SOC
FORUM
2023

Подводные камни экспертных утилит при работе SOC с инцидентами

Обо мне

SOC FORUM 2023

- Ранее работал в BI.ZONE, Group-IB
- 13 лет в сферах Digital Forensics, Incident Response, Malware Analysis, Threat Intelligence
- Занимался обратной разработкой форматов NTFS, FAT12/16/32, exFAT, реестра Windows, теневых копий
- Создаю утилиты и библиотеки:
  - yarp (парсер файлов реестра Windows),
  - dfir_ntfs (парсер томов NTFS, FAT12/16/32 exFAT, а также теневых копий),
  - winmem_decompress (распаковщик сжатых страниц из дампов памяти),
  - и многое другое…
- Ищу и нахожу уязвимости в парсерах файловых систем и загрузчиках

Максим Суханов

Ведущий эксперт CICADA8 MTC RED

masuhano@mts.ru

CICADA8

Про какие утилиты речь?

SOC FORUM 2023

- Агенты для мониторинга конечных точек
- Сборщики артефактов с конечных точек
- Парсеры артефактов
- IoC-сканеры

- А еще: библиотеки для всего этого

Два опасных типа ошибок

SOC FORUM 2023

Программа дает «ожидаемые»,
но ошибочные результаты

Программа приводит
к отказам на целевом хосте

КЕЙС

Отказ на целевом хосте

CICADA8

libyara

SOC FORUM 2023

- Всем известная библиотека для поддержки YARA
- Используется многими компаниями в собственных продуктах
- Поддерживает сканирование буферов, файлов и процессов

URL: https://github.com/VirusTotal/yara

Что и как хотели сделать?

SOC FORUM 2023

- Нужно просканировать память процессов на целевом хосте
- Целевая ОС: Linux
- Имеется простое YARA-правило, на базе строк
- Сканирование реализовано путем вызова утилиты yara в скрипте:
  ps + grep + xargs + yara

Жалоба клиента

SOC FORUM 2023

С утилитой yara есть сложности. ***При проверке наличия сигнатуры в запущенном процессе последний рестартится (проверял на сервисе kafka).*** *В документации и интернете такое поведение не описано и никаких дополнительных ключей для запуска я не нашел.*

Пока не могу запустить скрипт на всех ВМ – это чревато тотальными последствиями

Изучаем: dmesg

SOC FORUM 2023

```
[9323793.923902] [  75037]     0 75037   808927   405348  4333568    92245             0 java
[9323793.923904] [  75038]     0 75038  1226180   137972  1753088    16576             0 java
[9323793.923905] [  75039]     0 75039   813064    40716   876544    15850             0 java
[9323793.923906] [  34891]    89 34891    25680       53   208896      202             0 pickup
[…]
[9323793.923919] [  41554]     0 41554   271098   260155  2215936     3874             0 yara
[9323793.923921] oom-
kill:constraint=CONSTRAINT_NONE,nodemask=(null),cpuset=/,mems_allowed=0,global_oom,task_memcg=/
system.slice/zookeeper.service,task=java,pid=75037,uid=0
[9323793.923957] Out of memory: Killed process 75037 (java) total-vm:3235708kB, anon-rss:1621392kB,
file-rss:0kB, shmem-rss:0kB, UID:0 pgtables:4232kB oom_score_adj:0
[9323793.985446] oom_reaper: reaped process 75037 (java), now anon-rss:0kB, file-rss:0kB, shmem-rss:0kB
```

Изучаем: strace для java

SOC FORUM 2023

```
strace: Process 67185 attached
futex(0x7fe1cec289d0, FUTEX_WAIT, 67268, NULL) = ?
+++ killed by SIGKILL +++
```

Изучаем: strace для yara

SOC FORUM 2023

```
pread64(4, " \0\0\210\341\177\0\0\0\0\0\0\0\0\0\20\2\0\0\0\0\0\0\20\2\0\0\0\0\0"..., 135168,
140606626070528) = 135168
[...]
pread64(4, "\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0\0"..., 66973696,
140606626205696) = 30527488
[...]
pread64(4, 0x7f752ab26010, 135168, 140606693179392) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 66973696, 140606693314560) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 286720, 140606760288256) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 66822144, 140606760574976) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 135168, 140606827397120) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 66973696, 140606827532288) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 135168, 140606894505984) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 66973696, 140606894641152) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 16384, 140606968012800) = -1 EIO (Input/output error)
pread64(4, 0x7f752ab26010, 1036288, 140606968029184) = -1 EIO (Input/output error)
```

