OWASP™

Chapter talks, Billund

# Detecting malicious beaconing in enterprise environments

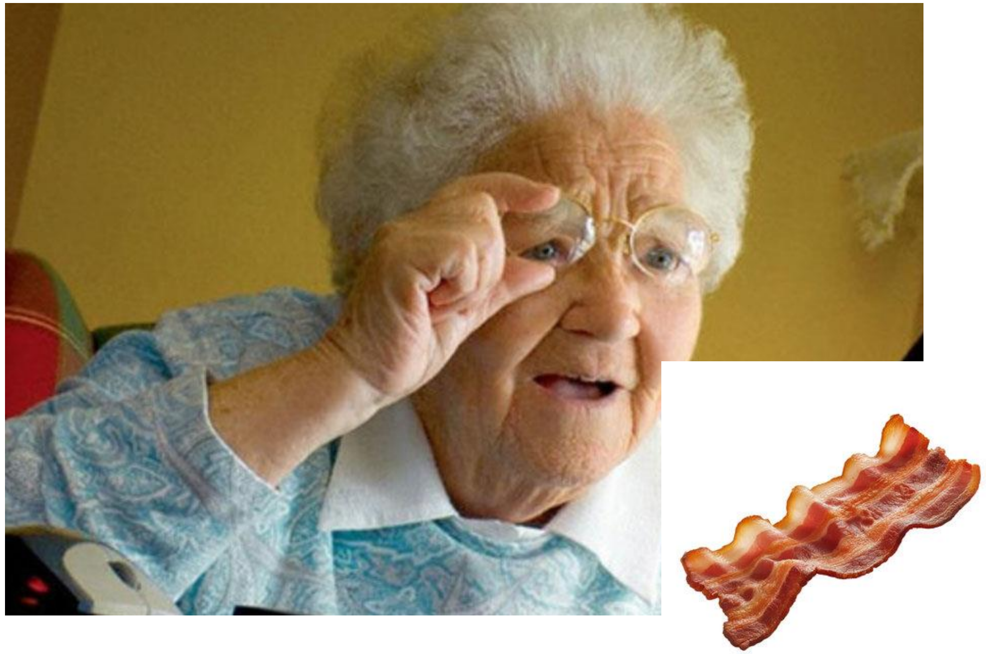
Dr. Daniel Varga

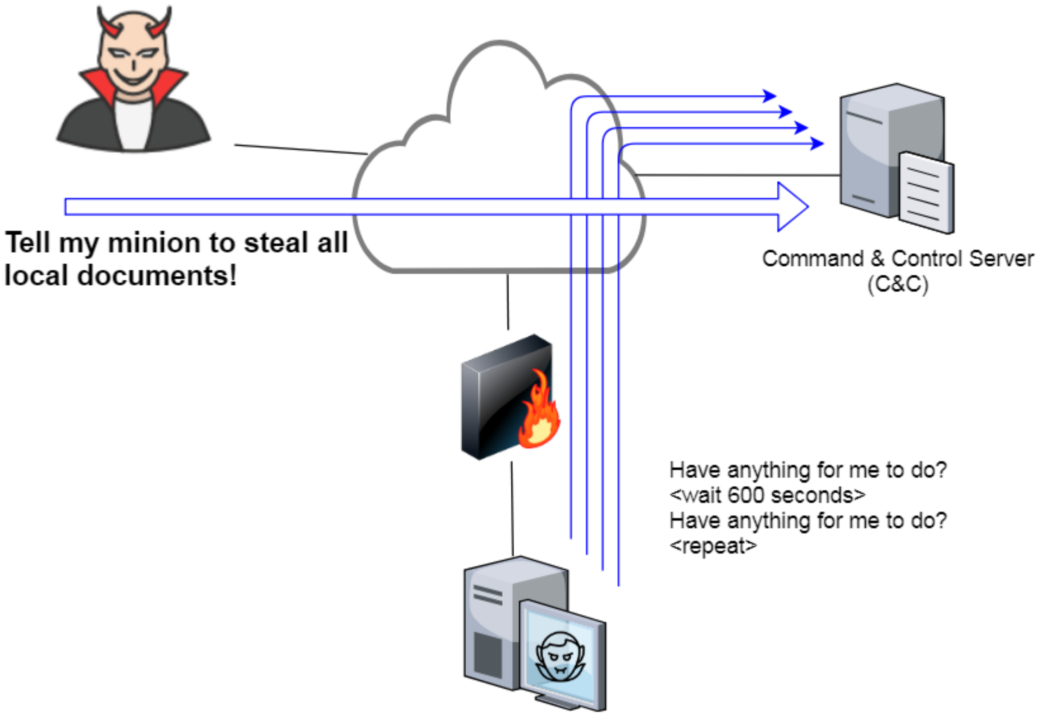
Associate Cyber Security Specialist

LEGO

daniel.varga@lego.com

# What’s a beacon?



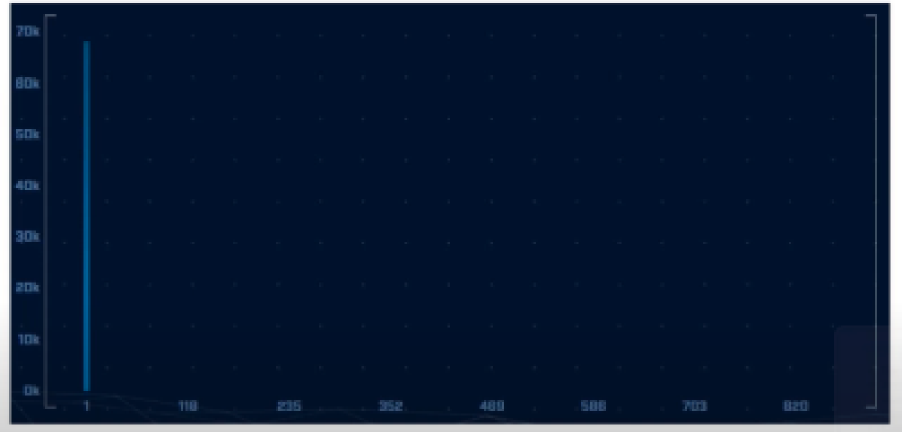


Tell my minion to steal all local documents!

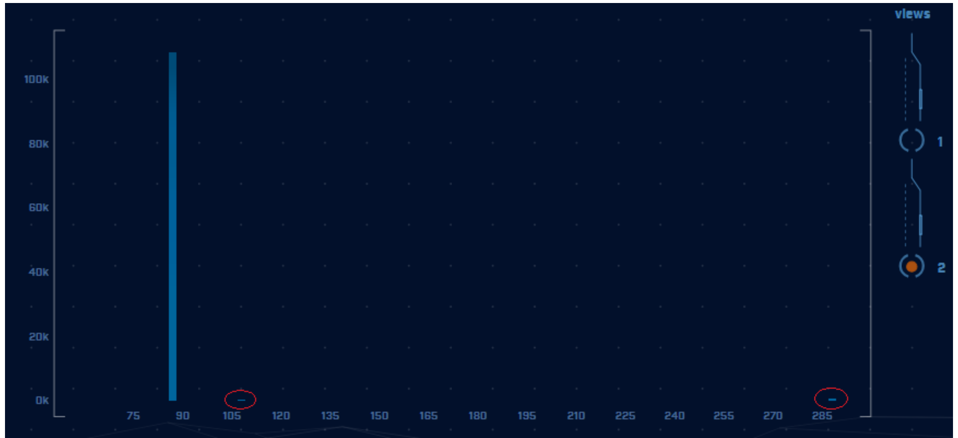
Command & Control Server (C&C)

Have anything for me to do?
<wait 600 seconds>
Have anything for me to do?
<repeat>

