
GREAT
IDEAS
Threat hunting in new kicks:
VirusTotal code similarity
with KTAE
kaspersky
Powered by
TheSAScon

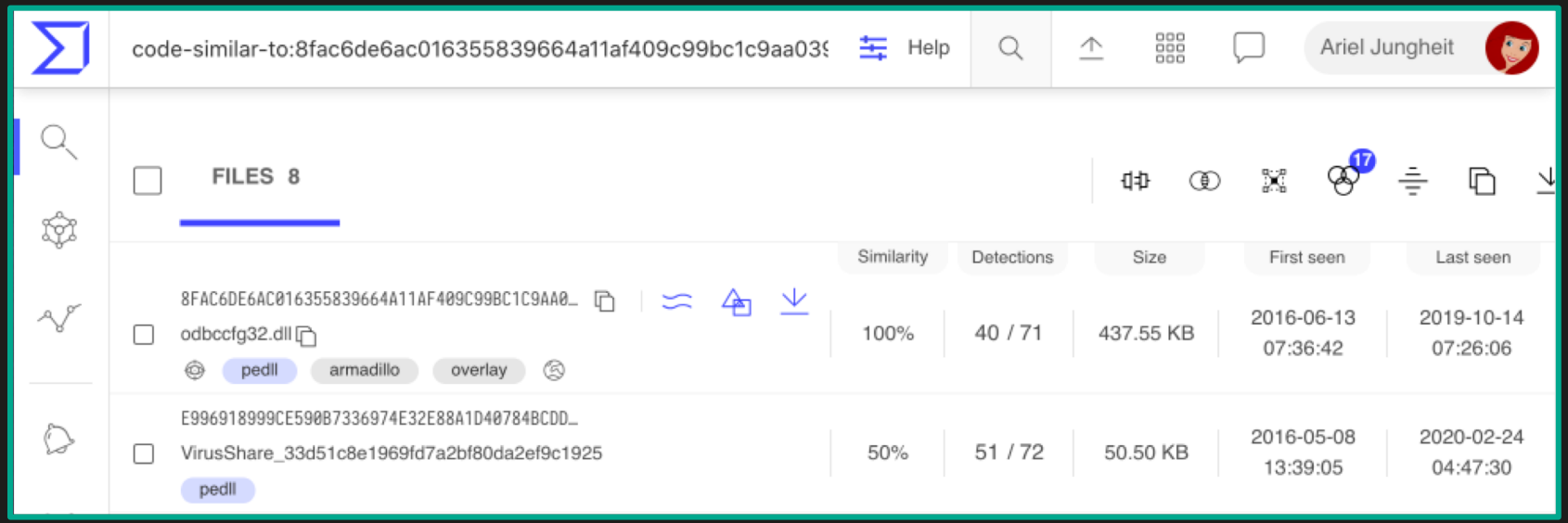
VirusTotal Beta feature – New search modifier

code-similar-to:

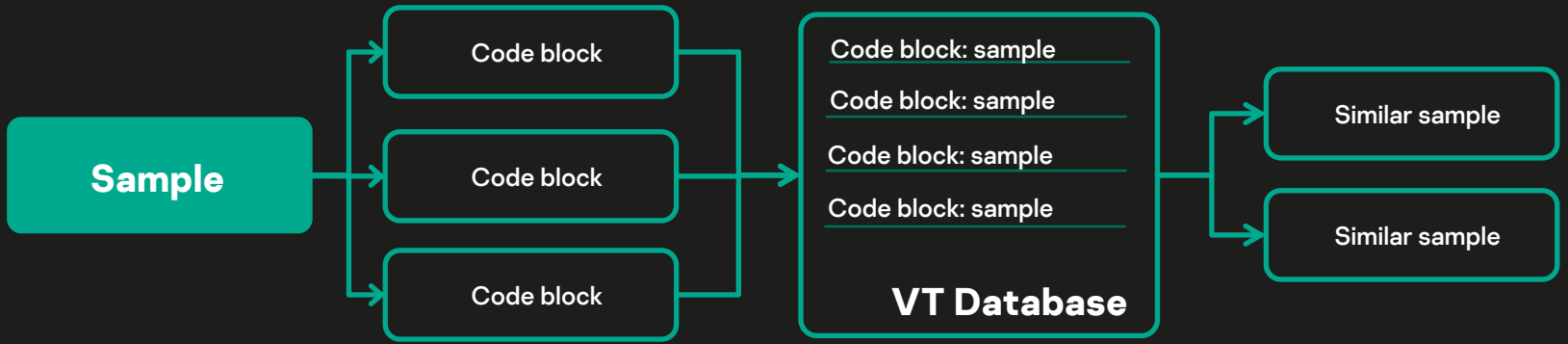
Juan Infantes Diaz

@jinfantesd

VirusTotal GUI



How it works?


