Comparison of identity theft in different countries

Francesco Di Ciccio

1. Introduction

1.1 What is identity theft?

Identity theft is a crime, a type of fraud in which an imposter steals individual information of another person (or a company) and uses it by pretending to be someone else in order to gain some benefit (e.g. money or goods). An identity theft can cause a victim both a financial and emotional damage and also a waste of time and energy needed to avoid consequences of the fraud.

The identity theft can also affect companies, causing not only economic but also reputational damages (e.g. the imposter offers lower quality services compared to the ones offered by the original company). Moreover the companies are supposed to protect beyond themself, also their employees, clients and suppliers, and so they are responsible to damages caused to the third parties in a case of bad handling of their sensible data. Depending on a different country the fine for not proper handling such data could be very high. Therefore, in order to prevent that risk, the most important practices are the training of the employees about the risks and manners in which identity theft can be performed, and the countermeasures to adopt. Also it is advisable to have a good information system provided with adequate security and constantly updated.

Having enough information about an individual, a criminal can take over that individual's identity to conduct a wide range of crimes: for example, false applications for loans and credit cards, fraudulent withdrawals from bank accounts, fraudulent use of telephone calling cards, or obtaining other goods or privileges which the criminal might be denied if he were to use his real name. If the criminal takes steps to ensure that bills for the falsely obtained credit cards, or bank statements showing the unauthorized withdrawals, are sent to an address other than the victim's, the victim may not become aware of what is happening until the criminal has already inflicted substantial damage on the victim's assets, credit, and reputation.

1.2 Types of frauds

The different types of frauds are:

- *Identity Cloning*: substitution of a person with the goal of creating a new identity;
- *Financial Identity Theft*: identity data of an individual or a company stolen to obtain credits, loans, open bank accounts in the name of the victim;
- *Criminal Identity Theft*: an imposter gives another person's name and personal information such as a driver's license, date of birth, ID card, etc. to a law

enforcement officer during an investigation or upon arrest. The imposter may also present to the law enforcement a counterfeit license containing another person's data;

- *Synthetic Identity Theft*: use of different subjects' personal data combined in order to create a new identity;
- *Medical Identity Theft*: use of someone else's data in order to obtain medical services or goods;
- *Ghosting*: creation of a new identity, different from the original one by exploiting the data of a deceased person;
- *Cyber Bullying – Impersonation*: impersonation in a different person, by means of cellular phones or web services 2.0 (social networks, blogs, etc.), with the purpose of sending messages with objectionable contents.

1.3 How identity thieves acquire personal information

1.3.1 Introduction

Personal information can be collected from a variety of sources and by a variety of methods, some relatively simple and low tech, which involve physical theft, and others are more sophisticated, which are performed by means of technology.

1.3.2 Techniques involving physical theft

- *Theft of Wallets, Purses, Cell Phones, Computers, Mail and Other Sources of Personal Information*: everything containing personal information can be stolen, lost or forgotten and discovered by an unscrupulous person;
- *Dumpster Diving*: identity thieves may sort through household garbage, searching for pieces of paper containing financial and other personal information (e.g. receipts, bills, bank statements, etc.). Certain businesses are especially vulnerable to dumpster diving. These include hotels, rental car companies and others that swipe credit cards for reservations and then discard, rather than destroy the copies, once the customer has paid the bill;
- *Change of Address*: thieves redirect mail because it is an abundant source of personal information, and because redirecting mail gives a thief more time to engage in fraudulent transactions before the victim detects any suspicious activity. This can be done in different ways depending on how a mail redirect is regulated in different countries;
- *Tombstone Theft*: the personal information of deceased persons can be accessed from newspaper obituaries and headstones. Obituaries provide birthdates, full names and frequently, critical family information. Careless funeral homes may provide personal information to thieves posing as the deceased's insurance company. An identity thief can use this information to create accounts and take out loans without repaying them;
- *Skimming*: scammers could clone a credit card, by means of an electronic device (skimmer), while using it in a shop or while withdrawing money to an ATM; this device allows to know all the necessary data so that the cloned card can be used instead of the original one. In addition to the skimmer, it could be used a micro

camera in order to record the code typed by the victim. A skimmer could be easily bought on the Internet;

- *Insider Theft*: identity theft often originates from within organizations holding personal information that may include ATM card numbers, PIN codes, credit card numbers and expiry information, passwords, account information, and other personal information of value to thieves. The security of this information is only as good as the integrity of the employees. Identity theft may originate with fraud by a disgruntled or financially strapped employee who sells personal information.

1.3.3 Techniques involving technology-based theft

- *Phishing*: it is an attempt to acquire sensitive information such as usernames, passwords, and credit card details (and sometimes, indirectly, money) by masquerading as a trustworthy entity (e.g. popular social web sites, auction sites, banks, online payment processors or IT administrators) in an electronic communication like email or instant messaging. Phishing emails and messages often directs users to enter details at a fake website whose look and feel are almost identical to the legitimate one;
- *Vishing or voice phishing*: it is an evolution of phishing. The first contact with the victim is by an email in which, this time, he is not asked to click on a link but to call a fake bank's telephone number. Here a fake operator will answer asking for the data to access the victim's bank account;
- *Spamming*: it is born like a type of advertisement, it is often exploited to induce the victim to click on a link or download a file which will install, without his knowledge, malwares on his pc;
- *Keylogging*: it is a malware that records all the keystroke a victim makes. There exists also a hardware version that has to be installed physically on a keyboard;
- *Spoofing*: is the creation of email messages with a forged sender address (easy in the plain Internet e-mail system, since original SMTP doesn't provide any authentication). Spam and phishing emails typically use such spoofing to mislead the recipient about the origin of the message;
- *Pharming*: it is an attack intended to redirect a website's traffic (usually a bank website) to another, fake site. In this way there can be stolen access keys for the victim's bank account. Pharming can be conducted either by changing the hosts file (file that maps hostnames to IP addresses) on a victim's computer or by exploitation of a vulnerability in DNS server software. The latter technique is called DNS poisoning. The result is that when an unsuspecting user enters the website address that has been changed into their browser, they will automatically be brought to the spoofed site. Their browser's address bar will show the correct address, but the site displayed will be a fake one;
- *Sniffing*: it is the bugging of packets that travel into the network in order to find sensible data. It can exploit the weaknesses of some websites (e.g. the use of http protocol instead of https, which introduces the SSL cryptographic protocol) in sending personal information submitted in forms unencrypted;

1.4 How to protect yourself from spam and identity theft

- *Use of an antivirus always updated*: it is important to keep an antivirus updated promptly and regularly. New malware can be found every day and they and it can spread extremely fast. Moreover, it is important to keep also the operating system updated in order to prevent eventual vulnerabilities that can expose the pc to viruses attacks;
- *Do not purchase goods, which are suggested by the non-requested emails*: there is a risk to have the email address inserted in mailing lists that are sold to spammers, with the chance to receive even more spam emails that could lead to a fraud;
- *Use of a firewall client on computers*: it protects computers that are connected to the Internet;
- *Do not answer the spam emails and ignore their links*: it is important not to answer spam emails or click links contained therein, not even to cancel the subscription for those emails: it will confirm the email address validity to the spammers;
- *Use of a secondary email addresses*: it is advisable to use a primary email addresses only with known persons (e.g. friends, colleagues, etc.) and a secondary email addresses for filling web forms. It is not advisable to publish the primary email addresses on forums, blogs, social networks and other public websites because they will be easily intercepted by spammers;
- *Do not answer to messages that ask personal and financial information*: it is strongly recommended not to give those information on messages and emails; banks and e-commerce companies usually never send those kind of messages;
- *Do not click on pop-ups*: it is dangerous to click on pop-ups like the ones that alert on the presence of a virus on your computer and propose solutions, because these links could lead to downloading malware;
- *Do not save passwords on the computer or other devices*: it is possible for hackers to access computer or other online devices and find the file containing passwords;
- *Do not trash papers containing personal information without destroying them*: it is recommended to make illegible papers containing any personal or financial information before trashing them in order to avoid *dumpster diving*;
- *Use of some good practices while using a credit card*: it is advised to not give away the card while paying, for example in shops, and always cover the hand which is entering the PIN code while paying or withdrawing money, in order to avoid *skimming*;

1.5 How thieves can use stolen personal information

Examples of unlawful use of personal information of victims are:

- *Selling personal information*: insiders or crackers who have stolen a good amount of personal information usually commit this form of fraud. Here the economic gain does not come from exploiting the information directly, but rather by selling it on the black market to others who will ultimately use it to commit fraud;
- *Forging identity documents*: thieves often use personal information to create fake credit and debit cards, driver’s licences, vehicle registration certificates and other identity documents. These forgeries will then be used to commit a fraud;

- *Taking over existing accounts*: once a thief has enough personal information of a victim, he can contact organizations with which the victim has existing accounts, pretend to be the victim, and take control of the accounts by changing the mailing address or the credentials used to access the account. A thief could take over a bank account and empty it out over a short period of time to avoid raising any suspicions.
- *Opening new accounts*: with a minimum amount of personal information (depending on the different countries), such as name, address and SIN (Social Insurance Number - Canada) or SSN (Social Security Number - USA), an identity thief can open all sorts of accounts, such as bank accounts, credit accounts (either credit cards, credit lines or loans), in-store accounts and cell phone accounts. Usually, once the account is opened, the thief will change the billing or correspondence address in order to conceal her activity from the victim, which usually will not realize that something is wrong until a credit application is refused or a debt collector contacts her.
- *Ordering goods online using a drop-site*: a thief may shop online exploiting stolen personal information, usually with a computer based in a different jurisdiction from that of the victim or using a different country: after ordering some items in the victim's name and using the victim's credit card number, the thief will ask the merchant to deliver them at a "drop-site", where a trusted third party or associate will receive them, repackage, and ship them to the thief. If the authorities check out the drop site, the associate has a defence of plausible deniability. Prosecuting the foreign conspirator(s) is almost impossible because of jurisdictional and resource issues [1];
- *Obtaining a passport*: transnational criminal and terrorist organizations may misuse fraudulently obtained travel documents to support their illegal activities;
- *Obtaining government benefits*: thieves may be able to obtain various government benefits such as Employment Insurance, welfare and Old Age Pension benefits, by masquerading as another person, using stolen personal information;
- *Medical identity theft*: a thief may obtain medical services in the name of a victim, causing addition of erroneous data to her medical records. Among all the countries this fraud is very spread in the USA, so it will be better analysed in the next chapter;
- *Hijacking email accounts*: imposters can take over the victim's email address, domain name, chat account or other computer based identifiers, and sending messages to others in the name of the victim. Usually this type of identity theft is related more to defamation than to fraud for economic gain, although it can be used for the latter. Another common purpose for hijacking Internet accounts is to send spam;
- *Concealing one's true identity*: the offender assumes another name to cover up past crimes and avoid capture, sometimes over many years. She can also use another name and identification to avoid arrest (easier in countries where there is no mandatory ID document);
- *Mortgage fraud*: it generally occurs when a thief provides fraudulent information, such as false employment records, to a lender in order to obtain a mortgage. Title fraud, another variant of real estate fraud, involves an individual falsely assuming the identity of another property owner. The criminal then uses the identity of the true owner to assume the title or sell or obtain other mortgages based on that property;

- *Taking over insurance policies*: the identity thief may make a change of address on the car insurance policy of a person after stealing her personal information. She will then make false claims for "pain and suffering" suffered from auto accidents.

