

A Guardian Angel Protects You From Identity Theft

Latanya Sweeney, *Carnegie Mellon University*

Meg Kemp recently graduated from college and is looking for a job, so she posted her resume on-line. Norton Steuben is a retired law professor who hasn't looked for employment in over 35 years and rarely uses the Internet; yet, his school maintains his curriculum vita on-line. The on-line resumes of both of these people initially put them in danger of being identity theft victims, but they received some unexpected protection from a computer program¹.

Identity theft is a growing problem, and personal information included in on-line resumes and vitae can increase a person's risk. Most people are unaware of the risk they place themselves or others when on-line resumes include Social Security numbers, dates of birth, and personal mailing addresses. In the United States, these fields of information can be used to fraudulently acquire new credit cards in someone else's name without their knowledge.

In many cases, resume information is willingly provided by people in pursuit of employment. In other cases, organizations post vitae of employees and affiliates, sometimes without their knowledge. Of course, no one is proposing resumes and vitae not be posted on the Internet. Instead, care should be taken in selecting information to include in an on-line resume or vita.

On-line resumes and vitae should not contain Social Security numbers. Age can be used instead of date of birth, but even age should be omitted whenever possible. Home addressees should be included only if absolutely necessary.

Remember, once a resume appears on the Internet, it is likely to be made eternally public thanks to search engine caches and Internet archiving organizations (e.g., archive.org). The best remedy is to never post sensitive information on-line.

If sensitive resume information is already on-line, then exposure needs to be limited. It would be helpful if the person who is the subject of the resume could be notified and made aware of the risk in the belief that informed subjects are likely to remove sensitive information. Notification and education can provide an effective means to help achieve this outcome. But how are these people found? How are they notified? There are over 14 million on-line resumes to consider².

Technology can help. Imagine a benevolent program that crawls through freely available information on the Internet and emails people for whom uncovered facts can be combined sufficiently to impersonate them in financial or credentialing transactions. This is the ambitious goal of the Identity Angel Project in the Data Privacy Laboratory at Carnegie Mellon University³. Identity Angel currently: (1) locates resumes that contain sufficient information to fraudulently acquire a new credit card; and then, (2) notifies the subjects of these resumes by email – encouraging them to remove sensitive information. Individuals like Meg and Norton have benefited from Identity Angel.