Time delta distribution



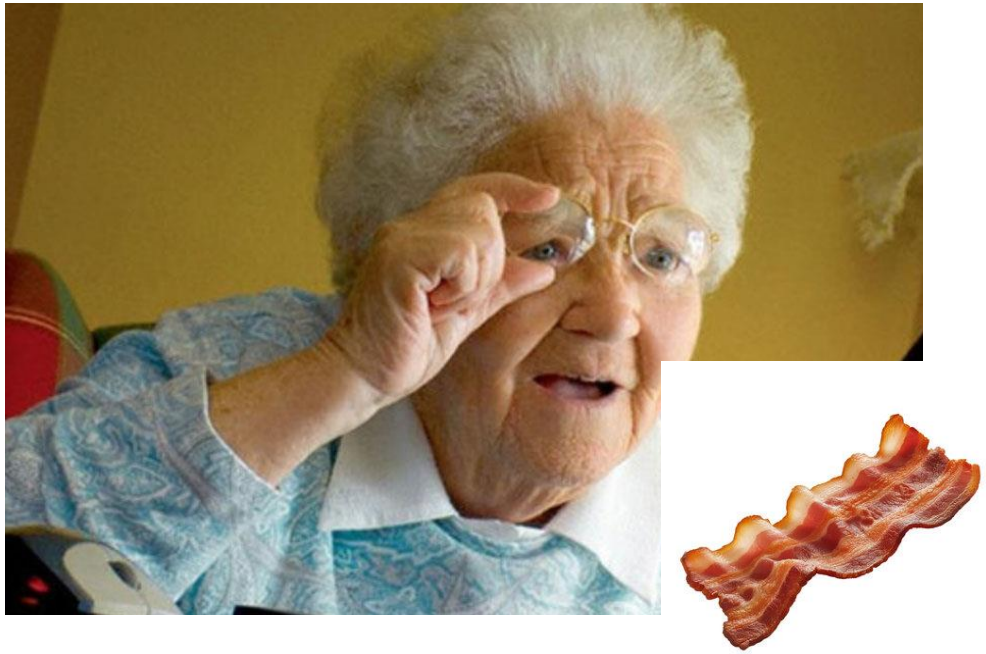
Data size distribution

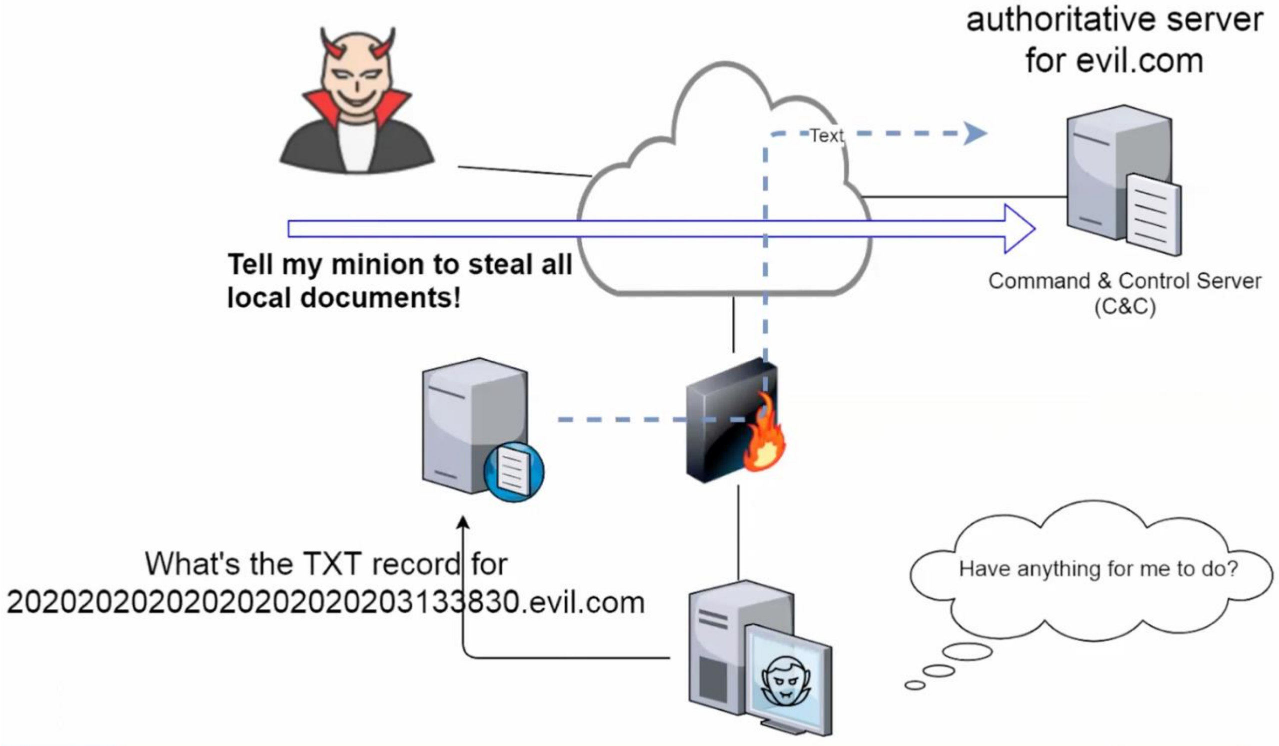


Ref:
https://www.activecountermeasures.com/identifying-beacons-through-session-size-analysis/

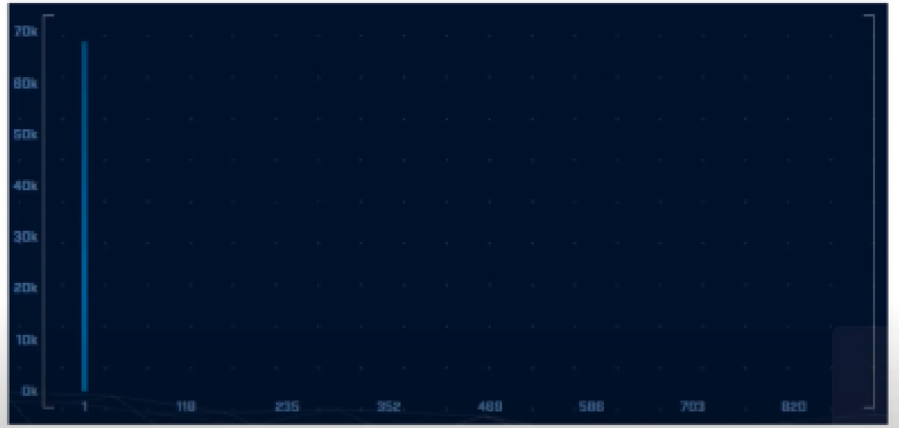
https://www.scworld.com/podcast-segment/4295-beacon-analysis-chris-brenton

# What’s a beacon?

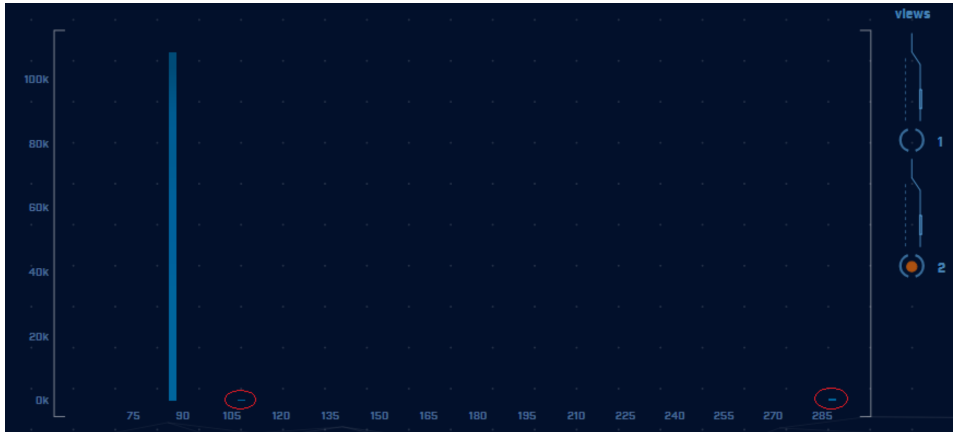



authoritative server for evil.com
Text
Tell my minion to steal all local documents!
Command & Control Server (C&C)
What's the TXT record for 20202020202020202020203133830.evil.com
Have anything for me to do?

