Carnegie Mellon developing online tools to detect and identify sex traffickers

January 13 2015

A contract from the Defense Advanced Research Projects Agency (DARPA) is funding researchers at Carnegie Mellon University's School of Computer Science who are seeking to undercut sex traffickers by developing online tools that target a key vulnerability: the need to advertise.

The three-year, $3.6 million CMU effort will develop machine learning algorithms for domain specific indexing and search, and will build on existing efforts by CMU researchers to analyze ads for sex services posted to websites.

The contract is part of DARPA's Memex program, a three-year research initiative to develop software that will enable domain-specific indexing of open, public Web content and domain-specific search capabilities. The contract is administered by the Air Force Research Laboratory in Rome, N.Y.

The sheer number of ads by sex traffickers and the high rate at which they are posted and reposted on websites can be daunting for investigators. Computer programs, however, can uncover commonalities that can help detect prostitution rings, identify children and adults who have been coerced into service and locate the human traffickers themselves.

"In the end, sex traffickers are trying to sell a service and ultimately they have to advertise those services. That's their Achilles' heel," said Jeff Schneider, the project's principal investigator and a research professor in the School of Computer Science's Auton Lab, which studies statistical data mining.

In addition to analyzing obvious clues, CMU experts in computer vision, language technologies and machine learning will develop new tools for such tasks as analyzing the authors of ads or extracting subtle information from images.

The project will build on the work of Artur Dubrawski, senior systems scientist and the project's co-principal investigator, and Emily Kennedy, a research analyst and alumna of CMU's Dietrich College of Humanities and Social Sciences, both of the Auton Lab. Their work began more than three years ago as part of Kennedy's undergraduate honors project.

"Originally, we looked for ways to help the victims of human trafficking," Dubrawski said. "But we quickly realized the best way to help the victims would be to help law enforcement."

They developed a program, called Traffic Jam, and have trained more than 200 law enforcement officers in its use.

Detective Darren Ruskamp of the Modesto (Calif.) Police Department used Traffic Jam to follow up on a tip about a Nebraska girl, identifying a sex trafficker who was travelling with prostitutes across the Midwest and West and culminating in his arrest. Traffic Jam enabled him to gather evidence by quickly reviewing ads the trafficker posted for several locales.

"It made a huge difference," Ruskamp said of Traffic Jam. "Not many investigators are going to have eight or 10 hours to search ads in multiple cities."

The new project at CMU is funded through DARPA's Memex program, which is developing new Web search technologies.

Schneider said the tools being developed in the CMU program will be useful not only in combating [human trafficking](link), but in creating indexing tools applicable for the Deep Web. The Deep Web includes not only notorious sites such as online market places for illicit drugs and other contraband, but also any site not routinely accessible through commercial search engines. The Deep Web is estimated to be many times larger than the surface Web and growing exponentially.

In addition to Schneider, Dubrawski and Kennedy, CMU's team includes computer vision experts Abhinav Gupta and Martial Hebert of the Robotics Institute, and Noah Smith of the Language Technologies Institute, an expert in automated analysis of human language. It also includes Claire Schmidt of Thorn, a foundation that leverages technology to combat child sexual exploitation.