Novel Techniques for Fraud Detection in Mobile Telecommunication Networks

Yves Moreau, Bart Preneel, Dept. Electrical Engineering-ESAT K.U.Leuven

{yves.moreau,bart.preneel}@esat.kuleuven.ac.be

Peter Burge, John Shawe-Taylor, Dept. Computer Science, Royal Holloway

{peteb,jst}@dcs.rhbnc.ac.uk

Christof Stoermann, Siemens AG

christof.stoermann@zfe.siemens.de

Chris Cooke, Vodafone

chris.cooke@vf.vodafone.co.uk

Group: Network

Abstract: *This paper discusses the status of research on detection of fraud undertaken as part of the European Commission-funded ACTS ASPeCT (Advanced Security for Personal Communications Technologies) project. A first task has been the identification of possible fraud scenarios and of typical fraud indicators which can be mapped to data in toll tickets. Currently, the project is exploring the detection of fraudulent behaviour based on a combination of absolute and differential usage. Three approaches are being investigated: a rule-based approach, an approach based on neural network, where both supervised and unsupervised learning are considered. Special attention is being paid to the feasibility of the implementations.*

Introduction:

It is estimated that the mobile communication industry loses several million ECUs per year due to fraud. Therefore, prevention and early detection of fraudulent activity is an important goal for network operators. It is clear that the additional security measures taken in GSM and in the future UMTS (Universal Mobile Telecommunications System) make these networks less vulnerable to fraud than the analogue networks. Nevertheless, certain types of commercial fraud are very hard to preclude by technical means. It is also anticipated that the introduction of new services can lead to the development of new ways to defraud the system. The use of sophisticated fraud detection techniques can assist in early detection of commercial frauds, and will also reduce the effectivity of technical frauds.

One of the tasks of the European Commission-funded ACTS project ASPeCT (Advanced Security for Personal Communications Technologies) [1] is the development of new techniques and concepts for the detection of fraud in mobile telecommunication networks. This paper intends to report on the progress made during the first year. For a more detailed status report, the reader is referred to [2].

The remainder of this paper is organised as follows: section 1 discusses the identification of possible fraud scenarios and of fraud indicators; section 2 discusses the general approach of user profiling; section 3 and 4 present respectively the rule based approach and the neural net based approach to fraud detection.

Section 1: Possible frauds and their indicators

A first step of the work consists of the identification of possible fraud scenarios in tele-communication networks and particularly in mobile networks. These scenarios have been classified by the technical manner in which they are committed; also an investigation has been undertaken to identify which parts of the mobile telecommunication network are abused in order to commit any particular fraud. Other characteristics that have been studied are whether frauds are technical fraud operated for financial gain, or they are fraud related to personal use - hence not employed for profiteering. A further classification is achieved by considering whether the network abuse is the result of administrative fraud, procurement fraud, or application fraud.

Subsequently, typical indicators have been identified which may be used for the purposes of detecting fraud committed using mobile telephones. In order to provide an indication of the likely ability of particular indicators to identify a specific fraud, these indicators have been classified both by their type and by their use.

The different types are :-

- usage indicators, related to the way in which a mobile telephone is used;
- mobility indicators, related to the mobility of the telephone;
- deductive indicators, which arise as a by-product of fraudulent behaviour (e.g., overlapping calls and velocity checks).

Indicators have also been classified by use:-

- primary indicators can, in principle, be employed in isolation to detect fraud;
- secondary indicators provide useful information in isolation (but are not sufficient by themselves)
- tertiary indicators provide supporting information when combined with other indicators.

A selection has been made of those scenarios which cannot be easily detected using existing tools, but which could be identified using more sophisticated approaches.

The potential fraud indicators have been mapped to network data required to measure them. The information required to monitor the use of the communications network is contained in the toll tickets.

Toll Tickets are data records containing details pertaining to every mobile phone call attempt that is made. Toll Tickets are transmitted to the network operator by the cells or switches that the mobile phone was communicating with at the time due to proximity. They are used to determine the charge to the subscriber, but they also provide information about customer usage and thus facilitate the detection of any possible fraudulent use. It has been investigated which fields in the GSM toll tickets can be used as indicators for fraudulent behaviour.

Before use in the fraud detection engine, the toll tickets are being preprocessed. An essential component of this process is the encryption of all personal information in the toll tickets (such as telephone numbers). This allows for the protection of the privacy of users during the development of the fraud detection tools, while at the same time the network operators will be able to obtain the identity of fraudulent users.

Section 2: User profiling

Absolute or differential analysis

Existing fraud detection systems tend to interrogate sequences of Toll Tickets comparing a function of the various fields with fixed criteria known as *triggers*. A trigger, if activated, raises an alert status which cumulatively would lead to an investigation by the network operator. Such fixed trigger systems perform what is known as an *absolute* analysis of the Toll Tickets and are good at detecting the extremes of fraudulent activity.

Another approach to the problem is to perform a *differential* analysis. Here we monitor behavioural patterns of the mobile phone comparing its most recent activities with a history of its usage. Criteria can then be derived to use as triggers that are activated when usage patterns of the mobile phone change significantly over a short period of time. A change in the behaviour pattern of a mobile phone is a common characteristic in nearly all fraud scenarios excluding those committed on subscription where there is no behavioural pattern established.

