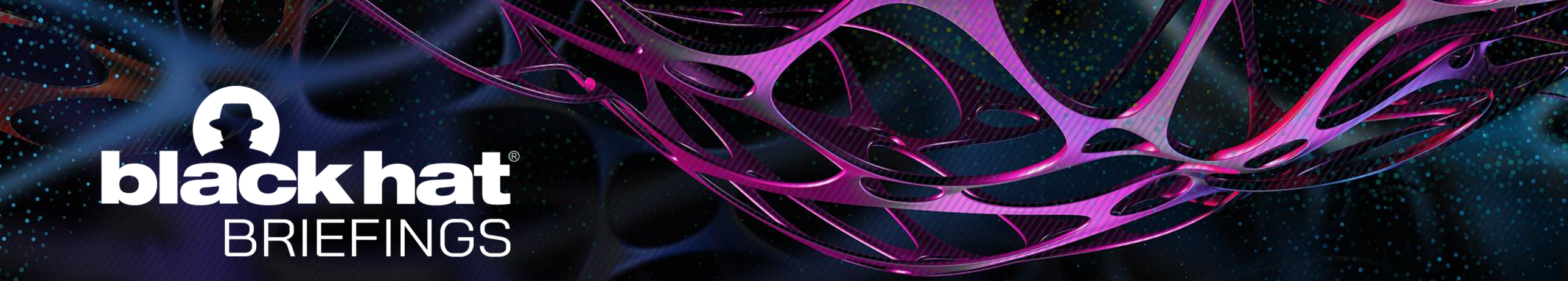



AUGUST 6-7, 2025

MANDALAY BAY / LAS VEGAS

# Protecting Small Organizations in the Era of AI Bots

Rama Carl Hoetzlein

"51% of Internet traffic is non-human, with
37% of Internet traffic from bad bots"

2025 Imperva, Bad Bot Report

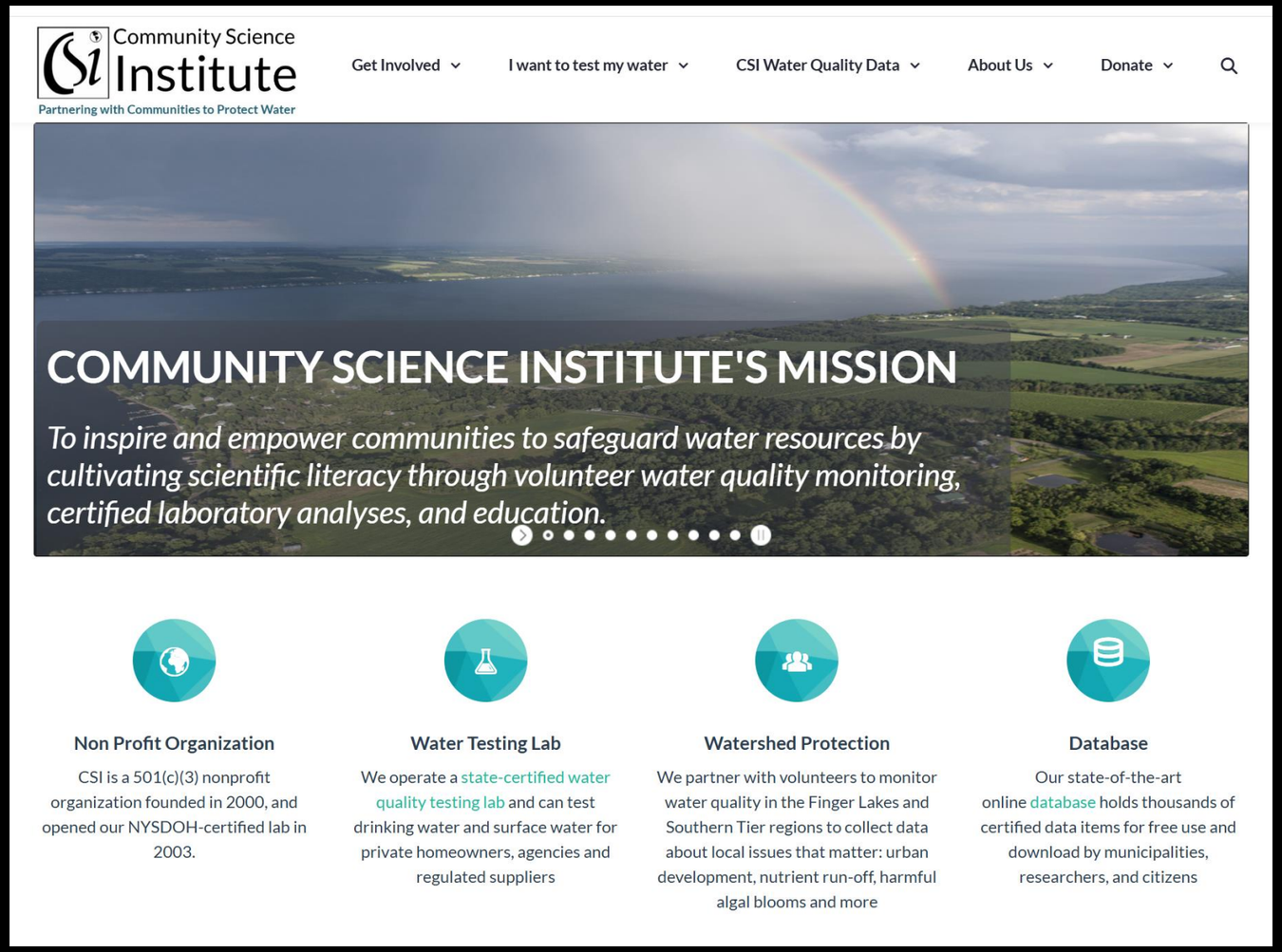
"51% of Internet traffic is non-human, with 37% of Internet traffic from bad bots"

2025 Imperva, Bad Bot Report

"87% of the malicious bot IPs [in our study] were not listed in popular IP blocklists."

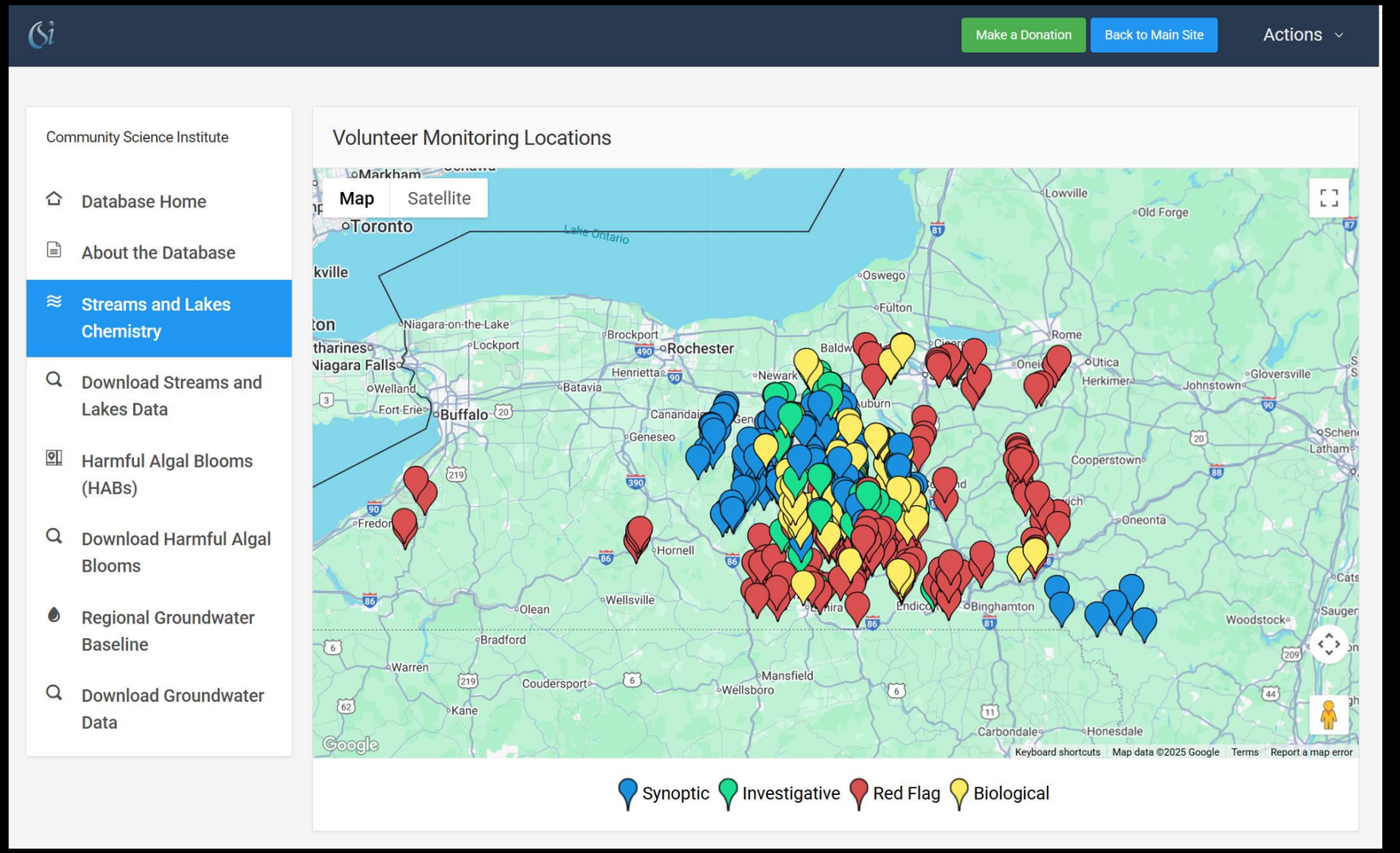
2021 Xigao Li et al., Good Bot, Bad Bot

# Client




The Community Science Institute is a public, non-profit that promotes scientific literacy, volunteer water quality monitoring and certified lab analysis for central New York.