Sample
Code block
Code block
Code block
Code block: sample
Code block: sample
Code block: sample
Code block: sample
VT Database
Similar sample
Similar sample

So what is this?

In our case

Turla Carbon implant WICACCESS.DLL

...
4c 8b c6 8b d3 49 8b cc ff d0 8b f8
48 8b 05 c6 98 00 48 85 c0 74 0c
33 c0 48 83 c4 28 41 5c 5f 5e 5b c3
4c 8b c6 8b d3 49 8b cc e8 92 fe ff ff 85
4c 89 44 24 18 89 54 24 10 48 89 4c 24 08
41 54 48 83 ec 28 48 8b 74 24 60 4c 8b
5c 24 58 85 db 75 14 8b c7 48 83 c4
...

12 code blocks

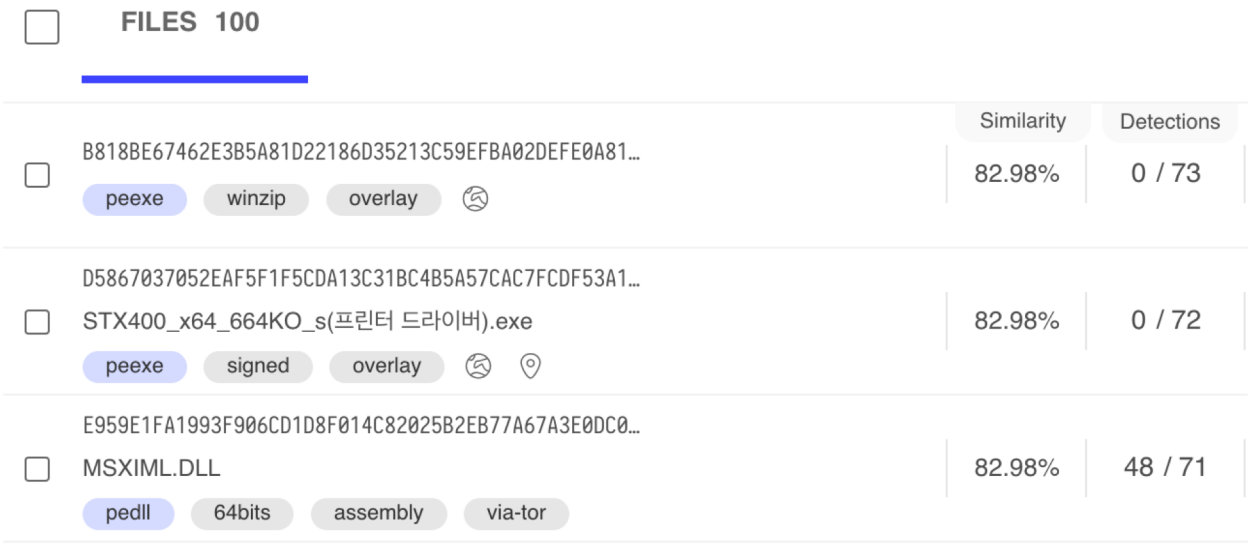
>100 Samples over 50% similarity

...

3176818e9ecfd218d096fabc6b78584a	83%
00ced150ba51846ebd53702a4b0b6c1b	83%
ecec74ff2c40038eab1ca50b38e86568	76%
e45ff81f38edd0c6d052bb5978517572	76%
ee239af759945b943bfa1bb0ce02c2f2	72%
17c57edac9ef3ee775920b4d9ecf9802	68%
98dad8feb6d666ac5b64271ffcf90fd9	65%
1fb407a20373f3970f08d3f3c086841d	60%
adf3c31e62b29961b12ddac1572b954b	58%
8eb15b499cb61cd31fada7cbb0ea4951	58%

...

