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

http://threatgeek.typepad.com/.a/6a0147e41f3c0a970b017d42bd6969970c-pi

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

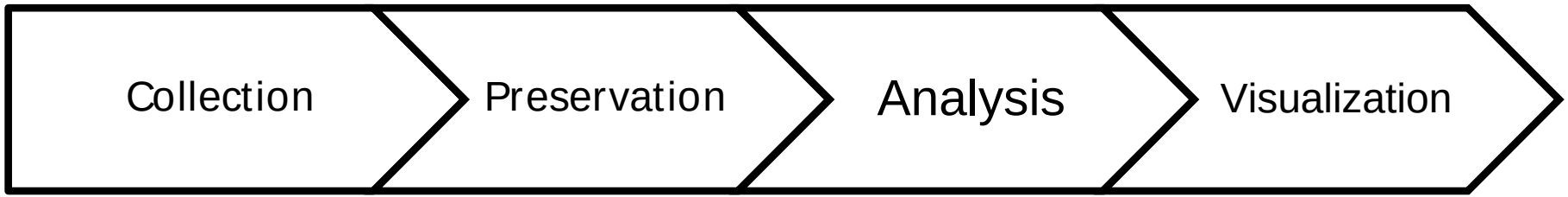
Solving crimes against people where evidence may reside on a computer

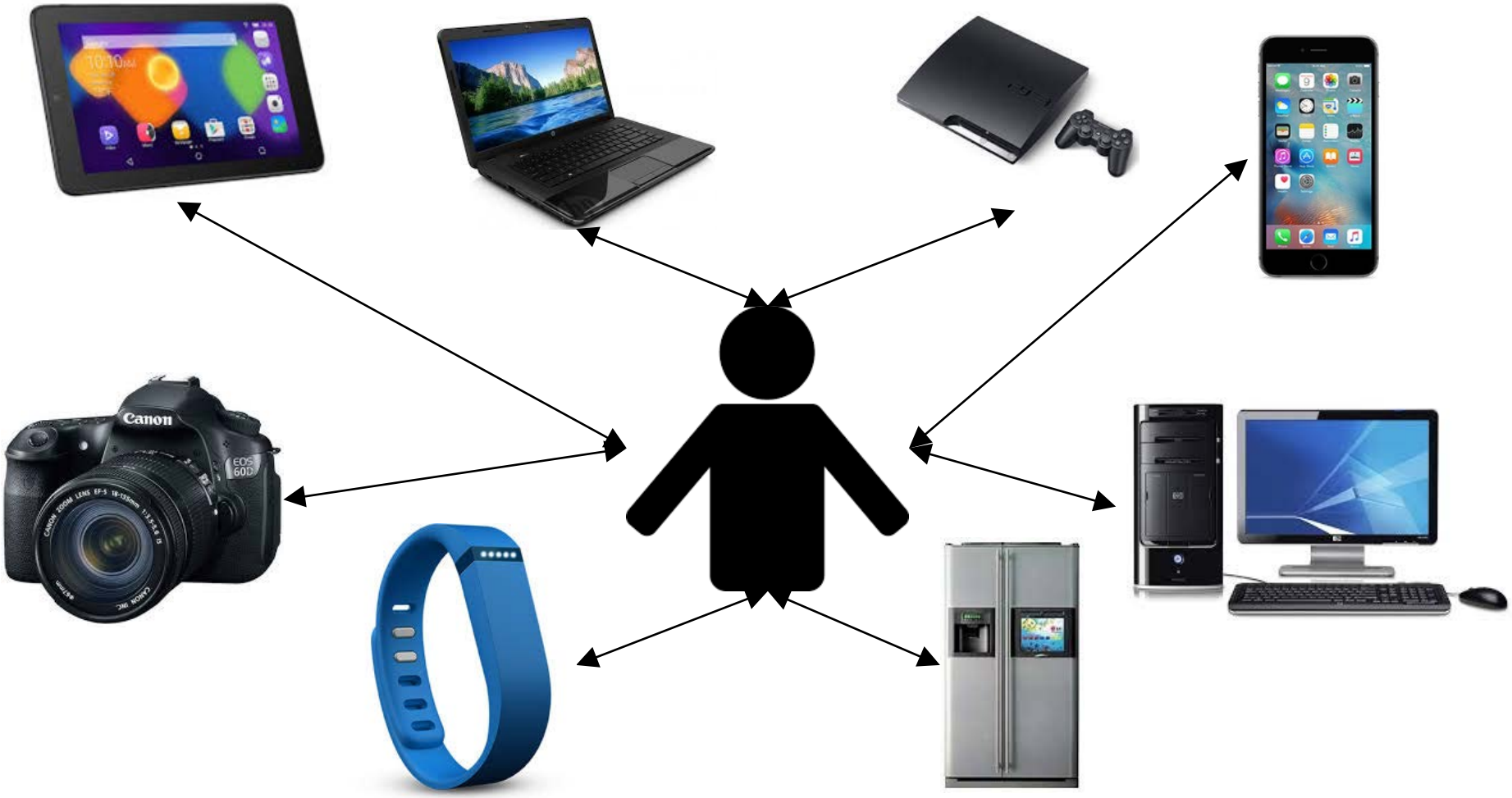
- money laundering and child exploitation

Providing information assurance

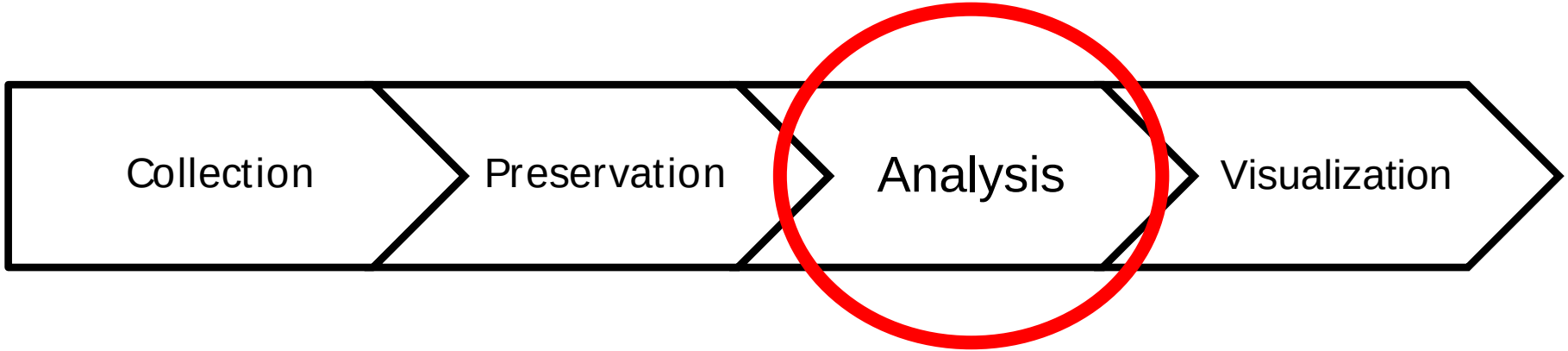
- Ability to reconstruct the evidence left by cyber attacks

