Online Banking Fraud Detection Based on Local and Global Behavior

Stephan Kovach
Laboratory of Computer Architecture and Networks
Department of Computer and Digital System
Engineering, Polytechnic School of São Paulo
São Paulo, Brazil
skovach@larc.usp.br

Wilson Vicente Ruggiero
Laboratory of Computer Architecture and Networks
Department of Computer and Digital System
Engineering, Polytechnic School of São Paulo
São Paulo, Brazil
wilson@larc.usp.br

Abstract – **This paper presents a fraud detection system proposed for online banking that is based on local and global observations of users' behavior. Differential analysis is used to obtain local evidence of fraud where a significant deviation from normal behavior indicates a potential fraud. This evidence is strengthened or weakened by the user's global behavior. In this case, the evidence of fraud is based on the number of accesses performed by the user and by a probability value that varies over time. The Dempster's rule of combination is applied to these evidences for final suspicion score of fraud. Our main contribution is a fraud detection method based on effective identification of devices used to access the accounts and assessing the likelihood of being a fraud by tracking the number of different accounts accessed by each device.**

Keywords- differential analysis; local and global behavior; device identification; Dempster-Shafer theory

I. INTRODUCTION

Fraud prevention describes the security measures to avoid unauthorized individuals from initiating transactions on an account for which they are not authorized [1]. In spite of many advanced mechanisms available for fraud prevention for online banking applications, it can fail. Fraud detection consists in identifying such unauthorized activity once the fraud prevention has failed. In practice, fraud detection must be used continuously, since the system is unaware that fraud prevention has failed [1]. Among the approaches used by fraudsters, phishing is one of the most common forms for stealing account details for authentication from the customers. Social engineering is the most common method used in phishing. Social engineering usually comes in the form of e-mails trying to convince users to open attachments or by directing them to some fraudulent site**,** and most of the time it is so well designed that many costumers are led to informing their account details**.**

This paper presents a framework, and the corresponding system, for online banking fraud detection in real time. It uses two complementary approaches for fraud detection. In the differential analysis approach, the account usage patterns are monitored and compared with the history of its usage, which represent the user's normal behavior. Any significant deviation from the normal behavior indicates a potential fraud [2].

In the global analysis approach, each device is monitored and classified as legitimate or fraudulent with certain probability based on global information. This is based on three assumptions. First, it is assumed that each device used for online banking has a single identification. The second assumption is based on the fact that the probability of a transaction being a fraud increases with the number of accounts accessed by the same source that requested the current transaction. The third assumption comes from the fact that the only way to know that a fraud has been perpetrated is when the customer reports it.

The major contribution of this paper is the finding, by empirical analysis of a real-world transaction dataset, that the effective identification of access devices and monitoring the number of different accounts accessed by each device is a very promising supplement for other methods in detecting fraudulent behavior in online banking applications.

This paper is organized as follows: Section 2 presents an overview of related work on fraud detection. Section 3 describes some characteristics of online banking frauds. Section 4 details the proposed fraud detection system. Section 5 concludes the paper and outlines future work.

II. RELATED WORK

There are few published works about fraud detection within the domain of online banking applications. This is most likely due to the privacy, the secrecy and the commercial interests concerning this domain, rather the absence of research [3]. Therefore, due to the limited exchange of ideas, the development of new fraud detection methods in the banking area is difficult. Most published work is related to the domain of credit card, computer intrusion and mobile communication. Some relevant works on fraud detection are reviewed next.

Credit card frauds- Most of the works on preventing and detecting credit card fraud were carried out with special emphasis on data mining and neural networks. Aleskerov, Freisleben and Rao [4] describe a neural network based database mining system in which a neural network is trained with the past data of a particular customer and the current spending patterns is processed to detect possible anomalies. However, Bolton and Hand [5] proposed a detection technique in which break point analysis is used to identify changes in spending behavior.

