Visual Analysis of Malware Behavior

October 11th, 2009
About Us

- University of Mannheim, Germany
- Laboratory for Dependable Distributed Systems
- Security research:
  - Analysis of malware and spam
  - Honeypots / IT-Forensics
Motivation

- Sandbox Service - CW Sandbox.org
- Dynamic malware analysis
  - API-Hooking
  - Monitors 121 API-Calls out of 20 sections
  - Detailed Behavior Report (XML)
- Up to 4000 new samples per day
- Manual processing not possible
Motivation

• What are the main Operations?
  • Network activities?
  • Registry access?
  • Filesystem operations?
  • ...

• Which operations do not occur?
Motivation

- What are the main Operations?
  - Network activities?
  - Registry access?
  - Filesystem operations?
  - ...

- Which operations do not occur?

Which samples are of interest?
<analysis
cwsversion="2.1.6" 
time="3/28/2009 10:06:45 PM" 
file="C:\1265393.xml"
md5="a9f6aa1649e9a0f1bfad8a576f0193a0"
sha1="1d13db7f757f7cc6c71588402a0660a464a7aabb50"
logpath="C:\cwsandbox\log\1265393\run_1"
>
calltree
<process_call
index="1"
pid="1164"
filename="C:\a9f6aa1649e9a0f1bfad8a576f0193a0"
filename_hash="hash_error"
starttime="00:00:485"
startreason="AnalysisTarget"
>
calltree
<process_call
index="2"
pid="1212"
filename="C:\\DeleteFileDos.bat"
filename_hash="hash_error"
starttime="00:01:16"
startreason="CreateProcess"
>
calltree
<process_call
index="4"
pid="1244"
filename="C:\WINDOWS\system32\attrib.exe attrib c:\a9f6aa1649e9a0f1bfad8a576f0193a0 -r -a -s -h"
filename_hash="6be7ccfc384b1b05b08b7fc5aebc3bb3365cc55"
starttime="00:01:547"
startreason="CreateProcess"
>
calltree
<process_call
index="3"
pid="1252"
filename="C:\\DeleteFileDos.bat"
filename_hash="hash_error"
starttime="00:01:793"
startreason="CreateProcess"
>
calltree
<process_call
index="5"
pid="1536"
filename="C:\WINDOWS\system32\attrib.exe attrib c:\a9f6aa1649e9a0f1bfad8a576f0193a0 -r -a -s -h"
filename_hash="6be7ccfc384b1b05b08b7fc5aebc3bb3365cc55"
starttime="00:01:547"
startreason="CreateProcess"
>
calltree
<process_call
index="1"
pid="1164"
filename="C:\a9f6aa1649e9a0f1bfad8a576f0193a0"
filename_hash="hash_error"
size="11776"
md5="a9f6aa1649e9a0f1bfad8a576f0193a0"
sha1="1d13db7f757f7cc6c71588402a0660a464a7aabb50"
username="Administrator"
parentindex="0"
starttime="00:00:485"
terminationtime="00:01:172"
startreason="AnalysisTarget"
terminationreason="NormalTermination"
executionstatus="OK"
applicationtype="Win32Application"
>
<thread
tid="1176"
>
<all_section
filename="c:\a9f6aa1649e9a0f1bfad8a576f0193a0"
successful="1"
address="#\x24;400000"" end_address="#\x24;416000" size="90112"
filename_hash="hash_error"
>
<load_dll
filename="C:\WINDOWS\system32\ntdll.dll"
successful="1" address="#\x24;7C910000" end_address="#\x24;7C9C6000" size="745472"
filename_hash="cc3461f7147042c14d739a77c19166ecc8139"
>
<load_dll
filename="C:\WINDOWS\system32\kernel32.dll"
successful="1" address="#\x24;7C800000" end_address="#\x24;7C908000" size="180344"
filename_hash="eb90c03a6c632446a93cb41883378oa3"
>
<load_dll
filename="C:\WINDOWS\system32\user32.dll"
successful="1" address="#\x24;7E360000" end_address="#\x24;7E3F1000" size="393920"
filename_hash="68e9f90f6fe9b8bf237eed9b10d5f6844b79fe5"
>
<load_dll
filename="C:\WINDOWS\system32\gdi32.dll"
successful="1" address="#\x24;77EF0000" end_address="#\x24;77F39000" size="990800"
filename_hash="0f3701a7f72c7635691f7317ade3c3d63904ec96"
>
<load_dll
filename="C:\WINDOWS\system32\advapi32.dll"
successful="1" address="#\x24;77DA0000" end_address="#\x24;77E4A000" size="696320"
filename_hash="f683eb855353e341ebf5f0a5335d9d6c4e4e3b2b2"
>
<load_dll
filename="C:\WINDOWS\system32\rpcrt4.dll"
successful="1" address="#\x24;77E50000" end_address="#\x24;77EE2000" size="598016"
filename_hash="5e8715672a065171a2f266c06089a04c784292"
>
<load_dll
filename="C:\WINDOWS\system32\secur32.dll"
successful="1" address="#\x24;77FC0000" end_address="#\x24;77FD1000" size="69632"
filename_hash="bic618466fa5ec767e901bed904f53727f85810"
>
<load_dll
filename="C:\WINDOWS\system32\oleaut32.dll"
successful="1" address="#\x24;77F00000" end_address="#\x24;77F17000" size="569344"
filename_hash="31d8a173d177f470928b466db776648861c3b32e3"
>
<load_dll
filename="C:\WINDOWS\system32\msvcr70.dll"
successful="1" address="#\x24;778E0000" end_address="#\x24;77C38000" size="360448"
filename_hash="70d5f9708cc9348bb9d1009fa0f73a696b96de9" />
Visualization