2. Identity theft in USA

2.1 Introduction

The term identity theft stem from the United States with which it always had an intrinsic connection, despite the spread of the phenomenon to other countries like the European Union. Unlike many (if not most) other countries, the United States actually has a legal definition of identity theft: "knowingly transfers or uses, without lawful authority, a means of identification of another person with the intent to commit, or to aid or abet, any unlawful activity that constitutes a violation of Federal law, or that constitutes a felony under any applicable State or local law." This definition comes from the Identity Theft and Assumption Deterrence Act, which Congress passed in 1998.

The United States has conducted perhaps more studies than any other country in the world, which were mainly based on consumer and victim surveys in an attempt to develop an accurate picture on the prevalence of identity theft inside of its geographical boundaries. The primary source of data is actually not a study but a complaint database called CSN (Consumer Sentinel Network) [2], which was introduced in 1999 after the Congress passed the Identity Theft and Assumption Deterrence Act. The reader can find some statistics about those complains in Figure 1, Figure 2 and Figure 3 regarding the Calendar Years from 2001 to 2013.

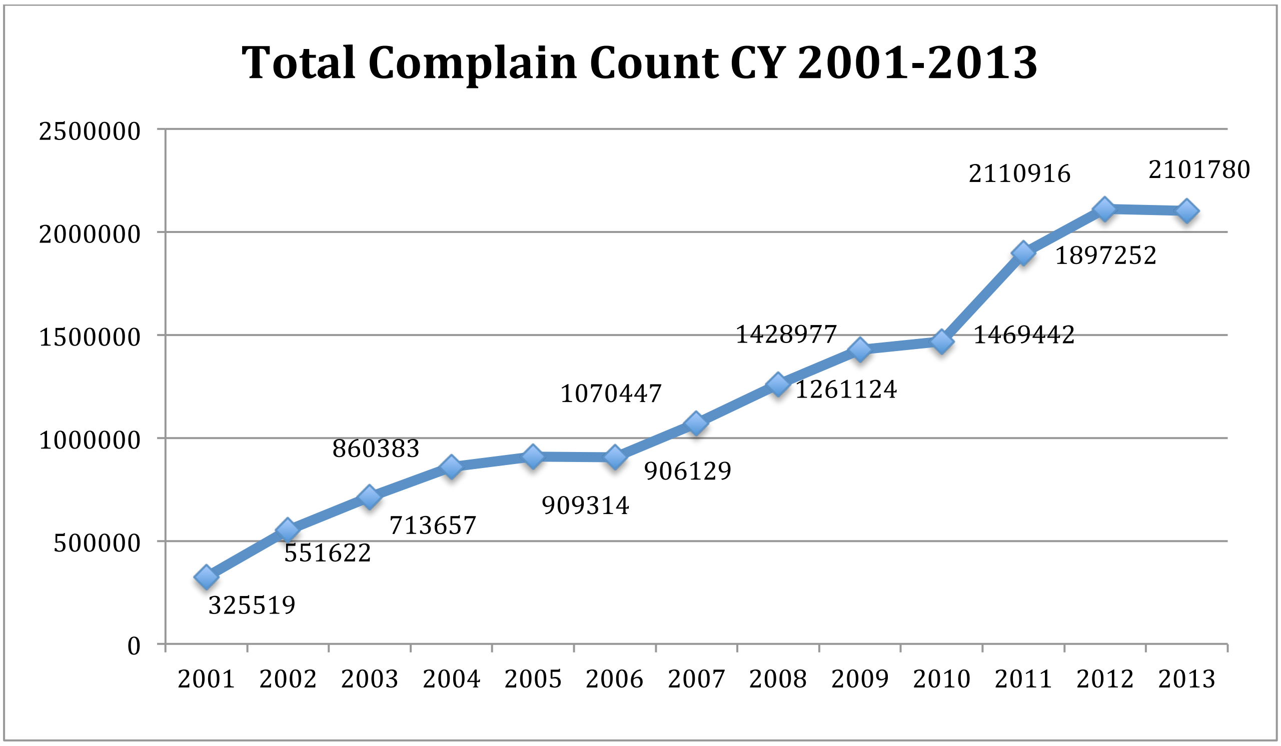


Figure 1. Consumer Sentinel Network complaint count for Calendar Years 2001-2013

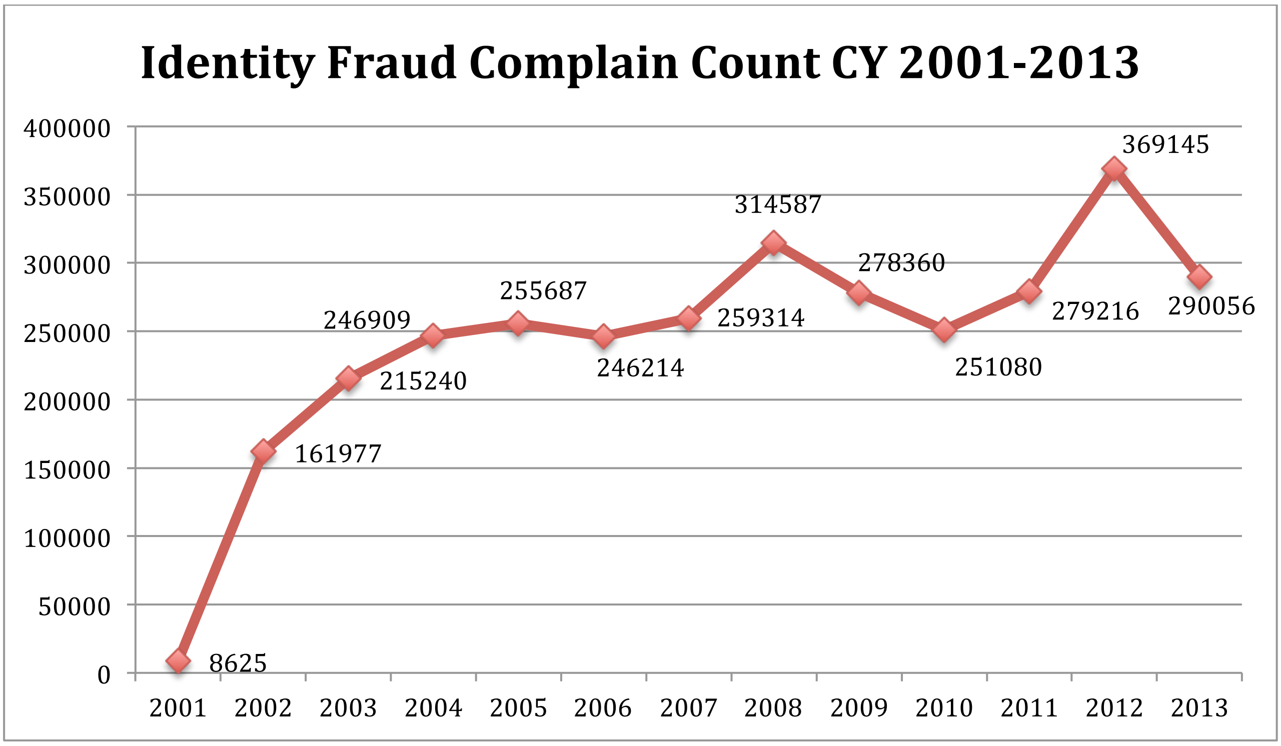


Figure 2. Consumer Sentinel Network complaint identity theft count for Calendar Years 2001-2013

