AYTOMATED FORENSIC ANALYSIS

2016 MSc Cyber Security Project Presentation

Anastasios Koutlis
MSc in Cyber Security
University of York

Supervisor
Prof. Howard Chivers

September
2016

Overview

- Introduction
- Existing Work
- Literature Review
- Design & Methodology
- Implementation
- Evaluation
- Future Work & Conclusion
- References

Introduction

Problem

- **Existence of a large number of artifacts in a computer that must be processed**
- **Time consuming process**
- **Danger of missing important connections**

Idea

- **Provide a tool that is capable of assisting an investigator during a forensic analysis**
- **Create rules to find these high-level events by looking into low-level artifacts**
- **Provide a visualization technique showing**
  - **Low-level events extracted**
  - **High-level events created**
  - **Their connections**
- **Evaluate the process based on test events provided and evaluate the efficiency of the developed framework**

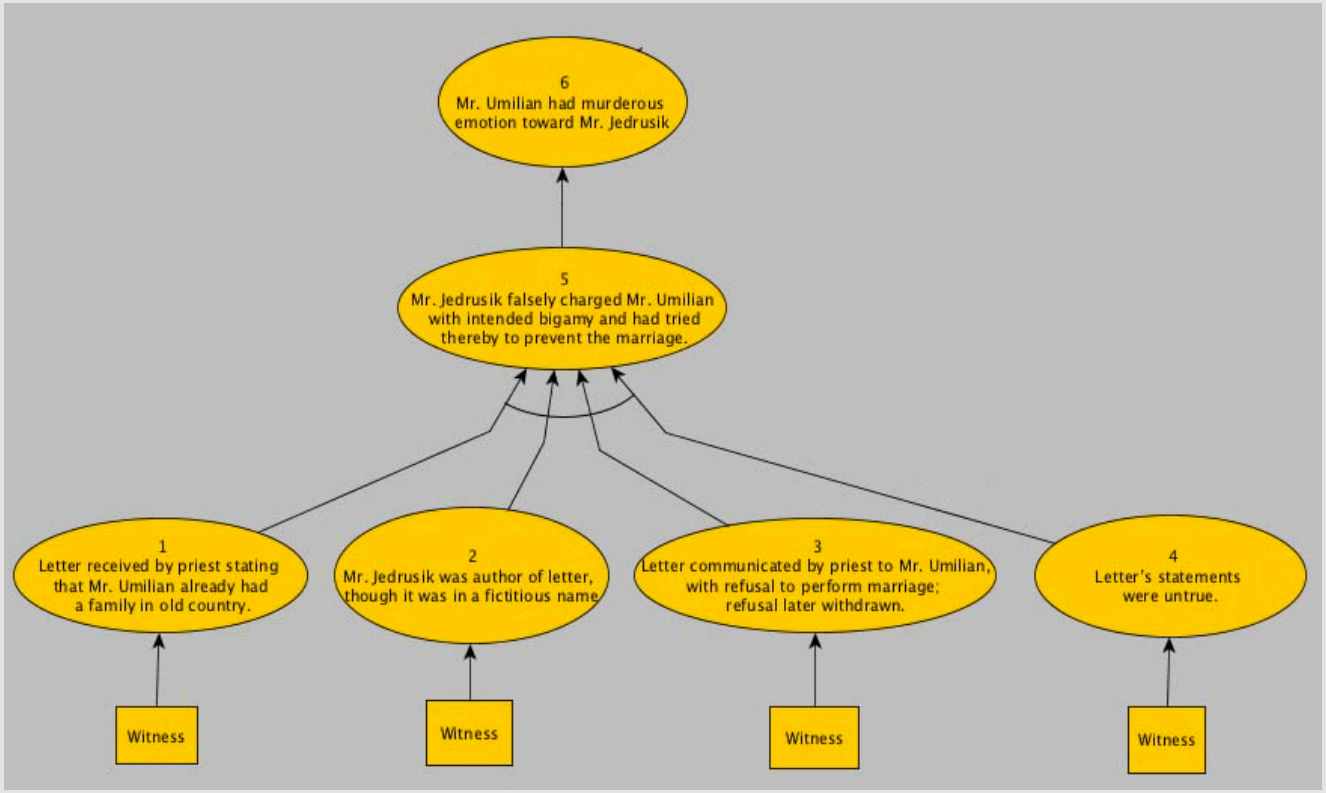
Existing Work

- **PyDFT**
  - **Able to create high-level events**
  - **Does not offer a graphical representation**
- **SADFC**
  - **Uses OWL and Data Mining to find relationships between evidence**
  - **Offers a timeline representation**
- **FORE**
  - **Uses OWL to find correlations between data in a system**
  - **Uses Event Graphs to present correlations**
- **Log2timeline**
  - **Is not able to create high-level events**
  - **Offers a graphical representation of evidence**

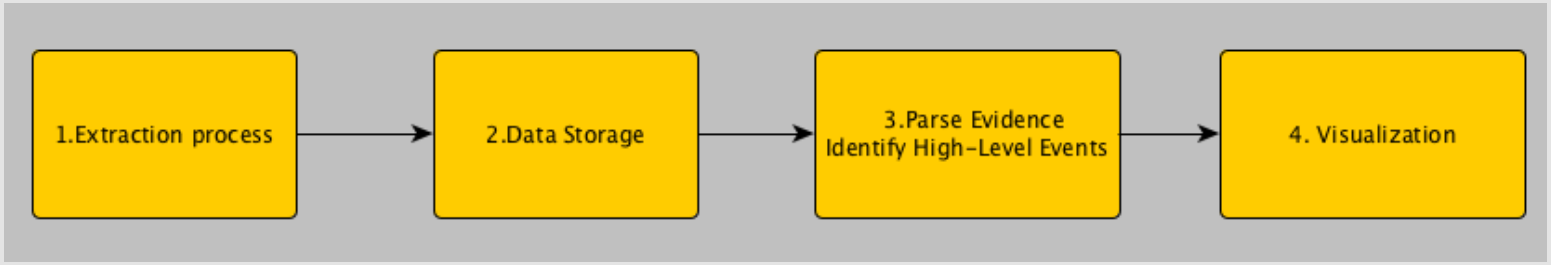
Literature Review

- **Importance of Graphical Representation**
  - **Network Forensics**
  - **Social Networks**

- **Evidence Representation**
  - **Wigmore's Chart Method**
  - **Pollock's Argumentation Scheme**



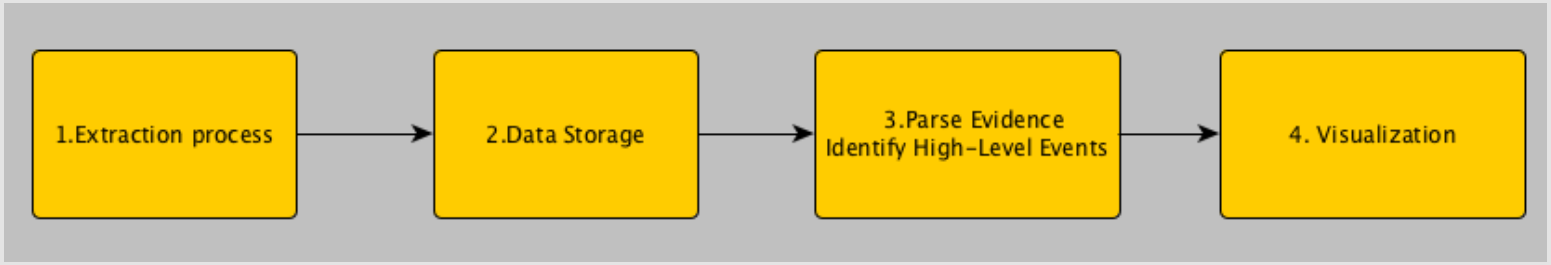
Design & Methodology (1)



