L. Sweeney, *AI Technologies to Defeat Identity Theft Vulnerabilities*, Carnegie Mellon University, School of Computer Science, Data Privacy Lab White Paper Series, LIDAP-WP12. Pittsburgh: 2004.

AI Technologies to Defeat Identity Theft Vulnerabilities

Latanya Sweeney
Data Privacy Laboratory
School of Computer Science
Carnegie Mellon University
latanya@privacy.cs.cmu.edu

Abstract

In 2003, the U.S. General Accounting Office identified SSN vulnerabilities as ripe for exploitation by terrorists, making SSN problems a serious concern to homeland security [1]. This writing proposes two AI technologies used to defeat identity theft vulnerabilities specifically related to Social Security numbers (SSNs). The "credential validation problem," involves matching the person presenting the credential to the subject of the credential. The first tool proposed herein, *SSNwatch*, is used to verify whether a given SSN matches the person presenting the SSN by exploiting semantics (knowledge representation) related to SSNs. The second tool, *Identity Angel*, crawls through Cyberspace, finding people at risk to identity theft (data mining) and notifies appropriate parties. Deploying these tools combats fraud related to financial and identity crimes that threaten the nation's economic prosperity and security. They demonstrate how AI technologies can improve security while simultaneously enhancing the privacy of citizens.

Both of these tools are non-invasive. They work for less sophisticated computer users, who are the most likely to fall victim to these kinds of attacks. They operate within the existing information infrastructure. They do not require companies or agencies to change existing practices. They do not require any new laws. They do not require access to any sensitive information that is not already in the public domain. And, they do not impose overt modifications to routine user behavior in order to operate. Yet, the deployment of these non-invasive tools promises to be an effective guard against some important financial crimes related to the theft or misuse of personal identifiers and homeland security.

Tool #1 SSNwatch Validation Server

The SSNwatch Validation Server uses publicly available information about people and about SSNs to verify whether a given SSN matches the person presenting the SSN, and vice versa. Using publicly available information about SSN encoding and inferences about SSN assignments, the current SSNwatch Validation Server identifies the issuing state, date issued, estimated age range of the recipient, and activity status of an SSN. Uses of the SSNwatch Validation Server spot identity inconsistencies. An example is a 25 year old man presenting an SSN issued to a 50 year old man. Spotting these kinds of inconsistencies can be useful in processing job applications, apartment rentals, insurance claims, medical claims, and student applications, and so on.

Tool #2 Identity Angel

Identity Angel provides an immediate means of reducing identity theft risk by identifying people for whom information, freely available on the Internet, can be combined sufficient to impersonate the person in financial or credentialing transactions. In the example of fraudulently

acquiring a new credit card, the imposter needs the {*name, SSN, address, date of birth, mother's maiden name*} of the subject. Identity Angel automatically identifies people (tens of thousands in the example application) to whom this information is available freely on-line and then notifies the person (or appropriate parties) in cases where this information in part (or whole) can voluntarily be removed. Identity Angel runs automatically, without human intervention, finding vulnerable people and automatically notifying them.

Having technologies such as SSNwatch and Identity Angel deployed provides a provably effective and immediate guard to defeating identity theft vulnerabilities, and they do so non-invasively.

Technical Approach

In 2003, the U.S. General Accounting Office identified SSN vulnerabilities as ripe for exploitation by terrorists, making SSN problems a serious concern to homeland security [1]. These risks are real and imminent, yet poorly understood (until now), thereby thwarting appropriate non-invasive technical remedies (until now). In the United States, SSNs are essential to identifying, recognizing and authenticating people in health, financial, legal, and educational information. The ease at which SSNs can be replicated and their widespread availability currently places thousands of Americans in imminent danger of identity theft. The Federal Trade Commission Report on Identity Theft shows rapid growth in victim complaints received at their clearinghouse [2]. People whose identities are stolen can spend years – and lots of money – cleaning up their credit report. Victims have reportedly lost job opportunities, been refused loans, and even been arrested for crimes they did not commit. What are needed are innovative tools that help thwart these vulnerabilities without requiring extensive changes to the financial and legal infrastructures that rely on related flows of information. The tools presented herein are two such solutions.

Tool #1 SSNwatch Validation Server

A few months ago, an academic version of the SSNwatch Validation server was provided, free on-line at < http://privacy.cs.cmu.edu/dataprivacy/projects/ssnwatch/index.html>. This writing describes uses of SSNwatch by law-enforcement and business personnel who use the current service. SSNwatch provides an effective means to help combat identity theft by matching the person presenting the SSN with inferences about the person's history of residence and age and about the validity of the SSN itself. Validation can be done with all or part of an SSN. Below are some sample runs from SSNwatch using only part of an SSN!

Results for SSN 078 - 05 –

Geography	New York
Date of issuance	Issued before 1993
Year of Birth (5-digit prefix)	64% born 1889 to 1910 98% born 1879 to 1921

If the person presenting the SSN fails to list or acknowledge New York as a prior residence, then it is extremely unlikely that the provided SSN was issued to that person.

If the person presenting the SSN is about age 20, then it is extremely unlikely that the provided SSN was issued to that person.

Results for SSN 221 - 98 –

Geography	Delaware
Date of issuance	Issued after December 2002

Results for SSN 221 - 02 –1023

Geography	Delaware
Date of issuance	Not issued. Not a valid SSN!

Results from the SSNwatch Validation Service as of January 2004 are shown above. SSNs beginning 221-98 have been issued, but not those beginning 221-02! It is extremely unlikely that the SSN on the right was issued to any person.

