Evading Malicious Code with Concurrent Programming in Parallel Architectures and Their Protection Methods

Caglar SAYIN

Master's Thesis Project Description
Master of Science in Information Security
5 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2013

Avdeling for
informasjonssikkerhet
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Information Security
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Revision history

Version #	Description of change (why, what where - a few sentences)
0.1	First version is made available via Fronter

Abstract

Abstract (1/2 page) This document provides format and guidelines for the MSc project descriptions. The document has been produced using MikTeX and TeXnicCenter.

The objective of the abstract is to provide the reader with an understanding of the work to be done and put him in the position to make a 'correct' decision regarding reading/not reading the report.

The abstract of the project description *must* include

- a summary of the problem description,
- motivation and
- a summary of the planned contribution from the master project in terms of *new* results.

Control questions

1. Does the abstract have a 'reasonable' length?
2. Is it clear to a non expert (e.g. a typical reader of a newspaper) what problem is addressed?
3. Does a person that has been working in the field find the text informative?
4. Do the results that might be obtained have the potential to be interesting to a lot of people? How interesting to how many and why?
5. Would a decision maker/manager be willing to pay NOK 400.000 to have the project completed (estimated salary costs + overheads) after having read the abstract? Why/why not?

Contents

1 Introduction (1-2 pages)

The purpose of introduction chapter is giving the readers blueprint of the subject, the problems that we face, the change in the solutions, as well as motivation of its importance. In addition, It also purpose to form proper research question which will guide thesis.

1.1 Topic covered by the project

The thesis purposes an architecture of the malware which process parallel, access memory concurrently, conceal itself systematically, shortly that it is likely to be rocket science. However, everything actually started with a simple mathematical theory by John Von Neumann [1] and the first example of practical malware is writen by Bob Thomas at BBN, and it was called Creeper

The malware is abbreviation of malicious software. It could be any piece of code which is defined malicious. There is no formal definition of malicious, it could be some software advertise without any contest or it could be self-producing code piece which aim to distribute itself and steal your private information, and it turned an arm race between power holders today.

With development of the first malware, their counter software are created and anti malware software have evolved with them so far. In this race, malware authors are always one step further, because of security's nature. This race between black and white side raised the bar above. The motivation of the information amount and severity influence both today, and that information can be sometimes vital.

1.2 Keywords

Security, Concurrent Malware Design, Malware, Concurrent, Parallelism and Concurrency

1.3 Problem description

The one of the main and indecipherable problem in security discipline is formulating general threat definition and recognizing malicious activity and all this problems unsurprisingly reflect on information and computer security concept. Security is defined by system's identification, which involve with purpose, crowd, design structures, network model and so on, and today's information system which is designed with various architectural forms is protected against malware by general purpose protection tools. In the market, The anti malware tools producers focused on pragmatic solutions to survive, but it leads to that most of these tools are utterly reverse engineering process which works on result instead of reason.

With usual and pragmatic signature based methods, there are two mainstream techniques to detect malicious code which are called static and dynamic analysis.Static analysis identifies malwares mainly with code flow graph and data flow graph on stored file which is not processing. However, On the dynamic side it is a bit more tricky to analyze process, because you are working on the running pieces of codes without knowledge of

structures and worse than this, it must concern race condition and memory coherency flaws.

The detection methods and techniques have been adequately worked so far because of the simplicity of architectures and usage of the massive generic computers, However, with increasing of the not standardized, parallel and popular devices like arm's SoC, it is not hard to estimate their new challenges. It is really likely to evade and obfuscate properly your on-the-fly processes with using uncertain charactership of parallel processing, complexity of concurrent programming, and structure of "Non Uniform Memory Architecture".

1.4 Justification, motivation and benefits

If malware designing is superficially considered, you could fall in usual fallacy that It is not beneficial and exactly opposite. However, if we can design it, there is always more skillful author who already abuse this vulnerabilities on the black side of the moon. The work we are obligated to actually proof this vulnerabilities and design counter measure against them. In this way, our blessed motivation is finding possible vulnerabilities, and mitigate or eliminate their risk. Otherwise; if we confront with unknown attack, it could be too late to fix and analyze it. For example, some of the most sensational and beneficial papers are criticize malware as same as the thesis ([2],[3],[4]), and their values are undoubted today.

1.5 Research questions

1. Can a malware model be designed with using parallel and concurrent architecture in order to conceal its presence from detection mechanism?
2. If we can design the mentioned malware, can we build a detection mechanism against these kind of malware's presence?
3. If we build the detection mechanism, What is detection complexity of the algorithms?

1.6 Planned contributions

This Master thesis is looking for better understanding on concurrent malware abilities and their counter-measure. Especially, It will try to show how possible to abuse concurrent memory accessing and how durable recent detection kits. It is quite unique work which we have to consider on the future. Ultimate goal is to eliminate any uncertainties which detection methods encounter with concurrent memory accessing.

2 Related work (3-10 pages)

The purpose of this chapter is to explain to the reader what knowledge is already available from the literature.

The purpose of the related work chapter is to:

- Identify to what extent information identified in the 'Research questions' section is provided in the literature.
- Give an overview of why/how the literature provides the answer to the research questions identified.
- Identify areas/ research questions where the literature appears to be weak or non-existent.

The Related Work Chapter is NOT:

- A list of abstracts and summaries of more-or-less-relevant literature.

If you have

- found some relevant literature
- made summaries of what you have written

you should

- reorganize these summaries to focus on the research questions you have identified.

