RSAConference2019
Asia Pacific & Japan
Singapore | 16–18 July | Marina Bay Sands

SESSION ID: SEM-T03D

Tracking Attack Campaigns using Data Science

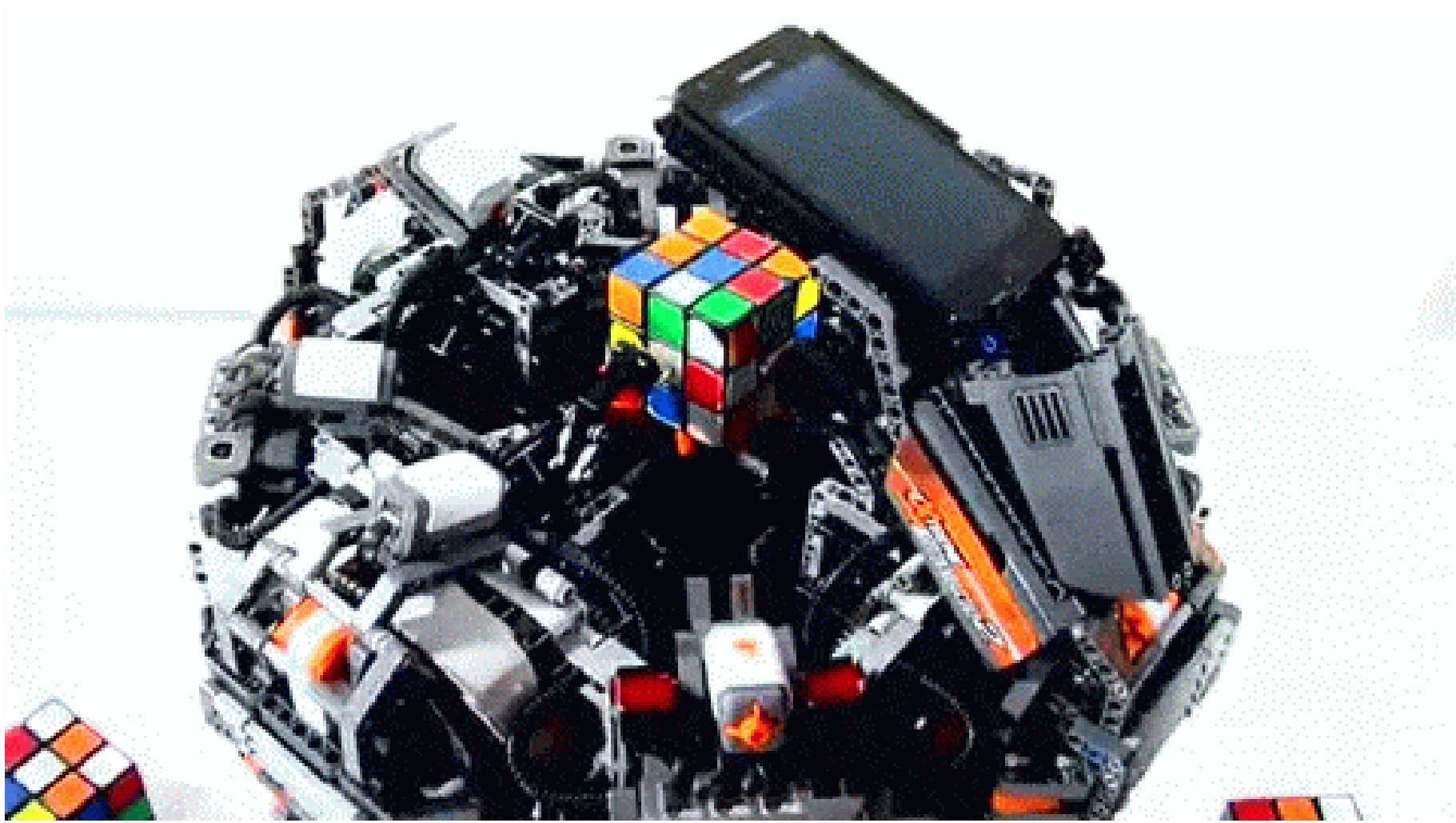
Vicky Ray

Principal Researcher
Palo Alto Networks
@0xVK

Vicky Ray | Principal Researcher, Asia Pacific

- Specialize in tracking cybercrime and APT campaigns.
- Outreach & building external relationships in APAC.
- Assigned to INTERPOL.
- Background in Security Operations, Threat Intelligence, Incident Response and building SOC & CERT Teams.
- Collaborating with Global Law-Enforcements.

UNIT 42



paloalto NETWORKS

UNIT 42

paloalto NETWORKS

Agenda

- Why Data Science?
- Challenges in finding similarities in malware
- Similarity analysis
- Unit 42 case study - OceanLotus
- Visualization using Data Science

What is Datascience?

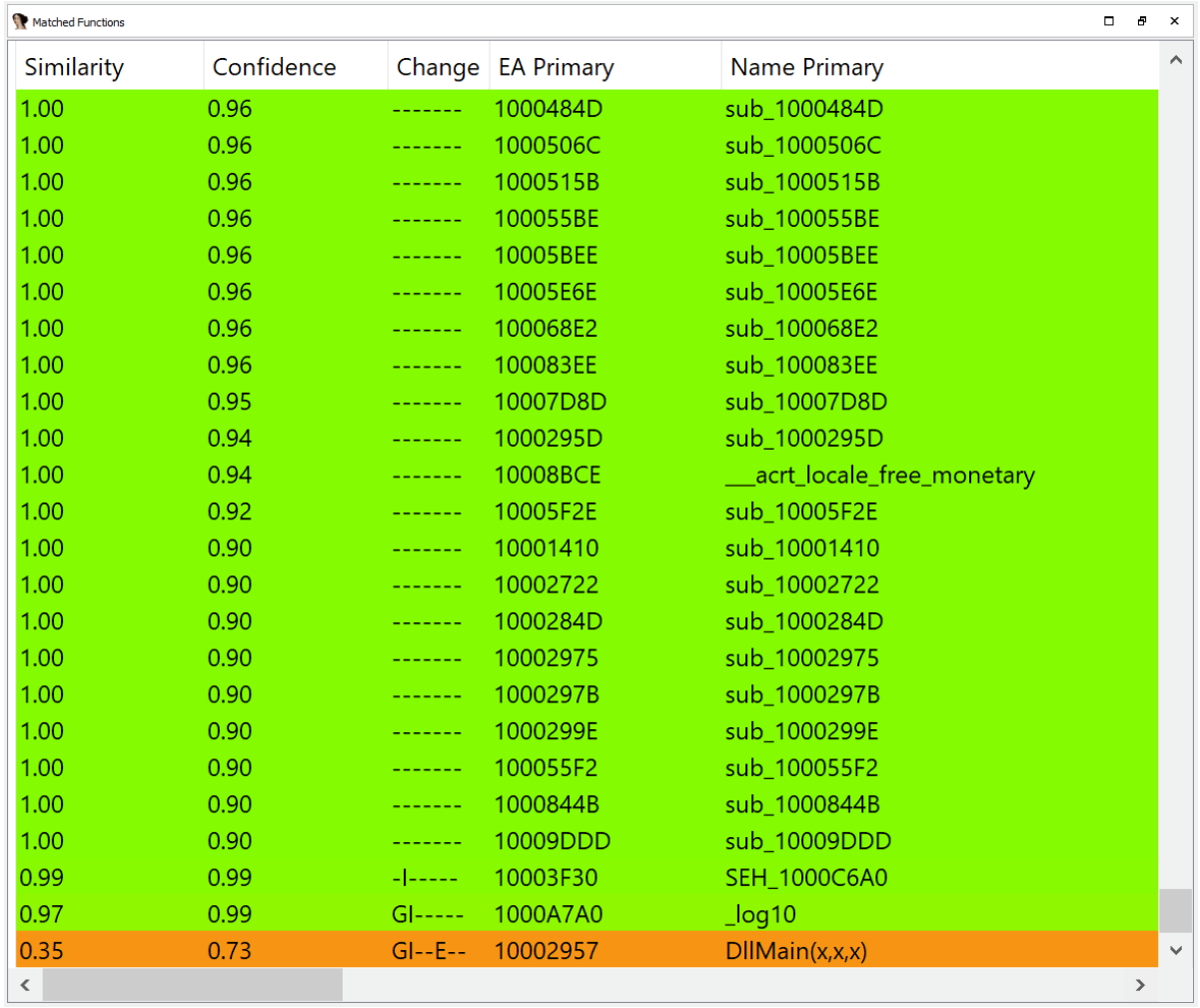
I am NOT a Data Scientist

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

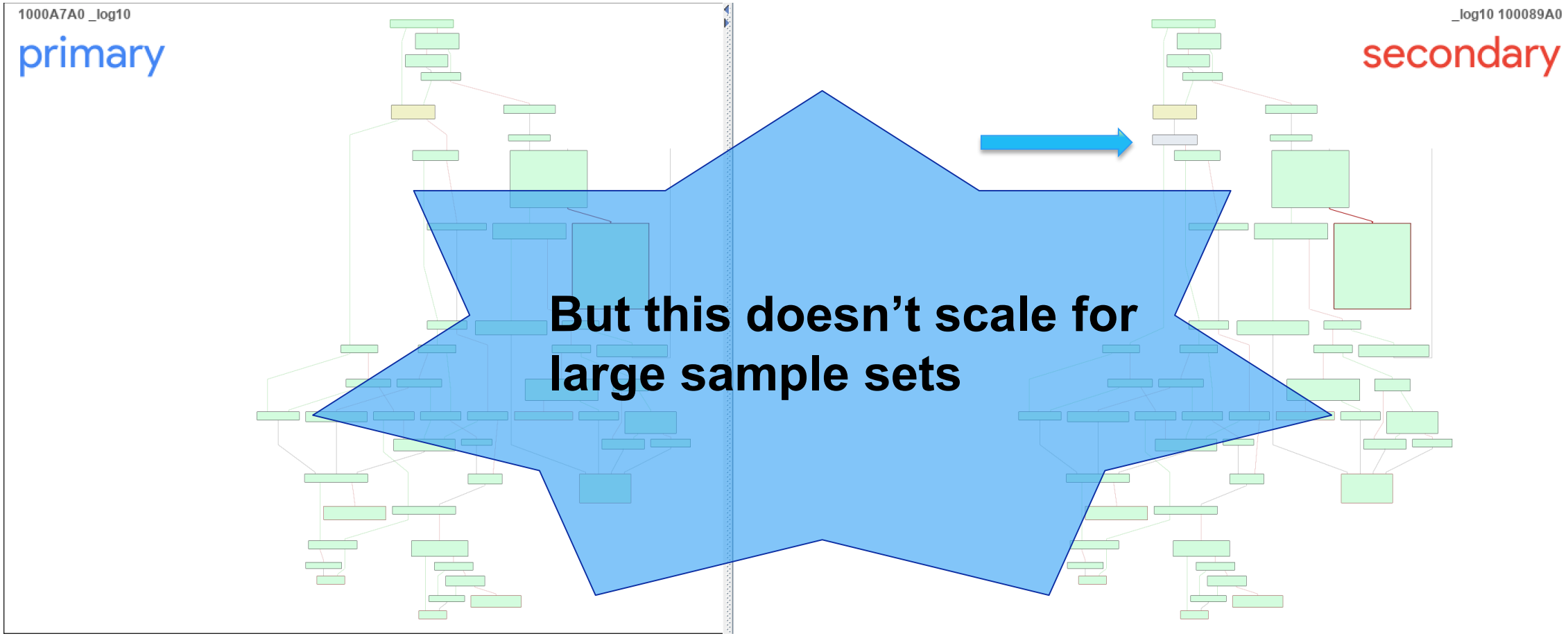
Unique attributes of two samples from same malware family

