Let your CSIRT do malware analysis, Recruit-CSIRT has done it!

Tatsuya Ichida

Recruit Technologies, Co. Ltd.

FIRST TC Amsterdam 2017, 25th April
Agenda

- Self-introduction
- Background and Motivation
- Malware Analysis System for Recruit-CSIRT
- Advantages and Disadvantages
- Conclusion
Self-introduction

Tatsuya Ichida (age 29)

Recruit-CSIRT since 2015

- Security Engineer
  - for developing useful tools
- Incident handler
  - at Recruit-CSIRT
- Loves Malware Analysis
- Splunk Log Analyst
- Tokyo Denki University CySec speaker
- In the past,
  - Security Operation Center, Malware Analysis Leader
- CISSP, GCIH, GPEN
No Incidents -> Can develop tools!

IRG
Incident Response Group
- Containment
- Incident Handling
- Internal and external coordination
- Emergency Forensic

QMG
Quality Management Group
- Prevention
- Vulnerability Assessment and Management
- Literacy Education

SOC
Security Operation Center
- Detection
- Security Log Monitoring
- Artifacts Analysis
- Forensic Investigation
Background and Motivation
Background 1

Malware explosion in the wild

500 million

Ref: https://www.av-test.org/en/statistics/malware/
Ransomeware explosion in our env

200

100

7/13

7/15

(C) Recruit Technologies Co., Ltd. All rights reserved.
We discovered a malware file over 100MB in size!

Can it be done Reverse Engineering rapidly? No way!
Background 4

Work time

Reducing Work Time
**FIRST Step**

- Using Commercial Malware Analysis Products.
  - Sandbox Product A → Advantage of Anti-Sandbox.
  - Sandbox Product B → Advantage of Mal-Signature.

**But .. When We got malware**

malware 「α」 was not analyzed by A because of the Browser Version.
malware 「β」 was not analyzed by B because of Anti-sandbox technique.
malware 「γ」 was not analyzed by both because of the size!

We paid a lot of money!
Motivation – Challenge

Let’s create our own malware analysis system

Purpose
- Reducing cost
- Reducing user work time
- Stored knowledge internally

It’s impossible to create a malware analysis system that can handle all samples perfectly.

government of the people, by the people, for the people
by Lincoln

Analysis of our CSIRT, by our CSIRT, for our CSIRT
by Recruit

Our system’s target is “our malware”
Malware Analysis System for Recruit-CSIRT
Malware Analysis System Overview

- **Manage Server**
- **Analysis Guest VM**
- **Windows Server (AD, FileServ)**
- **Internet**
- **C&C traffic**
- **Logs**
- **Malware Submit**
- **Connections**
- **Web**
- **Report View**
- **cuckooDB**
- **IntelligenceDB**
- **Real timeDB**
Malware Analysis Scheme

Analysis Scheme

Auto-Collect Malware

Auto-Optimize Analysis env

Real-time View of Behavior Changes

Post Intelligence

- From Introduced Malware Detection Sensor
  - Targeted Malware

- From Internet Malware DB
  - Newest Malware before being targeted

- Auto-time Sync
  - Over 100MB huge malware analysis
  - Auto-Defense against external attacks

- Auto-log collector
  - Pcap excluding normal

- Block The C&C traffic
  - C&C’s IP
  - C&C’s FQDN

Image Confidential In speaking only you can see

Auto-C&C Server Analysis

WebProxy/FW

http://malicious.jp

Anti-Virus management

Start Using Selenium WebDriver based

cuckoo based

mongoDB based

(C) Recruit Technologies Co.,Ltd. All rights reserved.
Advantages and Disadvantages
## Advantages and Disadvantages

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimized guest image env</td>
<td>Cannot handle a lot of malware</td>
</tr>
<tr>
<td>Capable of analyzing huge malware samples</td>
<td>No accelerated sleep bypass</td>
</tr>
<tr>
<td>Anti-Virus detection control</td>
<td>Weak to virtual env evasion</td>
</tr>
<tr>
<td>Auto-C&amp;C analysis</td>
<td></td>
</tr>
<tr>
<td>Real-time visualization</td>
<td></td>
</tr>
</tbody>
</table>
Advantages – 「Optimized guest Image env」

- the Same Image
  - Same Middleware, Same Applications, Same Versions
- Some Ground bait, Mouse Control and Real Date

