Hunting Adversaries with "rastrea2r" and Machine Learning

Ismael Valenzuela, @aboutsecurity
Principal Engineer
GSE 132, SANS Certified Instructor

SANS SIEM Summit 2017 – Scottsdale, AZ
Twitter: @aboutsecurity

My 16+ year career in cybersecurity

Computer Geek (Linux User)

PenTester (Shell Rockstar)

Forensicator, IR & Threat Researcher (Sleep Deprivation)

Principal Engineer @McAfee & SANS Certified Instructor (lots of Photoshop)
A THANK YOU EMAIL FROM SANS

AND I HAVEN'T EVEN FINISHED SIEM NETWARS!
Subject: SANS SIEM Summit - Your complimentary iPhone X

From: Ismael Valenzuela – ivalenzuela@sans.edu

Dear attendee,

Thank you for spending time with us in Scottsdale, AZ for the amazing first edition of the SANS SIEM Summit!

As promised, you can request your complimentary iPhone X by filling out this form: SANS iPhoneX Order

We look forward to seeing you at the next event!

PS. For questions around shipping or any other issues with the delivery of the phones, please contact Eric Conrad on his personal cell: +1 555 555-5555

Regards
Ismael.
Dear attendee,

Thank you for spending time with us in Scottsdale, AZ for the amazing first edition of the SANS SIEM Summit!

As promised, you can request your complimentary iPhone X by filling out this form: SANS iPhoneX Order<http://aboutsecurity.io/iPhoneX>

We look forward to seeing you at the next event!

PS. For questions around shipping or other logistic issues with the delivery of the phones, please contact Eric Conrad on his personal cell: +1 555 555-5555

Regards

Ismael.
Hey Justin, this is the invoice for the iPhone X I ordered for the Summit attendees: aboutsecurity.io/iPhoneX - Did you get that money you were waiting for? I think you said it was from one of your Nigerian relatives right? Wasn't he a prince or something? Anyways, I hope you get the money soon else I don't know who's gonna pay for all these phones...
Hey Seth, looking forward to seeing you at the SIEM Summit tomorrow! Look what I got from Black Friday. I bought one for each Summit attendee!! [aboutsecurity.io/iphoneX](http://aboutsecurity.io/iphoneX) - check it out, it's gonna blow you away! ;)}
What do you want to do with invoice.rtf?

Size: 8.04 KB
From: 52.179.101.199

- **Open**
  The file won’t be saved automatically.

- **Save**

- **Save as**
Speeding up collection with rastrea2r

Triage and Hunting for IOCs with ‘gusto’ and style 😊

Rastrea2r (pronounced rastreador):

- [https://github.com/aboutsecurity/rastrea2r](https://github.com/aboutsecurity/rastrea2r) (opensource!)

- Command line tool (coz command line is SEXY!)
- Python / Multiplatform (win32/64, linux and osx)
- Uses a REST API to report YARA scans
- Wrapper to sysinternal, system command and 3rd party tools
- Easy to integrate with McAfee ePO (but also distributable via SSCM, etc.)
- Packaged binaries available on GitHub

- Less time collecting! More time doing analysis!!
Current functionality in rastrea2r v0.7.1

And it works on Fridays too! (if you dare…)

- **yara-disk**: Yara scan for file/directory objects on disk
- **yara-mem**: Yara scan for running processes in memory
- **memdump**: Acquires a memory dump from the endpoint **Win only
- **tria**ge: Collects triage information from the endpoint **Win only
- **web-hist**: Collects and parses browsing history for all users in the system (ie, firefox, chrome and opera). **Win only **NEW
- **prefetch**: Collects and parses prefetch folder from the endpoint **Win only **NEW

Obtaining HELP:

```
rastrea2r_win32_v0.7.1.exe –h
```
Integration with CyLR

- Collects forensics artifacts from NTFS in a fast and secure fashion.
- Bypasses Windows API (raw file collection process)
- Artifacts are collected in memory

- By Alan Orlikoski and Jason Yegge
- https://github.com/rough007/CyLR

The standard list of collected artifacts are:

"C:\Windows\System32\config"
"C:\Windows\ProgramData\Microsoft\Windows\Start Menu\Programs\Startup"
"C:\Windows\Prefetch"
"C:\Windows\Tasks"
"C:\Windows\SchedLgU.Txt"
"C:\Windows\System32\winevt\logs"
"C:\Windows\System32\drivers\etc\hosts"
"C:\$MFT"
THAT NIGERIAN PRINCE NEVER EMAILED BACK