 ISBN: 978-1-61208-116-8

Computer intrusion- Intrusion detection approaches in computers is broadly classified into two categories based on a model of intrusions: misuse and anomaly detection. Misuse detection attempts to recognize the attacks of previously observed intrusions in the form of a pattern or a signature and then monitors such occurrence. Anomaly detection tries to establish a historical normal profile for each user, and then uses sufficiently large deviation from the profile to indicate possible intrusions [6]. Denning [7] presents a statistical model for real-time intrusion detection based in anomaly detection. Ghosh and Schwrtzbard [8] describe an approach that employs artificial neural networks used for both anomaly and misuse detection.

Mobile communication frauds - Fraud in communication networks refers to the illegal access to the network and the use of its services. Cortes and Pregibon [9] define statistical summaries, denominated signatures, of users over two time windows, namely, current and historical, respectively. The current network activity is compared with the historical activity for any deviation. Fawcett and Provost [10] present rule-based methods and neural networks for detecting fraudulent calls based on profiling subscriber behavior.

In all domains above mentioned, fraudsters tends to adapt to new prevention and detection measures. In the same way, legitimate users may gradually change their behavior over a longer period of time. Therefore, fraud detection techniques need to be adaptive and to evolve over time in order to avoid false alarms. Models can be updated at fixed time points or continuously over time [9][10].

Panigrahi, Kundu, Sural, and Majumdar [11] describe a framework for fraud detection in mobile communication networks using rule-based deviation method. The main point of this paper is the detailed description of the use of Dempster-Shafer theory in order to combine the evidences of fraud given by two rules.

The system proposed in this paper combines three different approaches: (1) differential analysis using statistical models in order to detect local evidence of fraud; (2) an innovative approach using a probabilistic model for evaluating the likelihood of a transaction being a fraud based on its global behavior; and (3) Dempster-Shafer theory for combining evidences of fraud.

III. ONLINE BANKING FRAUD CHARACTERISTICS

An empirical analysis performed on real-world transactions datasets revealed that most of frauds had the following behavior characteristics:

- Large number of different accounts accessed by a single fraudster;
- Transactions involving small values in many accounts;
- More payment transactions than usual in a single account;
- Increased number of password failures before the occurrence of frauds.

While the latter two characteristics can be detected by differential analysis using local attributes, the first two characteristics need information about similar attacks in other accounts. The fraud detection system described in the next section takes these characteristics into account.

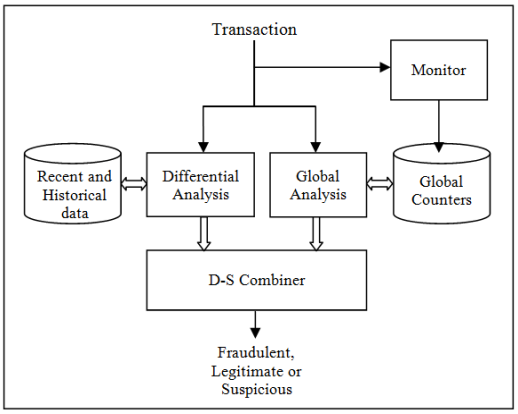

Figure 1. The general architecture of the system

IV. THE FRAUD DETECTION SYSTEM

The general architecture of the proposed fraud detector is illustrated in Fig. 1. In this architecture, each access device from which transactions are performed is supposed to have an identity. These identities are used along with a set of counters to monitor the number of different accounts accessed by each device. The system uses two independent approaches for detecting frauds: a differential analysis approach that detects significant changes in transaction patterns in individual accounts, and a global analysis approach that uses the set of counters to detect unusual number of accounts accessed by a single device. The fraud evidences determined by the two approaches are then combined in order to determine an overall score that may trigger an alarm depending on a prefixed threshold. The main issues of the architecture are discussed in the next subsections.

A. *Device Identification*

The proposed fraud detection technique has as a core concept, the notion of access device identity.

In the domain of online banking, where accesses are made through the Internet, the identification of source devices based in IP address only is difficult since it can change over time. In the proposed approach, the identification of the access device is made by a component that must be downloaded and installed in the client device. This component generates a *fingerprint* of the access device and sends it to the bank site as part of each transaction data. The *fingerprint* is calculated by applying a cryptographic function on the hardware and software information, as the processor and the operating system serial numbers, MAC address, and some configuration details.

