RSAConference2020
San Francisco | February 24 – 28 | Moscone Center

HUMAN ELEMENT

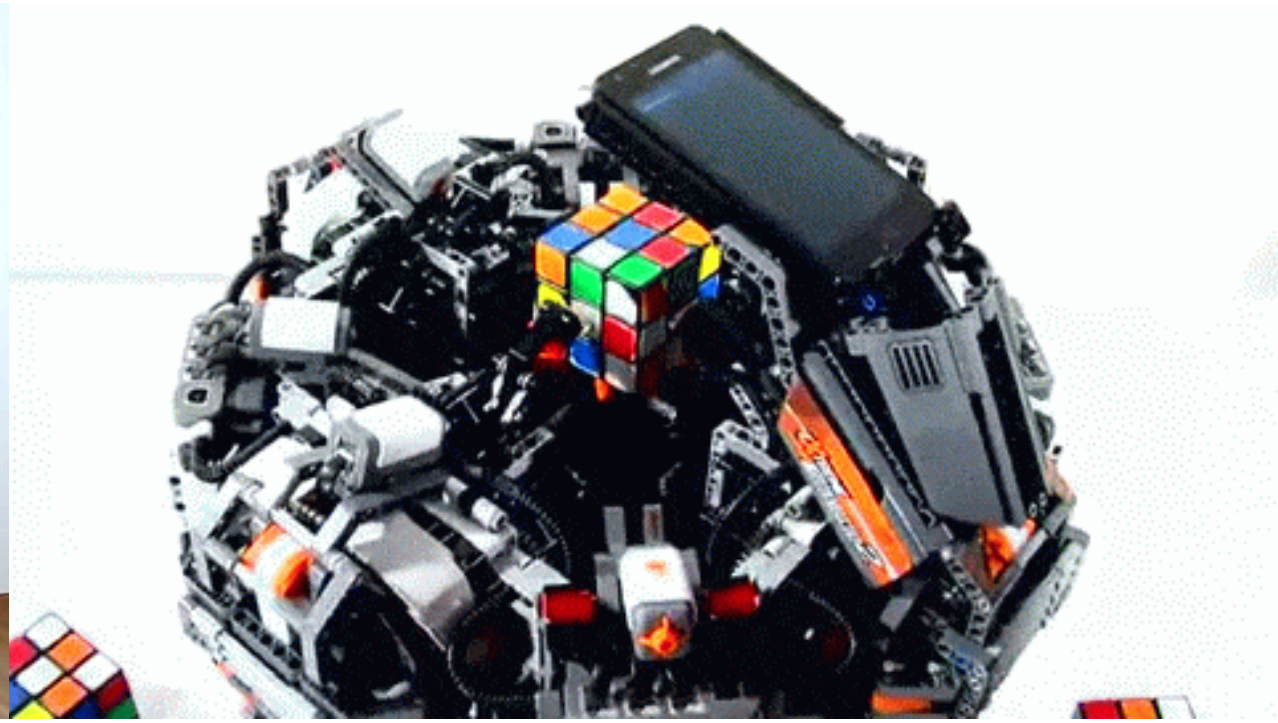
SESSION ID: HTA-W02

The Magic of Tracking Attack Campaigns using Data Science

Vicky Ray

Principal Researcher
Palo Alto Networks
@0xVK

UNIT 42



Agenda

- Why Data Science?
- Challenges in finding similarities in malware
- Similarity analysis
- Unit 42 case study
- Visualizations & Jupyter Notebooks

What is Data Science?

Data Science is a field which uses scientific methods, statistics and algorithms to discern and extract insights from various types of data sets and is a critical component of Machine Learning and AI based solutions.

Data Science to me…

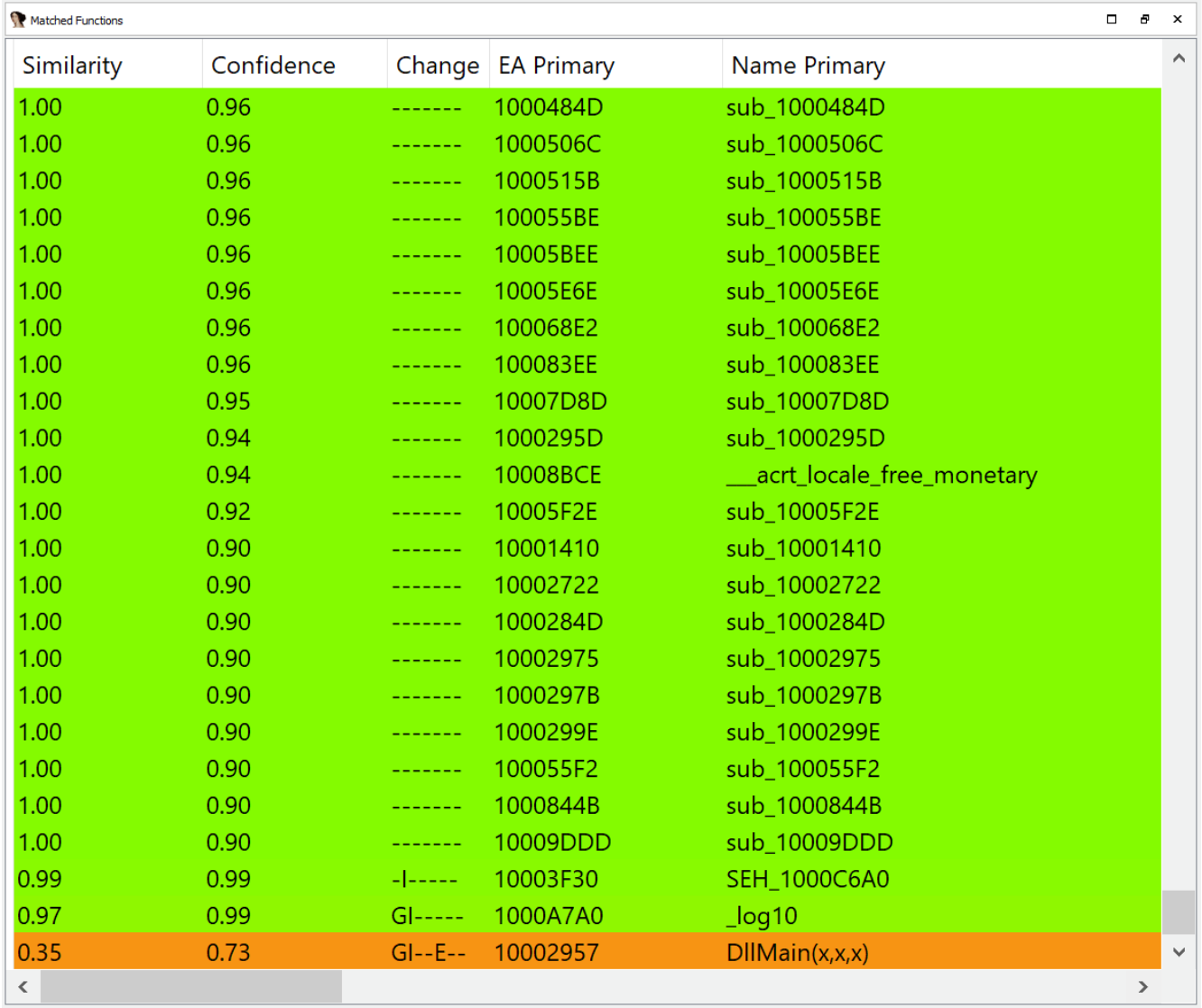
I am NOT a Data Scientist

But trying to use data science techniques to solve my own challenges

Unique attributes of two samples from same malware family

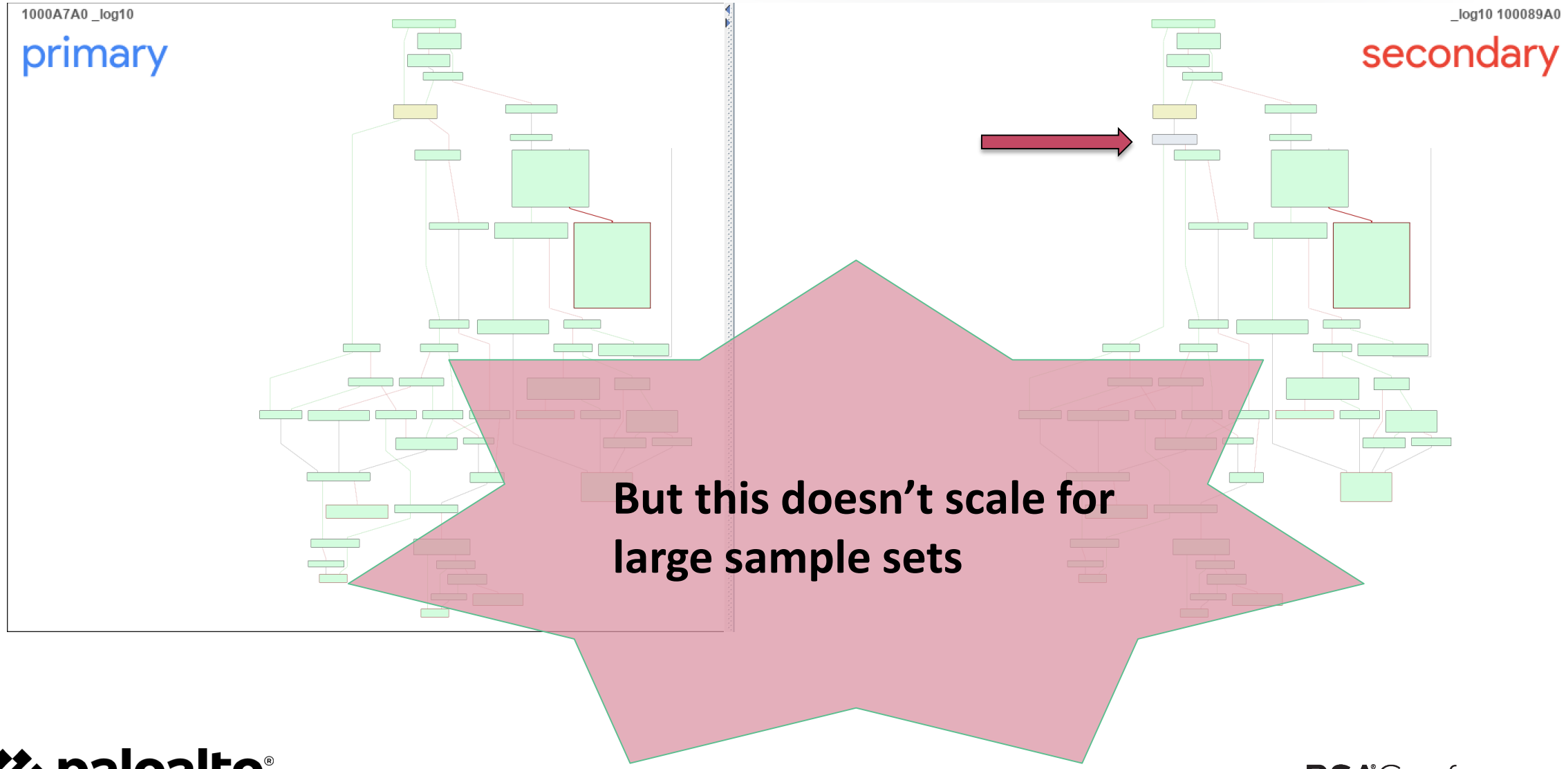
WildFire Verdict	MD5	File Size (Bytes)	File Type	Import Table Hash	Ssdeep Fuzzy Hash
Malware	35b4f04dd3df5defb3bd3040be880320	288,256	DLL	7bae0f81f57e10fd1249e9e2e448191f	3072:M9p+8WQZI1wLLr0B0wUYSYtRRox/pQJE1UxInIFtwSqWVOAg0FujMf2DuSaAY8/O:K+8W+AywWY3JE1UGYBOAOQvDgHsF
Malware	b9af036686b5745bec438d2f7c9dc3d6	401,920	DLL	8ca6613631a033689c4f5b0e3a0b7b06	6144:gm1p0nsblWJDgaQACffBrZabX3LbNE6UAscSYOYdtQ5EUpnEI75gkpGR4HtsvqKm:/jblWJfpCXBriX3cAdk5EUpn7yQsvbm