Изучаем: карта памяти для java

SOC FORUM 2023

```
7f00701de000-7f00702db000 rw-p 00000000 00:00 0
7f00702db000-7f00702df000 ---p 00000000 00:00 0
7f00702df000-7f00703dc000 rw-p 00000000 00:00 0
7f00703dc000-7f00703e0000 ---p 00000000 00:00 0
7f00703e0000-7f00704dd000 rw-p 00000000 00:00 0
7f00704dd000-7f00704e1000 ---p 00000000 00:00 0
7f00704e1000-7f00705de000 rw-p 00000000 00:00 0
7f00705de000-7f00705e2000 ---p 00000000 00:00 0
7f00705e2000-7f00706df000 rw-p 00000000 00:00 0
7f00706df000-7f00706e3000 ---p 00000000 00:00 0
7f00706e3000-7f00707e0000 rw-p 00000000 00:00 0
7f00707e0000-7f00707e4000 ---p 00000000 00:00 0
7f00707e4000-7f00708e1000 rw-p 00000000 00:00 0
И так далее!
```

Выводы

1. Целевой процесс (java) потребляет много памяти
2. Значительная часть памяти целевого процесса находится в swap
3. Чтение памяти целевого процесса со стороны утилиты yara приводит к перемещению страниц из swap в ОЗУ
   - В libyara есть оптимизация для чтения отображенных в память файлов, но не анонимных страниц…
4. Это увеличивает потребление оперативной памяти целевым процессом
5. И приходит OOM Killer!

Сообщение об ошибке

SOC FORUM 2023

msuhanov commented on Dec 17, 2021 • edited

Hello.

YARA version: 4.1.3.

In one environment (GNU/Linux), attempts to scan several processes result in the KILL signal being sent to *these processes* (not to the *yara* process) by the OOM killer.

Further investigation demonstrated that each process has a large amount of anonymous mappings, most of them have their pages swapped. Reading such memory brings swapped pages into the resident set, and the OOM killer is triggered (against the process being scanned).

I see some tweaks merged into the main branch (`605b2ed` and `063cbfb`), but they won't protect against this situation (because mappings are anonymous, there are many of them, and there are multiple processes sharing the same properties).

My proposal is to implement a command-line argument to disable scanning of swapped pages. Later, this functionality could be extended and swapped pages could be read directly from a swap file (thus, memory pressure would affect the scanning process only).

I'm attaching a patch for your review (2021-12-18 08:00 UTC: fixed a wrong type, uploaded again).
yara_patch_ignore_swapped_pages.txt

(I will make a pull request once the proposal is accepted.)

КЕЙС

Это же работающая система, данные могут меняться

CIGADA8

Velociraptor

SOC FORUM 2023

- Известная утилита для сбора телеметрии и артефактов с конечных точек
- Может работать в режиме «быстро собрать артефакты и выйти»
- Имеет поддержку NTFS (для чтения заблокированных файлов)

URL: https://github.com/Velocidex/velociraptor

Наблюдение

При сборе артефактов слишком часто повреждаются копии файлов: *от $MFT до кустов реестра*

ПРИМЕР: копия куста реестра Amcache имеет отсутствующие в ее середине данные, а имеющиеся в середине данные расположены по неправильным («сдвинутым») смещениям

Другие утилиты (например, KAPE) к такому не приводят – проблема точно не в ОС!

Отладка

SOC FORUM 2023

Успешное воспроизведение ошибки под отладчиком – когда копирование данных происходит очень медленно
Причина? Я не знал! *(Пока разработчик не пояснил.)*

Сообщение об ошибке

msuhanov commented on Oct 11, 2022

Hello.

When collecting data in real-world environments using Velociraptor, I observed the following corruption scenarios.

