



The Star-Ledger

Tragedy spurs renewed interest in mining Internet to spot killers

Monday, April 23, 2007

BY KEVIN COUGHLIN
Star-Ledger Staff

Can Internet search technology identify psychopaths before they commit atrocities like last week's Virginia Tech massacre?

The government already is pursuing a range of controversial "data-mining" projects, meant to scour Web sites and documents for subtle patterns and associations that might flush out terrorists.

At least one expert thinks these same automated methods could thwart the next Seung-Hui Cho, or snag sexual predators, by trawling blogs, audio and video on Web sites such as MySpace, Facebook and YouTube for hints of trouble.

"It's not inconceivable to try and do that," said Rohini Srihari, a computer scientist at the State University of New York at Buffalo. "Are we there yet? Probably no. But does the technology exist and is it feasible? Yes. And I think we have to, for the safety of people."

Srihari is an expert on teasing related tidbits from seemingly disparate documents. With funding from the National Science Foundation and the Federal Aviation Administration, she has been testing mathematical formulas on data compiled by the 9/11 commission to see if they could have helped prevent the attacks of Sept. 11, 2001.

She also is trying to predict roadside bombings in Iraq, with backing from the Air Force, by sifting through military field reports and other intelligence data to pinpoint important details that may have been overlooked.

For years companies have used data mining to parse financial and medical data; it's helped the government find fraud in aid programs to Hurricane Katrina victims.

Privacy advocates warn the government could compile dossiers on millions of Americans under the guise of hunting potential killers. An FBI system for filtering e-mail traffic, ominously dubbed Carnivore, was mothballed several years ago after its disclosure produced a public backlash.

And Congress cut public funding for a Pentagon program called Total Information Awareness (TIA), started in 2002, amid protests against Big Brother snooping. However, classified research continues, according to reporting by the National Journal magazine and the Washington Post.

When it comes to large-scale data mining, "the cost to law-abiding citizens is way too high given the remote possibility of benefits," said Jim Harper of the Cato Institute, a libertarian think tank. But the 9/11 commission identified an urgent need for faster ways to "connect the dots" in mountains of intelligence data.

By 2004, some 52 agencies were mining data, or planning to do so, according to the Government Accountability Office. The Department of Homeland Security has nine such programs up and running, with three more in the works, the GAO reported last summer.

One of those tools is dubbed ADVISE for Analysis Dissemination, Visualization, Insight and Semantic Enhancement. It's intended for crunching databases, e-mails, reports and news stories for patterns and for relationships among people, organizations and events. The result is a flow chart showing connections that otherwise might elude analysts.

The Senate Judiciary Committee questioned the accuracy of data mining in January. Weeks later, the GAO concluded ADVISE fails to address privacy concerns, as required by federal law. The GAO also cautioned against the temptation to apply ADVISE for purposes other than counterterrorism.

Tangram, another pattern-sniffing project, picks up where the controversial Total Information Awareness left off. Still in its infancy, Tangram is overseen by the Office of the Director of National Intelligence. It strives to transform "a set of disjointed, cumbersome" technologies "into a self-configuring, continuously operating intelligence analysis support system," according to a government solicitation document for contractors.

Tangram would analyze intelligence databases and private communications, transactions and activities to give individuals a "suspicion score," the National Journal reported in the fall.

Data mining sometimes can trace connections that involve "guilt by association," where a search target is specified. Still, these exercises can take days and produce "runaway false detections," according to the government solicitation document. Spotting suspicious activity without knowing any of the actors is harder; there is no benchmark for "normal" behavior.

In such "highly uncertain instances, a Tangram-like system may be the only method by which seemingly meaningless data becomes meaningful. The objective is to find the most important 'known unknowns,'" the document states.

Tangram would deploy something called Active Learning. An automated series of "what if" scenarios would point analysts in directions calculated to yield the best information.

Yet the odds of catching deranged college students or terrorists with such tools smacks more of Tom Cruise in the movie "Minority Report" than of reality, some experts say.

Many innocent people would be implicated, said Rebecca Wright, a computer science professor at the Stevens Institute of Technology in Hoboken.

"You just can't prevent every tragedy," said Harper of the Cato Institute. "Everybody is searching for a way to prevent (massacres) from happening. But I honestly doubt there is a way that is consistent with life in a free society."

In the case of Cho, who gunned down 32 students and faculty members last Monday before killing himself, no electronic sleuthing was needed. His written assignments, demeanor and alleged stalking already had alarmed classmates and faculty.

"It's very clear that lots of people knew this was a troubled kid," said Kim Taipale of the Center for Advanced Studies in Science and Technology Policy. Trolling online for college kids with bad attitudes probably is bad policy.

"The writings, the anti-social behavior -- unfortunately, that's a lot of what college is about," he said. "Working out those demons."

Kevin Coughlin may be reached at kcoughlin@starledger.com.

© 2007 The Star Ledger

© 2007 NJ.com All Rights Reserved.