FIGHTING IDENTITY FRAUD WITH **DATA MINING**

Groundbreaking means to prevent fraud in identity management solutions

SAFRAN
Morpho

Contents

Executive summary

According to the USDOJ[1] "Identity theft and identity fraud are terms used to refer to all types of crime in which someone wrongfully obtains and uses another person's personal data in some way that involves fraud or deception, typically for economic gain." Forgery and use of fraudulent identity documents are major enablers of Identity Fraud. Those are growing concerns for governmental bodies all over the world.

Indeed, the forgery and use of fraudulent identity documents are national security issues. Lack of reliable means to identify individuals has paved the way to terrorist attacks, since September 11th, 2001.

The trafficking of identity documents is also a major internal security issue, and a threat to democracy. The forgery of identity documents has become a lucrative and under-penalized way of making money for organized crime networks. Citizens have to prove their identity to benefit from key rights, that contribute to define good living conditions and democracy for everyone: the right to cross borders (passports and electronic identity cards), the right to vote (election cards), to benefit from social welfare and medical care (electronic healthcare and welfare cards), and the right to travel (electronic driving licenses).

Therefore, technical solutions and good practices, that are described in this paper, are essential to fight the enablers of Identity Fraud: the forgery and use of fraudulent identity documents. For governments and citizens, Identity Fraud is more than a security issue: it is a financial burden; it threatens the trust of the citizens and of other nations in security and democracy, and it causes severe psychological damages to the victims of fraud. In some cases, the best known security solutions are not sufficient to fight fraud efficiently. This paper describes those cases. Finally, innovative technical solutions and good practices, that complement the best known security solutions, are explained. This paper puts a special emphasis on how those technical solutions and good practices fill the security gaps of usual identity documents issuance and utilization solutions.

[1] http://www.justice.gov/criminal/fraud/websites/idtheft.html, retrieved on November, 24th 2010

The impact of identity fraud?

The manufacturing and use of forged identity documents is a financial burden for governments, social welfare institutions, and financial institutions.

The potential savings, and return of investment in efficient solutions to fight fraud, are significant: in fact, in 2009, EHCN (European Healthcare Fraud & Corruption Network) evaluated that healthcare fraud costs 56 billion in Europe annually and € 180 billion globally[2] ; in 2006, the cost of identity fraud in the US was valued at $ 15,6 billion by the US Federal Trade Commission[3].

The cost of fraud consists in:

- Direct costs for the governmental institutions or private organizations: payment of fraudulent claims…
- Indirect costs for the end-users, the governmental institutions or private organizations: higher premiums, payment of fraudulent debts, legal fees, lost wages, and time spent to fix the impact of fraud

Furthermore, for every government, granting a reliable and trusted identity to national citizens is a key mission of the state: suspicion of fraud in major national events, like elections, can undermine the trust of the citizens and the international community, in the power and in the legitimacy of the state. This is especially the case in nascent democracies.
Last but not least one should not neglect the psychological impact of fraud. The psychological damages of somebody committing crimes on your own name, and the damages on your reputation, may be long to recover.

[2] The financial cost of Healthcare fraud, EHCN, 2010
[3] Federal Trade Commission – 2006 Identity Theft Survey Report, Synovate, Nov. 2007

The forgery of identity documents disclosed!

The following "weak links" in the chain of trust can lead to issuance of fraudulent identity documents:

The theft of genuine blank documents, which are forged over fraudulent production lines. This risk should be mitigated by secure handling and transportation, traceability mechanisms for blank documents, and innovative security features on the documents themselves.

A corrupt staff.
The security policy of the personalization site and relevant management of enrolment and personalization personnel should guarantee that a high level of confidence can be applied to the personal data that is handled by the system. Besides, every action that is performed on every identity document should be imputable to each operator.

The modification of a genuine identity document or the forgery of an identity document "from scratch".
Insufficiently secure documents can be altered, or reproduced easily. A relevant combination of security features, security mechanisms implemented in the chip, and enhanced control means prevents the forgery of identity documents. Biometrics is a reliable way to control that the handler of the document and the person to who it was issued are the same person. Cryptographic objects in the identity documents allow governmental authorities to check "in the field" that a document was issued by an official entity.

The issuance of multiple identities.
The search of duplicates in the citizens database during the enrollment process thanks to biometrics ensure that each citizen is enrolled only once in the database.

Nevertheless, a mix of best-known security solutions and best practices in the issuance process is sometimes not sufficient to ensure a proper protection against the fraudulent use of the identity documents. ***Even state of the art features are no silver bullet for :***

- The illegitimate use of a genuine identity document.

- The issuance of a genuine identity document, under a "synthetic" identity. A "synthetic" identity is made after several altered or forged identity documents, which are used to prove one's identity at the enrollment step. "Synthetic identity" fraud is the act of creating a virtual identity, to perpetrate criminal activities.