Digital Forensic Process

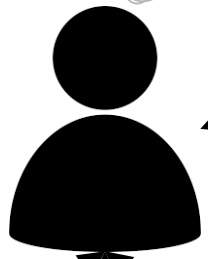



Canon
EOS 60D

Digital Forensic Process



Action 1
Action 2
Action 3



Certifications

Legal

Standards

- Google Search History
- Chat logs
- Email
- Photos
- Internet Activity Logs
- Executable Programs
- Internet Protocols Address
- Financial Asset Records
- Address Books
- Telephone Records
- Maps
- Movie Files
- Images
- Configuration Files

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-Specific Certifications

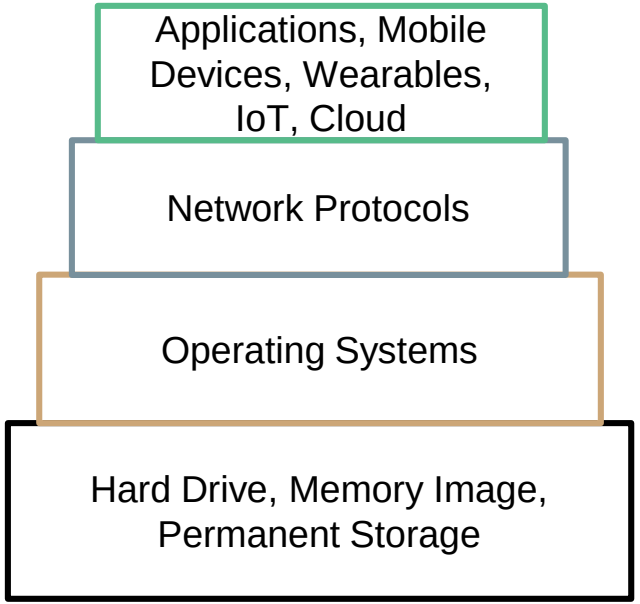
AccessData Certified Examiner
EnCase Certified Examiner
EnCase Certified eDiscovery Practitioner

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

Proliferation of Devices

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



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

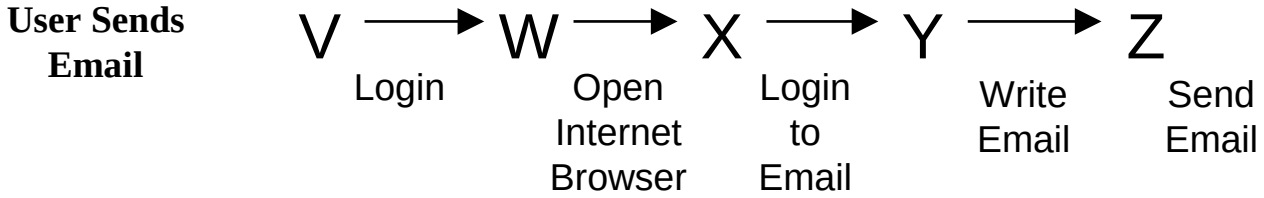
- Google Search History
- Chat logs
- Email
- Photos
- Internet Activity Logs
- Executable Programs
- Internet Protocols Address
- Financial Asset Records
- Address Books
- Telephone Records
- Maps
- Movie Files
- Images
- Configuration Files

Analysis Toolkit

This actor took action X is supported by facts & observations with strength and quantity