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

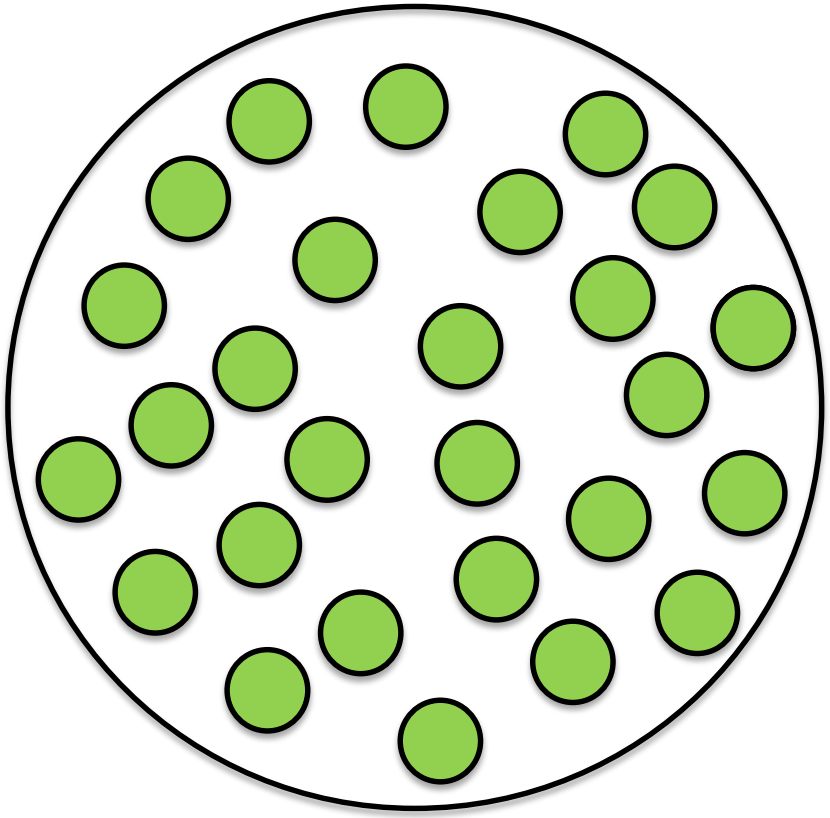


Similarity analysis on IDA using BinDiff

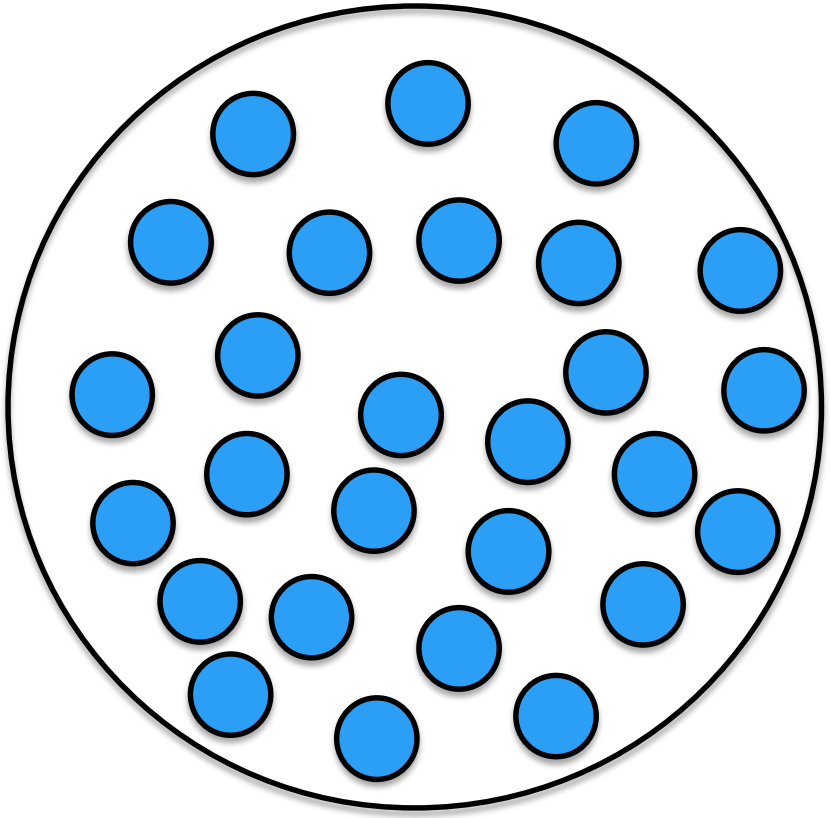


How can we reduce analysis times?

Feature Similarity



Sample A

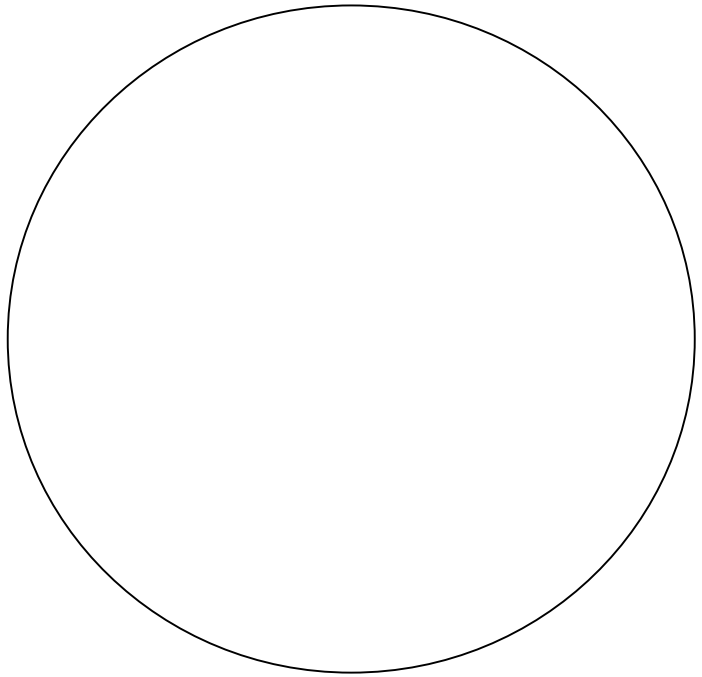


Sample B

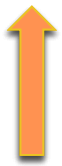
Types of features for malware

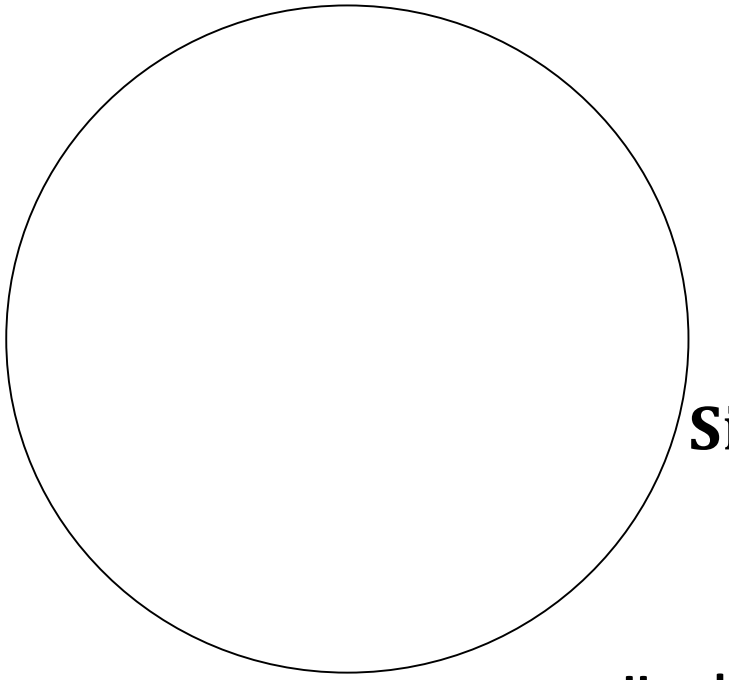
- String based features
- IAT
- PE header
- Instruction sequence
- Dynamic analysis

Jaccard Index to the rescue



Set A





Set B

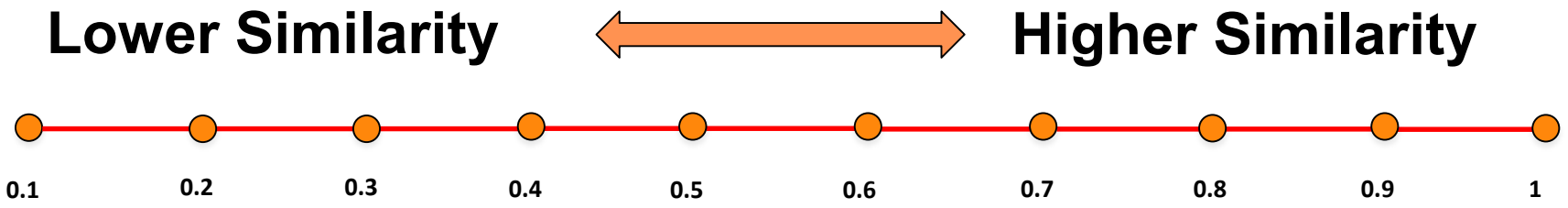
$$\mathbf{Sim(A, B)} = \frac{\mathbf{A \cap B}}{\mathbf{A \cup B}}$$

