

Advertisement

The Best-Run Companies Are Running To Dollar.com

Up To A 3% REBATE

INTRODUCING **DOLLAR 4Business**

Chrysler 300

DOLLAR RENT A CAR

[Print This Article](#)
[E-Mail This Article](#)
[Subscribe to The Post](#)

Advertisement



QUICK QUOTES

Enter Symbol go
[Tables](#) | [Portfolio](#) | [Index](#)

MOST VIEWED ARTICLES

Technology On the Site

Updated 4:00 p.m. ET

- [Candidate Profiles Pop Up on Facebook](#)
- [Data Mining Still Needs a Clue to Be Effective](#)
- [Be Prepared For ID Theft](#)
- [ING Financial to Notify Potential Identity Theft Victims](#)
- [Dozens Illegally Used Calif. Woman's ID](#)

RSS NEWS FEEDS

[Top News](#)
[Technology](#)
What is RSS? | [All RSS Feeds](#)

E-MAIL NEWSLETTERS

View a Sample and Sign Up
[TechNews Daily Report](#)
[Personal Finance](#)
[Personal Tech](#)
[Manage Your Newsletters](#)

Data Mining Still Needs a Clue to Be Effective

By [Guy Gugliotta](#)
Washington Post Staff Writer
Monday, June 19, 2006; Page A08

In the two decades or so since software scientists began "mining" computerized databases for information they were never designed to yield, the sophistication of their techniques has increased dramatically.

And although marketing companies today -- especially with the advent of the Internet -- can routinely predict who you will vote for, where you will eat dinner and, most of all, what products you will buy, experts say it is far less clear whether security agencies can sift mounds of data to track down terrorist networks -- unless they start with a useful lead.

More than a month has passed since USA Today reported that the National Security Agency had amassed a database of 2 trillion telephone calls since 2001, ostensibly as a tool to hunt al-Qaeda operatives.

Details of the NSA's activities remain unclear, but data mining experts say they are puzzled about how the information might be used. It would work best, they say, when investigators can trace telephone numbers of known suspects and build a web of contacts, in much the same way police use phone records to track drug traffickers.

But to discern suspicious call patterns from lists of dialed numbers, they will have to dig past the raw data into callers' identities, and, in the vast majority of cases, will find they have simply tapped into networks of law-abiding people involved in daily routines. This



(By Patterson Clark - The Washington Post; File and iStockPhoto Images)

PRIVACY

An entire industry has mushroomed during the past decade because of the ability of companies to gather and make sense of public records, criminal histories and other electronic details. What are they doing with it?

- [Data Mining Still Needs a Clue to Be Effective](#)
- [STOLEN PERSONAL DATA: ING Financial to Notify Potential Identity Theft Victims](#)
- [Dozens Illegally Used Calif. Woman's ID](#)
- [Be Prepared For ID Theft](#)
- [Research Explores Data Mining, Privacy](#)

[More News](#)

FROM SLATE

[Has Your Identity Been Stolen?](#)
Here's what you do if

TOP JOBS FROM LOCAL EMPLOYERS

- [Customer Service Rep](#) / HILLDRUP
 - [Clinical Liason - Nursing](#) / THE SPECIALTY HOSPITAL OF WASHINGTON
 - [Developer / Programmer Analyst](#) / ASI CORPORATION
 - [Applications Developer](#) / ASI CORPORATION
 - [Corporate Info. Tech. Director](#) / ASI CORPORATION
- [All Top Jobs](#)

FEATURED ADVERTISER LINKS

- [Refinance Rates As Low As 2.9% - FREE QUOTES!](#)
- [\\$200,000 loan for \\$917/mo. Bad credit OK](#)
- [For College Graduates, Reduce Your Student Loans](#)
- [Barbra Streisand Tickets, Tim McGraw Tickets w/ Faith Hill, NASCAR Tickets, Eric Clapton Tickets](#)
- [ReNu, Celexa, Car Accident, Vioxx, Bextra](#)
- [\\$145,000 Mortgage for Under \\$484/Month!](#)
- [Promotional Pens, T-shirts, Custom Hats](#)
- [Cool Gadgets, Great Deals, Visit CircuitCity.com](#)
- [Get Free Checking with direct deposit](#)

approach, several experts said, raises privacy questions even as it wastes time.

"When they look at a map of phone numbers, they have no idea what's going on," said Valdis Krebs, an expert in deriving "social networks" from databases. "It might not be a bad person you find; it may be that the soccer team and the softball team are calling the same pizza parlor."

Even though they have no firsthand knowledge of the NSA's program, Krebs and other data mining experts, some of whom requested anonymity because of the sensitivity of their work, agreed to discuss how such a mountain of information might be used.

The only clue offered by the Bush administration came when former NSA director Michael V. Hayden told the Senate during his confirmation hearing as CIA director that analysts used "targeting" to enter the database:

"Every targeting is documented," he told the Senate. "There is a literal target folder that explains the rationale and the answers . . . as to why this particular number we believe to be associated with the enemy."

His testimony suggested that NSA analysts are searching the database for telephone numbers of known suspects. These calls can be traced to other numbers, establishing a communications pattern and providing leads to other suspects.

"You start with a few bad guys, and you have to know where to look," said IBM distinguished engineer Jeff Jonas, a specialist in using software to track undesirables in the gambling industry. "Phone records can give you that."

But Jonas and others noted that tracking suspects' telephone records was a staple of good police work long before electronic search engines made it feasible to scan trillions of calls. And even now, just as 50 years ago, the search moves more quickly and effectively if the searchers can rule out useless information.