Time delta distribution



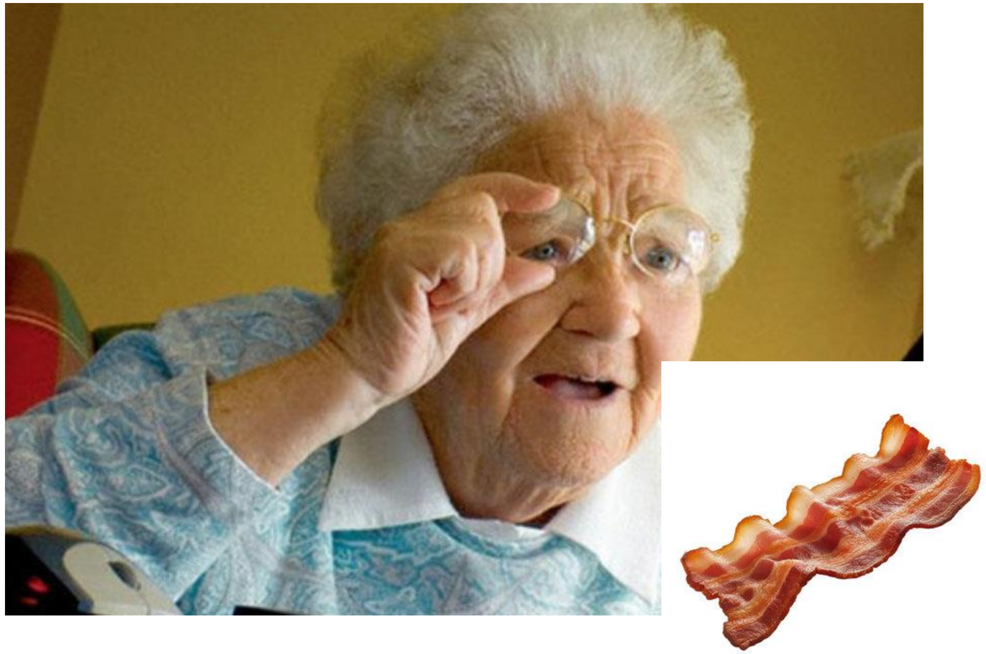
Data size distribution

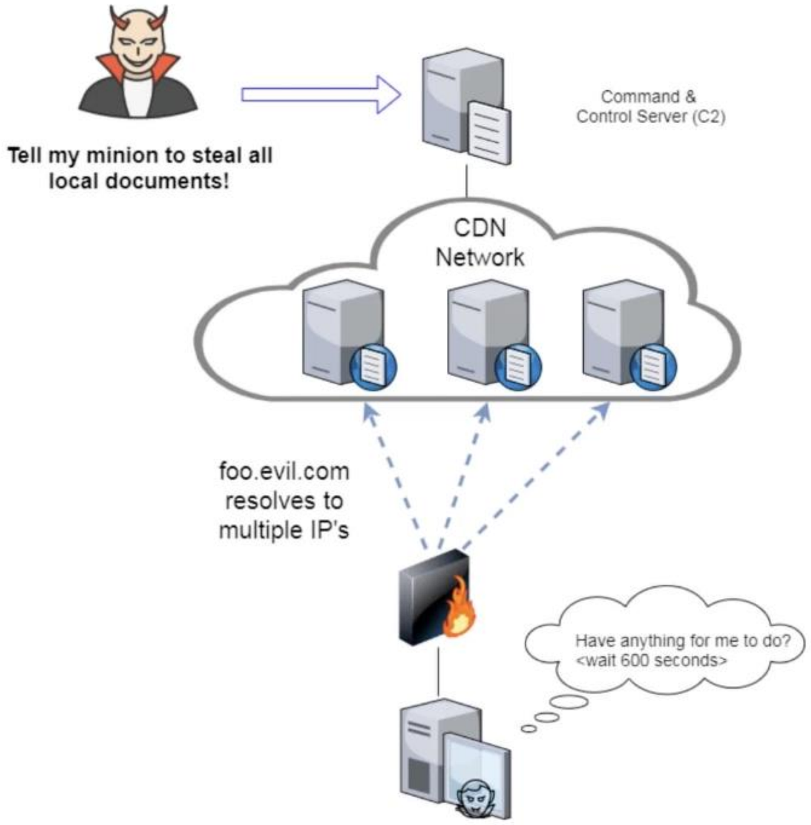


Ref:
https://www.activecountermeasures.com/identifying-beacons-through-session-size-analysis/

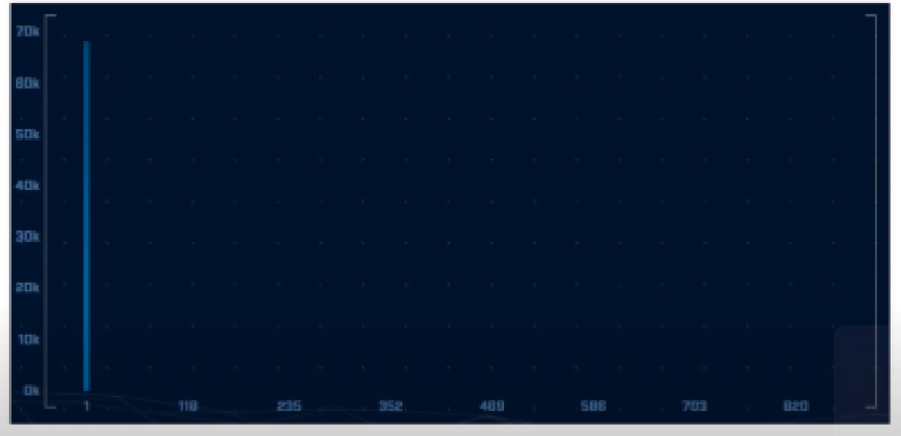
https://www.scworld.com/podcast-segment/4295-beacon-analysis-chris-brenton

# What’s a beacon?

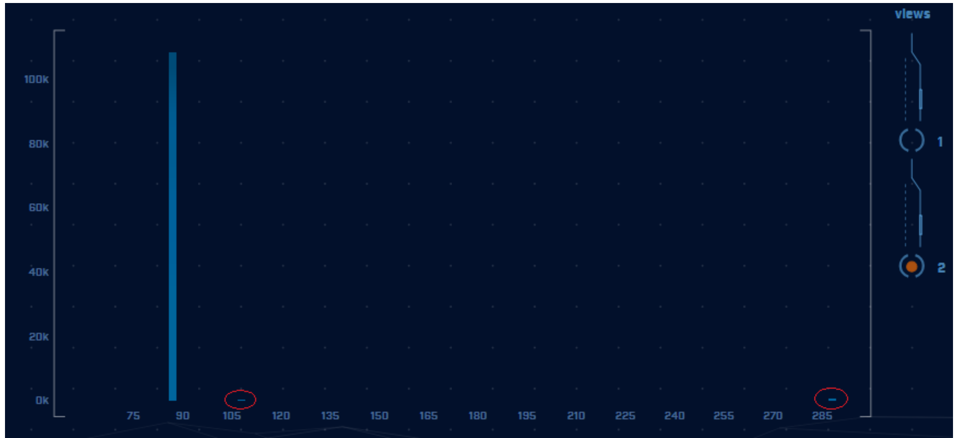



Tell my minion to steal all local documents!
Command & Control Server (C2)
CDN Network
foo.evil.com resolves to multiple IP's
Have anything for me to do? <wait 600 seconds>