JI value range - 0 to 1

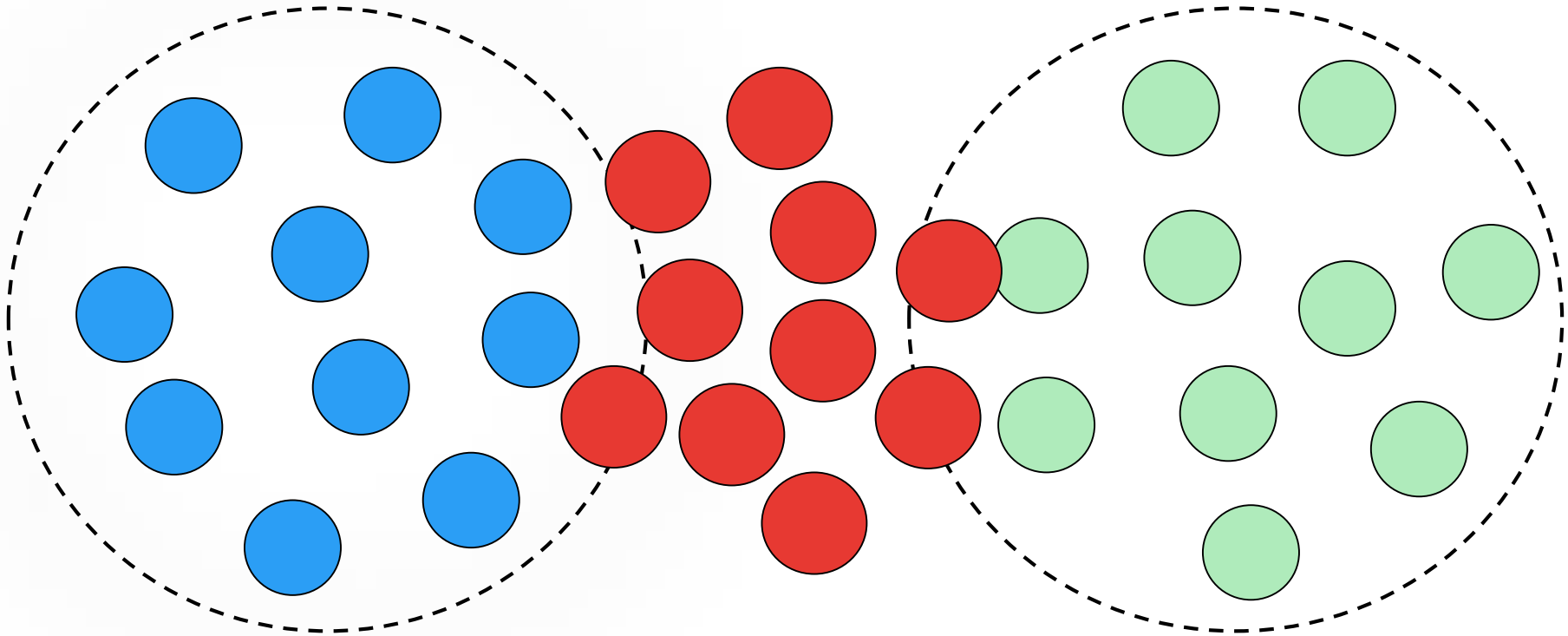
Values closer to 1 has higher similarity

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



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

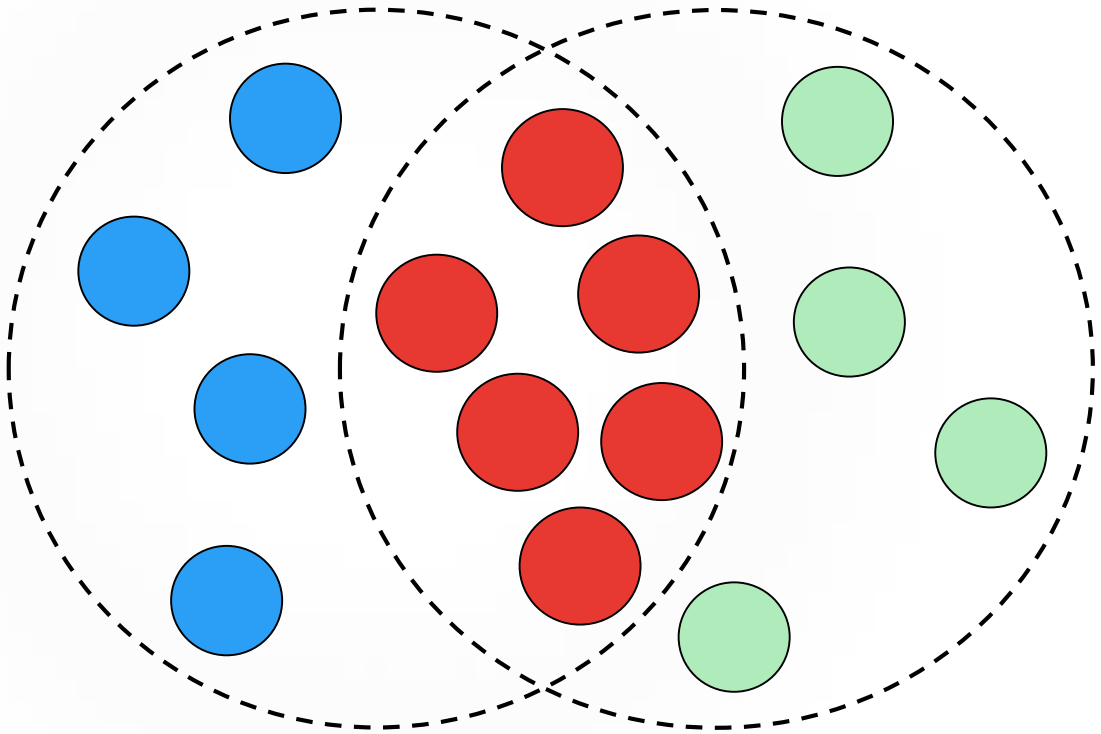
$$\mathbf{Sim(A, B)} = \frac{\mathbf{0}}{\mathbf{20}} = \mathbf{0}$$

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

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

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

Using Jaccard Index to extract similarity



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

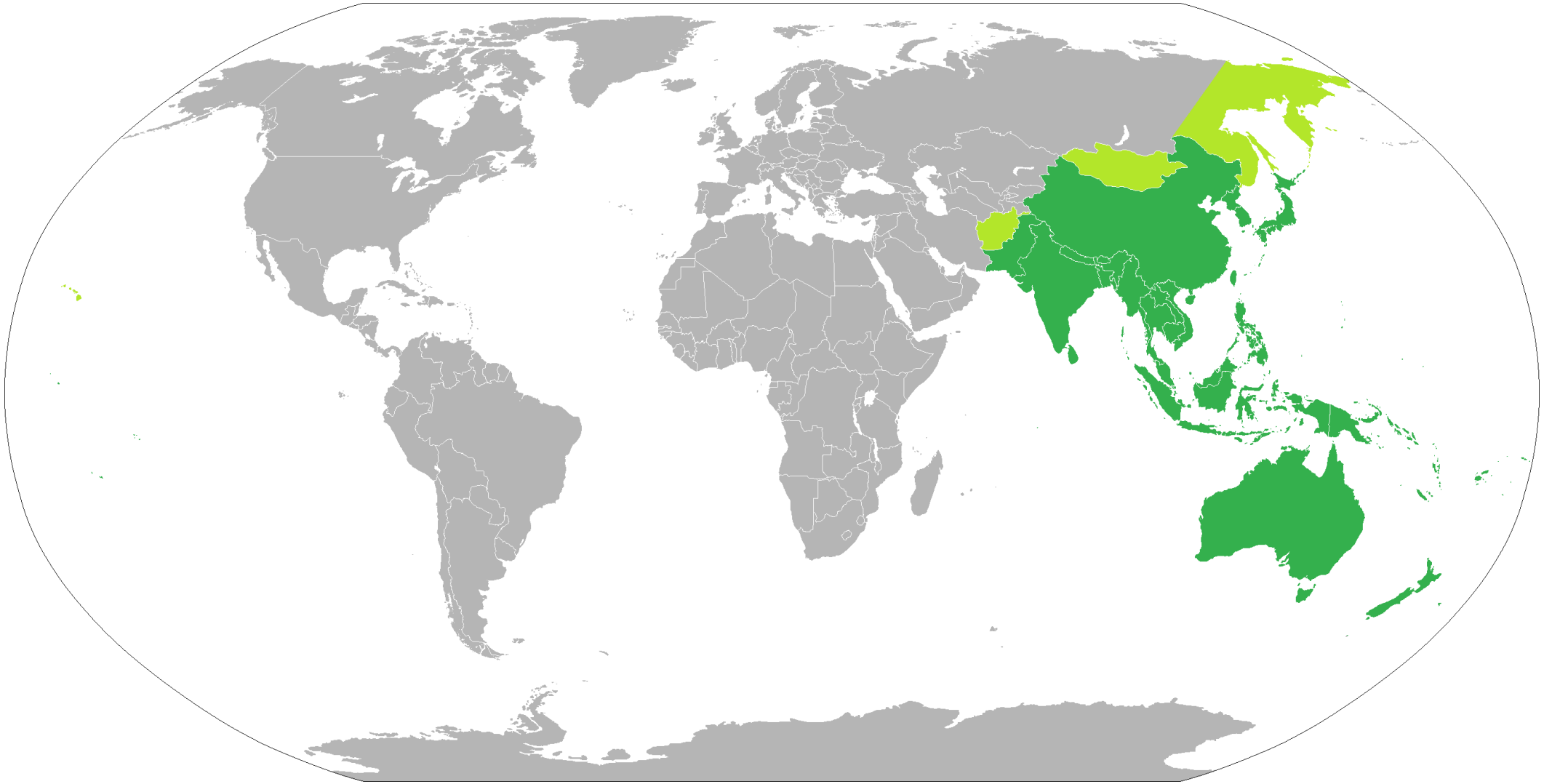
$$\mathbf{Sim(A, B)} = \frac{6}{14} = 0.42$$

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

Case study

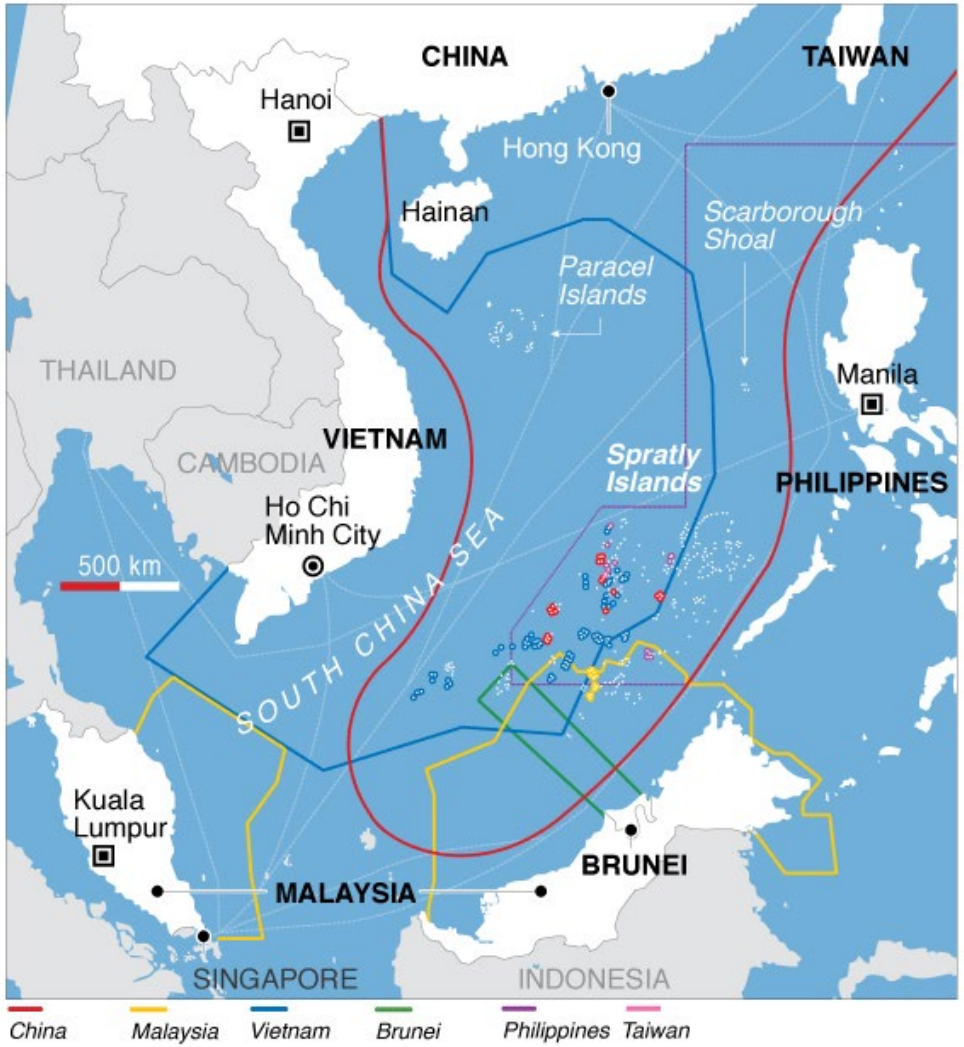
- **Implementing jaccard-index in real world case study and threat intel hunting.**