Similarity analysis on IDA using BinDiff

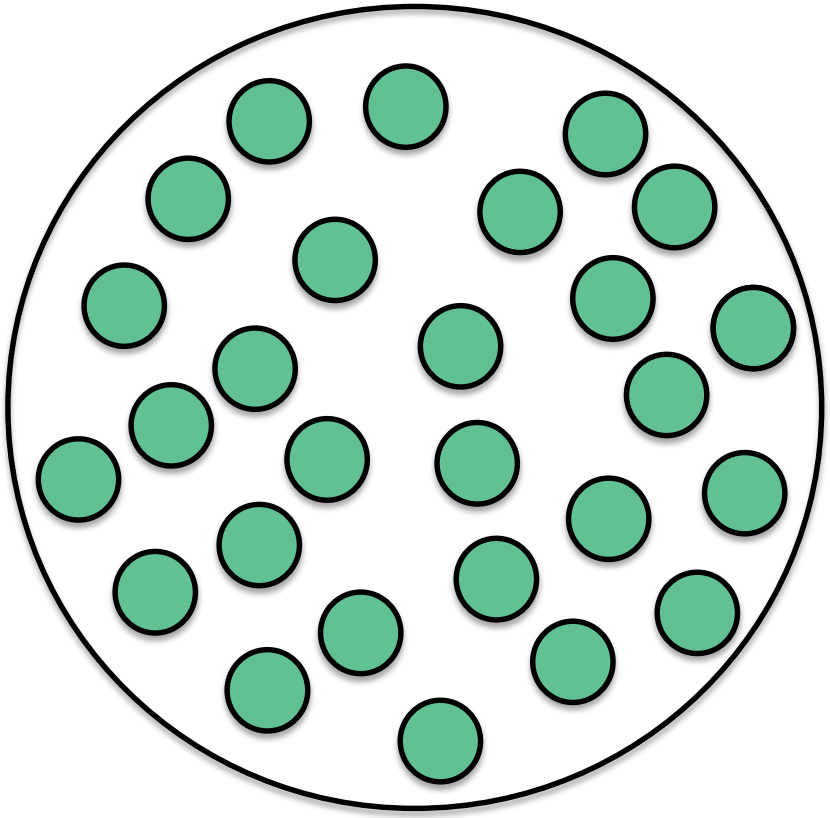


Similarity	Confidence	Change	EA Primary	Name Primary
1.00	0.96	-------	1000484D	sub_1000484D
1.00	0.96	-------	1000506C	sub_1000506C
1.00	0.96	-------	1000515B	sub_1000515B
1.00	0.96	-------	100055BE	sub_100055BE
1.00	0.96	-------	10005BEE	sub_10005BEE
1.00	0.96	-------	10005E6E	sub_10005E6E
1.00	0.96	-------	100068E2	sub_100068E2
1.00	0.96	-------	100083EE	sub_100083EE
1.00	0.95	-------	10007D8D	sub_10007D8D
1.00	0.94	-------	1000295D	sub_1000295D
1.00	0.94	-------	10008BCE	___acrt_locale_free_monetary
1.00	0.92	-------	10005F2E	sub_10005F2E
1.00	0.90	-------	10001410	sub_10001410
1.00	0.90	-------	10002722	sub_10002722
1.00	0.90	-------	1000284D	sub_1000284D
1.00	0.90	-------	10002975	sub_10002975
1.00	0.90	-------	1000297B	sub_1000297B
1.00	0.90	-------	1000299E	sub_1000299E
1.00	0.90	-------	100055F2	sub_100055F2
1.00	0.90	-------	1000844B	sub_1000844B
1.00	0.90	-------	10009DDD	sub_10009DDD
0.99	0.99	-I-----	10003F30	SEH_1000C6A0
0.97	0.99	GI-----	1000A7A0	_log10
0.35	0.73	GI--E--	10002957	DllMain(x,x,x)

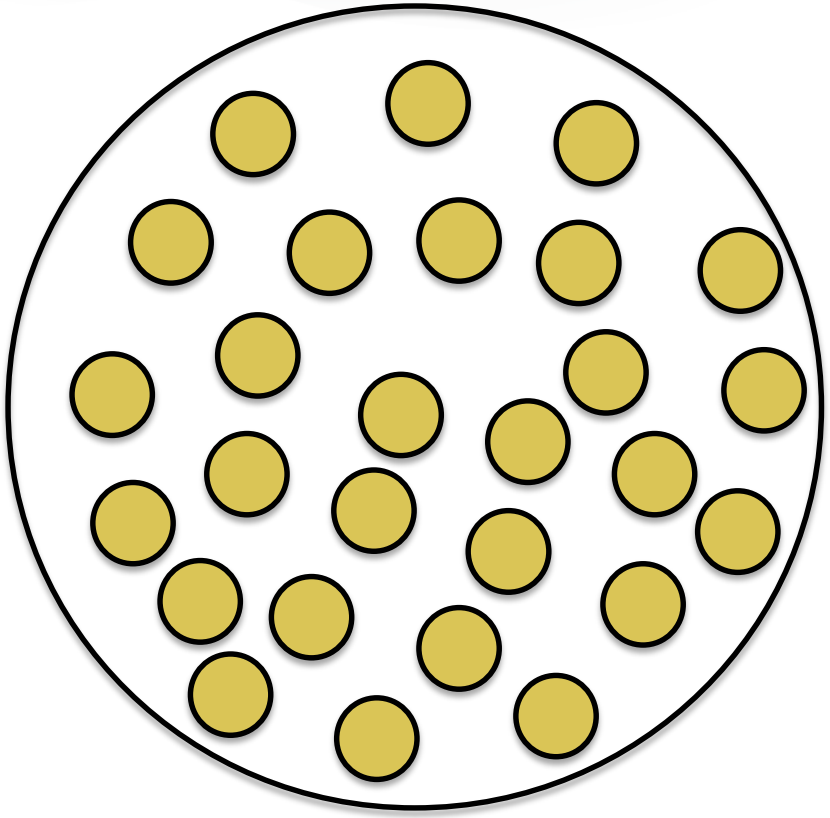
Similarity analysis on IDA using BinDiff



Feature Similarity



Sample A



Sample B

Types of features for malware

String based features

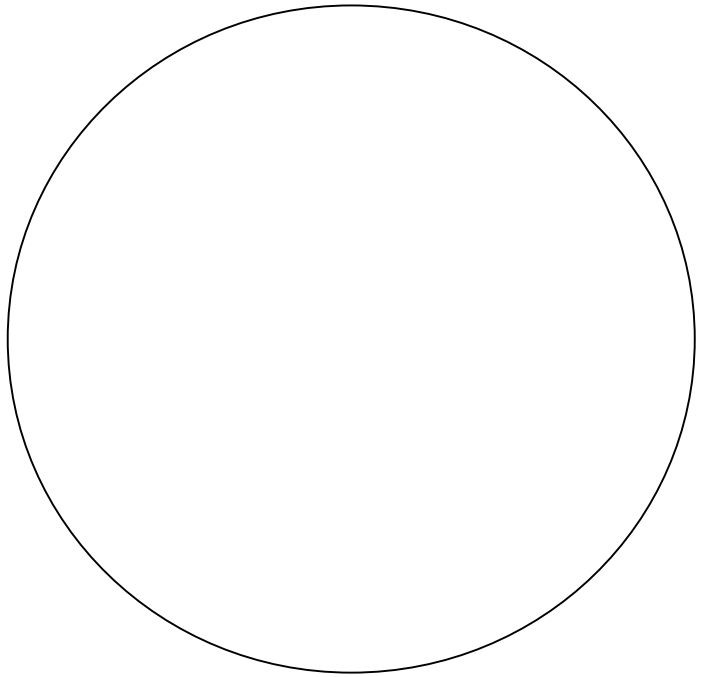
IAT

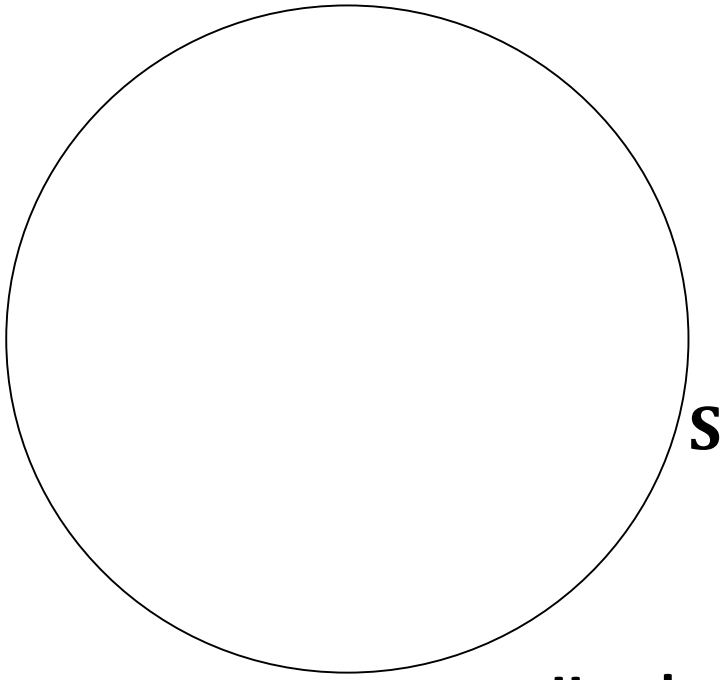
PE header

Instruction sequence

Dynamic analysis

Jaccard Index to the rescue





$$\mathbf{Sim(A, B)} = \frac{\mathbf{A \cap B}}{\mathbf{A \cup B}}$$

Set A

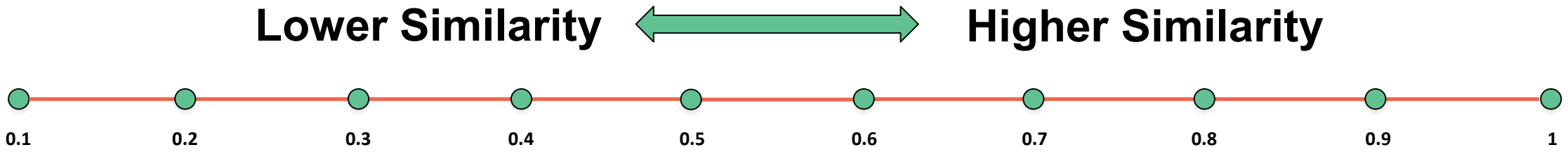
Set B

JI value range - 0 to 1

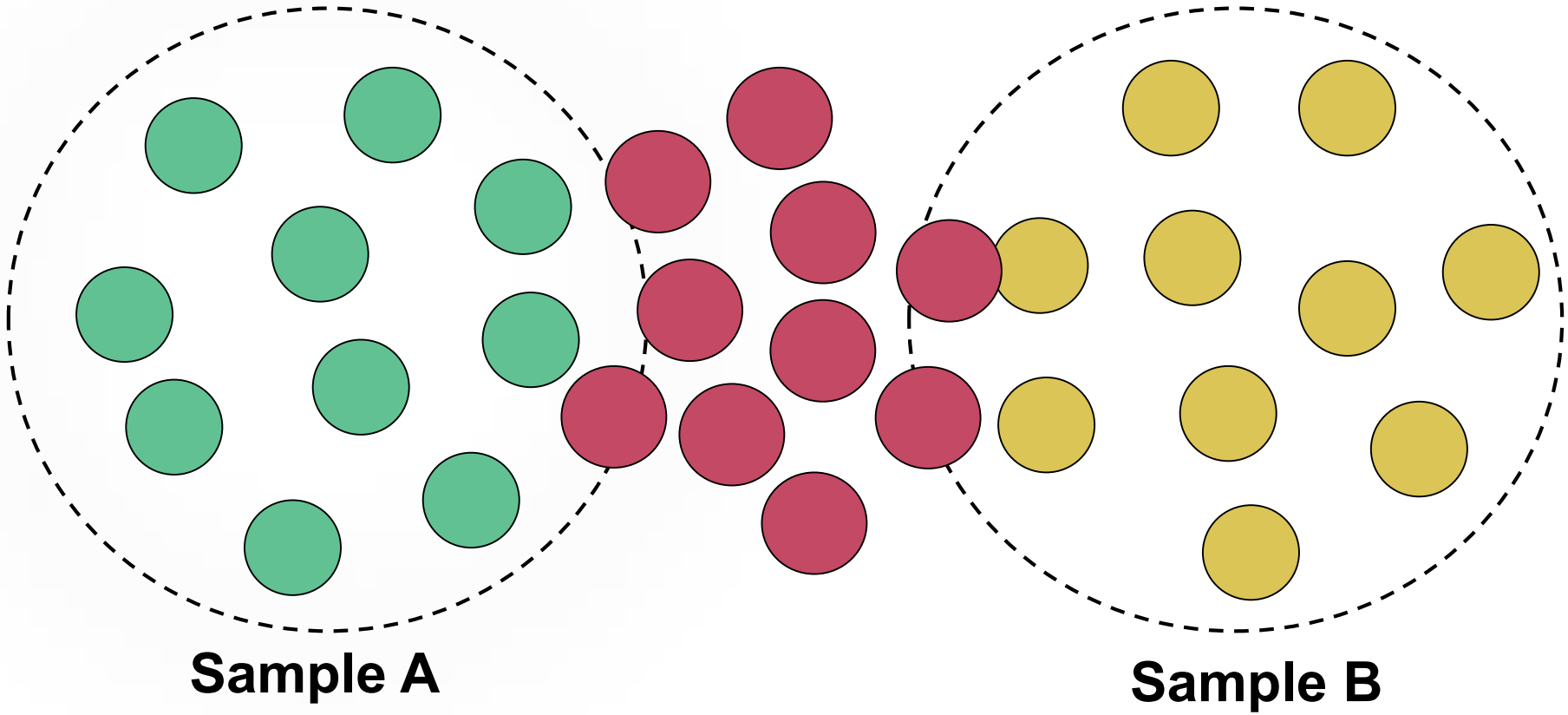
Values closer to 1 has higher similarity

Jaccard Index similarity value

$$\mathbf{Sim(A, B)} = \frac{\mathbf{A \cap B}}{\mathbf{A \cup B}} = \mathbf{0 \text{ to } 1}$$



