Social security numbers can be predicted with public information, researchers find

July 6 2009

Carnegie Mellon University researchers have shown that public information readily gleaned from governmental sources, commercial data bases, or online social networks can be used to routinely predict most — and sometimes all — of an individual's nine-digit Social Security number.

Project lead Alessandro Acquisti, associate professor of information technology and public policy at Carnegie Mellon's H. John Heinz III College, and Ralph Gross, a post-doctoral researcher at the Heinz College, have found that an individual's date and state of birth are sufficient to guess his or her Social Security number with great accuracy. The study findings will appear this week in the online Early Edition of

the *Proceedings of the National Academy of Science*, and will be presented on July 29 at the BlackHat 2009 information security conference in Las Vegas. Additional information about the study and some of the issues it raises is available at http://www.ssnstudy.org.

The predictability of Social Security numbers is an unexpected consequence of seemingly unrelated policies and technological developments that, in combination, make Social Security numbers obsolete for authentication purposes, according to Acquisti and Gross. Because many businesses use Social Security numbers as passwords or for other forms of authentication — a use not anticipated when Social Security was devised in the 1930s — the predictability of the numbers increases the risk of identity theft. ID theft cost Americans almost $50 billion in 2007 alone. The Social Security Administration could mitigate this vulnerability by assigning numbers to people based on a randomized scheme, but ultimately an alternative means of authenticating identities must be adopted, the authors conclude.

"In a world of wired consumers, it is possible to combine information from multiple sources to infer data that is more personal and sensitive than any single piece of original information alone," said Acquisti, a researcher in the Carnegie Mellon CyLab. Information that once was useful to make public may now be too available. An example is the Social Security Administration's Death Master File, a public database with Social Security numbers, dates of birth and death, and states of birth for every deceased beneficiary. Its purpose is to prevent impostors from assuming the Social Security numbers of deceased people. But Acquisti and Gross found that analyzing the death file enabled them to detect statistical patterns that would help them predict Social Security numbers of the living.

These statistical patterns can help narrow guesses of an individual's Social Security number, when combined with that person's date and state

of birth. Birth information can be obtained from various sources, including commercial databases, public records (such as voter registration lists) and the millions of profiles that people publish about themselves on social networks, personal Web sites and blogs.

The statistical patterns and the birth information can be used to predict Social Security numbers because the Social Security Administration's methods for assigning numbers, based in part on geography, are well-known. For most individuals born nationwide since 1989, Social Security numbers are assigned shortly after birth, making those numbers easier to predict.

Acquisti and Gross tested their prediction method using records from the Death Master File of people who died between 1973 and 2003. They could identify in a single attempt the first five digits for 44 percent of deceased individuals who were born after 1988 and for 7 percent of those born between 1973 and 1988. They were able to identify all nine digits for 8.5 percent of those individuals born after 1988 in fewer than 1,000 attempts. Their accuracy was considerably higher for smaller states and recent years of birth: for instance, they needed 10 or fewer attempts to predict all nine digits for one out of 20 SSNs issued in Delaware in 1996. Sensitive details of the prediction strategy were omitted from the article.

"If you can successfully identify all nine digits of an SSN in fewer than 10, 100 or even 1,000 attempts, that Social Security number is no more secure than a three-digit PIN," the authors noted.

When the researchers tested their method using birth dates and hometowns that students had self-reported on popular social networking sites, the results were almost as good despite the inaccuracies typical of social network data. Enrollment records were used to confirm the accuracy of the predictions, though the researchers did not receive

confirmation of any individual Social Security number, but only aggregate measures of accuracy.

"Dramatically reducing the range of values wherein an individual's Social Security number is likely to fall makes identity theft easier," Gross said. A fraudster who knows just the first five digits of an individual's number might use a phishing email to trick the person into revealing the last four digits. Or, a fraudster could use networks of compromised computers, or "botnets," to repeatedly apply for credit cards in a person's name until hitting the correct nine-digit sequence.

Future Social Security numbers could be made more secure by switching to a randomized assignment scheme, but protecting people who already have been issued numbers is harder, the researchers said. Given the ease with which Social Security numbers can be predicted — particularly the first five digits and particularly for the millions of Americans born since 1988 — legislative and policy initiatives aimed at removing the numbers from public exposure, or redacting their first five digits, may be well-meaning but misguided, Acquisti added.

"Given the inherent vulnerability of Social Security numbers, it is time to stop using them for verifying identities and redirect our efforts toward implementing secure, privacy-preserving authentication methods," Acquisti said. Methods to consider include two-factor authentication, similar to the PIN number/card combinations used for bank accounts, and digital certificates.

Source: Carnegie Mellon University (news : web)