Objective Analysis

Provide quantitative assessments to detect certain user actions



Framework

Framework

Extract Facts

Define relationships between facts

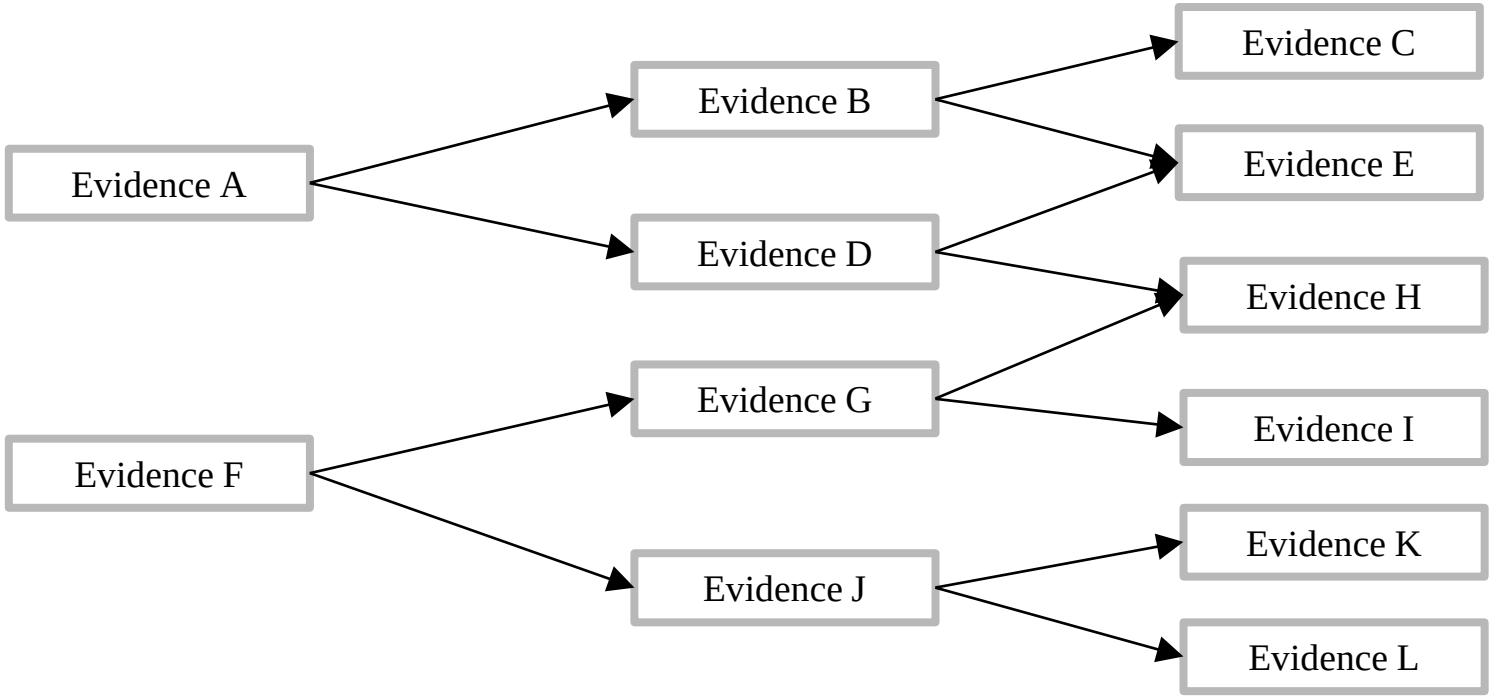
Construct user action mappings

Identify actions

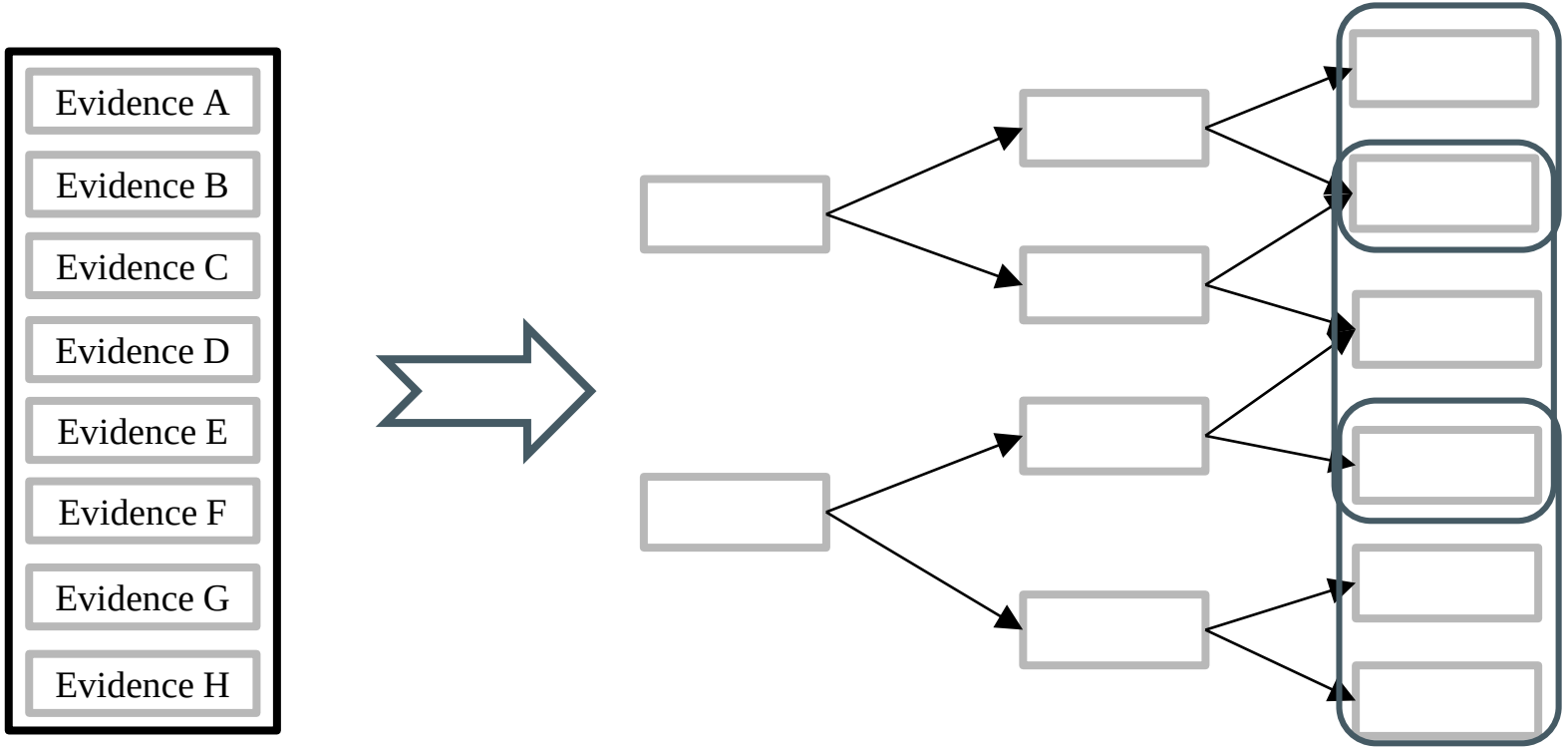
Extract Facts

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

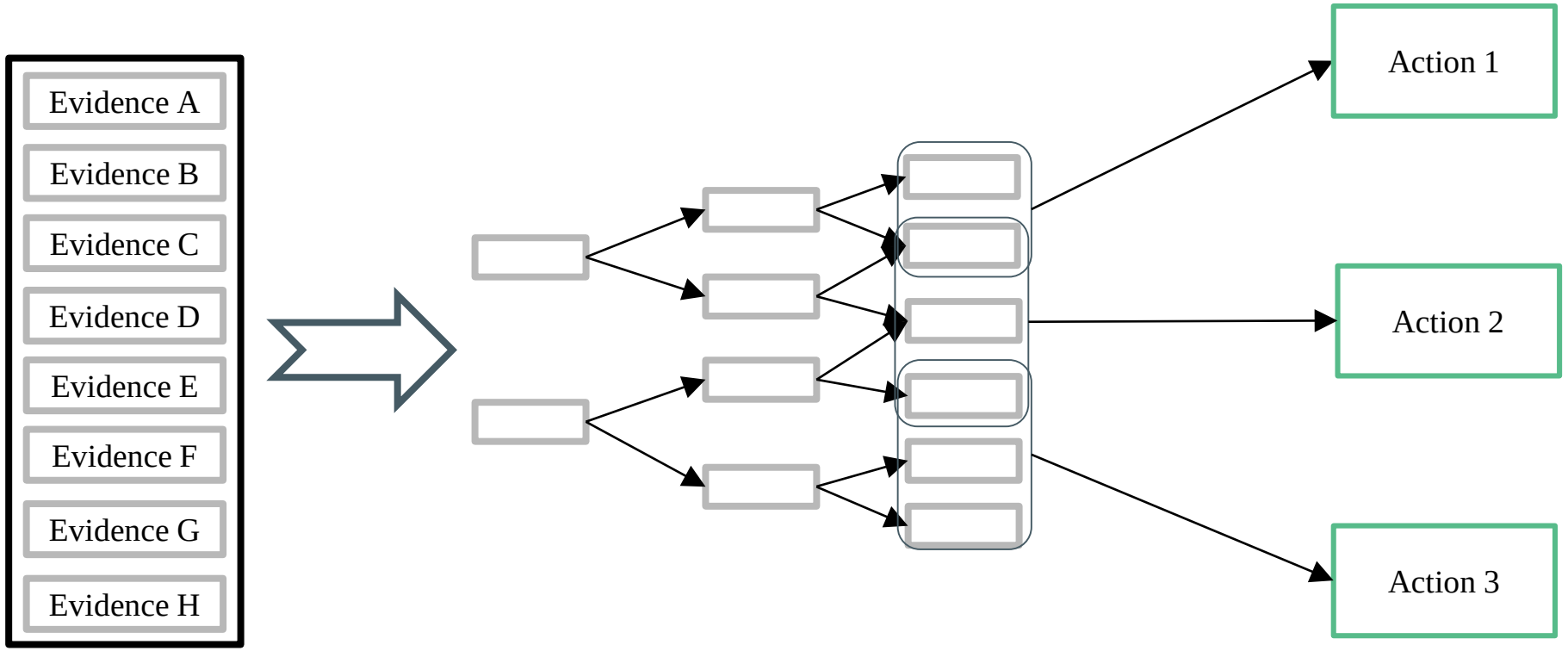
Define Relationships



Construct Mappings



Identify Actions



Implementation

Prototype

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-generic 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

