

[Home](#) / [News](#) / [March 10th, 2012](#); [Vol.181 #5](#) / [Article](#)

Scientists probe terrorist talk on 'Dark Web'

[Mathematical tools pry clandestine information from hidden realm of the Internet](#)

By Rachel Ehrenberg

March 10th, 2012; Vol.181 #5 (p. 8)

VANCOUVER — Deep within the recesses of the Internet, extremists discuss and plot terrorist acts. But new mathematical tools that combine web crawling techniques, sophisticated algorithms and human expertise are gaining access to this “dark side” of the Web and may help predict and prevent violence.

Researchers engaged in the “Dark Web Project,” a program started partly in response to the 9/11 terrorist attacks, have developed methods for tracking the spread of dangerous ideas through certain rogue and jihadi Web forums. Using a mathematical model known as SIR, used by epidemiologists to describe the transmission of disease, researchers have determined that the infection rate for becoming a suicide bomber is 2 in 10,000, Hsinchun Chen of the University of Arizona in Tucson reported February 18 at the annual meeting of the American Association for the Advancement of Science.

“Violence in social media is infections of the mind,” said Chen.

The Dark Web Project, housed at the University of Arizona, collects information from blogs, forums and other websites from hidden realms of the Web. Search engines typically explore only what’s known as the publicly indexable web. The invisible Web, which includes these Dark Web forums, is estimated to contain 500 times as much information as the surface web.

Dark Web forums are particularly tough to crack. There’s no centralized index of forums, and access is often restricted to people who have to apply and be approved, which can take weeks. Using mathematical approaches to identify and target forums from known extremist sites and less obvious places, such as an AOL group, people then apply for membership. If they can gain access, the researchers then need to assess things like how often the site downloads information and how many connections it has. Then they can use a crawling or “spidering” technique to collect and index information from the forums.

Authorship analysis techniques can then reveal what messages come from the same individual. The approach is up and running for English and Arabic, and in process for French, Urdu and Pashto, said Chen.

Such analyses, whether of Dark Web forums or mentions of influenza on Twitter, can be fraught with difficulty, said Vinton Cerf, chief Internet evangelist for Google. Words may be misspelled, and new phrases are invented every day.

The analyses of the Dark Web forums suggest that the longer participants are involved in a forum, the more violent their messages become.

Many of the collected threads are now available for researchers through a Dark Web Forum Portal, which contains more than 15 million messages. The team has also started a video portal where video content, who posted it and comments on it can be explored. In the two months that researchers have been working on the video portal, they've archived more than 7 million messages and 850,024 threads.

CITATIONS & REFERENCES :

H. Chen. Security Informatics: The Dark Web Experience. American Association for the Advancement of Science annual meeting, Vancouver, February 18, 2012. Abstract available: