“Botnet Battlefield”: A Structured Study of Behavioral Interference Between Different Malware Families

Bishwa Hang Rai
Supervisor: Prof. Dr. Alexander Pretschner
Advisor: Mr. Tobias Wüchner

Table of contents

Outline

Malware

- Malware is a general term to refer any malicious software that corrupts or steals data, or disrupt operations with illegitimate access to computer or computer networks
- It can be classified into self-replicating (virus,worms) and non-replicating (Trojan)
- Based on its ability to change its structure it can also be broadly classified into Polymorphic and Metamorphic
- Different variants of same malware with similar semantics or same author are regarded as to be from same family

Growth of Malware

- With the increase in growth of the Internet, many of our daily life activities such as email, banking, bill payment, and social networking are dependent on it.
- Malware authors are introducing new malware on daily basis to steal those valuable data and personal information and sell it illegally in the underground market.
- Annual loss caused by malware in 2006, 2.8 billion dollars in US and 9.3 billion euros in Europe
- Driven by monetary profit, high rise in numbers of new malware with 140 million new malware introduced in 2015 alone

Interference Between Malware Families

- There has been some anecdotal evidences of feud between the malware families
- In 2004, NetSky vs Bagle and MyDoom trying to remove each other along with message of profanity
- In 2010, SpyEye vs Zbot with KillZeus feature
- In 2015, Shifu malware family with AV like feature
- All of these interferences were to negate the presence of another malware
- Increase their own profit taking control of larger share of economy

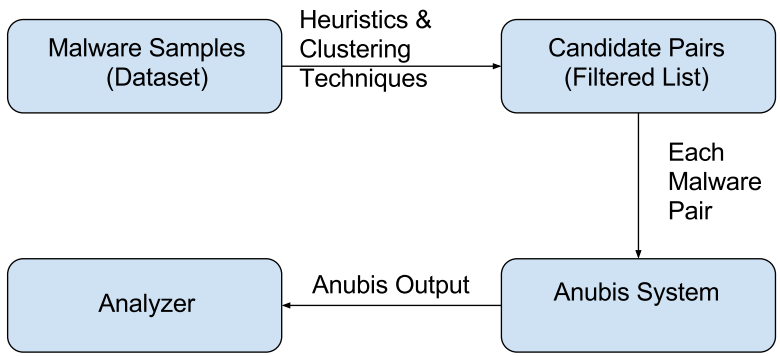
Problem Statement

- The purpose of our research is to identify the existence of aforementioned behavioral interference between the malware families
- The study will provide novel knowledge for understanding the dynamic aspect of modern malware, the inter-family relations, and their associated underground economy
- This behavior is also a case for environment-sensitive malware
- That is to say malware changing their behavior depending on different factors of their running environment, such as presence or absence of files, programs, or running services

Outline

Research Process

- Get wide variety of malware samples
- Use heuristics and clustering to get the candidate pair list
- Run each candidate pair in malware analysis system (Anubis in our case)
- Analyze the log of analysis run to detect behavioral interference



Outline

Contribution

Our research will provide the following contributions:

- To the best of our knowledge, we are the first to perform a systematic study of interferences between malware families
- A novel approach to malware clustering based on malware behavior profiles
- An automated system that detects interfering malware samples on a large scale

Outline

List of Candidate Pairs

- Value of N (maximum family cutoff) in algorithm chosen to be 10
- File with the highest number of candidate pair and Process the lowest
- No candidate pair from resource type Job, Device, Driver

Resource types	#candidate pairs
File	213,171
Registry	39,899
Sync	7,781
Section	2,786
Process	54
Total	263,691

Experiment Setup

- 7 Anubis instance
- Each instance emulates entire running PC with Windows XP Service Pack 3 as OS
- Uses Qemu and monitors process by invoking callback routine for every basic block executed in virtual processor
- Unpacker and Packer used to run the candidate pair
- 10 minutes as total run time of each candidate pair experiment
- 4 minute for each malware, and 2 minute to boot system

Result of Candidate Run

Resource types	# tested pairs	# true positive	prediction accuracy
File	5,000	1032	20.64%
Registry	5,000	731	14.62%
Sync	1,000	119	11.9%
Section	1,000	93	9.3%
Process	54	6	11.11%

- Highest Accuracy for File and Registry
- Lowest for Process
- Average accuracy rate 14.25%

Some Examples

- Artemis! vs Cosmu on resource `C:\Old.exe`
- VB.CB vs Startpage.AI on resource `C:\WINDOWS\window.exe`
- KeyLogger vs OnlineGames on resource `C:\windows\system32\svrchost.exe`

Outline

Threats to Validity

- Different values of N would give different candidate pairs and different results
- Random resource name
- Total execution time 10 minutes
- Sequence of execution
- Semantics of Malware

Outline

Conclusion

- Behavioral interference between malware families exists
- Malware checks for the presence of resource created by other malware and deletes it
- Our system could detect such interfering malware with average accuracy rate of 14.25%
- In our dataset, Files and Registries were the most interfered resource and Process was the least

Future Work

- Make the experiment more efficient to run multiple times with different parameters
- Research on other different approaches to clustering
- In depth analysis (static) of positive pair to know the true semantics of malware

Questions

Candidate Selection

R = Set of all interesting resource
A_r = Set of malware that creates a particular resource 'r'
B_r = Set of malware that delete/access (failed) particular resource 'r'
N = Maximum number of families to consider
E = Set of all probable candidate
function C (j)
    c_j = cluster id that malware j belongs to
    Return c_j
end function
for all $r \in R$ **do**
    if $|C(x_r) : x \in A_r| > N \vee |C(y_r) : y \in B_r| > N$ **then**
        continue
    end if
    for all $(x_r, y_r) \in A_r \times B_r$ **do**
        if $C(x_r) \neq (y_r)$ **then**
            $E \leftarrow (x_r, y_r)$
        end if
    end for
end for