Hunting C2 Beaconing at Scale in the Modern Age

Mehmet Ergene
Security Researcher & Data Scientist, Binalyze

About me

Mehmet Ergene

Security Researcher & Data Scientist, Binalyze

Handpan player

Lindy hopper

@Cyb3rMonk

https://github.com/Cyb3r-Monk

https://posts.bluraven.io

Agenda

- Current C2 Beaconing Hunting Process
- C2 Usage in Modern Attacks
- The Experiment
- Solution & Jupyter Notebook Release
- Q&A

Current C2 Beaconing Hunting Process

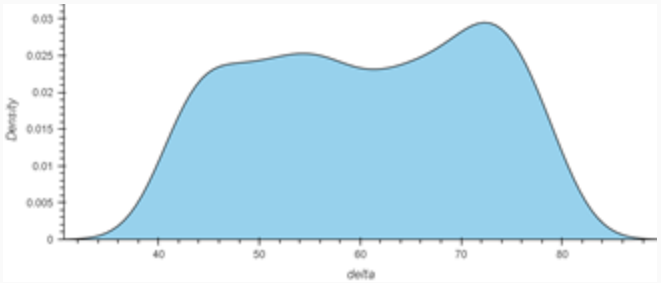
Current C2 Beaconing Hunting Process

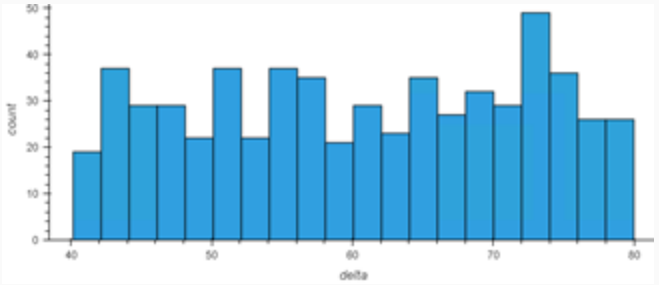
Proxy/Bro/Zeek Logs

Timestamp	SourceIP	DestinationHostName	SentBytes	ReceivedBytes
7/14/2023, 1:06:58.916 AM	10.35.20.61	twitter.com	2000	360
7/14/2023, 1:08:06.920 AM	10.35.20.61	kas2kjah13eas.cloudfront.net	800	430
7/14/2023, 1:09:13.932 AM	10.35.20.61	kas2kjah13eas.cloudfront.net	810	410
7/14/2023, 1:09:15.109 AM	172.18.5.30	www.amazon.com	600	5000
7/14/2023, 1:10:10.909 AM	10.35.20.61	kas2kjah13eas.cloudfront.net	815	425
7/14/2023, 1:11:02.921 AM	10.35.20.61	kas2kjah13eas.cloudfront.net	820	20000
7/14/2023, 1:11:30.409 AM	172.18.5.30	www.yahoo.com	700	3000
7/14/2023, 1:12:15.921 AM	10.35.20.61	kas2kjah13eas.cloudfront.net	800	40000

Current C2 Beaconing Hunting Process

- For each source-destination pair (Source IP/User - Destination IP/Host):
  - Generate list of connection intervals(time delta) and data size(packet size)
    time_delta = [0,0,14,16,25,30,30,25,15,13,22,60,68,10,100,150]
    packet_size = [600,610,600,605,680,700,760,900,20000,15000,600,640,620,250000,630,625]
  - Analyze time delta distribution
  - Analyze data size distribution
  - If both distributions are uniform and narrow, it's more likely a beaconing traffic
    - If false positive, whitelist the IP or Hostname





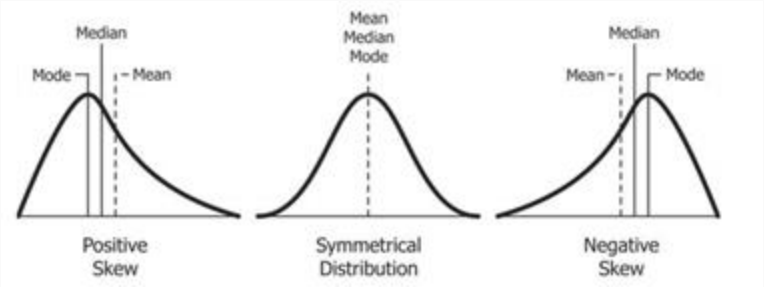
Uniform time delta distribution of a beacon with
60s sleep and 33% jitter

