Expert Systems with Applications 36 (2009) 11559–11569

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Designing an expert system for fraud detection in private telecommunications networks

Constantinos S. Hilas *

Dept. of Informatics and Communications, Technological Educational Institute of Serres, Terma Magnisias, Serres, GR-621 24, Greece

ARTICLE INFO

Keywords:
Fraud detection
User modeling
Expert systems
Telecommunications
Data mining applications

ABSTRACT

Telecommunications fraud not only burdens telecom provider's accountings but burdens individual users as well. The latter are particularly affected in the case of superimposed fraud where the fraudster uses a legitimate user's account in parallel with the user. These cases are usually identified after user complaints for excess billing. However, inside the network of a large firm or organization, superimposed fraud may go undetected for some time. The present paper deals with the detection of fraudulent telecom activity inside large organizations' premises. Focus is given on superimposed fraud detection. The problem is attacked via the construction of an expert system which incorporates both the network administrator's expert knowledge and knowledge derived from the application of data mining techniques on real world data.

1. Introduction

Telecommunications fraud can be simply described as any activity by which telecommunications service is obtained without intention of paying (Gosset & Hyland, 1999). This kind of fraud has certain characteristics that make it particularly attractive to fraudsters. The main one is that the danger of localization is small. This is because all actions are performed from a distance, which in conjunction with the mess topology and the size of networks makes the process of localization time-consuming and expensive. Additionally, no particularly sophisticated equipment is needed, if one is needed at all. The simple knowledge of an access code, which can be acquired even with methods of social engineering, makes the implementation of fraud feasible. Finally, the product of telecommunications fraud, a phone call, is directly convertible to money (Hoath, 1998).

Several categories of telecommunications fraud have been reported. The main are the technical fraud, the contractual fraud, the hacking fraud, and the procedural fraud (Gosset & Hyland, 1999). Technical, contractual and procedural fraud usually burdens the telecom service provider, while hacking fraud also harms the subscriber. The latter may happen in the form of the superimposed fraud where the fraudster (hacker) uses the service in parallel with the subscriber and burdens his account. All fraud cases can actually be viewed as fraud scenarios, which are related to the way the access to the network was acquired. Detection techniques tailored to one case may fail to detect other types of fraud. For example, velocity traps, which can identify the use of a cloned cell phone, will fail to detect a case of contractual fraud. So, fraud detection focuses on the analysis of users' activity. The related approaches are divided into two main subcategories, the absolute analysis and the differential one. The first searches for limits between legal and fraudulent behavior, while the second tries to detect extreme changes in the user's behavior. In both cases, analysis is achieved by means of statistical and probabilistic methods, neural networks and rule-based systems. In Moreau and Vandewalle (1997) the use of indicators of excessive usage is being criticized as they may not only imply fraud but they may also point to the best customers.

A comparison of probabilistic methods with those that use rules is given in Taniguchi, Haft, Hollmen, and Tresp (1998). In 1999, Fawcett and Provost (1997), proposed a combination of rules and profile extraction, in order to detect fraud. The outputs of their system are combined via a trained linear model in order to produce alarms. Rosset et al. report encouraging results from the use of rules that are exported with a variant of the C4.5 algorithm (Rosset, Murad, Neumann, Idan, & Pinkas, 1999). Alves et al. propose two anomaly detection methods based on the concept of signatures for the detection of superimposed fraud (Alves et al., 2006). The appropriate feature extraction procedure is dealed with in Dong et al. (2004). In a previous work (Hilas & Sahalos, 2006), the author of the present paper concluded to a user behavior characterization model that gives good results towards superimposed fraud detection.

The use of expert systems towards fraud detection has either not been published or is referred to under different names (Liao, 2005) with the most common one being "data mining". There is however a limited bibliography in relative subjects such as

* Tel.: +30 2321049314.
E-mail address: chilas@teiser.gr

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.03.031

intrusion detection in computer systems (Jackson, DuBois, & Stallings, 1991; Sebring, Shellhouse, Hanna, & Whitehurst, 1988), user profiling for credit card fraud detection (Kokinnaki, 1997), auto insurance fraud (Belhadji & Dionne, 1997), or consumer behavior analysis (Adomavicius & Tuzhilin, 1999). Some recent publications combine data mining or expert systems approaches towards telecom churn prediction (Wei & Chiu, 2002; Shin-Yuan Hung & Yen, 2006) and subscription fraud detection (Estevez, Held, & Perez, 2006).

In the present paper a rule-based expert system is presented which aims to the detection of superimposed fraud cases in the telecommunications network of a large organization. Rules are induced by both using the network administrator's expert knowledge and by applying data mining methods on real world data. The paper proceeds as follows. In the next chapter the telecommunications environment in which the expert system will operate is presented. In the third chapter the expert system's operating characteristics and specifications are outlined. In Chapter 4 a brief analysis of prior data mining analysis of the data in hand is given, while the structure of the expert system is presented with the use of flow charts in Chapter 5. Experimental results are given in Chapter 6. In the last chapter conclusions are drawn.