Results for SSN 078 - 05 –1120

Geography	New York
Date of issuance	Issued before 1993
Year of Birth (6-digit prefix)	66% born 1891 to 1909 98% born 1881 to 1919 100% born 1877 to 1919
Status	Retired number (due to death or termination)

This SSN has been retired from service (due to death or termination). SSNs are not recycled or re-used. If a person presents this number as their SSN, then it is extremely likely that the person has made an error.

Results for SSN 615 - 23 –

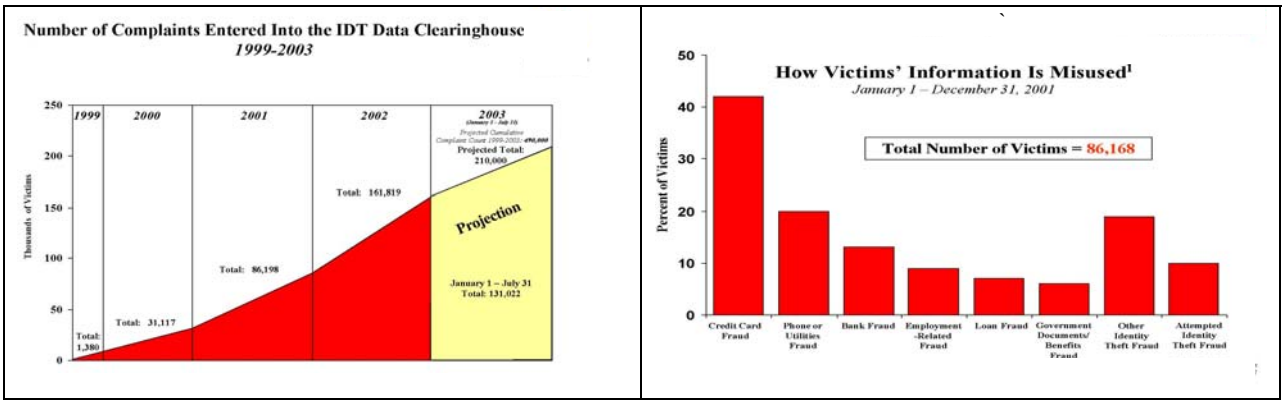
Geography	California
Date of issuance	Issued February 2001

If the person presenting this SSN reports work experience dating back to 1983, then it is extremely likely the person has made an error.

The algorithms within the SSNwatch server use publicly available information about SSN encodings, assignments and issuances to make inferences.

Tool #2 Identity Angel

The Identity Angel identifies people for whom information, freely available on the Internet, can be combined sufficient to impersonate the person in financial or credentialing transactions. It utilizes a set of detection algorithms, inference algorithms and a taxonomy of demographic and financial credentialing information. It consists of a crawler, a parser, and an email service. These components can be focused on particular kinds of information and threats. The example of identity theft for new credit card fraud is provided below as an example application.



Source:Federal Trade Commission Report, Victim Complaint Data (2002)

The Federal Trade Commission Report on Identity Theft shows rapid growth in the number of victim complaints received at their clearinghouse; see Figure above. Nearly half of the complaints involve credit card fraud; see Figure above. Of credit card fraud reports, more than half (or 26% of all thefts) involved new accounts. Incidents among young adults were reported to be high, which makes this group of particular interest. We consider young adults as being more likely to use the Internet, to have school related information about them available on the Internet, and to have resumes and facts about themselves posted on the Internet. They are also likely 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. For this reason, the case of new credit card fraud poses a good example application area for the Identity Angel.

During execution of the initial versions of Identity Angel, thousands of Americans were found at risk. Below are some very simple examples based on resumes that the Identity Angel found on-line. These include all the relevant information. When Americans (sampled from those found) were notified by email, they promptly modified the resumes or removed them altogether. Emails of gratitude were received in response (Sweeney, 2004).

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. 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**-… OBJECTIVE. ... userpages.umbc.edu/~slytle1/resume.html

Sample on-line resumes 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.

References

[1] U.S. Federal Trade Commission, *Report on Identity Theft, Victim Complaint Data: Figures and Trends January-December 2001*, Federal Printing Office, Washington, DC: 2002.

[2] United States Social Security Administration, *Historical highlights of the Social Security Number*, 2004. Available at www.ssa.gov/history/hfaq.html.

[3] U.S. General Accounting Office, *Improved SSN Verification and Exchange of States' Driver Records Would Enhance Identity Verification*, Washington, DC: 2003.

[4] Sweeney, L. Social Security Number Watch: the SSNwatch Validation Server. Carnegie Mellon University, School of Computer Science, Laboratory for International Data Privacy. Pittsburgh: 2003. Operational on-line at http://privacy.cs.cmu.edu/dataprivacy/projects/ssnwatch/index.html.

[5] United States General Accounting Office. *Social Security Numbers: Private Sector Entities Routinely Obtain and Use SSNs, and Laws Limit the Disclosure of This Information*. GAO-04-11, Washington, DC: January 22, 2004.

[6] Sweeney, L. The Identifiability of Data. Carnegie Mellon University, School of Computer Science, Laboratory for International Data Privacy, LIDAP-3, 2001. Forthcoming book.

[7] Sweeney, L. "Finding Lists of People on the Web," *Computers and Society*, ACM, 32 (8) March 2004. Available at privacy.cs.cmu.edu/dataprivacy/projects/rosterfinder/index.html