Digital Forensic Analysis: From Low-Level Events to High-Level Actions

Imani Palmer
Department of Computer Science
University of Illinois at Urbana-Champaign

Roy Campbell
University of Illinois at Urbana-Champaign

Boris Gelfand
Los Alamos National Laboratory
Outline

Motivation

Framework

Implementation

Evaluation

Conclusion & Future Work
Digital Forensics

Technology is more intertwined in daily life leading to an increase in court cases where digital evidence is vital.

Digital forensic has grown from an obscure tradecraft to an important part of investigations.

Digital forensic tools are used by examiners within:
- Local, state and Federal law enforcement
- Military and other US government organizations
- Private industry
Digital Forensics

Solving crimes committed with computer

- phishing and bank fraud

Solving crimes against people where evidence may reside on a computer

- money laundering and child exploitation

Providing information assurance

- Ability to reconstruct the evidence left by cyber attacks
Digital Forensic Process

Collection ➔ Preservation ➔ Analysis ➔ Visualization
Digital Forensic Process

- Collection
- Preservation
- Analysis
- Visualization
<table>
<thead>
<tr>
<th>Action 1</th>
<th>Action 2</th>
<th>Action 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Certifications</td>
<td>Legal</td>
<td>Standards</td>
</tr>
<tr>
<td>Google Search History</td>
<td>Chat logs</td>
<td>Internet Activity Logs</td>
</tr>
<tr>
<td></td>
<td>Email</td>
<td>Executable Programs</td>
</tr>
<tr>
<td></td>
<td>Photos</td>
<td>Internet Protocols Address</td>
</tr>
<tr>
<td></td>
<td>Financial Asset Records</td>
<td>Address Books</td>
</tr>
<tr>
<td></td>
<td>Telephone Records</td>
<td>Maps</td>
</tr>
<tr>
<td></td>
<td>Maps</td>
<td>Movie Files</td>
</tr>
<tr>
<td></td>
<td>Images</td>
<td>Configuration Files</td>
</tr>
</tbody>
</table>
The Problem

The digital forensic investigative process is marred by its lack of knowledge, accreditation, and human bias.
Commonwealth vs. Michael Fiola

Fiola returned his laptop to his employer
Child pornography was found and Fiola was charged with the possession of child pornography
Fiola’s defense team found that the laptop contained malware that was programmed to visit multiple child pornography websites
Charges were dropped after Fiola and his family spent thousands of dollars fighting the case
Connecticut v. Amero

Elementary school substitute teacher was convicted of contributing to the delinquency of minors.

A school computer in her class displayed pop-ups from a pornographic website.

Outside investigators found the school computer was infected with spyware.

Julie Amero was able to get the conviction overturned but not before her previous life was in shambles.
Digital Forensic Analysis

Legal system relied on the examiner and digital evidence in order to achieve these convictions

Digital forensic tools were accurately

Conclusions drawn from the evidence were incorrect
Certifications

No gold standard for professional certifications

Specific vendor product certifications

Increase the fragmentation

Misguided belief there is no generic conceptual approach

Every case is unique, standards are meaningless

Vendor-Neutral Certifications
- Certified Computer Examiner
- Computer Hacking Forensic Investigator
- Certified Forensic Computer Examiner
- GIAC Certified Forensic Examiner
- GIAC Certified Forensic Analyst
- GIAC Network Forensic Analyst
- GIAC Advanced Smartphone Forensics
- GIAC Reverse Engineering Malware
- CyberSecurity Forensic Analyst
- Certified Cyber Forensics Professional

Vendor-Specific Certifications
- AccessData Certified Examiner
- EnCase Certified Examiner
- EnCase Certified eDiscovery Practitioner
Proliferation of Devices

Cases increasingly require the analysis of multiple devices
Varying data sources
Tool development
Extraction of data

- Applications, Mobile
  Devices, Wearables,
  IoT, Cloud

- Network Protocols

- Operating Systems

- Hard Drive, Memory Image,
  Permanent Storage
Standards and Certifications

Qualifications of expert witnesses

No credentials or a formal educational process

Lower courts accept qualifications based on skills and previous experience

Need for national and internationally recognized certification and standardization
Analytical Methods

- Literal string searching
- Simple pattern matching
- Indexing data to speed up searching and matching
- Hash analysis
- Logical level file reviews
My Approach
### Analysis

