To catch a cyber-thief

Concordia University

 Montreal, June 5, 2013 – When local police came calling with child porn allegations last January, former Saint John city councillor Donnie Snook fled his house clutching a laptop. It was clear that the computer contained damning data. Six months later, police have finally gathered enough evidence to land him in jail for a long time to come.

With a case seemingly so cut and dry, why the lag time? Couldn't the police do a simple search for the incriminating info and level charges ASAP? Easier said than done. With computing devices storing terrabytes of personal data, it can take months before enough evidence can be cobbled together from reams of documents, emails, chat logs and text messages.
That's all about to change thanks to a new technique developed by researchers at Concordia University, who have slashed the data-crunching time. What once took months now takes minutes.

Gaby Dagher and Benjamin Fung, researchers with the Concordia Institute for Information Systems Engineering, will soon publish their findings in Data & Knowledge Engineering. Law enforcement officers are already putting this research to work through Concordia's partnership with Canada's National Cyber-Forensics and Training Alliance, in which law enforcement organizations, private companies, and academic institutions work together to share information to stop emerging cyber threats and mitigate existing ones.

Thanks to Dagher and Fung, crime investigators can now extract hidden knowledge from a large volume of text. The researchers' new methods automatically identify the criminal topics discussed in the textual conversation, show which participants are most active with respect to the identified criminal topics, and then provide a

visualization of the social networks among the participants.

Dagher, who is a PhD candidate supervised by Fung, explains “the huge increase in cybercrimes over the past decade boosted demand for special forensic tools that let investigators look for evidence on a suspect’s computer by analyzing stored text. Our new technique allows an investigator to cluster documents by producing overlapping groups, each corresponding to a specific subject defined by the investigator.”

Fung says that, “out of all the types of available data in cybercrime investigation, text data is the most common medium used by scammers, identity thieves and child exploitation criminals. But this type of data is also the most challenging to analyze. It’s really hard make a software program automatically interpret the underlying meaning of the text.”

The researchers have also developed a new search engine to help investigators identify the relevant documents from a large volume of text. Says Dagher, “In a normal search engine, a user enters some keywords and results can vary – widely. In contrast, our search engine captures the suspects’ vocabulary, and then uses it to improve the accuracy of the search results. Just like some cultures are said to have over 50 words for snow, criminals might have 50 words for… snow of a different kind! This search engine allows investigators to pick up on those nuances and quickly identify the incriminating documents.”

“Experiments using real-life criminal data already suggest that our approach is much more effective than the traditional methods,” says Dagher. This new method of quickly sifting through huge amounts of text to zero in on the evidence could soon be used by law enforcement agencies around the world, meaning future cybercriminals can go to trial much more quickly, saving time for the police – as well as money for tax-payers.

Source: http://www.concordia.ca/now/media-relations/news-releases/20130605/to-catch-a-cyber-thief.php [1]

Links:
[1] http://www.concordia.ca/now/media-relations/news-releases/20130605/to-catch-a-cyber-thief.php