That's cool



…Now what?

- Did we just get 100 samples which attribute to Turla?
- How many of them are relevant for us?
- Which should we analyze further?

Limitation and Caveats

As of now:

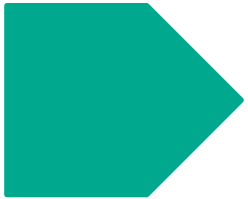
- Works with PE files
- Sample set is limited
- Packed samples are an issue
- Code blocks subsets
- ~~Can't chain with other modifiers~~
- Code blocks could be benign / no whitelist
- Bugs in code similarity calculation

Very much in BETA

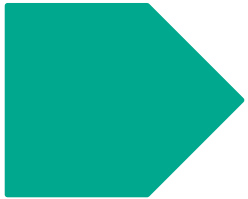
If you like it, let us know

What's our goal here?

One APT attributed sample



Collect similar samples



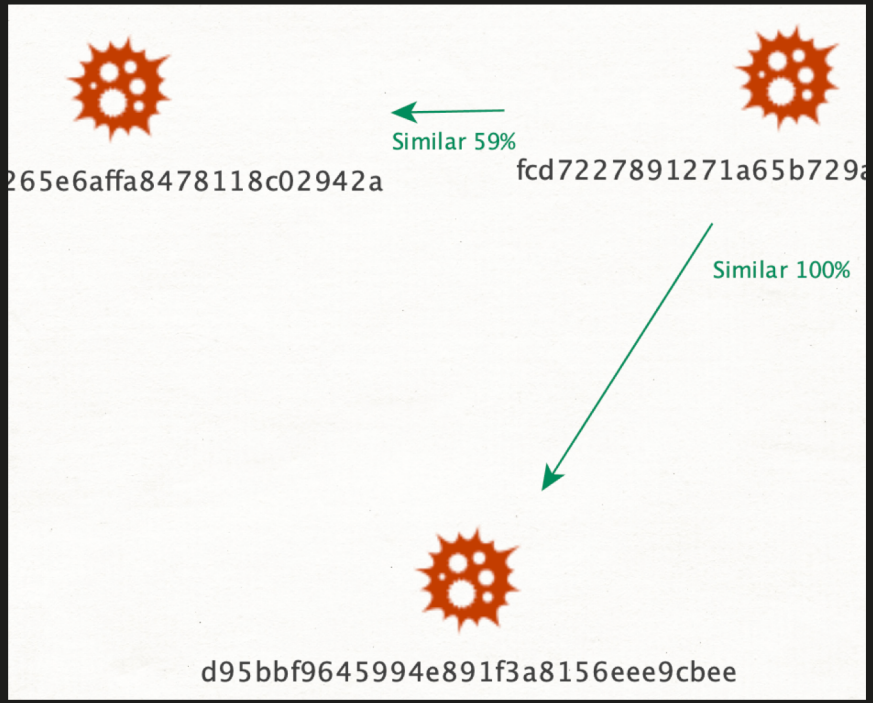
Filter samples of that APT

Maltego Transform

Yara rule generator

Maltego Transform

- Accepts hash and similarity threshold
- Returns samples over given threshold



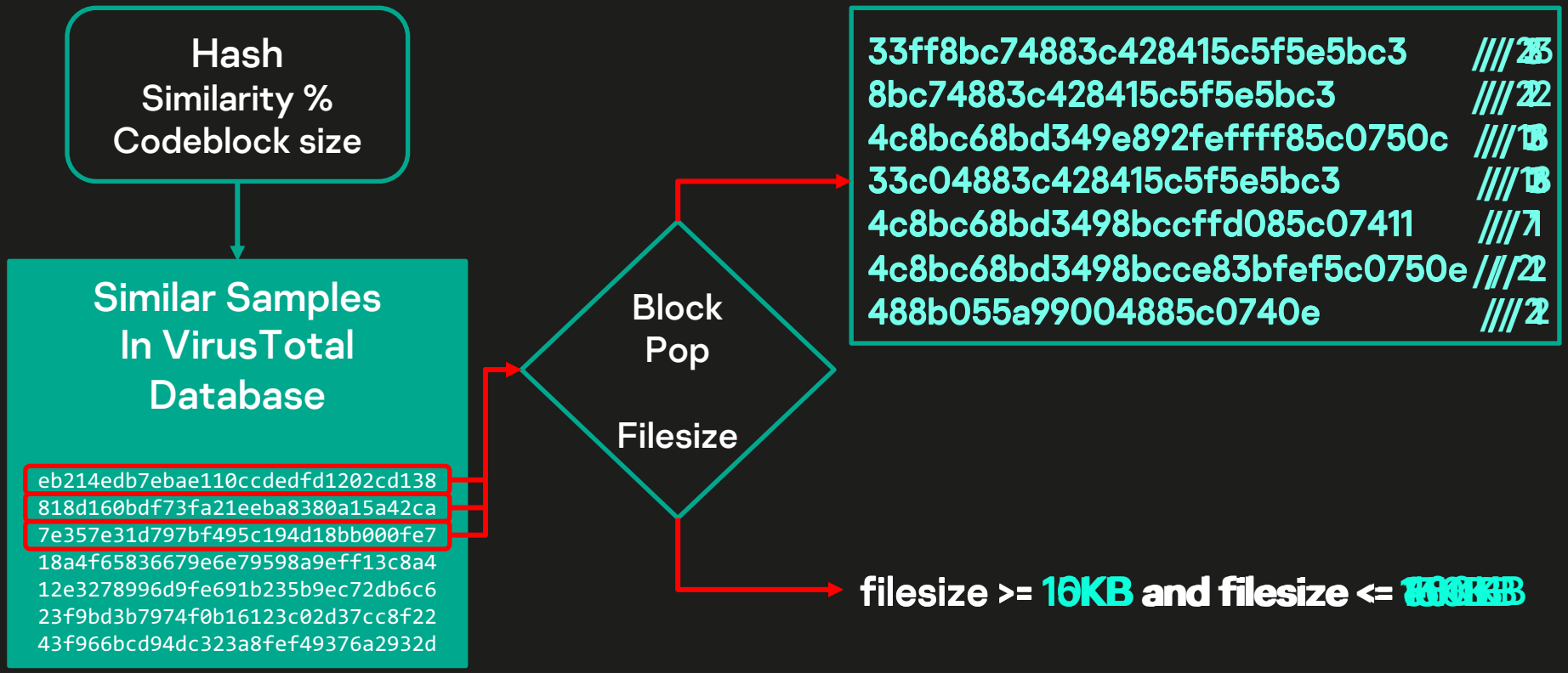
Yara generator

- Accepts hash, minimal codeblock size, similarity threshold

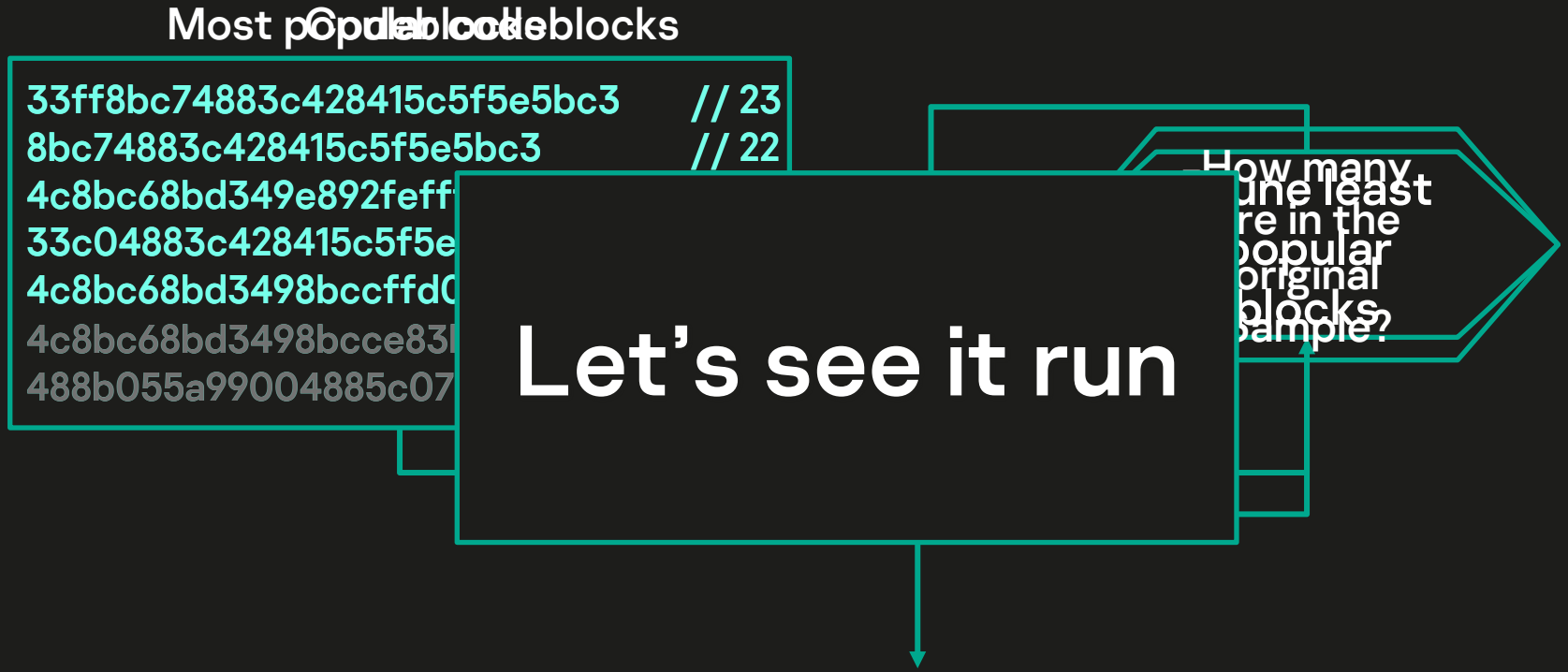
```
usage: VTSimilarity_YaraGen.py [-h] [--hash Hash] [--t 0.5]