"Before you search the world, make sure you're using your local resources properly," Jonas said. "And make sure your search complies with some process. Remember, the 'no-fly list' also includes people who are abusive to other people on planes."

This "number first" approach reflects traditional "deductive" law enforcement techniques, in which investigators begin with a fact -- a telephone number or a corpse -- and work backward to find the details needed to build a case.

In theory, experts said, modern computers also allow investigators to move in the other direction -- to identify telephone use patterns in a call record database and work forward until a suspect's name


Here's what you do if you find out your identity might have been stolen.

WHO'S BLOGGING?

Read what bloggers are saying about this article.

[Silly Humans](#)

[Full List of Blogs \(1 links\) »](#)

Most Blogged About Articles

[On washingtonpost.com](#) | [On the web](#)

POWERED BY **Technorati**

SAVE & SHARE

[Tag This Article](#)

▶ [Saving options](#)

■ Powered by **Del.icio.us**

drops out. But these "inductive" techniques are far more difficult and less reliable, the experts said, because it is virtually impossible to distinguish a web of suspicious linkages from a harmless one in an immense, unedited bundle of numbers.

"I'm sure the NSA is excellent at finding patterns and motifs in the data, but what do they mean?" Krebs asked. "Unless you start getting more information on the patterns, you're not going to be able to interpret them at all. Patterns alone won't tell you whether someone's good or evil."

What investigators get instead are lots of "false positives": "It's like if you Googled [Secretary of State] Condoleezza Rice and al-Qaeda. You'd get millions of hits, but they wouldn't be meaningful," one industry expert said. To milk whatever databases NSA may possess, "what they have today are glorified search engines."

Still, several experts suggested ways that pattern-making could be useful. "Suppose you looked at calls between two geographical points, and you could see what kind of pattern ordinary people had," said Olvi L. Mangasarian, co-director of the University of Wisconsin's Data Mining Institute.

"Then you compare it to another pattern of calls that you know" are suspicious and try to develop a "classifier" -- a software tool -- to distinguish between them, he said. "It would be difficult -- but it would be doable."

Another expert suggested comparing call patterns at different times. "Suppose that before the Super Bowl, there's a quickening of traffic in the phone system of the city that has the game," the source said. "Then suppose you see a similar pattern, except there's no Super Bowl. Then what's going on?"

In all cases, however, the technicians operating the system would have to be expert, even visionary, to avoid false positives and to root out meaningful patterns from the background "noise" of billions of innocent communications.

One authority compared the ideal analyst to an expert lie detector operator or to the sonar man who can identify a submarine's nationality just by listening to its screws turning. "Computers can jump to conclusions just like humans," this expert said. "To make the correct inference requires deep, intellectual thinking; these systems are significantly less reliable than lie detector tests."

Still, even the best technicians are going to find themselves searching multiple blind allies in navigating a mega-database such as telephone logs, the experts said, so much so that the time needed to clear false positives may outweigh the odds of finding a terrorist.

"Even if one out of 10 searches is a hit, the technique is useful," one expert said. "But one out of 1,000 or one in 1 million?" In these cases, experts suggest, maybe the technician would be more cost-effective by searching something besides phone logs.

MORE TECHNOLOGY ARTICLES

Most Viewed Technology Articles

- [Candidate Profiles Pop Up on Facebook](#)
- [Data Mining Still Needs a Clue to Be Effective](#)
- [Be Prepared For ID Theft](#)
- [ING Financial to Notifv Potential Identity Theft Victims](#)
- [Dozens Illegally Used Calif. Woman's ID](#)

» [Top 35 Most Viewed](#)

Editors' Picks: Technology

- [Data Mining Still Needs a Clue to Be Effective](#)
- [Potential Identity Theft Victims to Be Notified](#)
- [Candidate Profiles Pop Up on Facebook](#)
- [Artificial Intelligence](#)
- [Verizon Accuses Vonage of Infringement](#)

» [More in the Technology Section](#)

[Print This Article](#)

[E-Mail This Article](#)

[Permission to Republish](#)

© 2006 The Washington Post Company

Ads by Google

[Bush vs Clinton](#)

Who would you trust? Vote now to see results
[popular-survey.com](#)

[Coffee Exposed](#)

A shocking secret coffee co's don't want you to know
[www.coffeefool.com](#)

[George W. Bush](#)

Should He Step Down As President? Vote Now To See Survey Results!
[www.popularq.com](#)

SEARCH:  News Web results by 

[NEWS](#) | [OPINIONS](#) | [SPORTS](#) | [ARTS & LIVING](#) | [Discussions](#) | [Photos & Video](#) | [City Guide](#) | [CLASSIFIEDS](#) | [JOBS](#) | [CARS](#) | [REAL ESTATE](#)

[washingtonpost.com: Help](#) | [Contact Us](#) | [About Us](#) | [Advertise With Us](#) | [Site Index](#) | [Site Map](#) | [Make Us Your Homepage](#) | [mywashingtonpost.com](#) | [Work at washingtonpost.com](#)
[The Washington Post: Subscribe](#) | [Subscriber Services](#) | [Advertise](#) | [Electronic Edition](#) | [Online Photo Store](#) | [The Washington Post Store](#) | [About The Post](#)
[The Washington Post Company: Information and Other Post Co. Websites](#)

© Copyright 1996-2006 The Washington Post Company | [User Agreement and Privacy Policy](#) | [Rights and Permissions](#)