I HOPE HE'S OKAY
rastrea2r in action – 15 minutes of TRIAGE
rastrea2r in action

The **web-history** plugin returns a CSV file with the browsing history for a user (or all users on the computer) from Firefox, Chrome, IE and Opera:
Evidence of **persistence**: Sysinternals startup-list

```plaintext
Caption=zCLcIIJndaI
Command=C:\Users\DEMOUS~1\AppData\Local\Temp\WplgaQeK.vbs
Description=zCLcIIJndaI
Location=HKU\S-1-5-21-2126574468-858435778-3403648540-2604\SOFTWARE\Microsoft\Windows\CurrentVersion\Run
SettingID=
User=SCP\demouser1

Caption=VMware User Process
Command="C:\Program Files\VMware\VMware Tools\vmtoolsd.exe" -n vmusr
Description=VMware User Process
Location=HKLM\SOFTWARE\Microsoft\Windows\CurrentVersion\Run
SettingID=
User=Public
```
### Process Information for WIN81

<table>
<thead>
<tr>
<th>Name</th>
<th>Pid</th>
<th>Pri</th>
<th>Thd</th>
<th>Hnd</th>
<th>VM</th>
<th>WS</th>
<th>Priv</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>64</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>System</td>
<td>4</td>
<td>8</td>
<td>114</td>
<td>940</td>
<td>139412</td>
<td>1504</td>
<td>135424</td>
</tr>
<tr>
<td>smss</td>
<td>292</td>
<td>11</td>
<td>2</td>
<td>44</td>
<td>4212</td>
<td>1640</td>
<td>284</td>
</tr>
<tr>
<td>csrss</td>
<td>396</td>
<td>13</td>
<td>8</td>
<td>349</td>
<td>45856</td>
<td>3848</td>
<td>1716</td>
</tr>
</tbody>
</table>

**EQNEDT32**

cscript
conhost
gBiwSNwnbyAD

<table>
<thead>
<tr>
<th>Process</th>
<th>Pid</th>
<th>Pri</th>
<th>Thd</th>
<th>Hnd</th>
<th>VM</th>
<th>WS</th>
<th>Priv</th>
</tr>
</thead>
<tbody>
<tr>
<td>taskhost</td>
<td>1332</td>
<td>6</td>
<td>10</td>
<td>298</td>
<td>1192604</td>
<td>17340</td>
<td>11164</td>
</tr>
<tr>
<td>wuaucrt</td>
<td>2760</td>
<td>8</td>
<td>1</td>
<td>116</td>
<td>80396</td>
<td>6092</td>
<td>1360</td>
</tr>
<tr>
<td>taskhostex</td>
<td>3504</td>
<td>8</td>
<td>6</td>
<td>212</td>
<td>87552</td>
<td>7796</td>
<td>2140</td>
</tr>
<tr>
<td>svchost</td>
<td>956</td>
<td>8</td>
<td>28</td>
<td>1153</td>
<td>138212</td>
<td>23044</td>
<td>11900</td>
</tr>
<tr>
<td>svchost</td>
<td>1004</td>
<td>8</td>
<td>22</td>
<td>650</td>
<td>4194303</td>
<td>55200</td>
<td>45456</td>
</tr>
<tr>
<td>WuDFHost</td>
<td>2360</td>
<td>8</td>
<td>9</td>
<td>249</td>
<td>38032</td>
<td>5696</td>
<td>3052</td>
</tr>
<tr>
<td>dasHost</td>
<td>2904</td>
<td>8</td>
<td>3</td>
<td>332</td>
<td>72572</td>
<td>11508</td>
<td>3520</td>
</tr>
<tr>
<td>spoolsv</td>
<td>1644</td>
<td>8</td>
<td>8</td>
<td>429</td>
<td>77024</td>
<td>12084</td>
<td>4496</td>
</tr>
<tr>
<td>svchost</td>
<td>1068</td>
<td>8</td>
<td>22</td>
<td>516</td>
<td>123964</td>
<td>22912</td>
<td>16532</td>
</tr>
<tr>
<td>FrameworkService</td>
<td>1232</td>
<td>6</td>
<td>29</td>
<td>431</td>
<td>93236</td>
<td>1684</td>
<td>7168</td>
</tr>
</tbody>
</table>
Netlogon  
[System]  
TCP  0.0.0.0:49157  0.0.0.0:0  LISTENING  524  
Can not obtain ownership information  
TCP  0.0.0.0:49160  0.0.0.0:0  LISTENING  2232  
PolicyAgent  
[System]  
TCP  0.0.0.0:49163  0.0.0.0:0  LISTENING  532  
[System]  
TCP  127.0.0.1:445  127.0.0.1:50230  ESTABLISHED  4  
Can not obtain ownership information  
TCP  127.0.0.1:50230  127.0.0.1:445  ESTABLISHED  4  
Can not obtain ownership information  
TCP  192.168.20.104:139  0.0.0.0:0  LISTENING  4  
Can not obtain ownership information  
TCP  192.168.20.104:49934  192.168.20.5:445  ESTABLISHED  4  
Can not obtain ownership information  
TCP  192.168.20.104:49943  192.168.20.5:445  ESTABLISHED  4  
Can not obtain ownership information  
TCP  192.168.20.104:49944  192.168.20.5:445  ESTABLISHED  4  
Can not obtain ownership information  
TCP  192.168.20.104:49945  192.168.20.5:445  ESTABLISHED  4  
Can not obtain ownership information  
TCP  192.168.20.104:50182  192.168.1.221:9999  ESTABLISHED  12456  
[EQNEDT32.EXE]  
TCP  192.168.20.104:50229  192.168.1.221:9999  SYN_SENT  9320  
[gBlwSNwnbyAD.exe]  
TCP  [::]:135  [::]:0  LISTENING  676  
RpcSs
Can we **automate** this? Some engineering issues...

