

RESEARCH REPORT

Mining the Dark Web

MIS department develops terrorism informatics.

In the wake of the World Trade Center disaster five years ago, one of the issues that came into sharp focus was the unpreparedness of law enforcement and the intelligence community to collect, analyze, and interpret data on terror cells.

"Law enforcement was getting very deep information from their investigations," says McClelland Professor of MIS Hsinchun Chen, but by researching on a case-by-case basis, "they were doing it the old-fashioned way. They were behind by many years."

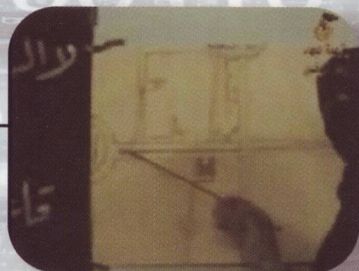
In 2003, Chen participated in a Sandia National Laboratory conference called KnowFest, which explored the possibility of creating a terrorism knowledge network.

"The idea was to bring a methodology to it, to look at the issue on a large scale through collection and analysis," he continues. Traditional law



Hsinchun Chen, professor of MIS and director of the Artificial Intelligence Laboratory.

MOSUL, IRAQ SUICIDE BOMBING VIDEO FROM THE DARK WEB ARCHIVE



PHASE 1: Planning Session



PHASE 2: Farewell



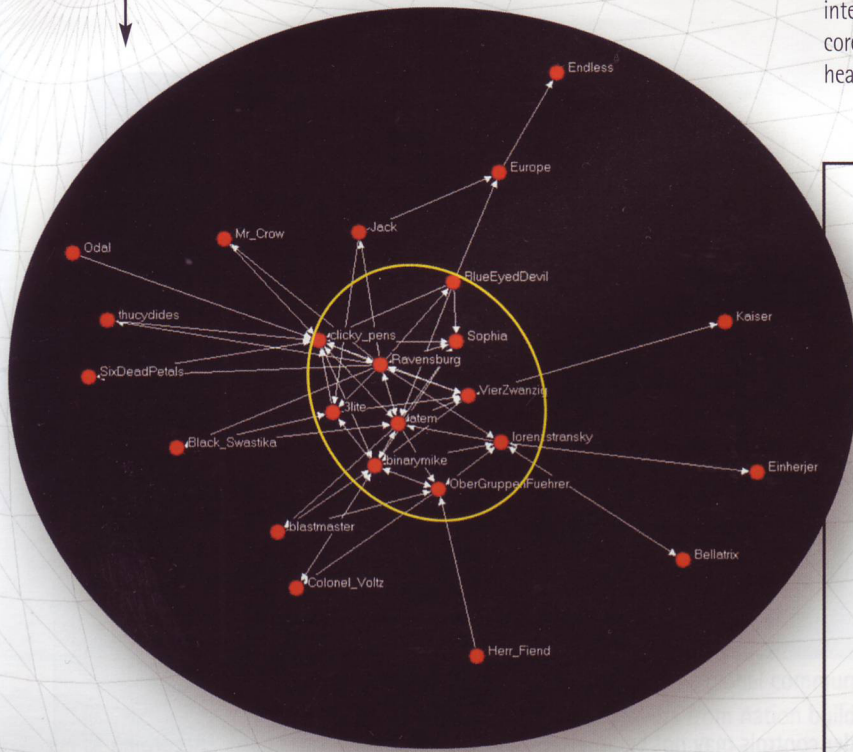
PHASE 3: Suicide Truck Preparation

enforcement and intelligence analyses monitor a specific person's online activities, but after the conference, Chen and a team of Eller graduate students created a series of programs and mechanisms to capture the big picture of terrorism informatics through what has become known as the Dark Web project.

The team designed a set of programs to comb through the Internet, collecting and tracking terrorist and extremist websites — resulting in a digital library of two terabytes' worth of material including websites, forums, and video clips. The database is one of the largest such collections in the world.

In the second phase of the project, Chen says his team performs analysis, including content analysis to address how terror cells use their web presence to recruit, fundraise, train, and disseminate propaganda internally and externally. Though the Dark Web project has collected some 1,500 websites to analyze, Chen notes, "Sites are stale — it's the forums that are more interesting."

THREAD INTERACTION NETWORK — An evaluation of the thread interaction network reveals a core set of participants at the heart of a forum discussion.



Even defining which groups can be considered terrorist organizations proved challenging. "The term 'terrorist' is politically charged," says Chen. One country's freedom fighters can be another country's terrorist group. Chen says that in their process of definition, he and his researchers considered the policies of the U.S. State Department, the European Union, and China, among others. The Dark Web project ultimately follows terror groups as defined by the U.S. State Department and extremist groups as defined by the Anti-Defamation League and the Southern Poverty Law Center.

Also challenging: many of Chen's team of master's and Ph.D.-level MIS students hail from the regions affected by terrorist activity, including Palestine and Lebanon. "We're an international group," he says, and part of the reason is necessity, since they might not otherwise have the resources to translate the range of languages represented in the Dark Web. But, he says, the project is emotional and people can't help but get involved.

In his own assessment, Chen notes that what's more alarming than the roughly 20% of fundamentalist Muslims who might be sympathetic to terror groups like al-Qaida are the 60% who are in the middle — those who may become "homegrown" terrorists. "They can be stirred up by events, but also flipped from off to on," he says. "They are not the hardcore people, but they are a huge population and that's the scary part."

Chen says the Internet has become a flattening tool, but is not without hope that the technology he has pioneered might be used by law enforcement officials and intelligence analysts to counter the ability of extremist groups to sway the people in the middle, and to predict and prevent terror attacks. "That's our noble goal," he says, "to make the world a better place."

rorists to take violent action. Chen says that by tracking those infectious ideas, it might be possible to seek remedies in the political arena and counter the message before the ideas escalate into violence.

But countering extremist measures falls outside Chen's expertise. "I'm a professor, not in the CIA," he says. "We don't give leads. What we do is help agencies who come to us for assistance to develop a systematic methodology and advanced technologies. We can teach them to use the tools we've created."

It's the same model Chen applied when he developed COPLINK, an integrated system of law enforcement databases to support crime analysis through a web-based interface. The program is now in use by more than 300 police departments across the U.S., including throughout the state of Arizona.

"COPLINK was a pretty black and white program," says Chen. "We are the good guys and they are the bad guys, and we are helping cops do their jobs. The Dark Web can be a gray area."



PHASE 4: Explosion

Terror groups that are sophisticated enough to host forums on their websites — which let visitors and administrators post messages online — have allowed the Dark Web team to explore authorship analysis in English, Chinese, and Arabic. Authorship analysis is the process of determining an individual's "writeprint" or unique habits of syntax and other considerations. It was the same assessment applied to the works of Shakespeare when the sole authorship of his plays was called into question.

Terror cell forums also enable sentiment analysis, monitoring how infectious ideas catch fire online and galvanize ter-