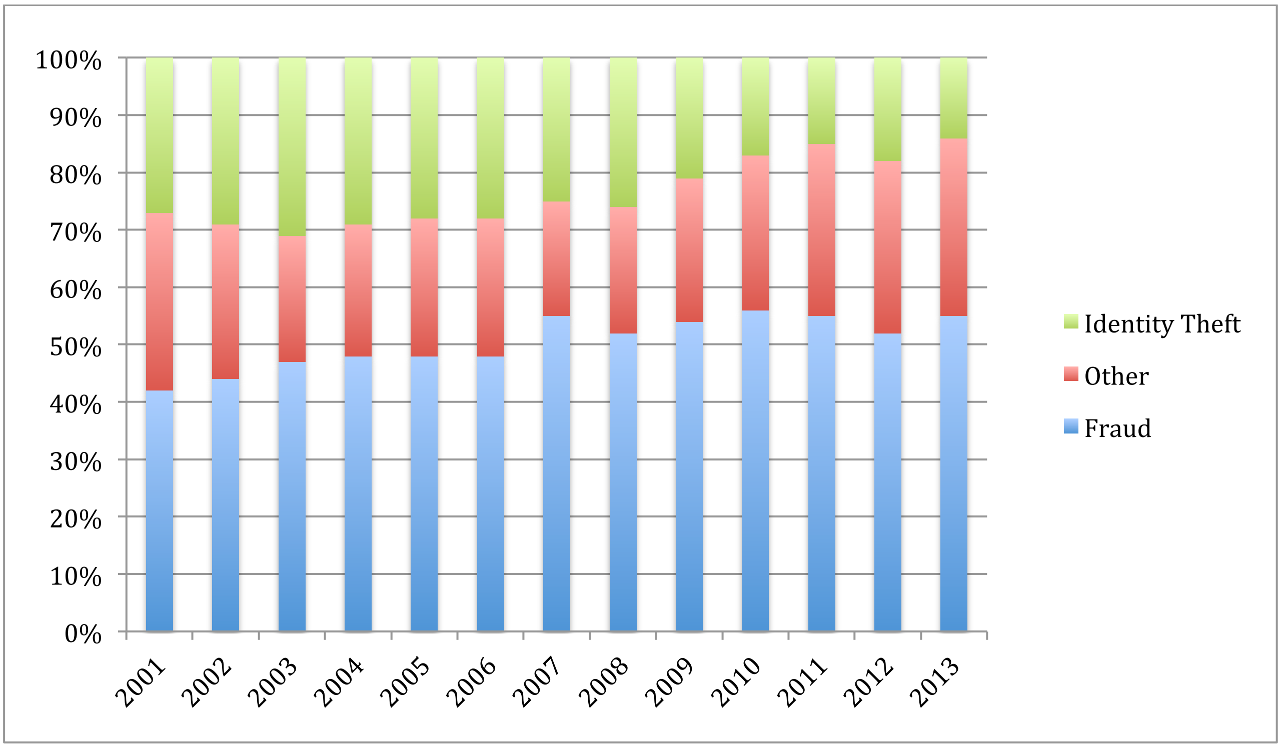


Figure 3. Consumer Sentinel Network complaint type percentages for Calendar Years 2001-2013

2.2 Vulnerabilities

2.2.1 Identity documents

Probably, the most important weakness of the US system regarding identity theft is that there is no mandatory national identity card for all American residents. All legislative attempts to create one have failed due to tenacious opposition from liberal and conservative politicians alike, who regard the national identity card as the mark of a totalitarian society.
At present, the only national photo identity documents are the passport and passport card, which are issued to U.S. nationals only upon voluntary application. Most people use state-issued driver's licenses as identity cards, but even those documents are not mandatory to have, if you are not driving.

We will see that in addition to the vulnerabilities within public infrastructure, there is also the private sector that plays an important role in a high frequency occurrence of the identity theft in the USA.

2.2.2 Social Security Number

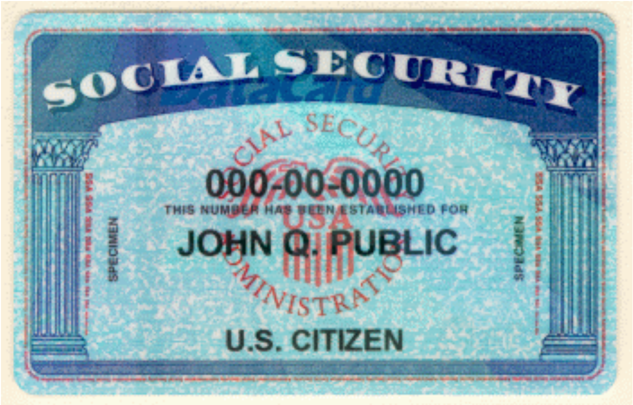


Figure 4 Example of Social Security Number card

The main vulnerability within the infrastructure is the use and availability of the Social Security Number (SSN) that is used as the main identifier for individuals, both citizens and permanent residents, within the United States. The main problem currently is the high level of usage and availability of this number.

When individuals present the number at either a public or a private sector institution, the institution accepts this number as a means of identification; financial institutions generally require a SSN to set up bank accounts, credit cards, and loans, partly because they assume that no one, except the person it was issued to, knows it. But this is clearly a weakness because the number is not really publically available, but can easily become available by mistake. An identity thief that has someone SSN needs very little additional information to effectively steal his identity.

Many citizens and privacy advocates are concerned about the disclosure and processing of Social Security numbers. Furthermore, researchers at Carnegie Mellon University have demonstrated an algorithm that uses publicly available personal information to reconstruct a given SSN (this algorithm is stressed in the next section) [3].

Exacerbating the problem of using the social security number as an identifier is the fact that the social security card contains no biometric identifiers of any sort, making it essentially impossible to tell whether a person using a certain SSN truly belongs to someone without relying on other documentation (which may itself have been falsely procured through use of the fraudulent SSN). Congress has proposed federal laws that restricts the use of SSNs for identification and bans their use for a number of commercial purposes, e.g., rental applications.

The Social Security number is a nine-digit number in the format "AAA-GG-SSSS". The number is divided into three parts. The *area number*, the first three digits, is assigned by geographical region, and it is based on the ZIP code in the mailing address provided on the application for the original Social Security card. The applicant's mailing address does not have to be the same as their place of residence. Thus, the area number does not necessarily represent the State of residence of the applicant. The next two digits, known as the *group number,* were assigned in a non-consecutive yet predictable order within each distinct area number. The final four digits were determined serially and issued in order of application. Thus an identity theft could predict someone’s SSN by knowing her date and location of birth. The Social Security Administration (SSA) regularly publishes the highest group number that had been issued for a given area code, so the set of total possible numbers was effectively divided into issued and unissued ranges. This actually facilitates the work of fraudsters trying to represent someone else’s SSN as their own, and also in the case of child identity theft (which we will discuss in the sequel).

On June 25, 2011, the SSA changed the SSN assignment process to "SSN randomization": the first three digits don’t represent anymore the area code (and are reintroduced numbers before unused like 000, 666, 900-999) and also the last numbers are no more depending on the issuing data. All numbers are now randomized to avoid the predictability of an individual’s SSN.

Unfortunately it had the unintended consequence of reducing the information available to risk managers, so that now it is nearly impossible for risk managers to distinguish between legitimately issued numbers and those that are being illegitimately asserted. Prior to randomization, a SSN in the unissued range sent a strong and explicit signal to risk managers.

2.2.3 Predict the SSN by using the public data

In [3] the authors show how from publicly available data they can gather enough information to build an algorithm that try to guess someone’s SSN.

They started by hypothesizing that the Enumeration At Birth initiative (EAB, an anti-fraud program which started extending nationwide in 1989 and for which the overwhelming majority of US new-borns obtain their SSNs shortly after birth), increased the likelihood that, for US-born applicants, the state of SSN application would be their state of birth, and the date of application would correlate with their birthday. Furthermore the SSA published the Death Master File, which is a publicly available file reporting SSNs, names, dates of birth and death, and states of SSN application for individuals whose deaths were reported to the SSA. Ironically, one of its applications is fraud prevention, because the DMF can be used to expose impostors who assume deceased individual SSNs.

With the increasing automation of SSN assignment systems, the DMF could be exploited to verify if there are some regularities in the SSN assignment scheme, and then predict individual SSNs based on the dates and states of birth of their applicants. Specifically, the authors used the DMF as an analysis set to identify assignment patterns and as a test set to test the accuracy of SSN predictions based on extrapolated patterns.

After grouping and sorting DMF data by state of assignment and date of birth, the analysis on those data confirmed the regularities expected: as hypothesized, a strong correlation exists between dates of birth and all 9 SSN digits.

In the algorithm authors predicted the first five digits of the each DMF record, namely the Area Number (AN, first three digits) and Group Number (GN, next two digits), based on the most frequent ANs and GNs assigned to the DMF records with the same state of application and born around the target record birthday; they predicted its last four digits, namely the Serial Number (SN), combining its birthday with coefficients estimated from linear regressions over individuals' SNs with similar birthdays and same state of application as the target.

The regression model is sketched in the following equation:

$$SN_i = \alpha + \beta_1 \cdot dd_{i,vw} + \beta_2 \cdot ANGN_{i,vw} + \epsilon_{i,vw}$$

Where SN_i is the SN assigned to individual i, born on day *dd* and whose record can be found within the window of days vw in a specific year and state; $ANGN_{i,vw}$ is a vector of dummies for the various ANGNs that can be found associated with the SSN records contained in the DMF within that variable window (the ANGN dummies account for the cyclical pattern of SN issuance); and ϵ is the regression error. The target individual's date of birth and its predicted ANGN are combined with the β_1 regression coefficient for the day $dd_{i,vw}$ and the β_2 dummy coefficients for the predicted $ANGN_{i,vw}$ from the regression conducted over the DMF records included within a window of days around the target's date of birth.

As hypothesized, they found widespread increases in prediction accuracies after 1989 (the onset of the nationwide EAB program), particularly for less populous states (fewer daily births determine more discernable patterns).

For instance, the authors of [3] accurately predicted the first five digits of only 2% of California records with 1980 birthdays versus 90% of Vermont records with 1995 birthdays; on average, they matched the first five digits for 44% of all records born nationwide after 1988. Imagining a brute force algorithm where, for each target, the attacker tries out the predicted SSN before moving up and down the SNs in 1-integer steps for the following attempts (while keeping the predicted ANs/GNs constant), fewer than 1,000 attempts may be sufficient to identify the SSNs of 8.5% of all individuals born after 1988.

When fewer than 1,000 attempts are sufficient to predict massive amounts of SSNs, various brute force attacks become economically feasible: an attacker could try to identify an SSN by testing subsets of variations predicted by the algorithm across different channels, including phishing emails, online instant credit approval services, or the SSA's own SSN Verification Service and the Department of Homeland Security's E-Verify system. Unlike traditional identity theft strategies, these channels allow attackers to test, covertly and cheaply, multiple variations of predicted SSNs for massive numbers of targets, while choosing them based on demographic traits, bridging the gap between statistical predictions and actual identity theft.

2.2.4 Verification in the private sector

In addition to weaknesses or vulnerabilities, which are the results of government introduced initiatives, the private sector also indirectly facilitates the occurrence of identity theft. With regard to financial service providers, for example, significant problems have occurred. Verification is a big issue, especially when it comes to new applications for credit cards or loans. The problem is that in today's very competitive financial marketplace, speed is essential, and thieves love fast credit approval, because speed is the enemy of accuracy. Credit card issuers, for their part, can be very superficial in releasing cards, failing to match SSNs and dates of birth and otherwise failing to take basic precautions in their enthusiasm to get cards in circulation.

