Government Increasingly Turning to Data Mining

Peek Into Private Lives May Help in Hunt for Terrorists

By Arshad Mohammed and Sara Kehaulani Goo
Washington Post Staff Writers
Thursday, June 15, 2006; D03

The Pentagon pays a private company to compile data on teenagers it can recruit to the military. The Homeland Security Department buys consumer information to help screen people at borders and detect immigration fraud.

As federal agencies delve into the vast commercial market for consumer information, such as buying habits and financial records, they are tapping into data that would be difficult for the government to accumulate but that has become a booming business for private companies.

Industry executives, analysts and watchdog groups say the federal government has significantly increased what it spends to buy personal data from the private sector, along with the software to make sense of it, since the Sept. 11, 2001, attacks. They expect the sums to keep rising far into the future.

Privacy advocates say the practice exposes ordinary people to ever more scrutiny by authorities while skirting legal protections designed to limit the government's collection and use of personal data.

Critics acknowledge that such data can be vital to law enforcement or intelligence investigations of specific targets but question the usefulness of "data-mining" software that combs huge amounts of information in the hopes of finding links and patterns that might pick someone out as suspicious.

Dialing for Data
Recent reports about the National Security Agency's effort to acquire phone call records highlights the government's growing interest in the technique.

"The only question we would have is at what rate would the demand be increasing," Wayne Johnson, a financial analyst at Raymond James Financial Inc., said of the government's interest in buying commercial data and related software.

It is difficult to pinpoint the number of such contracts because many of them are classified, experts said. At the federal level, 52 government agencies had launched, or planned to begin, at least 199 data-mining projects as far back as 2004, according to a Government Accountability Office study. Most of the programs are used to improve services, such as detecting Medicare fraud and improving customer relations. But a growing number of agencies are exploring the technology to analyze intelligence and assist in the hunt for terrorists.

Another GAO report released in April found that of $30 million spent by four government agencies last year on services from data-crunching companies, 91 percent was for law enforcement or counterterrorism.

The hope is that the technology can help to discern and thwart threats just as businesses have used it for years to predict consumer behavior on buying cosmetics or repaying mortgages, for example.

Companies keep an increasing amount of data about everyone -- tracking their buying, travel, bank transactions and bill-paying habits. Data mining uses mathematical formulas to look for patterns in those behaviors. The results could enable the grocery store to send out targeted coupons, or, in theory, help the government decide how likely it may be that someone is linked to terrorist groups.

The Education Department's Project Strikeback uses mining methods to compare its databases with the FBI and verify identities. The Defense Department's Verity K2 Enterprise program searches data from the intelligence community and Internet searches to identify foreign terrorists or U.S. citizens connected to terrorists. A Navy program analyzes data to try to predict where it might find small weapons of mass destruction and narcotics smuggling in the shipping industry.

Cogito Inc. sells software to the National Security Agency that the company says can find patterns in massive amounts of data, such as lists of telephone calling records.
The Utah-based company does not know how the super-secret agency is using the software, but it does know that data-mining technology once used primarily by commercial clients is now doing booming business with the federal government.

"What was surprising . . . was how aggressive and hot the intelligence and security market is for this," said William Donahoo, vice president of product management and marketing at Cogito. More than half of Cogito's clients are in the fields of intelligence, security and public safety, he said.

Donahoo said he believed the NSA could use the software to reveal patterns about how people deal with one another just from their calling records.

"There are gatekeepers and bridges and collaborators and leaders that could be identified just by the nature of the communications among the groups," he said. "You do not have to know the content of the conversation to identify this."

**False Positives**

Critics argue that catching terrorists is far different from predicting consumer purchases or preventing credit card fraud, saying that data mining is likely to provide so many false leads that its use is a waste of time and money.

"What you don't want is to get into the Kevin Bacon game, which is to say that you show that everybody is six degrees of separation from a terrorist," said James B. Steinberg, dean of the Lyndon B. Johnson School of Public Affairs at the University of Texas. Steinberg was a deputy national security adviser in the Clinton administration.

"Out of pure resource allocation, it is so unlikely to provide something useful and so likely to provide dead ends and false leads that you are going to spend an enormous amount of resources on things that don't pan out," he said. "Before you start searching haystacks for needles, you've got to have some reason to believe that the needles are there."

The federal government's most public experiment with data mining since the terrorist attacks in 2001 failed to get off the ground, after the Homeland Security Department spent $200 million on it and the technology failed to prove what it set out to do, according to several former U.S. officials familiar with the program.
The system, originally called CAPPS II, sought to comb airline passenger records and verify information that fliers provided about themselves with information provided by companies that aggregate data about consumers. The problem, according to several officials who worked closely on the program but declined to speak publicly about it, was that the information about consumers was never proved to be effective in evaluating the risk posed by an airline passenger.

At first, officials sought to identify passengers who were not "deeply rooted" in a community and, for example, moved often and did not have an established credit history. But the system always ended up scoring too many people as "risky" who really posed little threat.

"I am just not prepared to say that because someone can't get a mortgage, they are a terrorist threat to an airplane," said a former official, who spoke on condition of anonymity because he was not authorized to speak for the program. "These data aggregator products are used today in the financial world to identify certain things, and they're not designed to identify potential terrorist threats."

The former official said that the program still shows some promise, but that it needs more testing and should be considered only one tool of many to protect the nation's air travel system.

Despite privacy concerns about CAPPS II that were raised by groups such as the American Civil Liberties Union, top U.S. officials continue to express faith that the technology will prove to be useful for national security purposes.

"This issue of using data to ferret out evildoers, many administration officials believe very firmly this is the way we should be going and that the barriers there should be overcome because it will result in a greater good," said another former official, who spoke on condition of anonymity. "It's a philosophy that if you have nothing to hide, why do you care if I know what movies you rent? Who you are talking to? If you live a godly life, a perfect life, you don't have worry about 100 percent disclosure."

**Security vs. Privacy**

Even critics say data mining can be effective in targeted circumstances, such as gathering information about known suspects. But the government's wide interest in the
technology disturbs privacy advocates, who say the vast commercial data industry provides a ready-made window into private lives that the government would be unable to legally assemble on its own.

Jim Dempsey, policy director at the Center for Democracy and Technology, said risks include errors in the data, drawing incorrect inferences from the information and "the chilling effect that comes when a citizenry feels itself under scrutiny."

But since the 2001 terror attacks, a slim majority of the American public has favored protecting security over preserving civil liberties, according to opinion pollsters.

"The public is willing to bend the rules a little bit with respect to privacy," said Andrew Kohut, director of the Pew Research Center, adding that Americans showed similar tendencies during the "red scares" after World War I and World War II. "They are giving the government the benefit of the doubt in large part because they are concerned about terrorism."

© 2006 The Washington Post Company