Using Jaccard Index to extract similarity



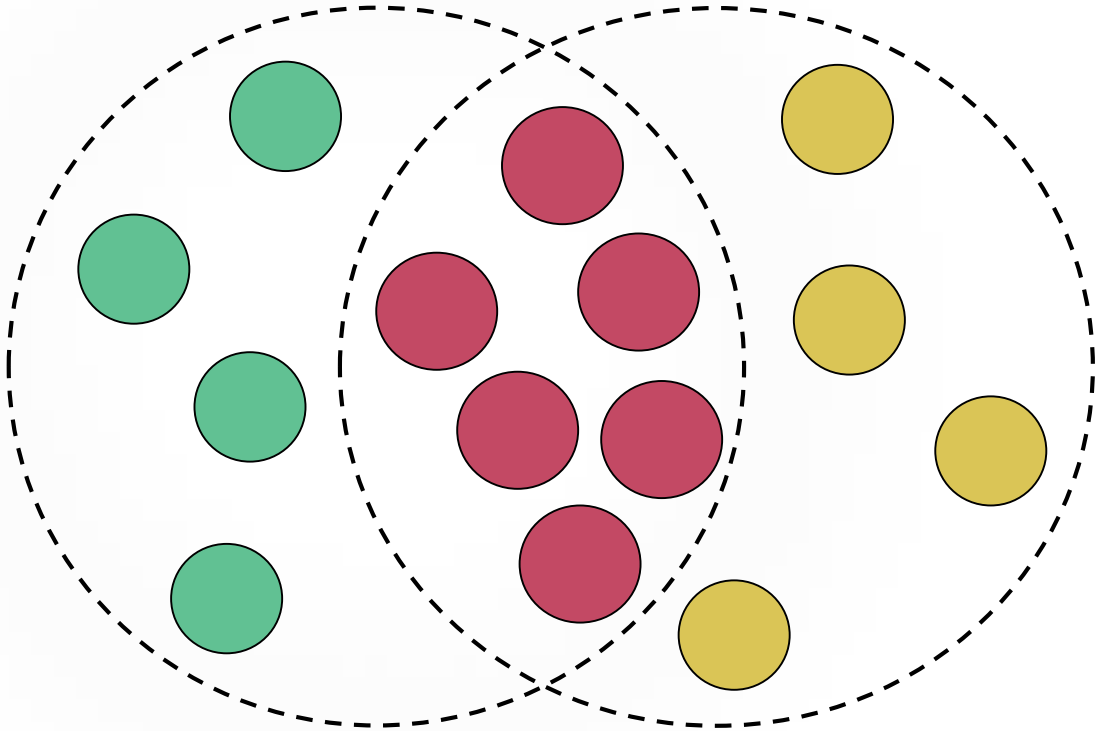
$$\mathbf{Sim(A, B)} = \frac{\mathbf{A \cap B}}{\mathbf{A \cup B}}$$

$$\mathbf{Sim(A, B)} = \frac{\mathbf{0}}{\mathbf{20}} = \mathbf{0}$$

$$\mathbf{Sim(A, B)} = \frac{\mathbf{10}}{\mathbf{10}} = \mathbf{1}$$

Total Attributes = 20
Common Attributes = 0
Jaccard Index = 0
Total Attributes = 10
Common Attributes = 10
Jaccard Index = 1

Using Jaccard Index to extract similarity



$$\mathbf{Sim(A, B)} = \frac{\mathbf{A \cap B}}{\mathbf{A \cup B}}$$

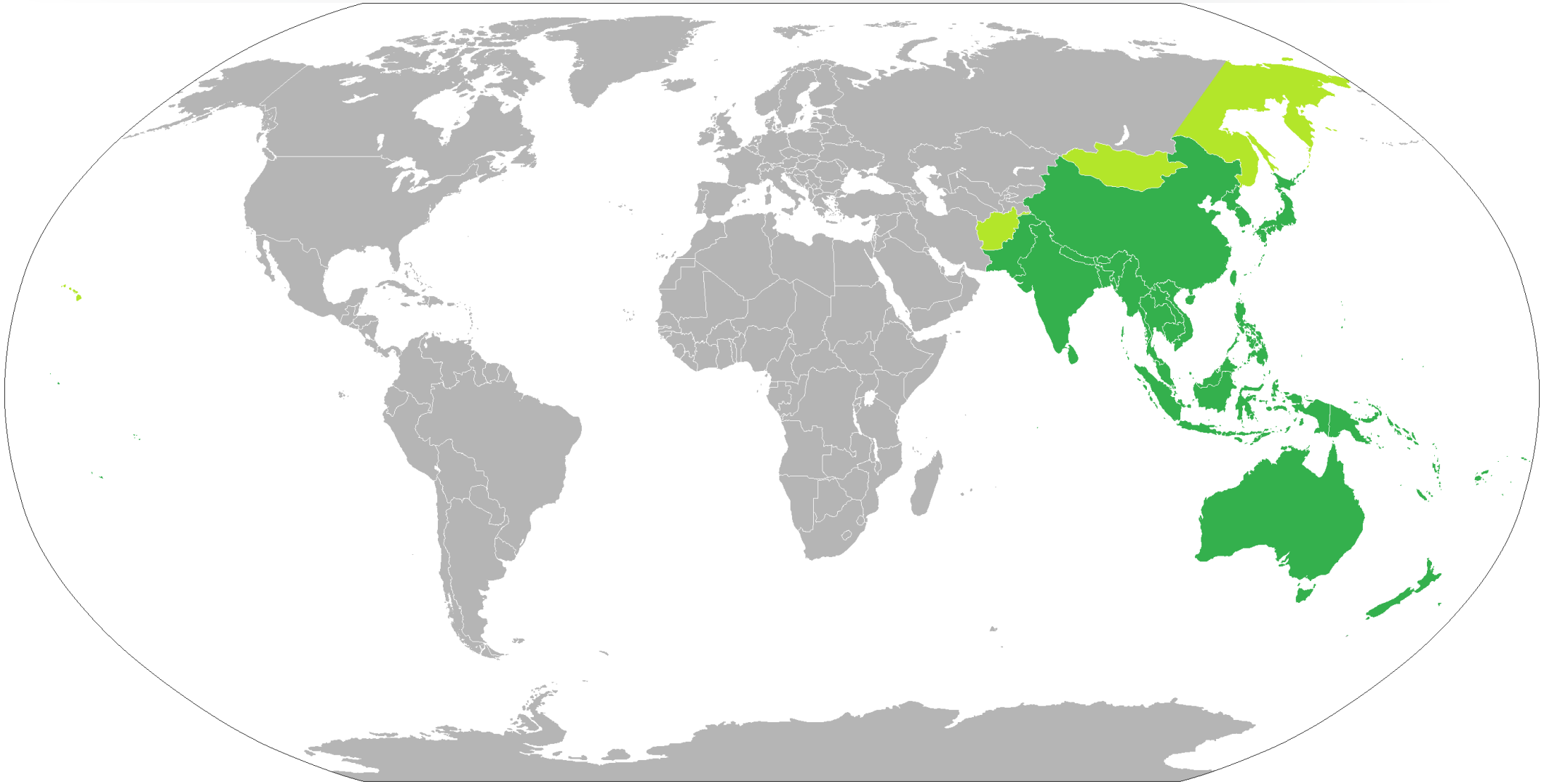
$$\mathbf{Sim(A, B)} = \frac{6}{14} = \mathbf{0.42}$$

Total Attributes = 14
Common Attributes = 6
Jaccard Index = 0.42

Case study

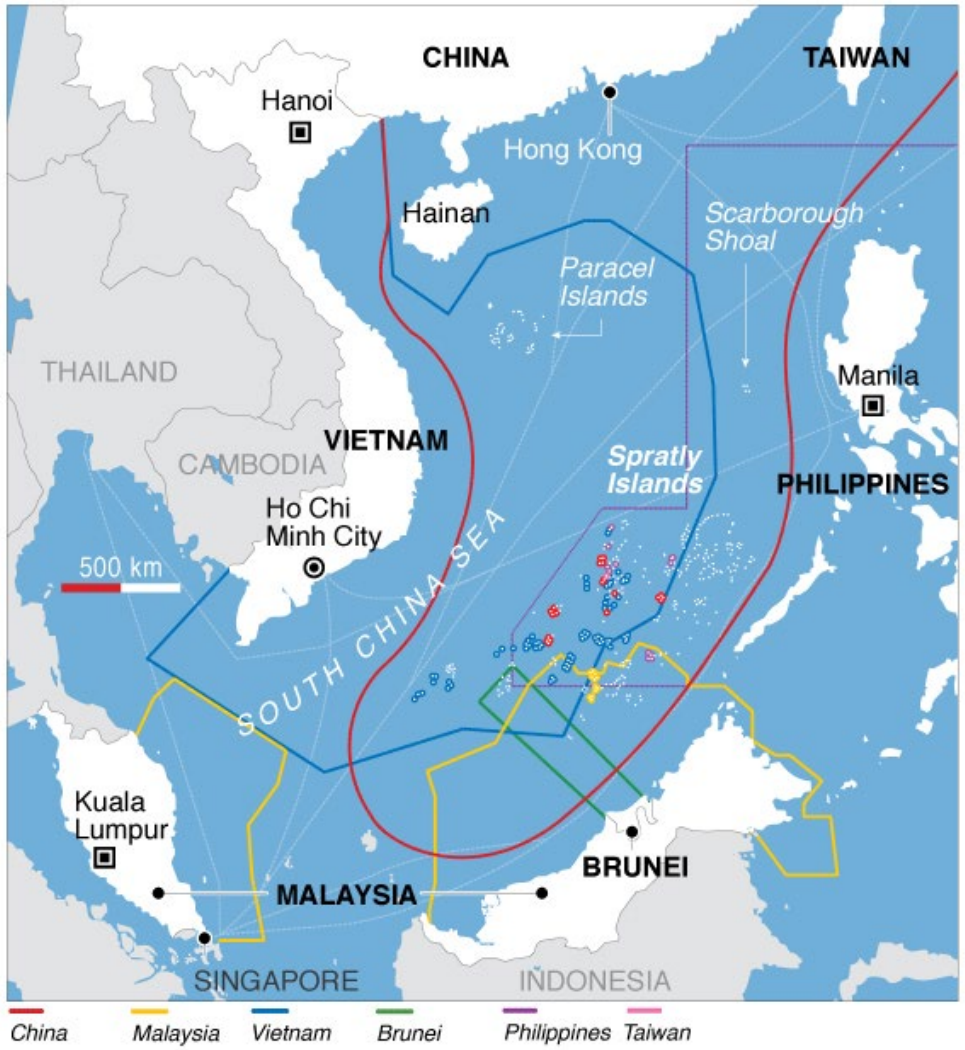
Implementing Jaccard-Index in real world case study and threat hunting.

APAC – A hot bed of APT activity



https://en.wikipedia.org/wiki/Asia-Pacific

APAC – A hotbed of APT activity



https://en.wikipedia.org/wiki/Territorial_disputes_in_the_South_China_Sea

OceanLotus

unit42

LATEST RESEARCH TOOLS PLAYBOOKS ABOUT US

SUBSCRIBE

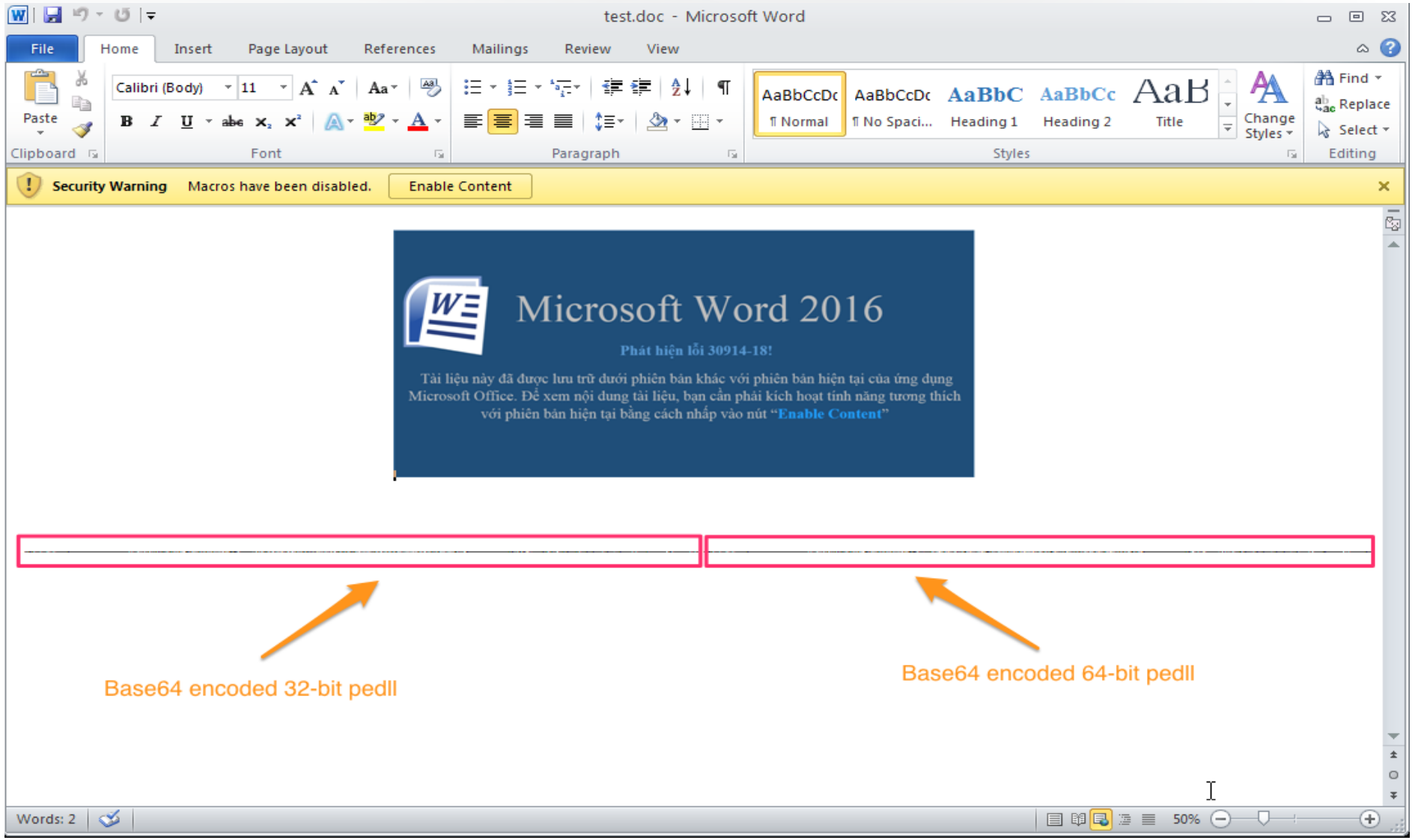
paloalto NETWORKS

UNIT 42 / TRACKING OCEANLOTUS' NEW DOWNLOADER, KERRDOWN

Tracking OceanLotus' new Downloader, KerrDown

By Vicky Ray and Kaoru Hayashi
February 1, 2019 at 6:00 AM
Category: Unit 42
Tags: KerrDown, OceanLotus