Time delta distribution



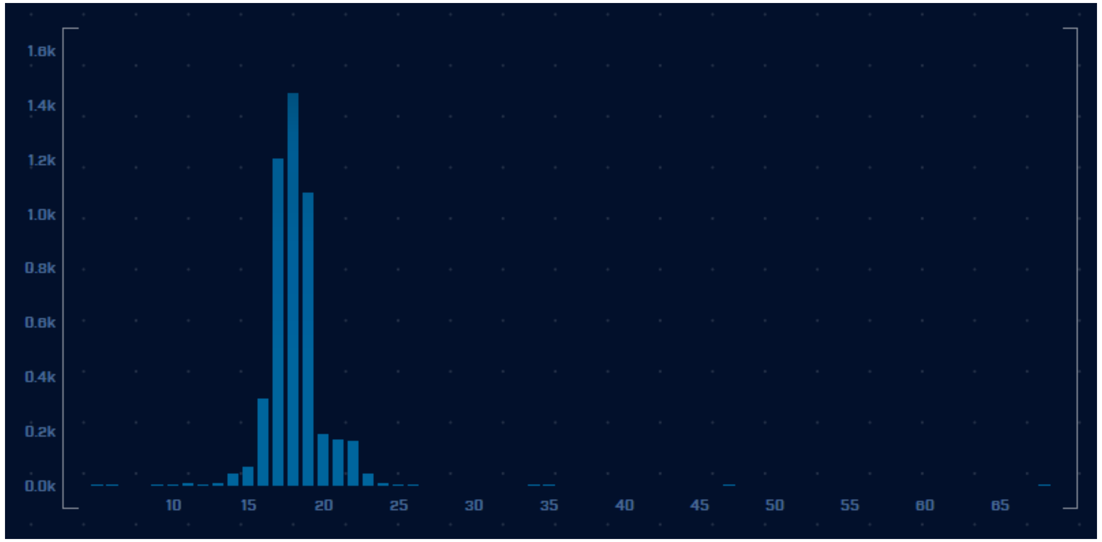
Data size distribution



Ref:
https://www.activecountermeasures.com/identifying-beacons-through-session-size-analysis/

https://www.scworld.com/podcast-segment/4295-beacon-analysis-chris-brenton

# Jitter



Ref: https://www.activecountermeasures.com/detecting-beacons-with-jitter/

# RITA (Real Intelligence Threat Analytics)

- Analyze skewness: skewness score
- Analyze dispersion: MAD score

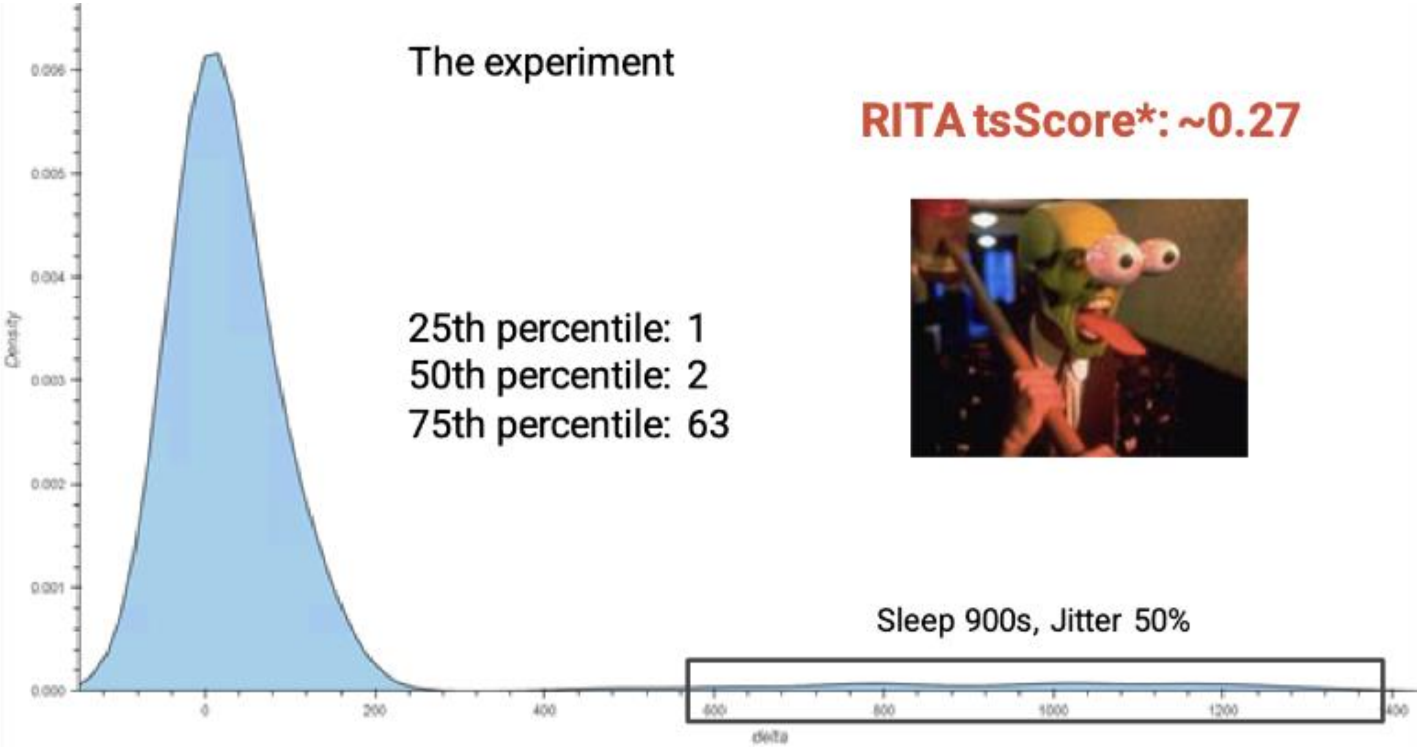
Score = (skewness_score + mad_score)/2

Score > 0.85 BEACONING!!!

Ref: https://github.com/activecm/rita

# Let’s have a slightly better SOP

| Phase | Duration (min) | Sleep (s) |
|---|---|---|
| Normal | 30 | 2 |
| Keyboard activity | 450 | 90 |
| Idle | 960 | 900 |



The experiment

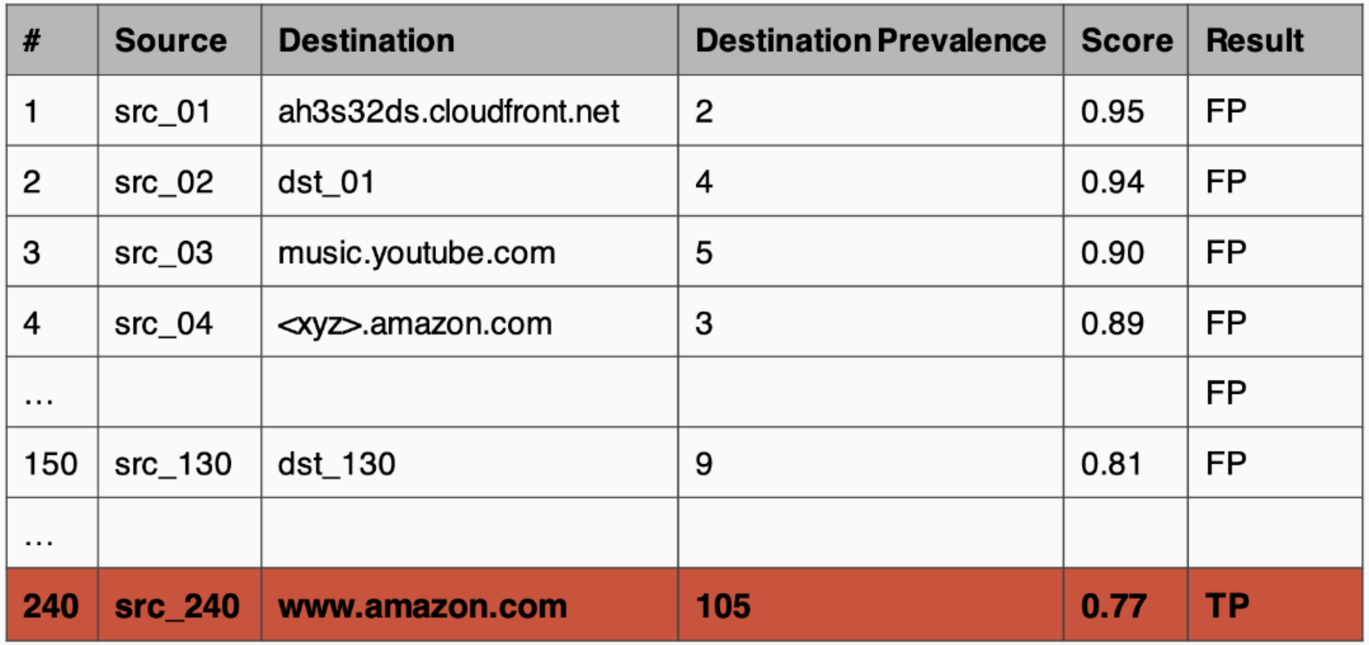
RITA tsScore*: ~0.27

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

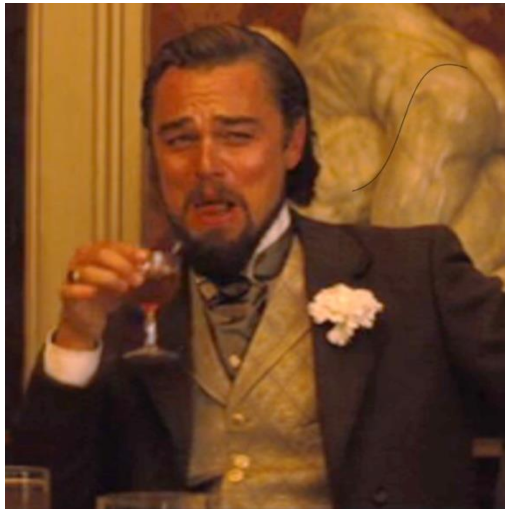
Sleep 900s, Jitter 50%

Ref: https://github.com/Cyb3r-Monk/ACCD/blob/main/Hunting%20C2%20Beaconing%20at%20Scale%20in%20the%20Modern%20Age.pdf