Data Size & Time Delta Distribution Analysis

- Percentile
  - Percentage of values below a specified point
    25th percentile (p_25) = the value x where 25% of the values are below x
- Median
  - Middle number of an ordered list
    [1,1,2,**2**,4,6,9] : 2
  - 50th percentile = Median
- Median Absolute Deviation(MAD)
  - Measurement of how wide or narrow the distribution is
    - Find median : 2
    - Calculate the absolute distance between the median and each item
      [1,1,0,0,2,4,7] –sort→ [0,0,1,**1**,2,4,7] (absolute deviation list)
    - Find the median of the absolute deviation list: 1

Data Size & Time Delta Distribution Analysis

- Mean
  - The average of a data set
- Mode
  - The value(s) that appears most frequently in a data set
    [1,1,2,2,4,6,9] ⇒ 1, 2
  - Doesn't have to exist in every data set
- Skewness
  - Asymmetry of a data distribution
  - Bowley's Formula: (p_25 + p_75 - p_50 * 2) / (p_75 - p_25) = x (-1 < x < 1)



1 > x > 0 x = 0 -1 < x < 0

Data Size & Time Delta Distribution Analysis

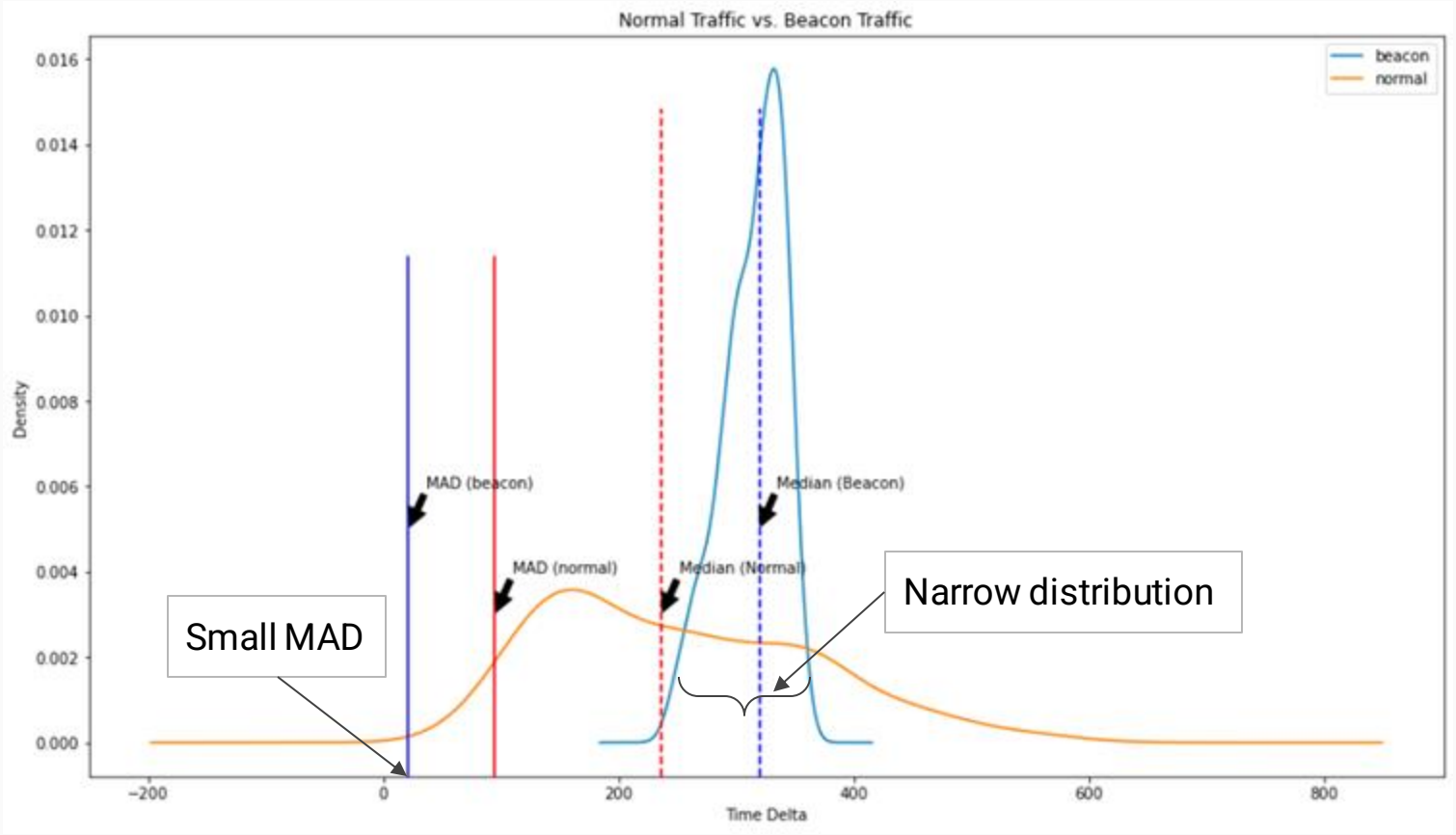
- Analyze skewness ⇒ skewness_score
  - Less skewed ⇒ higher score
- Analyze dispersion ⇒ mad_score
  - Dispersion is about MAD
  - Small dispersion = narrow distribution ⇒ higher score

