B16 Indexing the Windows® Registry for Software Detection

Alex J. Nelson, MS, 100 Bureau Drive, MS 8970, Gaithersburg, MD 20899-8970; Mary T. Laamanen, MS, NIST, 100 Bureau Drive, Gaithersburg, MD 20899; John Tebbutt, 100 Bureau Drive, STOP 8970, Gaithersburg, MD 20899-8970; and Darrell D. Long, PhD, 1156 High Street, Mailstop SOE 3, Santa Cruz, CA 95064*

After attending this presentation, attendees will learn about a technique to profile software on a Windows® computer to aid digital forensic triage. They will learn practical uses, investigative value, and structural characteristics of the Windows® Registry. They will learn a generic, information-retrieval-based measurement which can be used to attribute arbitrary components of the Registry to known software packages.

This presentation will impact the forensic science community by showing how being able to identify software usage on a computer is a boon to forensic triage. By detecting the software used on a computer, an investigator can receive key, rapidly actionable information, such as signs of malware or anti-forensic utilities. Searching for software presence, usage, and removal using the Windows® Registry allows one to translate what were purely machine-level artifacts back into human actions, illustrating the story of computers of otherwise unknown provenance.

The Windows® Registry is a central store for Windows® systems, recording configuration and system state. Forensic investigators find it a key resource in identifying system uses and reading specific records to support hypotheses about what a computer was used to do.

Using a document search model, the Windows® Registry's extensive namespace allows one to identify software histories associated with a computer—including applications installed, run, and/or removed—with a measurable certainty.

The Registry is a data structure analogous to a specialized file system. It contains a hierarchical namespace used to store configuration values varying in size from as small as a byte to values in the kilobyte range. This namespace acts much like a file system, where its analogies to "files" and "directories" can be generally referred to as "cells." With baseline cell tallies of one hundred thousand in Windows® XP to almost a half-million in Windows® 8, there are many locations in which telltale signs of software activity can be left behind.

One strategy to record software effects is to install, use, and uninstall applications on virtual Windows systems by snapshotting virtual machines—a process referred to as "diskprinting" in the spirit of fingerprinting—and enumerating the differences. Differences in the Registry show the particular cells affected in each step taken, and these cells can be grouped into change sets. Then, a well-known information retrieval technique, document search, identifies applications likely to have affected arbitrary Windows® systems, along with the actions those applications took. Comparing the cells affected in each change set shows the distinctness of each as an in-Registry fingerprint. Distinct fingerprints enables one to identify, from an arbitrary Registry namespace, the signs of application presence, use, and removal. For instance, a standalone application, run not by installing but instead from a thumb drive, could show signs of the browser having run while signs of installation would be absent.

Registry cell paths are distinct to the combination of the operating system, application, and basic user action. A single application's installation, use, and uninstallation was observed to affect over fifty thousand cells in the Registry. Without resorting to analysis of memory or deleted content, the document search approach was able to identify software used in a research scenario that recorded real computers' state. Applications that had been diskprinted were recognized in the scenario, even though the version diskprinted was not the precise version used in the scenario.

The Windows® Registry is a file system, configuration store, and log, often a critical source in forensic investigations. This presentation shows that by looking not at the data stored in the Registry, but instead at the structure of this data, the software history of a computer can be learned. Some software effects in the Registry can be detected based on observing behaviors of an older version of the software. This implies that significant Registry patterns are preserved as products evolve, much like a genetic lineage.

Being able to identify software usage on a computer is a boon to forensic triage. By detecting the software used on a computer, an investigator can receive key, rapidly actionable information, such as signs of malware or anti-forensic utilities. Searching for software presence, usage and removal using the Windows Registry allows one to translate what were purely machine-level artifacts back into human actions, illustrating the story of computers of otherwise unknown provenance.

** Presenting Author*

Windows® Registry, Information Retrieval, Differential Analysis