# The beacon will be lost in the ocean of false positives…



| # | 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** |

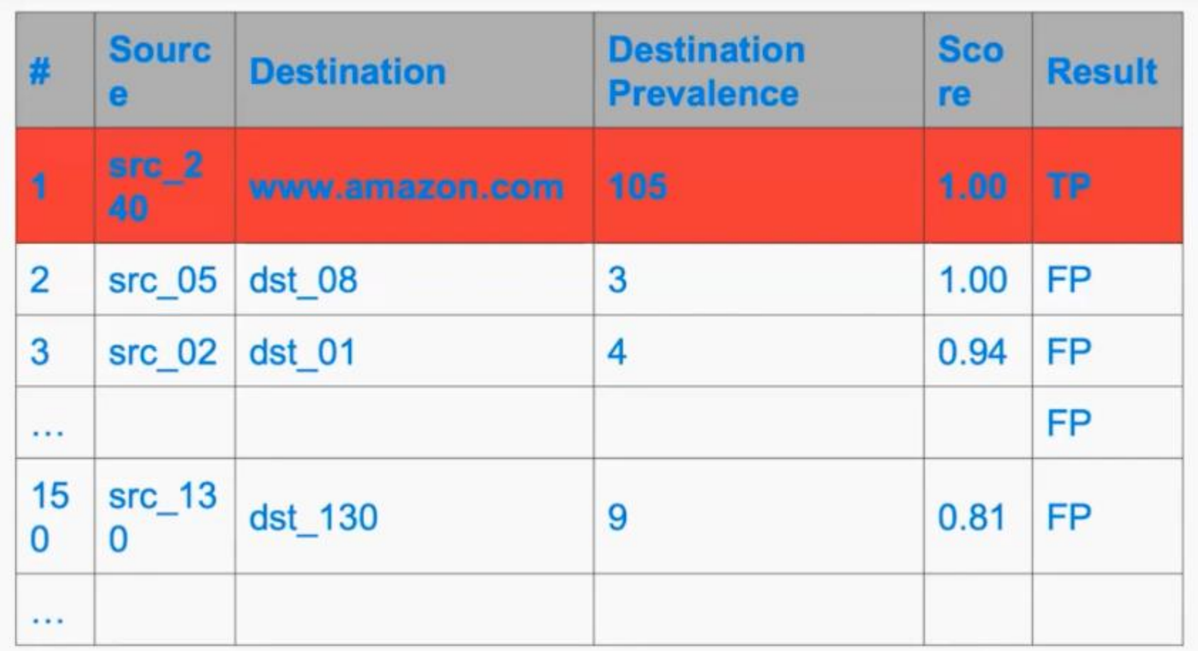


# AC&CD: Active C&C Detector by Mehmet Ergene

- Use 15th, 30th and 45th percentiles
- Use jitter
- Do not use skewness

Same with data-size distribution with one additional factor:

- At least 1 connection must have received data size > 20 kB

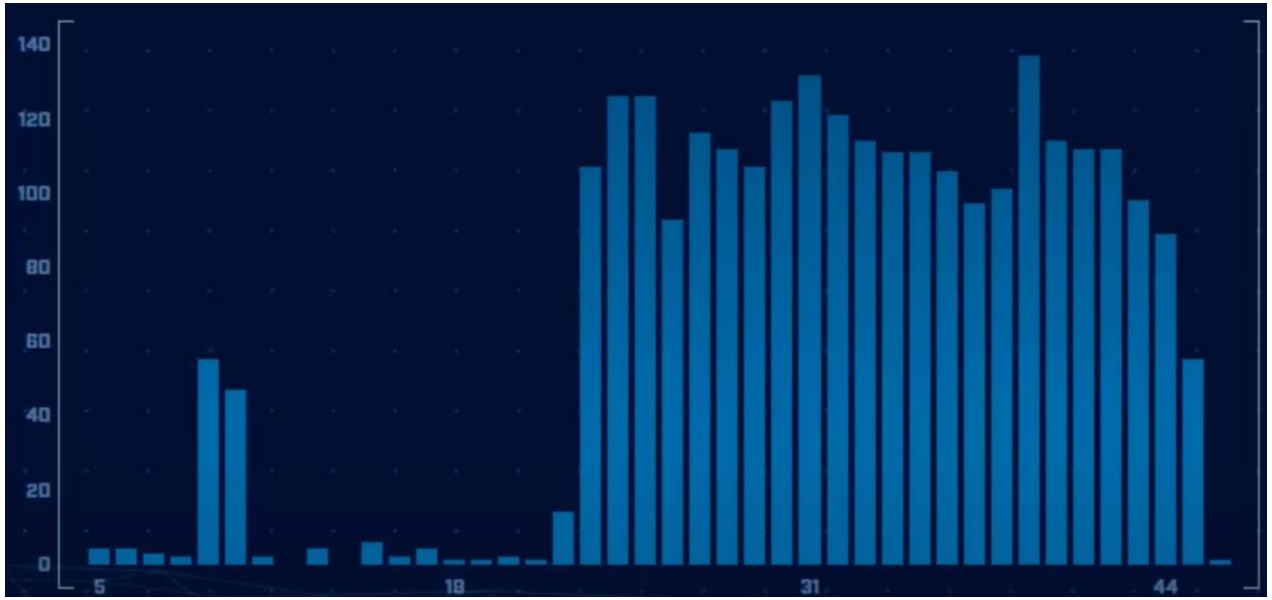


| # | Source | Destination | Destination Prevalence | Score | Result |
|---|---|---|---|---|---|
| 1 | src_240 | www.amazon.com | 105 | 1.00 | TP |
| 2 | src_05 | dst_08 | 3 | 1.00 | FP |
| 3 | src_02 | dst_01 | 4 | 0.94 | FP |
| … | | | | | FP |
| 150 | src_130 | dst_130 | 9 | 0.81 | FP |
| … | | |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

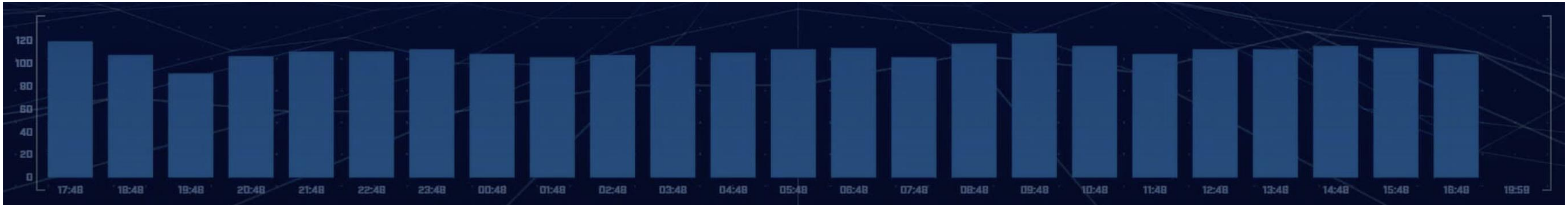
Ref: https://github.com/Cyb3r-Monk/ACCD/tree/main