                               [--hashlist hash_list.txt] [--min_block 4]
                               [--apikey Your VirusTotal API Key]

VTSimilarity Yara Generator

optional arguments:
  -h, --help            show this help message and exit
  --hash                MD5/SHA1/SHA256 hash to check on VTi
  --t 0.5               Minimum similarity threshold (default=0.5)
  --hashlist list.txt.  Path to a file containing list of hashes
  --min_block 4         Minimum desired codeblock size
  --apikey              VT API Key
```


Hash
Similarity %
Codeblock size
Similar Samples In VirusTotal Database
eb214edb7ebae110ccdedfd1202cd138
818d160bdf73fa21eeba8380a15a42ca
7e357e31d797bf495c194d18bb000fe7
18a4f65836679e6e79598a9eff13c8a4
12e3278996d9fe691b235b9ec72db6c6
23f9bd3b7974f0b16123c02d37cc8f22
43f966bcd94dc323a8fef49376a2932d
Block Pop
Filesize
Codeblocks
33ff8bc74883c428415c5f5e5bc3
8bc74883c428415c5f5e5bc3
4c8bc68bd349e892feffff85c0750c
33c04883c428415c5f5e5bc3
4c8bc68bd3498bccffd085c07411
4c8bc68bd3498bcce83bfef5c0750e
488b055a99004885c0740e
filesize >= 16KB and filesize <=

Tools - Yara generator



filesize >= 10KB and filesize <= 480KB and 4 of them

Tools - Yara generator


```
rule VT_Code_Similarity_7ec8a9641d7342d1a471ebcd98e28b62 {

meta:
        description = "rule to hunt for samples similar to Turla Carbon implant"
        similarity_threshold = "80%"
        minimal_codeblock_size = "4"
        similar_samples_analyzed = "93”
        hash = “7ec8a9641d7342d1a471ebcd98e28b62”

strings:
        $block1 = { 33ff8bc74883c428415c5f5e5bc3 } // Seen in 93 samples
        $block2 = { 4c8bc68bd3498bcce892feffff85c0750c } // Seen in 93 samples
        ...
        $block9 = { 4c8bc68bd3498bcce83bfeffff85c0750e } // Seen in 93 samples
        $block10 = { 488b055a99004885c0740e } // Seen in 1 samples

condition:
        (uint16(0) == 0x5A4D) and filesize >= 8KB and filesize <= 41496KB
        and 5 of them
}
```

Poll #1

Do you see value in using code similarity in hunting for specific malware?

kaspersky

But wait, there's more!