A famous story that proves the rather inaccurate verification mechanisms of credit card companies is the story of Steve Borba and his dog Clifford. Borba opened up an email account using his dog's name, and, as time passed, he received a pre-approved credit card application in his email inbox. For Clifford's social security number, Borba used 9 zeros and he explicitly wrote on the application that Clifford was indeed a dog. Despite this comment and the seemingly impossible social security number, Clifford received his credit card three weeks later [4].

A very similar story is Gary More's one, who, annoyed by the big amount of unwanted credit card applications he was receiving, tried to call the issuers in order to stop all the junk mail. Since that approach failed to achieve the desired results, he took one of the many unwanted solicitations in hand, wrote the words "Never waste a tree" across it (as his way of telling the sender to stop wasting paper), and mailed it in. As a result Mr. More received a credit card issued to "Never Waste Tree" [5].

Financial service providers carry a big responsibility with regard to the prevention of identity theft, because they form the most crucial link between the attackers, the financial benefits, and the victims. This inadequate verification mechanisms and aggressive marketing methods certainly help perpetrators along. If financial institutions start taking better precautions, they could certainly limit some of the identity fraud that occurs in the opening of bank accounts and the extension of credit. The problem is that in a competitive market, these institutions fear that a more rigorous screening process might scare consumers away to competitors who do not take such measures.

2.3 Child identity theft

Child identity theft occurs when an imposter uses a minor's SSN for his personal gain. The Social Security numbers of children are valued because they do not have any information associated with them, this means that there is a clean history associated to their SSN so that a thief can attach every name and date of birth to their records. Thieves can establish lines of credit, obtain driver's licenses, or even buy a house using a child's identity, and this fraud can go undetected for years, as usually parents don't check their children's credit, so they don't discover the problem until years later. This is possible because when someone applies for, say, a loan, usually the banks check if that SSN has a good credit history. They don't check if the name is associated to that SSN because the government charges a fee to do that and most banks don't want to pay.

Child identity theft is fairly common, and studies have shown that the problem is growing. The largest study on child identity theft, as reported by Richard Power of the Carnegie Mellon Cylab with data supplied by AllClear ID, found that of 40,000 children 10.2% were victims of identity theft. This child identity theft report is not based on survey results, but on identity protection scans on 42,232 children (age 18 and under) in the U.S. during 2009-2010. This pool of 42,232 child identities includes everyone under 18 in a database of over 800,000 identity records [6].

Prior to the SSN randomization it was quite easy for a thief to steal the SSN of a child, since the last digits were related to its issuing date (a clue to the person's birthday), but now that the all the digits are randomized is almost impossible to guess if a SSN belongs to a child or not.

The potential impact on the child's future is profound; it could destroy or damage a child's ability to win approval on student loans, acquire a mobile phone, obtain a job or secure a place to live.

2.4 Medical identity theft

Medical identity theft is the fraudulent acquisition of someone's personal information such as the name, SSN and health insurance number, for the purpose of illegally obtaining medical services or devices, insurance reimbursements or prescription drugs.

Medical identity theft is a crime that can cause great harm to its victims; for example it can results in erroneous entries being put into existing medical records, and can involve the creation of fictitious medical records in the victim's name. Yet despite the profound risk it carries, it is the least studied and most poorly documented of the cluster of identity theft crimes. It is also the most difficult to fix after the fact, because victims have limited rights and recourses.

A thief may use a victim's name or health insurance numbers to see a doctor, get prescription drugs, file claims with your insurance provider, or get other care. If the thief's health information is mixed with yours, your treatment, insurance and payment records, and credit report may be affected, and this can leave a trail of falsified information in medical records that can plague victim's medical and financial lives for years.

As the health care system transitions from paper-based to electronic, this crime may become easier to commit and victims may find it more difficult to recover from medical identity theft, as medical errors are disseminated through computer networks and other medical information-sharing pathways.

At the beginning of this year, the Identity Theft Resource Centre produced a survey showing that medical-related identity theft accounted for 43% of all identity thefts reported in the United States in 2013 [7].

This high rate of medical identity theft in the USA probably comes from the fact that the health insurances there could have high prices, and so could be convenient to receive medical care or prescription in the name of someone else. This, connected to the fact that in the USA there is no mandatory ID document with picture, makes it easier for a thief to obtain medical care instead of someone else. After stealing the insurance number and maybe the SSN with the methods described before in this document, a thief will need just some other basic information like the full name of her victim.

2.5 Comparison with Canada

2.5.1 Identity documents

Like in the USA, in Canada there is no mandatory identity card for residents. The most commonly used forms of identification are the driver's licence and the SIN card, which has become a national identification number in much the same way that the SSN has in the US.

The widespread usage of these two documents for identification purposes has made them de facto identity cards. However, unlike in the US, in Canada there are specific legislated purposes for which a SIN can be requested. An organization can request a person's SIN if law specifically permits it, or if no alternative identifiers would suffice to complete the transaction, otherwise they cannot deny or refuse a product or service on the grounds of a refusal to provide a SIN. Examples of organizations that legitimately require a SIN include employers, banks and investment companies, and federal government agencies. Giving a SIN when applying for consumer credit, such as buying a car or electronics, or allowing it to be used as a general-purpose identification number, such as by a cable company, is strongly discouraged.

2.5.2 Child identity theft

In Canada it seems that this kind of identity fraud is still not widely spread. However, since it is relatively common in the US, it is likely to spillover north of the border and so the police is keeping an eye out for it.

In Canada many parents apply on behalf of their child for a SIN in order to open a Registered Education Savings Plans (RESP), but the SIN itself leaves a child vulnerable to identity theft because the number isn't associated with a birth date. Scammers can easily use the SIN to create a new identity and apply for a credit.

Child identity theft is particularly heinous because it can take years to recover from the crime. It is also difficult to know if your child has been victimized, because children don't have a credit history. It takes sometimes 10 to 15 years to spot it, because kids shouldn't have a credit report, they shouldn't have any type of debt until they're 18.

2.5.3 Other kinds of identity fraud

For Canada, the comparable statistics with the US concerning the medical identity theft are not available. However, the Ontario Ministry of Health and Long-Term Care website makes mention of identity theft as a source of health care fraud in the province.

The proliferation of unsolicited pre-approved credit card applications with personal information already typed onto them, has made discarded mail an especially good target in the United States. Thieves complete these applications, substituting a new address, thus the credit cards obtained can then be used to collect charges in the name of the victim. In Canada, unsolicited credit card applications do not contain sufficient personal information for a thief to obtain a card in someone else's name.

The biggest debit-card fraud in Canadian history took place in August 2003: five Russian citizens were arrested for it. Their scam involved the purchase and subsequent modification of five ATM machines, which were modified to capture all the necessary information to reproduce the card. They also captured the PIN numbers entered. About 4,000 people fell victim to the scam and all of them were reimbursed by the banks.

Selling personal information to another individual in Canada, except under very limited circumstances, is not illegal. Currently, the Criminal Code only makes illegal to sell or transfer credit card data or computer passwords, while selling driver's licence numbers and social insurance numbers is not prohibited.

2.5.4 Some statistics

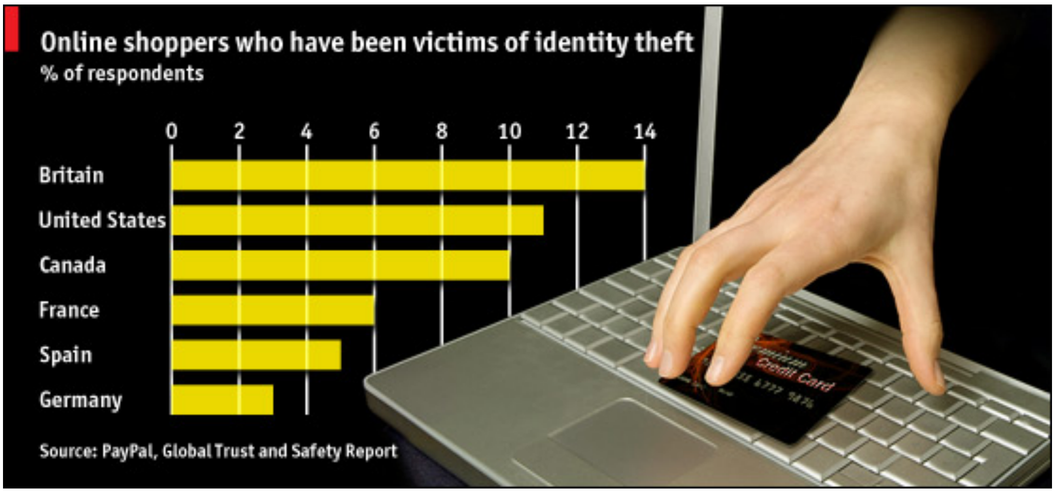
From the Table 1, made with data from Police-reported crime statistics in Canada from 2011 to 2013 [8], we can see how the identity theft fraud is less relevant with respect to the US. From the 2010 report going back, the identity theft data are not mentioned.

	TOTAL CRIMES	IDENTITY THEFT	PERCENTAGE	RATE per 100000 POPULATION
2011	1984790	12013	0,6%	35
2012	1957227	10807	0,55%	31
2013	1824837	11594	0,63%	33

Table 1 Police-reported crime for identity theft 2011-2013

3. Identity Theft in the EU

3.1 Introduction



Source: PayPal, Global Trust and Safety Report

Figure 5 Online shoppers who have been victims of identity theft from the Global Trust and Safety Report by Paypal

If one uses a narrow definition of identity theft/fraud (e.g. fraudulent applications and account takeover cases only), the problem in Europe is still important, but its scope is limited compared to the other fraud typologies. It is interesting to note that according to the intervention of a major payment card scheme at the high level conference of November 2006 organised by the European Commission, account takeover and application fraud cases in the UK accounted for around three quarters of all cases of these types of fraud in Europe (including non EU Member States). This seems to suggest at this stage that the identity theft/fraud problem (using a narrow definition) in Europe in connection to payment fraud is touching more severely the UK than any other country in Europe.