1. Version: 0.6.5

One case: the $MFT file is empty (0 bytes). Other files are collected properly.
Another case: the $MFT file does not start with expected data (many file record segments located in the beginning of the source $MFT file are not present in its copy – the first kibibytes of the file are absent).
Two more cases: the Amcache hive has missing hive bins and existing (successfully copied) hive bins become misaligned somewhere in the middle of the copied file.

The .idx file for one case with the Amcache hive is:

```
{"file_offset":0,"original_offset":0,"file_length":131072,"length":131072}
{"file_offset":131072,"original_offset":131072,"file_length":163840,"length":163840}
{"file_offset":229376,"original_offset":294912,"file_length":229376,"length":229376}
{"file_offset":458752,"original_offset":524288,"file_length":262144,"length":262144}
{"file_offset":720896,"original_offset":786432,"file_length":262144,"length":262144}
{"file_offset":983040,"original_offset":1048576,"file_length":131072,"length":131072}
{"file_offset":1114112,"original_offset":1179648,"file_length":0,"length":131072}
```

As you can see, `file_offset` is not equal to `original_offset` after the first two runs.

(Another .idx file has similar misaligned runs.)

2. Version: 0.6.6

I was able to reproduce this (or a very similar) bug using the debugger while slowing down the execution flow of the Velociraptor binary. In particular, the following condition is occasionally met (before the corresponding data run is fully read):

velociraptor/utils/copy.go
Line 70 in 999b62d

```
70      if n == 0 {
```

In the previous version (0.6.5), the `maybeCollectSparseFile` routine (`reporting/container.go`) does not add padding bytes for partial reads, thus creating misaligned data. In the current version (0.6.6), there is no misaligned data, but there are null bytes instead of real data, because of the padding:

Первичный ответ разработчика

SOC FORUM 2023

scudette commented on Oct 12, 2022 — Contributor

This is expected as sparse files mean that there is a gap in the file. Velociraptor does not pad the gap instead it copies the runs that contain data and create an idx file that tells it which runs are present. The file offset refers to the offset in the collected file while the original offset refers to the offset in the original file. If they are not equal this is because there is a sparse run in the original that was skipped. NTFS has many such sparse runs all the time so this happens in many cases. Examples I have seen include mft and registry files and occasionally evtx files.

In previous versions downloading the file from the uploads tab in the GUI will pad out the gaps while the export zip file will include the idx file.

In the next version users have a choice to also pad out files in the export download zip.

scudette commented on Oct 12, 2022 — Contributor

It looks like a read from the underlying NTFS parser is returning less data than we expect - so we pad it to maintain alignment. This might be an ntfs parser bug or just a side effect of the system changing while the acquisition is taking place (for example in the ntuser.dat file the MFT might have the runs changed between the time we parse the mft and try to copy the file out).

Are you saying that repeated acquisitions sometimes work better?

Итог

- Библиотека NTFS, встроенная в Velociraptor, поддерживает периодическое «обновление» дескриптора, используемого для чтения с диска
- Дескриптор закрывается и тут же открывается еще раз.
  Каждые 60 секунд!
- Попытка чтения из закрытого дескриптора возвращает ошибку
- Это приводит к неполной записи в целевой файл
  - В старых версиях это приводит к «сдвигу» смещений в целевом файле!
  - В более новых – «сдвига» нет, но «зазоры» заполняются нулевыми байтами…

Бонус

Если вендор заявляет, что при экспорте файла $MFT
с работающей системы (под управлением современной версии ОС Windows) в его копии могут оказаться ошибки, **то вендор не осознает, что это ошибка в его софте**

КЕЙС

Во все места можно добраться рекурсивно

CIOADA8

Loki

SOC FORUM 2023

- IoC-сканер с поддержкой YARA
- Ранее (до 04.2023) поставлялся только в 32-битном варианте
- Кое-кто рекомендует эту утилиту клиентам для сканирования инфраструктуры

URL: https://github.com/Neo23x0/Loki

Что хотели сделать?

SOC FORUM 2023

Порекомендовать клиенту эту утилиту (с нашими YARA-правилами)

Но ведь ее нужно проверить перед использованием!