- Assembly code per code block is stored
  - Making it easier to edit the codeblock
- All offsets of code blocks are stored
  - We can use their offset
    $block1 at 4205060
  - $block2 at 5100000
    $block3 in (5107800..filesize)
  - We could order then in a specific sequence
    $block2 in (@block1..@block3)

- Other neat ideas are welcome!

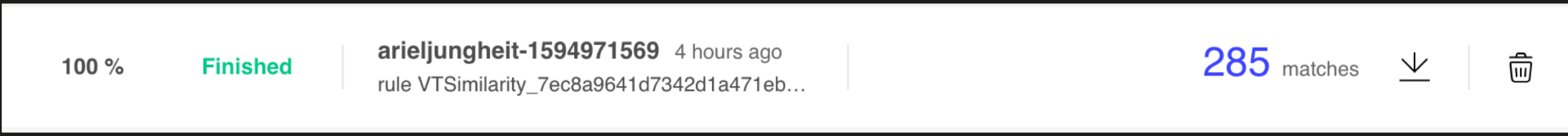
Open source

Fork it on Github

Let’s put our Yara rule to use

- Tweak it as you wish
- Hunt offline?
- Hunt online!

Retrohunt results



285 Samples in the last year

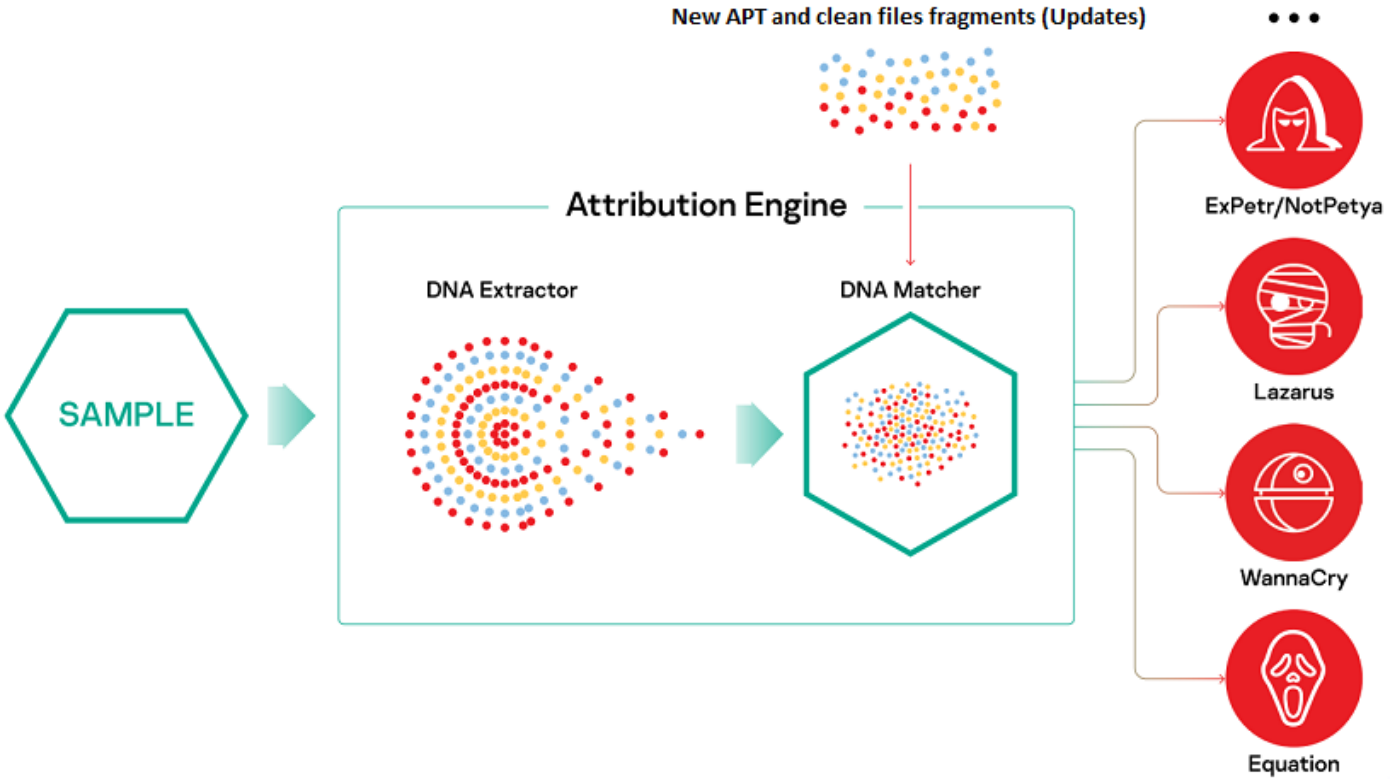
…But how many are relevant?