Score = (skewness_score + mad_score) / 2

Score > 0.85 ⇒ BEACONING!

RITA
(Real Intelligence Threat Analytics)

Data Size & Time Delta Distribution Analysis



C2 Usage in Modern Attacks

Domain Fronting

- Attacker host : myevilc2[.]com
- Fronted domain: jkahsfkjah13eas.cloudfront.net
  - **Globally unique** hostname given by the provider or set by the attacker
- How the traffic looks in the logs

Source	Destination Host	Destination IP
Victim User/IP	kasfkjah13eas.cloudfront.net	Multiple Cloudfront IPs

- Can’t block/whitelist the host when it’s a False Positive
  - It belongs to a Cloud Service Provider
  - Hard to maintain

Web Services

- Attacker host: ??
- Web Service : graph.microsoft.com
- How the traffic looks in the logs

Source	Destination Host	Destination IP
Victim User/IP	graph.microsoft.com	MS Graph IP

- Can’t block/whitelist the host when it’s a False Positive
  - It belongs to a SaaS or Cloud Service Provider

Malleable C2 Profiles

- Attacker host: ??
- Host header : www.amazon.com
- Attacker IP : An arbitrary IP that doesn't belong to www.amazon.com
- How the traffic looks in the logs

Source	Destination Host	Destination IP
Victim User/IP	www.amazon.com	5[.]4.23.34

- Can't block/whitelist the host when it's a False Positive
  - It's a benign hostname

SOCKS Tunneling

- Tunneling Post-Ex tool traffic (Evil WinRM, etc.) to the target network
  - No need to drop the malicious files on target
- HTTP/1.1 (Cobalt Strike SOCKS proxy)
  - Requires small sleep parameter (sleep 0) to function effectively
- HTTP/2 (gTunnel, etc.)
  - Single HTTP connection (no beaconing behavior)
  - Most likely(?) blocked in enterprise environments
- SSH
  - Most likely blocked in enterprise environments

★ Tunneling can be done over the C2 channel or a different channel

In-Memory Execution

- Transfer the command/code over the C2 to the Beacon
- Execute the code in the Beacon or sacrificial process memory
- Beacon sends the results back over the C2