There are many advantages to performing a differential analysis through profiling the behaviour of a user. Firstly, certain behavioural patterns may be considered anomalous for one type of user, and hence potentially indicative of fraud, that are considered acceptable for another. With a differential analysis flexible criteria can be developed that detect any change in usage based on a detailed history profile of user behaviour. This takes fraud detection down to the personal level comparing like with like enabling detection of less obvious frauds that may only be noticed at the personal usage level. An absolute usage system would not detect fraud at this level. In addition, however, because a typical user is not a fraudster, the majority of criteria that would have triggered an alarm in an absolute usage system will be seen as a large change in behaviour in a differential usage system. In this way a differential analysis can be seen as incorporating the absolute approach.

The differential approach

Most fraud indicators do not become apparent from an individual Toll Ticket. With the possible exception of a velocity trap, we can only gain confidence in detecting a real fraud through investigating a fairly long sequence of Toll Tickets. This is particularly the case when considering more subtle changes in a user's behaviour by performing a differential analysis.

A differential usage system requires information concerning the users history of behaviour plus a more recent sample of the mobile phones activities. An initial approach might be to extract and encode information from Toll Tickets and to store it in record format. This would require two windows or spans over the sequence of transactions for each user. The shorter sequence might be called the Current User Profile (CUP) and the longer sequence, the User Profile History (UPH). Both profiles could be treated and maintained as finite length queues. When a new Toll Ticket arrives for a given user, the oldest entry from the UPH would be discarded and the oldest entry from the CUP would move to the back of the UPH queue. The new record encoded from the incoming Toll Ticket would then join the back of the CUP queue.

Clearly it is not optimal to search and retrieve historical information concerning a user's activities prior to each calculation, on receipt of a new Toll Ticket. A more suitable approach is to compute a single cumulative CUP and UPH, for each user, from incoming Toll Tickets which can be stored as individual records, possibly in a database. So that we maintain the concept of having two different spans over the Toll Tickets without retaining a database record for each Toll Ticket, we will need to decay both profiles before the influence of a new Toll Ticket can be taken into consideration. A straight forward decay factor may not be suitable as this will potentially dilute information relating to encoded parameters stored in the user's profile. An important concern here is the potential creation of false behaviour patterns. Several decaying systems are currently being investigated.

Relevant toll ticket data

There are two important requirements for user profiling. At first, efficiency is of the foremost concern for storing the user data and for performing updates. Secondly, user profiles have to realise a precise description of user behaviour to facilitate reliable fraud detection. All the information that a fraud detection tool will need to handle is derived from the toll tickets provided by the network operator.

The following toll ticket components have been viewed to be the most fraud relevant measures:

- Charged_IMSI[1] (identifies the user)
- First_Cell_Id (location characteristic for mobile originating calls)
- Chargeable_Duration (base for all cost estimations)
- B_Type_of_Number (for distinguishing between national / international calls)
- Non_Charged_Party (the number dialed)

These components will continually be picked out of the toll tickets and incorporated into the user profiles in a cumulative manner.

It is also anticipated that the analysis of cell congestion can provide useful ancillary information.

Section 3: Rule-based approach to fraud detection

In ASPeCT, several approaches are taken to identify fraudulent behaviour. In the rule-based approach, both the absolute and differential usage are verified against certain rules. This approach works best with user profiles containing explicit information, where fraud criteria given as rules can be referred to. User profiles are maintained for the directory number of the calling party (A-number), for the directory number of the called party (B-number) and also for the cells used to make/receive the calls. A-number profiles represent user behaviour and are useful for the detection of most types of fraud, while B-number profiles point to hot destinations and thus allow the detection of frauds based upon call forwarding. All deviations from normal user behaviour resulting from the different analysing processes are collected and alarms will finally be raised if the results in combination fulfil given alarm criteria.

The implementation of this solution is based on an existing rule-based tool for audit trail analysis PDAT (Protocol Data Analysis Tool) [3]. PDAT is a rule based tool for intrusion detection developed by Siemens ZFE (Corporate Research and Development). PDAT works in heterogeneous environments, has the possibility of on-line analysis, and provides a performance of about 200 KB input per second. Important goals were flexibility and broad applicability, including the analysis of general protocol data, which is achieved by the special language PDAL (Protocol Data Analysis Language). PDAL allows the programming of analysis criteria as well as a GUI-aided configuration of the analysis at runtime.

Intrusion detection and mobile fraud detection are quite similar problem fields and the flexibility and broad applicability of PDAT are promising for using this tool for mobile fraud detection too. The main difference between intrusion detection and mobile fraud detection seems to be the kind of input data. The recording for intrusion detection produces 50 MB per day and user, but only for the few users of one UNIX-system. In comparison, fraud detection has to deal with a huge amount of mobile phone subscribers (roughly 1 Million), each of whom, however, produces only about 300 bytes of data per day. PDAT was able to keep all interim results in main memory, since only a few users had to be dealt with. For fraud detection, however, intermediate data has of course to be stored on hard disc. Because

[1] International Mobile Subscriber Identity

of these new requirements it was necessary to develop some completely new concepts such as user profiling and fast swapping for the updating of user profiles. Also, the internal architecture had to be changed to a great extent. The new architecture is depicted in figure 1.