# Client




CSI Database: Curated, certified, water quality data for Stream & Lake chemistry, Harmful Algae Blooms and Biomonitoring

# Client

We observed that a single server received over 150,000 page hits over 20 days, corresponding to **7,500 hits / day.**

# Client

We observed that a single server received over 150,000 page hits over 20 days, corresponding to **7,500 hits / day.**

Traffic was so severe that it was degrading server performance for CSI's known human users and clients.

# Early Investigation

IP B-class aggregation and org lookup

| | | | | | | |
|---|---|---|---|---|---|---|
| 1.83.255.255 | 2 | 2 | 0 | CHINANET-BACKBONE | Shaanxi | China |
| 2.57.255.255 | 1 | 4 | 0 | UNMANAGED-DEDICATED-SERVERS | North Holland | The Netherlands |
| 2.125.255.255 | 1 | 2 | 0 | BSKYB-BROADBAND-AS | England | United Kingdom |
| 2.136.255.255 | 1 | 1 | 0 | Telefonica_de_Espana | Andalusia | Spain |
| 3.12.255.255 | 14 | 45 | 4612 | AMAZON-02 | Ohio | United States |
| 3.80.255.255 | 1 | 3 | 0 | AMAZON-AES | Virginia | United States |
| 4.151.255.255 | 4 | 9 | 2454 | MICROSOFT-CORP-MSN-AS-BLOCK | Texas | United States |
| 5.58.255.255 | 1 | 1 | 0 | COLUMBUS-PE-TE | Ternopil Oblast | Ukraine |
| 5.101.255.255 | 1 | 2 | 0 | PINDC-AS | St.-Petersburg | Russia |
| 5.102.255.255 | 1 | 184 | 148 | CUSTDC | England | United Kingdom |
| 5.135.255.255 | 1 | 1 | 0 | OVH | Hauts-de-France | France |
| 5.161.255.255 | 3 | 3 | 0 | HETZNER-CLOUD2-AS | Virginia | United States |
| 5.181.255.255 | 3 | 135 | 51 | ORG-ISI14-RIPE | La RiviÃ¨re Angla | Seychelles |
| 5.185.255.255 | 1 | 1 | 0 | TPNET | Mazovia | Poland |
| 5.235.255.255 | 1 | 1 | 0 | TCI | East Azerbaijan P | Iran |
| 5.255.255.255 | 97 | 173 | 1491 | YANDEX | Moscow | Russia |
| 8.48.255.255 | 1 | 1 | 0 | GOGO | Colorado | United States |
| 8.210.255.255 | 8 | 26 | 5855 | ALIBABA-CN-NET | Central and West | Hong Kong |
| 17.241.255.255 | 330 | 455 | 2705 | APPLE-ENGINEERING | California | United States |

Visitor traffic is from the entire world, despite the fact that the CSI Database is entirely data for central New York State

# Background

What existing tools are available?

1. Throttling is ineffective – modern crawlers *observe* rate limits.
2. Public blocklists are ineffective – up to 87% not listed
3. GREP is ineffective – difficult to interpret, good for spot checks
4. GoAccess, AWStats – summary statistics hide details
5. OSSEC, CrowdSec – real-time monitoring, do not examine historic/log access patterns
6. AI/ML Detection (Meyer 2008) – requires non-attack baseline
7. Rank Analysis (Zang 2008) – requires good pre-filtering
8. Large Organizations (Yen 2013) – we focus on small organizations

# Recent Approaches & Limitations

AI/ML Detection (Meyer 2008) – requires non-attack baseline
Rank Analysis (Zang 2008) – requires good pre-filtering
Large Organizations (Yen 2013) – we focus on small organizations

GoAccess log analysis

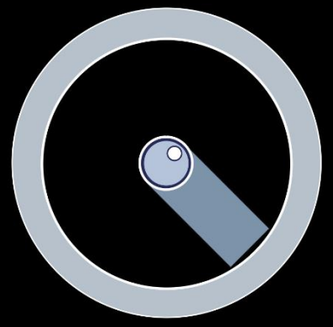
```
Dashboard - Overall Analyzed Requests (23/Jan/2025 - 13/Feb/2025)                [Active Panel: Visitors]

 Total Requests   119247 Unique Visitors   16874 Requested Files 22299 Referrers   0
 Valid Requests   116282 Log Parsing Time 1s    Static Files     1881  Log Size    22.53 MiB
 Failed Requests 2965    Excl. IP Hits     0    Not Found        0     Tx. Amount 0.0    B
 Log Source       csi_log_2025_02_12.txt

> 1 - Unique visitors per day - Including spiders                                Total: 22/22

Hits      h% Vis.     v% Data
----- ------ ---- ------ ----
 1335  1.15%  216  1.28% 13/Feb/2025 |||||||
13284 11.42%  805  4.77% 12/Feb/2025 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 5971  5.13% 1103  6.54% 11/Feb/2025 |||||||||||||||||||||||||||||||||||||||
 7133  6.13%  639  3.79% 10/Feb/2025 |||||||||||||||||||||||||||||||||||||||||||||
 3198  2.75%  711  4.21% 09/Feb/2025 |||||||||||||||||||
 1803  1.55%  621  3.68% 08/Feb/2025 |||||||||
 5702  4.90%  712  4.22% 07/Feb/2025 |||||||||||||||||||||||||||||||
 2631  2.26%  646  3.83% 06/Feb/2025 ||||||||||||||
 8817  7.58% 1303  7.72% 05/Feb/2025 |||||||||||||||||||||||||||||||||||||||||||||||||||
 7073  6.08%  930  5.51% 04/Feb/2025 |||||||||||||||||||||||||||||||||||||||||||||
 3217  2.77%  630  3.73% 03/Feb/2025 |||||||||||||||||||
 5004  4.30%  629  3.73% 02/Feb/2025 ||||||||||||||||||||||||||
 8848  7.61%  699  4.14% 01/Feb/2025 ||||||||||||||||||||||||||||||||||||||||||||||||||
 2768  2.38%  587  3.48% 31/Jan/2025 ||||||||||||||||
 3468  2.98%  893  5.29% 30/Jan/2025 ||||||||||||||||||
10442  8.98%  641  3.80% 29/Jan/2025 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 2990  2.57%  665  3.94% 28/Jan/2025 ||||||||||||||||||
 3786  3.26%  578  3.43% 27/Jan/2025 ||||||||||||||||||||||
 2832  2.44%  794  4.71% 26/Jan/2025 ||||||||||||||||
 9439  8.12% 1519  9.00% 25/Jan/2025 ||||||||||||||||||||||||||||||||||||||||||||||||||||||
 5683  4.89% 1463  8.67% 24/Jan/2025 |||||||||||||||||||||||||||||||
  858  0.74%   90  0.53% 23/Jan/2025 |||||
```

Statistical tools just tell us – yes – you have a lot of traffic, and it varies by day.

# Methods

# Question:

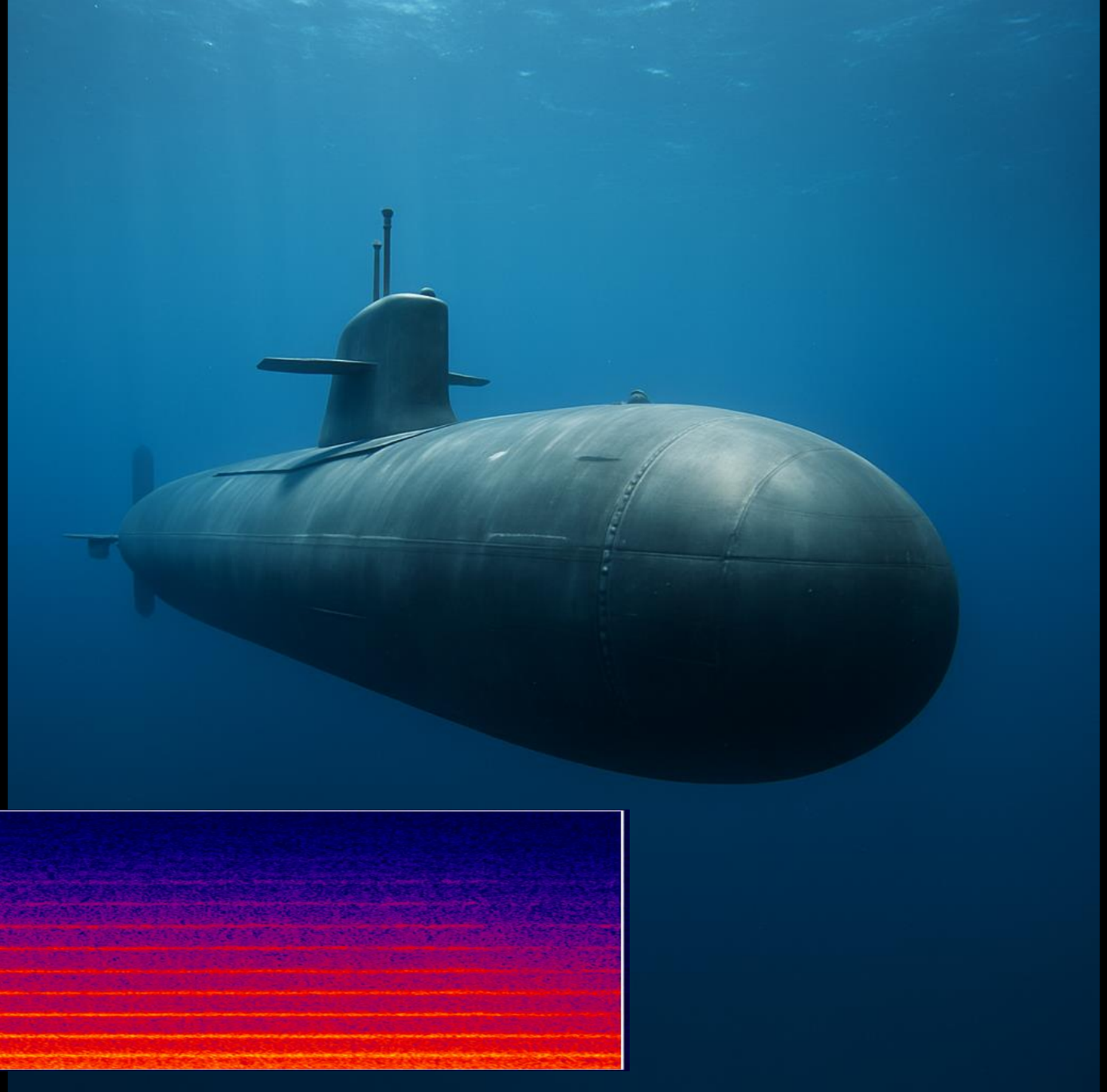
How can we distinguish human access patterns from machines?