Specific conditions in the field or lack of appropriate infrastructure explain the failure of the best-known security solutions and best practices in the issuance process :

- All means of verification may not be used at each point of control. Because of economic reasons, technical constraints, or a lack of a persistent political will to invest in security infrastructure, security features are not checked appropriately.

- For regulatory or cultural reasons, some countries may decide to postpone the use of biometrics in their identity documents.

- In some areas, people with a specific condition are not able to produce a useable biometrics (people with amputated fingers or hands, people with severely damaged fingerprints…).

- The enrollment process might not be under full control :

  - There might be a risk for the collusion of operators with a fraudulent documents production network. This favors gaps in the enrollment process (e.g. no biometrics enrolled), or inaccuracy of some information enrolled (e.g. fraudulent modification of the family name or of the birth date).

  - The process of verification of the identity of the citizen at the enrollment step leaves room for identity theft and/or enrollment under "synthetic identity".

Data Mining : a complement to biometrics for fraud detection and deterrence

How Data Mining differs from others database security solutions?

For those reasons, it is the whole process of document issuance and utilization which has to be secure, beyond the document itself. The purpose of Data Mining tools is to have a fine-grained, adaptable understanding of the personal data of the people, and of the activity logs of the document issuance system. Data Mining enables a holistic view on the data related to one citizen, from the enrollment step to each transaction made with the identity documents. In fact, each transaction made with an electronic identity document leaves an electronic trace, when the document is used for verification at a border check point, or to sign an electronic document on an e-government site for example. Those traces are usually stored for traceability purposes, but not analyzed. The purpose of Data Mining tools is to monitor these traces, to help detect fraud.
Data Mining tools enable **fraud deterrence** by detecting anomalies in the document issuance process, in real-time or near real-time. **Fraud deterrence** features are based on static business rules enforcement systems, and predefined consistency checks on the personnal data provided by the applicant at the enrollment step.
Even if **fraud deterrence** is necessary, fraudsters quickly work around security measures, because of leaks of confidential information, collusion or corruption. They easily adapt to statically defined checks on data consistency or process integrity. This is the reason why **fraud deterrence** features of Data Mining tools, are not sufficient to fight fraud efficiently, and to adapt to the newest fraudulent behaviors.

To adapt and learn previously unknown fraud patterns, any Data Mining solution should implement the following fraud detection features:

- **Association.**

It is the ability to identify and track patterns, where one event is connected to another event. For example, too many manual validations of documents for which a duplicate has been identified, by the same officer, in the same slot of the day (at the end of a production shift), is a suspicious pattern that may turn out to be fraudulent after investigation.

- **Clustering and classification.**

It is the ability to identify new patterns in the data. For example, a suspiciously high rate of travel documents issued for individuals coming from a low income neighborhood, which may home a fraud network.

What kind of data can be analyzed by Data Mining ?

A Data Mining solution should rely on sophisticated mathematic algorithms (neural networks, decision trees...) to analyze data coming from different sources:

- Personal data of the applicants to the documents issuance process. This data can include alphanumerical data: the name of the person, location of birth, the adress of residence, the personal data of relatives…and numerical data: the date of birth, size…
- Audit logs of the documents production process. Each step of the enrolment and issuance workflow is logged in a database, for each document issued. The actions performed by the operators on the database, and the actions performed on a document (for example, cancellation of a document reported as stolen) are also logged.

External databases such a Civil Registries, database of stolen or fraudulent documents, of convicted criminals, databases with geographic or sociologic profiles of convicted fraudsters, can be used to improve the accuracy of the analysis. For investigation, monitoring, or traceability purpose, IT services in governmental bodies produce huge amounts of data, and have no means to make the most of it. Data Mining tools natively handle this data from external databases: they merge and adequately prepare this data for an adequate statistical analysis. Each time new information are gathered on somebody, or each time a citizen uses his card, significant information are used to know the person better, and predict how he will behave more accurately.

Is it necessary to have a good understanding of fraud patterns to detect them?

Data Mining tools implement two techniques of pattern analysis and fraud discovery:

- **Unsupervised learning.**

Algorithms are able to differentiate outliers from nominal patterns, and isolate them for investigation. After human analysis by a fraud analyst, if the fraud is confirmed, search queries matching outliers are generated to identify automatically new fraudulent cases matching the newly identified fraudulent patterns. For example, a suspicious pattern in the behavior of a personalization operator, who has just been hired, can be automatically detected and confirmed thanks to the investigations of a supervisor, and monitored thanks to rules generated by the modeling part of the solution.

- **Supervised learning.**

Preidentified fraudulent patterns are tagged by antifraud experts and analyzed to improve the performance of the detection through experience. Search queries are generated to enable administrators to identify automatically new fraudulent cases matching the preidentified fraudulent patterns.

Why accuracy of Data Mining systems improve with experience ?

New documents to be issued or new transactions are scored using the statistical data consolidated after analysis and investigation of information coming from individuals who are already in the database. The quality of the data mining capability keeps improving as new data is added (see Figure 1 Typical Data Mining Business Process).