# What if the attacker took real efforts to hide the traffic?

Time delta distribution



Network connection histogram with 1 hour bin size

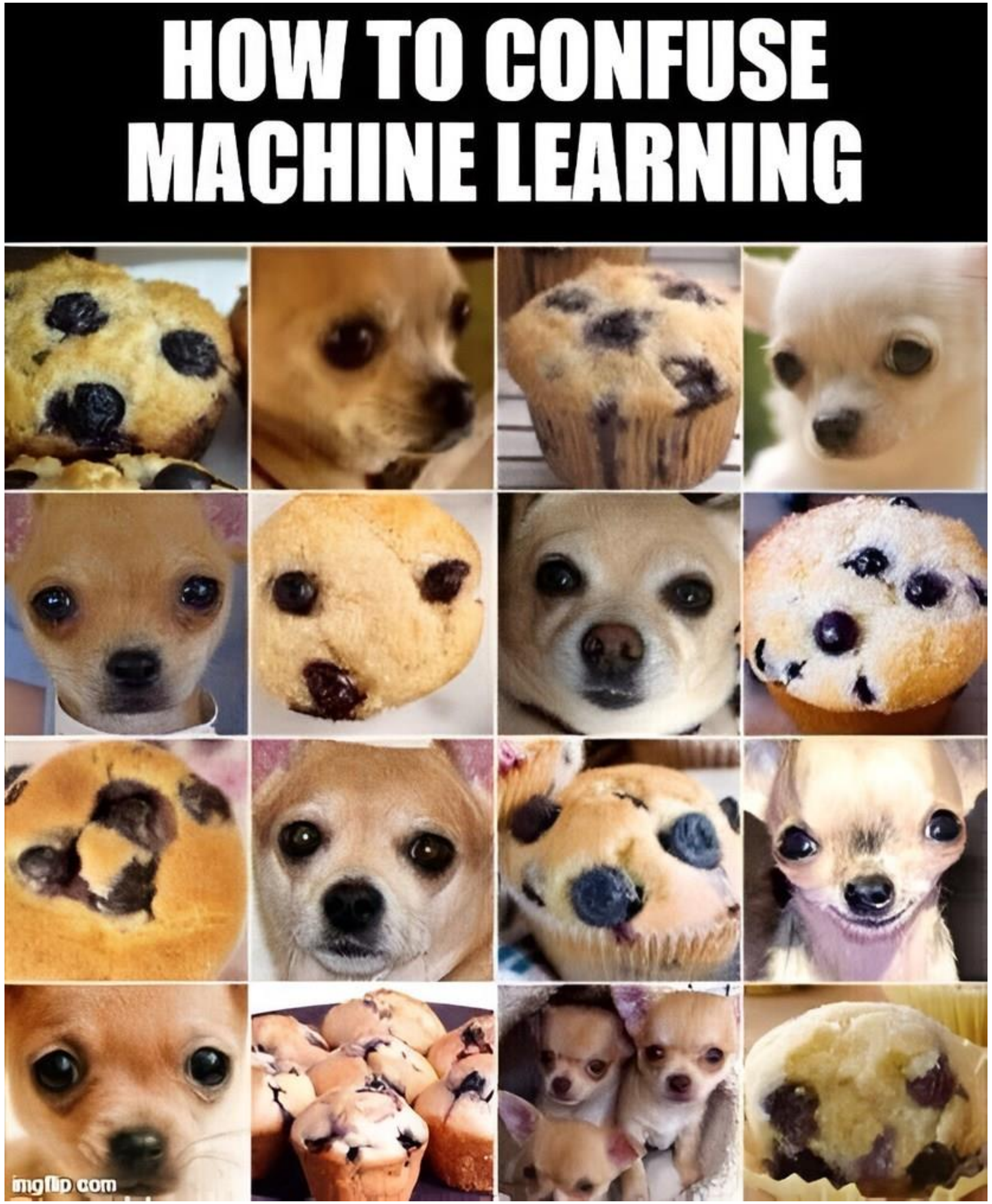


Ref: Chris Brenton, Active Countermeasures, Beacon Analysis – The Key to Cyber Threat Hunting, https://www.youtube.com/watch?v=0b1KPXEVJS0

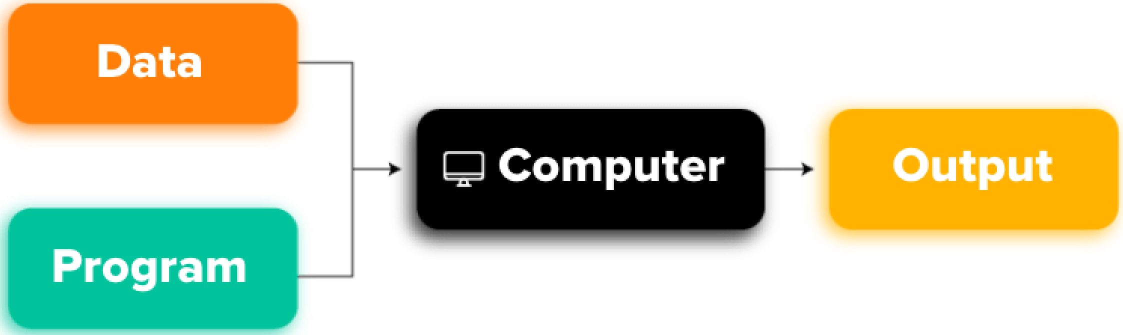
# But we are getting better… Right?

“Dwell time (time for persistent connections) used to be 6 months, now it’s only 4 months.”

Ref:
https://www.theguardian.com/science/2017/jul/26/cats-vs-dogs-in-terms-of-evolution-are-we-barking-up-the-wrong-tree
10.1083/jcb.201907026.