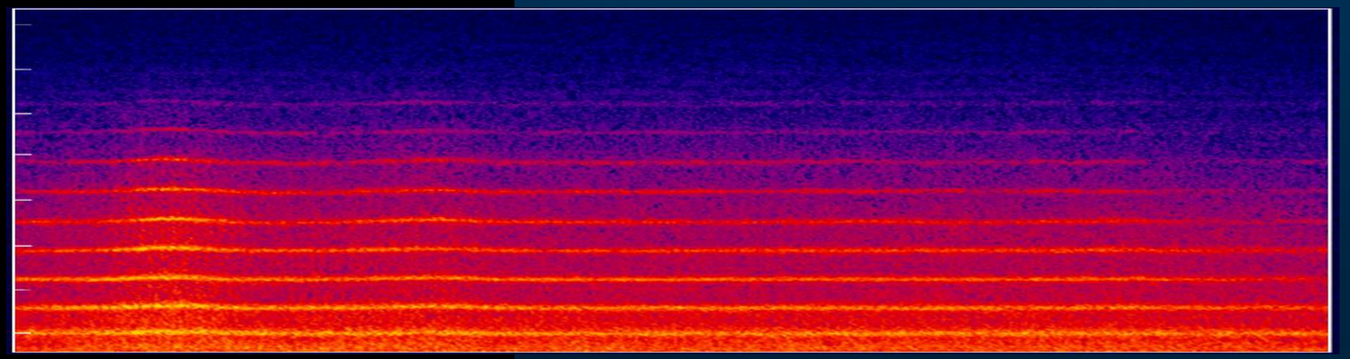


QUANTA

Quanta Sciences

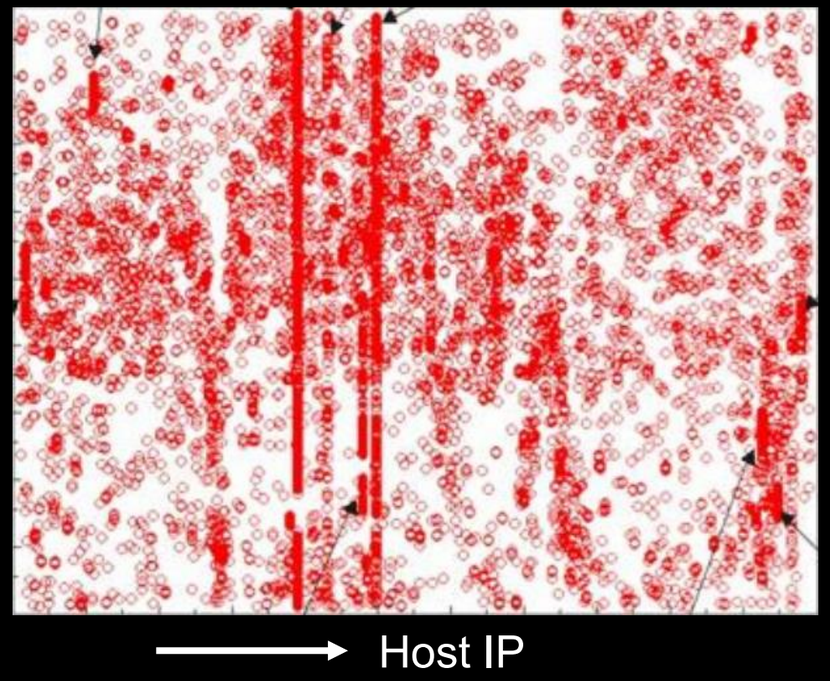
we are a knowledge systems, AI and data visualization startup





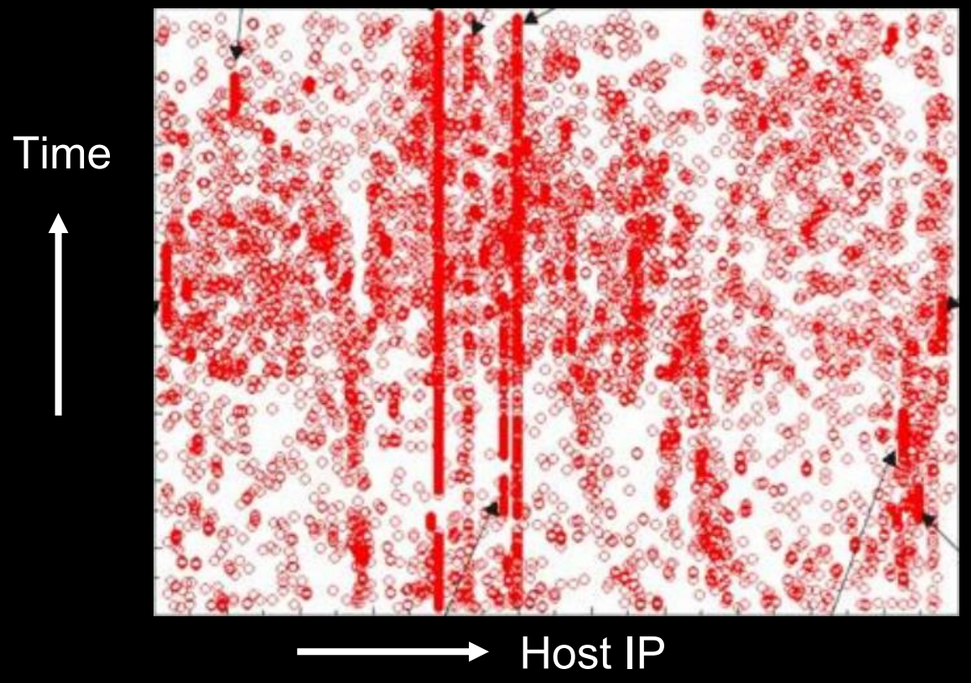
Does it *sound* mechanical?

# Investigation

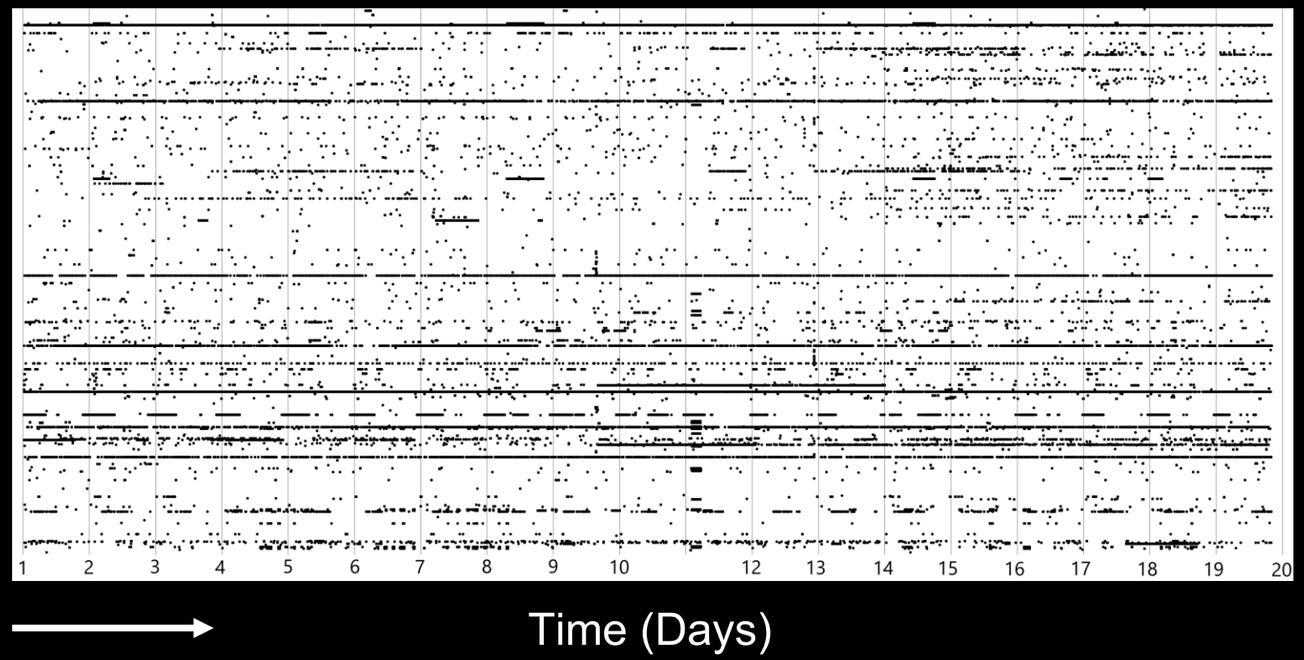



*From:*
Jungkee Kim, Web Server Log Visualization,
Intl. Journal of Advance Smart Convergence, 2018