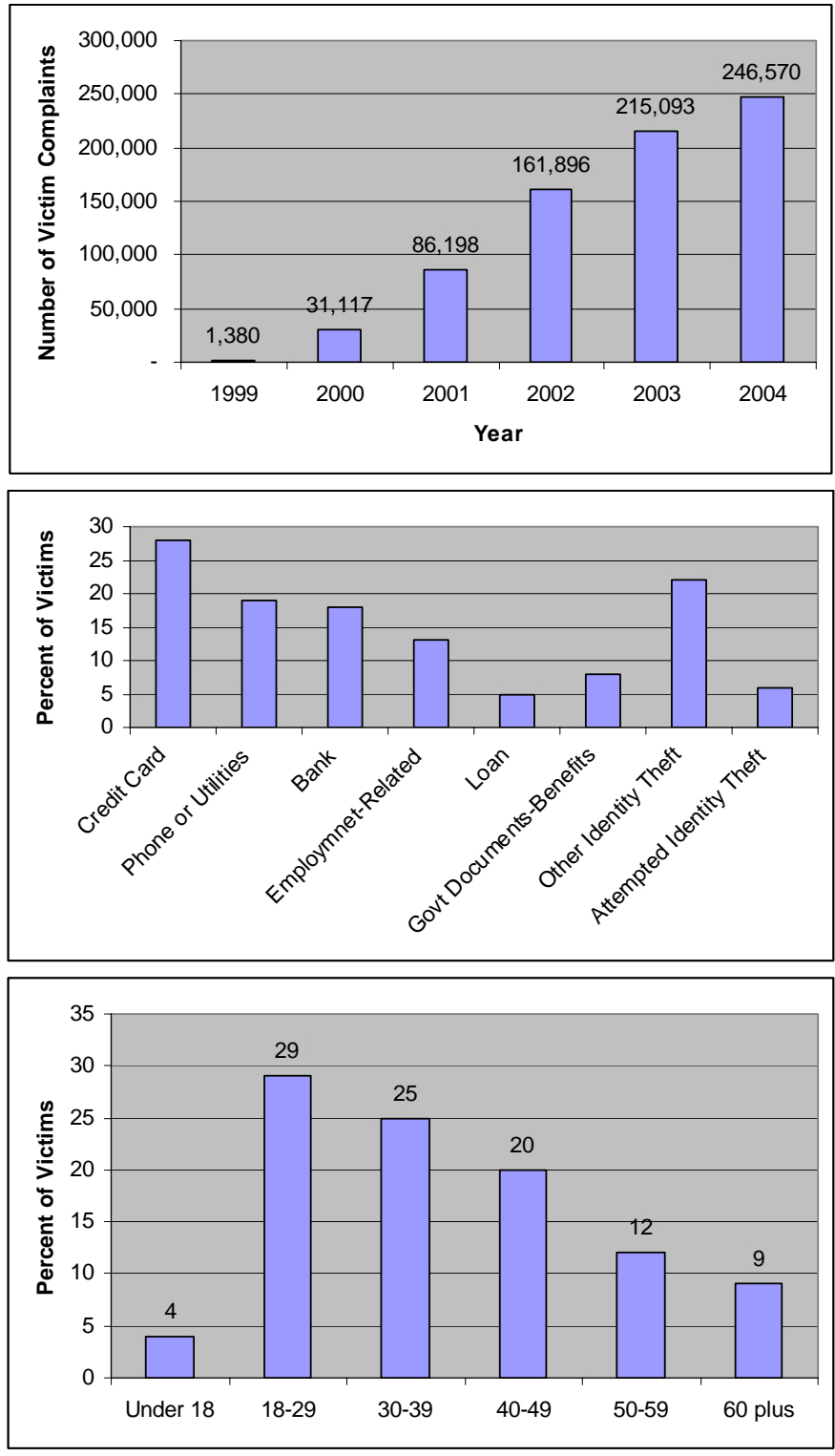


Figure 1. Information from the Federal Trade Commission Report on Identity Theft, based on victim complaints.

The Nature of Identity Theft Problems

Identity theft occurs when one person uses another person's personally-identifying information without permission to commit fraud or other crimes. The Federal Trade Commission Report on Identity Theft⁴ shows rapid growth in victim complaints of identity theft received at their clearinghouse. More than 86,000 cases were reported in 2001; that grew to 162,000 cases in 2002; and was 246,570 in 2004. More than one-fourth (28%) of these involved credit card fraud. Of credit card fraud, the report identified about half (or 17% of all thefts) as new accounts, making the acquisition of new credit cards, a major identity theft problem. See Figure 1.

Most incidents (29%) occurred among younger adults, who tend to have resumes and facts about themselves posted on the Internet. They also tend to have multiple residences in a short time period, making the issuance of a new credit card to a fraudulent address more difficult to determine. Identity Angel seeks to help this group, among others, by focusing on the risk of new credit card fraud related to on-line resumes.

There is some additional good news. Courts have maintained that a victim is not liable for charges made by a credit card issued to someone else who forged the victim's name on the credit card application—provided the victim knew nothing about the credit card⁵. So, a victim does not pay for the goods and services purchased on a fraudulently acquired card. Instead, the expense is paid by everyone through higher interest rates and aggressive fee structures.

