'Dark Web' Project Takes On Cyber-Terrorism

Friday, October 12, 2007

By Steven Kotler

FOX NEWS

There are currently over a billion Internet users in the world, but not all of them are friendly.

In recent years, the anonymous nature of the Web has turned it into a boomtown for all sorts of radicalized hate.

"Since the events of 9/11, terrorist presence online has multiplied tenfold," says Hsinchun Chen, director of the University of Arizona's Artificial Intelligence Lab. "Around the year 2000, there were 70 to 80 core terrorist sites online; now there are at least 7000 to 8000."

Those sites are doing everything from spreading militant propaganda to offering insurgency advice to plotting the next wave of attacks, making the net, as Chen also points out: "arguably the most powerful tool for spreading extremist violence around the world."

But thanks to Chen, that tide may be turning. He's the architect behind the newest weapon in the war on terror — a giant, searchable database on extremists known as Dark Web.

Using a bevy of advanced technologies, Dark Web is an attempt to uncover, cross-reference, catalogue and analyze all online terrorist-generated content.

This is a vast amount of material, posted in dozens of languages and often hidden behind the blandest of portals.

The more radical of these forums can host as many as 20,000 members and half a million postings, making the Web an increasing nightmare for the intelligence community, but a perfect prowling ground for a data-mining expert like Chen.

In fact, Dark Web is Chen's second foray into online crime-fighting. The first began in 1997, when he — already an expert at tracking social change online (crime and terrorisms being extreme examples of social change) — teamed up with the Tucson Police Department and the National Science Foundation (NSF) to help develop Coplink, a way for law enforcement forces around the country to link files and consolidate data.

It was Coplink that helped build the case against the Washington, D.C., Beltway snipers, John Muhammad and Lee Boyd Malvo. Because of this and other successes, in early 2002 the NSF asked Chen to try to build a similar system against terrorism.

He began with a modified version of Web-spidering. Typically, Web spiders are keyword-based followers of the hyperlinks between Web pages. This is essentially how search engines like Google and Yahoo do their work.

Unfortunately, a study done by the NEC Research Institute, the research arm of Japan's consumer-electronics giant NEC Corporation, found that existing engines cannot keep up with the Web's growth rate. Each one can only mine 16 percent of the available material.

The recent arrival of meta-search engines, capable of triangulating between several engines at once with a much higher success rate, solved this problem, but unearthed another.

"Information analysis was our goal," says Chen, "and information overload was the biggest hurdle."

To clear this hurdle, Dark Web relies on all sorts of analytical tools. It utilizes existing technologies such as statistical analysis, cluster analysis, content analysis and link analysis, as well as brand new technologies like sentiment analysis, which is capable of scanning documents for emotionally charged keywords such as "that sucks."

This form of analysis has proven effective in gauging the success of new consumer products. But instead of judging the fate of the latest movie, Chen uses sentiment analysis to look for emotions like rage and hate in an attempt to tease apart the social activists from the suicide bombers.

That's merely the beginning. Dark Web also employs social-network analysis to map extremist networks, determining the importance of each member and establishing the organizations' hierarchies.

To do this, Chen uses centrality and structural-equivalence measures to examine social-network components, such as the prestige allotted to any given poster by other members and the "closeness" — a given poster's access to information on the network coupled with his independence from others — among subjects in an attempt to further separate an organization's leaders from its outliers.

Researchers then explore things such as cohesiveness and group density — using a form of pattern analysis called blockmodeling — to help determine the stability of any given organization and, perhaps more importantly, the nodes most vulnerable to attack.

These methods were already in use before Dark Web. Chen and his cohorts also developed a few novel ideas of their own, including a technique called Writeprint which examines structural and semiotic content from anonymous postings in an attempt to determine authorship.

"The Web is a gargantuan series of diffused networks," says NSF spokesman Dana Cruikshank. "Dark Web finds the patterns that make it much less decentralized."

Chen says that if Dark Web had been online before the Iraq war, it could have determined whether the purported links between Al Qaeda and Saddam Hussein were fact or fiction.

Moreover, the database also offers a terrorism knowledge portal, essentially a search engine for extremism, and a terrorism expert finder, a database of the world's best anti-terrorism minds — two things that have been sorely missing in the war against extremism.

Despite all of this tantalizing potential, not everyone is convinced Dark Web is actually a tool for freedom.

Marc Rotenberg, executive director of the Electronic Privacy Information Center, an online civil-liberties group, says "the very same tools that can be used to track terrorists can also be used to track political opponents."

To make sure that doesn't happen, Rotenberg maintains that Dark Web must be used within the confines of our existing privacy laws — an idea that may be better in theory than in practice.

Though Chen strenuously denies it, there are a number of similarities between Dark Web and the Defense Advanced Research Projects Agency's controversial Total Information Awareness (TIA) initiative, for which funding was cut off by Congress in 2003 over civil-liberties concerns.

"Just because someone posts something we don't like on the Internet, doesn't mean they also suspend their First Amendment rights," says Mike German, the ACLU's policy counsel on national security, immigration and privacy. "Things like authorship analysis are particularly tricky. How could you know that someone was really intent on violence before that act of violence was committed?"

German, who spent years on the domestic-terrorism beat for the FBI before coming to work for the ACLU, feels that Dark Web is a great waste of critical resources.

"I know this from my time spent undercover, infiltrating exactly these kinds of organizations: Every terrorist training manual makes it clear that a huge separation should be kept between the bomb-makers and the propagandists. Between the action wing and the political wing. This means, by design, Dark Web is chasing the wrong people."

Chen disagrees.

"By design, we really only look into the contents of the propagandists of the jihadist movement," he says. "I think this is the bigger danger — the ability of the Web to attract and 'infect' young disgruntled men in the world.

"We do not get into the actual operational wings of their groups, as most of the secret operational communications are encrypted and moved off-line," Chen explains. "Tracking those secret member communications is the domain of NSA, not us."

Civil-liberties concerns may continue to dog the technological front of the war on terror, but Dark Web is already producing results.

A recent study by Chen's group of training manuals and methods to build and use improvised explosive devices posted online — including where in the world such manuals have been downloaded — has led to countermeasures that are currently keeping soldiers and civilians alike safer. Which is, after all, the point.