# Investigation

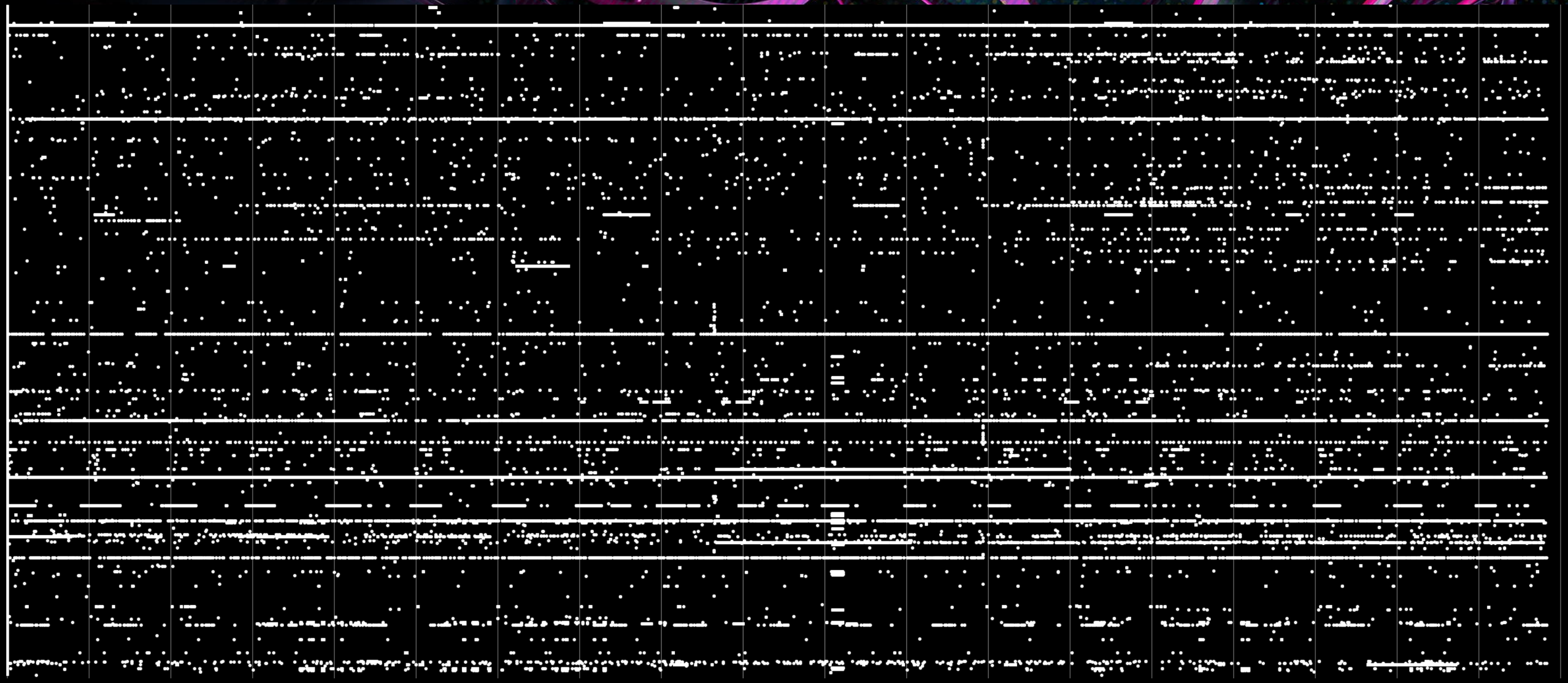


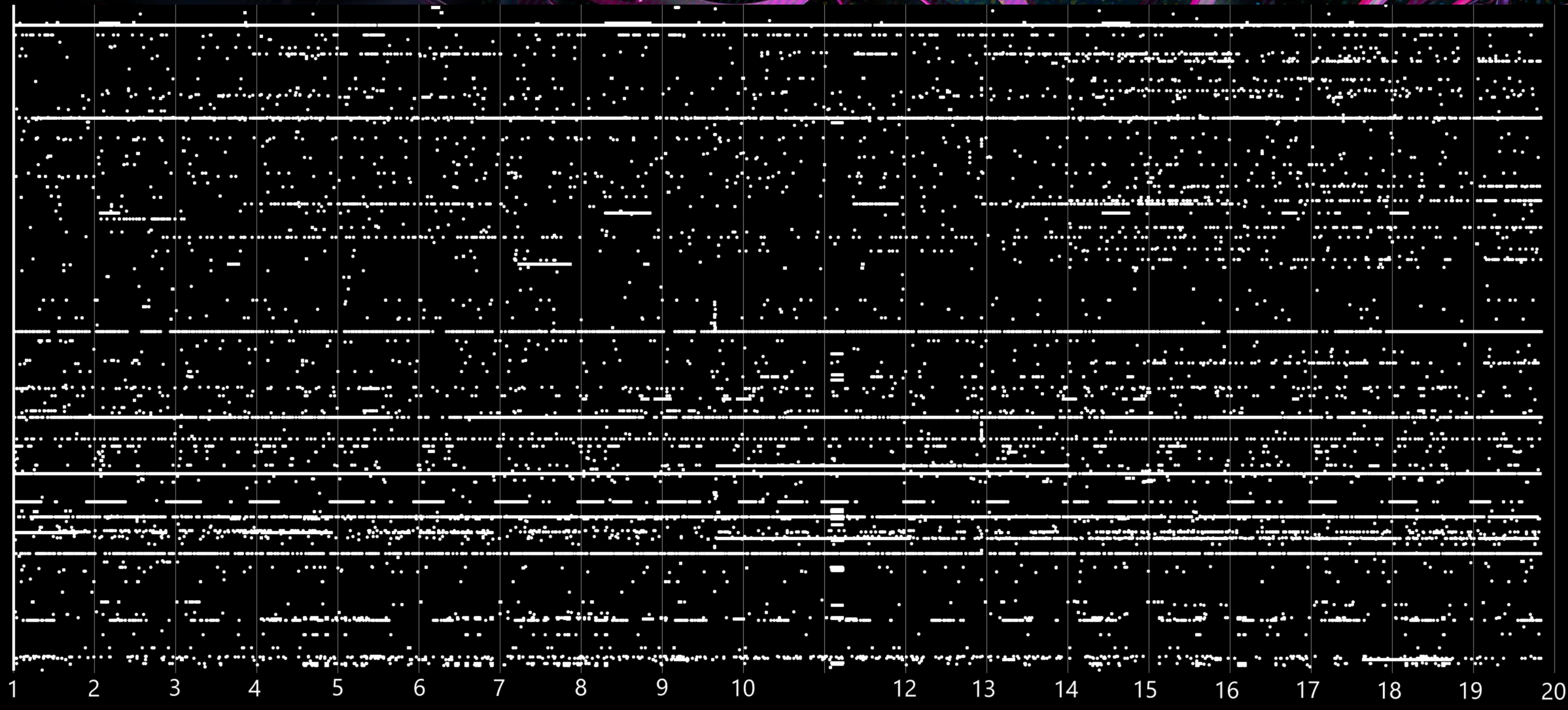

*From:*
Jungkee Kim, Web Server Log Visualization,
Intl. Journal of Advance Smart Convergence, 2018



