



## Digital & Multimedia Section – 2009

---

### **B12 Supporting Cyber Crime Investigation With the UAB Spam Data Mine**

*Chengcui Zhang, PhD\*, CH 127, 1530 3rd Avenue South, Birmingham, AL 35294; Chun Wei, MS, Wei-Bang Chen, MS, Richa Tiwari, MS, and Xin Chen, PhD, CH 128, 1530 3rd Avenue South, Birmingham, AL 35294; and Gary Warner, BS, CH 100, 1530 3rd Avenue South, Birmingham, AL 35294;*

After attending this presentation, attendees will understand how cybercrime investigations can be assisted and how additional evidence of guilt can be gathered through queries to the UAB Spam Data Mine. The UAB Spam Data Mine gathers millions of email messages together into a relational database which supports rich queries as well as complex data analysis to reveal non-intuitive relationships between the cybercrime events to be identified. Online for more than a year, the Data Mine has been successfully used to merge multiple phishing cases against several brands into single cases, and to provide additional data used in the sentencing portions of cases to prove dates and durations of criminal activity in several cases in multiple countries.

The Spam Data Mine will be explained, including the sources for the millions of emails, and the method of parsing, analyzing, and clustering the data. How the Data Mine has been used successfully as the starting point of a successful Malware Investigation will be demonstrated, proving that multiple seemingly unrelated malware attacks were actually a single attack aimed at stealing financial account information through keystroke logging of compromised computers, and leading to identification and arrest of involved perpetrators. In many cases, the Spam Data Mine was able to rapidly conclude that a malware attack was underway, even when the anti-virus products had not yet been updated to provide signatures to detect the emergent malware.

In the second part of this presentation, also discussed is how spam campaigns which use "image-based" spam can be successfully clustered into their appropriate campaigns, even when the images are obscured to prevent successful optical character recognition (OCR). A unique approach to separating complex images into several layers for deeper analysis and matching will be explained. A case study will be presented where image analysis was able to identify many spam messages all related to the same "Stock Market Pump and Dump" campaigns which will be used to illustrate the possibilities that Image-based clustering of emails can achieve to assist Law Enforcement. The ability to rapidly identify whether a new spam message is part of an existing criminal operation or is something new will be illustrated.

While the computer science aspects of data mining and image processing techniques will be discussed in some detail, many examples of the types of complex law enforcement queries that can be supported by the UAB Spam Data Mine will be provided. Illustrations of appropriate ways to use these new capabilities in support of investigative efforts will be discussed for many scenarios.

**Computer Forensics, Spam, Spam Images**