The Experiment

Scenario

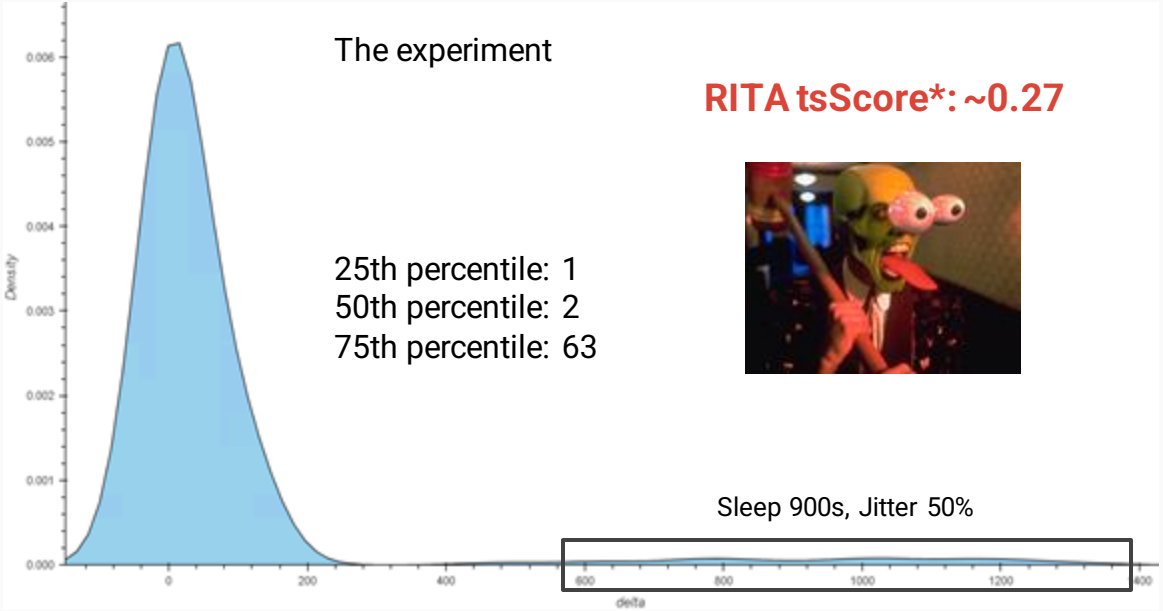
Beacon Usage

Phase	Duration	Sleep	Jitter
SOCKS Tunneling*	30 min	2s	50%
Keyboard Activity	450 min	90s	50%
No Activity(Idle)	960 min	900s	50%

State	Data Size(Bytes)	Jitter(Bytes)
Idle (just checking in)	800	500
Keyboard Activity (commands/tools)	20000, 40000, …	500

*HTTP/1.1 SOCKS Tunneling

Time Delta Distribution

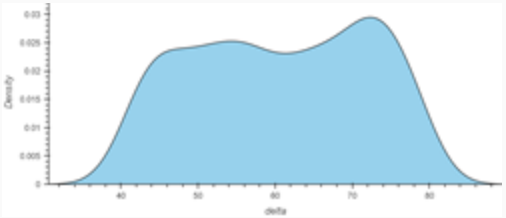


The experiment

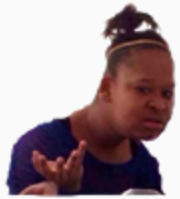
25th percentile: 1
50th percentile: 2
75th percentile: 63

