Cardiff School of Computer Science & Informatics

Automatic detection of Steganographic content

Author: James Taylor

Student number: C1129788

Supervisor: Mike Dailey

Moderator: Irena Spasic

Module title: One Semester Individual Project

Module number: CM3203

Credits due: 40 credits

1 PROJECT DESCRIPTION

In the digital world, steganography has received much attention over the last few years, due to its suspected use in criminal and terrorist organisations (New York Times, 2001). Steganography comes from the Greek words '*steganos*' and '*graphein*' which literally translates to '*covered writing*' and relates to any medium which hides a secret message within an ordinary message. Due to the nature of the technique, it is intrinsically difficult to detect unless the investigator knows exactly what to look for, as the message essentially hides in plain sight.

This project aims to make the detection of various steganographic techniques much more simple by automating the steganalysis process. This should stop seemingly innocuous files containing crucial information potentially "slipping through the cracks" during a forensic investigation of some digital media. It is also extremely hard to take further action in an investigation based solely on steganographic evidence, unless a payload has been completely recovered, and this project will aim to address this problem where possible.

I aim to solve this problem through the design of a system that can automatically process either single or multiple files of various types and scan each one for known steganographic techniques using various methods of steganalysis. If such techniques are found, it will display with a degree of certainty which techniques have potentially been used. In circumstances where it is certain that a particular technique has been used, the system would allow the user to automatically retrieve the concealed information. On the other hand, in cases where it is not certain but there is sufficient suspicion, it would highlight these files so they can put under higher levels of scrutiny.

As digital media is a very fast-paced medium, with the modification of existing file types to accommodate updated software, similarly the art of steganography keeps up with this trend with new methods of concealing data being discovered all the time. This means that the system will also need to be extensible, allowing new techniques to be added to a library of existing techniques as they are discovered, without affecting the core functionally of the program. This would make sure that during a forensic investigation the system stays current and no seemingly innocuous files are missed.

2 PROJECT AIMS AND OBJECTIVES

- The system should be able to detect data concealed in files using various steganalysis techniques.
    - I am aiming for between five to seven different forms of the most commonly used steganalysis techniques in the core functionality of the system focusing mainly on the *'stego-only'* or *'blind'* steganalysis techniques, where the existence of stego-text is unknown.
- The system should be extensible so new steganalysis techniques can be added.
    - This will be implemented in the form of a host/add-in architecture.
- The system should have a simple, user friendly GUI.
- The system should be able to process batches of files of various types or single files depending on the situation.

- The system will store the hash value of files when steganography is detected and keep this information in a database.
    - If possible, this database should also contain the recovered stegotext.
    - Therefore, if a particular file is seen more than once it will not need to be processed again. Matching the hashes would be enough to show this is the same file and the stegotext could be returned immediately.

2.1 Ethical implications

Although there are many uses of steganography, both legitimate and illegitimate, most of the focus in recent years has been on the criminal and terrorist organisations that use it. It is because of this that it is perceived as a threat, potentially more so than encryption due to the nature of it being difficult to detect. This is especially pertinent now given that the newly amended section forty-nine of the RIPA act forces the disclosure of encryption keys when it is in the interest of national security (Regulation of Investigatory Powers Act, 2000) which could result in a greater use of steganography from the criminal community. As steganography is not widely used in the mainstream security community, it therefore seems that an automated steganalysis system does not propose a threat to the average person and could be of benefit to law enforcement. This project does not require any user data to be stored, so there are no privacy concerns and for these reasons, this project does not require ethical approval by the department.

2.2 Project Constraints

Given the scope of the project and the fact that steganalysis is a massive field, the main constraint in this project is going to be time. Assuming I can achieve successful implementation of one method of steganalysis each week, that should mean the system will have a baseline of around six techniques. Given more time, more could be implemented. Hence why the system should be extensible, so if it is ever used later down the road by myself or other, techniques that were missed during the original implementation could easily be added to the application.

