Criminal Investigations and Network Analysis
A DHS CENTER OF EXCELLENCE AT GEORGE MASON UNIVERSITY

A Machine Learning-Based Approach to Analyzing and Triaging Encrypted Data Containers in Law Enforcement Applications

Lead PI: Kim-Kwang Raymond Choo, The University of Texas at San Antonio

Traditional and Digital Forensics

SUMMARY

Online sexual exploitation and abuse of children is a problem growing exponentially in the U.S. DHS requires improved digital forensic and investigative capabilities in cases that involve child exploitation and abuse materials. This project will provide a machine learning model for detecting, analyzing, and triaging encrypted data containers, without the need to first decrypt the content, allowing law enforcement agencies to build probable cause for a court order, facilitating investigation of child sexual abuse materials (CSAM).

PROBLEM STATEMENT

Existing technical approaches for detection of CSAM generally focus on the detection and recognition of individual objects. However, such an approach is ineffective when dealing with encrypted data. This reinforces the importance of designing systems that can be used to analyze and triage encrypted data containers. Despite recent advances in artificial intelligence research, there have only been limited attempts to explore the use of machine or deep learning in the detection of file(s) of interest contained in an encrypted container, for example based on the encryption pattern of a particular file type. This research seeks to improve investigative capability in cases involving child sexual abuse materials (CSAM), by providing DHS and other law enforcement agencies with a machine learning model for detecting, analyzing, and triaging encrypted data containers, without the need to first decrypt the content, to build probable cause for a court order to unlock the device.

APPROACH

The machine learning model will utilize deep neural fuzzy classification techniques that provide a certainty rate for similarities between contents, based on their file types. Such a model will facilitate the investigation of CSAM, and is designed to complement existing systems such as Microsoft's PhotoDNA and the Child Exploitation Tracking System.

ANTICIPATED IMPACT FOR DHS

The *Riley v. California* case, and many other cases that may have gone unreported, reinforce the importance of designing technical solutions to detect, analyze and triage encrypted data containers to build probable cause without the need to first decrypt the content. When the probable cause has been established, enforcement agencies can then either apply for a court order to unlock the device or conduct an electronic device search.

The research team will provide the law enforcement community and other relevant stakeholder groups with a deployable system that can be used immediately to supplement law enforcement and other relevant efforts.