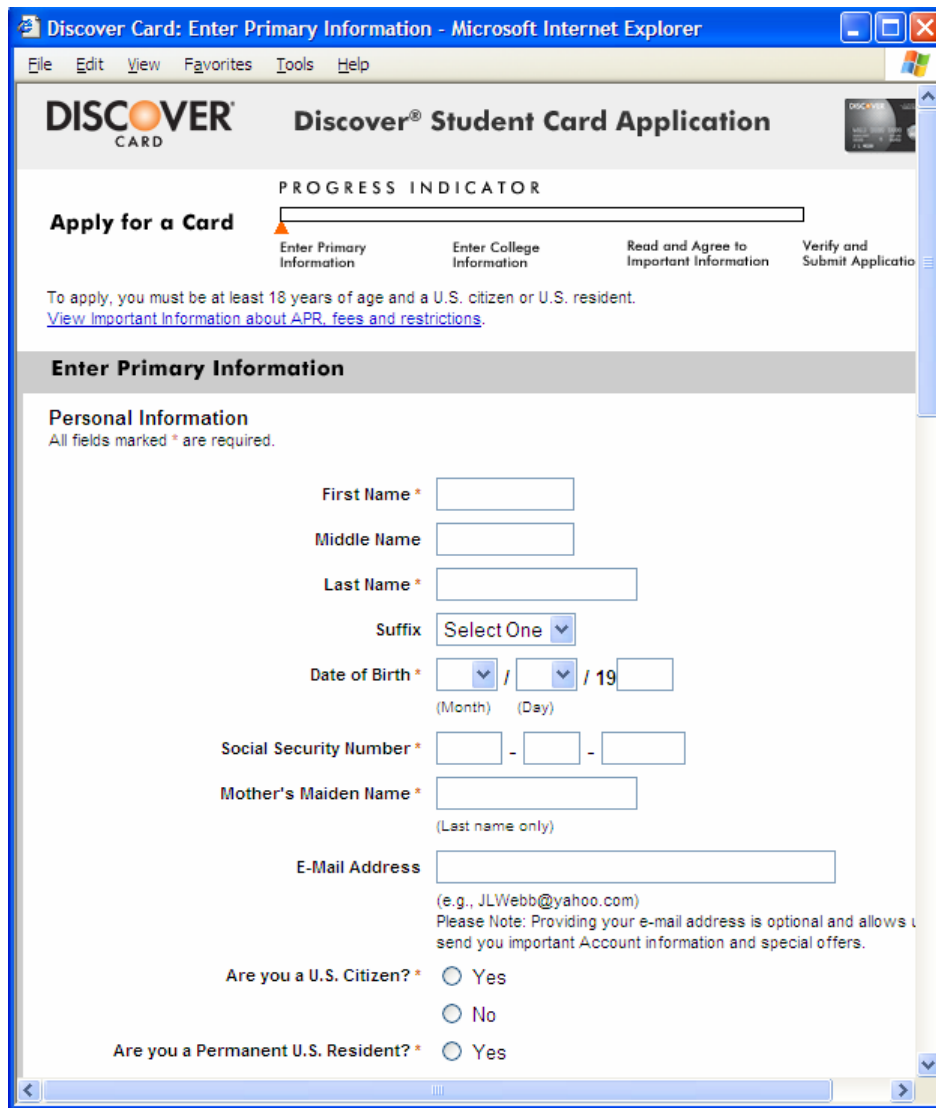
The personally bad news for the victim relates to trying to clean-up his credit report. People whose identities have been stolen have spent years—and lots of money—cleaning up the mess thieves made of their credit record. Several on-line logs (e.g., Identity Theft Spy) archive cases appearing in the press⁶.

Identity thieves will often get several cards in the victim's name. Each fraudulently acquired card has to be corrected separately. The existence of the cards can go undetected by the victim for years, making correction even more difficult once discovered. For many years later, victims report lost job opportunities, refused loans for education, housing and cars, and have even been arrested for crimes they did not commit^{4,5,7}.

The Federal Trade Commission encourages people to review their credit reports annually⁸. Overall, this is a good practice so problems aren't detected just when credit is most needed, but there is a caveat. If an unknown credit card appears on a credit report, the victim must immediately begin correction and notification procedures, else risk being forever responsible for the debt. Once the credit report is received by the victim, the victim has been "notified." Inaction on the part of the victim may be interpreted as authorizing use of the card. An editor at Consumer Reports examined credit card reports and found half those checked contained errors; the two reasons cited were being mistaken for another person with a similar name and fraud⁹.

Applying for a New Credit Card

When applying for a new credit card, the applicant is only represented by the information provided on the application itself. The basic Information necessary for a credit card application is: {*Name, Social Security number, Address, Date of birth, Mother's maiden name*}. Figure 2 shows an example.



The screenshot shows a web browser window titled "Discover Card: Enter Primary Information - Microsoft Internet Explorer". The page header includes the Discover Card logo and "Discover® Student Card Application". A progress indicator shows four steps: "Enter Primary Information" (current), "Enter College Information", "Read and Agree to Important Information", and "Verify and Submit Application". Below the progress indicator, a note states: "To apply, you must be at least 18 years of age and a U.S. citizen or U.S. resident. [View Important Information about APR, fees and restrictions.](#)"

The main section is titled "Enter Primary Information" and "Personal Information". A note says "All fields marked * are required." The form includes the following fields:

- First Name *
- Middle Name
- Last Name *
- Suffix (Select One dropdown)
- Date of Birth * (Month/Day/19__ format)
- Social Security Number * (___-___-___ format)
- Mother's Maiden Name * (Last name only)
- E-Mail Address (e.g., JLWebb@yahoo.com)

Below the fields are two questions with radio button options:

- Are you a U.S. Citizen? * (Yes/No)
- Are you a Permanent U.S. Resident? * (Yes)

Figure 2. An on-line credit card application. Requested demographic information is shown and also includes phone and address (not shown). In student applications, the name of the college and expected year of graduation is also requested. [https://www.discovercard.com/cardmembersvc/discovercard/apply-for-a-card/primaryinformation]

How might an imposter gather the necessary information freely over the web? Mother's maiden name is used as a challenge question "after" the credit card is issued and not verified beforehand. So, in most cases, any name will suffice as Mother's maiden name.