KTAE

Kaspersky Threat Attribution Engine

kaspersky

Few words about KTAE



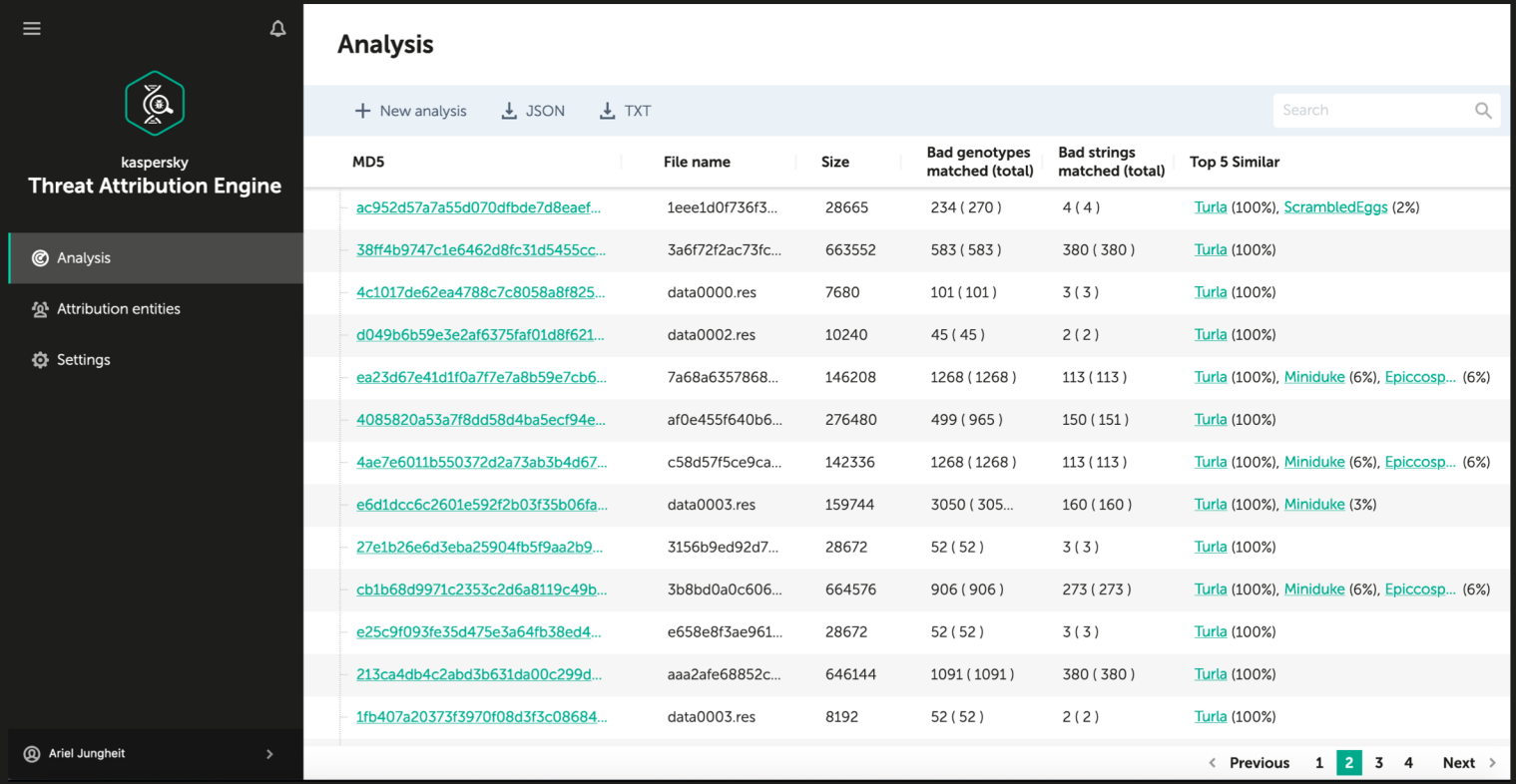
> 800

How about we

285 Matches

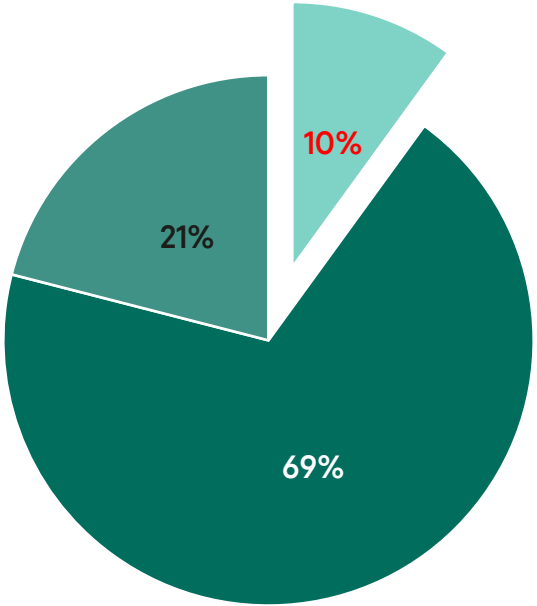
KTAE

KTAE Magic


kaspersky
Threat Attribution Engine
Analysis
Attribution entities
Settings
Ariel Jungheit
Analysis
New analysis
JSON
TXT
Search
MD5
File name
Size
Bad genotypes matched (total)
Bad strings matched (total)
Top 5 Similar
ac952d57a7a55d070dfbde7d8eaef...
1eee1d0f736f3...
28665
234 (270)
4 (4)
Turla (100%), ScrambledEggs (2%)
38ff4b9747c1e6462d8fc31d5455cc...
3a6f72f2ac73fc...
663552
583 (583)
380 (380)
Turla (100%)
4c1017de62ea4788c7c8058a8f825...
data0000.res
7680
101 (101)
3 (3)
Turla (100%)
d049b6b59e3e2af6375faf01d8f621...
data0002.res
10240
45 (45)
2 (2)
Turla (100%)
ea23d67e41d1f0a7f7e7a8b59e7cb6...
7a68a6357868...
146208
1268 (1268)
113 (113)
Turla (100%), Miniduke (6%), Epiccosp... (6%)
4085820a53a7f8dd58d4ba5ecf94e...
af0e455f640b6...
276480
499 (965)
150 (151)
Turla (100%)
4ae7e6011b550372d2a73ab3b4d67...
c58d57f5ce9ca...
142336
1268 (1268)
113 (113)