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

LATEST RESEARCH TOOLS PLAYBOOKS ABOUT US

unit42

SUBSCRIBE

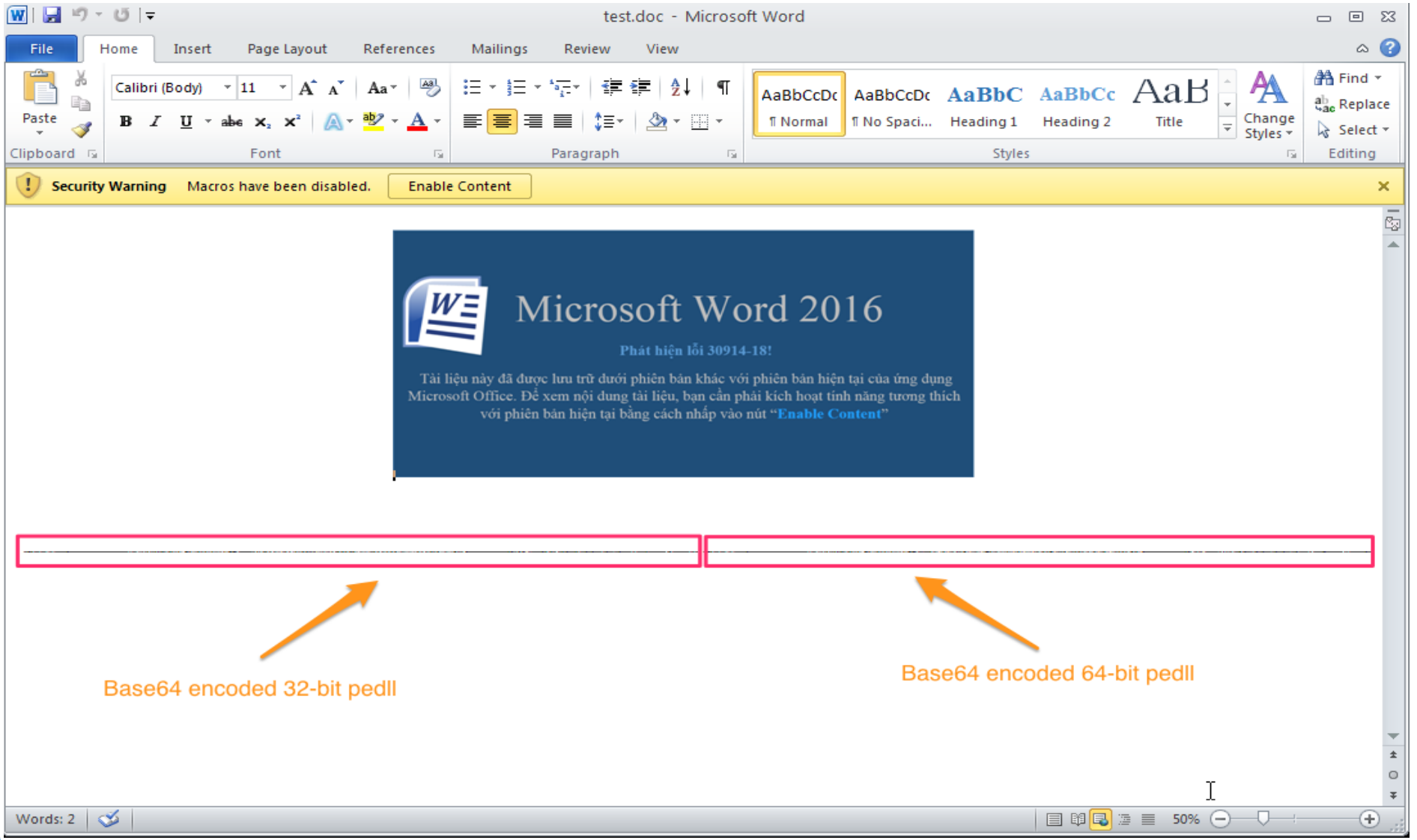
paloalto NETWORKS

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

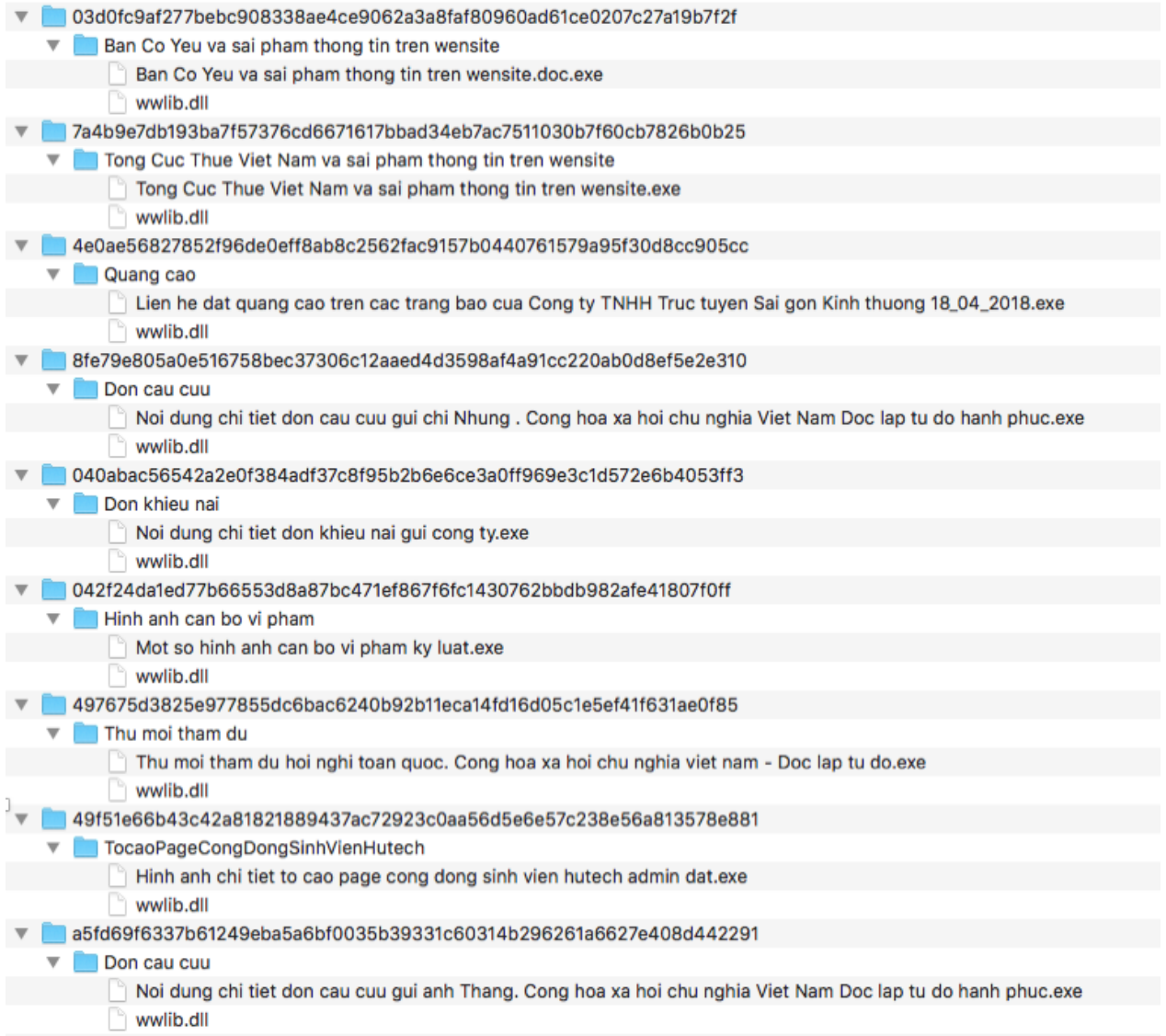
```
TVqQAAMAAAAEAAAA//8AALgAAAAAAAAAQAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAEAEAAA4fug4AtAnNIbgBTM0hV
GhpcyBwcm9ncmFtIGNhbm5vdCBiZSBydW4ga
W4gRE9TIG1vZGUuDQ0KJAAAAAAAAAAqpnrWb
scUhW7HFIVuxxSF2lvlhWfHFIXaW+eFG8cUhdpb
5oV2xxSFVZkXhIzHFIVVmRGEdMcUhVWZEIR+x
xSFZ7+HhW3HFIVuxxWFNccUhfyZHYRvxxSF/Jnrh
W/HFIVux4OFb8cUhfyZFoRvxxSFUmljaG7HFIUA
AAAAAAAAAAAAAAAAAAAAAAAAAAAABQRQ
AATAEGAENVo1sAAAAAAAAAAOAAAiELAQ4AAK
IAAADu/
AEAAQA
ADgBAA
EAAAEAAAAAAABAAAAAAAAAAAAAAAKQQAQ
AoAAAAAMAEAEgCAAAAAAAAAAAAAAAAAAAA
AAAAANAEALQNAADACQEAOAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAPgJAQBAAAAAA
AAAAAAAAAAwAAAHAEAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAC50ZXh0AAAAC6EAAA
AQAAAAogAAAAQAAAAAAAAAAAAAAAAAAACAA
AGAucmRhdGEAAPhWAAAAwAAAAFgAAACmAA
```

```
TVqQAAMAAAAEAAAA//8AALgAAAAAAAAAQAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAEAEAAA4fug4AtAnNIbgBTM0hV
GhpcyBwcm9ncmFtIGNhbm5vdCBiZSBydW4ga
W4gRE9TIG1vZGUuDQ0KJAAAAAAAAAAAjfZU3Zxz
7ZGcc+2RnHPtk04AKZGMc+2TTgAhkEhz7ZNOAC
WRqHPtkXEL4ZW8c+2RcQv5lfbz7ZFxC/2V3HPtk
bmRoZGQc+2RnHPpkORz7ZPVC8mVmHPtk9UIE
ZGYc+2RnHGxkZhz7ZPVC+WmHPtkUmljaGcc+2
QAAAAAAAAAAAAAAAAAAAAAAAAAAABQ
RQAAZIYHAEVVo1sAAAAAAAAAAPAAIiALAg4AAJ
CAAQA
AAAAAA
AAAAQA
EAAAAAAAAAAQAAAAAAAEAAAAAAAAAAAAA
QAAAAAAAAAAAAABoLwEAKAAAAACwBABIAg
AAAJAEAKgMAAAAAAAAAAAAADABAAMBgAA
kCEBADgAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAADQIQEAlAAAAAAAAAAAAAAALAAAE
gCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAudGV4dAAAAPCdAAAAEAAAAJ4AAAAEAAA
AAAAAAAAAAAAAAAAgAABgLnJkYXRhAAAchwA
```

```
MZ@ !L!This program cannot be run in DOS mode.
```

```
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



First Seen	WildFire Verdict	MD5	File Size (Bytes)	Import Table Hash	Ssdeep Fuzzy Hash
04/25/2019 3:37:19am	Malware	208e984792e322d401602d45a043c9db	83,968	dd7918d309ef66eaf0dd80849153d3b1	1536:sHxfHdMZvkznEcl+QtSbM2TsbToqo7EbDwQIusWscdaHiah68nTq4i:sHHRFR8M2TaFctlqaC8NnTqD
04/22/2019 5:11:43am	Malware	097978b6d22be6b2ed739e6c7b4d567a	4,278,784	dd7918d309ef66eaf0dd80849153d3b1	98304:i9Bnj6niKDTKn4BVnAk9+TFKFrKOnLYWUUvVnJv4vTv:i9hjAiKDTw4BdAk9+TFKFrKsLXUUvdJG
04/18/2019 1:30:34am	Malware	635e563c8620dff00176df7efdd74dfe	71,146	7bae0f81f57e10fd1249e9e2e448191f	1536:QrI3wkTs3xHkpXrWMUqrLZtENQBKsWzcdw/LTax:shArWQKI1w/LTa
04/17/2019 1:48:21am	Malware	9af9be96c14ef18e641486f51f840bc2	342,016	544ea5a03970edeca3e3ed55c2bd1a16	6144:ZpYDTYetamawVcYH5gEqycnJoaOWDnFEFW1c2CNohb8jXG:ZpY9taRwqYH+Tia6olE
04/17/2019 1:01:06am	Malware	c81bac80b928f489bdd02b02cf4fcac5	287,744	7bae0f81f57e10fd1249e9e2e448191f	3072:shArWQKI1w/LTFB0wUYSYtRRox/pQJE1UxInIFtwSqWVOAg0FujMf2DuSaAY8/fG:JrWlpywWY3JE1UGYBOAOQvDgHsF
04/15/2019 9:30:15pm	Malware	f5978aab68abe95bd00c77a6e2d07627	1,258,567		24576:E2dlMD1fuMB5HP09RkNDT2O6w5Cj+ECjmXCGZ3cUdXtX4:EA+LHP0kX6/C0N3c9
04/12/2019 2:01:16am	Malware	b8d588e9b272d262dbd12e849833f888	297,472	7bae0f81f57e10fd1249e9e2e448191f	3072:/tuNa+nI1wjLPPXV6r5H7YM03kIqsMc6RH4+qhAg0FuKAg0FuQMBSWYjtsEu:cNaygr5Hs/1O5KLhAOKAOPSWYjtsF
04/12/2019 2:01:15am	Malware	5bb2a74e135fedf0997484a1a8a14976	292,352	544ea5a03970edeca3e3ed55c2bd1a16	6144:4AcSaU0s8Q0U5aTD/+n1VTzyjx/sohaQo:4Ac7LQ0UzIx/so
04/05/2019 11:01:24pm	Malware	88eae0d31a6c38cfb615dd75918b47b1	762,880	8ca6613631a033689c4f5b0e3a0b7b06	12288:lM732tz7fyC7TA+fBV+eoMNa9JvV2/KiGJmjA5IkyNaOaCq0ACGF9m:2732tzzyAT/JY3vT9IKRJmqV8faWpG
04/02/2019 1:00:24pm	Malware	35b4f04dd3df5defb3bd3040be880320	288,256	7bae0f81f57e10fd1249e9e2e448191f	3072:M9p+8WQZI1wLLr0B0wUYSYtRRox/pQJE1UxInIFtwSqWVOAg0FujMf2DuSaAY8/O:K+8W+AywWY3JE1UGYBOAOQvDgHsF
04/01/2019 5:01:01pm	Malware	b9af036686b5745bec438d2f7c9dc3d6	401,920	8ca6613631a033689c4f5b0e3a0b7b06	6144:gm1p0nsblWJDgaQACffBrZabX3LbNE6UAscSYOYdtQ5EUpnEI75gkpGR4HtsvqKm:/jblWJfpCXBriX3cAdk5EUpn7yQsvbm
03/29/2019 4:31:54am	Malware	6af1625b7f9a4b3fe934f52714a46afb	82,944	dd7918d309ef66eaf0dd80849153d3b1	1536:RoveeMxk4cUHmqrmTkxqomWsqOQIusWscd7HMmkRE4i:RovebBmTkExHlq7smkuD
03/29/2019 12:19:29am	Malware	4a0144c7436e3ff67cf2d935d82d1743	161,808		3072:1IWzdBu9+0ZZHmdGHQgracn2mc9DXDmUpErLNDQtR9Ghi:1IkjX0ZZHmd/Ujnzc9DXCx1cEhi