32-битная программа в 64-битной ОС

SOC FORUM 2023

- Что хранится в директории «*C:\Windows\System32*»?
- Там хранится содержимое из «*C:\Windows\SysWOW64*»!
  Оригинальное содержимое – в «*C:\Windows\Sysnative*»…
- Но этой директории нет в листинге «*C:\Windows*»! **Рекурсивным обходом до нее не добраться!**

Original Path	Redirected Path for 32-bit x86 Processes
%windir%\System32	%windir%\SysWOW64
%windir%\lastgood\system32	%windir%\lastgood\SysWOW64
%windir%\regedit.exe	%windir%\SysWOW64\regedit.exe

32-bit applications can access the native system directory by substituting %windir%\Sysnative for %windir%\System32. WOW64 recognizes Sysnative as a special alias used to indicate that the file system should not redirect the access. This mechanism is flexible and easy to use, therefore, it is the recommended mechanism to bypass file system redirection. Note that 64-bit applications cannot use the Sysnative alias as it is a virtual directory not a real one.

Сообщение об ошибке

SOC FORUM 2023

The tool doesn't scan the Sysnative folder #206

Closed **msuhanov** opened this issue on Dec 15, 2021 · 2 comments

msuhanov commented on Dec 15, 2021 • edited

Hello.

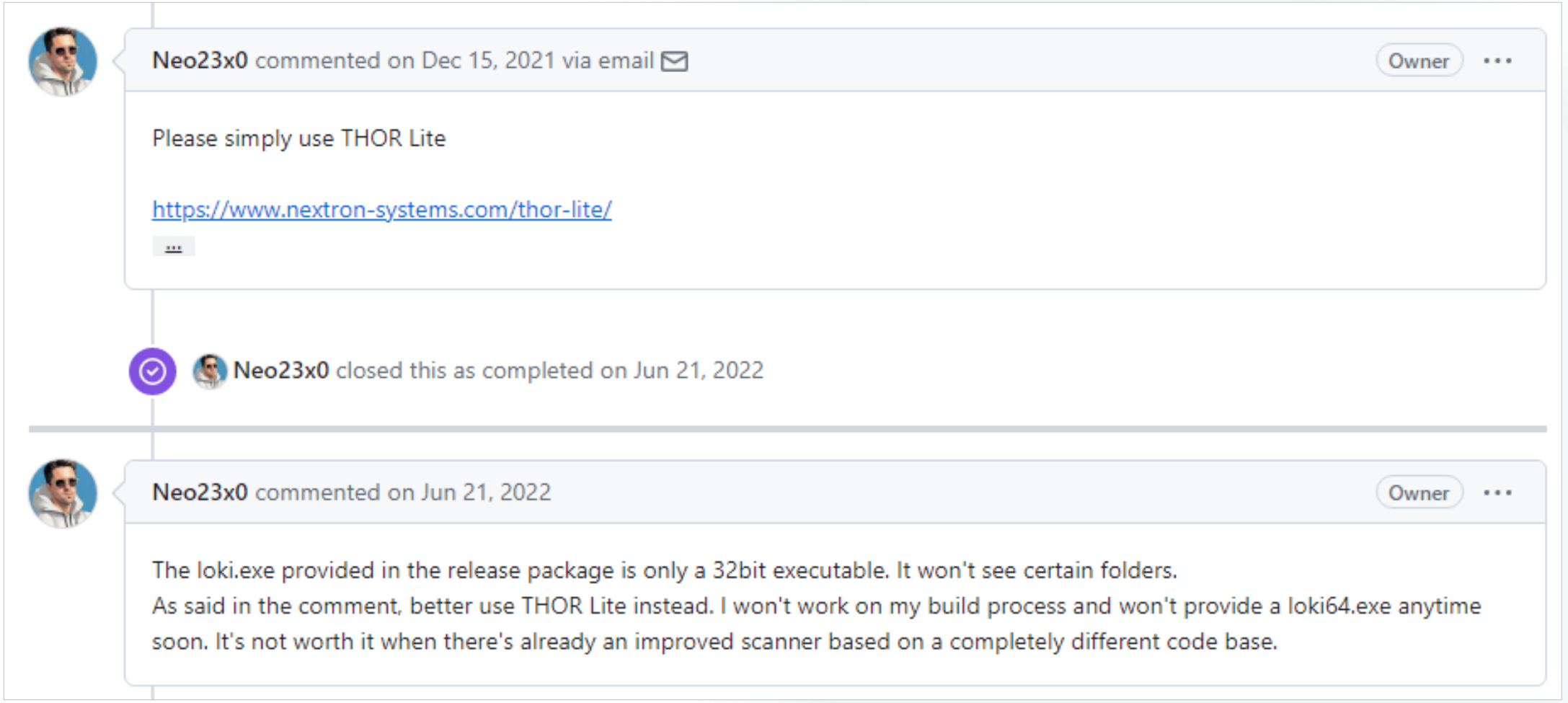
I have created a path-based rule for two files, one placed in the C:\Windows\ folder and another one placed in the C:\Windows\System32\ folder (both match the same rule used for this test). Since the tool is running under the Wow64 subsystem, the real System32 folder is called Sysnative (C:\Windows\Sysnative\). And the tool doesn't inspect that folder during the scan.