2.3 Deliverables

By the end of this project, I will produce the following items:

1. A program that can:
    a. Automatically process multiple files
    b. Detect steganography in these files and, where possible, automatically retrieve stegotext
    c. Store hash values of known steganographic files in a database to avoid processing the same file twice
    d. Be extended by developing a host/add-in architecture where new techniques can easily be added to the software
2. A report that demonstrates the capabilities and limitations of the above program and describes the whole development process

3 WORK PLAN

3.1 WEEKLY OBJECTIVES

Below are the objectives I hope to achieve as the project progresses. I will evaluate myself against these objectives to make sure that the project and myself are on track. Any difficulties or problems I may run into throughout the project can hopefully be addressed at the weekly supervisor meetings if possible and where they not, correspondence via email would be the first port of call.

WEEK	OBJECTIVE	PROJECT MILESTONE
1	Initial meeting with supervisor. Submit initial project plan. Perform initial market research.	Start of project.
2-3	Research methods of hiding data and research methods of steganalysis.	
3-4	Start implementation, develop extensible framework with database for hashes. Begin testing framework.	Start of implementation.
4-10	Implement as many steganalysis techniques as possible in the timeframe. Continually test newly implemented steganalysis techniques.	End of framework development. Start implementing steganalysis techniques. Start testing steganalysis techniques.
7-8	GUI design and development.	Start & finish implementation of GUI.
9-11	Begin evaluation of software and shift focus of the project to writing the report.	End of implementation & testing. Start evaluating the software and writing the final report.
12	Proof reading and submission.	End of project.

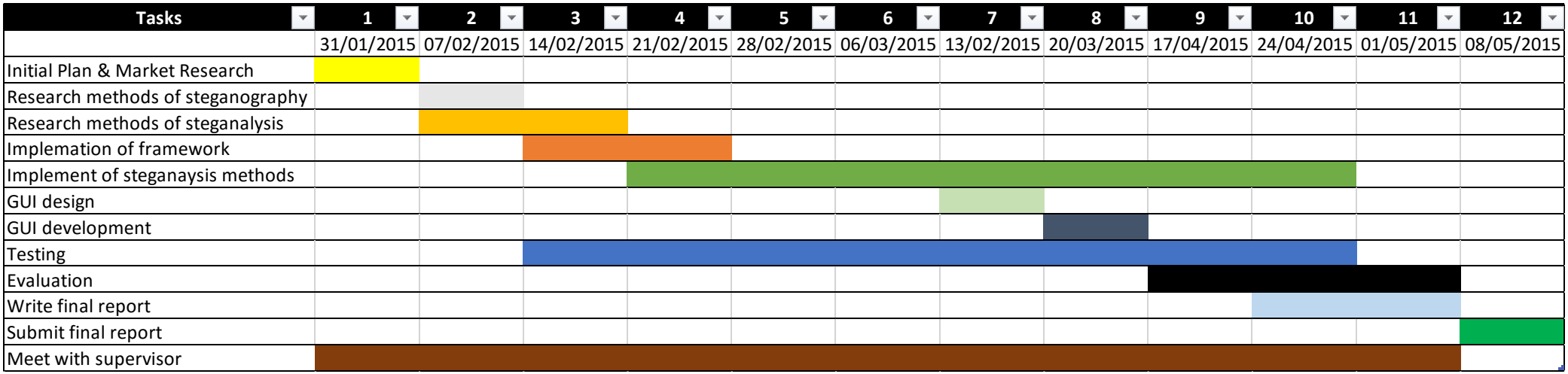
4 REFERENCES

New York Times. (2001, October 30). *http://www.nytimes.com/*. Retrieved from http://www.nytimes.com/2001/10/30/science/physical/30STEG.html?pagewanted=all

Regulation of Investigatory Powers Act. (2000). *http://www.legislation.gov.uk/*. Retrieved from http://www.legislation.gov.uk/ukpga/2000/23/section/49

5 PROJECT GANTT CHART

The Gantt chart below is a visualisation of the above work plan and will be used to evaluate my progress at any given time.



Tasks	1	2	3	4	5	6	7	8	9	10	11	12
	31/01/2015	07/02/2015	14/02/2015	21/02/2015	28/02/2015	06/03/2015	13/02/2015	20/03/2015	17/04/2015	24/04/2015	01/05/2015	08/05/2015
Initial Plan & Market Research												
Research methods of steganography												
Research methods of steganalysis												
Implemation of framework												
Implement of steganaysis methods												
GUI design												
GUI development												
Testing												
Evaluation												
Write final report												
Submit final report												
Meet with supervisor												