B18 Changes in Approach for Scalability in Digital Forensic Analysis Emphasizing Distributed, Collaborative, and Automated Techniques

Brian Roux, MS, 701 Poydras Street, Suite 150P, New Orleans, LA 70065*

After attending this presentation, attendees will understand the difficulties currently faced in digital forensics, the mounting problem increasing storage capacities pose for the field, the failings inherent in current generation digital forensic tools' approach to analysis, the necessity of adopting a collaborative and staged approach to analysis, and see an example of how a new such approach can be implemented to address these problems.

This presentation will impact the forensic science community by proposing a more scalable architecture for digital forensic analysis and exposing failings in current generation practices as they relate to time constrained scenarios.

Digital forensics is traditionally performed using "dead" analysis where storage devices are imaged for later processing while the device is in an off state. Modern digital forensics is moving toward the employment of "live" analysis techniques for memory capture, among other things, prior to power off or solely live techniques in cases where systems cannot be powered down. The former approach requires extensive time for thorough analysis and the later requires highly skilled individuals able to deal with unexpected situations arising from live analysis where malicious content could be triggered. Both practices require finite resources: time and expertise.

In a criminal setting the expanding need for immediate triaging of devices is beginning to be articulated as backlogs expand. In England, for example, the Association of Chief Police Officers is searching for a system to act as a "digital breathalyser" to deal with evidence backlogs approaching two years in some areas. The United States has had a number of high profile cases dealing with laptop searches at border crossings with the methodologies exposed painting a haphazard picture of their practices.

Modern digital forensic tools, both commercial and open source, employ a single user paradigm wherein the evidence, once acquired, is worked on in an individual workstation context. The prevailing approaches are also pointed toward analysis in the lab rather than in the field. This design fails to apply modern advances in distributed and high performance computing to speed analysis, and is overly reliant on static features rather than allowing for dynamic insertion of automated processes into the analysis.

Reconstruction of the forensic process is proposed as a staged approach using a pipelined system for the automated portion of analysis. The proposed process treats forensic data as a server centered, rather than workstation centered, resource. By centralizing control over the forensic data, information can be used by multiple systems in tandem. In the example presented, a triage stage takes files from the loaded disk image and distributes them to processing nodes which push the read data through an expandable pipeline of automated processes including file hashing, text extraction, indexing, file type identification, etc. Experimental results show a significant decrease in processing time versus the traditional single station approach.

This distributed approach also allows for node specialization for dealing with proprietary data types requiring specialized or in-depth second pass processing such as extracting individual emails from email stores or performing cross analysis of multiple systems for similarities. Post-triage stages can continue using automated processing while making previous stage data available for one or more analysts to examine allowing preliminary reports to be generated before the data is completely processed. Likewise, limited preliminary analysis can be performed in the field during acquisition or initial inspection with that information integrated with the triage stage.

An initial overview will be presented of the prototype "Black Friar" framework which implements the staged approach to forensic analysis with performance results. Results will be followed by examination of the future development road map as an implementation of the staged forensic approach with special emphasis placed on the flexibility open source frameworks for forensics provide for analysts to integrate new tools into the process. After becoming "production ready" Black Friar will be available as an open source digital forensic tool.

It is recommended attendees be familiar with current generation digital forensic practices, have a working knowledge of file systems and common file types, and some understanding of distributed computing, distributed file systems, and general concepts in high performance computing.

Digital Forensic Analysis, Distributed Processing, Digital Forensic Frameworks

** Presenting Author*