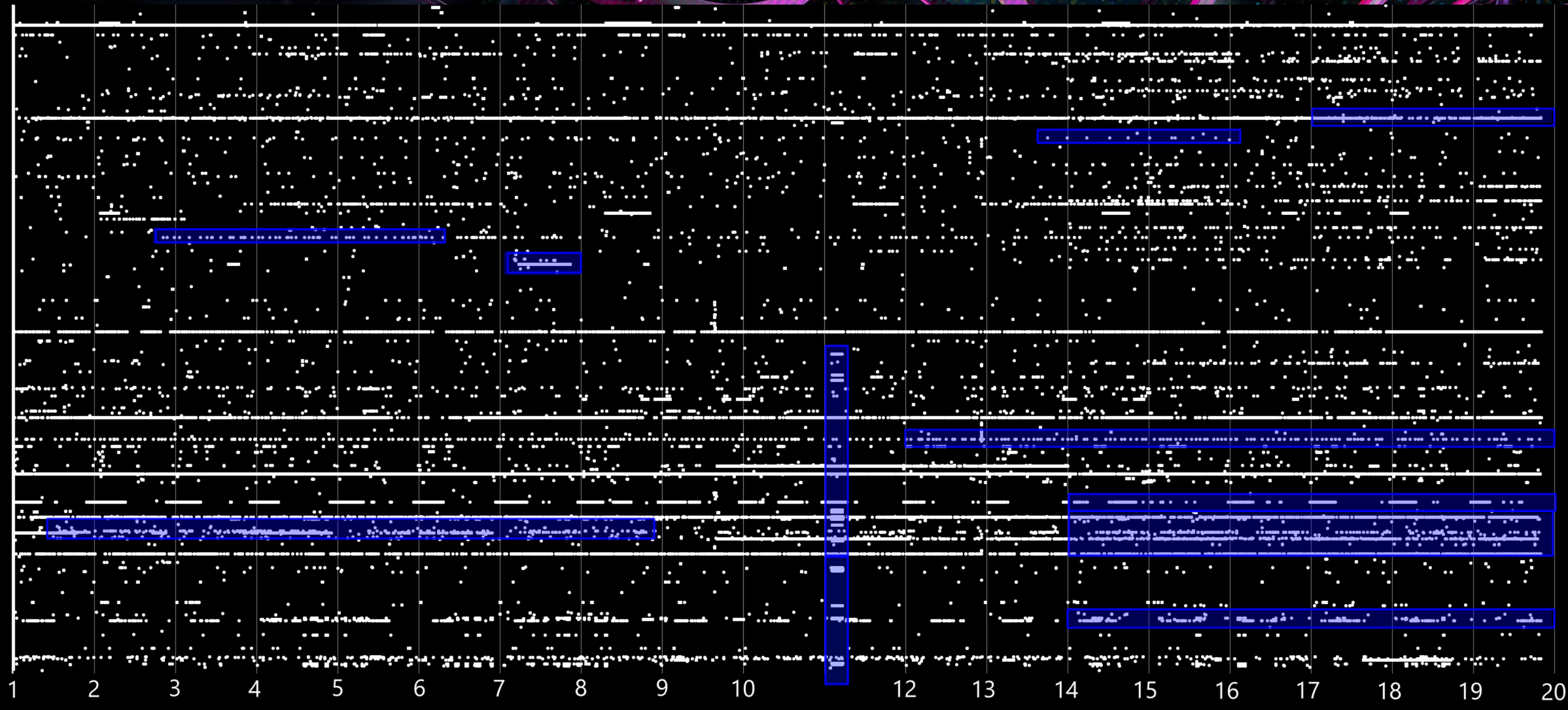

Benefits of Visualization:

- Entire log in one snapshot
- Everything is there, no statistical summary
- Easy for humans to see patterns


1
2
3
4
5
6
7
8
9
10
12
13
14
15
16
17
18
19
20
Time (Days)




What do you think is human here?


1
2
3
4
5
6
7
8
9
10
12
13
14
15
16
17
18
19
20
Time (Days)
Probably not human

# Methods

We are interested in distinguishing mechanical access patterns regardless of whether they are benign or malicious.

# Methods

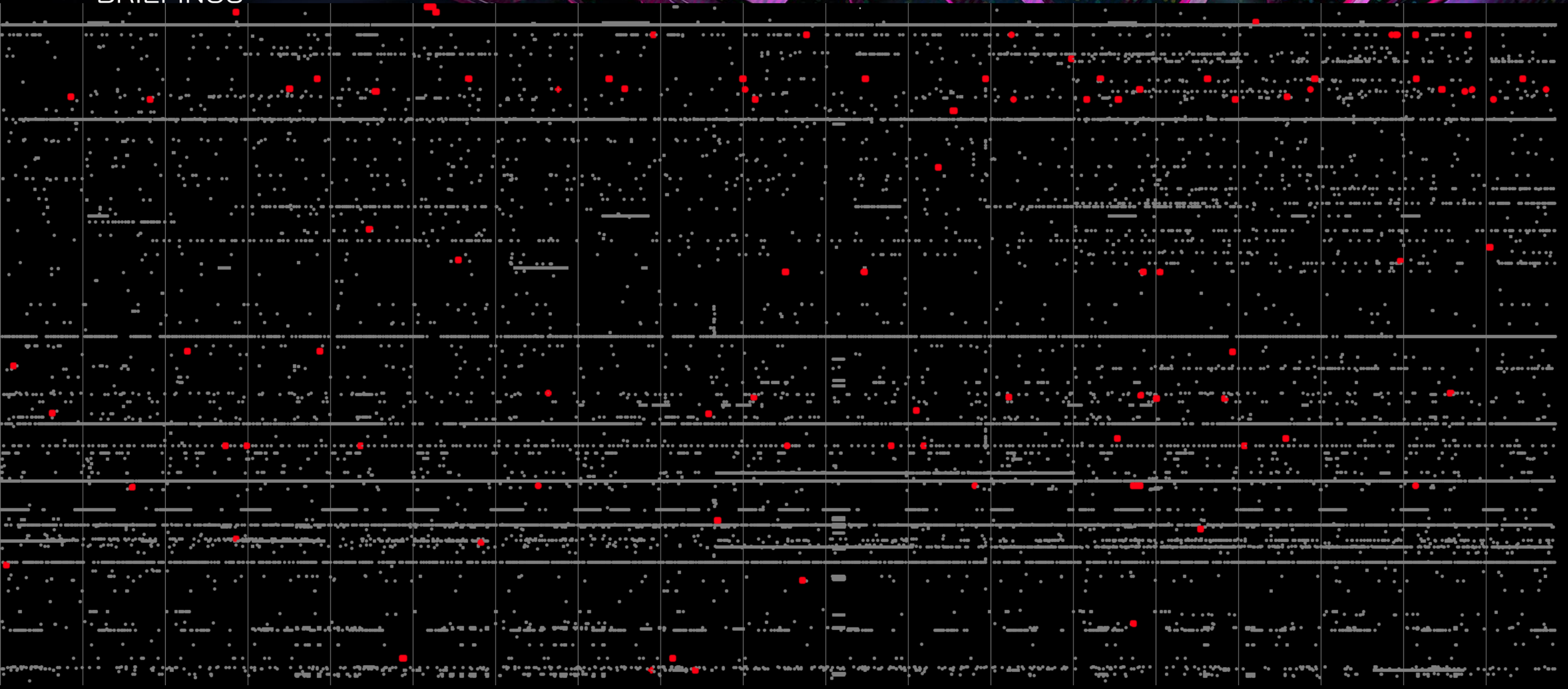
Throttling:

How fast are you?

Block based on frequency of visit.

e.g. no more than 20 pages/minute

We found that most traffic was observing rate limits.



- Throttle limited IPs

Reduced traffic by only 33%

# Methods

What are other patterns that humans would follow?

| | | *Human* |
|---|---|---|
| 1. Throttling | - How fast are you? | <20 page/min |

# Methods

What are other patterns that humans would follow?

| | *Human* |
|---|---|
| 1. Throttling - How fast are you? | <20 page/min |
| 2. Consecutive - How often do you visit? | <5 days consec. |

# Methods

What are other patterns that humans would follow?

| | *Human* |
|---|---|
| 1. Throttling - How fast are you? | <20 page/min |
| 2. Consecutive - How often do you visit? | <5 days consec. |
| 3. Daily Range - How long can you work? | <6 hours/day |

# Methods

What are other patterns that humans would follow?

| | *Human* |
|---|---|
| 1. Throttling - How fast are you? | <20 page/min |
| 2. Consecutive - How often do you visit? | <5 days consec. |
| 3. Daily Range - How long can you work? | <6 hours/day |
| 4. Daily Hits - How much do you look at? | <100 hits/day |

# Methods

What are other patterns that humans would follow?

| | *Human* |
|---|---|
| 1. Throttling - How fast are you? | <20 page/min |
| 2. Consecutive - How often do you visit? | <5 days consec. |
| 3. Daily Range - How long can you work? | <6 hours/day |
| 4. Daily Hits - How much do you look at? | <100 hits/day |

