Efficient Tracking, Logging, and Blocking of Accesses to Digital Objects

Cyber Security Division
2013 Principal Investigators’ Meeting

September 16th, 2013

Fabian Monrose
University of North Carolina at Chapel Hill
fabian@cs.unc.edu
919-962-1763

Michael Bailey, University of Michigan
Charles Schmitt, Renaissance Computing Institute
Team Profile

- Fabian Monrose: **Professor** of Computer Science at University of North Carolina at Chapel Hill. Prior to joining UNC, he was an Associate Professor at John Hopkins, and a founding member of their Information Security Institute.

- Michael Bailey: **Research Professor** at University of Michigan. He currently directs and contributes to research on the security and availability of complex distributed systems. Prior to working at the university, he was Director of Engineering at Arbor Networks, and a programmer at both Amoco Corporation and Andersen Consulting.

- Charles Schmitt: **Director of Informatics**. He provides technical leadership and management for biological and medical science related projects RENCI. Prior to joining RENCI, he was the senior computer scientist at BD Technologies, where he assisted in software development and bioinformatics support for programs in medical diagnostics and genomics.
Researchers and practitioners routinely require access to large corpora of sensitive data for a wide host of scientific activities.

Cleanrooms are secure, but cumbersome to use.

Do not meet today’s needs.
Virtual data enclaves have emerged as solutions for hosting sensitive data (e.g., medical records, call meta-data, network traffic, etc.)

Yet, few solutions provide a secure environment for reducing the risks of unauthorized access to, and loss of, information hosted in these enclaves.
Goals

- To design and implement techniques for tracking the chain of custody of sensitive data hosted in enclaves.
- Particularly interested in an evaluation within the **Secure Research Workspace** at the Renaissance Computing Institute (RENCI).
Goals

**Deliverables:** An object tracking platform that will enable DHS and its customers to *(a)* identify and authenticate access to digital objects that originate from disk *(b)* track accesses to these objects on disks and in memory and *(c)* track changes to these objects via a provenance-aware audit trail.
Approach

- Monitoring framework implemented within a Hypervisor
  - extends TrailOfBytes prototype

- Spans three layers: storage, memory, and system-call modules

- Key idea is in monitoring access to physical memory when data is first loaded from a datastore

- Semantic linkages captured in a provenance-aware filesystem
Benefits

- Enhance state of the art in digital provenance in virtual data enclaves

  - Improved data provenance representation: A rich interface for managing and mining the recorded information, thereby providing deeper insights into how objects were manipulated

  - Limiting breaches: Capabilities to not only record, but also to deny, unauthorized accesses or transfer of data from datastores for which provenance tracking has been enabled
Several research prototypes or government-led efforts for hosting sensitive data.

But, no commercial ventures (that we are aware of).
### Current Status

<table>
<thead>
<tr>
<th>Task Description</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design and Implementation</td>
<td></td>
</tr>
<tr>
<td>Ontological Provenance Model (Task 4.1)</td>
<td></td>
</tr>
<tr>
<td>Development</td>
<td></td>
</tr>
<tr>
<td>KVM port (Task 4.2)</td>
<td></td>
</tr>
<tr>
<td>Dynamic Provisioning (Task 4.3)</td>
<td></td>
</tr>
<tr>
<td>Provenance Tracking (Task 4.4)</td>
<td></td>
</tr>
<tr>
<td>Capturing Semantic Linkages (Task 4.5)</td>
<td></td>
</tr>
<tr>
<td>Tamper Resistant Audit Trail (Task 4.6)</td>
<td></td>
</tr>
<tr>
<td>Origin Attestation (Task 4.7)</td>
<td></td>
</tr>
<tr>
<td>Lightweight Process Snapshots (Task 4.8)</td>
<td></td>
</tr>
<tr>
<td>Multi-host Tracking (Task 4.9)</td>
<td></td>
</tr>
<tr>
<td>Selective Blocking (Task 4.10)</td>
<td></td>
</tr>
<tr>
<td>Testing, Deployment &amp; Outreach</td>
<td></td>
</tr>
<tr>
<td>Deployment and Case Study (§4.4.1)</td>
<td></td>
</tr>
<tr>
<td>Software Release</td>
<td></td>
</tr>
<tr>
<td>Publications &amp; Documentation</td>
<td></td>
</tr>
</tbody>
</table>

Milestones: Origin Attestation and fast process snapshots

- **Attestation**: Implemented a technique for accurately determining what applications are accessing a monitored object
  
  - use information gathered from the guest VM and target process’ image (as available **within the hypervisor**)

- **Multi-host process snapshots**: Implemented a technique to support process snapshots and fast transfers (at **1255 MB/s**)

Milestones:
(Provenance tracking and selective blocking)

Explored several **within-guest OS** techniques for interrupting a running application:

1. inject code
2. inject new thread
3. inject DLL
Milestones:
(Provenance tracking and selective blocking)

Adapted solution to a within-hypervisor mechanism that uses a system-call detection technique coupled with a code injection and process-redirection approach.
Next steps
Deployment & Evaluation

- Technical assessment: several testing criteria including
  - accuracy (e.g., auditing and logging capabilities)
  - performance impact,
  - usability,

- System-level test with one of two existing multi-institutional studies using de-identified medical data:
  - medical visualization techniques to guide a medical decision support application for pediatric patients with epilepsy
  - medical visualization techniques to guide a medical decision support application for patients with a major depressive disorder