HOW TO CONFUSE MACHINE LEARNING
imgflip.com

# TRADITIONAL PROGRAMMING



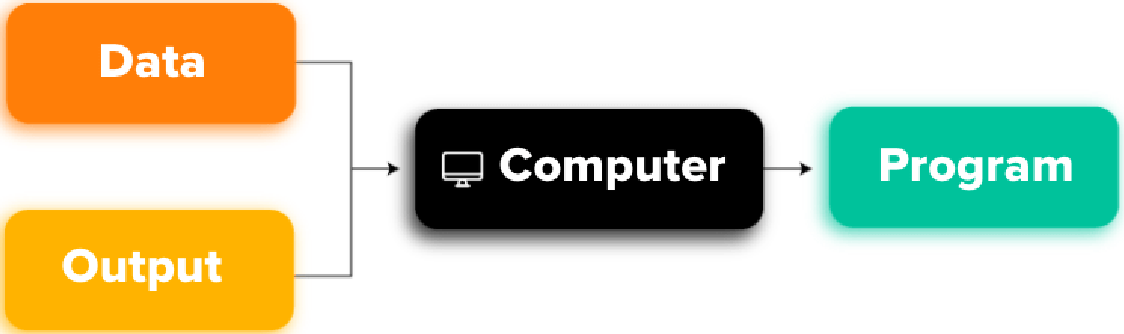
Data

Program

Computer

Output

# MACHINE LEARNING



Data

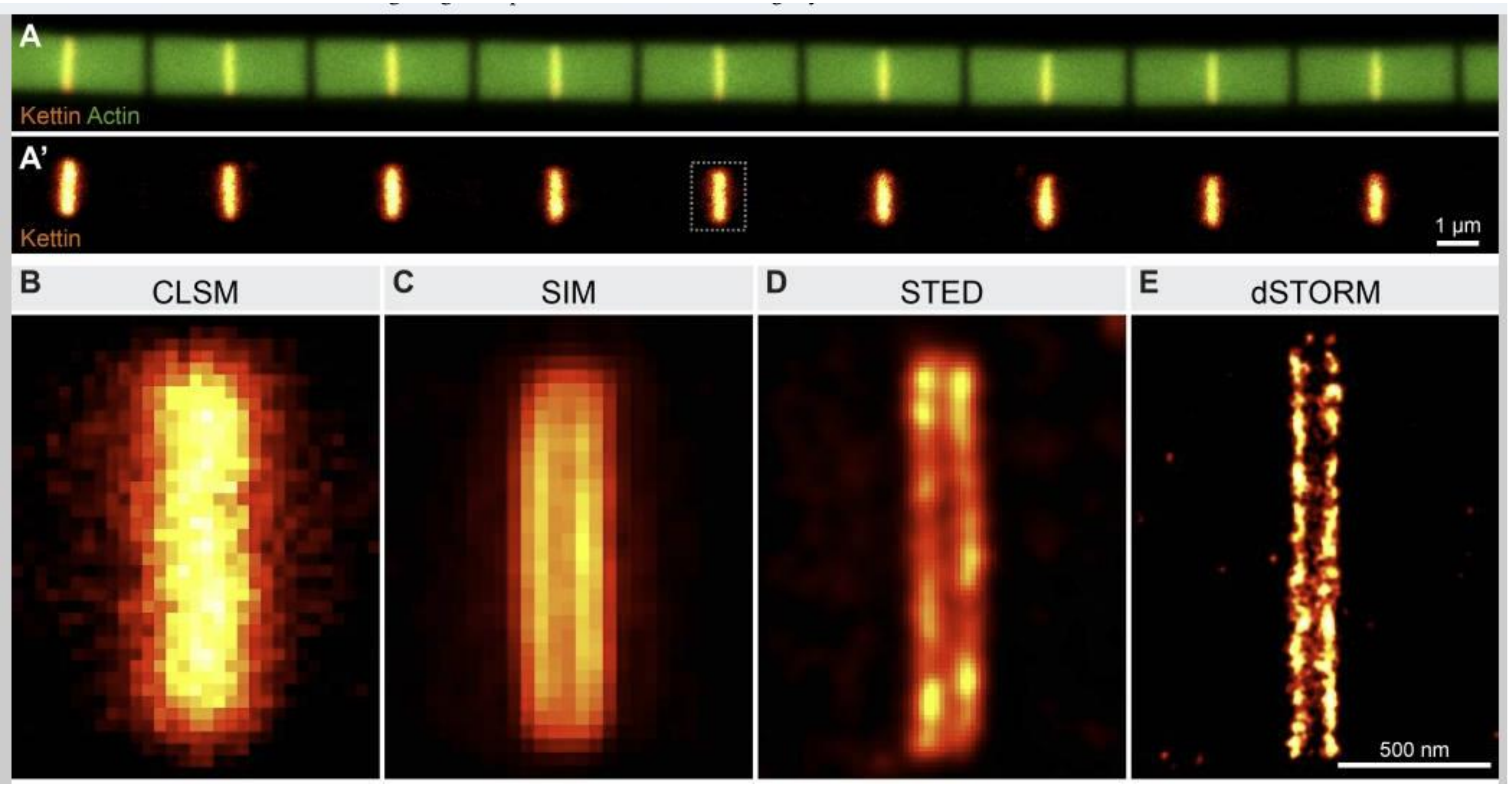
Output

Computer

Program

Ref: https://www.avenga.com/magazine/machine-learning-programming/

# Example from image analysis

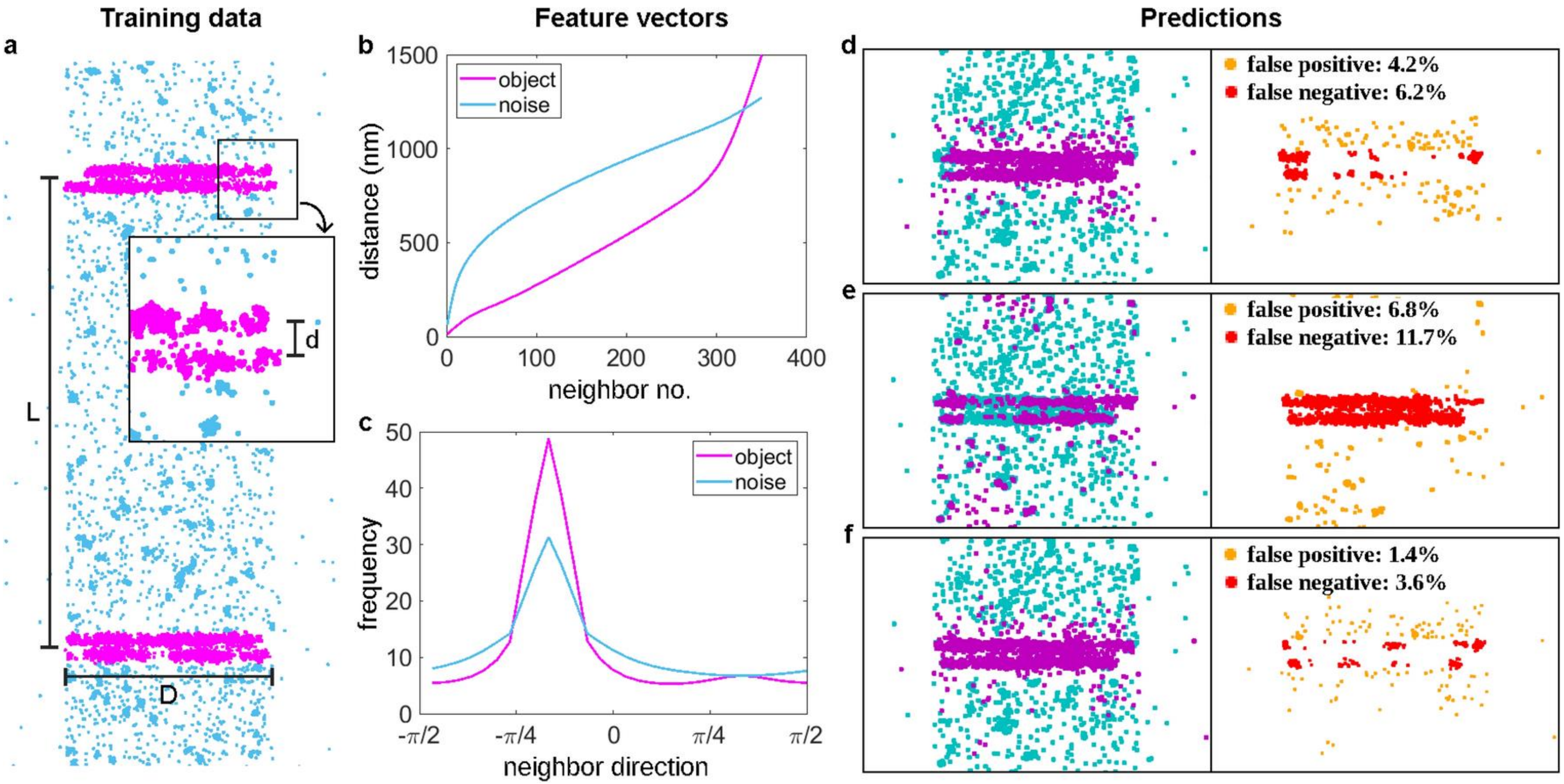


How to find thousands of those double line structures in SMLM images for statistical analysis?