The implementation details of the component are out of the scope of this paper. It assumes that the component is implemented with three main requirements:

1. It generates a different fingerprint for each different access device;

 ISBN: 978-1-61208-116-8

2. It introduces some randomness during the fingerprint generation in order to difficult its spoofing by other devices;
3. It informs the new fingerprint whenever the configuration of the device changes.

Actually the proposed system is based on the component that is already being used by the actual online banking system.

B. Global Behavior and the Monitor

The observation of the user's global behavior plays a major role in the fraud detection system proposed herein. An example of global behavior that may evidence a fraud is the large number of different accounts accessed by a single device. Another example is the occurrence of login fails over many accounts using a single trial password. A set of counters are used in order to verify the global behavior of the users. As shown in Fig. 1, the monitor accounts for updating these counters at each incoming transaction.

C. Differential Analysis

In the differential analysis approach, an incoming transaction is examined against a set of profiles that characterize the normal usage pattern of a legitimate customer. If the current usage pattern deviates significantly from the customer's average usage pattern, it may indicate a potential fraud. In order to calculate this deviation, two buffers are used in such a way that all transactions submitted in the current session enter the first buffer. The second buffer keeps a certain number of most recent transactions. The transactions in the first and second buffer are used to calculate the current usage pattern and the customer's average usage pattern, respectively. Then, the deviation is calculated using a statistical method the result of which is a probabilistic value that gives a degree of belief in the evidence of fraud. If this session is classified as legitimate, all the transactions of the first buffer are inserted in the second buffer and the oldest transactions are removed from it [12].

Some of the profiles monitored by this module are described below:

- *Payment transaction frequency*. This profile is monitored in order to detect the sudden increase of payment transactions, which is unusual to legitimate a user.
- *Password failures*. The measure of password failures at login time is compared against a fixed limit that is determined from prior observations. This profile is useful for detecting attempted break-ins [7].
- *Login frequency*. Profiles for login frequencies by day and time are monitored to detect fraudsters who try to log into an unauthorized account during a period of time when the legitimate user is not expected to be using the account [7].

D. Global Analysis

The purpose of the global analysis module is for strengthening or weakening the evidence of a fraud determined by the differential analysis module. It is performed by evaluating a new evidence of fraud based on global observation of the user's device behavior. The evidence of fraud, given by a probability, is determined by means of three lists: Black List, White List and Suspect List. The Black List contains the identity of devices associated to transactions that have already been classified as fraudulent. The White List contains the identities of the devices, as well as the account numbers accessed by them, associated to transactions classified as legitimate. The Suspect List contains the identities of devices the transactions of which have not yet been classified. The assignment of the devices to one of those lists and the determination of its fraud probability score are driven by rules described as follows:

For each incoming transaction,

- If the current device is in the Black List, then the fraud probability is assigned to one meaning that the transaction is fraudulent with a high level of evidence;
- If the current device and the account number accessed by it are in the White List, then the fraud probability is assigned to zero denoting that the transaction is legitimate with high level of evidence. Note that the device identity may be associated with one or more accounts in the White List. This is the case in which a single user has many accounts;
- If the current device is neither in the Black List nor in the White List, then the device identity and the accessed account number are included in the Suspect List. While in this list, the fraud probability of this transaction is determined by an exponentially decaying function described in the next section.

E. The Suspect List and the Exponentially Decaying Function

If the incoming transaction device is inserted in the Suspect List, it will remain there until explicitly classified as fraudulent or legitimate, when the associated device identity and account number are inserted in the Black or White List, respectively. The idea behind this rule comes from the fact that a given transaction can only be assured as fraudulent by the customer himself/herself.

If no fraud is reported until the end of a prefixed period of time, nothing can be said about the trustiness of this device. In this case, the device will be moved to the White List since it is more likely to be legitimate based on the analysis made on real-world transactions dataset. However, a flag is set indicating that this device was moved to the White List at the end of predefined period of time and not explicitly classified as legitimate. This flag is used by the fraud analyst if a fraud performed from this device is detected later. Since the device had been moved to the White List, the next transactions performed by this device will be regarded as legitimate by this module.