However, it would be a wrong conclusion to believe that the problem does not concern other EU countries.

3.2 United Kingdom

3.2.1 General description of the problem

In the EU, the country that is closer to the US with respect to identity theft is the UK, which has an extensive history when it comes to identity-related crime. In February 2004, the European Commission held a Forum on identity theft where the attendees noted how "Identity theft is growing fast outside the EU (e.g. US and Canada) and is very relevant in the UK. For now, it does not seem to be equally prominent in the other Member States."

One of the reasons why the UK is closer to the US is probably due to limited use of the ID documents: identity cards for British nationals were introduced in 2009 on a voluntary basis, so driver's licenses and passports are the most widely used ID documents in the UK as well. Only workers in certain high-security professions, such as airport workers, were required to have an identity card, and this general lack of ID being compulsory tends to remain the case today.

In UK there are two different definitions for identity fraud and identity theft: the latter is considered as a preliminary action conducted before committing an identity fraud, which certainly changes the overall debate but also the legal countermeasures introduced.

The predominant part within the UK is occupied by the identity frauds committed for obtaining financial benefits. The focus is mainly on the private sector but to a certain extent also on the public sector, with regard to potential for government benefit fraud.

Data on the prevalence of identity fraud in the UK continues to grow and is significantly more comprehensible than in other European countries. The Credit Industry Fraud Avoidance System (CIFAS), for example, has records of consumer complaints about identity fraud dating back to 1999 [9]. The table below demonstrates how from 1999 until 2006 identity fraud complaints continued to rise consistently, in 2007 there was a small decline extended also in 2008 but, from 2009, it started to increase significantly again until last year, where there was a substantial decrease.

Year	Cases Recorded
1999	9,000
2000	16,000
2001	24,000
2002	34,000
2003	46,000
2004	56,000
2005	66,000
2006	80,000
2007	77,500
2008	77,600
2009	102,300
2010	102,650
2011	113,250
2012	123,600
2013	108,500

Table 2 Number of victims of identity froud per year according to CIFAS.

3.2.2 Vulnerabilities

Within the UK, various vulnerabilities exist which facilitate the occurrence of identity fraud, and regarding identity theft, the public availability of information is a significant issue. Kevin McNulty, head of the Identity Fraud Reduction Team, acknowledged how there is a "huge availability of public data" including records of births, marriages and deaths.

Furthermore, several websites exist (like social networks, blogs, etc.) which provide perpetrators with extensive personal information on potential victims. Staff members, especially within call centres, also have access to significant amounts of sensitive personal data including credit card numbers and CCV numbers. These employees are a major vulnerability because they can either access the information voluntarily to commit dishonest acts or organized crime networks can bribe them into handing over the customer information.

Nicola Westmore, from the UK Ministry of Justice, described an important case that emphasized this vulnerability in data availability, which concerns online publication of land registry. New legislation allowed them to put details of people's titles and properties on the Internet, creating a massive database of sensitive personal information, including photocopies of mortgage deeds and signatures. Thankfully enforcement came to take it off, due to the fact that individuals felt as though there was the potential for identity theft, since anyone had access to all that information.

In addition, some high-profile data losses within the UK have introduced an entirely different dimension to the availability of sensitive personal data. A prime example is the data loss which occurred in November 2007, when news broke about how the HM Revenue & Customs lost disks which contained records of 25 million child benefit recipients. These disks were lost in the mail, when they were handed to TNT by courier mail and were supposed to arrive at the National Audit Office. A junior employee at the HMRC office in Washington, Tyne & Wear, apparently put this material in the post. The disks contained rather sensitive information, which were password protected but not encrypted: names, addresses, dates of birth, child benefit numbers, National Insurance numbers and bank or building society account details.

While this data loss receives tremendous attention from both the government and the media, it was not the first data loss at HMRC. In October 2007, an HMRC member left a laptop in his car, which was subsequently stolen. The computer contained records from finance houses revealing the identity of high value customers who had invested in Individual Savings Accounts. Ever since the media caught on to the HMRC data loss, many other organizations have come forward about data losses. The data loss also led to, perhaps expected and inevitable, new phishing attacks.

Problems within the private sector do also exist. The Information Commissioner's Office describes how a freelance journalist, based in Southampton, decided to check local banks in his area. He went along to several banks and a post office, looked in the bins placed outside them and found a significant number of discarded personal data (cut up debit/credit cards, torn up bank statements/insurance application forms, etc.). He contacted the banks, did not receive a very favourable response and as result contacted the ICO, which commenced an investigation but, before that was completed, the journalist contacted the BBC Watchdog program. Watchdog visited several towns in the UK and their 'researchers' found similar discarded personal data in bins outside banks/building societies.

About a month later, BBC Watchdog researchers repeated the operation in other towns and again got similar results. In addition, a journalist in Scotland carried out a similar operation in his local town and recovered personal data. All the documentation recovered was forwarded to the ICO and it resulted in the undertakings being obtained from the Post Office, eleven Banks and the Immigration Advisory Service. The ICO hopes that these undertakings signed by chief level executives will assist in the prevention of sensitive data.

Another vulnerability consists in how individuals in the UK can rather easily change their name. A name change does not require a legal deed, which means perpetrators who want to commit an identity fraud can simply change their name to correspond with whatever documentation they have managed to obtain.

While particular vulnerabilities exist with regard to the identity theft stage, societal factors also facilitate the actual occurrence of identity fraud. Within one research experiment, Martin Gill, Professor of Criminology at Leicester University, along with a colleague tried to assess to what extent they could accomplish certain activities with a voter registration card. Within the UK, any eligible citizen has a voter registration card, which is not a form of identification. Nevertheless Gill and colleague tried to see whether people (i.e. employees at a financial institution and the post office) would be willing to accept it as a form of identification. As they went to withdraw money, the bank employee did not even bother asking for identification and neither did the post office employee when they went to pick up a parcel. This clearly hints at potential problems with regard to verification of a client's identity, and this is certainly a vulnerable area in the UK.

3.2.3 Countermeasures

The Home Office created the Identity Fraud Steering Committee (IFSC) and the Identity Fraud Forum (IFF) in 2003, which developed a framework to identify effective measures to prevent and react to subsequent occurrences of identity fraud.

As a result of these priorities, the IFSC and the IFF managed to develop a number of regulatory instruments in their battle against identity fraud. Important measures introduced by the Government primarily include aligning penalties, defining a new criminal offense, developing and sharing good practice, and raising public awareness. The Government decided to increase the maximum punishment for fraudulently obtaining a driver's license from a maximum fine of £2,500 to a maximum two years prison sentence.

Furthermore, in 2003 the Government decided to introduce a new criminal offense, for which any individual who is either in possession or in control of false identity documents, whether genuine documents illegally obtained or derived from another person, is in violation of the law and subject to criminal sanctions. The underlying motive for the introduction of a new criminal offense is the connection between use of false identity documents and organized crime.

On 15 January 2007, the Fraud Act of 2006 introduced a new offence of fraud that can be committed in three ways: by making a false representation (causing loss or risk of loss to someone with intent to make a gain), by failing to disclose information, and by abuse of position. Several new offences were defined in the act: obtaining services dishonestly, possessing equipment to commit frauds, and making or supplying articles for use in frauds.

This act facilitates the prosecution of identity theft, so probably this is the reason why in 2007, there was a decrease in the amount of identity fraud victims since 1999, as we can see from Table 2.

According to a report published by the CIFAS in October 2009 in order to respond to the significant increase of identity theft complains, around 81% of the British public are concerned about becoming a victim of identity theft. Nevertheless most consumers and businesses are not taking steps to protect themselves. The 22% access their bank details at work or in Internet cafes, while the 79% of businesses make no effort to destroy sensitive material that is thrown away or recycled [10].

The explanation of the decrease in 2013 of identity fraud comes from the 11% decrease in fraud levels recorded in the same year. CIFAS Communications Manager, Richard Hurley, notes: "The decrease in fraud comes on the back of The Audit Commission's Protecting the Public Purse 2013 which also reported a reduction in the number of frauds, and is good news for those who participate in the collective effort to prevent fraud".

While identity crimes such as identity fraud or the hijacking of an existing account fell during 2013, they still represented over 60% of all frauds recorded during the same year.

Richard Hurley states: "This is the third year that identity crimes have accounted for such a huge chunk of fraud in the UK. Sadly, it also confirms that still not enough is being done by individuals and organisations. Consumers have the right to demand that organisations handle their data securely, and increase their anti-fraud efforts and stop fraud before someone financially loses out. But collectively, every one of us has to be expected to take responsibility to do all that we can to keep ourselves safe. Not disabling firewalls, installing anti-virus software, using good online practice and strong passwords: these have long been messages that CIFAS and others have delivered".

3.3 France

3.3.1 Document fraud

Identity-related crime concerns in France mainly focus on document and financial fraud.

First of all, problems arising from the production of false documents mainly concern passports and driving licenses since they do not benefit from a centralised procedure of issuance, contrary to the national identification card (which is compulsory for everyone over the age of 16).

There is a relative freedom in the proof of one's identity: the law (Art. 78-1 to 78-6 of the French Code of criminal procedure, namely Code de procédure pénale) mentions only that during an ID check performed by police, gendarmerie or customs, one can prove his identity "by any means", the validity of which is left to the judgment of the law enforcement official. Though not stated explicitly in the law, an ID card, a driving licence, a passport, a visa, a Carte de Séjour, a voting card, are sufficient according to jurisprudence.

This system provides great flexibility. However, as a consequence of some failures in the hand-out and issuance procedure, a market of false identity documents has developed where the price of a false driver's license was estimated some years ago to be around 500 Euros and a false passport around 2,000 euros.