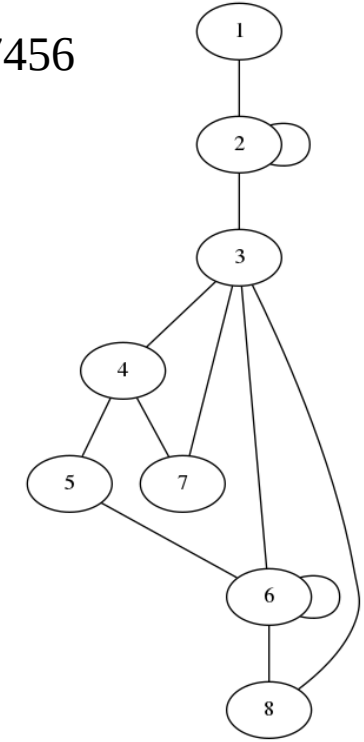
Sequence Alignment

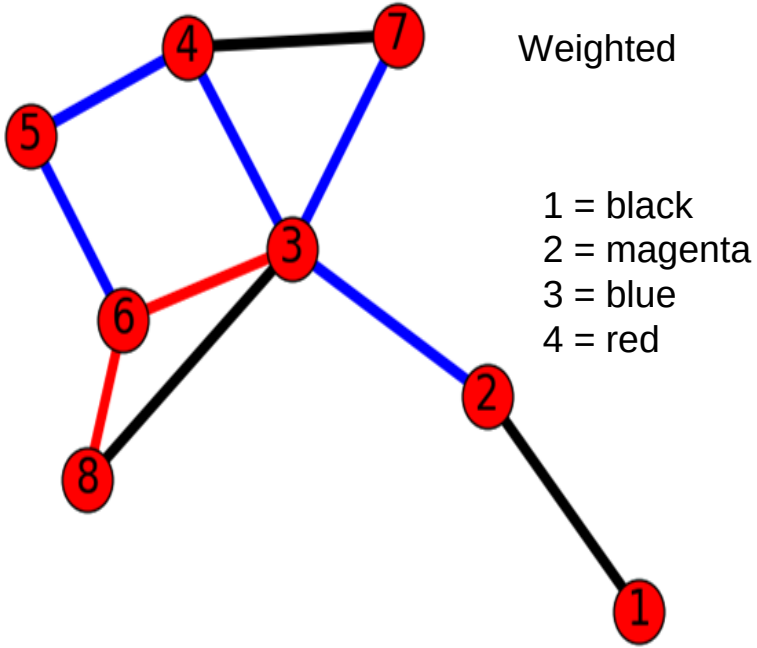
- Get global and local alignments between two sequences
  - Global alignment finds the best concordance between all characters in two sequences
  - Local alignment finds just the subsequences that align the best.
- Match score indicates the compatibility between an alignment of two characters in the sequences
  - Highly compatible characters given positive scores
  - Incompatible ones given negative scores or 0