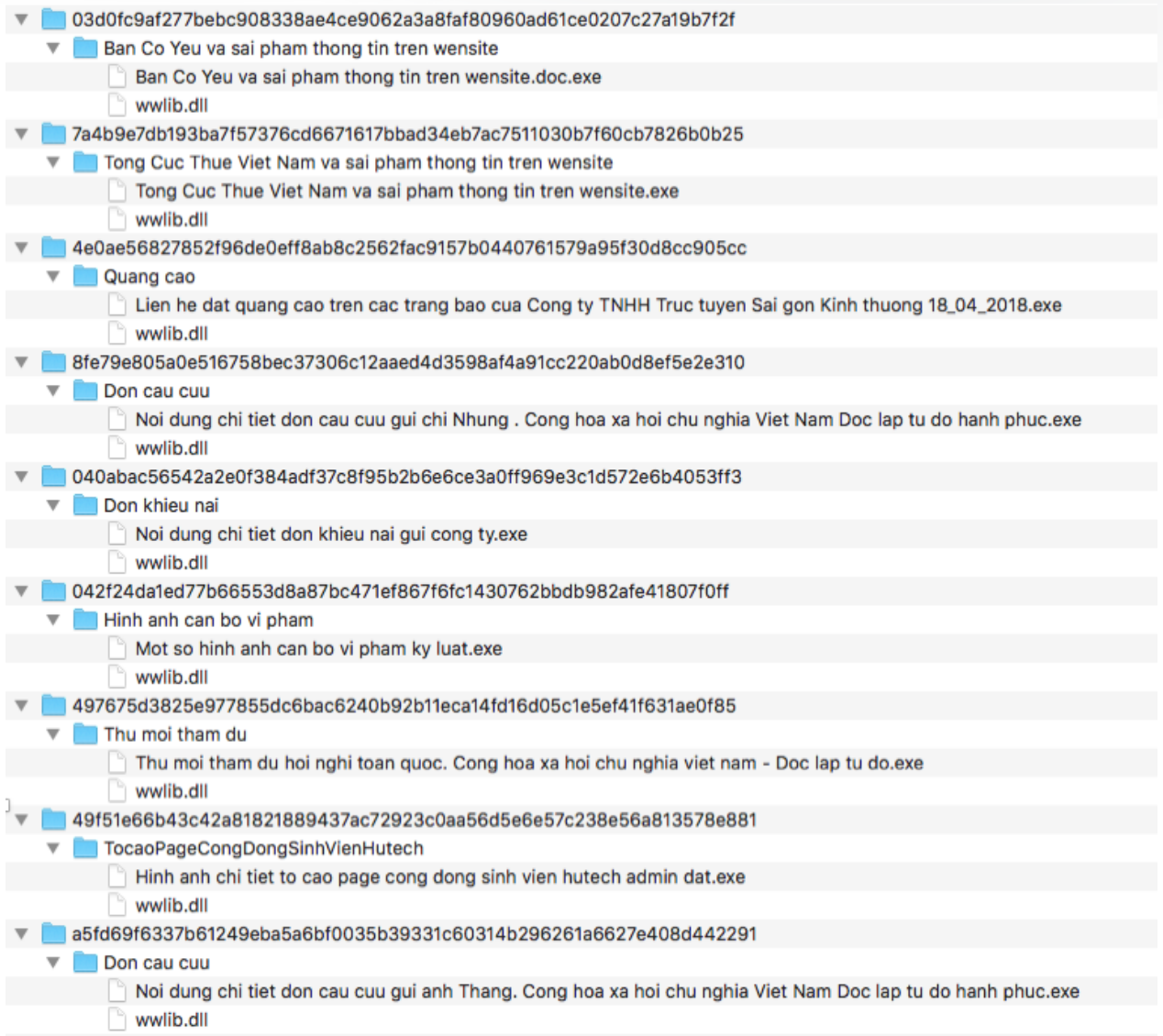
OceanLotus phishing doc file



Embedded DLL file


```
45      Dim b As String
46      Dim a As String
47      Dim tableNew As Table
48      Set tableNew = ActiveDocument.Tables(1)
49      If (iCheck = True) Then
50          a = tableNew.Cell(1, 1).Range.Text
51          a = Left(a, Len(a) - 2)
52          b = Base64Decode(a)
53      Else
54          a = tableNew.Cell(1, 2).Range.Text
55          a = Left(a, Len(a) - 2)
56          b = Base64Decode(a)
57      End If
```

RAR archives used by OceanLotus



Samples in Autofocus

FIRST SEEN	WILDFIRE VERDICT	MD5	FILE SIZE (BYTES)	FILE TYPE	IMPORT TABLE HASH	SSDEEP FUZZY HASH
02/04/2020 10:00:44pm	Malware	67d20d1baa7c38a0ecf9ec81b9bb29bd	92,672	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:npPPtkXxNTVcy1dHQWLETMM+lk7QOsWxcdQWNaumcdG0cRFeFAjg+mM:npKnVcadH5Y4k74QgmzuWjDmM
02/04/2020 10:00:35pm	Malware	58002c897188a16493d2832acf06e3e5	92,704	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:npPPtkXxNTVcy1dHQWLETMM+lk7QOsWxcdQWNaumcdG0cRFeFAjg+mV:npKnVcadH5Y4k74QgmzuWjDmV
01/26/2020 4:11:07am	Malware	b246b492bba57a6ed6c2e23a3c3c4243	92,672	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:7GPhkY673nHNaAtxWLELs3+p7QOsWxcdQwgcdGLLMcVZrFAjg+md:7YMnHg0AYJ74QwgecVZrWjDmd
12/18/2019 4:51:43pm	Malware	10eed552e0306a171482cd1f6ad26633	131,072	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:5EXVkreLLvHQi0vWLE/pO+o7QOsWxcdQXkhGFOSNdpc7OojshK2hjcJkIz2WM9Sn:5RcvHz0OYu74QXkA+qTN2pA/eJ51
12/18/2019 2:04:54am	Malware	4d596c44da9e3816c8d658ebb1bb4138	339,968	DLL	8ca6613631a033689c4f5b0e3a0b7b06	6144:EvHzsK+qTsAyuHgsLgDnQow/ITRj7HpvYKdnin2F4O/ltETBi:X3qTsAxAsLgTQow3jDtZAfO88
11/04/2019 12:30:28am	Malware	6e92cfd7b5fed8849392430221e7be8f	141,312	DLL	393639a60c6cd8a32838f37476501ec3	1536:g0EIHZhkJAR5IUlg7mheyVXKfIluXnYgCsWLPcdgQj6DN+Eqq0525dQrBEebOfnk:g0EI36UlmmhTx4XntUegmlM6BvyNar
11/03/2019 2:00:48pm	Malware	04e6321801672421 42f3f3de63b2079f	93,184	DLL	f18b0a8717c1b7e198c122affcaee2ab	1536:GY0XksLA41saJyYKw++BiQpsW8cdQDpdG+LM5veueki:GLOakTOiWQDe4b7
11/03/2019 2:00:36pm	Malware	efcb1907df4cb72f5a0b1379de2a637c	92,672	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:Op/5khi9z/Td6ivlAoWLEzE0+u7QOsWxcdQTDsdGrOukFAjg+md:OSi/TsgPBYf74QTDsukWjDmd
10/23/2019 11:13:54pm	Malware	8adfd63de516fcb142ea443fd5ab3b95	613,376	DLL	a22e95768170b57babebc0ace44b436a	12288:AQqT7V96B2Dxhj/P/tMxfVka+E/pbBfOlNCCIketfnvXiKQUaMN:mT7VMohj/P/tMxfV7J/pONlkufvXi/U
10/23/2019 10:31:08pm	Malware	ad43d67ed35472d4d6541d9c555f05db	1,008,128	DLL	a22e95768170b57babebc0ace44b436a	24576:OyZ7FBpcehIpkRDs3IB2nZsrexfX7Dir:FtSku3bere5
10/11/2019 6:59:06am	Malware	d5de9c7c03a1fbda6ac7de636cd52817	92,672	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:OM/3v01kJBObQutBBKz4WAN5ZQEV7QOsWxcdw0MdG5cEo17ig+mM:/3F83qzxA574w04lmDmM
10/11/2019 1:01:46am	Malware	be726f75c41ea39677fd2f2b14b97bbf	93,184	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:M7Kuokh5i3wGnyKg1WANGOQlV7QOsWxcdwhEdGCblQp8big+mD:UP0rPgMAc74whJiGDmD
10/11/2019 1:01:43am	Malware	9a15d92acb3e217457fa425a287afb10	93,184	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:B7KuokN5i3wGnyKg1WANGOQlV7QOsWxcdwwEdG6/57ig+mb:hL0rPgMAc74wwlmDmb
10/10/2019 10:34:17pm	Malware	7ff4b219545d54bba39afa72bf14ee1f	1,972,224	DLL	8ca6613631a033689c4f5b0e3a0b7b06	24576:H3HUKEmS9tbbYgfucMxaZCm/11nTXUWEZvj56cUc1VVLGxWexGjn1CCqnkySuZJy:J8bYYf/wvj5vQAekyKT9R/
10/10/2019 10:32:59pm	Malware	85d3dd22725e1d924e4ec347c43d2fb2	2,189,824	DLL	8ca6613631a033689c4f5b0e3a0b7b06	49152:rEXzROBe35YypVzbFufzt5FvY4FAKctavqQQ:
10/10/2019 10:00:50pm	Malware	5a9a6565aa340a9b9eae3c7d18f3d66a	1,517,568	DLL	8ca6613631a033689c4f5b0e3a0b7b06	24576:meXnVEYExRdAN+eqnJt1fYdWmYQi3f7TUcy+L1ieGYLIhlhUZlinhb:12Dv2etWz874hUTQd
09/25/2019 10:21:05pm	Malware	895d57720a7b182c13b6bb2786dfd12f	84,480	DLL	3740e2a27304e9151a5cbb50092b2ab7	1536:4SyD/4hyklRy161UFoEFqFDi51dYRQ4SSsW+acdyHRvvB4U:4xD/A1JEo9QYRVSl7yxvvBl
09/16/2019 8:41:12pm	Malware	2945adb7619f001ec53bc02dc8c6206a	82,432	DLL	3740e2a27304e9151a5cbb50092b2ab7	1536:PgLLWwjI7kAeoglN7Xw61jcq2q1Q4hxsWXGcd/abV7Rn5EoD:IXWLNttD1hj/abVRGG
09/11/2019 1:27:41am	Malware	b2ec1cd946491a82a4b6a56df5e3b304	5,327,360	DLL	3740e2a27304e9151a5cbb50092b2ab7	49152:H1PWqx3h2hGUE0EhK0NxuxUw2XhGE27GP8cu:HH8sZ5h5UwXhG5g2
09/11/2019 1:24:59am	Malware	dc212fb3bf0c87804440636668b837f2	84,480	DLL	3740e2a27304e9151a5cbb50092b2ab7	1536:4SyD/4hyklRy161UFoEFqFDi51dYRQ4SSsW+acdyHRvvB4U:4xD/A1JEo9QYRVSl7yxvvBl
09/06/2019 12:31:57am	Malware	53df13cac99ace723ea55cefb42d8510	2,185,216	DLL	3740e2a27304e9151a5cbb50092b2ab7	49152:GXkLxq5BXXNkL+sq5Xgq5Eq5HkLeWq5hXLxRTO:Kk+BXXNkKL/Hk2hX
09/05/2019 11:31:39pm	Malware	2a183f9124ec5b223f5386c3f05065db	2,168,832	DLL	8ca6613631a033689c4f5b0e3a0b7b06	49152:ULSDm9ShDn53GJ3yuXnfyFjAEB4h+i+iIZyV8LfC:
09/03/2019 11:11:55am	Malware	e585134f619bf1a0f686c859d4ba99b3	762,880	PE	8ca6613631a033689c4f5b0e3a0b7b06	12288:vM732tz7fyC7TA+fBV+eoMNa9JvV2/KiGJmjA5lkyNaOaCq0ACGF9m:E732tzzyAT/JY3vT9IKRJmqV8faWpG
09/02/2019 3:21:22pm	Malware	69e3c413c5b6ab8bd7ea5bdf8c047474	93,184	DLL	f18b0a8717c1b7e198c122affcaee2ab	1536:Rm0Cak8rgg1cCeZyYKgO+hiQpsW8cdQipdGCLM2ueki:RuW+CeUTuiWQi22b7
08/23/2019 1:35:02am	Malware	83143e9fd120ac5d62a9f34c97518b49	551,936	DLL	1c8d5c82ae8a6b3be790d2fcc12d9597	12288:gtBekr6+BCeZgI1DGlAf1WceaRu6m5NiIvln3uc6/g0gmOBMo9:VkrZMeA1Suf1jj9ye40gmEN9
08/22/2019 2:03:43pm	Malware	8ac2841fbb960a36739a958783ad1694	81,920	DLL	3740e2a27304e9151a5cbb50092b2ab7	1536:E4yzvOtdW10kHxQIPNDXw61jcR2qpQojasWWUTcdWlGgUsox:TyDOtmFNtOFjuiWg7sE
08/21/2019 6:36:41am	Malware	741e4cf71ef5a172dce34fbf4a27372f	220,672	DLL	1c8d5c82ae8a6b3be790d2fcc12d9597	3072:LsJaBtoB1tvlXgFShvlkt1DneEyHnpZlnHSm6acMhI:LsYBto3tVg+dkfDLyNHSEcy
08/21/2019 6:13:51am	Malware	9eb55481a0b5fcd255c8fb8de1042f88	72,192	DLL	1c8d5c82ae8a6b3be790d2fcc12d9597	1536:rskNuyWEjJrBNbIXy784J15SrptZLL3lAOFyKLuRe:rsJaBtoB15UN3yyycuRe
08/15/2019 5:31:06pm	Malware	6875f307d95790ca25c1da542ea736a8	82,432	DLL	3740e2a27304e9151a5cbb50092b2ab7	1536:PgLLWwjI7kAeoglN7Xw61jcq2q1Q4hxsWXGcd/abV7Rn5EoD:IXWLNttD1hj/abVRGG
08/08/2019 6:31:19am	Malware	7460ad4891556da13e0d2aead6abc4c2	92,672	DLL	8ca6613631a033689c4f5b0e3a0b7b06	1536:qVs0wmkiKbR0P3VWpWUVyK8+H7QOsWxcdQfwHdGc1fxlil67g+m:qVm4i0PVWoAI74QYbtI0Dm
08/07/2019 2:23:13pm	Malware	8d42d9fd3a4d32bc0474d07052ce8984	81,920	DLL	3740e2a27304e9151a5cbb50092b2ab7	1536:tvw5Z2spVktxkKRNDXw61jXQ2qlmQojasWWUTcd7/h3ZWFHox:K5ZmFNtlmFjui7/zWFHE
08/06/2019 10:36:39pm	Malware	d6b1b505a33c4db63b176bd7a2bc0a71	3,437,056	DLL	3740e2a27304e9151a5cbb50092b2ab7	98304:5ZV259PaNe4/wsr6kcXv69PvQP1vF+T52fSR:5ZQCcHsJcQvQaT0u
08/02/2019 3:00:41am	Malware	b4f02f50d5b25eec622c6f53fe278f06	721,408	DLL	8ca6613631a033689c4f5b0e3a0b7b06	12288:QlLOYrplrKv/ijE6n9EvlOQ9dhSQklQKaRl+3hxFxSUEmbClOmim:QlL9kijE6adOCOxaRc3hxF5EI
07/25/2019 7:37:40pm	Malware	38f63ed1a604f683a15522efa617f2ca	5,327,872	DLL	3740e2a27304e9151a5cbb50092b2ab7	49152:PaE+7+aEcO6y688oEDEytZ7REBE++EoodE+aHCwam7E:rSQy78gC
07/23/2019 2:31:21am	Malware	1312b06ee44a26d7bf45647a33a365d8	367,104	DLL	8ca6613631a033689c4f5b0e3a0b7b06	6144:tE4tk+qTsAyuHgsLgDnQow/ITRj7HpvYKdnin2F4O/ltE9Brnm:BhqTsAxAsLgTQow3jDtZAfO8jm
07/22/2019 10:05:13pm	Malware	8bcae5a118b5ccfb9e740899b351a10c	373,248	DLL	3740e2a27304e9151a5cbb50092b2ab7	6144:4VpVDvVKFpCaqO2zWPd0Cg3gXVBg9ta9uwu1vJK6rtr572Vs+0nuESB:Kputczcd0bwzg9Uo9I6prOd0uEI