The elapsed time since the occurrence of a fraud and its detection by the customer can take more than a month. According to the information from analysts of a real banking institution, there are some cases in which it takes up to two

 ISBN: 978-1-61208-116-8

month to be reported by the customer. The reason for this delay is due to the fact that many fraudulent transactions go unnoticed since the values involved in individual transactions are usually very small.

When a device is included in the Suspect List, an initial value is assigned to the fraud probability. This value is calculated by an exponentially decaying function that depends on the number of different accounts that were accessed by this device the transactions of which have not yet been classified. If a fraud on any of these accounts is reported by the customer, the associated device identity will be moved to the Black List.

The exponentially decaying function was chosen due to the fact that most of the frauds are reported as soon as they were committed and very few at the end of some period of time, for example, two months later. In other words, the probability of being a fraud is higher at the beginning of a transaction, decaying at a fast rate along the time.

The exponentially decaying function is expressed as

$$P(t) = P_{max} \cdot e^{-\lambda t} \quad (1)$$

where,

P_{max} is the maximum probability value assigned to the device when it is included in the list. It depends on the number of different accounts accessed by the device (N), since the probability of being a fraud increases with this number. For the initial trial, P_{max} was chosen as being equal to N/10 for $1 < N < 8$, and 0.9 for $N \geq 9$. The maximum value of 0.9 was chosen since 1.0 is reserved for assured fraudulent devices, i.e., included in the Black List.

λ is calculated such that at the end of the period (t_{end}), the probability value reaches an arbitrary low value. Assuming $t_{end} = 60$ days and $P(t_{end}) = 0.01$, then

$$\lambda = - (1/60).\ln(0.01/P_{max}) \quad (2)$$

Fig. 2 shows the exponentially decaying curves for each value of N.

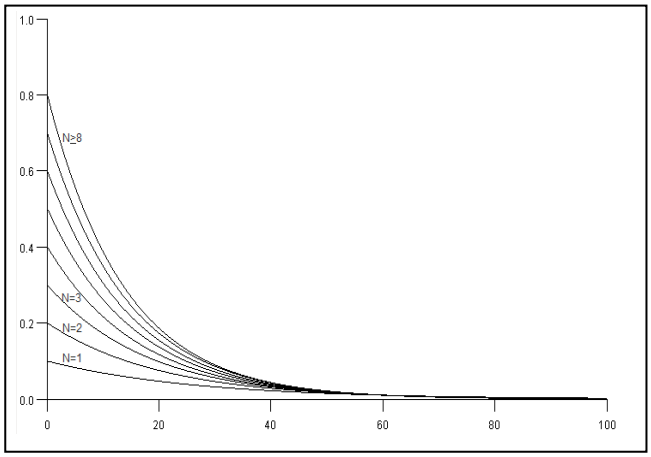

Figure 2. Exponentially decaying curves

The dashed line in Fig. 3 shows an example of the probability values assigned to a device that varies over the time. Note that when the number of accounts accessed by it increases, the probability value jumps to its correspondig maximum value.

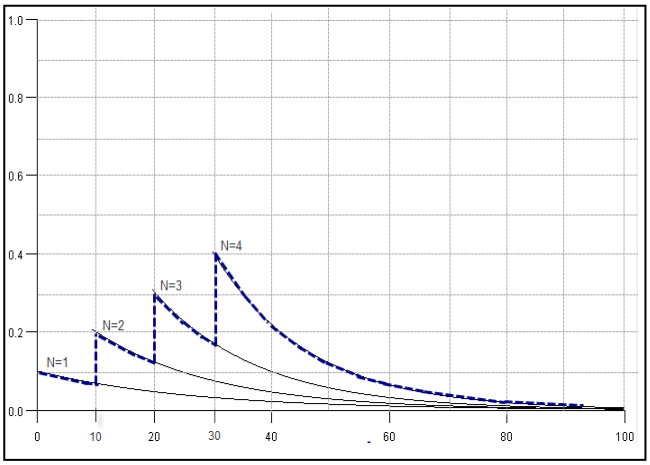

Figure 3. The fraud probability under exponential decaying function

F. *Dempster-Shafer Combiner*

This module uses the Dempster-Shafer theory to combine the evidences of fraud estimated by differential and global analysis modules and computes the overall suspicion score of a transaction.