Behavioral Science in Human-Computer Interaction

# Methods

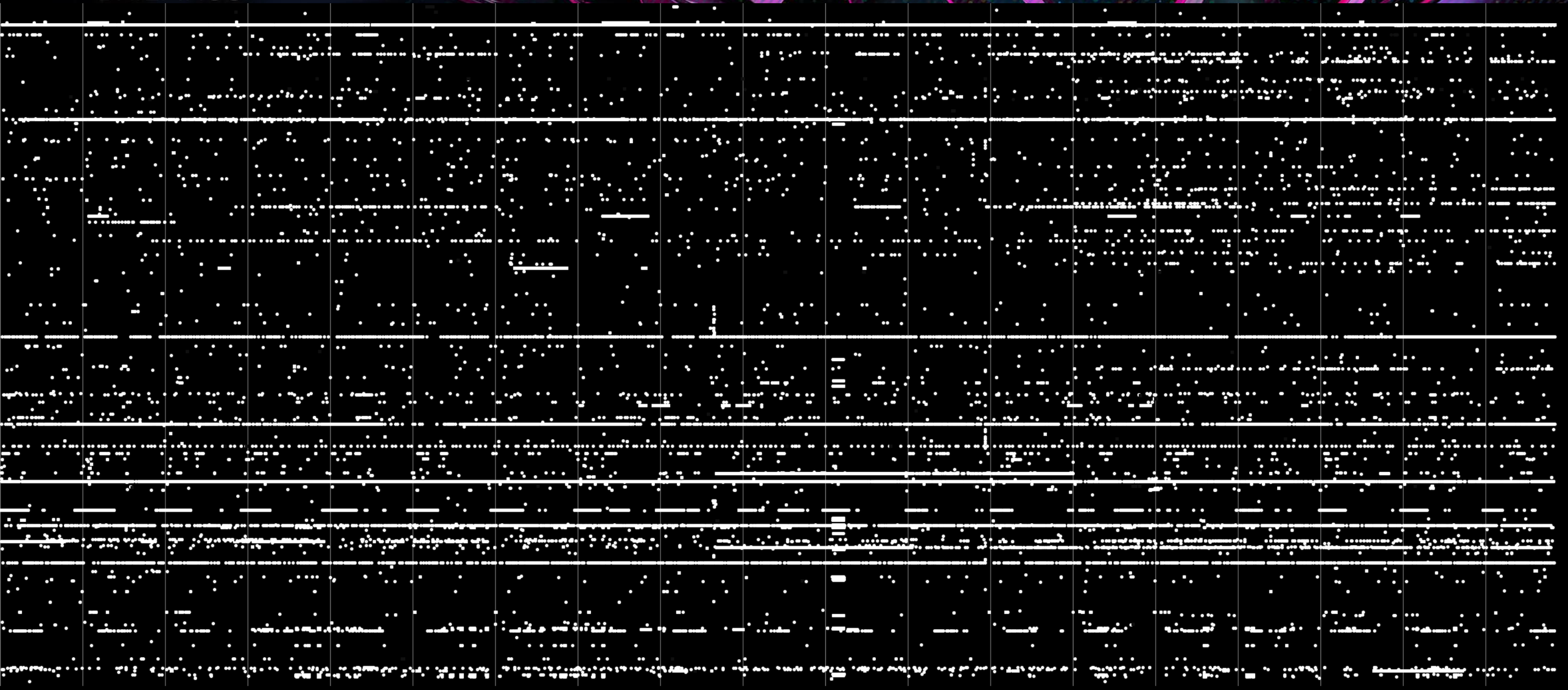
**LOGRIP**

Let's use Human Behavioral Metrics to develop a…

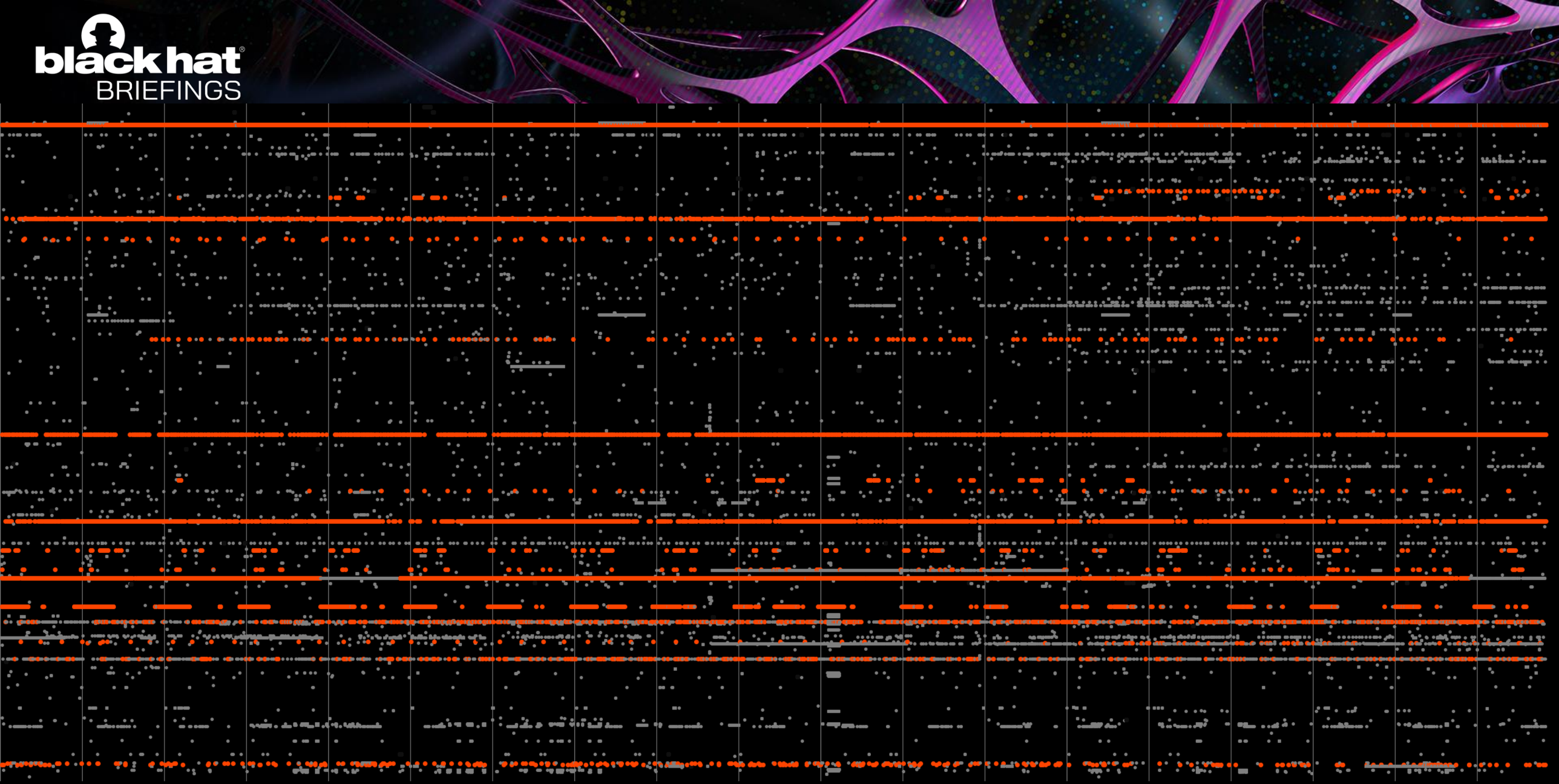
Scoring Algorithm:

1. IP Hashing - key-value map of IPs from raw pages
2. Sort page hits by day & time
3. Apply behavioral metrics
4. Score based on a weighted contribution of metrics