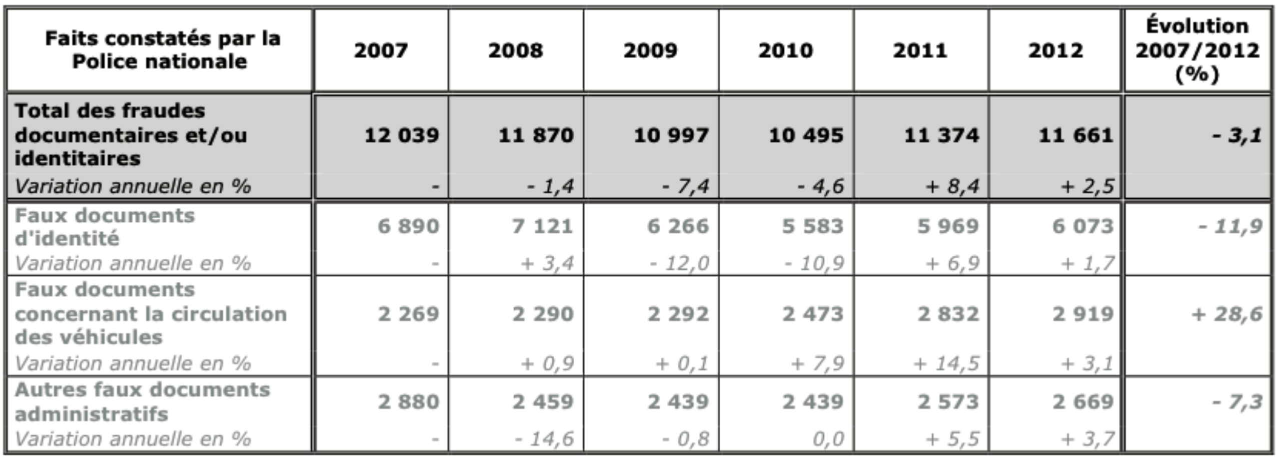
It is worth noting that there is no obligation to formally change the address on official documents such as the identification card or passport, it depends on the subject's will.

Regarding this problem of falsification, a suggestion could be giving legal validity for identification exclusively to documents such as the national identification card and passports.

Another problem resides in the fact that it is possible to obtain official identification documents by providing forged, therefore unofficial, documentation. This facilitates the obtaining of a false identity certified by an authentic document.

In view of obtaining an identification card or passport, individuals need to prove that they are entitled to be delivered such document. To that effect, they must produce an authentic document that states their date and place of birth and their filiations. These documents are usually delivered by municipalities, in charge of the Civil Registers, or abroad by Consulates. There is no centralised Civil Register in France.

An informative report to the French Senate in 2005 [11] observes that the staff in charge of handling these documents is sometimes insufficiently prepared and do not respect the existing guarantees of the procedure such as the persons whom this document can be handed to (the beneficiary itself, its ascendants, descendants or a third person with a valid mandate). This facilitates the issuance of authentic documents to unauthorized persons. In foreign countries, two main issues are identified as facilitating fraud, namely cases of corruption of civil officers and the lack of quality of Civil Registers (several persons recorded under the same name, out-dated registers, etc.). The Ministry of Foreign Affairs has observed an increase in the production of irregular or forged "justificatifs d'état civil" issued in foreign countries.

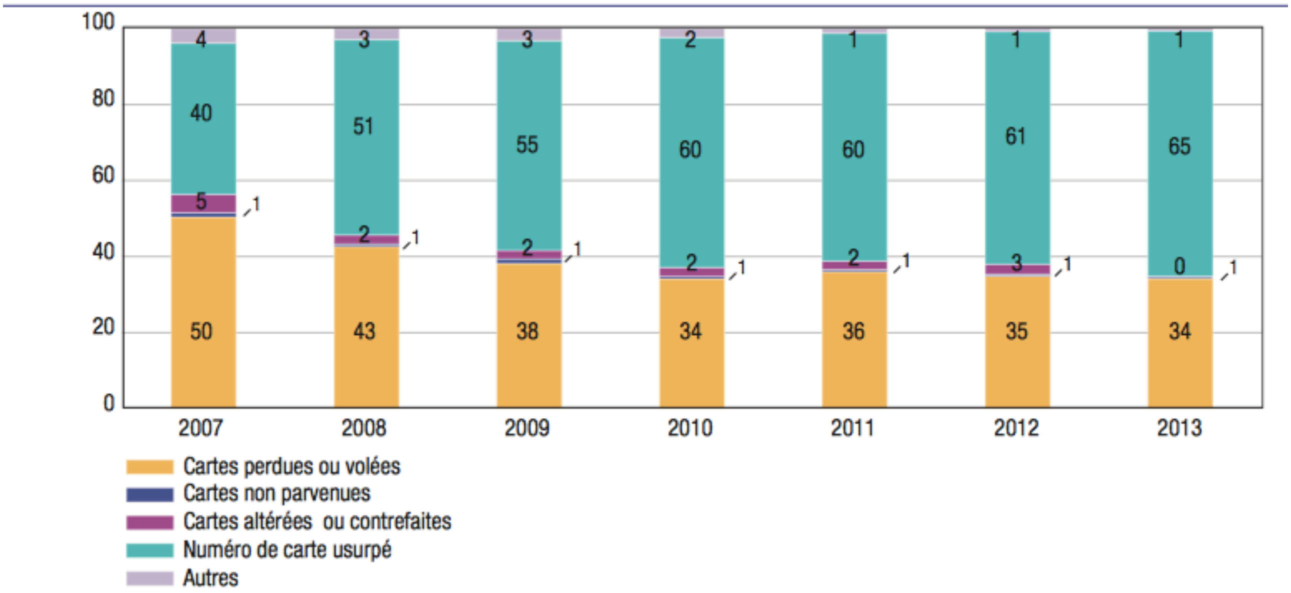


Faits constatés par la Police nationale	2007	2008	2009	2010	2011	2012	Évolution 2007/2012 (%)
Total des fraudes documentaires et/ou identitaires	**12 039**	**11 870**	**10 997**	**10 495**	**11 374**	**11 661**	***- 3,1***
Variation annuelle en %	-	*- 1,4*	*- 7,4*	*- 4,6*	*+ 8,4*	*+ 2,5*	
Faux documents d'identité	**6 890**	**7 121**	**6 266**	**5 583**	**5 969**	**6 073**	***- 11,9***
Variation annuelle en %	-	*+ 3,4*	*- 12,0*	*- 10,9*	*+ 6,9*	*+ 1,7*	
Faux documents concernant la circulation des véhicules	**2 269**	**2 290**	**2 292**	**2 473**	**2 832**	**2 919**	***+ 28,6***
Variation annuelle en %	-	*+ 0,9*	*+ 0,1*	*+ 7,9*	*+ 14,5*	*+ 3,1*	
Autres faux documents administratifs	**2 880**	**2 459**	**2 439**	**2 439**	**2 573**	**2 669**	***- 7,3***
Variation annuelle en %	-	*- 14,6*	*- 0,8*	*0,0*	*+ 5,5*	*+ 3,7*	

Source : État 4001 annuel, DCPJ.

Figure 6 Number of documents and/or identity frauds certified by the French police between 2007 and 2012

Looking at the Figure 6 it is clear how the document fraud in last six years has been almost constant. The only significant decrease started in 2009 and has continued also for the next year. A possible explanation for this trend could be that in 2009 entered into force the biometric passport, which, according to the report, "marked a milestone in the dissemination of these identity documents. It has also been a key term in the tasks committed to the national agency of documents security (ANTS - Agence Nationale des Titres Sécurisés), established in February 2007. The development of a new generation of securities responds to a purpose of fight against forgery and counterfeiting" [12].



Source : Observatoire de la sécurité des cartes de paiement

Figure 7 Distribution of France national frauds according to its origin

Since the Figure 6 shows an increase in document fraud occurrence in 2011, it is suggested to make the eID card an official man of identification, in order to introduce a more secure hand-out procedure, mainly through the centralisation of the procedures and the introduction of biometric identifiers.

In conclusion, identity fraud in France is considered a big problem only from a document fraud point of view. This is actually the main argument advanced by the government to legitimise the introduction of an eID card with biometrics identifiers and based on centralised databases. It still does not exist yet in France but it is already fuelling a heated controversy. Different organisations, including the French League of Human Rights, immediately condemned the eID, calling it "an infringement of human rights". The French government argues that the eID will help fight against impersonations (200000 cases of which are recorded in France every year) but some organisations are worried about the commercial exploitation of data and the creation of a database with the fingerprints and personal information of every French citizen [13].

3.3.2 Credit card fraud

Since 2003 the Observatory for Payment Card Security, under the resort of the French National Bank, issues an annual activity report that includes a survey of credit card fraud. The typologies defined to distinguish the origin of fraud are: lost or stolen cards, non-received cards, forged or counterfeit cards, stolen card number and a category "other", which concerns the fraudulent opening of an account upon an identity theft.

Figure 7 shows how the most important source of fraud (65%), constantly increasing since 2007, is the use of the stolen credit card numbers in the fraudulent remote transactions. For this kind of fraud, three main issues have been identified:

- For *fraudulent interception of card numbers*, the report reveals that none has been registered when the payment is made in a secure on-line environment. The reasons seem to rely, on the one hand, on the fact that service providers are allowed in very restrictive cases to store users' credit card information, and on the other hand, on the encryption mechanisms used.

- For *automatic generation of credit card numbers*, the cases remain marginal and not exclusive to the on-line environment. Off-line payments also suffered from this type of fraud. In that sense, safeguards have been implemented by actors of on-line payment channels (see Section 3.2.3). They managed to keep the risks originating from these practices under control.
- For *fraudulently obtaining credit card numbers through off-line payments*, the problem resides in the fact that credit card numbers appear on the receipt kept by the merchant. This number is currently necessary, in case of technical problems, to re-enter the transaction realised via the credit card. The solution has consisted in using a visual cryptogram printed in the credit card, which allows the merchant to ensure that the user has the credit card in their possession. Another weakness resides in purchases made by telephone, as the number is given to the merchant by the buyer. Police investigations demonstrated that embezzlers got the numbers through these practices in certain shops such as computers shops or petrol stations. The Observatory recommends a progressive process to delete credit card numbers on all receipts.

