RERSEARCH IN ACTION

Cybercrimes involving anonymous emails have been on the rise over the past few years. The transmission of threats, viruses and child pornography; drug trafficking; cyber terrorism or using e-mail as a safe channel for communications by cybercriminals: these are just some of the harmful ways in which these messages act. Luckily, researchers at Concordia are working hard to put a stop to it.

For recent PhD graduate Farkhund Iqbal, this subject has formed the basis for fascinating postdoctoral research. Working closely with supervisors Mourad Debbabi and Benjamin Fung, both professors with the Concordia Institute for Information Systems Engineering, he has helped develop an effective new technique to determine the authorship of anonymous emails. What's more, this special method can also be used in plagiarism detection, anti-aliasing in online trading and business, authorship characterization, and even in authorship verification disputes.

To determine whether a suspect has authored an email, they first identify the patterns found in writing samples that have been identified as belonging to the subject. Then, they filter out any of these patterns that are also found in the emails of other suspects. The remaining patternsare unique to the author of the emails being analyzed and constitute the suspect's "Writeprint," a distinctive identifier like a fingerprint. Common writing style patterns such as lexical and syntactic features, structural features found in the layout of an email, content-specific features; and idiosyncratic features can be used to create the Writeprint. The method even allows the determination of the nationality, gender, and education level of the anonymous author.

Iqbal is now developing a way to visually represent the Writeprint, in order that lay people may quickly identify common traits within anonymous emails. Law officials can then use the Writeprint software to generate an according picture, and rely on this visualization technique to identify commonalities between messages -- be they emails, blog posts, Internet chat logs or other anonymous textual transmissions -- thus leading to identifying their author. "A person's Writeprint is unique," explains Iqbal. "When we extract the features, we delete the overlapping patterns. So it's distinct. Everyone has one and, just like a finger-print, your Writeprint is unique to you."

This reliable, effective way to determine which of several suspects has written the emails under investigation will give law enforcement officials the cyber-forensic evidence needed to prove authorship -- and guilt -- to a high degree of accuracy. To test the accuracy of their technique, Iqbal and his supervisors examined the Enron email dataset, a collection that contains more than 200,000 real-life emails from 158 employees of the Enron Corporation. They were able to identify authorship with an accuracy of 80 to 90 percent.

These impressive results have led law enforcement agencies to use the technology in prosecuting cybercrimes. Iqbal is now developing this visualization technique for "non-technical people so that they have the feeling where they don't have to say okay,

what are the features: you are using semi-colons a lot or you sign off with cheers instead of best regards. The Writeprint program gives notations to these different features so that you can see with your naked eye which message is the most like the malicious one. We are expecting that it will be greatly appreciated by the cybercrime industry."

The next step is to develop Writeprints for electronic text sources written in languages other than English. "A multilingual approach is very necessary in the present international climate," explains Iqbal. "With the Internet, people who are collaborating may speak several different languages. So this is one of the research directions that I am planning to work on. We need to deal with this kind of multilingual text. The same text may contain two or three languages. We need to design parsers that can parse any language and identify the semantic meaning of words in different languages."

Iqbal has been studying Computer Science since 1989, when he pursued undergraduate studies in his native Pakistan. Fascinated by cybersecurity, he came to Canada in 2002 to pursue his master's degree at Concordia. He successfully defended his PhD thesis last January to a standing-room crowd and now continues to work closely with Fung and Debbabi, who he says, are "excellent supervisors. They are both demanding and generous and have really challenged me to produce the best work possible."

This has paid off in spades for Iqbal, whose research has been widely published in prestigious journals around the world. Now that the word is out about this nascent Writeprint technology, Iqbal hopes to translate the attention he has received from peers and the media into further success at international conferences. Although he hopes to continue his career at Concordia, attention from internationally respected researchers like Hsinchun Chen, anti-terrorism expert from the University of Arizona, may lead him farther afield in the ongoing hunt for cyber-criminals.

To learn more about Iqbal's work and for sample papers and media coverage, visit www.ncfta.ca.