User Action Maps

File Create Representation

Sequence: 12223456634566837456





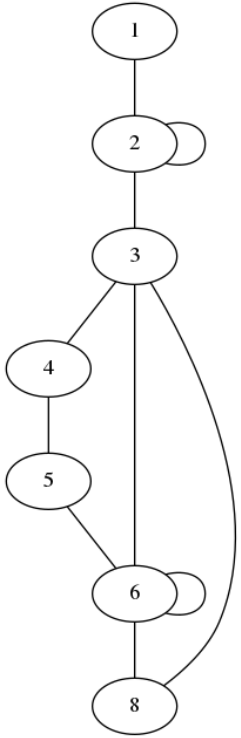
Weighted

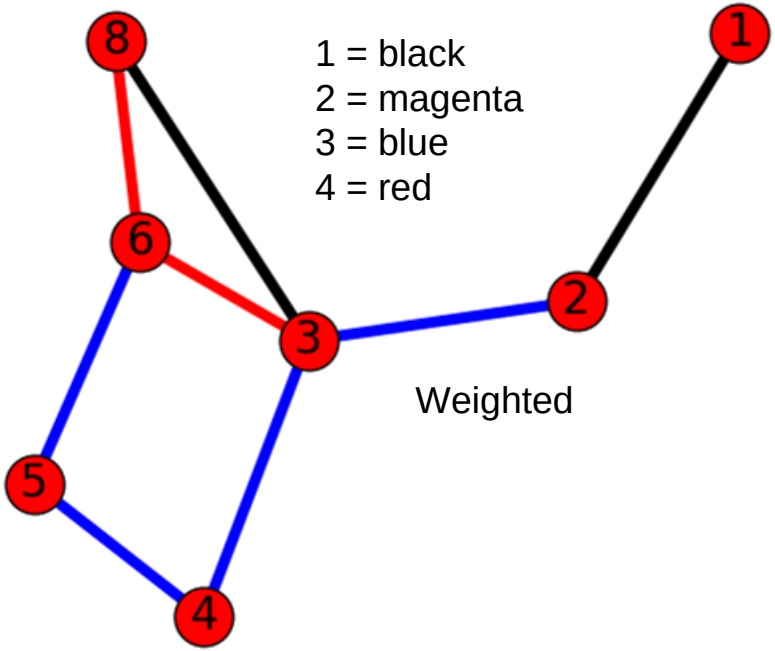
1 = black
2 = magenta
3 = blue
4 = red

Unweighted

File Delete Representation

Sequence: 12223456634566834566



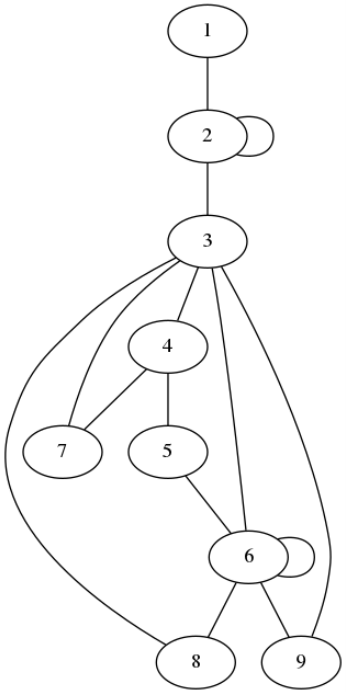


Weighted

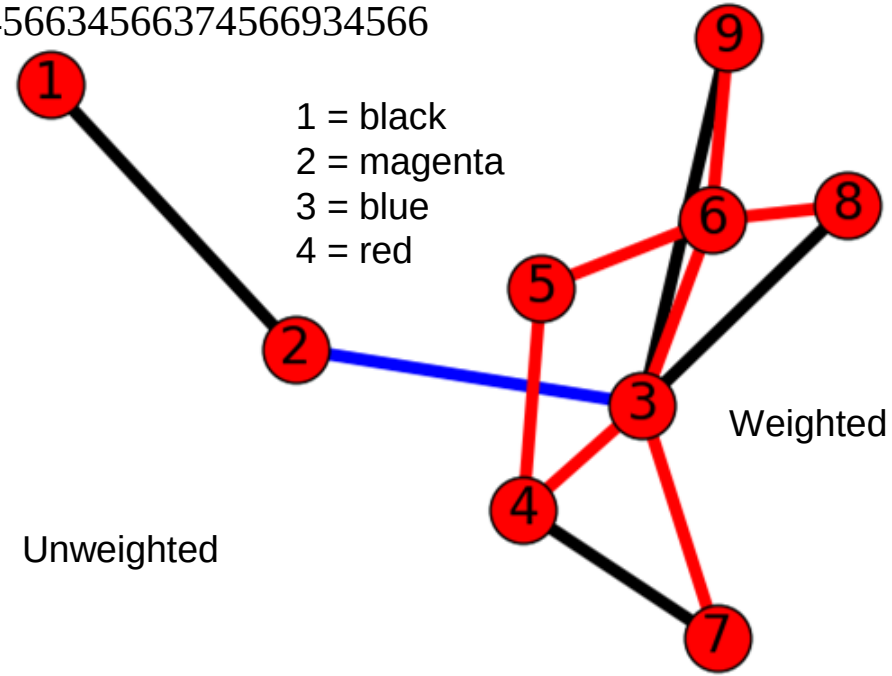
Unweighted

Application (emacs24-nox) Representation

Sequence: 1222345663456683456634566345663456634566374566934566



Unweighted

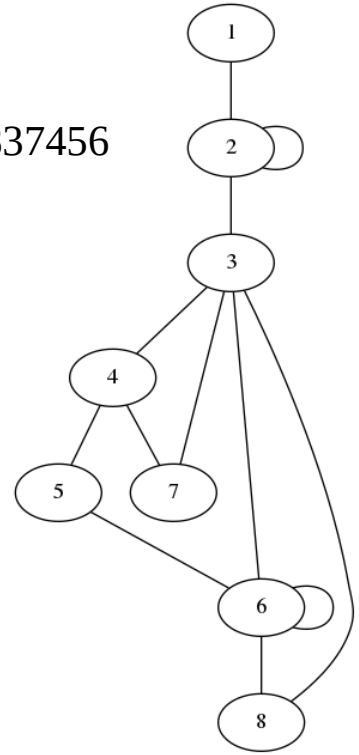


Weighted

Ping Localhost Representation

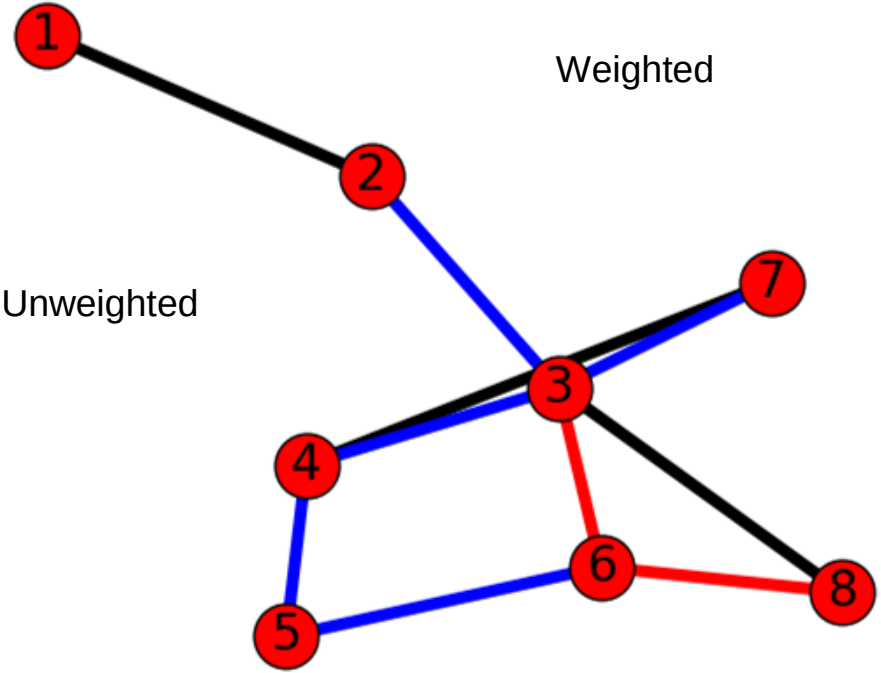
1 = black
2 = magenta
3 = blue
4 = red

