

The Philadelphia Inquirer

New Tool Automates Analysis of Jihad Sites

11.18.07

By: Arthur H. Rotstein

TUCSON, Ariz. - The quivering images and militant writings are frightening: an exploding Humvee; a lab technician making explosives, step by step; and "A guide to kill Americans in Saudi Arabia."

Tens of thousands of Web pages are now devoted to terrorist propaganda. On the surface, the messages and videos reveal little about their creators. But programmers and writers leave clues: the words they choose, their punctuation and syntax, and the way they code attachments and Web links.

Researchers at the University of Arizona are developing a tool that uses these clues to automate the analysis of online jihadism. The Dark Web project aims to scour Web sites, forums and chat rooms to find the Internet's most prolific and influential jihadists.

Lab director Hsinchun Chen hopes Dark Web will crimp what he calls "al-Qaeda University on the Web," the mass of Web sites where potential terrorists learn their trade. Experts said they were not aware of any comparable effort, though some said the project might have only limited applications.

The project in the university's Artificial Intelligence Lab will not identify people outside cyberspace, "because that involves

civil liberties," Chen said, preferring to let law enforcement and intelligence analysts take over from there. Instead, it will help identify messages with the same author and reveal links that are not obvious.

"Our tool will help them I.D. the high-risk, radical-opinion leaders in cyberspace," Chen said. A few agencies are on the verge of using some of his team's techniques, he said, but he would not name them.

Former FBI counterterror chief Dale Watson, who noted that terrorist Web sites and communications were now analyzed manually, said the ability to sort through so much data electronically "would be a great asset in the fight against terrorism."

"It would greatly enhance the speed and capability to sort through a large amount of data," Watson said. "That would be the key here. The issue will be where is the Web site originating and where are the tentacles going?"

The bulk of a \$1.3 million grant the National Science Foundation gave Chen's group will focus on who produces improvised explosives and what they talk about - such as U.S. troop movements and terrorist tactics. Before getting the NSF funding, Chen started the project with about \$3 million from other Artificial Intelligence Lab programs.

Dark Web's software, which

Chen calls Writeprint, samples 480 factors to identify whether the same people are posting to multiple radical forums. It can analyze everything from a fragment of an e-mail to videos depicting U.S. soldiers blown up in Humvees and fuel tankers.

Writeprint is derived from a program originally used to determine the authenticity of William Shakespeare's works. It looks at writing style, word usage and frequency and greetings, and at technical elements ranging from Web addresses to the coding on multimedia attachments. It also looks at linguistic features such as special characters, punctuation, word roots, font size, and color, he said.

Currently, intelligence analysts cannot effectively analyze writing style in cyberspace, particularly multilingual writings, he said.

"But using our tool . . . we can get about 95 percent accuracy, because I'm utilizing a lot of things your naked eye cannot see," Chen said.

He and counterterror specialists said that what he termed a tenfold increase in the last two years in jihadist content appearing online has outstripped intelligence analysts' abilities.

"Automating this is absolutely necessary," said Evan Kohlmann, a terrorism expert with the Washington-based Investigative Project on Terrorism. "We're reaching that finite limit" of what can be done manually by humans.