The Dempster-Shafer (D-S) theory is a mathematical theory of evidence that provides a formal framework for combining sources of evidence [13].

The main difference between the D-S theory and the probability theory is that the former allows the explicit representation of uncertainty. The other difference is that the D-S theory requires no knowledge of prior probabilities.

The D-S framework is based on a view that hypotheses can be regarded as a subset of a given set of mutually exclusive and exhaustive possibilities named a *frame of discernment* [13]. For the fraud detection domain, the frame of discernment Θ is consisted by two mutually exclusive values, given as: $\Theta = \{fraud, \text{-}fraud\}$. The set of all possible hypotheses of Θ corresponds to all subsets of Θ including itself, denoted by 2^{Θ}. In the case of fraud, the power set is consisted by three possible hypotheses: {*fraud*}, {*-fraud*} and $\Theta = \{fraud, \text{-}fraud\}$ (denoting the uncertainty).

A probability number *m(h)* between 0 and 1, expressing an estimative of confidence or belief, is assigned to a hypothesis *h.* This number is called *basic probability assignment (bpa)* or *mass*. In our system, the probabilistic values computed by the local and global analysis modules are applied to basic probability assignments.

Two functions are defined in the D-S theory in order to express uncertainty: *Belief function (Bel) and Plausibility function (Pls)*

Belief function (Bel) is the total belief committed to a hypothesis. It sums the mass of all non-empty subsets of the hypothesis and the mass of hypothesis itself.

Plausibility function (Pls) takes into account all the masses assigned to a hypothesis and those that can be plausibly transferred to it in the light of new information [13]. It defines the maximum belief that can be committed to a hypothesis.

Belief and Plausibility functions are related as follow:

$Pl(H) = 1 - Bel(-H);$

$U(H) = Pl(H) - Bel(H).$

where, *Bel(-H)* means *disbelief of H*, i.e., belief that refutes the hypothesis *H*; and *U(H)* means uncertainty of *H*.

Bel(H) and Pls(H) represents the upper and lower bounds in the evidence of hypothesis *H*.

The Dempster's rule of combination gives a function for evaluating an overall score from two evidences. Given two basic probability assignment of evidences $m_1(h)$ and $m_2(h)$, they may be combined into a third basic probability assignment $m_3(h)$ by the expression below:

$$m_3(h) = m_1(h) \oplus m_2(h) = \frac{\sum_{x \cap y = h} m_1(x) \cdot m_2(y)}{1 - \sum_{x \cap y = \varnothing} m_1(x) \cdot m_2(y)}$$

where, the symbol $\oplus$ denotes orthogonal sum.

This rule can be used for combining basic probability assignment of all features monitored by the local and global analysis modules and then obtaining overall summary values for each module. These summary values from both modules are then combined to provide the final suspicion score [14]. Based on this score, a transaction on a given account can be detected as fraudulent, legitimate or suspicious.

V. CONCLUSION AND FUTURE WORK

In this paper, we have introduced a novel approach for fraud detection in online banking transactions by using global counters and an effective identification of access devices. The idea behind this approach comes from the fact that fraud suspicion in a transaction increases with the number of accounts accessed from the same source. The effective identification of devices is made by a component that is downloaded and installed in each device during its first access to the bank. A monitor counts the number of different accounts accessed by each device. These counters are then used by the global analysis module that estimates the likelihood of a transaction being a fraud. The paper describes the details of how this likelihood is evaluated. A differential analysis is also performed on each transaction against a set of customer profiles. This approach is based on the proposition in which any significant deviation from the normal behavior indicates an evidence of fraud. The Dempster's rule combines the resulting fraud evidences from global and differential analysis to calculate the overall suspicion score of each transaction.

