Detecting Fraud in Mobile Telephony Using Neural Networks

H. Grosser, P. Britos, and R. García-Martínez

Intelligent Systems Laboratory. School of Engineering, University of Buenos Aires,
Software & Knowledge Engineering Center (CAPIS) Graduate School, ITBA,
Computer Science PhD Program, Computer Science School, University of La Plata
rgm@itba.edu.ar

Abstract. Our work focuses on: the problem of detecting unusual changes of consumption in mobile phone users, the corresponding building of data structures which represent the recent and historic users' behaviour bearing in mind the information included in a call, and the complexity of the construction of a function with so many variables where the parameterization is not always known.

1 Description of the Problem

The existing systems of fraud detection try to consult sequences of CDR's (Call Detail Records) by comparing any field function with fixed criteria known as Triggers. A trigger, when activated, sends an alarm which leads to fraud analysts' investigation. These systems make what are known as a CDR's absolute analysis and they are used to detect the extremes of fraudulent activity. To make a differential analysis, patterns of behavior of the mobile phone are monitored by comparing the most recent activities to the historic use of the phone; a change in the pattern of behavior is a suspicious characteristic of a fraudulent act. In order to build a system of fraud detection based on a differential analysis it is necessary to bear in mind different problems: (a) the problem of building and maintaining "users' profiles" and (b) the problem of detecting changes in behavior. Pointing the first problem, in a system of differential fraud detection, information about the history together with samples of the most recent activities is necessary. An initial attempt to solve the problem could be to extract and encode Call Detail Records (CDR) information and store it in a given format of record. To do this, two types of records are needed; one, which we shall call CUP (Current User Profile) to store the most recent information, and another, to be called UPH (User Profile History) with the historic information [1, 2]. When a new CDR of a certain user arrives in order to be processed, the oldest arrival of the UPH record should be discarded and the oldest arrival of the CUP should enter the UPH. Therefore, this new, encoded record should enter CUP. It is necessary to find a way to "classify" these calls into groups or prototypes where each call must belong to a unique group. For the second problem, once the encoded image of the recent and historic consumption of each user is built, it is necessary to find the way to analyze this information so that it detects any anomaly in the consumption and so triggers the corresponding alarm.

M. Ali and F. Esposito (Eds.): IEA/AIE 2005, LNAI 3533, pp. 613 – 615, 2005.

2 Description of the Suggested Solution

In order to process the CDR's, a new format of record must be created containing the following information: IMSI (International Mobile Subscriber Identity), date, time, duration and type of call (LOC: local call, NAT: national call, INT: international call). For constructing and maintaining the "user's profiles", we have to fix the patterns that will make up each of the profiles. The patterns must have information about the user's consumption. We propose the use of SOM (Self Organizing Map) networks to generate patterns (creating resemblance groups) to represent LOC, NAT, and INT calls respectively [3]. The user's profile is built using the patterns generated by the three networks. The data used to represent a pattern are the time of the call and its duration. The procedure to fill the patterns consists of taking the call to be analyzed, encoding it and letting the neural network decide which pattern it resembles. After getting this information, the CUP user profile must be adapted in such a way that the distribution of frequency shows that the user now has a higher chance of making this type of calls. Knowing that a user's profile has K patterns that are made up of L patterns LOC, N patterns NAT and I patterns INT, we can build a profile that is representative of the processed call and then adapt the CUP profile to that call. If the call is LOC, the N patterns NAT and the I patterns INT will have a distribution of frequency equal to 0, and the K patterns LOC will have a distribution of frequency given by the equation $v_i = e^{-\|X-Q_j\|} / \left(\sum_{j=1}^{L} e^{-\|X-Q_j\|} \right)$ [2] where X is the encoded call to be processed; v is the probability that X call could be i pattern and Qi is the pattern i generated by the neural LOC network. If the call were NAT, then L must be replaced by N and the distribution of LOC and INT frequencies will be 0; if the call were INT, then L must be replaced by I and the distribution of LOC and NAT frequencied will be 0. The CUP and UPH profiles are compared using the Hellinger distance [3] in order to settle whether there have been changes in the pattern of behavior or not. The value of distance will establish how different must CUP and UPH be, in order to set an alarm going. By changing this value, there will be more or fewer alarms set off.

3 Results

The generated patterns after the training of the neural networks (LOC, NAC, INT) are shown as follows: Fig.1 shows 144 patterns corresponding to local calls, Fig.2 shows 64 patterns corresponding to national calls and Fig. 3 shows the 36 patterns corresponding to international calls. The construction of profiles and detection of changes in behavior are shown as follows: Fig. 4 shows a user's CUP at the moment an alarm was set off. It can be observed that the distribution of frequencies indicates a tendency to make local calls (patterns 1 to 144) and International calls (patterns 209 to 244), Fig. 5 shows the same user's UPH at the moment the alarm was set off. It can also be observed that the distribution of frequencies indicates a major tendency to make INT calls only (patterns 209 to 244).

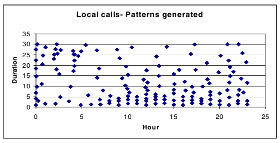

Fig. 1. LOC Patterns

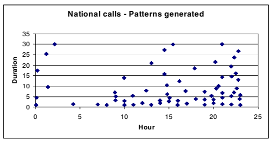

Fig. 2. NAC Patterns

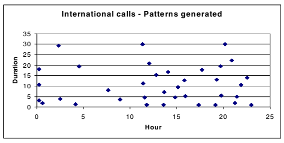

Fig. 3. INT Patterns

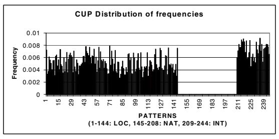

Fig. 4. User's CUP when an alarm was set off

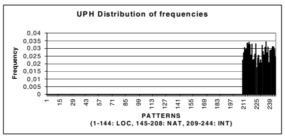

Fig. 5. User's UPH when an alarm was set off

By analyzing the detail of this user's calls from dates previous to the triggering of the alarm to the day it was set off, there is evidence that the alarm responded to the user's making only international calls till the moment that he started making local calls. When the number of local calls modified the CUP in the way illustrated by the graph, the alarm was triggered. If the user pays his invoice for international calls, this alarm is not an indicator of fraud, but it is an indicator of a sensitive change of behaviour in the pattern of user's consumption, and that is exactly what this system searches.

4 Conclusions

Though the change in behaviour does not necessarily imply fraudulent activity, it manages to restrict fraud analysts' investigation to this users' group. Applying to this group other types of techniques [1], it is possible to obtain, with a high degree of certainty, a list of users who are using their mobile phone in a " not loyal" way. It is also proven, with the experiences carried out, that the differential analysis provides with much more information than the absolute analysis, which can only detect peaks of consumption and cannot describe the user behavior in question.

References

1. ASPeCT. 1996. Advanced Security For Personal Communications Technologies. http://www.esat.kuleuven.ac.be/cosic/aspect/
2. Burge, P. and Shawe-Taylor, J. 2001. An Unsupervised Neural Network Approach to Profiling the Behaviour of Mobile Phone Users for Use in Fraud Detection. Journal of Parallel and Distributed Computing 61(7):pp. 915-925.
3. Hollmen J. 1996. Process Modeling using the Self-Organizing Map, Master's Thesis, Helsinki University of Technology, 1996.