03/24/2019 1:32:13pm	Malware	3348fa9c806d261c9d50c6c8384be28d	17,400		384:tmto1VCMfAs6oxxMfSIHkhh6NX44X6Ujnw+3WhaqnqQzU:qo1UWvpXIq6GH+3JYw
03/22/2019 9:30:20am	Malware	6aa3115fa1f3adb8f0539e93d2cf21ca	111,616	d9a6501ae5d568852f285058b5855d32	1536:D+MkwqykWUVxihhe764UqD+pp/lfnYgCsWLPcdgQhDauBHTm+f8dGxLMm4kASA:D+loh464hKdntUeggauBK+fQm4kAr
03/22/2019 4:00:27am	Malware	00b31e20cad362df56d47cc7aa6c726b	111,616	d9a6501ae5d568852f285058b5855d32	1536:s+MkwqykWUVxihhe764UqD+pp/lfnYgCsWLPcdgQjDauBHTm+f8dGhLMJCASA:s+loh464hKdntUeg6auBK+fgJCAr
03/19/2019 7:30:22am	Malware	ddd161a6bb63ca46e8cb0663587920fe	111,616	d9a6501ae5d568852f285058b5855d32	1536:C+MkwqykWUVxihhe764UqD+pp/lfnYgCsWLPcdgQxDauBHTm+f8dG4LMvASA:C+loh464hKdntUeg0auBK+fjvAr
03/15/2019 1:30:11am	Malware	5bff246eb66eda703bcb7742f658ce36	111,616	d9a6501ae5d568852f285058b5855d32	1536:1+MkwqykCUVxihhe764UqD+pp/lfnYgCsWLPcdgQUDauBHTm+f8dGfLMRNPgSA:1+hoh464hKdntUegNauBK+f2Hgr
03/14/2019 6:49:26pm	Malware	a39f4768aaf0c30968da27c6c3f6f20c	17,394		384:tmto1VCMfAs6oxxMfSIHkhh6NX44X6Ujnw+3WhaqnqQzG:qo1UWvpXIq6GH+3JYC
03/12/2019 4:41:09am	Malware	9d4c0bd0b4025cbea9100db4863bc2ff	82,944	dd7918d309ef66eaf0dd80849153d3b1	1536:78jcdlPgtkW0YcTLbUncqowt8nquQIusWscdXHNm/U4i:Aj+PbUnJ7ulqXtm/UD
03/11/2019 12:00:53pm	Malware	2e645c7971436417a2a6b4774bce59df	111,616	d9a6501ae5d568852f285058b5855d32	1536:s+MkwqykCUVxihhe764UqD+pp/lfnYgCsWLPcdgQDDauBHTm+f8dGCLMeNZgSA:s+hoh464hKdntUegWauBK+fNe7gr
03/05/2019 3:03:19am	Malware	13324fd1aa2edc18113658c8a25df4c4	82,944	dd7918d309ef66eaf0dd80849153d3b1	1536:id0N8kTBEXI4xQnJs8lMWqaU++gTkQIusWscdAH6UJNH/JT4i:idwLpJs8lsBWklqAaUJlpD
03/04/2019 8:55:47pm	Malware	60190ab01b53602bcf1ffad71eb8f385	82,944	dd7918d309ef66eaf0dd80849153d3b1	1536:nMVrX6otvkIpkOKj2g2Kmjz0zqomtExeQ4SSsW+acd2HY3WrNh5/Tt4U:ncrq+g21jz0WxXVSl724mrpRl
03/04/2019 12:00:16am	Malware	00ac0d7337290b74bdd7f43ec4a67ddb	958,059		12288:qtWzZQrUnNp6V9BQpvkRzVMm2O0UyncAf5YW+i3Cnt4dMCA:qtWqwnNwV9iarMq9cvCuMH
03/01/2019 5:41:24pm	Malware	724e7d4f5dd3ab20adba850d9938f974	1,741,295		24576:1dJVdI1ReEhK2Oq5vHmVsGl6unbfXycsg5l:1ywEA2FMNagr
03/01/2019 1:37:04pm	Malware	a100c1f728153e19a5d2960a43939543	1,741,295		24576:1dJVdI1RWEhK2Oq5vHmVsGl6unbfXycsg5l:1ywEA2FMNagr
03/01/2019 7:00:34am	Malware	0deb78c1edae349f6ff0e9c3ff5bc561	111,616	d9a6501ae5d568852f285058b5855d32	1536:I+MkwqykWUVxihhe764UqD+pp/lfnYgCsWLPcdgQTDauBHTm+f8dG4LMw2ASA:I+loh464hKdntUegWauBK+fjHAr
03/01/2019 4:12:36am	Malware	da14eece6191551a31d37d1e96681cd1	178,406		3072:1d52pb77M230VwZtW2Y0N3y/FAsqbtCgLhNmIZ4l9471kGUUAEToFM:1dC7gNR2w/tSog91is1tTTX
03/01/2019 1:00:12am	Malware	dd8b36f8f967a26314820c35632fa0d0	111,616	d9a6501ae5d568852f285058b5855d32	1536:k+MkwqykCUVxihhe764UqD+pp/lfnYgCsWLPcdgQrDauBHTm+f8dGCLM1dgSA:k+hoh464hKdntUeguauBK+fN1dgr
02/28/2019 7:06:31pm	Malware	76289f02a0b31143d87d5e35839fb24a	178,403		3072:cd52pb77M230VwZtW2Y0N3y/FAsqbtC+O0UdTrqrjKzJ6:cdC7gNR2w/tSot1xq86
02/28/2019 3:57:13am	Malware	6aab0ff5cb3189505ca9aaf1d602ab28	84,992	dd7918d309ef66eaf0dd80849153d3b1	1536:p5R5mqCgWgIkI0fpeVLsUvyqom4jRqQQIusWscdrHjzt4i:SqC6usUvPxi1lqrDZD
02/28/2019 12:32:24am	Malware	2294e82b60adf51d1cc8e307c0d7c466	99,036		3072:/1m+AbPC7JMIb5vAGBpqngQ0HM2A6ciLzHUpk:/qxSVtLqN0H46zHUu
02/27/2019 3:36:45am	Malware	9020db586f060b0301039ca2b5afc05c	82,944	dd7918d309ef66eaf0dd80849153d3b1	1536:ErFUd8kjiHEVZBT9NEqaG8IoOg7QIusWscdAHE+0UKpiH4i:QF8T9Nx78IslqAk+bKQD

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’.

Vicky Ray @0xVK · 19 Nov 2018

While #ssdeep and #IAT is a great way to cluster and perform similarity analysis on large malware sample datasets, some other similarity analysis methods like #jaccardindex and #minhash does yield interesting results.
#reducereversingtime

1 2 5

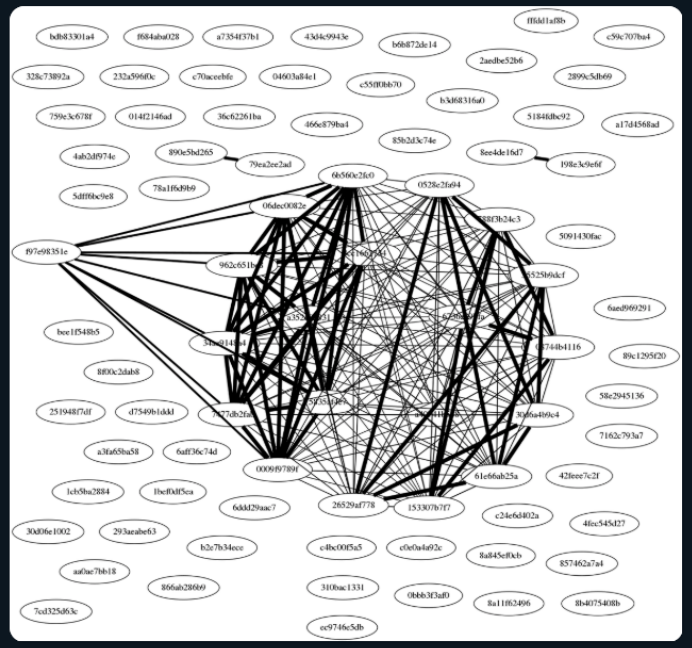
Show this thread

New family

Windshield

Vicky Ray @0xVK · 19 Nov 2018

Another method - String based similarity analysis of #badcake #malware samples using #jaccardindex. Graph created using #networkx.
#reducereversingtime #datascience

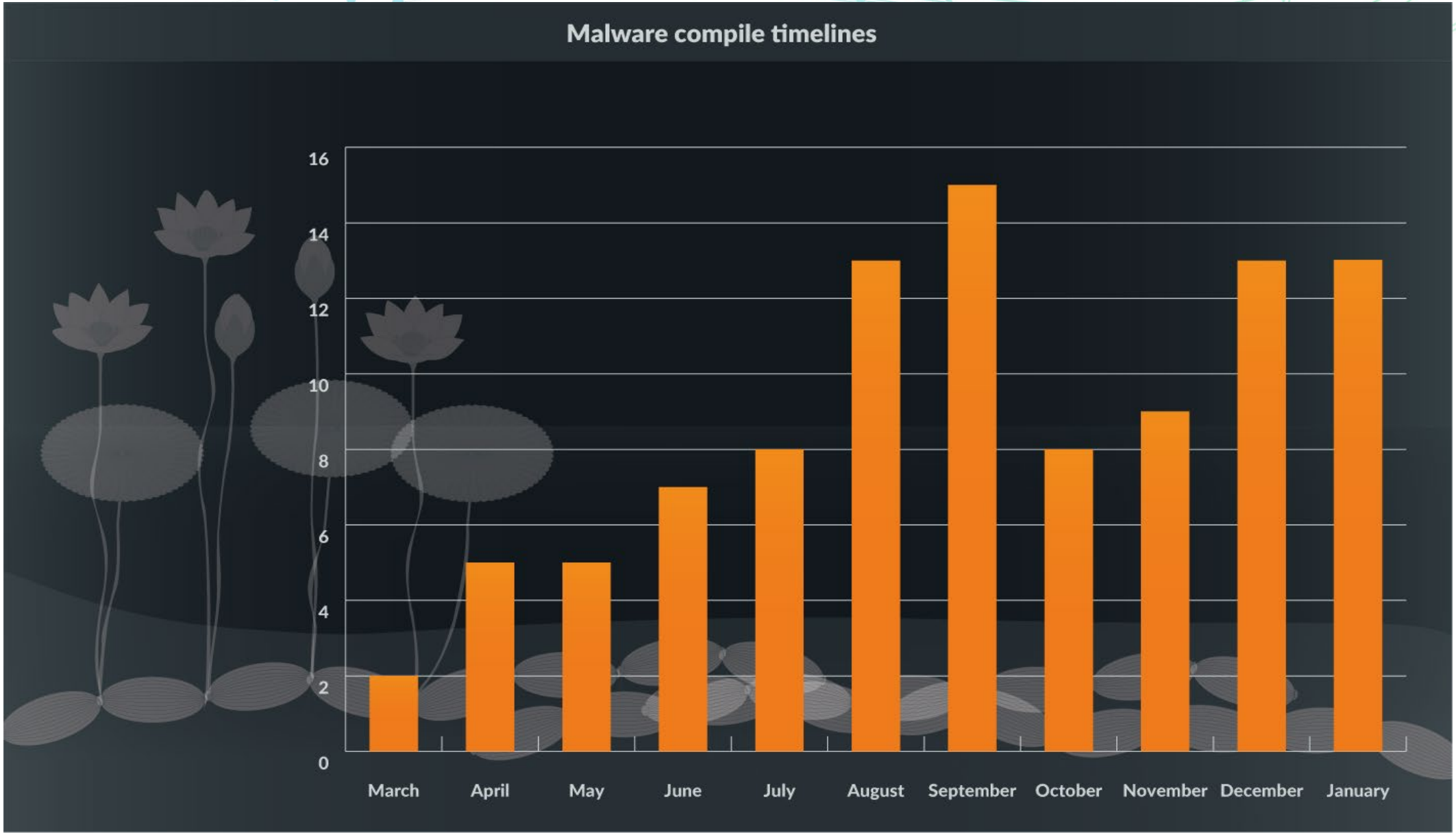


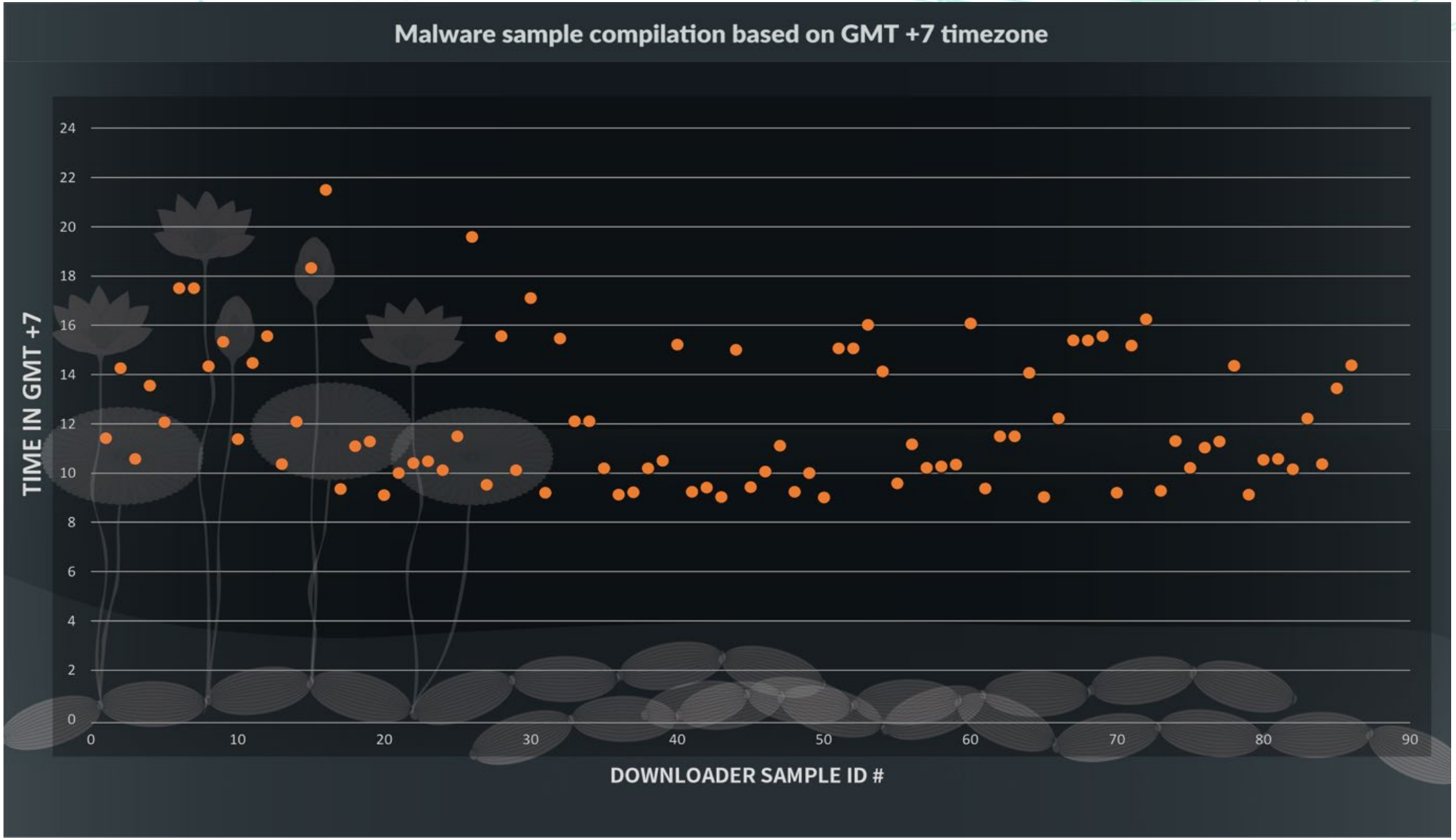
Vicky Ray @0xVK · 19 Nov 2018

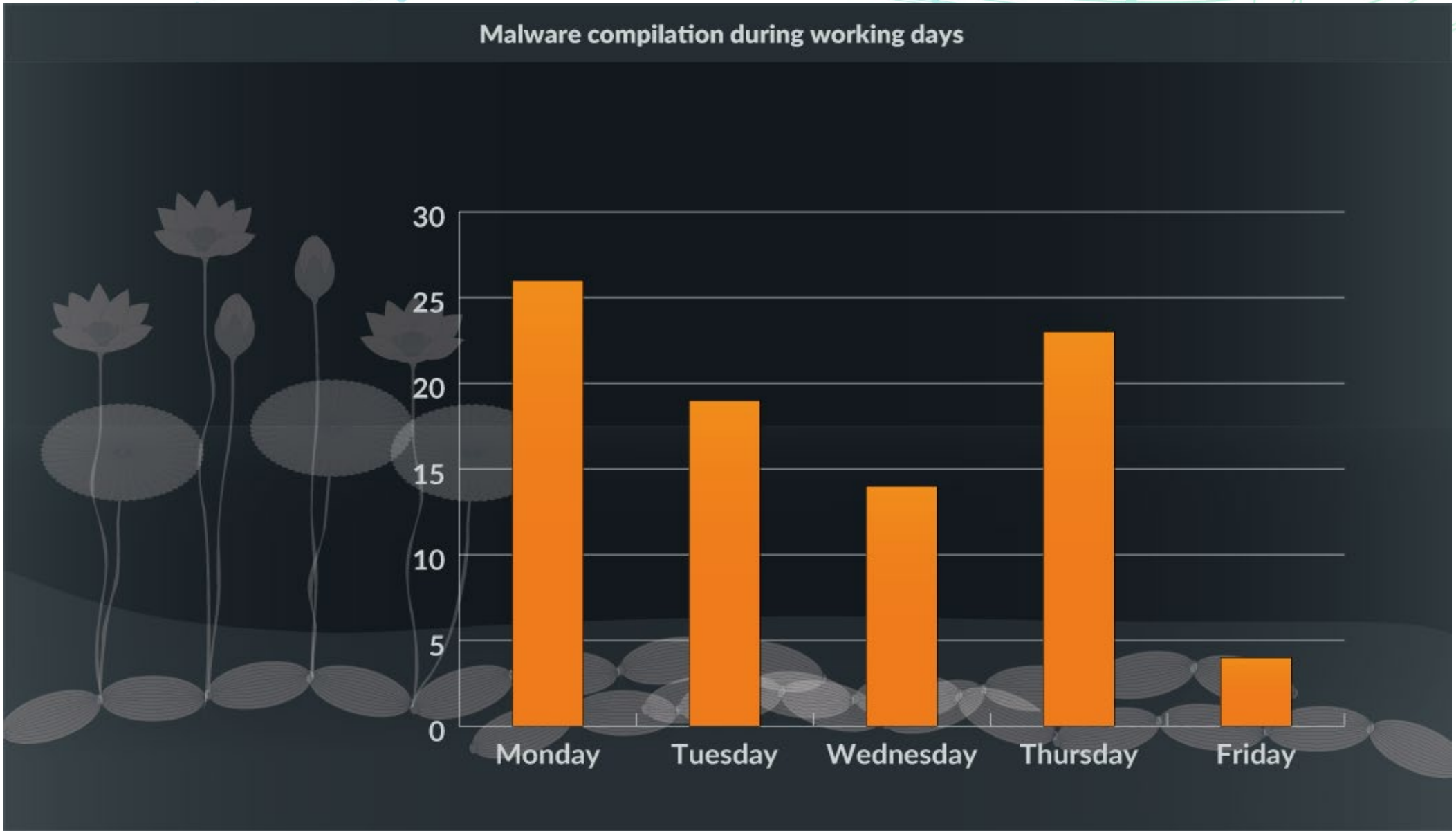
This is a similarity % of #badcake samples from #OceanLotus group using #minhash and #jaccardindex.
#reducereversingtime