3 **different analysts** writing scripts to parse the output of rastrea2r:

- Web History (naked IPs)
- Process name prevalence (NSRL)
- Process name entropy

Can the 3 analysts **reuse** each other’s scripts?

- No input/output standardization = no piping
- No agreement on how objects are manipulated
- No data traceability
From Collection to Mining
Solution: **OpenCNA** (Collect, Normalize & Analyze)

- CSV Definition of entities (file, user, system, etc)
- Python SDK
  - containing python wrappers for CSV
- Simple Infrastructure to pipe and report
- Sample Miners provided today as dockerized images:
  - Webhistory
  - Process prevalence (NSRL)
  - Process entropy
DOCKER?

DOCKER DOCKER DOCKER DOCKER
Pipelining Process
Pipelining Process

Sources: Entity extractors for rastrea2r.

Distribution: Using Docker

Stacking process: Signature Definition

Parsing: Logs to CSV

Traceability: Extending types
ls
README.md  analyzer  build.sh  cli  csvtools  data  normalizer  setup.py

$ cat build.sh

echo "Creating the docker images..."

echo "Creating the normalizer"
docker build -t opencna/normalizer normalizer/

echo "Creating the analyzers"
docker build -t opencna/analyzer/process-uncommon-nsrl analyzer/process-uncommon-nsrl/
docker build -t opencna/analyzer/random-process-name analyzer/random-process-name/
docker build -t opencna/analyzer/web-history analyzer/web-history/

pip install .
I know what you're thinking...

where is the MACHINE LEARNING??
Machine Learning (sorry, sorry, sorry)

**Why** do we use ML?
- To complement the human abilities through Human Machine Teaming

**What** for?
- Random string identification, anomalous web behavior
- More coming soon

**What algorithms?**
- Random Forests
- Principal Component Analysis (PCA)
- Bayesian statistics (Multinomial-Dirichlet models)
- Standard statistics (ANOVA)
- Markov processes
- Information theory (Shannon entropy)
CVE-2017-11882

- Buffer overflow allows RCE in EQNEDT32.EXE.
- Currently exploited by criminal groups through malicious RTF documents.
- Works on all versions of Office (inc. 365) all architectures and all versions of Windows released in the last 17 years!! Patch now!!

Open Source Tools - Sponsored by McAfee Foundstone

Repositories 11  People 7  Teams 1  Projects 1  Settings

Search repositories...  Type: All  Language: All

OpenCNA
OpenCNA (Collection & Normalization & Analysis)
Python  Updated 32 seconds ago

ExpertInvestigationGuides
Expert Investigation Guides
2 ★  Updated on Sep 26

InvestigationPlaybookSpec
OpenCNA

OpenCNA (Collection, Normalization and Analysis) tool, to collect data from endpoints (actually, we use rastrea2r to do that), normalize (parse) that data and analyze it.

Requirements

In order to make it easy to deploy, we rely on docker:

- Install and start Docker CE (https://www.docker.com/community-edition)
“Git clone” and send us feedback!

Checkout our public repos on GitHub:
https://github.com/Foundstone/OpenCNA

OpenCNA is developed and maintained by an amazing team including Gabriel Infante-Lopez, Matias Marenchino, Jorge Couchet, Nicolas Rivero and Ismael Valenzuela.

Follow us on twitter for updates @aboutsecurity @Foundstone