Sequence: 12223456634566837456

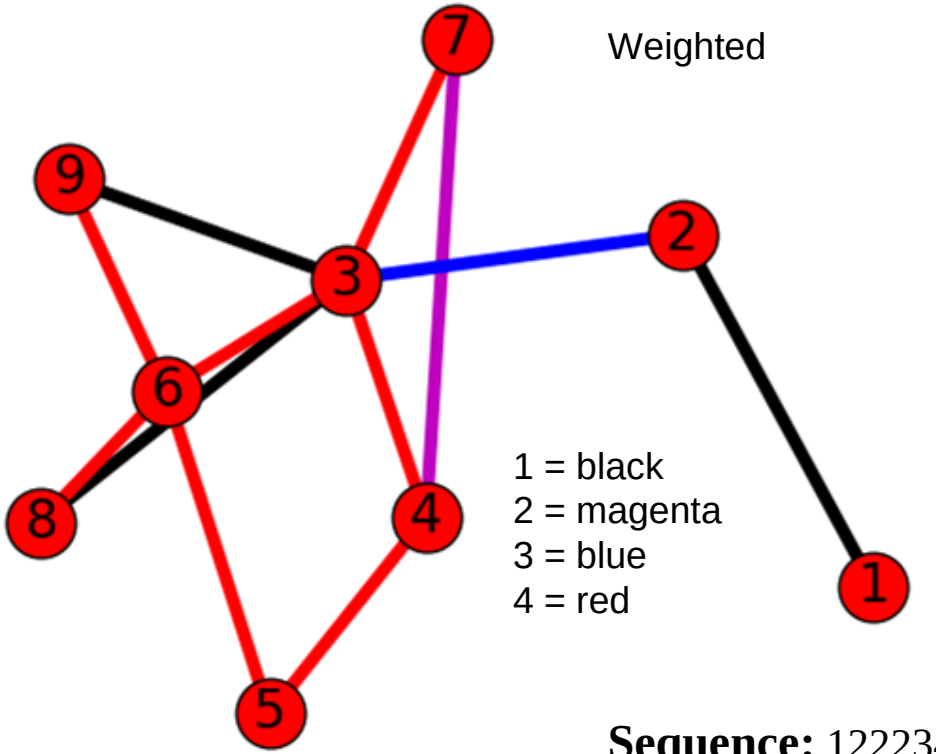


Unweighted

Weighted

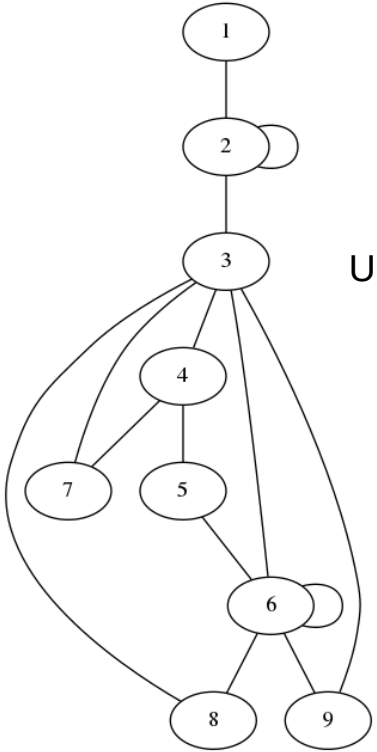


Ping google.com Representation



Weighted

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



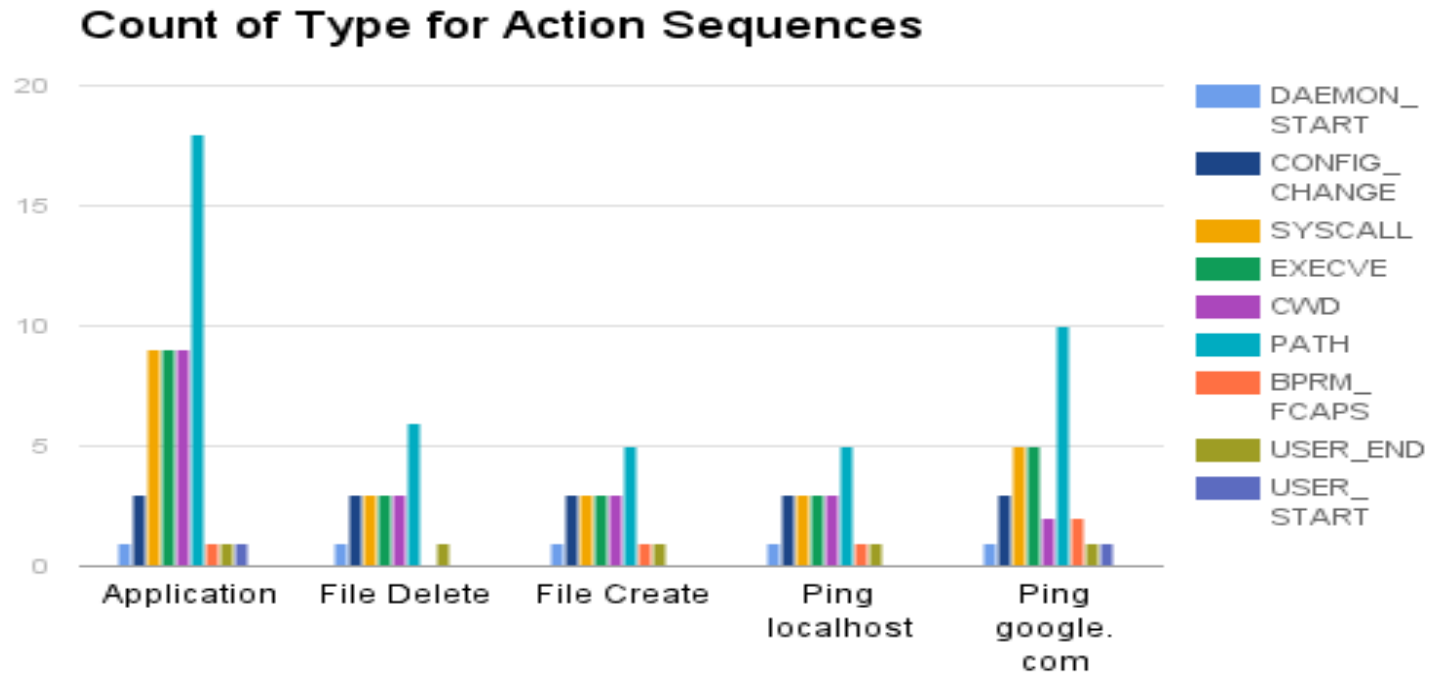
Unweighted

Sequence: 1222345663456683745663745669345 66

Comparison of Sequences

	DAEMON_START	CONFIG_CHANGE	SYSCALL	EXECVE	CWD	PATH	BPRM_FCAPS	USER_END	USER_START
Application	1	3	9	9	9	18	1	1	1
File Delete	1	3	3	3	3	6	0	1	0
File Create	**1**	**3**	**3**	**3**	**3**	**5**	**1**	**1**	**0**
Ping localhost	**1**	**3**	**3**	**3**	**3**	**5**	**1**	**1**	**0**
Ping google.com	1	3	5	5	2	10	2	1	1