The second most important source of fraud, despite its decreasing tendency since 2007, is still the one regarding the lost or stolen cards (34%). In this case, only good practices by the users could be the best countermeasures.

Counterfeit cards constitute only 0,2% of fraudulent domestic payments. This significant decrease is mainly due to the adoption of chip technology by some credit cards and by the strengthening in the security of existing EMV (Europay, MasterCard and Visa) smart cards.

EMV is a global standard for inter-operation of integrated circuit cards (IC cards or "chip cards"), IC card capable point of sale (POS) terminals and automated teller machines (ATMs), for authenticating credit and debit card transactions. The strengthening in EMV's security mentioned consists in the migration from the identification technology SDA (Static Data Authentication) to DDA (Dynamic Data Authentication). SDA is an improvement on the old magnetic stripe cards, since the former are much harder for fraudster to attack and counterfeit. However, SDA cards have known security weaknesses: fraudsters may still be able to collect the necessary chip data from SDA cards at the POS to produce counterfeit chip data, since the signature used for SDA cards is the same every time. DDA cards come as an improvement of SDA ones as they store an encryption key that generates a unique number for each transaction that is only valid for one authentication.

3.3.3 Technical countermeasures against credit card fraud

Various technical and organisational safeguards have been dictated by financial institutions and online merchants in order to secure the payment procedure:

- *The virtual dynamic card*: the e-credit card [*e-carte bleue*]. This service allows the consumer to create in real time a new credit card number for each transaction. This number remains valid for a certain time and is deactivated once it has been used. It thus prevents the reuse of the credit card number and double invoicing;
- *The system Sympass*: Sympass is a company created in 2001 that has developed a tool relying on the use of both the computer and the telephone keyboards. When buying online, the user gives the 8 first digits of his credit card and a phone number. He then immediately receives a phone call of an automated voice service asking him to key the last 8 numbers of the credit card;

- *Payment by card without any indication of the number (the ID Tronic solution)*: in this system it is not necessary to provide the credit card number when conducting a transaction. When registering, the user provides his payment data to the bank, which provides him with a password. When making the payment, the user provides his password or email address and receives a text message with a second password to authenticate the user.
- *3D secure system*: this system integrates an additional step in the payment procedure. When the card number is sent by the merchant to the bank for authorisation, this entity will request the cardholder to authenticate to the system before sending such authorisation to the merchant. The merchant will thus not be held liable in case of identity fraud due to the additional check made by the bank during the payment. However, full implementation of the system faces strong opposition from both merchants and banks on the basis of economical and technical reasons.

3.4 Estonian ID card

In all the other European countries the situation is more or less similar to that in the UK and France, not with respect to the numbers, which in those two countries are more relevant, but regarding the kind of issues that generate the identity fraud. We can say that for the countries that don't adopt a compulsory ID document the situation is more similar to the UK one, while all the others can be compared to France.

While speaking about national ID cards and, more specifically about electronic ID card, it is worth to mention the Estonian case.

Known as a technological pioneer in Europe, Estonia adopted the eID in 2002. Each citizen had to get the card when it was introduced and today almost 90% of the population (1 million people) owns one. The country seems to be the one using the electronic identity card to its fullest potential.

All Estonian citizens and permanent residents are legally obliged to possess such a card from the age of 15. The card is equipped with a chip, which stores private and public keys and public key certificate. Every user is provided with two pin codes: the first to authenticate and the second for digital signature. Public keys are published online and signed by the state to certify them. It is claimed that the secret keys never leave the card.

The card's compatibility with standard X.509 infrastructure has made it a convenient means of identification for use of web-based government services in Estonia. All major banks, many financial and other web services support ID-card based authentication. It is also possible to use it to encrypt e-mails and files, purchase "virtual" transportation tickets linked to the ID cards and for authentication in Estonia's ambitious Internet-based voting programme. In February 2007, Estonia was the first country in the world to institute electronic voting for parliamentary elections. Over 30000 voters participated in the country's first e-election. By 2014, at the European Parliament elections, the number of e-voters has increased to more than 100,000 comprising 31% of the total votes cast.

Under Estonian law, since 15 December 2000 the cryptographic signature is legally equivalent to a manual signature.

This card should be taken as an example for all the other nations; it is considered to be resilient to cloning and so the only possibility to commit an identity fraud with this card is to steal it together with the PIN numbers. Nevertheless, the Estonian ID number (isikukood) is quite public information, and this can represent a potential vulnerability. For instance, someone could use this number for medical identity theft since the Estonian hospitals seems not to electronically verify people ID cards. More in general it can represent a vulnerability for every institution where people need to apply with the ID number and this is not electronically verified. Of course this situation can easily be avoided by verifying everyone's identity card with the card reader.

It is worth mentioning that the card do not store biometric information. Therefore, for the countries that fear a breach in people's privacy it can be a good example to follow.

4. Automated identity theft attacks on social networks

4.1 Introduction

Social networks are an attractive target for identity thieves because of the big amount of information and sensitive data that people use to put in their profiles, like for example e-mail addresses, education, friends, professional background, activities they are involved in, their current relationship status and sometimes also their previous relationships.

In the paper [14], the authors investigate how easy it would be, for a potential attacker, to launch automated crawling and identity theft attacks in some popular social networking sites, in order to gain access to a large volume of personal user information.

They performed two kinds of attacks. The first one consists in cloning an existing profile on a social network in order to gain its contacts, and therefore their personal information, by sending them a friendship request. Their experimental results show that typically a user tend to accept a friend request from a profile that is already a confirmed contact in his friends list. In the second attack, they show that it is effective and feasible to launch an automated, cross-site profile cloning attack. In that attack, they automatically identify users who are registered in one social network, but not in another one. Then they clone the profile of a victim in the social network where he is registered, and forge it in another one where he is not registered yet. After creating the forged identity, they automatically try to rebuild the social network of the victim by contacting his friends that are registered on both social networking sites.

The results show that this attack is more effective than the previous one, since in this case the profile of the victim exists only once in a social network, and therefore, the friendship requests are more likely to be accepted without raising suspicion.

An identity thief that performs successfully those kind of attacks to a given amount of people, obtaining therefore their friendship on social networks, can not only gain all the information that they share with their friends, but also send private messages containing links to fake websites. In this way they can induce them to give more information with the methods explained earlier in this paper.

The implementation for those two attacks has been developed in a prototype system called iCloner (identity Cloner). It consists of several components that are able to crawl popular social networking sites, collect information on users, automatically create profiles, send friend requests and personal messages, and it also supports CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) analysis and breaking capabilities.

Those attacks can potentially be launched on a large scale, allowing an attacker to control hundred of thousands of cloned accounts and thus reaching millions of real user profiles. Furthermore, if the attacker has a high number of different IP addresses at his disposal (such as a botnet that consists of thousands of compromised hosts), the detection of an automated attack like the ones presented in this paper may become more difficult.

4.2 Architecture of iCLONER

iCloner consists in four main components: the crawler component is responsible for crawling the target social networking site and collecting the information in the contact lists and profiles if these are accessible to the public. The identity matcher analyzes the information in the database and tries to identify profiles in different social networks that correspond to the same person. The profile creator component can then use that information in order to create accounts in a social network. Finally, the message sender component is responsible to login into the created accounts and automatically send friend requests to the people that are known to be friends with the victim. Depending on the social networking site that is being targeted, CAPTCHAs might need to be solved in order to create accounts, to send friend requests, and sometimes even to access a user's profile (if a user sends many requests, a social networking site might request to verify that the user is a real person and not a script). The CAPTCHAs are analyzed by the CAPTCHA analysis component. In particular, the authors have analyzed the CAPTCHAs that are displayed by SudiVZ, MeinVZ, and Facebook and have designed techniques to break these CAPTCHAs with a success rate that makes automated attacks feasible in practice. They have not encountered CAPTCHAs on LinkedIn, and did not need to solve CAPTCHAs with XING.

4.3 Breaking CAPTCHAs

4.3.1 The environment

In order to automate their attacks the authors developed a number of CAPTCHA breaking techniques based on a set of open source tools and custom-developed scripts. They used ImageMagick for image filtering, Tesseract for the text recognition using OCR (Optical Character Recognition), and wrote a number of Python and Perl scripts to partition the CAPTCHAs and to apply manipulations at the pixel level.

4.3.2 Mein VZ and Studi VZ CAPTCHAs

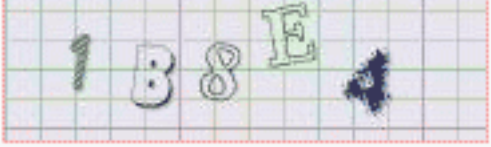


Figure 8 Example of a Mein VZ or Studi VZ CAPTCHA

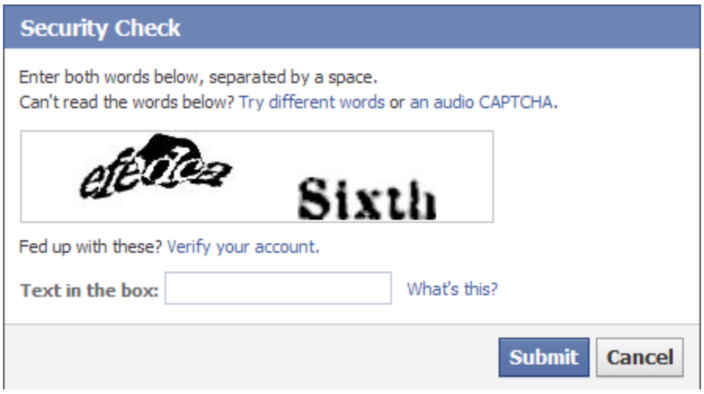
In both those social networks each CAPTCHA always contains exactly five letters, each letter is written in a different font, with differing foreground and background colors, and furthermore, it is often tilted, scaled, or blurred. In addition, a simple grid-based noise is added to the image.

The authors wrote a Perl script to detect and remove the grid noise and to replace the background with white pixels. A second script, then, attempts to identify the connected areas and partitions the image around them in order to isolate the single letters. In a case when the number of connected regions is different from the five letters (e.g. because two or more letters are partially overlapping, which is by the way not common), the old CAPTCHA query is discarded and new query is asked.