#### Analysis Toolkit

<table>
<thead>
<tr>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google Search History</td>
</tr>
<tr>
<td>Chat logs</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>Photos</td>
</tr>
<tr>
<td>Internet Activity Logs</td>
</tr>
<tr>
<td>Executable Programs</td>
</tr>
<tr>
<td>Internet Protocols Address</td>
</tr>
<tr>
<td>Financial Asset Records</td>
</tr>
<tr>
<td>Address Books</td>
</tr>
<tr>
<td>Telephone Records</td>
</tr>
<tr>
<td>Maps</td>
</tr>
<tr>
<td>Movie Files</td>
</tr>
<tr>
<td>Images</td>
</tr>
<tr>
<td>Configuration Files</td>
</tr>
</tbody>
</table>

![Analysis Tools - Image](image-url)
This actor took action X is supported by facts & observations with strength and quantity

Objective Analysis

Provide quantitative assessments to detect certain user actions

User Sends Email

V → W → X → Y → Z

Login → Open Internet Browser → Login to Email → Write Email → Send Email
Framework
Framework

Extract Facts
Define relationships between facts
Construct user action mappings
Identify actions
Define Relationships

Evidence A
- Evidence B
  - Evidence C
  - Evidence E
- Evidence D
  - Evidence H
- Evidence G
  - Evidence I
  - Evidence K
- Evidence J
  - Evidence L

Evidence F

Evidence H
Construct Mappings

Evidence A
Evidence B
Evidence C
Evidence D
Evidence E
Evidence F
Evidence G
Evidence H
Identify Actions

Evidence A
Evidence B
Evidence C
Evidence D
Evidence E
Evidence F
Evidence G
Evidence H

Action 1
Action 2
Action 3
Implementation
Determine sequence from audit logs

File Create

File Delete

Application

Ping localhost

Ping google.com
Audit Log Example

type=DAEMON_START msg=audit(1468507482.500:6969): auditd start, ver=2.3.2 format=raw kernel=3.13.0-91-generix auid=1000 pid=3719 subj=system_u:system_r:kernel_t:s0 res=success

type=CONFIG_CHANGE msg=audit(1468507482.600:1786): audit_backlog_limit=1024 old=1024 auid=1000 ses=1 subj=system_u:system_r:kernel_t:s0 res=1

type=CONFIG_CHANGE msg=audit(1468507482.600:1787): auid=1000 ses=1 subj=system_u:system_r:kernel_t:s0 op="add rule" key=(null) list=4 res=1

type=CONFIG_CHANGE msg=audit(1468507482.600:1788): auid=1000 ses=1 subj=system_u:system_r:kernel_t:s0 op="add rule" key=(null) list=4 res=1

type=SYSCALL msg=audit(1468507482.600:1789): arch=c000003e syscall=59 success=yes exit=0 a0=7f46bc2823c8 a1=7f46bc282368 a2=7f46bc2823a0 a3=7f46bba6ca10 items=2 ppid=3697 pid=3726 auid=1000 uid=0 gid=0 euid=0 suid=0 fsuid=0 egid=0 sgid=0 fsgid=0 tty=pts0 ses=1 comm="plymouth" exe="/bin/plymouth" subj=system_u:system_r:kernel_t:s0 key=(null)

type=EXECVE msg=audit(1468507482.600:1789): argc=2 a0="plymouth" a1="--ping"

type=CWD msg=audit(1468507482.600:1789): cwd="/"

type=PATH msg=audit(1468507482.600:1789): item=0 name="/bin/plymouth" inode=786567 dev=fd:01 mode=0100755 ouid=0 ogid=0 rdev=00:00 obj=system_u:object_r:file_t:s0 nametype=NORMAL
Convert Audit Log to a Sequence

type=DAEMON_START msg=audit(1468507482.500:6969): auditd start, ver=2.3.2 format=raw

type=CONFIG_CHANGE msg=audit(1468507482.600:1786): audit_backlog_limit=1024 old=1024 auid=1000

type=CONFIG_CHANGE msg=audit(1468507482.600:1787): auid=1000 ses=1

type=SYSCALL msg=audit(1468507482.600:1789): arch=c000003e syscall=59 success=yes exit=0

type=EXECVE msg=audit(1468507482.600:1789): argc=2 a0="plymouth" a1="--ping"

type=CWD msg=audit(1468507482.600:1789): cwd="/"

DAEMON_START = 1
CONFIG_CHANGE = 2
SYSCALL = 3
EXECVE = 4
CWD = 5
Graph Representation of Audit Log

Undirected Graph
  Relationship represents the order of commands

Weighted Graph
  Weights is based on the number of times type is repeated in the sequence

Sequence
User Action Map Construction
Graph Isomorphism

Two graphs which contain the same number of graph vertices connected in the same way are said to be isomorphic.