This chapter should include one subsection for each of the research questions identified in section 1.5.

2.1 Handling Potential problems

When searching for literature, you usually get too many hits or none at all...

Question 1

I don't find any relevant literature.

Answer 1.A

Make a list of words, phrases, applications, abbreviations, organizations, terminology etc. relevant for your area of interest. Ask a librarian to sit with you for 20 minutes to formulate relevant queries to available databases. Record your findings.

Answer 1.B

Go to the ACM (www.acm.org) or IEEE (www.ieee.org) web pages. Identify the SIGs (Special Interest Groups) of these organizations. Select the SIGs which looks the most interesting. Most SIGs publish one or more journals and/or organize workshops or conferences. Get hold of a few journals or proceedings and see if they're any interesting.

Question 2

I've found a lot of papers. They all look interesting, but I don't have time to read them all.

Answer 2.A

Narrow your search. Be more specific in your search. Read the abstracts of the relevant articles before you read the full papers.

Answer 2B

Find a citation index (e.g. `http://citeseer.ist.psu.edu/`. Read those papers with a high citation score first (a citation index rates papers according to 'academic popularity'). Alternatively, read those papers published in 'prestigious' conference proceedings or journals first.

Control questions:

1. Why can we have confidence that the most relevant literature has been identified?
2. is the related literature grouped in a sensible way such that the reader gets a good understanding of 'existing knowledge' relating to th research questions/problem description?
3. Is the chapter sufficiently comprehensive?

3 Choice of methods (2-5 pages)

This section is to include a description of the methods to be used, including references to literature describing the methods to be used (e.g. qualitative, quantitative, interviews, surveys, questionnaire, model building etc.) For each of the research questions to be addressed, the chapter is to explain why the method is

- appropriate
- likely to provide the desired knowledge/information.

4 Milestones, deliverables and resources (2-5 pages)

The purpose of this chapter is to convince the reader that you know exactly what to do. This chapter gives a description of how the project is to be broken down into smaller parts and activities.

1. What is it you have to do in order to obtain the desired knowledge?
2. What deliverables are to be produced (MSc thesis report, software,...)
3. When are the various deliverables going to be available?

For each deliverable, identify 4 versions, having an 'increasing' degree of completeness/quality. Students are strongly recommended to review each others drafts. For each version of a deliverable explain why and how this version is to be better/more complete. E.g. v1.0: my first draft - chapter text includes 1/2 page summaries only. v.2.0: Like v1.0, but comments by NN(who? fellow student) has been incorporated. v3.0:....

This section is to include a preliminary table of contents for the MSc thesis (only include 2 levels).

For each of the activities identified, specify

1. the time you need to complete each activity both calendar time and 'man-hours'.
2. hours needed by you
3. things you need to buy (consumables)
4. equipment, lab space or facilities you need access to
5. contributions from others (e.g. survey/interview participants) and how much each will have to contribute in terms of resources (probably time)

At the beginning of this section, provide a 2-3 line summary of the resource requirements. This is particularly useful if you have broken down the task into a lot of small tasks.

5 Feasibility study (1/2-3 pages)

An analysis of why it is likely that the desired results can be produced within the given time and resource bounds. This may include a description of

- similar projects completed by others and their 'resource consumption',
- an attempt to answer parts of the research questions
- the 'difficult' elements of the work and an explanation of why/how these problems can be solved. Alternatively you can explain an 'approximate' solution.

6 Risk analysis (1/2-2 pages)

In this project, there five inevitable risks which we can face during development.

- The thesis is highly dependent on the hardware, and the cost of the hardware constitute risk on its own. Any case of hardware defect leads to comprise obstacle.
- Hardware dependency is also leads to logistical and time consuming risk which could result with latency on submit time.
- Firmware codes which we are planning to work on are mostly undocumented. We could discover their usage by proper reverse engineering and fuzzing process when required, however it is obviously manpower.
- Most important and highlighting risk is there isn't proper research on this particular area. That means there are strongly possibly hidden risks which could cause other mental and physical result.
- During testing and purification part, Anti-malware tools could come out with unreliable result. To analyze result properly, we may need to inspect mentioned tools with reverse engineering process which could violate proper usage agreement. To mitigate that kind of risks, we could request research agreement from companies.

7 Ethical and legal considerations (1/4-1 page)

The content of this document could be used in order to malicious purpose, but any matter or information could be misused in the life and ignorance is not known well as a defense strategy. In this purpose, this thesis aims to enlighten security specialist and system developers against recent way of the possible attacks.

However, in order to act ethical responsibility, we tried to eliminate practice of tools and piece of codes which could leads malicious usage. In any case, there is no doubt that it is critical to discover and publish vulnerabilities which could cause deep impact before malicious people discover and abuse them.

> "Virus don’t harm, ignorance does."
> - VxHeaven

Bibliography

[1] Von Neumann, J., Burks, A. W., et al. 1966. Theory of self-reproducing automata.

[2] Moser, A., Kruegel, C., & Kirda, E. 2007. Limits of static analysis for malware detection. In *Computer Security Applications Conference, 2007. ACSAC 2007. Twenty-Third Annual*, 421–430. IEEE.

[3] Cavallaro, L., Saxena, P., & Sekar, R. 2008. On the limits of information flow techniques for malware analysis and containment. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, 143–163. Springer.

[4] Egele, M., Scholte, T., Kirda, E., & Kruegel, C. 2012. A survey on automated dynamic malware-analysis techniques and tools. *ACM Computing Surveys (CSUR)*, 44(2), 6.