```
Sample name or hash                                    Shared code estimate(1.0=100% 0.0=0%)
=============================================================================================
[*] 55525b9dcf342c7f323f941bd117bcc20b25f9a2728cb999532dbec011bfefbc.MLWR 1.0
[*] 61e66ab25afb6a3cfca8327d597151534584 0bdfdca33079e907c0657926827c.MLWR 0.99609375
[*] 26529af7782a902c04ae01898c8b14c9f01302165335858ad666b10532584254.MLWR 0.99609375
[*] 0528e2fa94f3b1253fe6c6a5345236456876725395 4630ab5cc141e41690ea43.MLWR 0.99609375
[*] 30d6a4b9c41225c22b3d1bf2f1eab3d1c57c8b1a69502eab076a4f97f14023ac.MLWR 0.99609375
[*] a352d0e831e5013e4f558fcd4a4daa390f5af3350e795f47fbc8ba50ae2300e6.MLWR 0.5703125
[*] 75835af4e772ead0e9faddd59328c44ab9a5b80f7df64f7d2ef18f94483c08de.MLWR 0.5703125
[*] 06dec0082eac094dc0b4b3de8854f190f1d3112dada0d414d9a085a0ee309199.MLWR 0.5703125
[*] 3cc166273476ebaf4d083e444914bdecf39a3faac5d0498008 59988b9c9c91b1.MLWR 0.5703125
[*] 962c651be81b447b3a7fcf2a58752bfac005f35157fa017a0a2a175ea4e2439e.MLWR 0.5703125
[*] 34ae9148a4db9993110e4fe4a0f8e9db17790b036ea0f5c236f53cbf845dd2a3.MLWR 0.5703125
[*] 0009f9789f0b3fd20e9a2c48ab36bbca322cdf050fc8d3ebe7e12b470a0e4551.MLWR 0.5703125
[*] 7477db2fab4dc77213008682e3302d6dd30e3963885f0a156d14bd067fa5b5cc.MLWR 0.5703125
[*] 6b560e2fc0be10d0ffd9e5440101f083ed7f5328735df79fd6c537c61bfcfe88.MLWR 0.5703125
```

1 IMG


Malware compile timelines
16
14
12
10
8
6
4
2
0
March
April
May
June
July
August
September
October
November
December
January

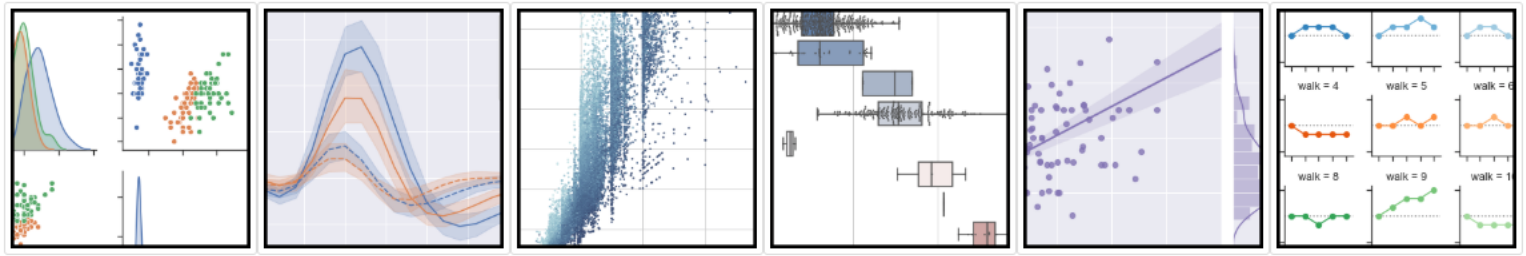

Malware sample compilation based on GMT +7 timezone
TIME IN GMT +7
0
2
4
6
8
10
12
14
16
18
20
22
24
0
10
20
30
40
50
60
70
80
90
DOWNLOADER SAMPLE ID #

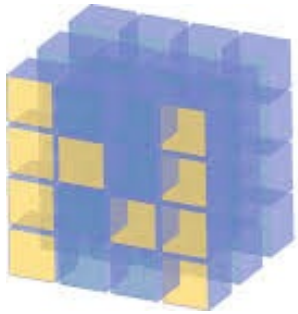

Malware compilation during working days
30
25
20
15
10
5
0
Monday
Tuesday
Wednesday
Thursday
Friday

Data Visualization in DataScience

matplotlib

seaborn: statistical data visualization

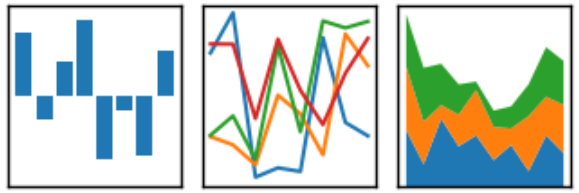




NumPy

pandas

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



Movies


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

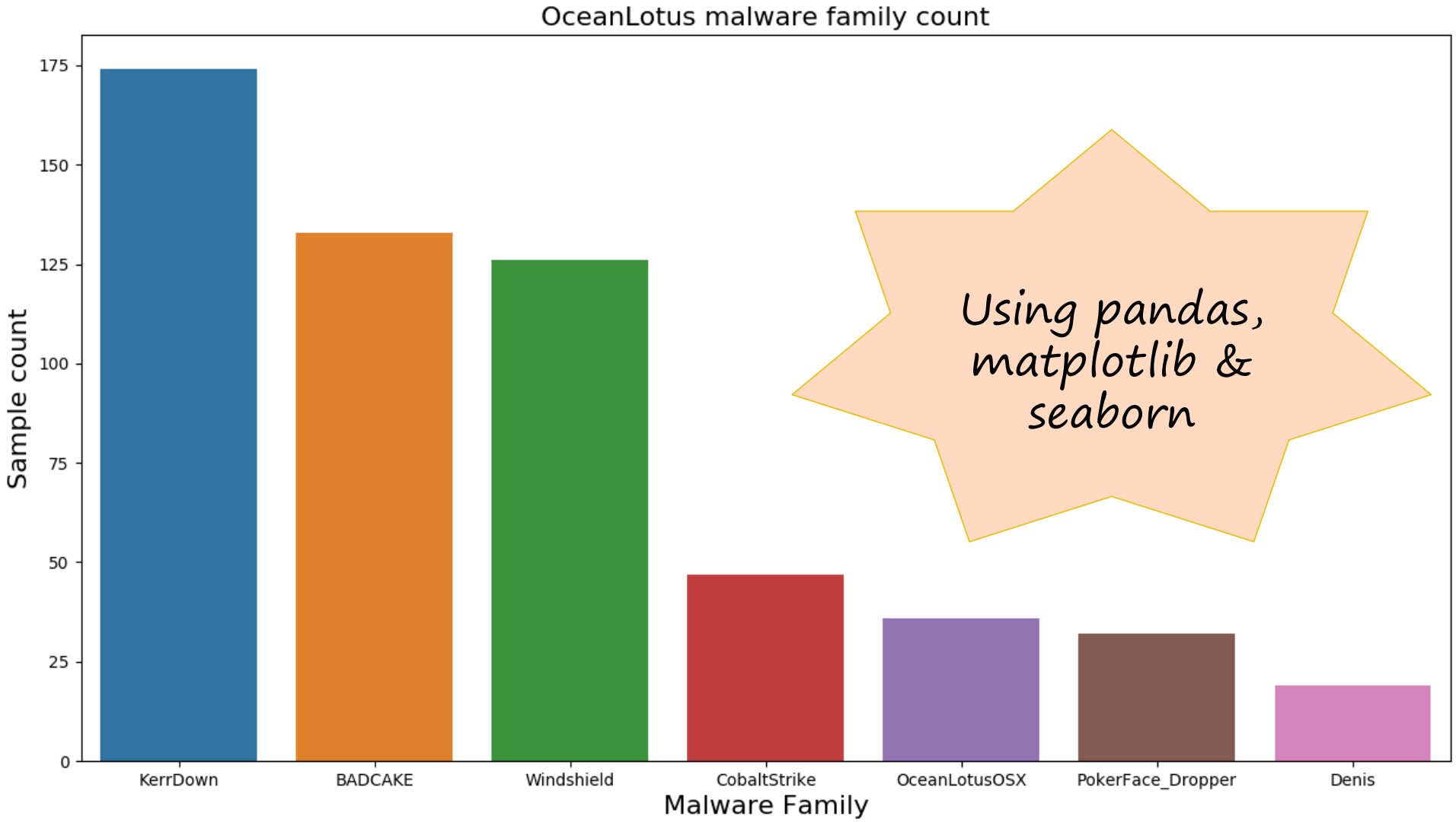
Data from Autofocus Lenz (Reality)