It help us to focus only on **malware infecting our env**
Advantages – 「Capable of analyzing huge malware」

- **Cuckoo Sandbox 2.0 rc1**
  - Agent Default
    - `/cuckoo/agent/agent.py`
      - XMLRPC based connection to host
        » Huge malware samples cause memory exceptions
        » Because of oversize XMLRPC’s memory…
  - Manager Default
    - `/cuckoo/lib/cuckoo/core/guest.py`
      - Has two managers
      - OldGuestManager Class(default) and GuestManager Class(for new agent)

- **We enhanced Cuckoo Sandbox 2.0 rc1**
  - New Agent
    - We added functions to the agent: time-sync, etc.
      - HTTP based Connection: Agent launches SimpleHTTPServer
      - No limit on Chunk Data to submit
  - Manager uses “GuestManager Class” in guest.py
Advantages – 「Anti-Virus Detection Control」

Almost Sandbox system -> **Antivirus disabled**
Usually prevents analysis

Our system permits Antivirus to delete the sample.
We observe while the malware and its child processes exist in our env.

Image Confidential
In speaking only you can see
Advantages - 「Auto-C&C Analysis」

Monitor Traffic

- Process behavior (Based on Win API calls)
  - Certainly malware emits traffic
  - But it may not include C&C traffic
  - Rootkit traffic cannot be caught

Pcap (Catch on host)
  - Include all traffic even from rootkits
  - Include much normal traffic
  - Huge volume

- Except Windows and Normal App traffic
- Monthly Update to "tcpdump exception rule"

Store IntelligenceDB (IPv4 or FQDN)

Analysis target

Multi Thread
- Check whether is on hosting server via HTTP
- Check its whois info from cymru
- Check virustotal reputation (downloaded malware from)

Result Update DB

IntelligenceDB

mongoDB
Advantages - 「Real-Time Visualization」

Default cuckoo report cannot be watched until analysis finished.

Created ‘running page’ to see the real-time behavior

Analysis Finish Button

1min

F5

update

Real-time view tells us the behavior changes ASAP -> Rapid Block Action & Rapid Re-Analysis
Advantages - 「Real-Time Visualization」 2

IntelligenceDB → FQDN, IP, WhoisOwner, CC, WebResponse, WebTitle, VT MalDownloadUrls, VT MalDownloadFile

SpentSecond
Thread id
Called API name
Points
Category
Is Success?
Time

bson

Real-timeDB

ajax

pcap

SpentSecond
DstIP
DstPort
SrcIP
SrcPort
Protocol
Dump Top 128 byte
Time

page1.json
page2.json
page3.json

・・・
Disadvantages - 「Cannot Handle a lot of Malware」

Our System handles in “Single Thread”

Malware can make us wait…

Generally, preprocessing seems to be important for this system.

- reducing the input sample
  - (auto) duplicate hash
  - (auto) untargeted extension and file-type
  - (manual) ‘targeted’ or ‘common’ by analyst
- reducing during analysis
  - Handle Anti-Virus detection

But we Focus Deeper Analysis than shallow and efficient

When we catch APT malware through forensic, We analyze long-term to observe the changes

C&C’s domain, IP spawn files, Attacker’s visit etc…
Disadvantages – 「No accelerated sleep bypass」

- Malware often calls ‘Sleep’ to wait for some time
- Some Sandbox products have functions…
  - Accelerated sleep bypass
    - In order to analyze the sample efficiently
- However malware is evolving…
  - Have Anti-sandbox techniques for this
    - Ex. CPU Clock difference using GetTickCount etc..

Our human resource is limited. We don’t take this into account.

i.e. Raw Analysis
Disadvantages – 「Weak to virtual env evasion」

✧ Malware often checks whether it runs on a virtual machine or not, halts its execution in analysis envs.
✧ There may be also endless Anti-Sandbox techniques employed.

Recruit changed Office PCs to VDI Thin Client.
Virtual env = Our env

Some Signature should be removed, but not all.
It is important to imitate VDI’s Virtual env.
Conclusion
Concluding

✓ Effective for our malware which is affected
✓ Can be used flexibly
  ✓ Theoretically no limit, since it is developed by ourselves
✓ Our System is not perfect to analyze all malware.
Thanks to FIRST and OSSs.