- **Extraction Process**
  - **Timestamps and information from files in the file system**
  - **Information from the Windows Registry**

- **Data Storage**
  - **Store metadata information in .csv**
  - **Parse .csv using Python**

Design & Methodology (2)

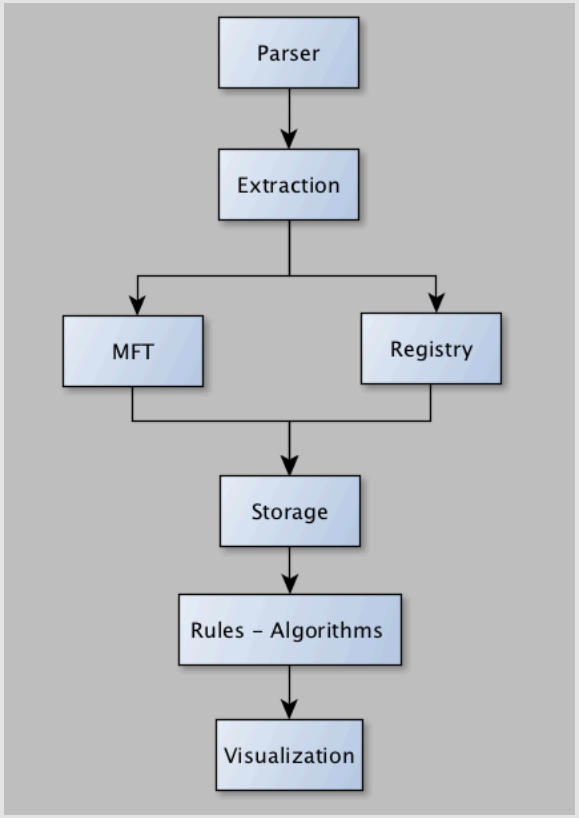


- **Rules Set - Three Algorithms created**
  - **"Is the framework able to show what happened in a particular time frame?"**
  - **"Is the framework able to show what happened in a user's session?"**
  - **"Is the framework able to present information that is related to a file?"**

- **Visualization**
  - **Use of Gephi tool**
  - **Every node contains information like evidence name, id and timestamp**
  - **Able to show clusters created**

Implementation

- Extract file system information using Sleuth Kit Python library (pytsk3)
- Extract MFT and use analyzeMFT to extract information to a csv file
- Extract Windows Registry information and store it in csv files
  - SYSTEM
  - SOFTWARE
  - SAM
  - NTUSER.DAT
- Run rules to parse the csv files
  - Identify connections between evidence
  - Store results in an Array
- gdf file implementation for Gephi to simulate



Experiments (1)

- **Preliminary Investigation**
  - **Able to identify**
    - **Software information**
    - **Device Information**
    - **Time Zone information**
    - **User information**