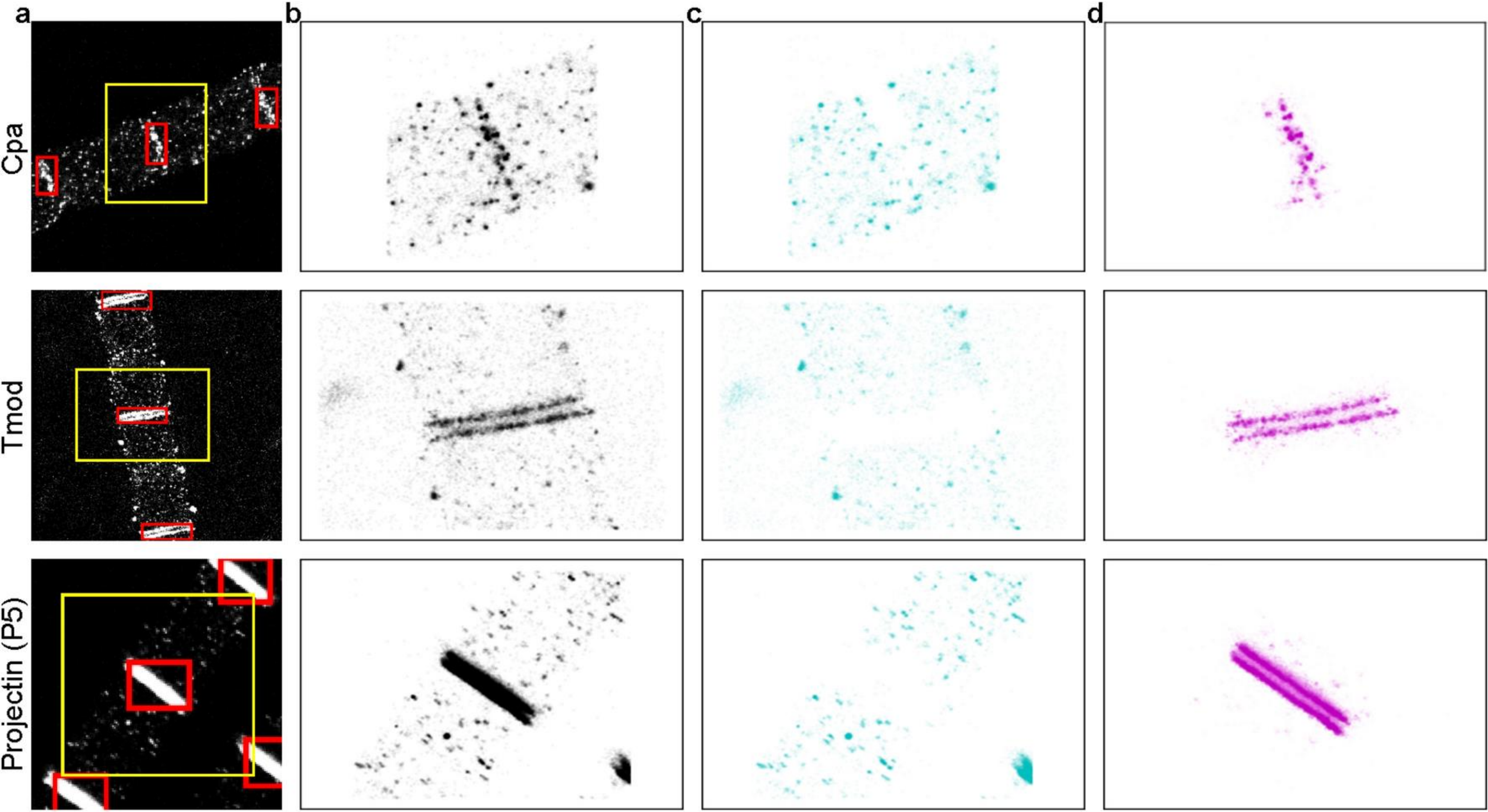
Let's use machine learning!

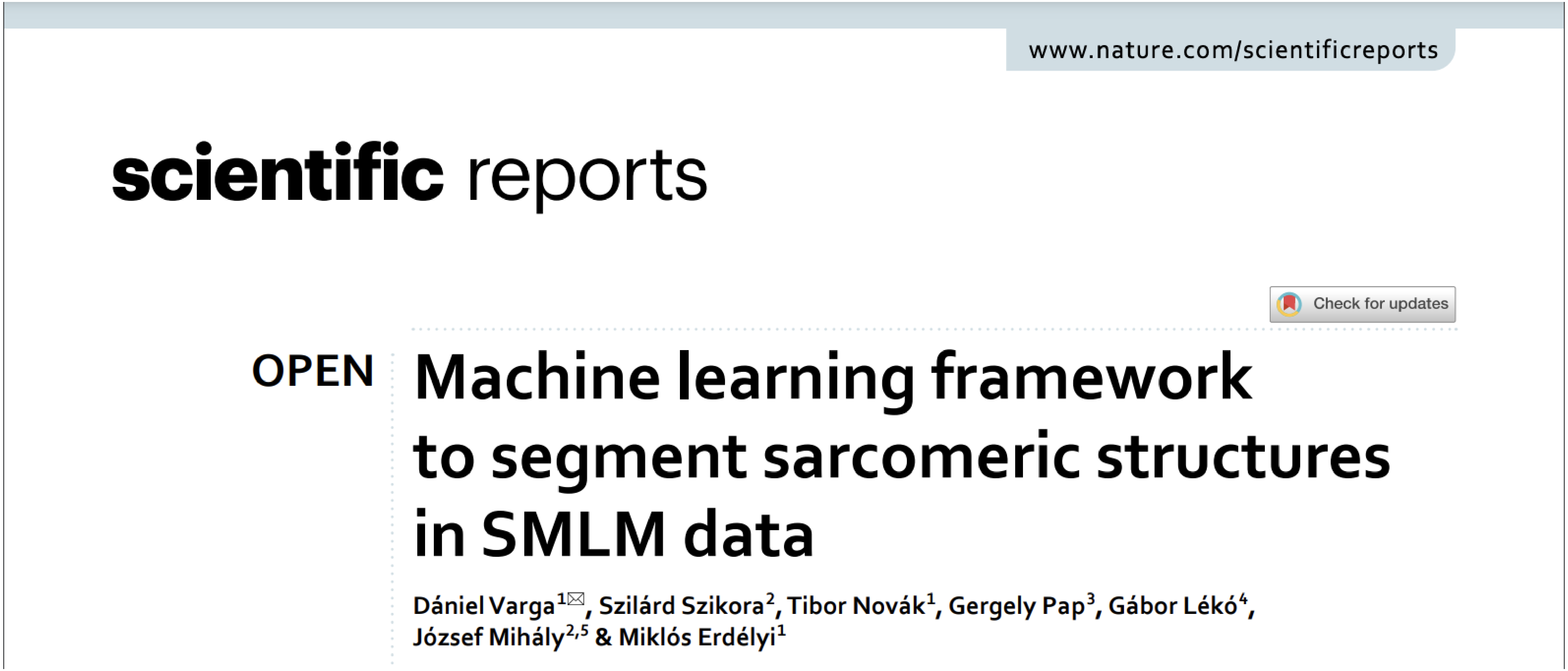
Ref: Szikora, Szilárd, et al. "Nanoscopy reveals the layered organization of the sarcomeric H-zone and I-band complexes." *Journal of Cell Biology* 219.1 (2019): e201907026.

# Solution: simulation!



Ref: Varga, Dániel, et al. "Machine learning framework to segment sarcomeric structures in SMLM data." *Scientific reports* 13.1 (2023): 1582.


a
b
c
d
Cpa
Tmod
Projectin (P5)



www.nature.com/scientificreports

scientific reports

Check for updates

OPEN

# Machine learning framework to segment sarcomeric structures in SMLM data

Dániel Varga[1]✉, Szilárd Szikora[2], Tibor Novák[1], Gergely Pap[3], Gábor Lékó[4], József Mihály[2,5] & Miklós Erdélyi[1]

You can do the same with logs…

# Implementation in our environment: BYO-ML utilities in Azure Machine learning Studio

1. Set up Azure Databricks environment

I. Create an Azure Databricks Workspace

II. Configure a Databricks cluster

2. Data preparation and ingestion

I. Use existing network logs as the baseline for non-malicious activity, since no malicious actor is inside our network

II. Generate artificial malicious beaconing network logs

III. Combine the "not malicious" logs and synthetic malicious logs into a labeled dataset

IV. Split the dataset into training and testing subsets

3. Build and Train ML model

I. Copy-paste from Microsoft Sentinel GitHub repository

II. Build ML model with BYO-ML libraries and templates

III. Train the model

4. Model Scoring Workflow for Log Analytics Integration

I. Configure the trained model to score incoming network logs in real time or on a scheduled basis

II. Use the BYO-ML utilities to write detection scores to Log Analytics in Microsoft Sentinel

5. Operationalize the Detection Rule

I. Set up the analytics rule based on the ML results

6. Monitor and Maintain

I. Refine during solving the generated incidents to reduce false positives

II. Regularly retrain the model with updated network logs

DALL·E

STEP 1
TOTALLY SAFE
STEP DATA
NETWORK PEPERATION
MALICIOUS LOG!
DETECTIOUS BEACORICKS
NETWORT NEURAL NETWARK
AZURE
Azure
DATA PREPARATION
SET UP MALICOUS SAFE!
MELICIOUS LOG
MALICOUS AZURE
SET UP DataBricks
BUILD & TRAIN NEURAL NETWAK
SUIR DETECTING NEURAL NETWERK
STEP ?
MALICIOUS LOG
SEP B 3: NALICIOUS LOGS
BUID AND TRAIN A NETWOUS LOG!
OUS LOG

# Thank you for your attention!

daniel.varga@lego.com