The problem is that the Sysnative folder isn't visible when enumerating the Windows folder, so one needs to explicitly try that name.

Here is related output (the tool is running with administrator privileges):

Ответ разработчика

SOC FORUM 2023



Neo23x0 commented on Dec 15, 2021 via email — Owner

Please simply use THOR Lite

https://www.nextron-systems.com/thor-lite/

...

Neo23x0 closed this as completed on Jun 21, 2022

Neo23x0 commented on Jun 21, 2022 — Owner

The loki.exe provided in the release package is only a 32bit executable. It won't see certain folders.
As said in the comment, better use THOR Lite instead. I won't work on my build process and won't provide a loki64.exe anytime soon. It's not worth it when there's already an improved scanner based on a completely different code base.

КЕЙС

Новый способ хранения данных исключает старый

CICADA8

Amcache

Куст реестра, хранящий некоторые данные телеметрии ОС Windows

В нем можно найти записи об исполняемых файлах

- В т. ч. о ранее запущенных исполняемых файлах!
- И пути к исполняемым файлам, и SHA1-хеши от первых 31.457.280 байт каждого зафиксированного исполняемого файла…

Два способа хранения данных – старый и новый (различается дерево ключей и значений)!

Два способа хранения данных

SOC FORUM 2023

```
Key path: Root\InventoryApplicationFile\
0000009af2c85d3391d41ce56094d48b56489f99d853
Last written timestamp (UTC): 2016-11-13 23:56:52.851288
Access bits: 2
Owner SID: S-1-5-18

Value name: ProgramId
Value type: REG_SZ
Data size: 90
Data (decoded):
000050cfd8ca67cf7180687da7e541d77ed500000000

Value name: FileId
Value type: REG_SZ
Data size: 90
Data (decoded):
00008b8db3434cac650ba795ff9aa8f89a24e4b83410

Value name: LowerCaseLongPath
Value type: REG_SZ
Data size: 122
Data (decoded):
c:\program files (x86)\adobe\reader 11.0\reader\acrord32.exe

[…]

Value name: Size
Value type: REG_SZ
Data size: 18
Data (decoded):
0x147e88
```

```
Key path: Root\File\a2c3a639-3770-11e3-9713-
806e6f6e6963\300001a337
Last written timestamp (UTC): 2016-11-12 02:08:22.131144
Access bits: 2
Owner SID: S-1-5-18

Value name: 17
Value type: REG_QWORD
Data size: 8
Data (decoded):
131220344719980488

Value name: 15
Value type: REG_SZ
Data size: 100
Data (decoded):
c:\program files\DAUM\potplayer\PotPlayerMini.exe

[…]

Value name: 101
Value type: REG_SZ
Data size: 90
Data (decoded):
00008976ca7fa6b29f51a9b7a81b3488222a3e66cd52

Value name: 100
Value type: REG_SZ
Data size: 90
Data (decoded):
0000ebb563493e2c2ffe6278417fd2f711b20000ffff
```

AmcacheParser

SOC FORUM 2023

- Популярный парсер для куста реестра Amcache
- Поддерживает оба способа хранения данных. Выбирает один из двух самостоятельно, при обработке куста

URL: http://ericzimmerman.github.io/

Ошибка

Оба способа могут использоваться одновременно – в одном кусте реестра Amcache!