The proposed system is very promising in detecting fraudulent transactions in online banking applications with low rate of false alarms.

The benefit of this approach comes from the fact that most of fraudsters do not attack a single account, but many accounts from a single device. Therefore, the simple observation of a device's global behavior, such as the number of different accounts that has been accessed by it, can bring more evidences rather than just applying complex statistical methods on its local parameters.

Currently, the system is in its final stage of development. It is being validated and its parameters adjusted using a real-world transaction dataset.

Among the directions for future work we are regarding the development of a *simulator* that produces different patterns of legitimate and fraudulent transactions in any proportion and randomness in order to evaluate the best threshold values for low rate of false alarms, and the study of new algorithms and probabilistic functions for global analysis.

REFERENCES

[1] R. J. Bolton and D. J. Hand, "Statistical fraud detection: A review". Statistical Science. Vol. 17, No. 3, 2002, pp 235-255.

[2] U. Murad and G. Pinkas, "Unsupervised profiling for identifying superimposed fraud", in Proceedings of the 3rd European Conference on Principles of Data Mining and Knowledge Discovery, 1999, pp. 251-266

[3] K. N. Karsen and T. G. Killingberg, "Profile based intrusion detection for Internet banking systems", Master Thesis, Norwegian University of Science and Technology, Norway, 2008

[4] E. Aleskerov, B. Freisleben, and B. Rao, "CARDWATCH: A neural network based database mining system for credit card fraud detection", in Computational Intelligence for Financial Engineering. Proceedings of the IEEE/IAFE, 1997, pp 220- 226. IEEE, Piscataway, NJ.

[5] R. J. Bolton and D. J. Hand, "Unsupervised profiling methods for fraud detection", in Conference on Credit Scoring and Credit Control 7", Edinburgh, UK, 5-7 Sept., 2001.

[6] Y. Kou, C.T. Lu, S. Sirwonqattana, and Y.P. Huanq, "Survey of fraud detection techniques", in Proceedings of the IEEE International Conference on Networking, Sensing and Control, vol. 1, 2004, pp. 749-754.

[7] D. E. Denning. "An intrusion detection model". IEEE Transactions on Software Engineering, 13:222-232, February 1987.

[8] A. K. Ghosh and A. Schwartzbaxd. "A study in using neural networks for anomaly and misuse detection", in Proceedings of the 8th USENIX Security Symposium, 1999.

[9] C. Cortes and D. Pregibon, "Signature-based methods for data streams," Data Mining and Knowledge Discovery, vol. 5, no. 3, pp. 167-182, 2001.

[10] T. Fawcett and F. Provost, "Adaptive fraud detection", Data Mining and Knowledge Discovery Journal, Kluwer Academic Publishers, Vol. 1, No. 3, 1997, pp. 291-316.

[11] S. Panigrahi, A. Kundu, S. Sural, and A. K Majumbar, "Use of Dempster-Shafer theory and Bayesian inferencing for fraud detection in communication networks", Lecture Notes in Computer Science, Spring Berlin/ Heidelberg, Vol. 4586, , 2007, p.446-460.

[12] P. Burge and J. Shawe-Taylor, "Detecting cellular fraud using adaptive prototypes", Proceedings of the AAAI-97 Workshop and AI Approaches to Fault Detection and Risk Management. Mento Park, CA: AAAI Press, 1997, pp. 9-13.

 ISBN: 978-1-61208-116-8

[13] Y. Wang, H. Yang, X. Wang, and R. Zhang, "Distributed Intrusion detections Based on data fusion method.", in Proceedings of the 5th World Congress on Intelligent Control and Automation, 2004, pp. 4331–4334.

[14] Q. Chen and U. Aickelin, "Anomaly detection using the Dempster-Shafer method," in Proc. of the 2006 International Conference on Data Mining, DMIN 2006, 2006, pp. 232–240.

 ISBN: 978-1-61208-116-8