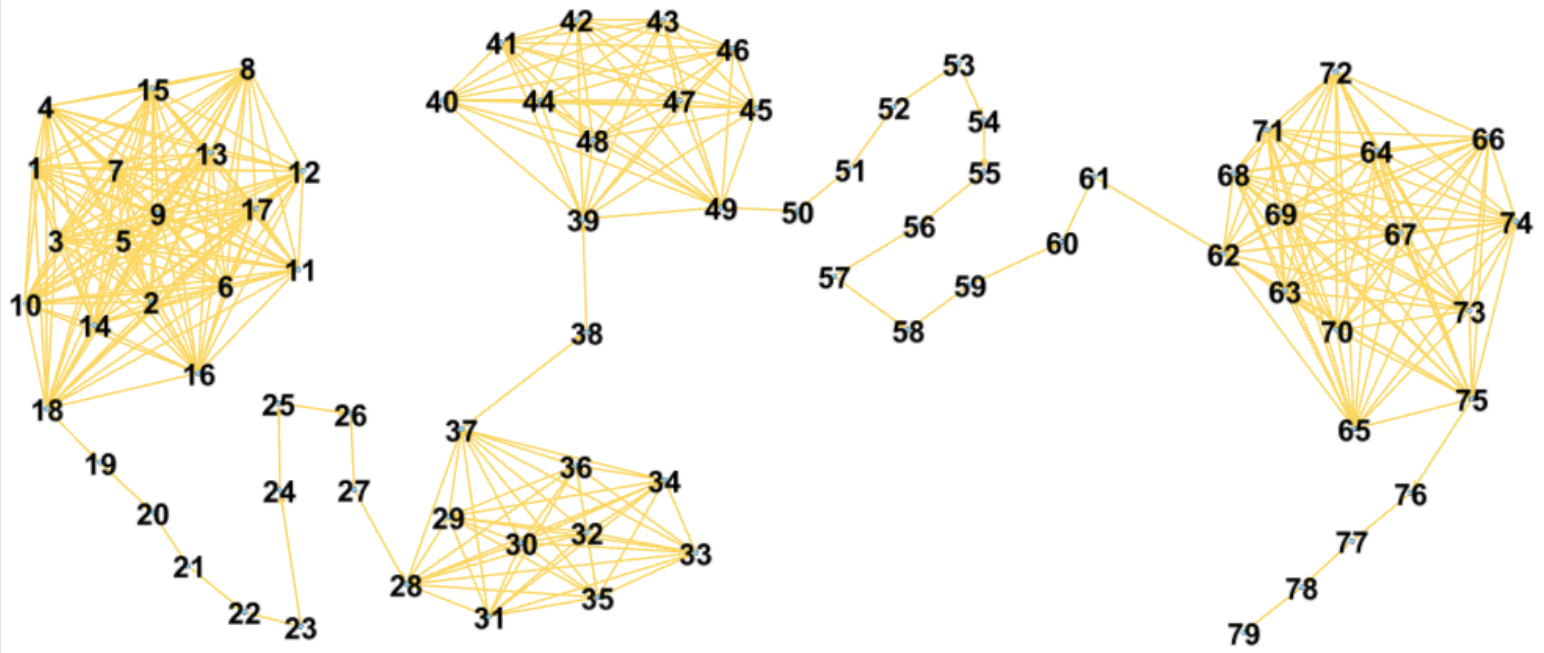
Files	Analysis
MFT	mft.csv
SAM	preliminary.csv services.csv usb.csv
SYSTEM	
SOFTWARE	
SECURITY	
NTUSER.DAT (1)	lastvisitedmru_Autolycus.csv mru_ Autolycus.csv userassist_ Autolycus.csv recent_ Autolycus.csv

Description	Value	Description	Value
Current control Set	1	**InstallDate**	04/07/2011 20:01:13
Computer Name	WIN-HG28CJ57ACD	**RegisteredOwner**	Windows User
Product Name	Windows 7	**SystemRoot**	C:\Windows
EditionId	Enterprise	**Time Last Shutdown**	15/07/2011 20:53:58
CurrentVersion	6.1	**Backup**	No backup found

Daylight Start Date				
0x0000	0x0003	0x0005	0x0001	0x0000
Every Year	March	5th week (ie last week of month)	1am	Sunday

Experiments (2)

- **Algorithm 2 – User Activity**



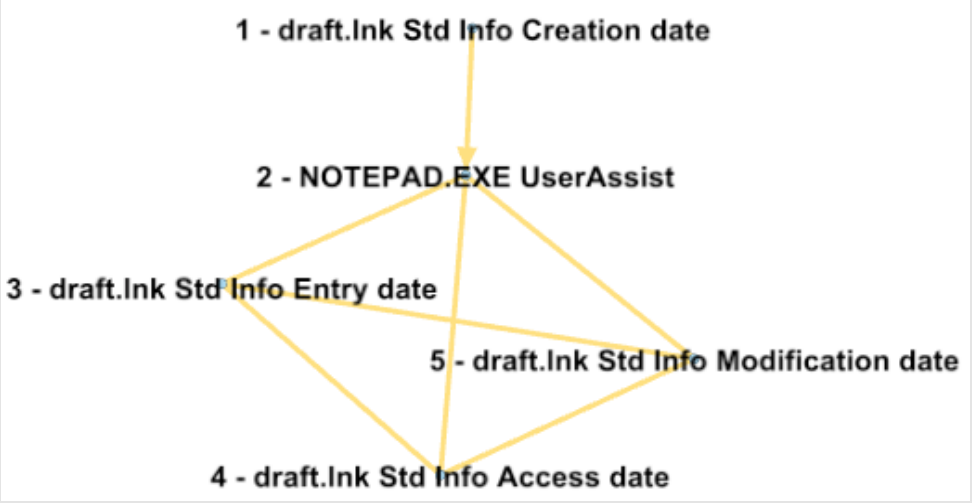
- **Four major clusters created**
  - **Allows analysis of large graphs to smaller ones**