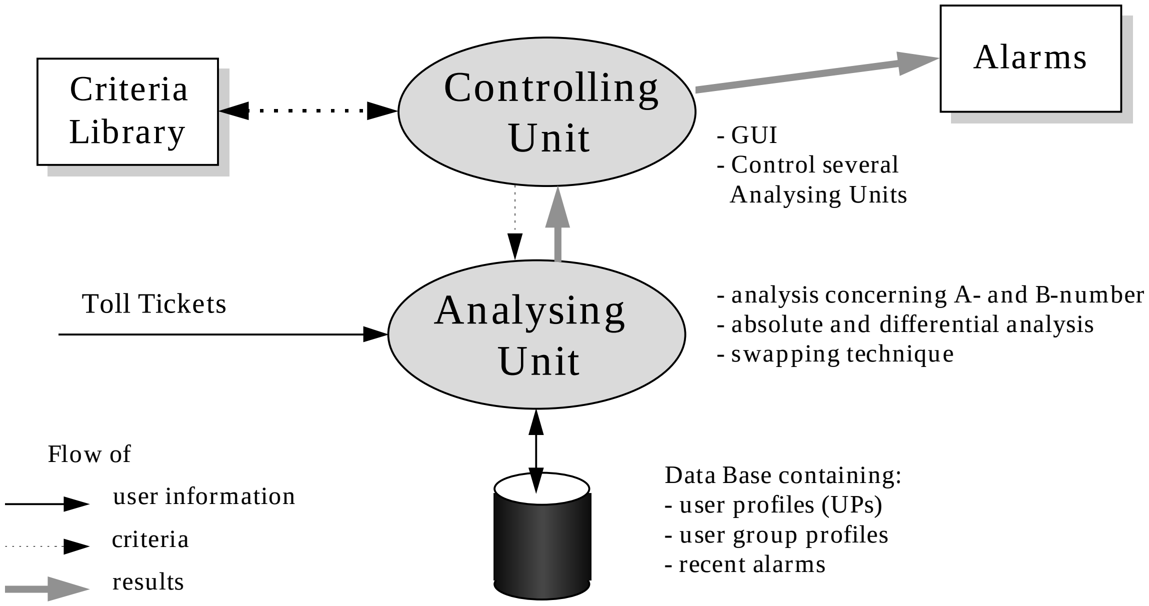

Figure 1 Architecture of rule-based fraud detection tool

Section 4: Neural network based approach to fraud detection

A second approach to identify fraudulent behaviour uses neural networks. The multiplicity and heterogeneity of the fraud scenarios require the use of intelligent detection systems. The fraud detection engine has to be flexible enough to cope with the diversity of fraud. It should also be adaptive in order to face new fraud scenarios, since fraudsters are likely to develop new forms of fraud once older attacks become impractical. Further, fraud appears in the billing system as abnormal usage patterns in the toll ticket records of one or more users. The function of the fraud detection engine is to recognise such patterns and produce the necessary alarms. High flexibility and adaptivity for a pattern recognition problem directly point to neural networks as a potential solution. Neural networks are systems of elementary decision units that can be adapted by training in order to recognise and classify arbitrary patterns. The interaction of a high number of elementary units makes it possible to learn arbitrarily complex tasks. For fraud detection in telephone networks, neural network engines are currently being developed worldwide. As a closely related application, neural networks are now routinely used for the detection of credit card fraud.

There are two main forms of learning in neural networks: unsupervised learning and supervised learning. In unsupervised learning, the network groups similar training patterns in clusters. It is then up to the user to recognise what class or behaviour has to be associated to each cluster. When patterns are presented to the network after training, they are associated to the cluster they are closest to, and are recognised as belonging to the class corresponding to that cluster. In supervised learning, the patterns have to be *a priori* labelled as belonging to some class. During learning, the network tries to adapt its units so that it produces the correct label at its output for each training pattern. Once training is finished the units are frozen, and when a new pattern is presented, it is classified according to the output produced by the network.

Unsupervised learning presents some difficulties. The problem is that patterns have to be presented - that is, encoded - in such a way that the data from fraudulent usage will form groups that are distinct enough from regular data. On the other hand, these systems can be trained using clean data only. With supervised learning, the difficulty is that one must obtain a significant amount of fraudulent data, and label it as such. This represents a significant effort. Further, it is not clear how such systems will handle new fraud strategies. Therefore, none of the approaches appears to be a priori superior to the other, and both directions are being investigated.

References

[1] ACTS AC095, project ASPeCT, *Initial report on security requirements*, AC095/ATEA/W21/DS/P/02/B, February 1996 .

[2] ACTS AC095, project ASPeCT, *Definition of fraud detection concepts*, AC095/KUL/W22/DS/P/06/A, September 1996.

[3] PDAT: Author 1, Author 2, Author 3, etc: Reference Title. Source, City, Date..