Leveraging technology to implement end-to-end fraud deterrence and fraud detection capabilities enables governmental bodies and the stakeholders of document issuance programs to:

- Safeguard against the risk of fraud, through the use of fraud deterrence tools
- Track, score and monitor suspicious activities or individuals, through the use of fraud detection tools. Those tools analyze audit data and personal data that is produced by the documents issuance and utilization systems. They help discover previously unknown patterns and relationships in large data sets.
- Reduce the time to uncover fraudulent activity. The use of Data Mining technology help reduce the workload of analysts and enables them to focus on investigating activities or individuals that have been tagged as suspicious, instead of spending their working hours looking for a needle in a haystack.

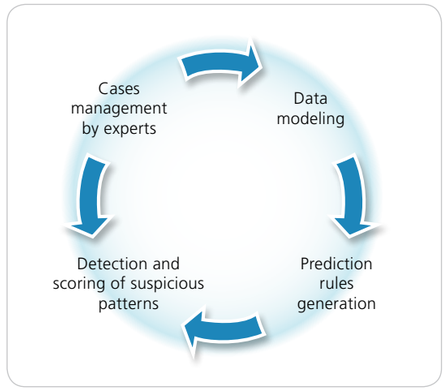


Figure 1 Typical Data Mining Business Process

Final guidelines for a comprehensive fraud-proof process

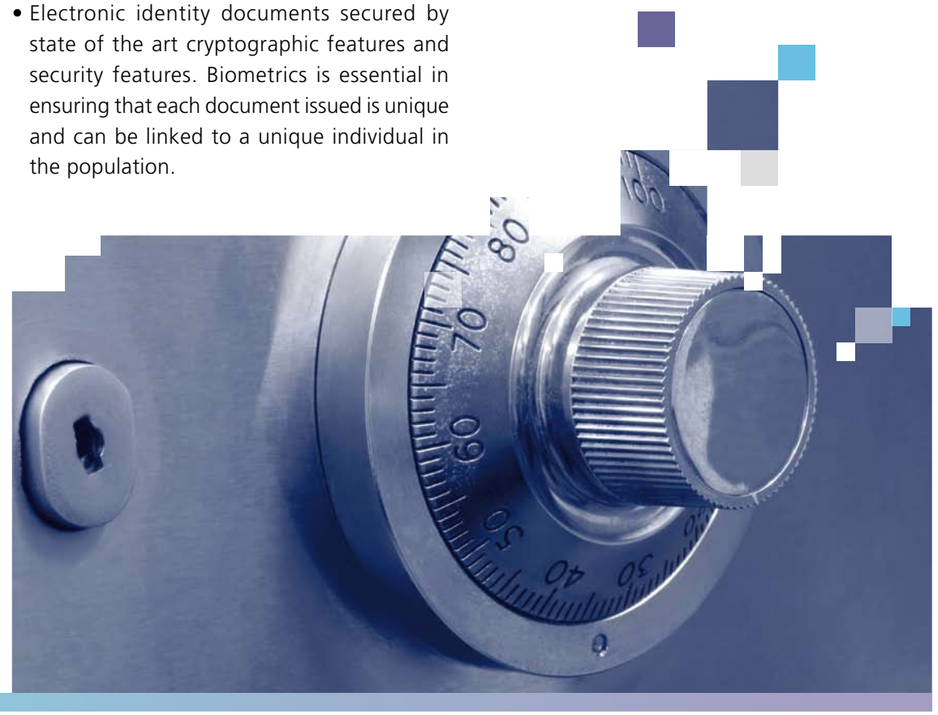
An efficient electronic document personalization process is made of the following building bricks:

- A prime integrator mastering an end-to-end document issuance solution, with a significant experience in securing personalization processes "from scratch": from the design of the solution to the operation of the enrolments and personalization.

- Strict security policies and site governance relying of lessons learnt from major governmental programs.

- Electronic identity documents secured by state of the art cryptographic features and security features. Biometrics is essential in ensuring that each document issued is unique and can be linked to a unique individual in the population.

- Innovative security solutions, enabling to proactively detect new and unsuspected fraudulent patterns in the issuance and utilization of electronic identity documents. Those tools enable to turn seemingly incoherent heaps of data produced by IT systems into powerful means to classify and score the risk of fraud, for any request to issue a document and any transaction originated by the bearer of an electronic identity document.



Printed in France - Polyvalence.com - Photos fotolia - D10184 - 11/10

SAFRAN
Morpho

Phone: + 33 (0)1 58 11 87 08 - Fax: + 33 (0)1 58 11 87 81 - www.morpho.com
Registered Office: Le Ponant de Paris - 27, rue Leblanc - F-75512 PARIS CEDEX 15 - FRANCE
Société anonyme au capital de 159.876.075 €
440 305 282 R.C.S. PARIS