Initial Plan

CM3203 – One Semester Individual Project – 40 Credits

A Machine Learning Approach for Memory Forensic Investigation

Author : Amna Mohd A Hameed

Supervisor : Michael Daley

Moderator : Luis Espinosa-Anke

School of Computer Science and Informatics

Cardiff University

2020

Contents

Project Description 3
Background 3
Brief Description 3
Project Aims and Objectives 4
Supervisor Roles 5
Ethics 5
Work Plan 5
Deliverables and Milestones 8
Conclusion 9
Reference 10

Project Description

Background

Nowadays, the field of digital forensic investigation is growing and advancing, as it plays a critical part in solving cyber-crimes and tracing criminal's digital activities. With the rapid pace of technological change has enabled the law enforcement and digital forensic investigators to take into consideration of digital evidence to be used as evidence in criminal proceedings or to present in the court. The digital evidence ranges from documents, email to images and videos which are extracted and gathered from any digital devices that contains data.

However, various suspected devices and evidence are regularly gathered from almost every crime scene as a source of intrusion; this has increased in a backlog of devices to be examined and analysed. In addition, some of the major challenges faced by most of the investigators and analysts are the ongoing forensic workload, storage issues and difficulty to discover information of forensic value in new devices and operating systems (Boast and Harriss ,2016). As numerous reported cases are inconsistency delayed or being in 'waiting to be allocated' list as police backlogs are measured in years due to these challenges and limitations (Goldberg, A. 2015). Obtaining evidences from the individual devices does not only require a lengthy process of both extraction and result analysis. But also, requires special knowledge and skills; this has enabled criminals with new opportunities to commit further crimes and increased a potential risk to the community (Quick and Choo, 2014).

Therefore, many forensic communities are nowadays more focused in prioritizing the memory forensic. As the bit-to-bit images are acquired of volatile memory from a live running system is major part of any forensics investigation and analysis of cybercrimes. As it contains significant number of artefacts that provides relevant clues to investigators such as recent system activity, system processes, open network connection and ports. These volatile artefacts only exist in memory dumps, or RAM, of the running system. Several tools that enable for analysing and examining these volatile data requires manual access to retrieve any artefacts and evidence (Fichera and Bolt, 2013).

Researches are introducing new approaches to a digital forensic investigation that involves the use of machine learning (ML), a science branch of the artificial intelligence (AI) field. Mainly aims to automation in digital investigation procedure with significant potential to aid and assist investigators and analysts and to reduce manual process and to minimize human errors for any forensic investigation operations (Iqbal and Abed Alharbi, 2019).

Brief Description

This project aims to create a tool using machine learning approach for memory forensics investigation. A machine learning model will be developed to automate assistance in identifying vital areas in memory dump for forensic investigation and analysis.

A memory acquisition and analysis tool, Volatility, will be used to acquire memory dump snapshots and aid to identify and understand memory dump data. Additionally, the project will be developing a machine learning model that will work as automate assistance for the memory dump logs and provides an output as a summary log detailing the memory dump analysis. The implementation of the tool and machine learning model will be developed using python-based programming language and all the process of implementation and testing will carry on a windows operating system. The overall aim of the tool is to support first-line analysts and investigators, as the tool will provide them automated assistance to locate at areas where evidences are likely to be found and to improve both the average time and manual forensic analysis of a memory dump.

Project Aims and Objectives

This section describes the proposed project main aims and objectives :

1. Investigate and analysis of memory dump and areas where the vital artefacts are likely to be hidden
    1. Locate areas of where valuable artefacts are hidden for a memory dump
    2. Identify types of the artefacts found in memory dump
    3. Output a summary for memory dump analysis log

2. Implement Machine Learning model as a tool to automate the identification of vital artefacts in memory dump for forensic investigation and analysis
    1. Using python-based libraries and modules to implement Machine Learning models
    2. Implement Machine Learning algorithms for memory dump that :
        1. Predict and detect areas where most likely vital artefacts are hidden in memory dump
        2. Assist locations of where they are hidden
        3. Sort memory dump to urgency of memory analysis and whether memory dump require further deep memory analysis
        4. Retrieve of artefacts required from further analysis
    3. Train different Machine Learning models with dataset

3. Implement a range of possible malicious artefacts in memory dumps and use to the implemented machine learning models
    1. Research about possible artefacts that can be implemented such as malware
    2. Assist analysis for given malicious memory dump
    3. Provide a set of information and most relevant data that should be investigated by memory forensic analyst.

4. Documentation and reflective practice on the achievements

1. Critical analysing the critical path of the project and justification and evaluation of the results from the testing of the implementations.

Supervisor Roles

Arranged weekly meetings with my supervisor, Michael Daley, on Wednesdays for approximate of 20-30 minutes. However, it won't be necessary to have a meeting each week. These meetings are an opportunity to show the project progress, seek advice and guidance.

The project involves a co-supervisor, Amir Javid, for supporting guidance for any concerns and issues on Machine Learning approach for the proposed project.

Ethics

As discussed with the supervisor regarding ethical consideration for the use of the dataset in this project; concluded that it is unlikely to require and seek ethical approval. However, ethics will be considered for the entire project process and the final report.

The given dataset for this project, will be prepared from the samples of the memory dump that is provided by the supervisor. It does not involve any human or user participation such as interviews, questionnaires or personal data to be stored.

Work Plan

To develop a successful project and to produce an organised professional report. A project plan has been designed as the following table shows a weekly workplan of 15 weeks, including Easter recess, for the project to be complete.

Each week is specified with the tasks that are needed to develop the project. Also includes progress review meetings and weekly deliverables by milestones .