Steps taken to apply Jaccard-Index algorithm

Extract all suspected samples.

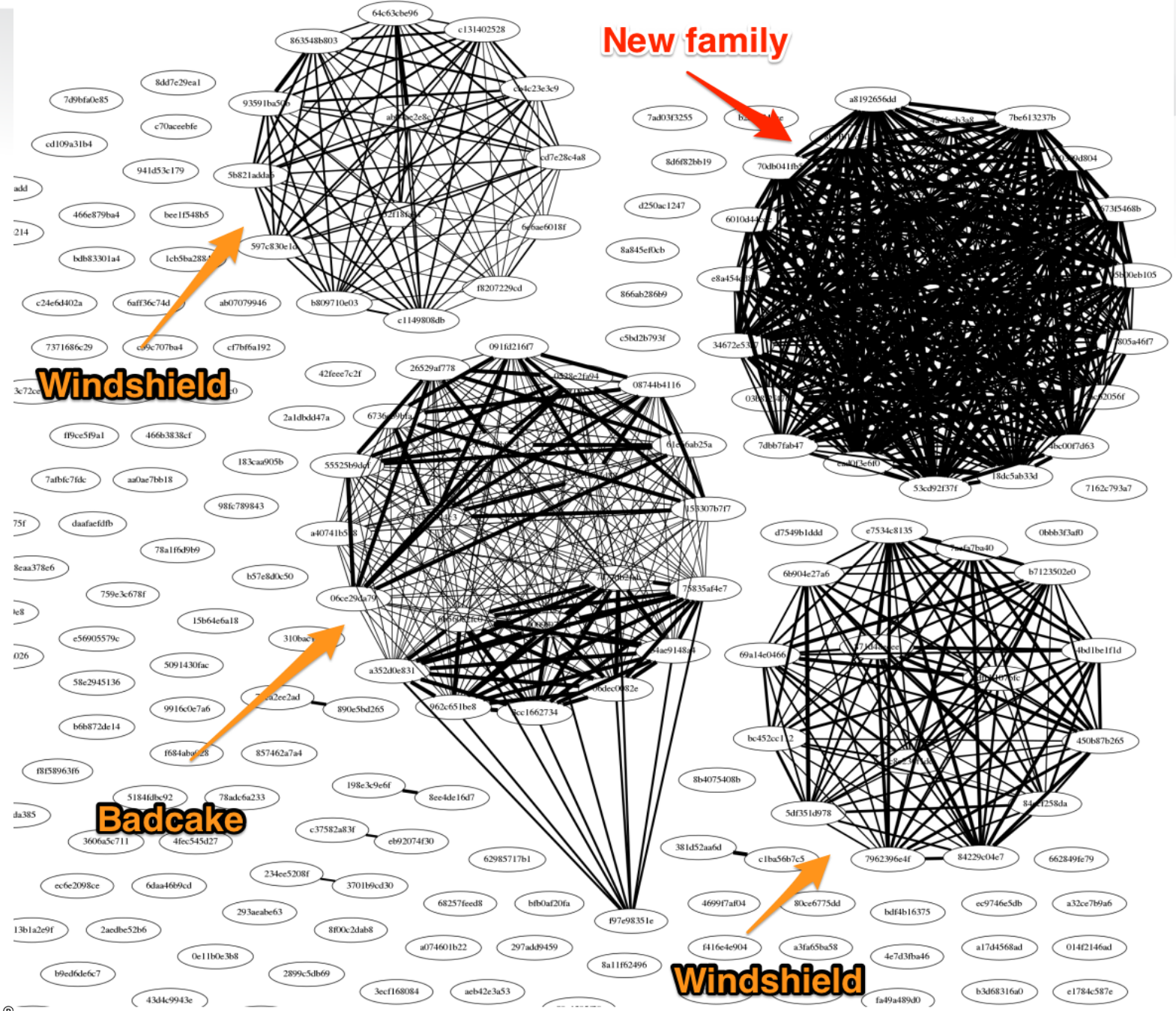
Extract all other known OceanLotus samples (along with other malware families).

Extract string features.

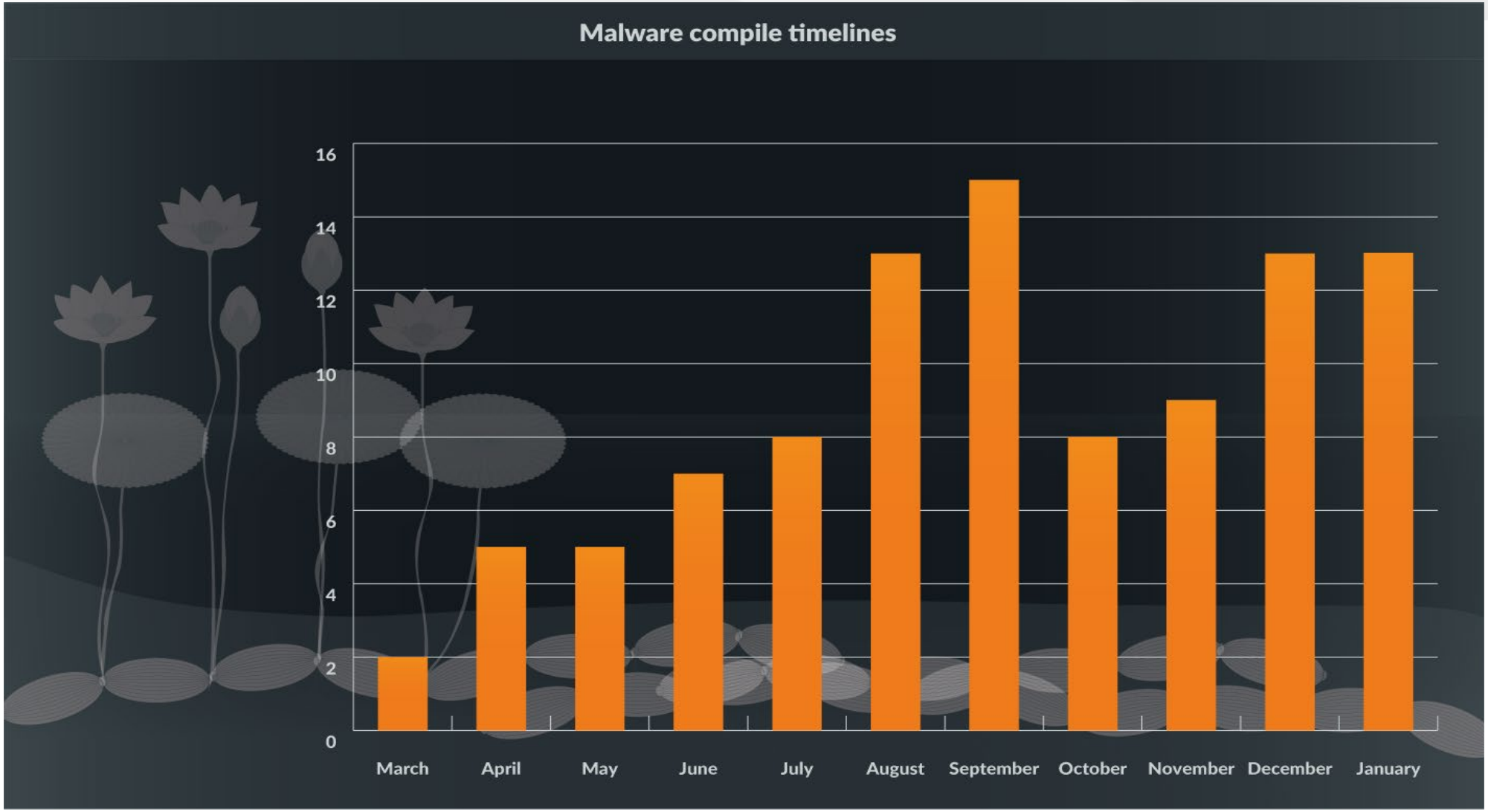
Apply Jaccard-Index on every sample features with other samples features.

Use python ‘networkx’ library for connecting the similarities.

