An Application Footprint Reference Set:

Tracking the Lifetime of Software

John Tebbutt
National Software Reference Library
National Institute of Standards and Technology

The National Software Reference Library

RDS 2.35 (December 2011):

13,500+ Products

74,555,829 Files

22,502,929 Unique SHA-1 Values

Motivation

Gather data on the specific effects of individual software packages on a system over the software's lifetime.

Provide investigators with new reference data from “living” software from the NSRL collection.

Question:

What changes occur in a system when a piece of software is

- Installed?
- Executed?
- Uninstalled/Deleted?

Application Footprint

We can measure the what, where, when and how:

- Nature of changes
- Location of changes
- Stage in application “life cycle”
- Actions causing changes

Nature of Changes

Filesystem (file hashes, MAC times, etc)

- Executables
- Libraries
- Documents/Images/Multimedia
- etc.

Configuration information

- Registry

Memory mapping information

- Portable Executable mappings

Location of Changes

Windows Registry

System RAM

Other locations (future)

Stage in Software Lifecycle

Depends on the package. At least:

- Post-installation
- Post-activation (if necess.)
- During Execution (possibly several)
- Post-execution
- Post-uninstallation/Deletion

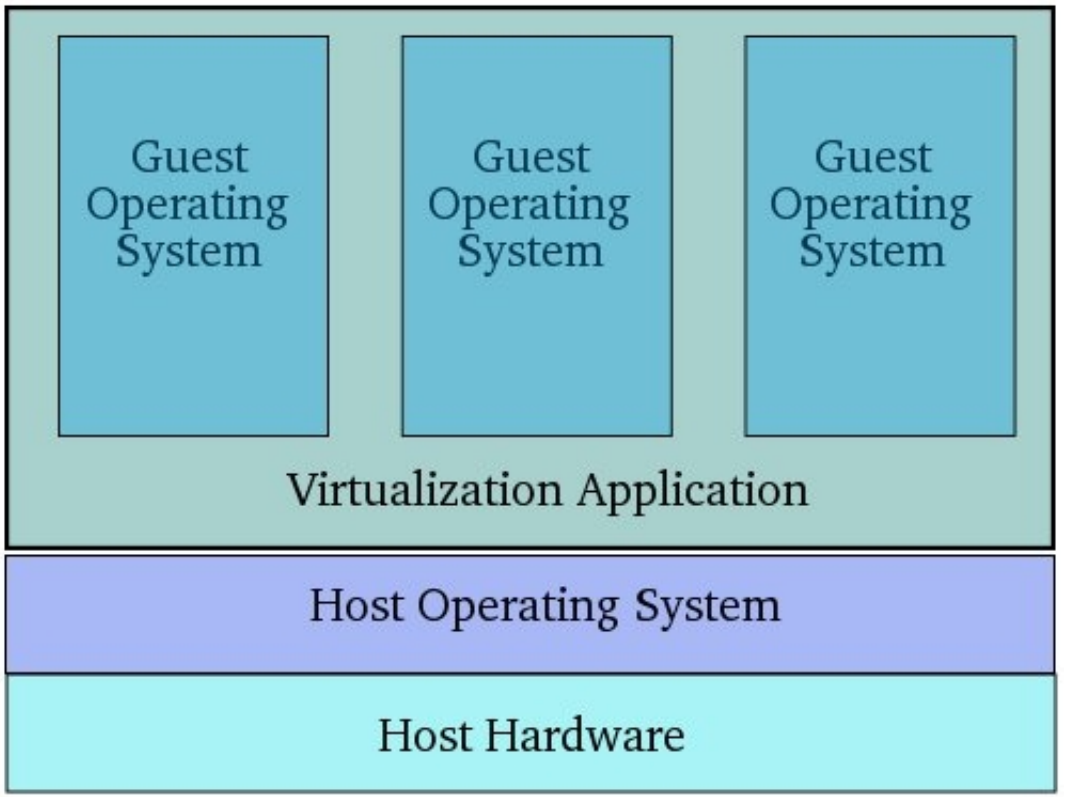
Actions Causing Changes

Particular actions during software execution may result in specific changes

e.g. visiting a web page in a browser will almost certainly add elements of the page to the browser cache. However there may be other less obvious changes...

Method

Virtual Machine Installation



Advantages

VM state can be captured at any time
- VM may be “paused” or “suspended”

VM is “frozen” as a set of files
- Hard drive, RAM contents, etc

Can be copied off for external processing...

...and saved for future reference

Sliced Footprints

Suspend VM after each action to record the Action's effects

Capture the lifecycle of an application as a series of suspended VMs, copied off and saved

A “slice” is a collection of metadata computed from a suspended VM

- file hashes, registry dumps, RAM contents
- etc

Application Footprint is the sequence of slices derived from the stored VMs

Capturing Application Footprints

Default set of slices for each Footprint is:

After installation
After activation/registration
During execution
-The application is started, left for a short time, and the slice taken
After execution
After uninstallation
After restart/shutdown of the VM
- to capture any housekeeping artifacts

How Do They Do It?

By hand.

Need to record:

- Information about the application's behavior/ state at the time of the slice
- Which slice is this?
  - # During execution, many slices may be created. What happened before each?
  - # Installation and activation may be separate events
- Unexpected behavior

Example

For each software package:

LOOP:

Retrieve a baseline VM image for the operating system.
Install the package.
Save VM .
Launch the software. Wait a short time.
Save VM.
Quit software.
Save VM.
Uninstall s/w.
Save VM.
Shutdown/restart VM.
Save VM.

END

Application Footprint Data

NSRL data on the footprinted package

- name, version, manufacturer, etc.
- date/time stamp information of the Footprint's creation (installation, execution, etc.)

Virtual machine metadata

VM software name and version

Application Footprint Data, contd.

Host data:

- operating system name/version/patch level
- hardware information

Description of each slice, and the stage in the software's life cycle that it represents

Sequence of slices recording the application lifecycle

Application Footprint Data Format

XML-based.

Schema/DTD TBD at time of submission

- Various efforts are underway to consolidate a representation of digital forensic data

Possible candidate is CybOX (Cyber Observable Expression)

- http://cybox.mitre.org/

Thank You

Questions?

John Tebbutt
National Software Reference Library
NIST
Gaithersburg, MD 20899
tebbutt@nist.gov