September 17, 2006 - 6:18AM

DIGITAL SPYING EXPERT: University of Arizona associate professor Hsinchun Chen's software, among other things, provides super-speed database searches.

U of A honing online intelligence

Ryan Gabrielson, Tribune

Hsinchun Chen goes where terrorists gather. He monitors what they say and, particularly, how they say it. He tracks who they are talking to, and whether they are spreading propaganda or providing training. "They're hiding," Chen said. "You need to dig them out."

Chen is not operating undercover, or eavesdropping on nearby conversations in an unmarked van on a darkened street. He is a middle-aged, polite computer science professor at the University of Arizona, working from a closet-sized office in Tucson. A supercomputer in his building's basement uses data-mining to scour the Internet and analyze the intelligence it gathers for the U.S. government. Chen is one of dozens of university computer researchers serving as online intelligence officers training the next generation of digital spies.

In the five years since the Sept. 11, 2001, terrorist attacks, many academics have joined ranks with the intelligence community, said Peter Freeman, a National Science Foundation official.

"This time we're on the front lines," he said.

The role is unusual and, at times, awkward for academics. After building a career on the premise of publishing what they know, their funding is coming from agencies that specialize in secrets.

It also is a new course for the intelligence community which has been forced to look outside of itself to devise effective battle tactics against elusive extremist groups, Freeman said.

CONNECTING THE DOTS
Terrorists have made the Web a headquarters where they communicate largely without detection. The Internet is a veritable blizzard of electronic activity, making it highly difficult to track any

one person or group, Chen said.

So agencies such as the Department of Homeland Security and the Central Intelligence Agency have pumped millions of dollars into computer science research.

The work is focused on data-mining, the process used to connect the dots when there are billions of dots from which to choose. Analysts construct formulas, known as algorithms, to unearth patterns in a huge mass of data.

That data can be financial records, news articles, video files, message boards and personal e-mails.

Banks and cell phone companies have for years used it to recognize abnormal spending or calling behavior, which signals that a phone or credit card might have been stolen.

The intelligence community's hope, Chen said, is that data-mining can streamline the digital chatter to show where and how terrorist plots are unfolding.

The tool has become so popular that in 2004, the Government Accountability Office, Congress' watchdog agency, counted 199 data-mining programs under way or in development.

While the spy agencies' interest has been a boon for computer science researchers, it comes with conditions. Potentially the most significant, several of the academics say, is their discoveries have to be handed to agents who might use the information in ways they oppose.

"What the government's going to do with it after is an issue that everyone should worry about," said Janyce Wiebe, a University of Pittsburgh computer science researcher. "We've already got that problem."

SAFETY VS. PRIVACY

Government forays into using data-mining to hunt terrorists have ignited controversy and concern that citizens' privacy rights will be violated.

The Bush administration has had to fend off questions about its warrantless surveillance of international phone calls and e-mails by the National Security Agency, disclosed in news reports last December. That operation was recently ruled illegal by a federal judge.

Three years ago, the intelligence community dismantled a massive data-mining operation dubbed "Total Information Awareness" that sought all forms of electronic information on people, including such personal information as health history and travel records, to uncover terror connections. The program came under deafening public criticism for its alleged monitoring of innocent citizens.

Because data-mining allows analysts to get the information they are looking for from a massive pile of data relatively quickly, it is becoming far easier to monitor a whole population.

Computer science is developing tools that limit online privacy, Wiebe said, and the American people “have to figure out how they’re going to deal with that.”

Using multiple intelligence community grants, Wiebe is working to develop ways for a computer to discern opinion from fact in text. Last month, she was awarded part of a $10.2 million grant from Homeland Security that, in addition to paying for research, requires her to instruct the agency’s analysts to use her technology.

SPY SCHOOLS

Most university researchers are devising better ways to find and analyze online extremist information, but do not collect it themselves. The aim of recent federal funding, in addition to advancing datamining, is to establish spy schools within universities.

Students are already receiving training in the field, said Edward Hovey, director of the Information Sciences Institute at the University of Southern California, but most of them go on to search-engine companies like Microsoft and Google.

To date, data-mining has primarily been applied to solve business problems.

Teaching intelligence agents presents new complications, said Hovey, who has received a grant to establish a spy school at USC.

“Somebody comes to you and says, ‘You know, I really need this because I want to stop some guy from blowing you up tomorrow.’ It’s really hard to say no,” he said. “On the other hand, to say, ‘I’m going to go behind the curtain and do all kinds of nefarious illegal things,’ like the NSA did with the phone numbers, then absolutely I refuse to even talk there.”

The university researchers say they are not working with classified material.

A majority of computer science researchers are not American citizens, Hovey said, severely limiting how close they can become to the spy agencies.

However, public Internet data, called “opensource,” is rich with information on extremist groups, said Fred Roberts, a computer science researcher at Rutgers University.

Public Web sites and message boards allow terrorists to connect for free with those they hope to lure, mainly disaffected young men, Chen said. The UA research team found multiple extremist organizations from around the world using Yahoo and Google messaging groups to recruit.

From there, the extremists can direct each other to increasingly sophisticated Web sites that hold multimedia files explaining how to carry out attacks. Chen’s team found an online video that gives step-by-step instructions on how to execute a suicide car bombing.

“If they don’t get the tools they are just disgruntled young men,” Chen said. The UA team passes its intelligence on to spy agencies. Chen declined to name which ones.

COVERT CASH

Chen's grant funding comes from every corner of the intelligence community.

But while the researchers' work is to be made public, little information is disclosed about what money the spy agencies are providing the universities. The Tribune reviewed a database of federal grants for the previous three years and found that intelligence community cash is rarely handed directly to researchers.

One major program — "Knowledge, Discovery and Dissemination" — involves nine spy agencies and has funneled research money through the science foundation since 2001, government documents show. While the program's grants are listed in the database, it was never mentioned by name.

The grant that provides much of the funding for Chen's work came from Homeland Security for a project labeled "Border Safe," he said. Again, the database does not include a grant from the agency for UA to search out terror groups.

Instead, Homeland Security provided almost $2 million in 2003 to the Corporation for National Research Initiatives, a nonprofit with close ties to the Pentagon, for Border Safe.

Buried deep in the nonprofit's public tax records for 2003 is a document listing $663,566 provided to UA. The transaction is listed only as "information management."

DATA MINE FIELD

There are several fields within data-mining. Computer science researchers are hunting for the means to analyze video as effectively as they do numbers, to understand what a text is saying regardless of its language and identify its author by how he uses his words.

Though Chen boasts his research is already netting important information on extremists, as an anti-terror tool, data-mining has significant limitations.

"It's the hot thing to stop terrorism," said Bruce Schneier, a computer security expert with Counterpane Internet Security in Mountain View, Calif. But, he added, "It is completely ineffective."

The greatest flaw that Schneier sees is that datamining is likely to generate a huge number of false alarms, patterns the computer misinterprets as terror activity. Those false alarms waste law enforcement's time and prompt unnecessary surveillance, or possibly detainment, of innocent citizens. The university researchers acknowledge data-mining's current shortcomings. As part of her research, Wiebe said she is attempting to upgrade computers' ability to understand how humans use language. A classic example of the problem is the phrase, "killed two birds with one stone," she said. A computer would interpret that as an actual event where the stone was a weapon, the birds victims.