The original address needs to be known, so a change of address can be included with the fraudulent application. Name searches on phone directories can often be used to find addresses.

Several websites provide a date of birth, given a person's name (e.g., anybirthday.com). So, the most sensitive information to acquire is the Social Security number and its matching name.

Social Security Numbers

A key element to fraudulently acquiring a new credit card is the Social Security number. These have evolved into national identifying numbers for individuals living and working in the United States. They are essential to identifying, recognizing, and authenticating people in health, financial, legal, and educational information. Some people might falsely believe that access to Social Security numbers, while available within many financial, health, employment, and government institutions, are not publicly available.

In 2003, the U.S. General Accounting Office¹⁰ identified Social Security number vulnerabilities as ripe for exploitation by terrorists [and other criminals], making Social Security number problems a serious concern to homeland security and a grave threat to the country's economic prosperity. These groups may gain funds for their activities while simultaneously causing havoc to the country's economy and citizens. The risks are real, yet poorly understood.

The California-based Foundation for Taxpayer and Consumer Rights said for \$26 each it was able to purchase the Social Security numbers and home addresses for Tenet, Ashcroft and other top Bush administration officials¹¹.

In contrast, Identity Angel seeks to obtain Social Security numbers for larger numbers of ordinary people using easily accessible and free on-line sources. In terms of risk, the less expensive sensitive information is to obtain and the greater the number of people having access to it, usually the greater the opportunity for abuse. Therefore, Identity Angel focuses on Social Security numbers available freely on the Web.

Methods for Mining Resume Information

Acquiring information sufficient to fraudulently obtain a new credit card using on-line resumes consists of locating on-line resumes, extracting relevant information and emailing subjects. Accomplishing these tasks builds on prior work.

In 2004, Sweeney¹² introduced a system that locates on-line lists of names of people ("rosters"). Rosters are evasive to search engine retrieval because they do not lend themselves to keyword lookup. Using expressions such as "employees" or "students" returns hundred of pages, but finding the rosters among them previously required many hours of human inspection. Sweeney's approach ("filtered searching") executes a simple predicate function on each page retrieved from keyword searches to determine whether a given page is an instance of the kind of page sought (e.g., a roster).

Identity Angel uses filtered searching to locate on-line resumes, because merely entering “resume” in a search engine yields thousands of web pages about resume writing and resume submissions, as well as, resumes themselves. Filtered searching confirms whether a given web page has format (using layout cues) and content (using headings) consistent with that of a resume or vita.

In 1996, Sweeney¹³ used a system of entity detectors in a black board architecture to extract personally-identifying information from text files (e.g., letters and clinical notes). To date, the system continues to out-perform statistical and linguistic based approaches.

Identity Angel also uses entity detectors, which in this case, are simple regular expressions, to identify instances of dates of birth, email addresses and 9-digit Social Security numbers appearing in resumes. These detectors take advantage of the ways in which dates, email addresses and Social Security numbers are traditionally written. The “@” dot (“.”) notation in email addresses is a writing convention in resumes unique to email addresses. Similarly, Social Security numbers are written sometimes as 9 contiguous digits and other times as 3 digits, dash (“-“), 2 digits, dash, followed by 4 digits. In resumes, key phrases, such as :”date of birth,” “SSN,” or “Social Security number,” appear adjacent to their related values. Identity Angel exploits these writing conventions to harvest sensitive information from resumes.

Experimental Results

Materials

Experiments were based on: (1) FilteredSearch Java code; (2) two resulting resume databases; and, (3) entity detectors, as described below.

FilteredSearch. Java code that uses the Google API to perform a series of searches, pruning out duplicate pages.

Resume Databases. Using FilteredSearch on keywords {“resumes”, “vitae”}, the first n distinct actual resumes containing Social Security Numbers (“SSNs”) were compiled into a database. DBA has $n=150$ resumes from December 2003 and DBB has $n=75$ resumes from December 2004 (excluding any in DBA).

Detectors. Regular expressions that identify ways of writing dates, SSNs and email addresses, as described previously.

SSNwatch. To confirm whether a provided number was actually an SSN, the SSNwatch Validation Server was used to confirm that the number was likely to be valid¹⁴.

Experiment 1: Finding Sensitive Resumes

The first 2000 suspect resume web pages retrieved from Google were reviewed using FilteredSearch and the SSN detector to produce DBA. Later, DBB was produced similarly. Figure 3 provides some examples.

