Using Simulation for Computer Security

By Edgar Lopez-Rojas
Blekinge Institute of Technology

ABSTRACT

We present a financial simulation model covering three related financial domains: Mobile Payments, Bank Transactions and Retail Stores systems. One of the advantages of using simulators is the generation at will of synthetic data sets. These synthetic data set can be used to further advance fraud detection research, without leaking sensitive information about the underlying data. Using statistics and social network analysis (SNA) on real data we can calibrate the relations between staff and customers, and generate realistic synthetic data sets. The generated data represents real world scenarios that are found in the original data with the added benefit that this data can be shared with other researchers for testing similar detection methods without concerns for privacy and other restrictions present when using the original data.

The problem we address in each of these domains concern different types of financial fraud. We limit ourselves to isolated cases of relatively straightforward fraud. However, in this research the ultimate aim is to cover more complex types of fraud, such as money laundering, that comprises multiple organizations and domains. Therefore the variety of domains we covered are all related with financial transactions.

Fraud is an important problem that impact the whole economy. Currently, there is a general lack of public research into the detection of many different types of fraud. One important reason is the lack of transaction data which is often sensitive. To address this problem we present a Mobile Money Simulator (PaySim), a Bank Simulator (BankSim) and Retail Store Simulator (RetSim), which allow us to generate synthetic transactional data.

Our simulations are based on the paradigm of multi agent based simulations (MABS). Hence, we developed agents that represent the clients, staff and merchants. The normal behavior is based on behavior observed in data from the field, and is codified in the agents as rules of transactions and interaction between clients, or customers and salesmen. Some of these agents were intentionally designed to act fraudulently, based on observed patterns of real fraud. We introduced known signatures of fraud in our model and simulations to test and evaluate our fraud detection results. The resulting behavior of the agents generate a synthetic log of all transactions as a result of the simulation. This synthetic data can be used to further advance fraud detection research, without leaking sensitive information about the underlying data.

One case of the use of simulation for computer security is RetSim. We used the RetSim simulator to model two common retail fraud scenarios to ascertain exactly how effective the simplest form of statistical threshold detection commonly in use could be applied to effectively choose appropriate thresholds (i.e. Thresholds for a *Triage* fraud detection model). The preliminary results show that threshold detection is effective enough at keeping fraud losses at a set level, that there seems to be little economic room for improved fraud detection techniques.