summary statistic. They compared subscriber's past behavior and current behavior to detect any abnormalities from the past behavior.

Combination of two unsupervised neural network based on user profile, was another work by Moerau, et. al (1999). They attempt to monitor user profile based on A-number analysis and B-number analysis. Monitoring a few indicators may not reflect the real fraud scenario. Burge and Shawe-Taylor, (2001) implemented recurrent neural network techniques by comparing past behavior and current behavior for fraud detection. Comparison between both of them using Hellinger distance method, show certain measurement for triggering alarm.

Bourkeche and Notare, (2002) also used behavior profile information by comparing recent information and past information of mobile phone's user, which they extracted from usage logs. Radial basis function neural network (RBF NN) was used due to its simplicity and flexibility to adapt to pattern changes. RBF NN is widely used for solving classification and pattern recognition problem.

Most techniques applied in fraud detection system use calling activity to create behavior profile for subscriber and try to detect deviations from these profiles (Yufeng, et. al,

2004). Two main approaches normally implemented in fraud detection system: differential analysis and absolute analysis. Combination of these approaches capable to verify certain rule against component of data set in calling activity. Flexible criteria can be developed to detect any usage change based on user behavior history.

From the above explanation, it is evident that the detection mechanisms of the first generation soon became inadequate. In the last few years, most of the works in fraud detection system are based on usage profile, subscriber behavior or calling pattern. The more advanced detection mechanisms must be based on individual subscriber behavior as different subscriber may generate different calling pattern. Some of the calling pattern may considered normal to certain subscriber but not to others.

As a complementary to detection techniques in fraud detection system, usage profiling efforts in understanding the subscriber behavior or calling pattern in telecommunication network may be required to improve operation performance in detecting fraudulent activity. With an additional method of measuring the severity of fraudulent activity, the fraud detection system even useful and important to network operator. Using usage profiling method, individual subscriber behavior can be studied and examined through interrogation of sequence call details record (CDR).

Even though, we are able to study the calling pattern up to individual level, yet there is no specific sequence of CDR would be guaranteed as 100% fraudulent. Therefore, with an additional method of measuring the fraudulent activity, network operator can properly

manage available resources they have to respond to the fraudulent subscriber based on the fraud severity level generated by the measurement method.

In this study, two main approaches are used which are differential analysis and absolute analysis. Differential analysis approach is used to detect changes between subscriber's calling behavior history and subscriber's recent calling behavior history, which may indicate fraudulent activity. Subscriber's calling behavior history or also known as Usage Profile History (UPH) and subscriber's recent calling history or also known as Current Usage Profile (CUP) are two types of profile generated by usage profiling. Mean while, the absolute analysis is used as the mechanism to detect the fraudulent activity by comparing the individual standard UPH and CUP, where the standard UPH values are considered as threshold value for that particular behavior attribute.

This study also conducts analysis on three type of standard UPH generation methods: minimum method, maximum method and average method. Standard UPH generation method refers to the method to create a standard UPH for every subscriber based on a series of subscriber's UPH.

1.2 Problem Statement

In fraud detection system, it is very important to define the performance metric carefully. Several detection techniques use metric like detection rate, false alarm rate and average time of detection (Yufeng, et. al, 2004). The typical fraud detection techniques try to maximize accuracy rate and minimize false alarm rate.

Accuracy of fraud severity measurement is one of the important aspects in fraud detection system. Fraud severity measurement will be the main indicator for fraud analyst to properly handle fraud cases. It will also allow the fraud analyst to prioritize the fraud cases for investigation as the priority of the fraud cases play important role in fraud detection.

Implementations of Hellinger distance for measuring a fraud severity have been carried out in some previous works. (Fawcett and Provost, 1997), (Taniguchi, et. al., 1998) (Burge and Shawe-Tylor, 2001)

The Hellinger distance is defined as following (Lachaud ,2005),(Guha, McGregor and Suresh, 2005), (Poland and Hutter, 2005):

$$d = \sum_{i=1}^{k} \left(\sqrt{C_i} - \sqrt{H_i}\right)^2, \qquad (1.0)$$

where C and H are the UPH and CUP, respectively, and K is the number of entries or attributes in the profile record.