# Intermediate Results



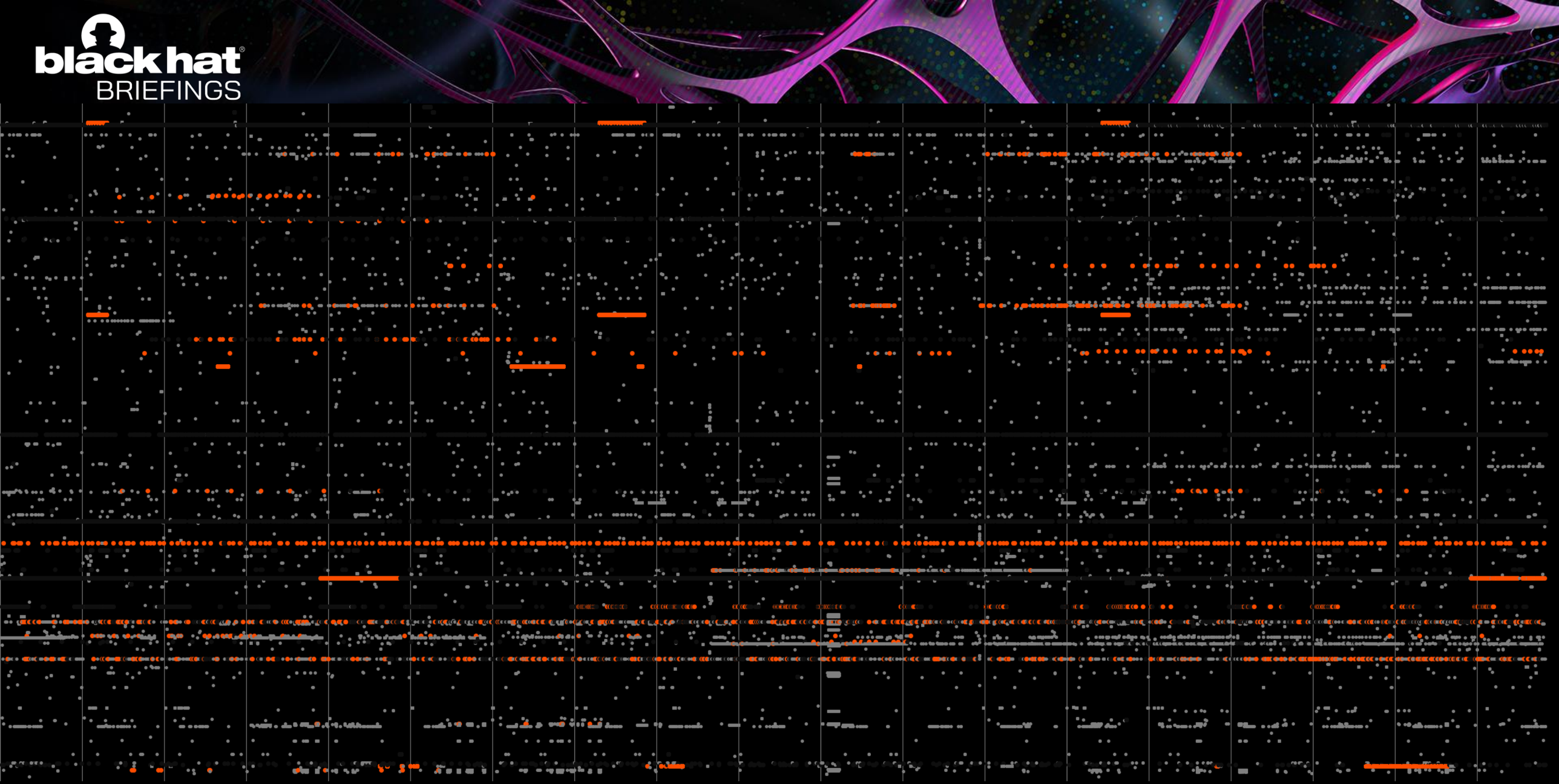
## Original Traffic



# Blocked by Consecutive Days /w Daily Range

RAMA CARL HOETZLEIN

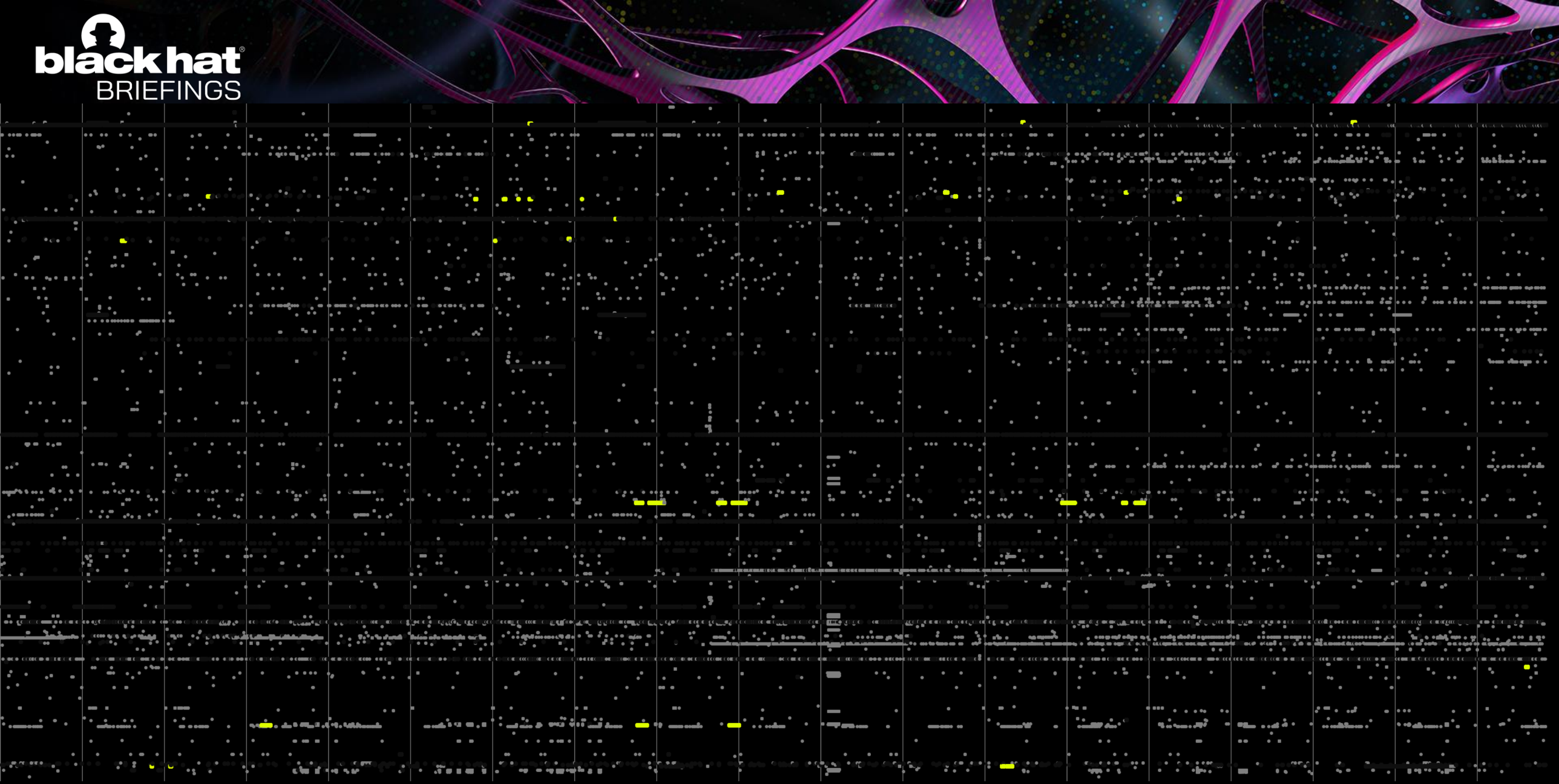
PROTECTING SMALL ORGANIZATIONS IN THE ERA OF AI BOTS



# Blocked by Daily Range with Freq

RAMA CARL HOETZLEIN

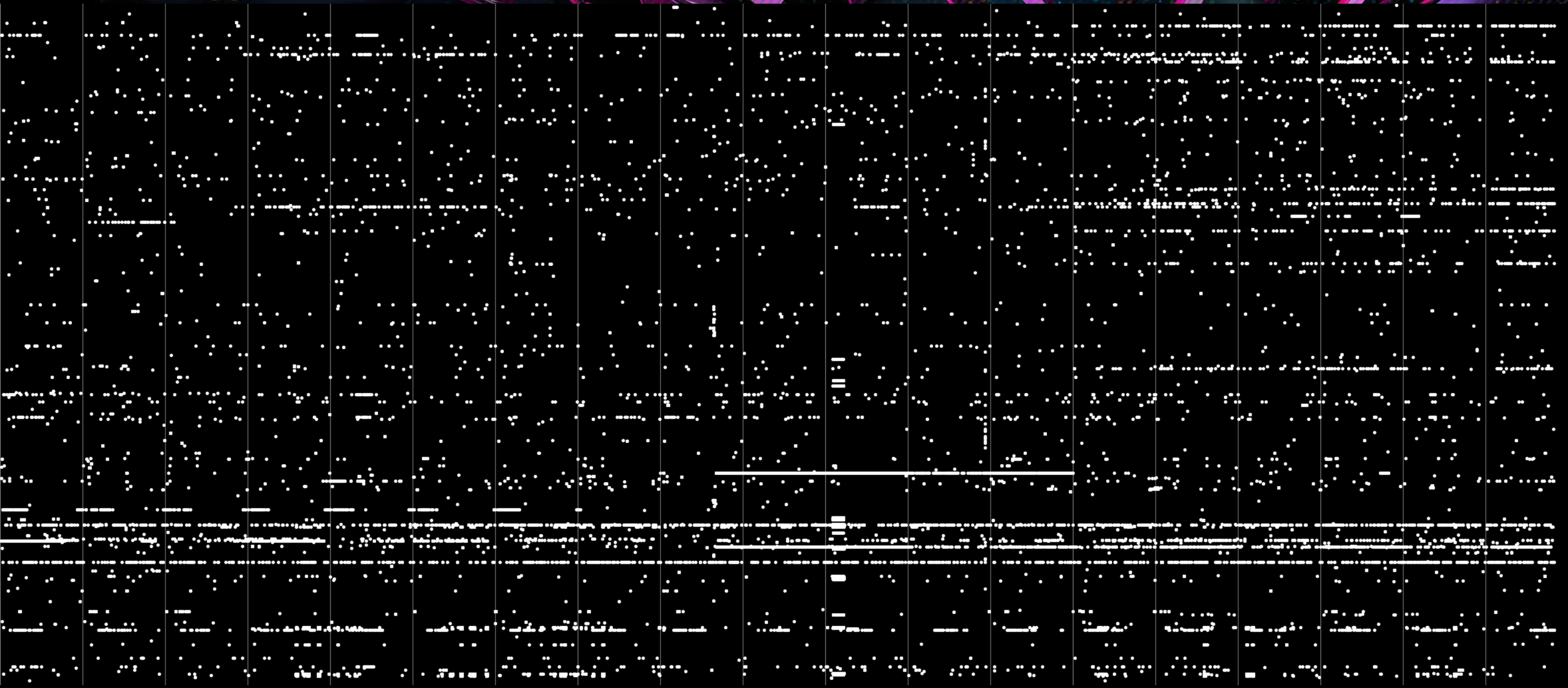
PROTECTING SMALL ORGANIZATIONS IN THE ERA OF AI BOTS