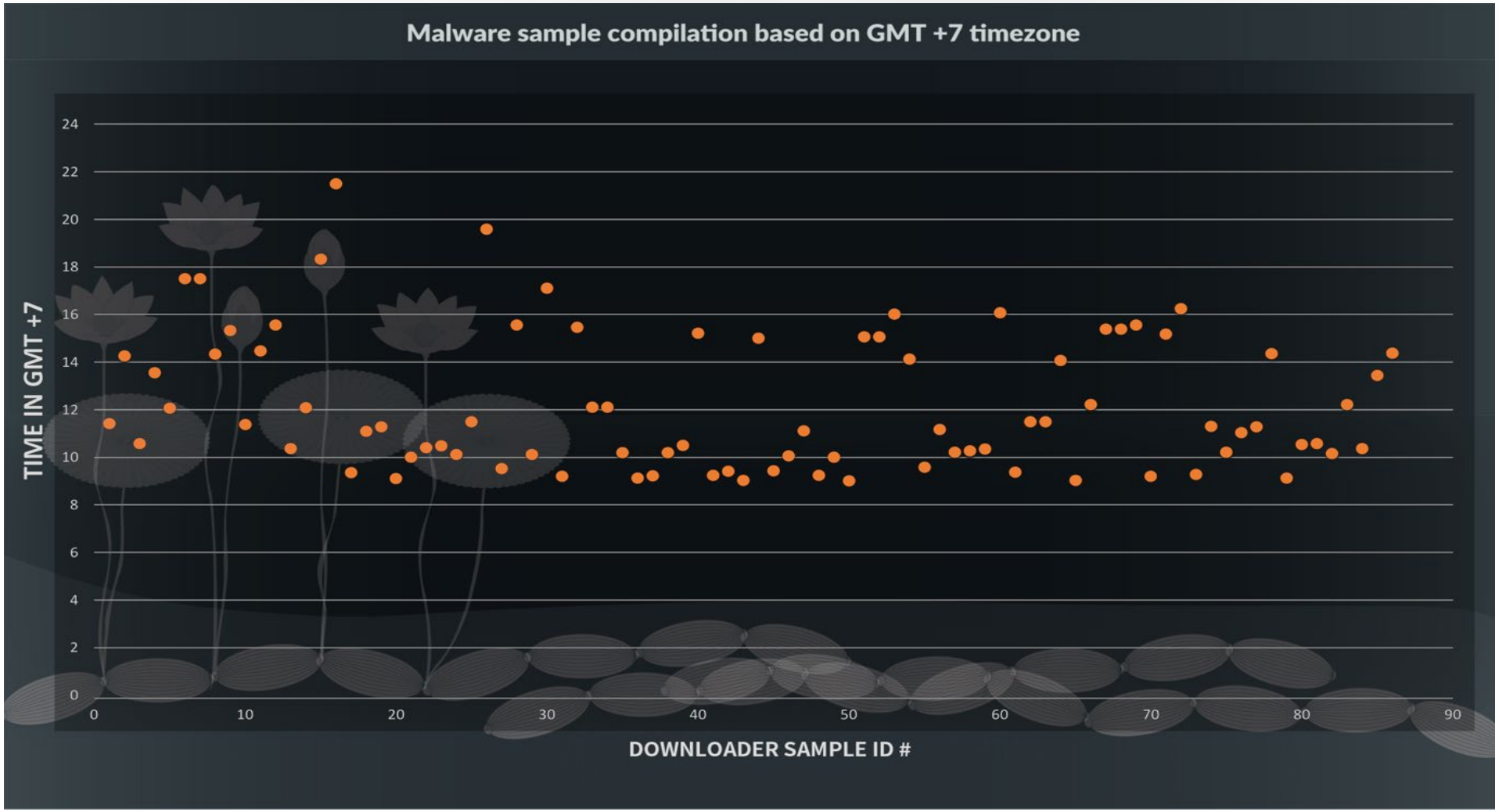
Use python GraphViz tool to visualize the graph generated by ‘networkx’.


New family
Windshield
Badcake
Windshield

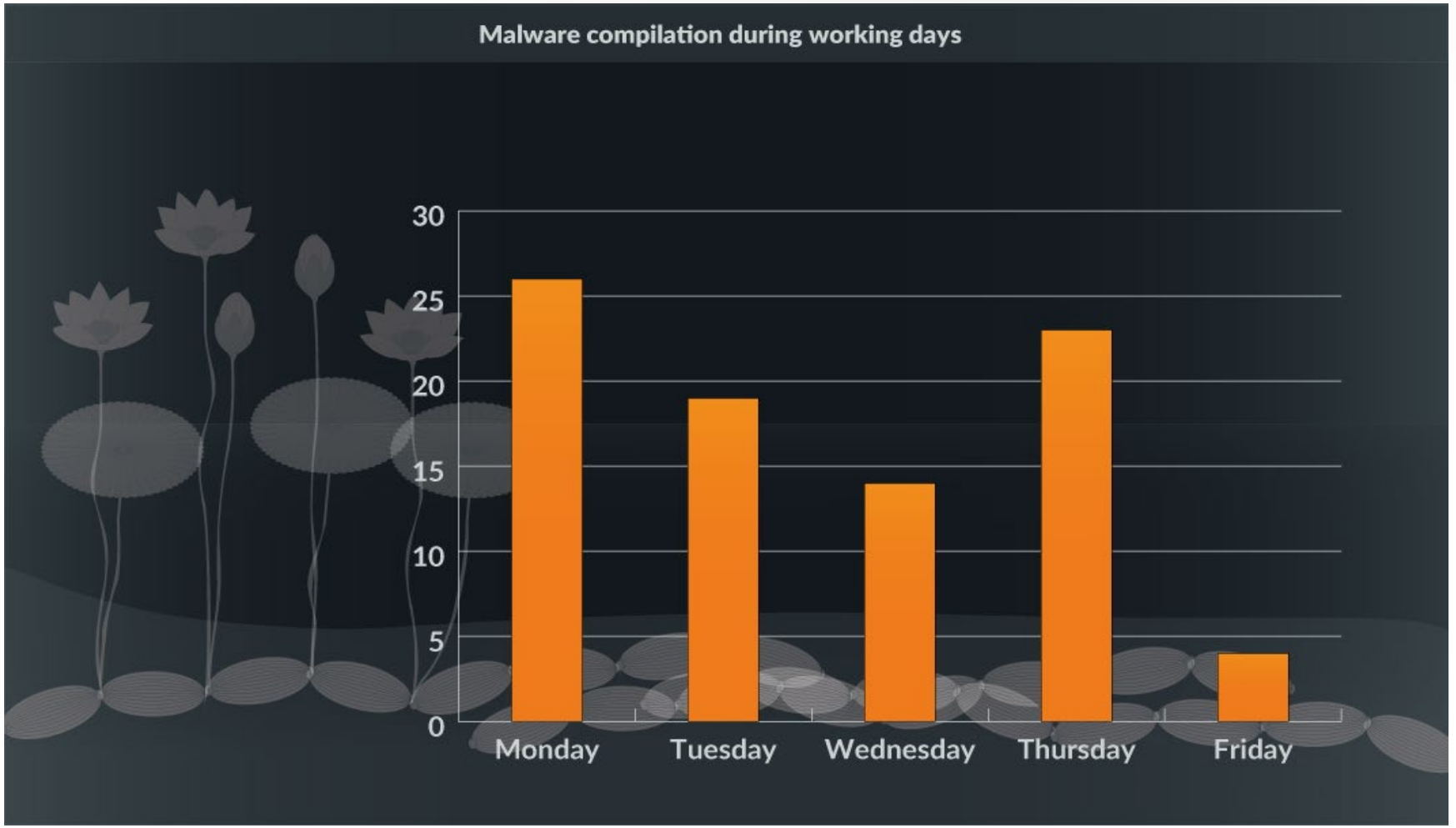
Initial campaign timeline



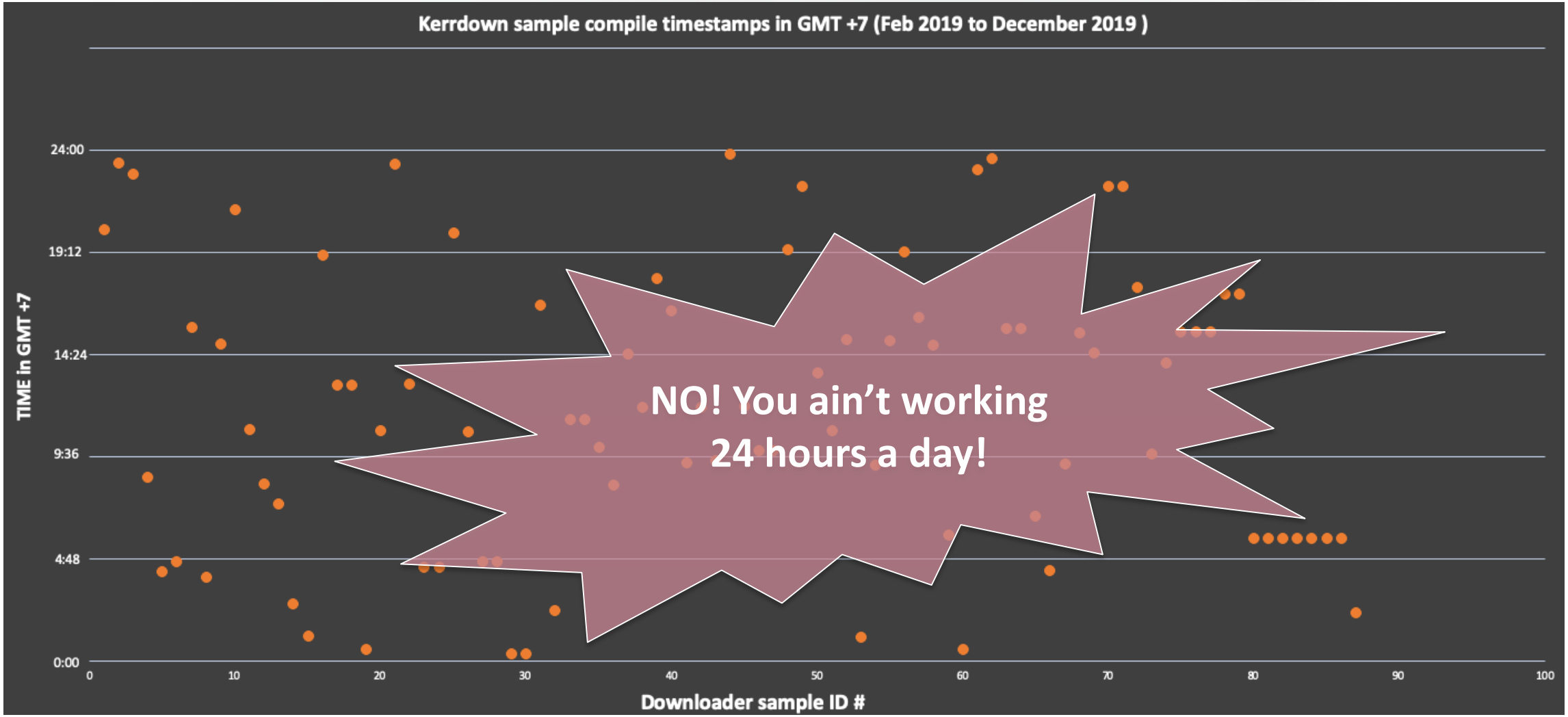
The ”Human Element” – Working hours of the OceanLotus group



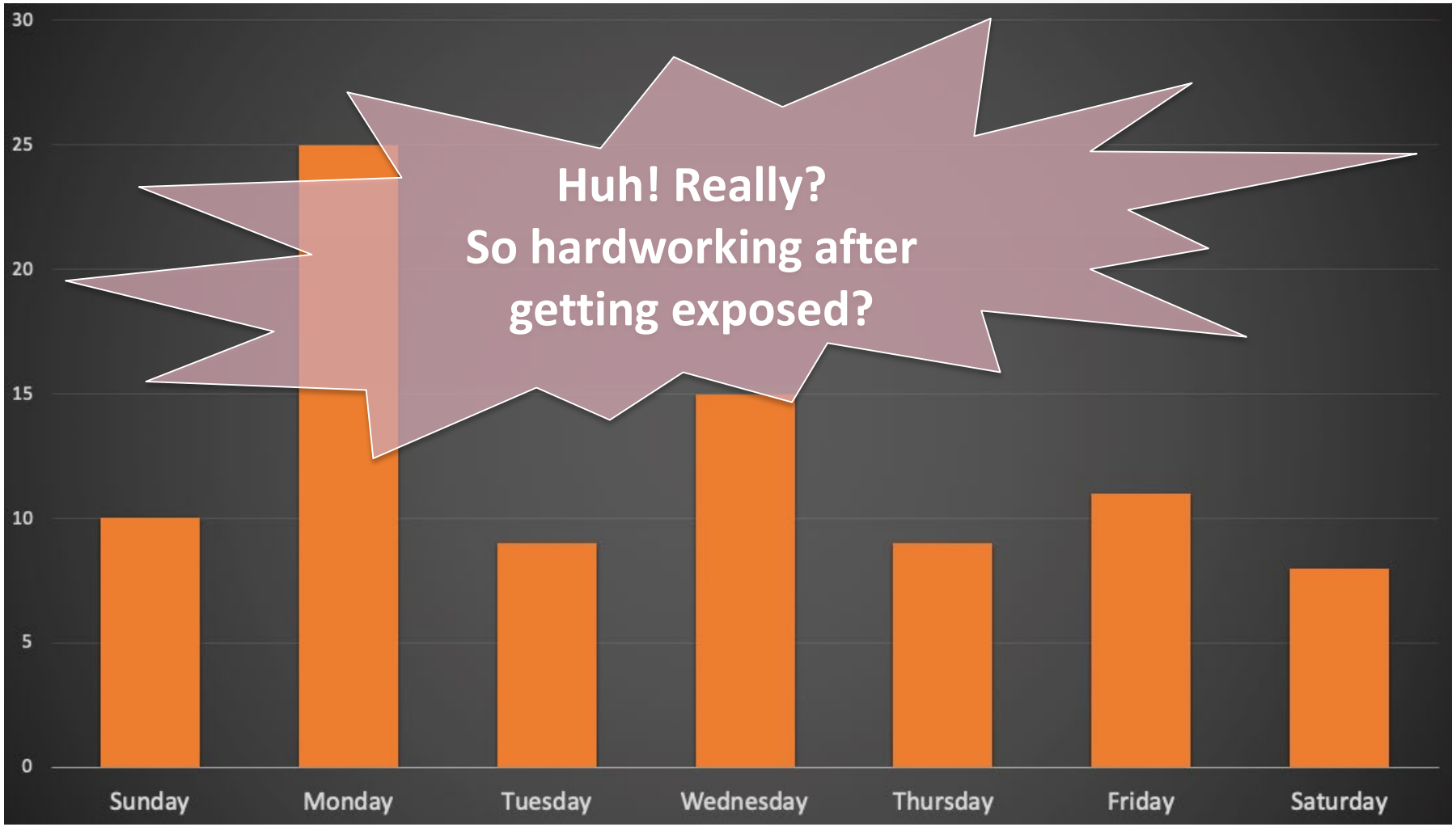
The ”Human Element” – Working days of the OceanLotus group