Experiments (3)

- **Algorithm 3 – File Investigation**

Id	filename	timestamp
1	draft.lnk Std Info Creation date	2011-07-15 12:16:45.758789
2	NOTEPAD.EXE UserAssist	2011-07-15 13:07:09.039000
3	draft.lnk Std Info Entry date	2011-07-15 13:07:09.070311
4	draft.lnk Std Info Access date	2011-07-15 13:07:09.070311
5	draft.lnk Std Info Modification date	2011-07-15 13:07:09.070311

- **Presents information for a file called 'draft.txt'**
  - **Ability to show file timestamps**
  - **Ability to show program used to open the file**



Evaluation

- Achieved project objectives
- The Framework is capable of assisting in a forensic investigation
  - Able to automatically extract evidence from a disk image
  - Able to identify connections between evidence
  - Able to provide a good visualization technique
- An Automated Forensic Analysis tool able to defend against false positives
  - Third Algorithm Restrictions
    - The program used to create the file the investigator is investigating must be opened up to 300 seconds before the file is created.
    - To find a file in the Recycle Bin it must be deleted up to 200 seconds before the file's timestamp.

Future Work & Conclusion

- **Import extra functionality to correlate more sources of information**
  - **Internet Evidence**
  - **Hiberfil.sys**
  - **Windows Event Log**
  - **Data Carving Techniques**
- **Conclusion**
  - **Results show ability to assist in investigations**
  - **The framework was tested using specific experiments and identified connections between artifacts correctly**
  - **Further improvements can result**
    - **In a framework able to correlate evidence from a lot of sources of information**
    - **Reduce false positives**

References

- S. L. Garfinkel, “Digital forensics research: The next 10 years.” Digital Investigation, vol. 7, 2010.
- C. Hargreaves, J. Patterson, “An automated timeline reconstruction approach for digital forensic investigations”, Digit Investing, 9 (2012) 69-79.
- Y. Chabot, A. Bertaux, C. Nicolle, and M.-T. Kechadi, “A complete formalized knowledge representation model for advanced digital forensics timeline analysis,” Digital Investigation, vol. 11, 2014.
- T. Anderson and W. Twining, Analysis of evidence. Northwestern.
- “The Sleuth Kit: The Sleuth Kit (TSK) Library User's Guide and API Reference,” The Sleuth Kit: The Sleuth Kit (TSK) Library User's Guide and API Reference. [Online]. Available: http://www.sleuthkit.org/sleuthkit/docs/api-docs/4.3/index.html. [Accessed: Jul-2016].
- “analyzeMFT Python Library,” GitHub. [Online]. Available: https://github.com/dkovar/analyzemft. [Accessed: Jul-2016].
- “The Open Graph Visualization Platform,” Gephi. [Online]. Available: https://gephi.org/. [Accessed: Jun-2016].
- Oh, Junghoon, Seungbong Lee, and Sangjin Lee. "Advanced evidence collection and analysis of web browser activity." digital investigation 8 (2011): S62-S70.
- “Data Carving Concepts,” SANS. [Online]. Available: https://www.sans.org/reading-room/whitepapers/forensics/data-carving-concepts-32969. [Accessed: Aug-2016].

Questions?

Thank You