```
:autofocus-lenz $ 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_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,Unit42.WindShield
381d52aa6df750f23c5b53d2fcb563002a5a5ea6c32ad3367338c50774e38c18 | PE                                | 2015-07-31 09:15:19 | malware | 1250816 | Unit42.RenameOnReboot,Unit42.WindShield
f8f58963f6fd9e9c08f5527f176cfc5fe379594a5cb57096e5180390266da968 | PE                                | 2015-08-11 11:43:24 | malware | 473600  | Unit42.RenameOnReboot,Unit42.WindShield
7afbfc7fdce80755557c8753d345eed277db8a7597041571653fbd7967f26ca9a | 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
```

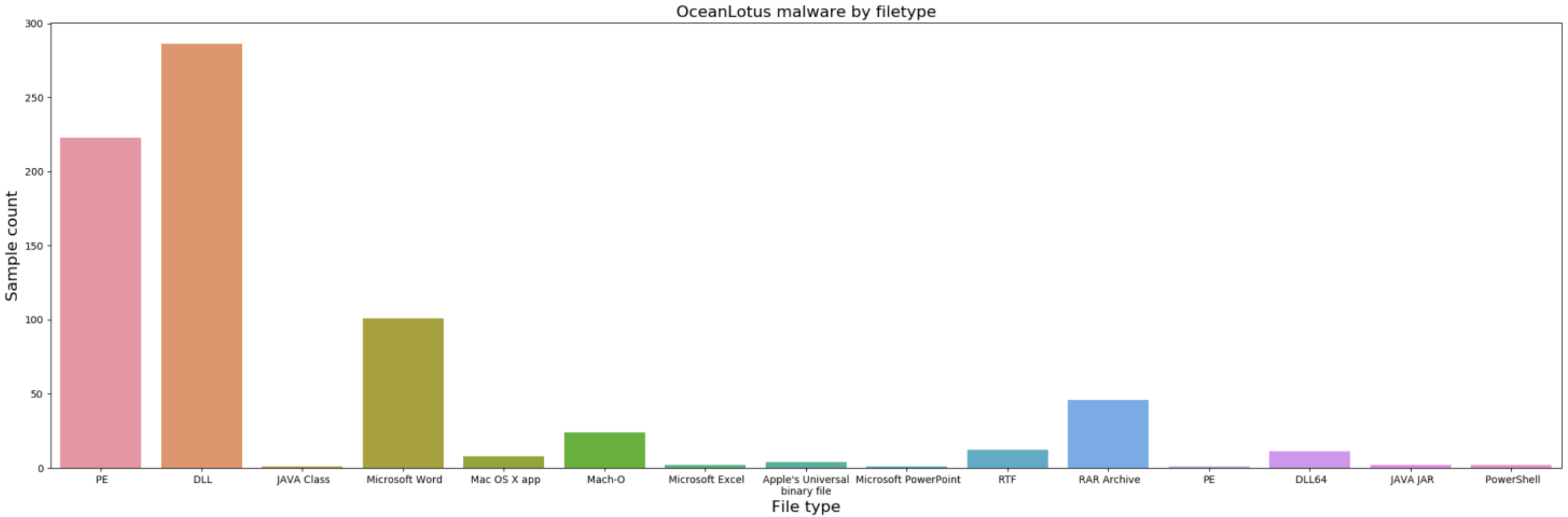

OceanLotus malware family count
Sample count
175
150
125
100
75
50
25
0
KerrDown
BADCAKE
Windshield
CobaltStrike
OceanLotusOSX
PokerFace_Dropper
Denis
Malware Family
Using pandas, matplotlib & seaborn

Sample python code

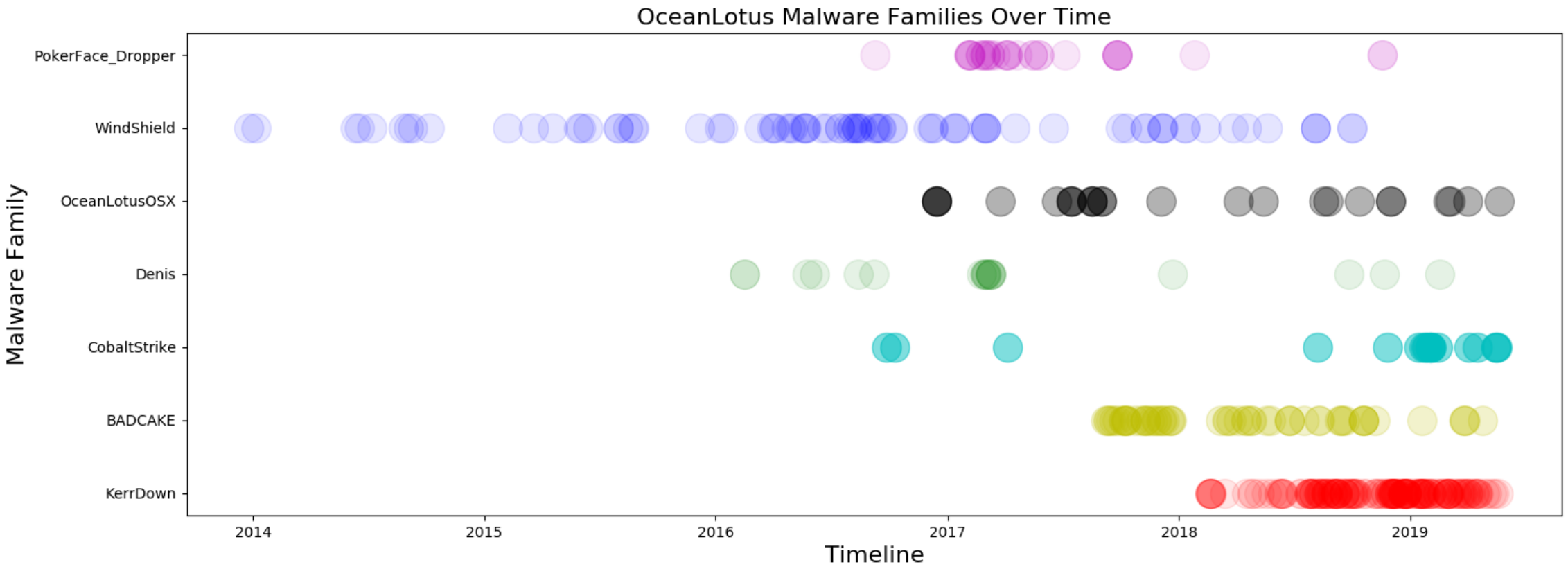
```
import pandas
from matplotlib import pyplot
import seaborn
pyplot.rcParams["figure.figsize"] = (15,8)