Sequence Type Count



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

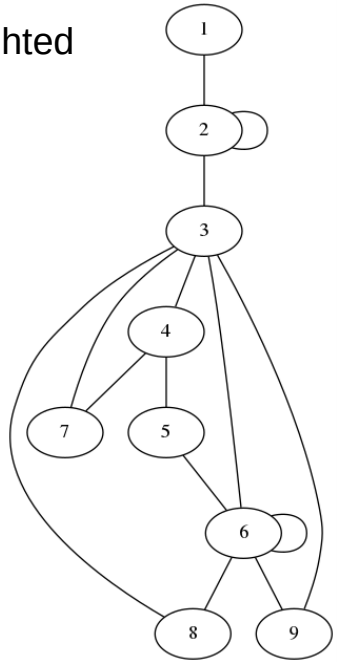
- Identification of False Negatives
- Identification of False Positives

Known Commands Representation

Sequence: 1222345663456683456634566345663456634566345663456374566934566

Commands
File Create
Run Application
File Delete

Unweighted

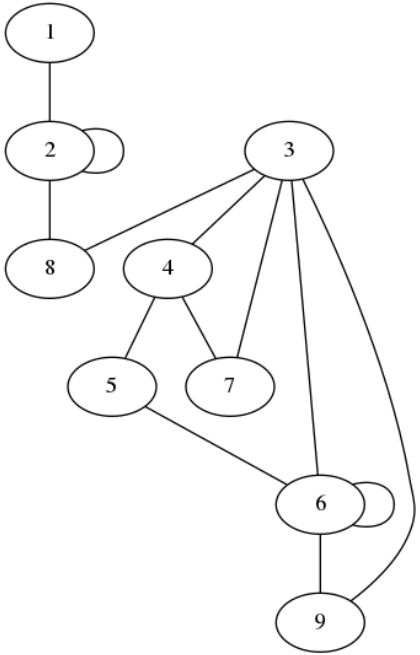


Unknown Commands Representation

Sequence: 1222834566345663456634566345663456634566345663456634566345663456
6345663456634566345663745669345 66

Commands
Wget
Echo

Unweighted

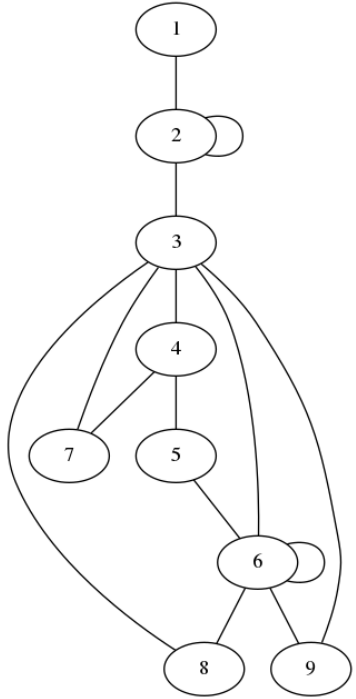


Combine Commands - Sequence - Graphs

Sequence: 1222345663456683456634566345663456634566345663456634566345663745669345

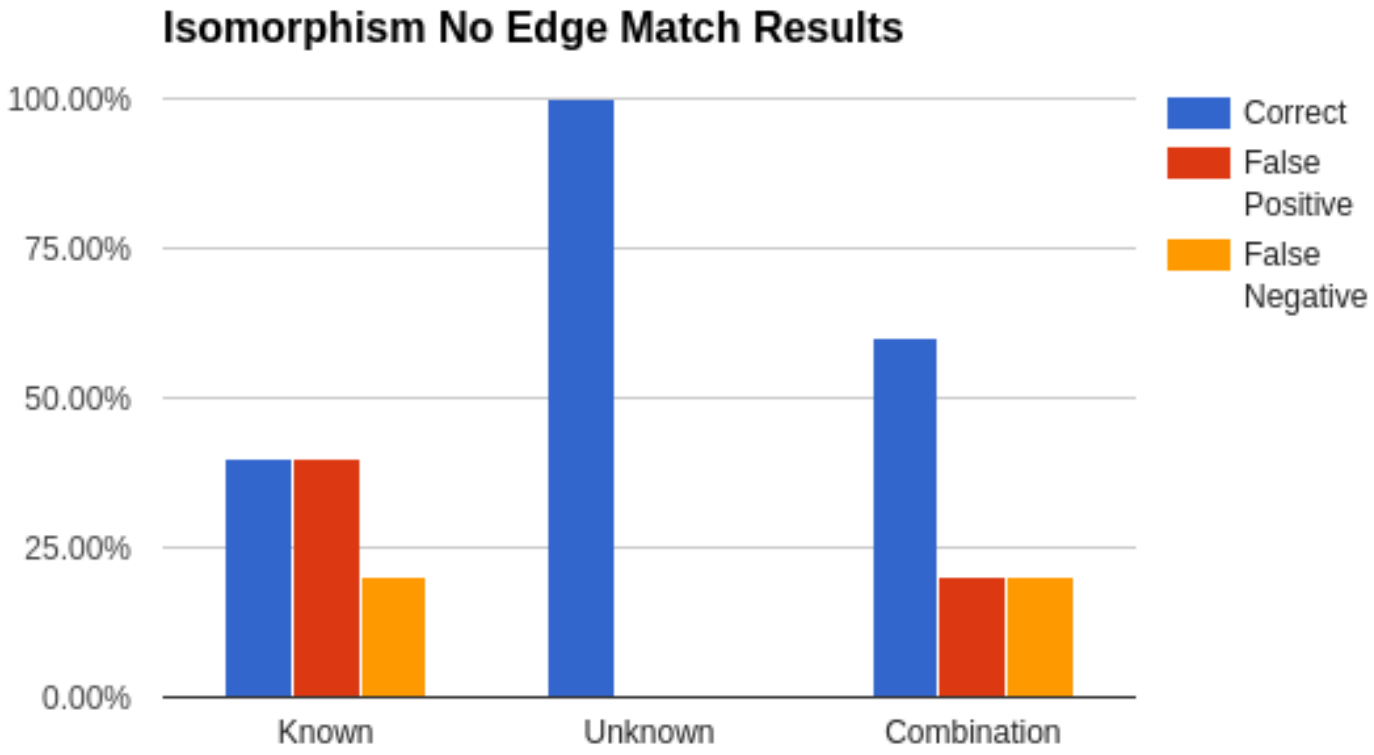
Commands
Wget
Run Application
Remove File

