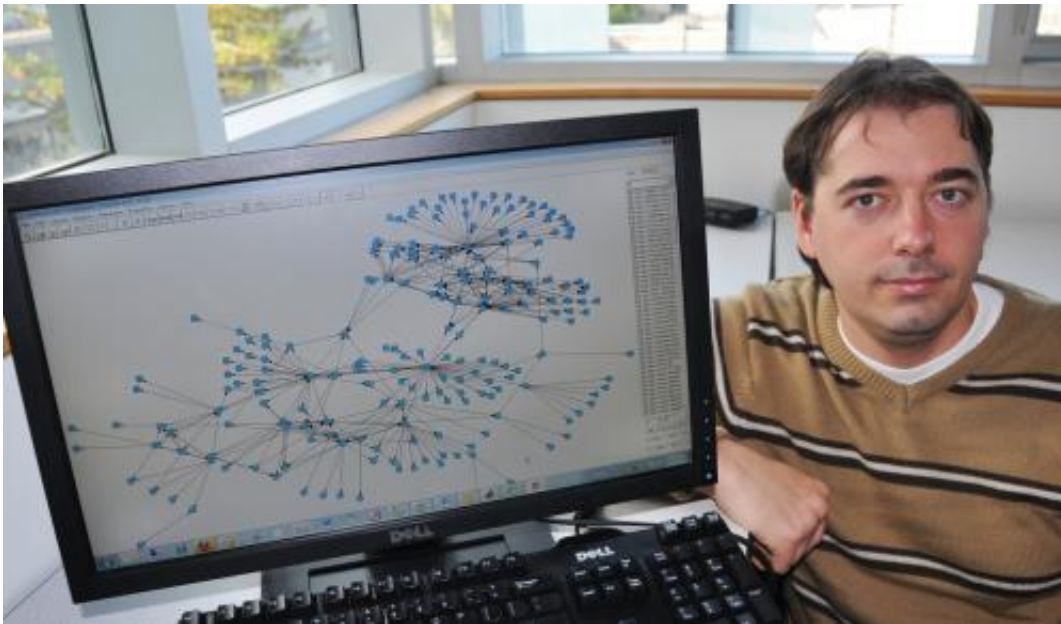


Researchers to expand child exploitation web-crawler

June 3 2014, by Richard Frank



Richard Frank

Researchers in Simon Fraser University's International Cybercrime Research Centre are expanding their Child Exploitation Network Extractor (CENE)—an online "web crawler" that identifies and tracks child exploitation networks online—to determine where networks are located.

They've secured a grant from the Canadian Internet Registration Authority (CIRA) through its inaugural Community Investment

Program. It's one of 28 projects shortlisted from 149 submissions to share more than \$1 million in funding.

CIRA seeks to build a stronger internet in Canada by providing funding to community groups, not-for-profits and academic institutions for projects deemed to enhance the internet for the benefit of all Canadians.

At present, CENE enables researchers to track online child-exploitation (CE) networks—a series of websites that are hyperlinked through URLs and lead consumers of CE [content](#) from one website to another.

Using the [web crawler](#), researchers can download webpages and recursively follow their links while collecting statistics on the content, keywords, images and videos.

"This allows for automatic identification of sites that contain CE content, without having users to view the content," says CENE's co-developer Richard Frank, an assistant professor in SFU's School of Criminology. "Through collaborations with the RCMP, their database of CE-image identifiers was integrated into CENE, allowing CENE to detect known CE images."

Previous research in online CE was aimed at developing and testing CENE to collect data on websites that disseminate CE content. Researchers found it was possible to map the networks of CE websites—a measure called "network capital"—and take into account both the harmful nature of website content and their connectivity.

They also discovered that "bridges" between otherwise unconnected parts of the network lead to the most damage, and that many of the millions of images seized by the RCMP can still be found on openly accessible websites.

"Although we can identify these websites, we cannot identify where they are," explains Frank. "This deficiency was pointed out by the RCMP in discussions concerning the next steps, as the agency is primarily concerned with illegal content within its own jurisdiction that it can directly act on."

With CIRA funding, researchers will extend CENE in two critical ways. They will add a "geolocation" capability into CENE that would enable it to determine the approximate physical location of the server. This would be accomplished through the use of a third-party service or an "ip2location" database.

CENE will also be able to look up the registered owner of the domain to identify the registered location of the owner. This could be accomplished by services offered by CIRA.

Frank says once CENE becomes location-aware, researchers will focus on the global distribution of CE content and registered owners to identify the countries where the content is hosted.

Meanwhile Canadian websites will be analyzed by comparing their structure, survival and content to websites not located in Canada, and determine their relative importance globally.

Provided by Simon Fraser University

Citation: Researchers to expand child exploitation web-crawler (2014, June 3) retrieved 24 April 2024 from <https://phys.org/news/2014-06-child-exploitation-web-crawler.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
