Open access • Book Chapter • DOI:10.1007/978-3-642-39884-1_35

Synthetic logs generator for fraud detection in mobile transfer services

— Source link

Chrystel Gaber, Baptiste Hemery, Mohammed Achemlal, Marc Pasquet ...+1 more authors

Institutions: Orange S.A., Télécom ParisTech

Published on: 01 Apr 2013 - Collaboration Technologies and Systems

Topics: Mobile payment

Synthetic logs generator for fraud detection in mobile transfer services

Chrystel Gaber, Baptiste Hemery, Mohammed Achemlal, Marc Pasquet, Pascal Urien

Synthetic logs generator for fraud detection in mobile transfer services

C. Gaber[1 2], B. Hemery[2], M. Achemlal[1 2], M. Pasquet[2], and P. Urien[3]

[1] Orange Labs, France Telecom, 42 rue des Coutures, BP 6243, F-14066 Caen, France,
[2] UNICAEN, ENSICAEN, CNRS, UMR 6072 GREYC, F-14032 Caen, France
[3] Telecom Paristech, UMR 5141, 37/39 rue Dareau 75014, Paris, France
{chrystel.gaber,mohammed.achemlal}@orange.com

Abstract. This article presents a simulator which generates synthetic data for fraud detection. It models fraudsters and legitimate users.

Keywords: synthetic data, simulation, fraud detection

Mobile payments become more and more popular and, thus, are very attractive targets for fraudsters. As new ways to commit crimes and avoid detection appear, research in the field of fraud is always evolving. Yet, research in fraud detection is limited as publicly available transactional databases containing frauds and groundtruth are scarce [1]. The main cause is that stakeholders are very reluctant to disclose information about frauds and their clients. We address this issue by generating synthetic data of a mobile transaction system. Done in the scope of the European FP7 project MASSIF, our model is based on the mobile-based transaction system and its users described by the MASSIF scenario providers in [2]. We create a tool which can be used by the project contributors and the fraud detection community. Synthetic data are not commonly used in the field of fraud detection although there is a lack of test data. To our knowledge, only one method is used to generate synthetic data for the training and testing of fraud detection systems [3]. Compared to it, ours enables to highlight specific characteristics of fraud detection algorithms because it is not necessarily set up with parameters driven from real data. We model the mobile money transfer platform and the behavior of regular users and fraudsters. Regular users behavior is modeled as a set of habits whereas fraudsters behavior is based on attack patterns. Only superimposed frauds were modeled [1]. A first prototype based on multi-agent models was implemented. We evaluated the model and the data created with the implemented simulator. The period and amount parameters are overfitted. It results in frequent transactions of higher value but a correct value of the total amount of money spent during the simulation. The outcome is encouraging and sets a reference in this field.

References

1. Richard J. Bolton and David J. Hand. Unsupervised profiling methods for fraud detection. In *Conference on credit scoring and credit control*, 2001.
2. M. Achemlal, S. Gharout, C. Gaber, M. Llanes, E. Prieto, R. Diaz, L. Coppolino, A. Sergio, R. Cristaldi, A. Hutchison, and K. Dennie. Scenario requirements. http://www.massif-project.eu/, March 2011.
3. E.L. Barse, H. Kvarnstrom, and E. Jonsson. Synthesizing test data for fraud detection systems. In *Computer Security Applications Conference, 2003. Proceedings. 19th Annual*.