Unweighted



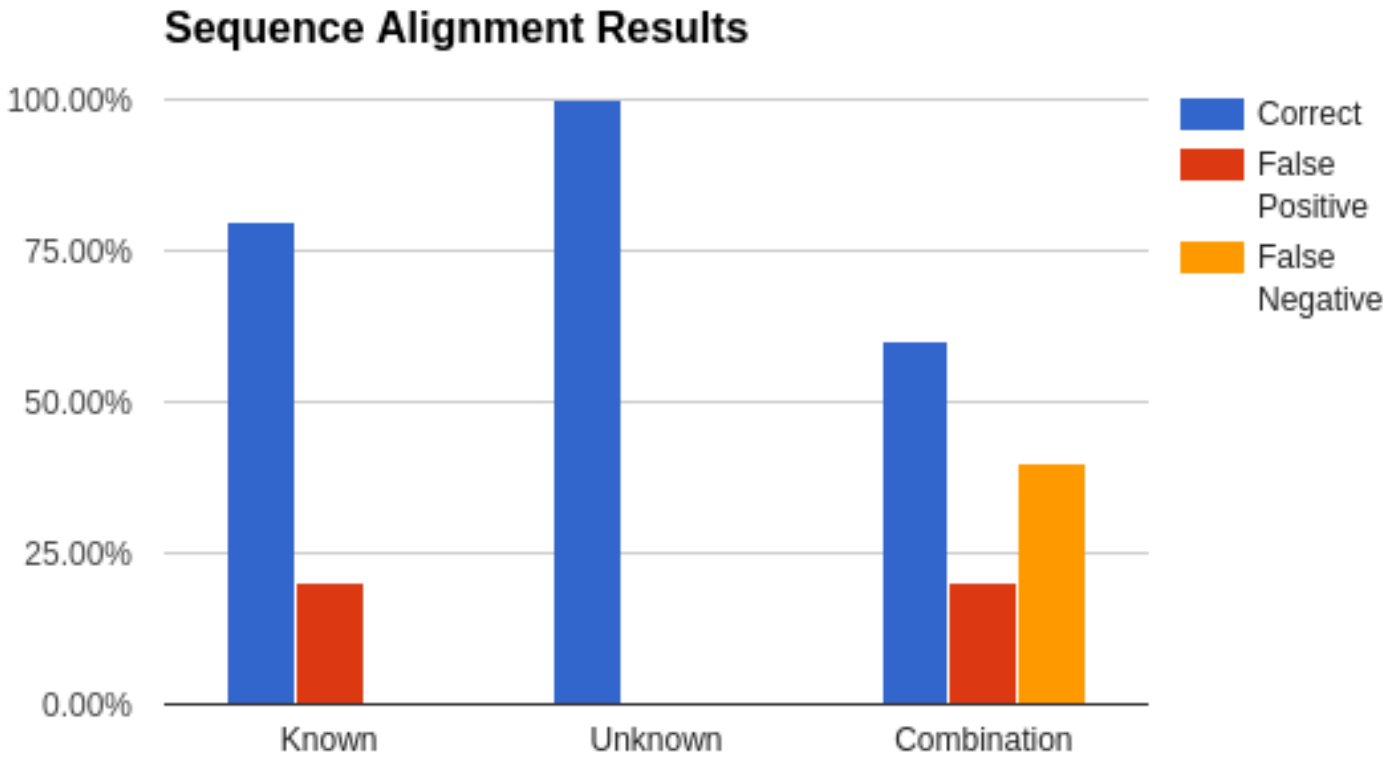
Isomorphism No Edge Match Results

	Known_Seq	Unknown_Seq	Combine_Seq
app_log	True	False	True
file_create	**False**	False	False
file_delete	**False**	False	**False**
ping_localhost	False	False	False
ping_google	**True**	False	**True**


Isomorphism No Edge Match Results
100.00%
75.00%
50.00%
25.00%
0.00%
Known
Unknown
Combination
Correct
False Positive
False Negative

Sequence Alignment Results

	Known	Unknown	Combination
Run Application	TRUE	FALSE	**FALSE**
File Create	TRUE	FALSE	TRUE
File Delete	TRUE	FALSE	TRUE
Ping Google	FALSE	FALSE	**FALSE**
Ping Localhost	**TRUE**	FALSE	**TRUE**


Sequence Alignment Results
100.00%
75.00%
50.00%
25.00%
0.00%
Known
Unknown
Combination
Correct
False Positive
False Negative

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

Forenscope

- 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 tain 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

1. M. Rigby. “Child Porn Investigations may Snare the Innocent.” Online. November, 2010.
2. Schroeder, Stephen C. "How to be a digital forensic expert witness." *First International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE'05)*. IEEE, 2005.
3. http://www.digital-detective.net/digital-evidence-discrepancies-casey-anthony-trial/
4. A. Alva and B. Endicott-Popovsky. Digital evidence education in schools of law. The Journal of Digital Forensics, Security and Law: JDFSL, 7(2):75, 2012.
5. S. Peisert, M. Bishop, and K. Marzullo. Computer forensics in forensis. In Systematic Approaches to Digital Forensic Engineering, 2008. SADFE’08. Third International Workshop on, pages 102–122. IEEE, 2008.
6. Yasinsac, Alec, et al. "Computer forensics education." *IEEE Security & Privacy Magazine* 1.4 (2003): 15-23.
7. Ball, Craig. "Cross-examination of the Computer Forensics Expert." (2004).
8. Garrie, Daniel B. "Digital Forensic Evidence in the Courtroom: Understanding Content and Quality." *Nw. J. Tech. & Intell. Prop.* 12 (2014): i.
9. Meyers, Matthew, and Marc Rogers. "Computer forensics: the need for standardization and certification." *International Journal of Digital Evidence* 3.2 (2004): 1-11.
10. Garfinkel, Simson L. "Digital forensics research: The next 10 years." *digital investigation* 7 (2010): S64-S73.
11. Kessler, Gary Craig. *Judges’ awareness, understanding, and application of digital evidence*. Diss. Nova Southeastern University, 2010.

References

12. Chan, Ellick, et al. "Forenscope: a framework for live forensics." *Proceedings of the 26th Annual Computer Security Applications Conference*. ACM, 2010.
13. Chan, Ellick M. *A framework for live forensics*. Diss. University of Illinois at Urbana-Champaign, 2011.
14. Francia, Guillermo, et al. "Visualization and management of digital forensics data." *Proceedings of the 3rd annual conference on Information security curriculum development*. ACM, 2006.
15. Reith, Mark, Clint Carr, and Gregg Gunsch. "An examination of digital forensic models." *International Journal of Digital Evidence* 1.3 (2002): 1-12.
16. Beebe, Nicole. "Digital forensic research: The good, the bad and the unaddressed." *IFIP International Conference on Digital Forensics*. Springer Berlin Heidelberg, 2009.
17. Rogers, Marcus K., and Kate Seigfried. "The future of computer forensics: a needs analysis survey." *Computers & Security* 23.1 (2004): 12-16.
18. L. P. Cordella, P. Foggia, C. Sansone, M. Vento, “An Improved Algorithm for Matching Large Graphs”, 3rd IAPR-TC15 Workshop on Graph-based Representations in Pattern Recognition, Cuen, pp. 149-159, 2001.
19. http://biopython.org/DIST/docs/api/Bio.pairwise2-module.html