Richard Allen Brown. PO Box 782. Kayenta, AZ 86033. Home Telephone-520-697-3513. NAU Telephone-520-523-4099. DOB: 03-10-77. SSN: 527-71 ... dana.ucc.nau.edu/~rab39/RAB%20Resume.doc
...2843. DOB: 10-10-48 New Britain, CT 06050-4010. F: (860) 832-3753. SSN: 461-84-... H: (203) 740-7255: (203) 561-8674. Education. Ph.D. www.math.ccsu.edu/vaden-goad/resume.htm
Scot Patrick Lytle. Home: (301)-249-5330 2116 Blaz Court School: (410)-455-1662 Upper Marlboro, MD 20772 SSN: 578-90-... userpages.umbc.edu/~slytle1/resume.html

Figure 3. Sample on-line resumes (unformatted) that include SSNs, Two of the resumes include dates of birth. All three include address and phone number. SSNs have been truncated for this writing but were fully available.

Based on exhaustive manual inspection and SSNwatch, the following results were found. Of the 150 resumes in DBA, 140 (or 93%) had complete 9-digit SSNs. 10 resumes had partial, invalid, or some other country's SSN. All of the 75 resumes in DBB had 9 digit SSNs.

Experiment 2: Extracting Sensitive Information

Applying Detectors to DBA, and then manually inspecting each resume, provided the following results. All email addresses (113 of 113 or 100%) were found. The '@' and dot (.) notation worked well. All dates of birth ("DOB") were found (110 of 110 or 100%), but some dates, which were not dates of birth were incorrectly reported as such; this happened in 20 cases (but only 7 where the proper DOB was not also found). SSN results were reported above.

In terms of combinations: 104 (or 69%) resumes had {SSN, DOB}; 105 (or 70%) had {SSN, email}, and 76 (or 51%) had {SSN, DOB, email}.

Experiment 3: Behavioral Impact

A single email message was sent to each of the 105 people in DBA having {SSN, email} alerting them to the risk. A year later, 102 (or 68% of all of DBA) no longer had the information available. In DBB, 46 were notified, and within a month, 42 (or 55% of all of DBB) no longer had the information publicly available.

Discussion

Entering {"resume" "vitae"} on Google in 2003 yielded tens of thousands of web pages. About 8% of these were real resumes listing actual Social Security numbers. A malicious person can manually review these resumes to find those containing sufficient information to fraudulently complete a credit card application, but someone with a little programming background can harvest this information by the hundreds. To help combat this vulnerability, Identity Angel attempts to contact the subjects of these resumes beforehand and encourage them to remove the sensitive information before it is exploited. There are many activities in which a person cannot help but risk being a victim of identity theft, but a person's on-line resume or vita does not have to be one of them.

Acknowledgements

Thanks to Elisa Bertino for the opportunity to present this paper, and to Sylvia Barrett, Kishore Madhava, Yea-Wen Yang and Nicholas Lynn for data assistance. This work was conducted by

volunteer service from the Data Privacy Laboratory in the School of Computer Science at Carnegie Mellon University. Government and/or business sponsors are welcomed.

References

1. "Angel Protects Those Who Might be Targets for ID Theft," *CBS News-Denver*, October 20, 2005. cbs4denver.com/video/?id=10164@kcnc.dayport.com
2. Career Builder, January 6, 2006. www.CareerBuilder.com
3. Identity Angel Project, Data Privacy Lab, Carnegie Mellon University. privacy.cs.cmu.edu/dataprivacy/projects/idangel/index.html
4. United States Federal Trade Commission, *Report on Identity Theft, Victim Complaint Data: Figures and Trends January-December 2004*, Federal Printing Office, Washington, DC: 2005.
5. Szwak, D. Understanding Credit Cards, Credit Reports And Fraud, *Lectric Law Library*. January 2006. www.lectlaw.com
6. Identity Theft Spy. www.identitytheftspy.com
7. Zeller, T. "Waking up to recurring ID nightmares," *New York Times*, January 9, 2006.
8. United States Federal Trade Commission, January 6, 2006. www.consumer.gov/idtheft/con_minimize.htm
9. *Consumer Reports*, July 2000.
10. United States General Accounting Office, *Improved SSN Verification and Exchange of States' Driver Records Would Enhance Identity Verification*, Federal Printing Office, Washington, DC: 2003.
11. *Social Security numbers sold on Web*. Associated Press, 8/2003.
12. Sweeney, L. Finding Lists of People on the Web. *ACM Computers and Society*, 34 (1) April 2004.
13. Sweeney, L. Replacing Personally-Identifying Information in Medical Records, the Scrub System. *Proceedings, Journal of the American Medical Informatics Association*. 1996.
14. Sweeney, L. *SSNwatch Validation Server*, 2004 <privacy.cs.cmu.edu/dataprivacy/projects/ssnwatch/index.html>.