RITA tsScore*: ~0.27

Sleep 900s, Jitter 50%



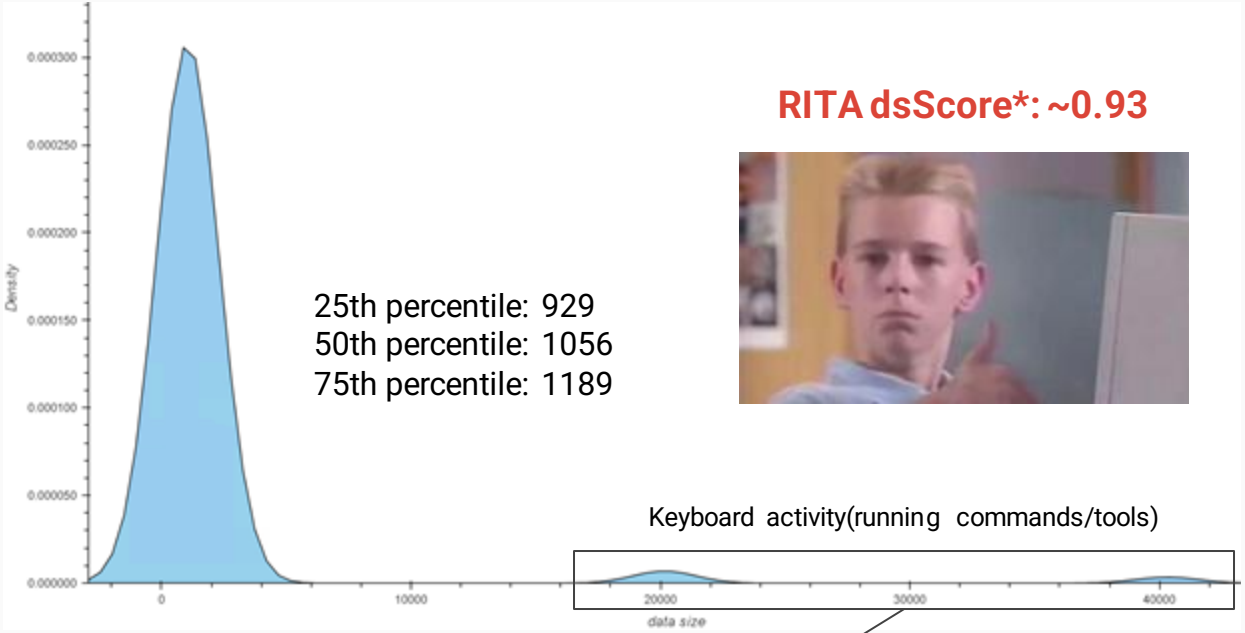
Uniform distribution we're looking for



Doesn't look like a uniform distribution?

*Time delta distribution score

Data Size Distribution



25th percentile: 929
50th percentile: 1056
75th percentile: 1189

RITA dsScore*: ~0.93

Keyboard activity(running commands/tools)

Falls into the 75% - 100% percentile, doesn't impact the scoring

*Data size distribution score

Detected Beaconing in an Enterprise

#	Source	Destination	Destination Prevalence	Score	Result
1	src_01	ah3s32ds.cloudfront.net	2	0.95	FP
2	src_02	dst_01	4	0.94	FP
3	src_03	music.youtube.com	5	0.90	FP
4	src_04	<xyz>.amazon.com	3	0.89	FP
…					FP
150	src_130	dst_130	9	0.81	FP
…					
240	**src_240**	**www.amazon.com**	**105**	**0.77**	**TP**

Solution

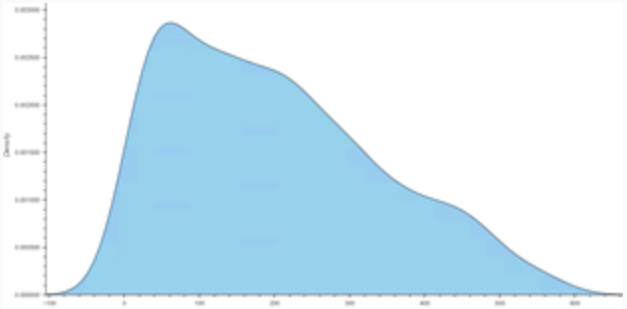
Time Delta Analysis

- Use 15th, 30th, and 45th percentiles
  - Large sleep values doesn't impact the scoring
- Use jitter as a scoring parameter

30th percentile : 58s
MAD : 30s
Jitter : (30/58) * 100 = 52 (52%)

Jitter <= 55 ⇒ Jitter score = 1
Jitter > 55 ⇒ Jitter score = 1 - (Jitter * 0.004)

- Do **NOT** use skewness for scoring
  - Jitter in C2 doesn't guarantee uniform distribution



Beaconing with skewed time delta distribution

Data Size Analysis

- Use 15th, 30th, and 45th percentiles of the Sent Bytes
  - Sending command/code over C2 doesn't impact the score
- Use jitter as a scoring parameter
- Do **NOT** use skewness for scoring
- At least 1 connection must have Received Bytes > 20.000
  - Beacon receives the command/code to execute
  - Adjustable as a threshold (do your own risk analysis!)
    - Nation State TA : 1
    - Ransomware TA : 5-10

