With computing devices now storing terrabytes of personal data, it could take months to find any serious types of forensic data from the huge amount of documents, emails, chat logs and text messages.
But the speed of the search is about the change because of a new technique developed that can slash data crunching time. What once took months can now takes minutes, said researchers at Concordia University.

**RELATED STORIES**
- Energy Firm Suffers Breach
- Spear Phishing: Energy Sector Targeted
- Manufacturing Most Attacked Industry
- Simulated Attacks Hike Security Awareness

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

Crime investigators can now extract hidden knowledge from a large volume of text. 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.”

“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,” Fung said. “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 relevant documents from a large volume of text. “In a normal search engine, a user enters some keywords and results can vary – widely,” Dagher said. “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,” Dagher said.

This new method of quickly sifting through huge amounts of text to zero in on the evidence could soon be in full-scale use 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 taxpayers.

Leave a Reply

You must be logged in to post a comment.

« Schneider Mitigates PLC's Holes
Global Cybercrime Botnet Breached »