Guilty Until Proven Innocent?

I n April 2003, the US Justice Department administratively discharged the FBI of its statutory duty to ensure the accuracy and completeness of the National Crime Information Center (NCIC) database. This enormous database contains over 39 million criminal records and information

BRUCE SCHNEIER
Counterpane Internet Security

on wanted persons, missing persons, and gang members, as well as information about stolen cars and boats. More than 80,000 law enforcement agencies have access to this database. On average, the database processes 2.8 million transactions each day.

The US Privacy Act of 1974 requires the FBI to make reasonable efforts to ensure the database records' accuracy. However, in April, the Justice Department exempted the system from the law's accuracy requirements.

This isn't just bad social practice, it's bad security. A database with more errors is much less useful than a database with fewer errors, and an error-filled security database is much more likely to target innocents than it is to let the guilty go free.

To see this, let's walk through some examples. Assume a simple database—names and a single code indicating "innocent" or "guilty." When a policeman encounters someone, he looks up that person in the database, and then arrests him if the database says "guilty."

Example one: Assume the database is 100 percent accurate. If that were the case, there wouldn't be any false arrests because of bad data. It would work perfectly.

Example two: Assume a 0.0001– percent error rate: one error in a million. (An error is defined as a person having an "innocent" code when guilty, or a "guilty" code when innocent.) Furthermore, assume that one in 10,000 people are guilty. In this case, for every 100 guilty people the database correctly identified, it would mistakenly identify one innocent person as guilty (because of an error). And the number of guilty people erroneously listed as innocent would be tiny: one in a million.

Example three: Assume a 1 percent error rate—one in a hundred—and the same one-in-10,000 ratio of guilty people. The results would be very different. For every 100 guilty people the database correctly identified, it would mistakenly identify 10,000 innocent people as guilty. The number of guilty people erroneously listed as innocent would be larger, but still very small: one in 100.

The differences between examples two and three are striking. In example two, one person is erroneously arrested for every 100 people correctly arrested. In example three, one person is correctly arrested for every 100 people erroneously arrested. The increase in error rate makes the database all but
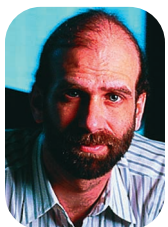
useless as a system for figuring out whom to arrest. And this is despite the fact that, in both cases, almost no guilty people get away because of the database error.

The reason for this phenomenon is that the number of guilty people is a very small percentage of the population. If one in 10 people were guilty, then a 0.0001 percent error rate would mistakenly arrest one innocent for every 100,000 guilty, and a 1 percent error rate would arrest approximately one innocent for every guilty. And if the number of guilty people were even less than one in 10,000, then the problem of arresting innocents would be magnified even more because the database has more errors.

Now, these are simple examples, but the NCIC database has far more complex data and tries to make more complex correlations. And I am assuming that the error rate for false positives is the same as the error rate for false negatives, and that there aren't any data dependencies that complicate the analysis. But even with these complications, the problems are still the same. Because there are so few terrorists (for example) among the general population, an error-filled database is far more likely to identify innocent people as terrorists than it is to catch actual terrorists.

Too far-fetched, you say? Well, this kind of thing is already happening. There are 13 million people on the FBI's terrorist watch list. That's ridiculous—it's simply inconceivable that a number of people equal to 4.5 percent of the US population are terrorists. There are far more innocents on that

Further resources

Sign this online petition (www.petitiononline.com/ncic/petition.html) to require accuracy in the National Crime Information Center (NCIC) database.

Other articles that describe what happens to innocents on the government's "no fly" list:
Due Process Vanishes in Thin Air," www.wired.com/news/privacy/0,1848,58386,00.html
"Please Step to the Side, Sir," www.salon.com/tech/feature/2003/04/10/capps/index_np.html

The risks of being named "David Nelson":
www.oregonlive.com/news/oregonian/margie_boule/index.ssf?/base/living/1051877124142830.xml

General risks of large law-enforcement databases:
"Putting the Blinders Back on Big Brother," www.securityfocus.com/news/3482
"Why We May Never Regain the Liberties That We've Lost," www.siliconvalley.com/mld/siliconvalley/5571471.htm

list than there are guilty people not on that list. And these innocents are regularly harassed by police; one recent article chronicled the problems anybody named "David Nelson" has boarding an airplane. But even without these problems, any watch list with 13 million people is basically useless. How many resources can anyone afford to spend watching about one-twentieth of the population, anyway?

That 13-million-person list feels a whole like CYA on the FBI's part. Adding someone to the list probably has no cost and, in fact, may be one criterion for how an FBI employee's performance is evaluated. Removing someone from the list probably takes considerable courage because someone is going to have to take the fall when "the warnings were ignored" and "they failed to connect the dots." What's the incentive to make life easier on all of those innocent Doug Nelsons?

Best to leave that risky stuff to other people, and to keep innocent people on the list forever.

Many argue that this kind of thing is bad social policy. I argue that it is bad security as well. ☐

Bruce Schneier is chief technology officer of Counterpane Internet Security and the author of over seven books on cryptography and computer security. You can subscribe to his free email newsletter, Crypto-Gram, at www.counterpane. com/crypto-gram.html, or contact him directly at schneier@counterpane.com.