Active C&C Detector

https://github.com/Cyb3r-Monk/ACCD

AC&CD Performance

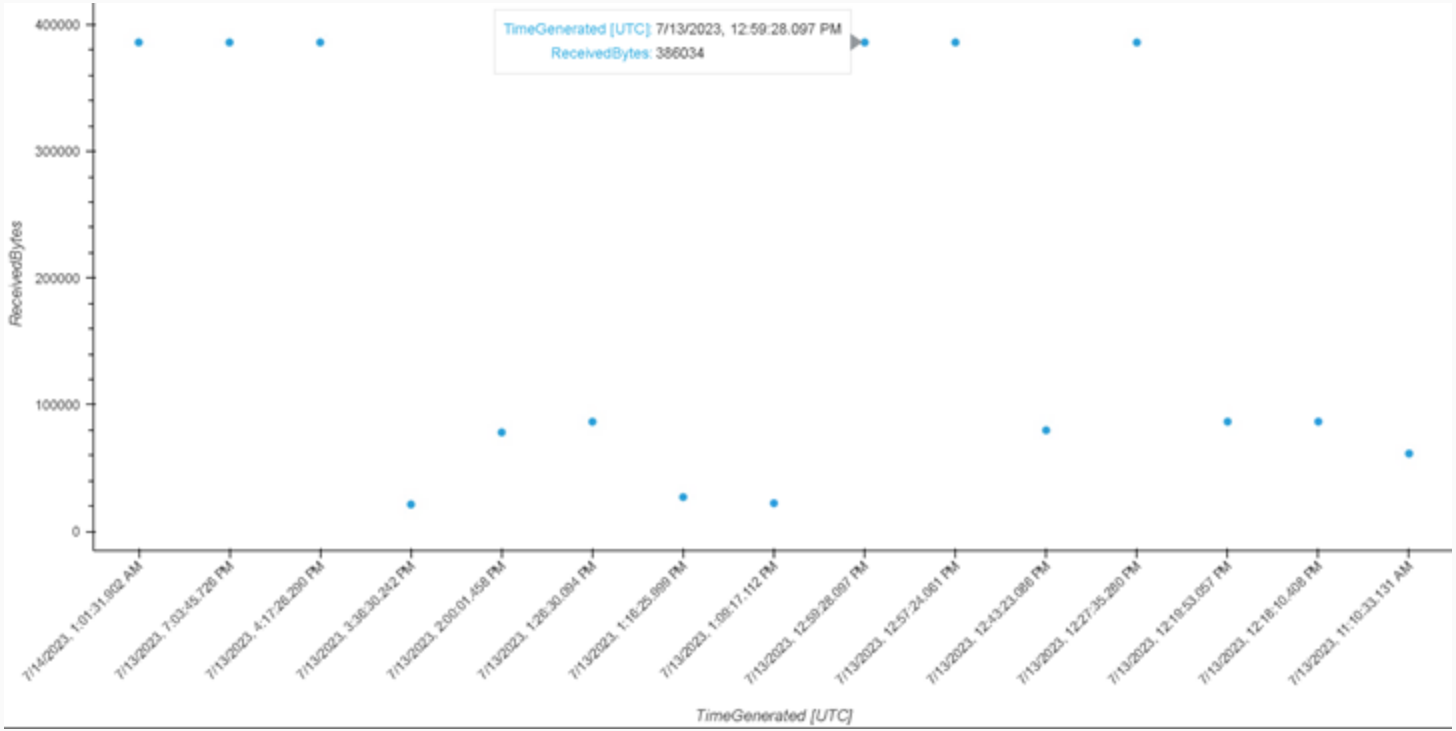
#	Source	Destination	Destination Prevalence	Score	Result
1	**src_240**	**www.amazon.com**	**105**	**1.00**	**TP**
2	src_05	dst_08	3	1.00	FP
3	src_02	dst_01	4	0.94	FP
…					FP
150	src_130	dst_130	9	0.81	FP
…					

Future Work

- Implementation in MSTICPy
- Working on a ML algorithm

DFIR Bonus

- Time Series graph shows when the attacker executed commands



Q&A