2. Operating environment

The present paper describes the construction of an expert system that aims towards the detection of superimposed fraud cases in the telecommunication network of a large organization. It is a tailored made application, which can also be applied to similar networks after the incorporation of any proprietary network policy and expert knowledge. Here the organization under study is a large University with more than 5000 employees (administrative, teaching and research stuff).

Each employee that holds a permanent post is supplied with a telephone set (terminal) and a unique Personal Authorization Code (PAC) that overrides the terminal's class of restrictions (COR) and authorizes him to place costly outgoing calls. The PAC is also used in order to properly charge users for the calls they place. According to the organization's charging policy, only calls to national, international and mobile destinations are charged. Calls to local destinations are not charged so they are not included in the study. If anyone (e.g., a fraudster) finds a valid PAC he can use it to place his own calls from any telephone set within the organization and charge the calls to the legitimate PAC owner.

Although a user's PAC can unlock any telephone set in the organization one expects its use to be highly correlated with the owner's telephone set. This could be used as a powerful rule to imply legitimate use. Adding to this, the user may also be related with a fax machine (e.g. in a Department's secretariat) and/or a third telephone set placed in a laboratory.

The PAC can be used concurrently from two telephone sets. This observation may be used as a clue in velocity traps. However, after the analysis of the real world data it was found that there is a case where the concurrent use of a PAC from two telephone sets is legitimate. This is the case when a user uses her PAC to send a multipage fax message and at the same time she uses it from another terminal to place a voice call.

The University's personnel may be divided into two main categories, namely the Administrative – Technical Staff and the Teaching – Research Staff. The majority of the Administrative – Technical Staff works from 07:00 a.m. to 15:00 p.m. 5 days a week (Monday–Friday). Two exceptions are the cleaning staff (06:00–14:00), and the security staff (works in three shifts, 24 h a day – 7 days a week). Personnel that belong in the last two categories have access to offices, laboratories, classrooms, etc. so one may expect their PAC to be loosely correlated with a single terminal. Teaching and Research staff does not have fixed office hours and they usually work after hours. It is not surprising for research to be conducted after Midnight or during Weekends.

There are also people that have some temporal labor relation with the University. These are graduate and post-graduate students, part-time support personnel, visitors, personnel on detachment, etc. Due to their temporal labor relation these people do not have a PAC. Teaching staff (e.g. Professors, Lecturers, etc.) are very likely to give their PAC to secretaries or students in order to assist them to several administrative jobs. The sharing of a PAC, even with people one trusts, makes the PAC prone to fraud. The analysis of the fraud cases that were studied is enhancing this remark. The most defrauded PACs are those that belong to the Teaching staff.

The experience of the network administrators along with the analysis of real fraud cases revealed that the fraudsters share some common characteristics. First of all they are greedy and tend to make many and expensive calls. Common destinations are premium services such as phone auctions, "party" lines, matchmaking lines, etc. Long duration or long distance calls with friends and close relatives are also common. Another important attribute is that fraudsters exhibit great mobility within the organization. They constantly change terminals probably in fear of localization. In a particular case, the fraudster revealed a hacked PAC to more people which started a "hail" of phone calls from the organization premises to premium phone services.

3. Expert system operating characteristics and specifications

The development of the expert system (ES) was based on the network's administrator's expert knowledge and the application of data mining techniques on several detailed user accounts. The analysis yielded appropriate tests that must be performed in order to identify new fraud cases. These tests are expressed by means of IF ... THEN ... ELSE rules.

User demographic data were also incorporated into the ES in order to enhance its accuracy. These are the user's labor relation with the organization (e.g. administrative, teaching staff, etc.), his telephone number, his fax number and his home telephone number. All these data are public domain and their use cannot be considered as an intrusion into user's privacy. Moreover, according to the Greek legislation (Greek Law 2472/1997, 2008) one may analyze private data as long as the analysis is conducted for administrative; security or research reasons, within an organization's premises and the raw data are protected from unauthorized access. It is also stressed that during the expert system's design and implementation process there is no need to have access to the raw data. One may only need to know their structure. Adding to this, during the data analysis process, all data may be anonymous and the system's output may only give warnings and alarms that are forwarded to the authorized personnel for visual inspection.

An additional specification for the expert system is its ability to perform both real-time tests and batch tests on historical data. It should be pointed out that in a Private Branch Exchange (PBX) phone call analysis cannot actually happen on real-time. This is because all the details of the call are available to the system only after the call has been completed. These details are written in the Call Detail Record (CDR), which is outputted by the PBX to a peripheral logging system. This is a much different procedure compared to a credit card validation scheme (Kokinnaki, 1997), where the transaction is completed only after the credit card is first checked for its validity and available credit limit. Additionally, during the credit card validation check one may also compare the current purchase with the owner's profile in order to diagnose probable fraud. This crosschecking cannot be made in a PBX. A PBX will only check the validity of the PAC prior to unlocking the outgoing trunk. The