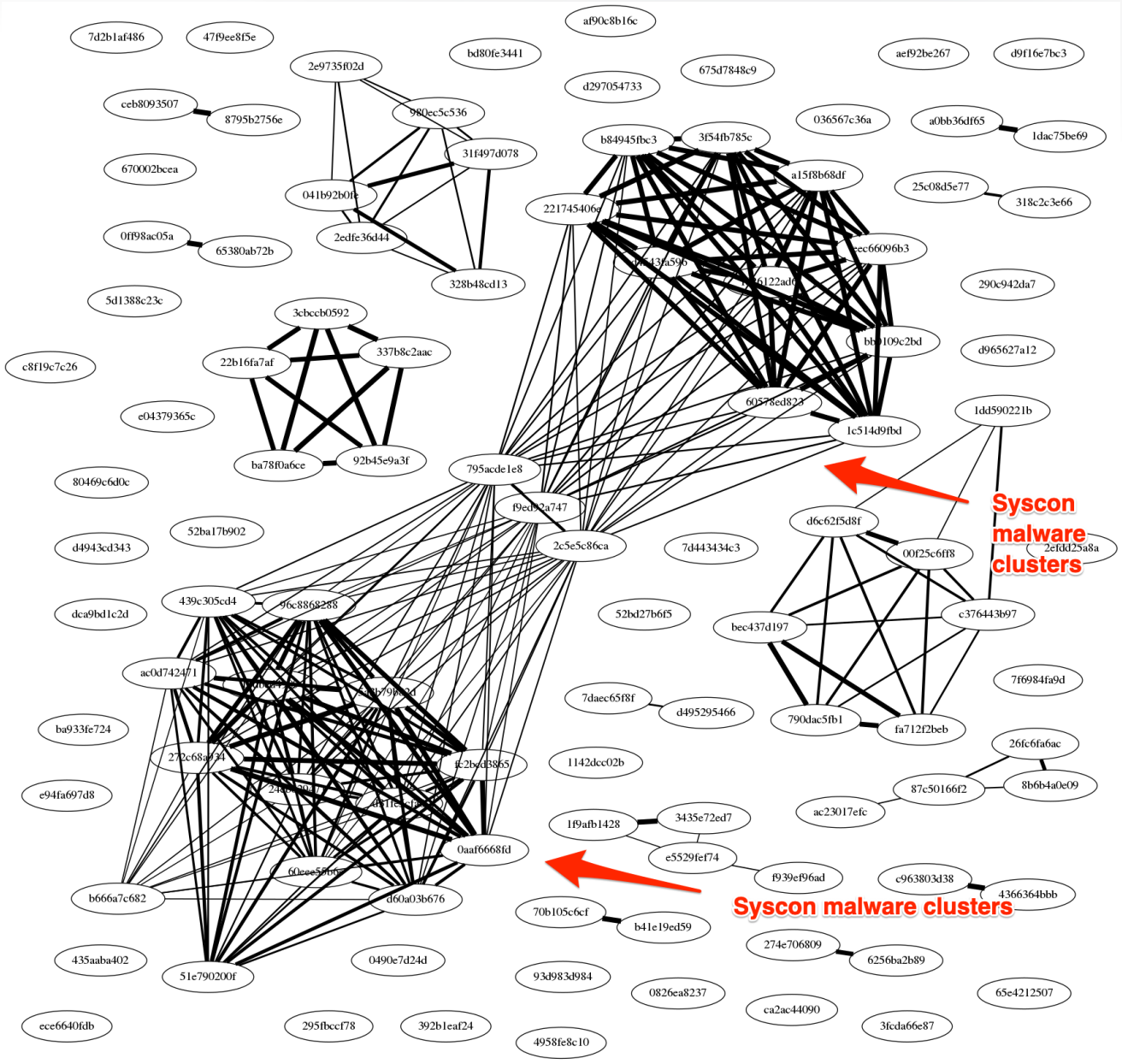
The ”Human Element” – Compile timestamps after publications.



The ”Human Element” – Manipulated timestamps

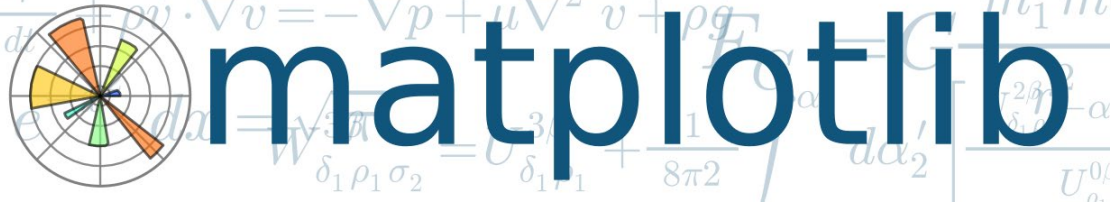


Fractured Statue Campaign – Targeting US Gov Agency

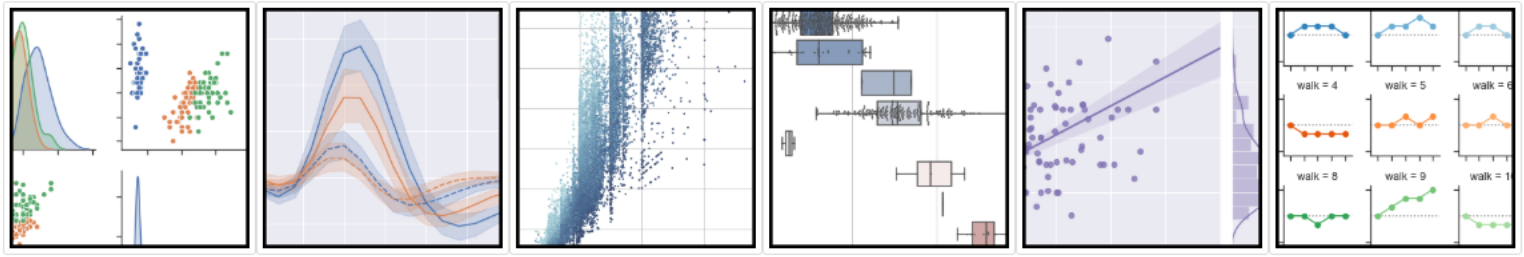


The “Human element” ☺

Data Visualization in DataScience

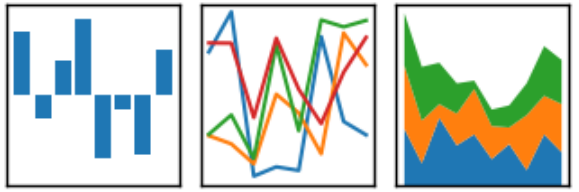


seaborn: statistical data visualization



pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



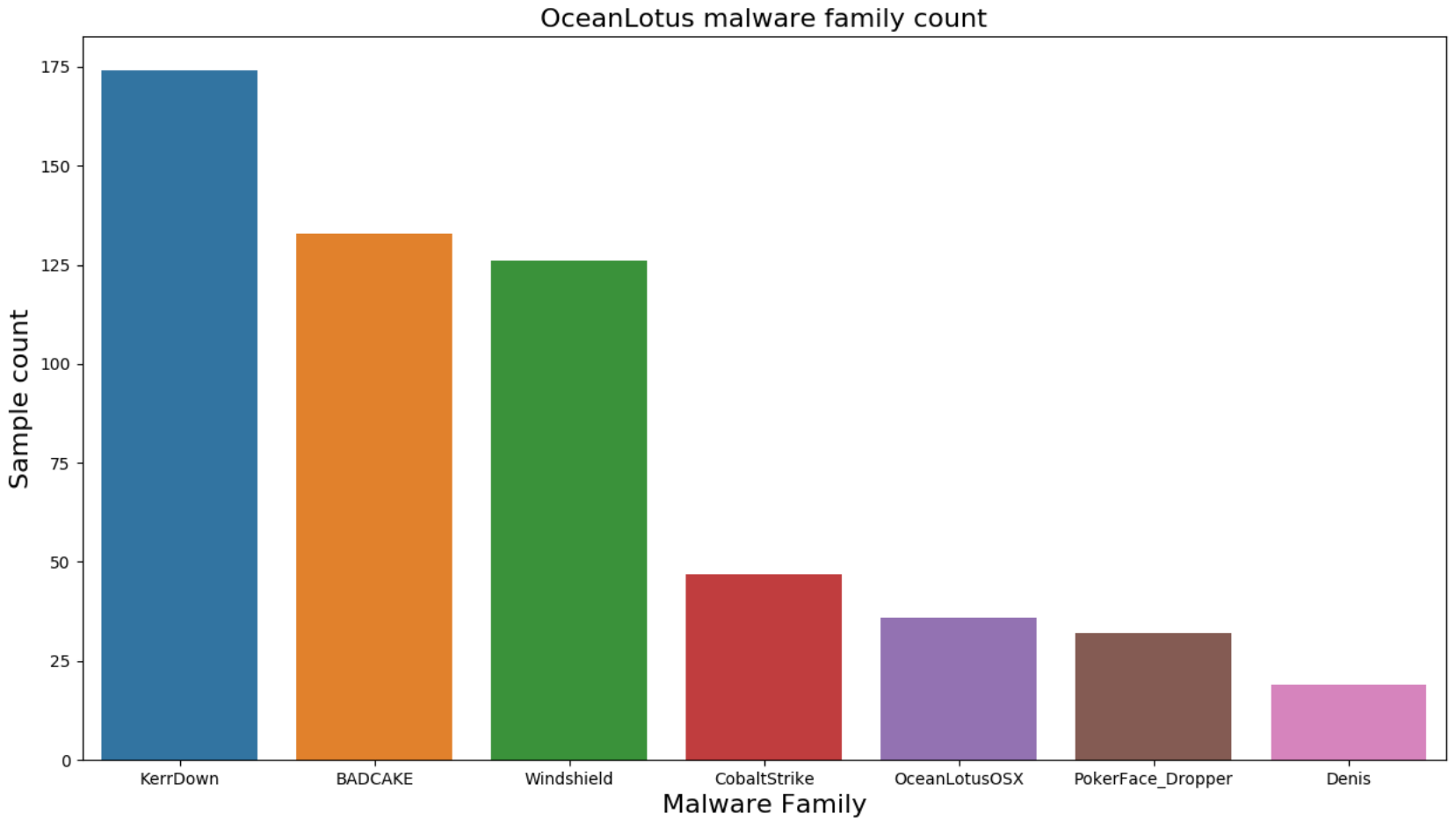
Movies


brushTip.x=event.x
brush.y=event.y-10
brushTip.y=event.y-10
pickColor(event.x, event.y)
--print("moved off rect")
brush.x=-200
brushTip.x=W-200
brush.y=event.y-10
brushTip.y=event.y-10
pickColor(W-200, event.y)
end
elseif event.phase=="cancelled" t

