

# Chicago Tribune

## Right questions key to data mining Finding phone links possible but difficult

**By Jon Van**

Tribune staff reporter

*Published May 12, 2006*

Connecting the dots is difficult, but for homeland security agents, the real trick is figuring out where the dots are and which ones need connecting.

That analogy may be at the core of the federal government's interest in keeping tabs on telephone calls Americans make to each other every day. Government agents reportedly hope that computers can sift through the mountains of phone data to extract nuggets of information revealing terrorist plotters.

Only within the past decade has a subset of computer science called link mining even become available to attempt such a daunting task, though some researchers believe that even the most powerful computers will never deliver the answers that the government seeks.

Congressional leaders were demanding answers from the Bush administration Thursday about a specific type of connecting-the-dots activity: whether the secretive National Security Agency had collected extensive phone call records, and whether the privacy rights of individuals had been violated.

Behind those questions is the arcane science of using superpowerful computers to mine data of all types for information.

"It's a massive data problem, but you can do it," said Kris Hammond, a Northwestern University professor of electrical engineering and computer science. "If it were impossible to get specific answers to specific questions from a huge database, Google couldn't exist."

Specific questions helpful

The likelihood of success, Hammond said, is higher if agents have specific questions, such as hypothetically what mobile phones in Washington, D.C., made calls to Tehran during a given period, and whether calls were made from those phones to San Francisco during another period.

But if officials don't know what they're looking for, they can't expect a data mining program to connect all the dots.

"If you approach the data without specific questions and just look for patterns, you can

# Chicago Tribune

find hundreds of millions of patterns," Hammond said.

Despite advances in artificial intelligence, computers aren't like human detectives who can make inferences and shift assumptions on the fly, said Yali Amit, a University of Chicago professor of statistics and computer science.

Government agents may not understand this, he said.

"They have records from millions of innocent people and perhaps a few thousand terrorists who might make phone calls," said Amit. "The size of the data set of interest--the terrorists--is too small. You get reliability rates that make the whole endeavor pretty ridiculous."

The White House hasn't confirmed the NSA program, but in December, an official of DARPA, a Defense Department agency that funds advanced research, published a paper in an academic journal that suggests an ambitious role for link mining.

"Metaphorically, link mining offers the potential not only for connecting the dots, but for determining which dots to connect, a far more difficult task," wrote Ted Senator, who stipulated he was expressing his own views, not those of DARPA or the government.

The science of connecting the dots began decades ago when sociologists began studying social networks, charting people's connections with each other by hand, said Karrie Karahalios, an assistant professor of computer science at the University of Illinois at Urbana-Champaign. They looked at things such as how gossip spread through a community.

With instant messaging, buddy lists and other Internet-related phenomena, the attention of computer research has turned to social networking.

"Previously, we only analyzed a person in the context of themselves, but now we're looking at links between that person and others," said Paul Bradley, data mining principal for Chicago-based Apollo Data Technologies LLC.

Who calls whom?

The NSA presumably is looking for patterns among the millions of phone calls Americans place to one another daily. With A calling B and B calling C, there might be some interest if C later calls A. Such data tied to phone numbers could be combined with census data, driver's license lists and other databases to produce useful information.

"They might look at how frequently, how recently the calls were made and where they were geographically," said Bradley.

When commercial firms do link mining, their goals are usually well-defined.

# Chicago Tribune

"A portal company like Yahoo, Google, or AOL might look at patterns among people using their e-mail services and instant messaging," said Jeff Kaplan, client services principal for Apollo Data.

The federal government agents may use the same basic software as private enterprise, but it has more resources.

"Unless you have some other information about a person beyond just calling patterns, there's not that much you can say," said Bradley. "But there are a lot of other databases out there that provide personal information . . . and the government has access to databases that people in the private sector don't even know exist."

-----

[jvan@tribune.com](mailto:jvan@tribune.com)