mal = pandas.read_csv("tags_only.csv")
seaborn.countplot(x='tag', data=mal)
pyplot.title("OceanLotus malware family count", fontsize = 16)
pyplot.xlabel("Malware Family", fontsize = 16)
pyplot.ylabel("Sample count", fontsize = 16)
pyplot.savefig("histogram_malware.png")
```

Stats by filetype



OceanLotus malware family timelines using Matplotlib & pandas



Sample python code

```
import pandas
import dateutil
from matplotlib import pyplot
pyplot.rcParams["figure.figsize"] = (17,6)

stats = pandas.read_csv("malware_list_oceanlotus.csv")
stats['fs_date'] = [dateutil.parser.parse(d) for d in stats['timestamp']]

kerrdown = stats[stats['tag'] == 'KerrDown']
badcake = stats[stats['tag'] == 'BADCAKE']
cobaltstrike = stats[stats['tag'] == 'CobaltStrike']
denis = stats[stats['tag'] == 'Denis']
komprogo = stats[stats['tag'] == 'Komprogo']
oceanosx = stats[stats['tag'] == 'OceanLotusOSX']
windshield = stats[stats['tag'] == 'WindShield']
pokerdrop = stats[stats['tag'] == 'PokerFace_Dropper']


pyplot.plot(kerrdown['fs_date'], kerrdown['tag'], 'ro', label="KerrDown", markersize=20, alpha=0.1)
pyplot.plot(badcake['fs_date'], badcake['tag'], 'yo', label="BADCAKE", markersize=20, alpha=0.2)
pyplot.plot(cobaltstrike['fs_date'], cobaltstrike['tag'], 'co', label="CobaltStrike", markersize=20, alpha=0.5)
pyplot.plot(denis['fs_date'], denis['tag'], 'go', label="Denis", markersize=20, alpha=0.1)
pyplot.plot(komprogo['fs_date'], komprogo['tag'], 'yo', label="Komprogo", markersize=20, alpha=0.1)
pyplot.plot(oceanosx['fs_date'], oceanosx['tag'], 'ko', label="OceanLotusOSX", markersize=20, alpha=0.3)
pyplot.plot(windshield['fs_date'], windshield['tag'], 'bo', label="WindShield", markersize=20, alpha=0.1)
pyplot.plot(pokerdrop['fs_date'], pokerdrop['tag'], 'mo', label="PokerFace_Dropper", markersize=20, alpha=0.1)


pyplot.title("OceanLotus Malware Families Over Time", fontsize = 16)
pyplot.xlabel("Timeline", fontsize = 16)
pyplot.ylabel("Malware Family", fontsize = 16)
pyplot.savefig("OceanLotus_malware.png")
```

Few lines of simple python code

Learnings & future work

- Datascience 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?

THANK YOU!

Twitter: 0xVK

https://unit42.paloaltonetworks.com/tracking-oceanlotus-new-downloader-kerrdown/
https://unit42.paloaltonetworks.com/