The next step consists into trying to match each letter against a set of known fonts, in particular, each font character (tilted from -10 to +10 degrees) is compared against the extracted letter and the number of matching pixels is counted. If this number is larger than a certain threshold (dynamically calculated as a percentage of the total number of pixels in the character), the match will be considered positive. If the match is not positive, six variations of the unknown letter are generated by applying a chain of ImageMagick's filters (adaptive-blur, contrast, contrast-stretch and black threshold in different combinations). Then the Tesseract engine is run on each variation, and if at least three equal results are obtained, a letter is considered to be recognized.

Finally, if there is a positive match for each of the five extracted letters, they will be concatenated to compose the answer to submit. Since in those two social networks a user can make only three errors in submitting CAPTCHAs before being banned, but they can be asked an arbitrary amount of times, the authors decided to discard solutions where there are ambiguous letters like “I” that could be confused with “1”, “S” could be confused with “5”, “0” can be confused with “O”, and “8” can be confused with “B”. In this way, in their experiments 71% of the CAPTCHAs weren't recognized, but between those that were recognized the correct answers were 88,7%, so the CAPTCHAs could be solved in 99,8% of the cases in one of the three consecutive permitted attempts.

4.3.3 Facebook CAPTCHAs



Security Check

Enter both words below, separated by a space.
Can't read the words below? Try different words or an audio CAPTCHA.

Fed up with these? Verify your account.

Text in the box:

Figure 9 Example of a Facebook CAPTCHA

Facebook adopts the reCAPTCHA solution that is a state-of-the-art approach developed at Carnegie Mellon University. It consists in using words that are encountered while digitizing books, but that cannot be correctly recognized by the OCR program. Using these words as CAPTCHAs has two main advantages: first, since a computer has failed to recognize them in the first place, it means that they are more difficult to break by automated programs;

second, when a human solves a CAPTCHA by reading the words, he contributes to the effort to increase the accuracy of the digitalized text of the book.

This type of CAPTCHA is composed of two words displayed at the same time: one of the words is an unknown word that the system was unable to read while digitizing a book; the other one is a known word that a number of other users have already identified. As the reader could guess the words presented to the user are meaningful and composed by a variable number of letters, but in order to complicate the CAPTCHA solving, they are slightly distorted and covered by a curved line. When the answer of the CAPTCHA is submitted, if the user correctly recognizes the known word, there are good chances that also the answer of the unknown word is correct.

In this case, the authors decided to perform a word-based analysis, since partitioning each character would be complicated, and also not very helpful.

The first step with this technique consists in attempting to unbend the word back to the original shape by extracting the middle line of each word (i.e., the sequence of pixels that are half way between the top and the bottom of each letter). Since this line is very irregular (e.g., it goes up for letters such as "t" and "l" and down for letters such as "g" and "p"), it is smoothed by approximating it with a third degree polynomial curve. After this process, each pixel column is translated up or down so that the approximating curve becomes a straight line.

The second phase is similar to the one used for the MeinVZ and StudiVZ CAPTCHAs: a number of different versions of the images containing the extracted word is generated by applying different combinations of ImageMagick filters and then is run Tesseract on each one. The text collected from the Tesseract output is then analyzed by a lexical module where each word is compared with the content of the English dictionary, if it does not match any known word, the program attempts again with an edit-distance spell correction algorithm to compensate for small errors in the text extraction routine. Since a large fraction of the words used in reCAPTCHA are not present in the English dictionary (e.g. person or geographical names, different languages, etc.), if the previous two tests fails, the word will be submitted to Google. If the number of the Google results is higher then a configurable threshold, the word is considered to be correct, otherwise is performed an attempt to substitute it with the Google suggestion, if Google makes one. If all the tests fail, the CAPTCHA will be drop and will be asked new one, since also in this case the user can ask it for an arbitrary amount of times.

In their experiments, the authors manually verified the result of their system when submitting 2000 reCAPTCHA words. On average, their tool was able to correctly recognize 14% of them. That is, 26% of the CAPTCHAs submitted, correctly identified at least one of the two words.

After a number of failed attempts, reCAPTCHA seems to become more resilient breaking attempts. A possible reason is that after a certain number of errors, the system starts to send CAPTCHAs containing two known words, thus verifying that both words are recognized correctly.

However, since the number of CAPTCHAs that this attack requires to break is fairly limited (i.e., we need to solve CAPTCHAs only when creating accounts and sometimes when sending friend requests), it is still feasible. Indeed, Facebook never banned their accounts even after submitting thousands of wrong answers. If we also consider that an attacker could use a botnet to have access to thousands of different IPs and distribute the

CAPTCHA breaking effort among many hosts, she would still be able to send thousands of friend requests every day.

4.3.4 Summing-up

The experiment results in the paper [14] show how easy can be to crawl profiles in some social networks. For example in Studi VZ and Mein VZ because of the easy-to-break CAPTCHA system that they use; in XING there is no CAPTCHA protection against automatic crawling but at least a better system for blocking profiles which generates a higher number of requests. Nevertheless there will be enough profiles crawled before an account can be banned.

In the profile cloning attack, the authors, in the experimentation part, tried to send friendship requests to the friends of some of the profiles that they successfully forged. They tried also to send friendship requests to the same profiles by some invented profiles they created to test the different behavior. The result is that for the forged profiles the acceptance rate was always over 60%, while for the fictitious profiles was below 30%. This confirms that by forging profiles, an attacker can achieve a higher degree of success.

They then tried to send private messages, containing a link to a fake website, to the profiles that accepted the request from the fictitious accounts, then, the same messages to the remaining contacts from the forged accounts. In both cases, interestingly, the difference is not significant (around 50%). This demonstrates that those presented attacks can be effectively used for spamming and directing a large number of users to web sites under the control of the attacker.

The cross-site profile cloning was performed between XING (source) and LinkedIn (destination). The result is that 56% of the requests were accepted. In this case, by the way, the most difficult part is to find profiles that are present in one social network but not in the other and friends of them that are present in both social networks.

In conclusion we can say that something can be done to improve the CAPTCHA system used by some social networks and the security measures adopted against the automated crawling, but the weakest thing in this field is the users' behavior.

5. Conclusions

The United States are clearly the country in which identity theft is more common compared the other countries. It is also the more interesting case of study.

Government, media and general public in the US are aware of the importance and the spread of this crime in their country, but despite their effort to fight it, still a lot has to be done.

It seems to be clear that the most important weakness in the US system is the use of the SSN as the main identifier. The first reason is because it is not an identity document since the SSN card has no picture. Second, because the algorithm used to produce it, before the randomization introduced in 2011, has been proved to be weak and predictable.

For what concern the private sector, certainly the main suggestion is to strengthen the controls and verifications in order to prevent frauds.

From the analysis conducted in this document, it seems that the countries that are more affected by identity theft are those in which there is no compulsory identity document, like

USA, Canada and UK. In those cases the first suggestion in order to fight identity theft is clearly the introduction of a mandatory identity document for all the residents.

In general, we can state that there are three main countermeasures to adopt and improve, within all the countries, in order to fight against identity theft:

- *Legal countermeasures*: there is the need of specific laws that define identity theft-related crimes and therefore adequate punishments for criminals. In this way it is also easier for a nation to quantify the number of cases of identity fraud and the damages that it causes;
- *Technical countermeasures*: they consists in the strengthen of the security measures adopted, both online and offline, like for example the ones introduced in France which are descripted in section 3.2.3;
- *Behavioural countermeasures*: this is probably the most important one. The adoption of good practices (like the ones introduced in section 1.3) from the population susceptible to identity theft is probably the countermeasure that can be more effective for the fight against this kind of fraud. For this purpose, every country should promote campaigns to inform people with any means: media, squares, courses in working places and schools, etc.

References

[1] CIPPIC Working Paper No. 2 (ID Theft Series), Techniques of Identity Theft, March 2007. https://cippic.ca/sites/default/files/bulletins/Techniques.pdf

[2] Consumer Sentinel Network reports. http://www.ftc.gov/system/files/documents/reports/consumer-sentinel-network-data-book-january-december-2013/sentinel-cy2013.pdf

[3] Alessandro Acquisti and Ralph Gross, Predicting Social Security numbers from public data, Carnegie Mellon University, Pittsburgh, PA 15213. http://www.pnas.org/content/106/27/10975.full.pdf+html

[4] FIDIS Deliverables, Identity-related Crime in Europe – Big Problem or Big Hype? http://www.fidis.net/resources/fidis-deliverables/hightechid/d127-identity-related-crime-in-europe-big-problem-or-big-hype/doc/30/

[5] Snopes.com, Money Tree. http://www.snopes.com/business/bank/treecard.asp

[6] Bill Hardekopf, Your Child’s ID is a Big Target for Identity Thieves, March 24, 2014. http://www.lowcards.com/childs-id-big-target-identity-thieves-23332

[7] Kaiser Healt News, Rise of Identity Theft. http://kaiserhealthnews.org/news/rise-of-indentity-theft/

[8] Government of Canada, Statistics Canada http://www5.statcan.gc.ca/subject-sujet/result-resultat?pid=2693&id=-2693&lang=eng&type=STUDIES&pageNum=1&more=0

[9] CIFA’s Annual Reports. http://www.cifa.in/web/PUBLICATIONS/ANNUALREPORTS.aspx

[10] Mercedes Bunz, Identity fraud is the UK's fastest growing crime in 2009, theguardian.com, October 12, 2009. http://www.theguardian.com/media/blog/2009/oct/12/ukcrime-id-theft-rising

[11] Rapport d’information au Sénat n°439 sur la nouvelle génération de documents d’identité et la fraude documentaire

[12] Rapport d'information de Mme Michèle ANDRÉ, fait au nom de la commission des finances n° 486 (2008-2009) - 24 juin 2009

[13] Peter Morrison, The complicated rise of the electronic identity card in Europe, myeurop.info, April 6, 2012. http://en.myeurop.info/2012/04/06/complicated-rise-electronic-identity-card-europe-5145

[14] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, Engin Kirda, All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks, EURECOM, Sophia Antipolis, France