- CWSandbox report is “too” detailed
- Use visualization for abstraction
  - Simple to create
  - Comprehensible

- Treemaps and Threadgraphs

- Static version (Python, matplotlib)
- Dynamic version using JavaScript (flot, Jit)
Treemaps

- Quick overview
- Displays the distribution of all operations
- No information about the sequence of operations
- Use case: Malware clustering
HOOKSHELL-207706_1265393
- set_file_attr
- write_value
- get_file_attr
- query_value
- enum_modules
- enum_process
- get_system_time
- load_dll
- find_file
- open_key
- create_key
- create_process
- read_value
- load_image
- create_open_file
- get_system_directory

ADULTBROWSER-227269_1265984
- com_get_class_object
- enum_keys
- query_value
- create_thread
- load_dll
- find_file
- open_key
- get_system_time
- create_socket
- get_system_directory
- com_create_instance
- open_file
- get_system_directory
- sleep
- connect_socket
Assignments

Cluster 1 (968 members, 3 prototypes, 23.6% of data, 1 label, 23.6% cumulative)

Cluster 2 (298 members, 8 prototypes, 7.3% of data, 1 label, 30.8% cumulative)

Cluster 3 (184 members, 2 prototypes, 4.5% of data, 1 label, 35.3% cumulative)

Cluster 4 (178 members, 2 prototypes, 4.3% of data, 1 label, 39.6% cumulative)

Cluster 5 (145 members, 1 prototype, 3.5% of data, 1 label, 43.1% cumulative)

Cluster 6 (134 members, 3 prototypes, 3.3% of data, 1 label, 46.4% cumulative)

Cluster 7 (133 members, 1 prototype, 3.2% of data, 1 label, 49.6% cumulative)

Cluster 8 (123 members, 4 prototypes, 3.0% of data, 1 label, 52.6% cumulative)

Cluster 9 (103 members, 3 prototypes, 2.5% of data, 1 label, 55.1% cumulative)
Cluster 5 (145 members, 1 prototypes, 3.5% of data, 1 label, 43.1% cumulative)
Cluster 17 (30 members, 5 prototypes, 0.7% of data, 1 label, 64.6% cumulative)
Threadgraph

- Detailed view of the monitored behavior
- Every monitored operation of all executed threads
- Sequential order preserved

- Truncated Reports only (static version)
- Zoom necessary (JavaScript)
Threadgraph of SampleID: 129956 AnalysesID: 665508 (level2)

< open_key key = HKEY_LOCAL_MACHINE\Software\Microsoft\Windows NT\CurrentVersion\AppCompatFlags\Layers />

Operation 36 of Thread 1 (open_key)
Analysis of malicious PDFs

- Monitor Adobe Acrobat Reader within CWSandbox
- Autoupdate disabled

- Short experiment:
  - 200 benign PDFs
  - 17 malicious PDFs

- Two different Treemaps
- Malicious operations are observable at once
Conclusion

• First attempt on visualization - need Feedback
• Plots are accessible on CWSandbox.org
  • Static + JavaScript version

• Study on image clustering and classification
• Development of further visualization types
Questions?

trinius@uni-mannheim.de