Turla (100%), Miniduke (6%), Epiccosp... (6%)
e6d1dcc6c2601e592f2b03f35b06fa...
data0003.res
159744
3050 (305...
160 (160)
Turla (100%), Miniduke (3%)
27e1b26e6d3eba25904fb5f9aa2b9...
3156b9ed92d7...
28672
52 (52)
3 (3)
Turla (100%)
cb1b68d9971c2353c2d6a8119c49b...
3b8bd0a0c606...
664576
906 (906)
273 (273)
Turla (100%), Miniduke (6%), Epiccosp... (6%)
e25c9f093fe35d475e3a64fb38ed4...
e658e8f3ae961...
28672
52 (52)
3 (3)
Turla (100%)
213ca4db4c2abd3b631da00c299d...
aaa2afe68852c...
646144
1091 (1091)
380 (380)
Turla (100%)
1fb407a20373f3970f08d3f3c08684...
data0003.res
8192
52 (52)
2 (2)
Turla (100%)
Previous
1
2
3
4
Next

285 m
on Yar
in the
year

nem
nes
at
9%



285 similar samples in last year:

- 28 samples are >99% Turla
- 195 samples are Whitelisted
- 61 are unidentified

16 Turla samples were new to us

VT Code Similarity + KTAE = Win

VirusTotal
Code Similarity
BETA

->

Yara

Hunting using
Code blocks

->

KTAE

Filtering for
APTs

Let's talk

About VirusTotal code similarity → Juan @jinfantesd
About Tools / Methodology → Me @arieljt

Ariel Jungheit
Senior Security Researcher
Kaspersky GReAT