Two graphs and with graph vertices are said to be isomorphic if there is a permutation of such that is in the set of graph edges iff is in the set of graph edges.
Sequence Alignment

Get global and local alignments between two sequences

Global alignment finds the best concordance between all characters in two sequences

Local alignment finds just the subsequences that align the best.

Match score indicates the compatibility between an alignment of two characters in the sequences

Highly compatible characters given positive scores

Incompatible ones given negative scores or 0
User Action Maps
File Create Representation

Sequence: 12223456634566837456

Unweighted

Weighted

1 = black
2 = magenta
3 = blue
4 = red
File Delete Representation

Sequence: 12223456634566834566
Application (emacs24-nox) Representation

Sequence: 12223456634566834566345663456634566345663456634566345663456634566934566

1 = black
2 = magenta
3 = blue
4 = red

Unweighted

Weighted
Ping Localhost Representation

**Sequence:** 1223456634566837456

1 = black  
2 = magenta  
3 = blue  
4 = red
Ring google.com Representation

Weighted

1 = black
2 = magenta
3 = blue
4 = red

Unweighted

Sequence: 122234566345668374566374566934566934566
<table>
<thead>
<tr>
<th></th>
<th>DAEMON_START</th>
<th>CONFIG_CHANGE</th>
<th>SYSCALL</th>
<th>EXECVE</th>
<th>CWD</th>
<th>PATH</th>
<th>BPRM_FCAPS</th>
<th>USER_END</th>
<th>USER_START</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application</td>
<td>1</td>
<td>3</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>18</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>File Delete</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>File Create</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ping localhost</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Ping google.com</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Sequence Type Count

Count of Type for Action Sequences

- Application
- File Delete
- File Create
- Ping localhost
- Ping google.com
Evaluation
Experiments

Matching of Known Commands

Baseline

Matching of Unknown Commands

Identification of False Positives

Matching of a Combination of Both Known and Unknown Commands

Identification of False Negatives

Identification of False Positives
Known Commands Representation

Sequence: 1222345663456683456634566345663456634566345663456634566374566934566

Commands
- File Create
- Run Application
- File Delete

Unweighted
Unknown Commands Representation

Sequence: 122283456634566345663456634566345663456634566345663456634566345663456634566374566934566

Commands
Wget
Echo

Unweighted
Combine Commands - Sequence - Graphs

Sequence: 1222345663456683456634566345663456634566345663456634566345663745669345669345

Commands
Wget
Run Application
Remove File

Unweighted
<table>
<thead>
<tr>
<th></th>
<th>Known_SEQ</th>
<th>Unknown_SEQ</th>
<th>Combine_SEQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>app_log</td>
<td>True</td>
<td>False</td>
<td>True</td>
</tr>
<tr>
<td>file_create</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>file_delete</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>ping_localhost</td>
<td>False</td>
<td>False</td>
<td>False</td>
</tr>
<tr>
<td>ping_google</td>
<td>True</td>
<td>False</td>
<td>True</td>
</tr>
</tbody>
</table>
## Sequence Alignment Results

<table>
<thead>
<tr>
<th></th>
<th>Known</th>
<th>Unknown</th>
<th>Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run Application</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>File Create</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>File Delete</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
<tr>
<td>Ping Google</td>
<td>FALSE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Ping Localhost</td>
<td>TRUE</td>
<td>FALSE</td>
<td>TRUE</td>
</tr>
</tbody>
</table>
Next Steps in Implementation
Live Forensics

Many elements of a computer’s state are kept in volatile memory

- Running processes
- Open network connections
- Browser sessions
- Mounted encrypted disks
Leverages DRAM memory remanence to preserve the state of the running operating system across a state-preserving reboot which recovers the existing OS without going through the full boot-up process.

Gain complete control over the system and perform taint-free forensic analysis using well-grounded introspection techniques from the virtual machine and simulation community.

To maintain fidelity, it operates exclusively in 125 KB of unused legacy conventional memory and does not taint the contents of extended memory.
Conclusion

Possible to aid in the knowledge of digital forensic investigative process

Current Challenges:

Need a greater representation

Logs are not enough

Network connections

Executables

Determine a more robust evaluation methods

Comparison of information retrieval overheads

Time required to execute a search
Future Work

Apply methods with memory analysis

Forenscope Module Cloner to retrieve volatile data

Create an additional module for Forenscope to outline user action maps

Predict actions from Sequence Alignment and Graph Isomorphism
References