*No weekly meetings scheduled in Easter recess.

	Tasks	Milestones
Week 1 (27th Jan/3rd Feb)	– Prepare initial plan draft – Initial supervisor meeting – Submit initial plan after receiving feedback from supervisor	**Prepared initial plan draft**
Week 2 (3rd Feb/10th Feb)	– Submit Initial plan – Prepare Final report structure	**Submission initial plan report by 3rd Feb 2020 ✓**

	– **Joint meeting with supervisor and co-supervisor for the project approach** – Begin background research on : o Memory forensics acquisition, investigation and analysis o Memory dump artefacts o Machine Learning (ML) o ML algorithms and process for forensic classification o Python libraries and modules for ML o Volatility - Memory acquisition and analysis tool o Memory dump dataset	**Progress review meeting** **Background research draft for Final report**
Week 3 (10th Feb/17th Feb)	– Review Volatility tool – Review, understand and learn ML Python libraries and modules – Start tool specification and design: o Define the approaches of ML models o Describe and justify chosen approaches and design for developing tool o List requirements and specification – Determine Python libraries and modules for the implementation	**Written Approach draft for Final report**
Week 4 (17th Feb / 24th Feb)	– Outline and design UML / Class diagram of the tool ML models – Initial Implementation Setup: o Set up a Git repository for the project o Set up a workspace environment o Install required tools o Start memory dump acquisition and analysis o Prepare dataset	**UML / Class Diagrams for the tool** **Initial implementation**
Week 5 (24th Feb/2nd Mar)	– **Progress review meeting**	**Progress review meeting**

	– Identify and analyse memory artefacts data – Start Implementation : ML models	

Week 6
(2nd Mar/ 9th Mar)
	– Continue Implementation :
	  o Finish ML model
	  o Develop another ML model
	– Start implementation draft
		Working ML model

Week 7 (9th Mar/16th Mar)	– Finish Implementation of all ML models – Implement artefacts in memory dump	**All ML models works properly**
Week 8 (16th Mar/23rd Mar)	– Finish implementation of artefacts in memory dump – Train ML models with the dataset – Testing and debugging ML models with dataset – Finish Implementation	**Finish implementation code** **Implementation final report draft**
Week 9 (23rdMar/ 27th Mar)	– **Progress review meeting** – Design testcases – Start tool testing and evaluation – Record tool testing and results	**Progress review meeting**

Easter Recess
Week 10-12
(27th Mar/20th Apr)
	– Finish testing and evaluation of the tool
	– Review implemented ML models
	– Debug and fix any issues while testing
	– Improve tool functionality
	– Start writing final report draft for remaining section
	  o Abstract , Acknowledgement , Introduction
	  o Future Work and Reflection
	  o Conclusion
		Tool tested and evaluated
		
		Evaluation and Results final report draft

Week 13
(20th Apr/ 27th Apr)
	– Restructure and organise all Final Report sections
	  o Add table of figures and abbreviation, glossary, appendices
	  o Arrange used references
		Prepared Final Report first draft

	– Prepare Final Report in a standard format – Finish and Review Final Report first draft all sections	
Week 14 (27th Apr / 4th May)	– **Progress review meeting** – Review Final Report first draft with supervisor – Adjust any errors and typos	**Progress review meeting**
Week 15 (4th May / 7th May)	– Proof-read all the Final Report – Submit the Final Report and the source code	**Submission of Final Report and the source code by 7th May 2020 ✓**

Deliverables and Milestones

The following list illustrates the predicted deliverables by milestones outcome at end of the proposed project:

- ✓ **Initial Plan Report Submission** : Final version to be submitted by **3rd Feb 2020**.

- ✓ **Draft sections of the Final report**
  - Draft version of the final report sections (background research , approach, implementation and testing , results and evaluation and future work) to be completed during the project as mentioned in the work plan.

- ✓ **Tool Implementation**
  - Python source code of ML models implementation
  - Include short guide of memory dump analysis

- ✓ **First Final Report draft**
  - Structure the Final report
  - Add the previously completed sections and finish the remaining sections (Title page, abstract, acknowledgements, table of contents, table of figures, glossary, table of abbreviations, appendices and references)

- ✓ **Final Report Submission :** Final version to be submitted by **7th May 2020**.

Conclusion

Overall , this report provides an initial plan overview for the proposed project. It includes project description, aims and objectives with a project work plan and defined deliverables by milestones.

Reference

1. Boast, K. and Harriss, L. (2016). Digital Forensics and Crime – POSTNOTE Number 520 March 2016 - UK Parliament. [online] Available at: https://researchbriefings.parliament.uk/?ContentType=&Topic=Science+and+technology&SubTopic=Internet+and+cybercrime&Year=2016&SortByAscending=false [Accessed 30 Jan. 2020].

2. Fichera, J. and Bolt, S. (2013). Volatile Data Analysis. Network Intrusion Analysis, pp.71-117.

3. Goldberg, A. (2015). Child Abuse Cases Delayed by Police Backlog. BBC News. [online] Available at: http://www.bbc.co.uk/news/uk-34713745 [Accessed 29 Jan. 2020].

4. Iqbal, S. and Abed Alharbi, S. (2019). Advancing Automation in Digital Forensic Investigations Using Machine Learning Forensics. Digital Forensic Science [Working Title]. [online] Available at: https://www.intechopen.com/online-first/advancing-automation-in-digital-forensic-investigations-using-machine-learning-forensics [Accessed 29 Jan. 2020].

5. Quick, D. and Choo, K. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. Digital Investigation, 11(4), pp.273-294.