НЕВЕРНОЕ ДОПУЩЕНИЕ:

В одном кусте реестра Amcache данные телеметрии могут храниться только одним способом

Хотите ли вы пропустить Mimikatz из-за этого?

Сообщение об ошибке

SOC FORUM 2023

Open

AmcacheParser ignores "old format" entries if "new format" entries are found #203

msuhanov opened this issue on Mar 29 · 1 comment

Here is the output of your tool:

```
AmcacheParser version 1.5.1.0

Author: Eric Zimmerman (saericzimmerman@gmail.com)
https://github.com/EricZimmerman/AmcacheParser

Command line: -f Amcache.hve --csv .

Warning: Administrator privileges not found!

Error parsing ProgramsEntry at {11517B7C-E79D-4e20-961B-75A811715ADD}\Root\Inventor

[...]

C:\Users\MS\Desktop\amcache\Amcache.hve is in new format!

Total file entries found: 486
Total device containers found: 141
Total device PnPs found: 105
Total drive binaries found: 284
Total driver packages found: 16

Found 18 unassociated file entry

Results saved to: .

Total parsing time: 1,323 seconds
```

Here is a key that isn't parsed by your tool:

```
Key path: Root\File\dd31621c-4f8f-11e4-9f49-806e6f6e6963\1000004513b
Last written timestamp (UTC): 2020-07-29 16:59:41.147582
Access bits: 0
Owner SID: S-1-5-32-544

[...]

Value name: 15
Value type: REG_SZ
Data size: 112
Data (decoded):
c:\program files (x86)\opera\69.0.3686.77\installer.exe^@

[...]

---
```

("^@" marks the null byte.)

So, the Amcache hive contains entries having both the new and old formats. Those using the old format aren't parsed (e.g., the path quoted above is missing in the CSV files).

Unfortunately, I can't share the hive file.

Обобщение

Обобщение

- Специалист склонен чересчур доверять результатам работы программы, если они соответствуют его ожиданиям

  Или если наблюдаемые аномалии могут быть объяснены чем-то ожидаемым!

- Ошибки могут оставаться незамеченными годами

- Часть ошибок очевидна, если сопоставить реализацию (свою) и документацию (чужую)
  - Но документация к программным интерфейсам ОС уже давно за гранью обозримости

- Невозможно преодолеть последствия ошибок за счет повторной обработки тех же данных программой «получше»
  - Кто это будет делать и когда?

Тест-кейсы

Существуют тест-кейсы для сложных случаев обработки NTFS и файлов реестра Windows

- https://github.com/msuhanov/ntfs-samples
- https://github.com/msuhanov/regf-samples
- https://github.com/msuhanov/yarp/tree/master/hives_for_tests
- https://github.com/msuhanov/dfir_ntfs/tree/master/test_data

Но они не покрывают «высокоуровневые ситуации»
И они не охватывают работу утилиты в «реальной инфраструктуре»

Но…

SOC FORUM 2023

1

Одни и те же ошибки имеют свойство повторяться в разных местах, в разных программах

2

Можно охватить тест-кейсами имеющиеся баг-репорты для открытого ПО

3

И создать чек-лист для их применения «на своем ПО»

Примеры

- Если мы сканируем содержимое диска рекурсивно, то актуален ли для нас тест-кейс с Sysnative?
- Если мы сканируем память другого процесса, то защищаем ли мы его страницы от выхода из swap? Добро пожаловать в тест-кейс на базе виртуальной машины
- Если мы парсим Amcache, то поддерживаем ли мы оба способа хранения данных, примененных одновременно в одном кусте реестра? Добро пожаловать в тест-кейс на базе публичных данных
- Velociraptor?

SOC FORUM 2023

Вопросы?

CIOADA8