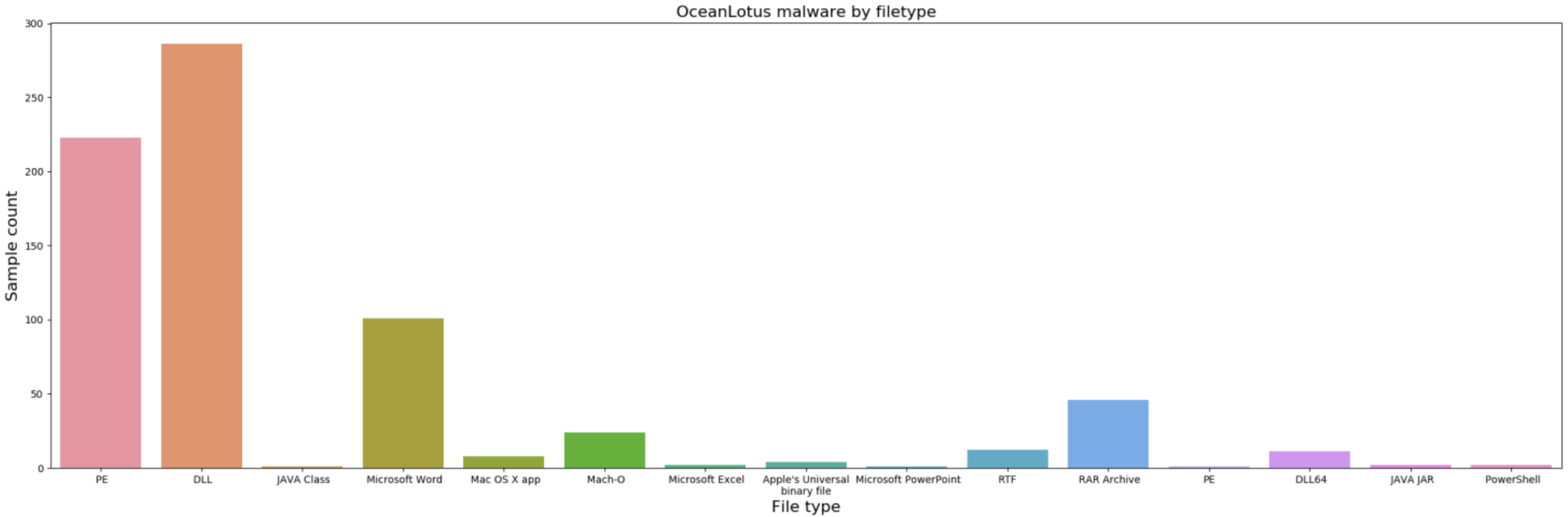
Data from Autofocus Lenz (Reality)

```
[redacted]:autofocus-lenz [redacted]$ python af_lenz.py -i query -q '{"operator":"all","children":[{"field":"sample.tag","operator":"is in the list","value":["Unit42.BADCAKE","Unit42.KerrDown","Unit42.OceanLotusOSX","Unit42.PokerFace_Dropper","Unit42.Komprogo","Unit42.Phoreal","Unit42.WindShield"]},{"field":"sample.tag_scope","operator":"is","value":"unit42"},{"field":"sample.tag_family"},{"field":"sample.tag_scope","operator":"is not","value":"private"},{"field":"sample.malware","operator":"is","value":1}]}' -r meta_scrape -l 20

{"operator":"all","children":[{"field":"sample.tag","operator":"is in the list","value":["Unit42.BADCAKE","Unit42.KerrDown","Unit42.OceanLotusOSX","Unit42.PokerFace_Dropper","Unit42.Komprogo","Unit42.Phoreal","Unit42.WindShield"]},{"field":"sample.tag_scope","operator":"is","value":"unit42"},{"field":"sample.tag_class","operator":"is","value":"malware_family"},{"field":"sample.tag_scope","operator":"is not","value":"private"},{"field":"sample.malware","operator":"is","value":1}]}

[+] sample_meta [+]

62985717b1d4164c017cd948605080a6f0fc1e69e8eb446a62117265b53adf42 | PE                                | 2013-12-29 16:10:19 | malware | 473600  | Unit42.WindShield
698c85a0266bfce5a79ee8967b13123963f7710be0e6215c677ab963cdfd7dc2 | PE                                | 2014-10-10 10:45:01 | malware | 1354240 | Unit42.WindShield
297add945921af3dfbec75601c437ffec682458761b609378f06a8f62b240f52 | PE                                | 2015-02-01 03:45:43 | malware | 443904  | Unit42.WindShield
c1ba56b7c5388d870951cae1986169cbfdd4a8a1ad002960aa0527485564fbe4 | PE                                | 2015-07-31 05:15:33 | malware | 1248768 | Unit42.RenameOnReboot,Uni
381d52aa6df750f23c5b53d2fcb563002a5a5ea6c32ad3367338c50774e38c18 | PE                                | 2015-07-31 09:15:19 | malware | 1250816 | Unit42.RenameOnReboot,Uni
f8f58963f6fd9e9c08f5527f176cfc5fe379594a5cb57096e5180390266da968 | PE                                | 2015-08-11 11:43:24 | malware | 473600  | Unit42.RenameOnReboot,Uni
7afbfc7fdce8075557c8753d345eed277db8a7597041571653fbd7967f26ca9a | PE                                | 2016-01-09 00:27:06 | malware | 552960  | Unit42.WindShield
a7271ebd48001f19c6ea8b64602ded790881b6d3379968a6fe4b34df76977492 | PE                                | 2016-01-29 01:55:43 | malware | 1189376 | Unit42.WindShield
941d53c179513b53a8f8b6079ac0e843dae0656da213efb9c5080b76c00ed94f | PE                                | 2016-04-17 02:46:48 | malware | 543744  | Unit42.WindShield
8d6f82bb19ea584473ddc79067d8e7db84f1acc07373ce61a1c4af4da3d851e5 | DLL                               | 2016-07-22 05:20:28 | malware | 2923008 | Unit42.WindShield
ab0707994603071d2aed0c0c229f3eeea1e6c67ac85ff8089b5b7d639c4311c1 | DLL                               | 2016-07-25 22:24:09 | malware | 2921984 | Unit42.WindShield
a32ce7b9a6a55fa4605a171384b1e11b7e75355c9062933a4c24b3a09469e294 | PE                                | 2016-11-24 19:15:52 | malware | 2278912 | Unit42.WindShield
d5ed6f95550f9660982279068f4859f9f66b52696a0998cd0579324abbcd2fe7 | PE                                | 2017-01-07 08:43:19 | malware | 2868224 | Unit42.WindShield
8a62781aba50d4cacb384e623283207ca975c4fe6b7ba6ce1050b68c75d69c46 | Microsoft Word Document           | 2017-02-03 02:45:08 | malware | 36305   | Unit42.PokerFace_Dropper
7a5773691c48435d07841f811734d44dde4c371a2f6d61b4f46111856dcb04a2 | Microsoft Word 97 - 2003 Document | 2017-02-03 06:36:06 | malware | 59392   | Unit42.PokerFace_Dropper
4883905ff9a44ed0ed4c45f66ea4d83be88825521f6da3a5819f91b396ba4efb | Microsoft Word Document           | 2017-02-05 23:24:14 | malware | 30459   | Unit42.PokerFace_Dropper
3a748339a7e681125e8a24e3a297497c52a6a63a876536ba59dbf26e127cecc9 | Microsoft Word Document           | 2017-02-05 23:24:15 | malware | 30525   | Unit42.PokerFace_Dropper
a67929c2b90403d07204fbd47f553827df530a5d1768e83a03f97fcf2c0294ad | Microsoft Word 97 - 2003 Document | 2017-02-06 02:45:11 | malware | 36352   | Unit42.PokerFace_Dropper
f7dd78538b7cdecc61e9f0a9bc7cb98ef93203925f83f62e5312e691b15e73d2 | Microsoft Excel 97 - 2003 Document | 2017-02-22 02:54:00 | malware | 98304   | Unit42.PokerFace_Dropper
9916c0e7a610de3cf20ee3bac7a0688ecb027b3c8c34e05e8846c6e7f68d4766 | DLL                               | 2017-04-27 07:58:20 | malware | 221184  | Unit42.WindShield
```

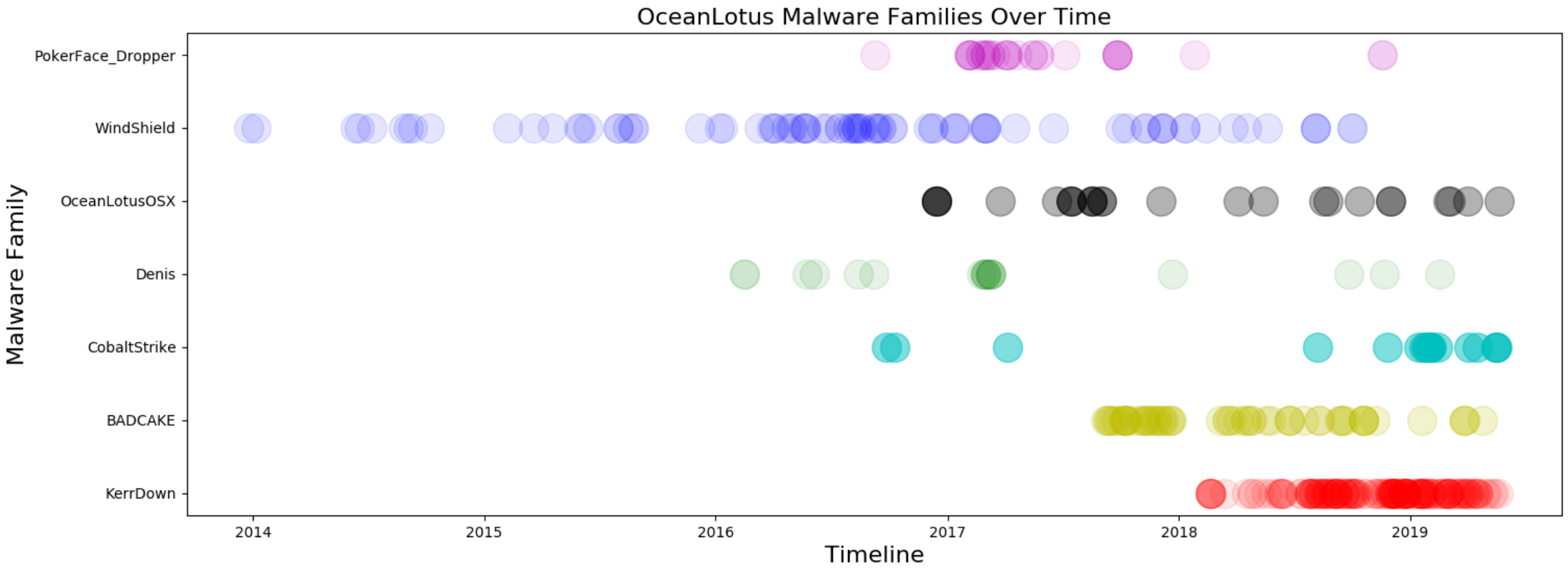
Stats by malware family



Stats by filetype



OceanLotus malware family timelines using Matplotlib & pandas



RSAConference2020

Quick Demo

Some magic with Jupyter Notebooks..

Jupyter Notebook uploaded to Github

https://github.com/vicky-ray/RSAC2020-OceanLotus-JupyterNotebook

Learnings & future work

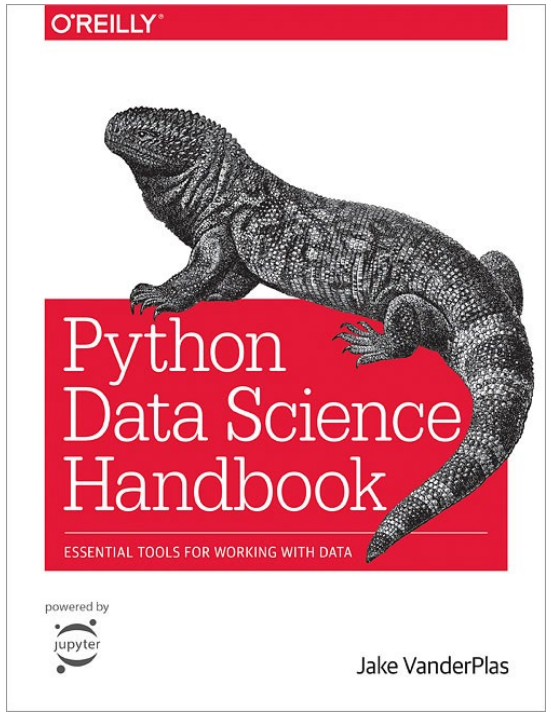
- Data Science allows us to discern insights which is otherwise difficult with traditional tools used by threat analysts & researchers.
- Learn about TTPs faster
- Explore better ways to visualize data with python libraries like pandas, matplotlib, seaborn etc.
- Essential to have a data scientist to be part of threat team, SOC etc.
- Explore more ways to extract features and apply similar algorithms-> Machine Learning

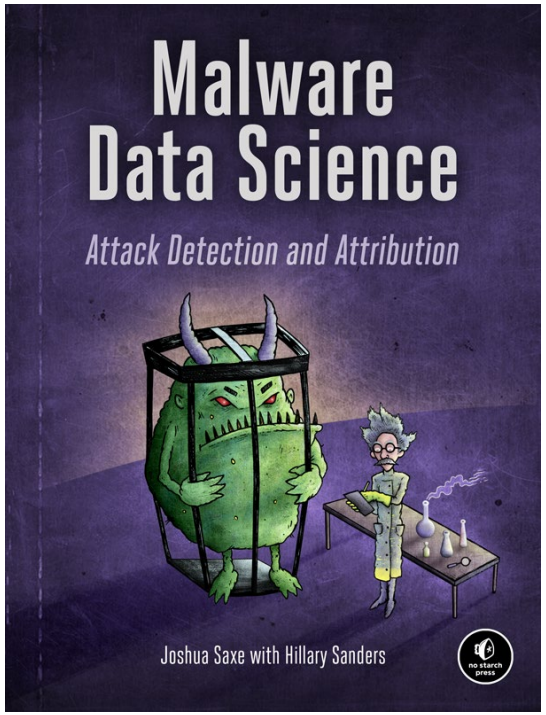
Who is the missing piece in the A team?

References & Credits

https://www.blackhat.com/docs/us-14/materials/us-14-Saxe.pdf

https://www.cs.bgu.ac.il/~frankel/TechnicalReports/2017/17-05.pdf





THANK YOU!

Twitter: @0xVK

https://unit42.paloaltonetworks.com/