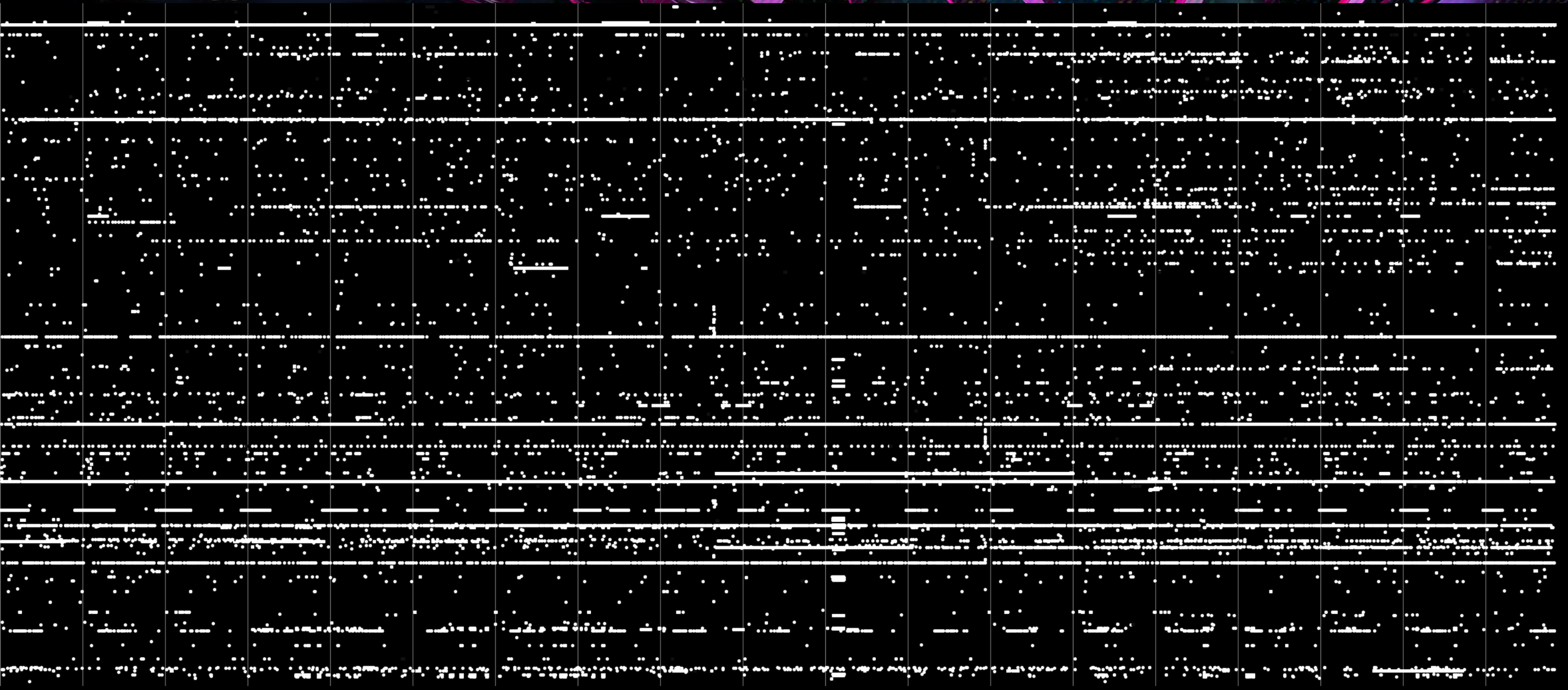
# Blocked by Daily Maximum

RAMA CARL HOETZLEIN

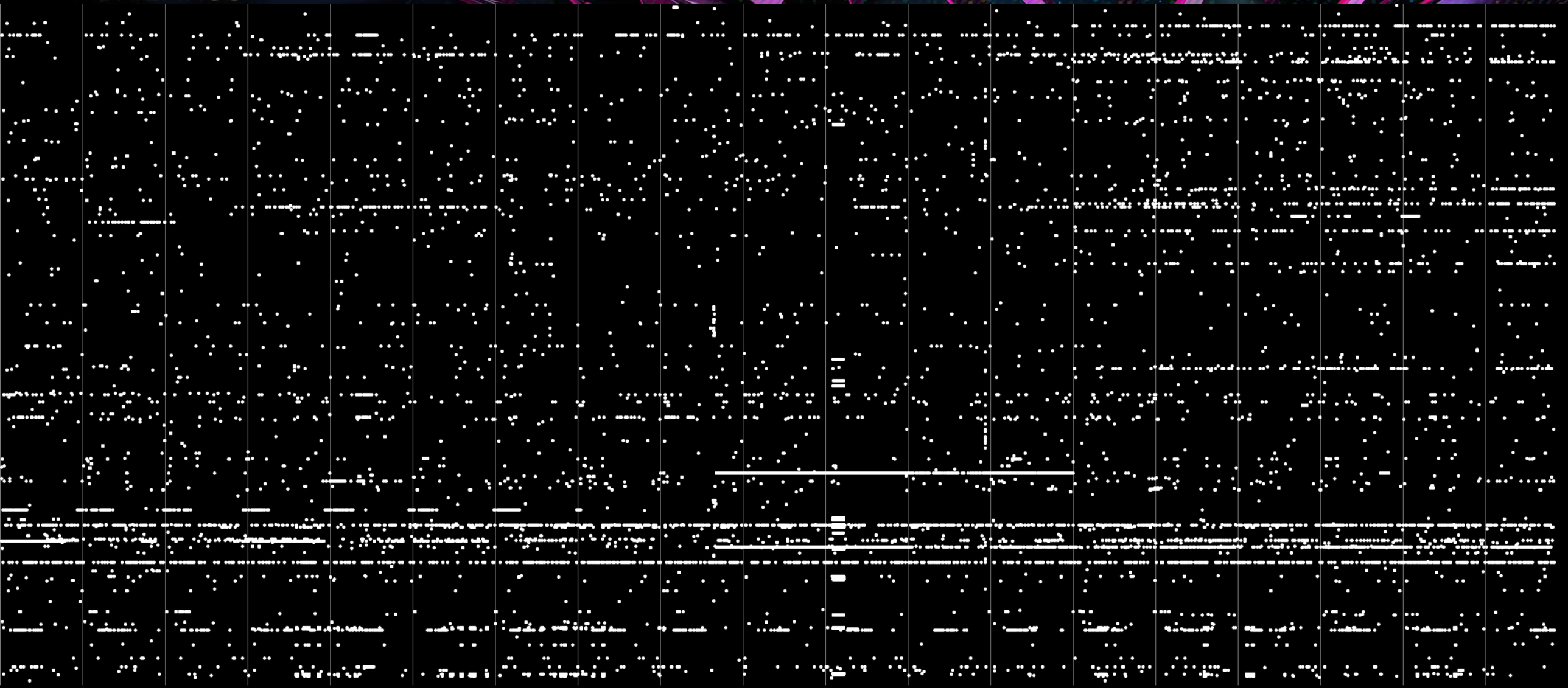
PROTECTING SMALL ORGANIZATIONS IN THE ERA OF AI BOTS



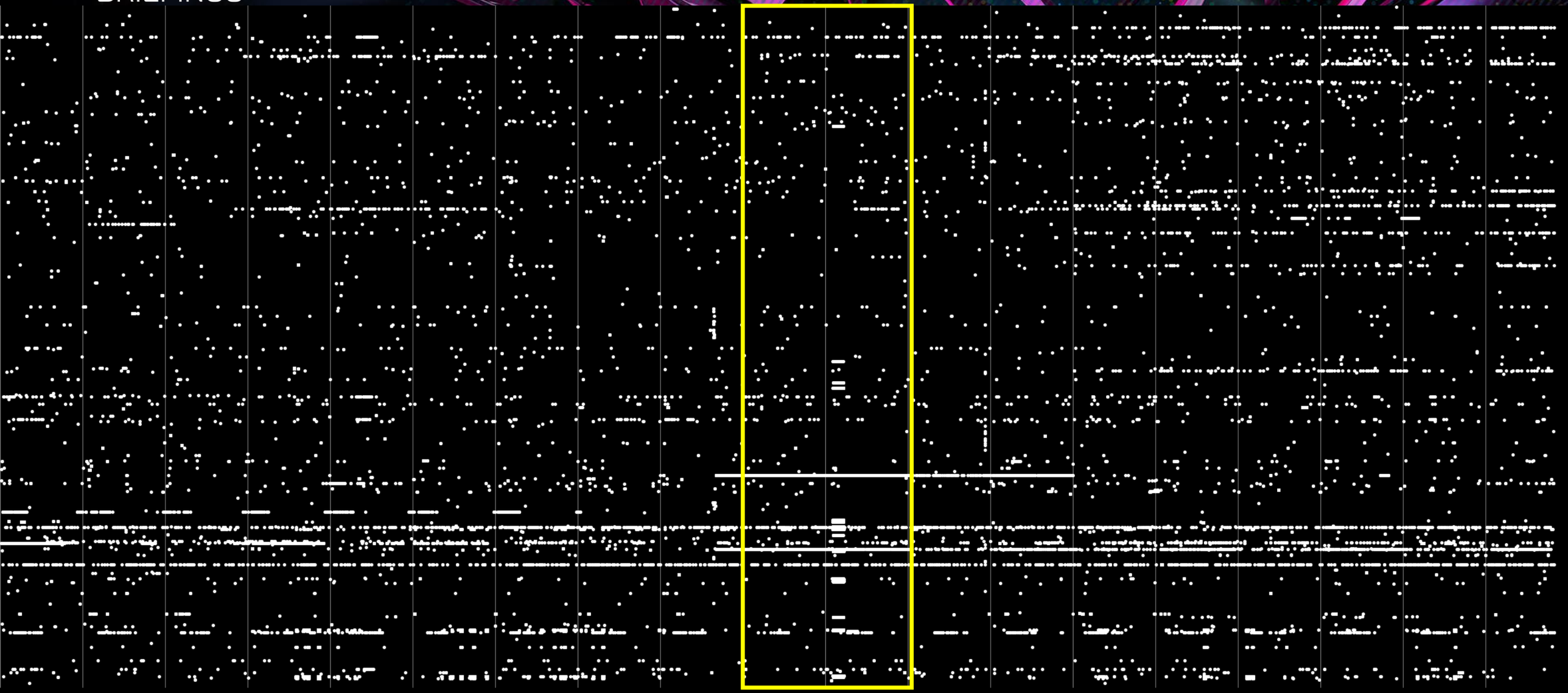
## Cumulative Filtered Results



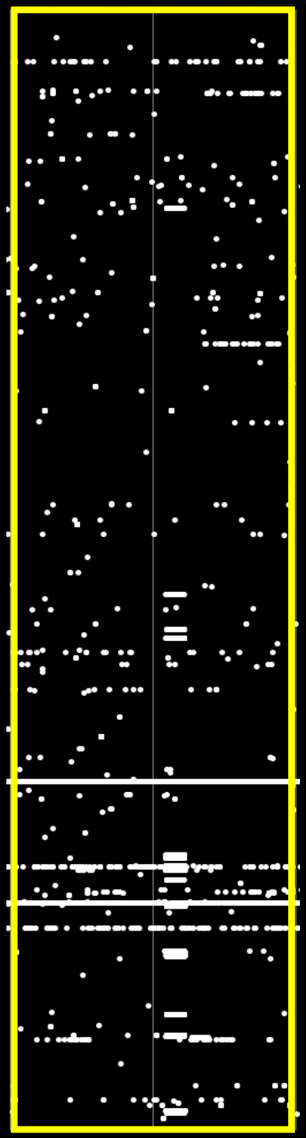
## Original Traffic

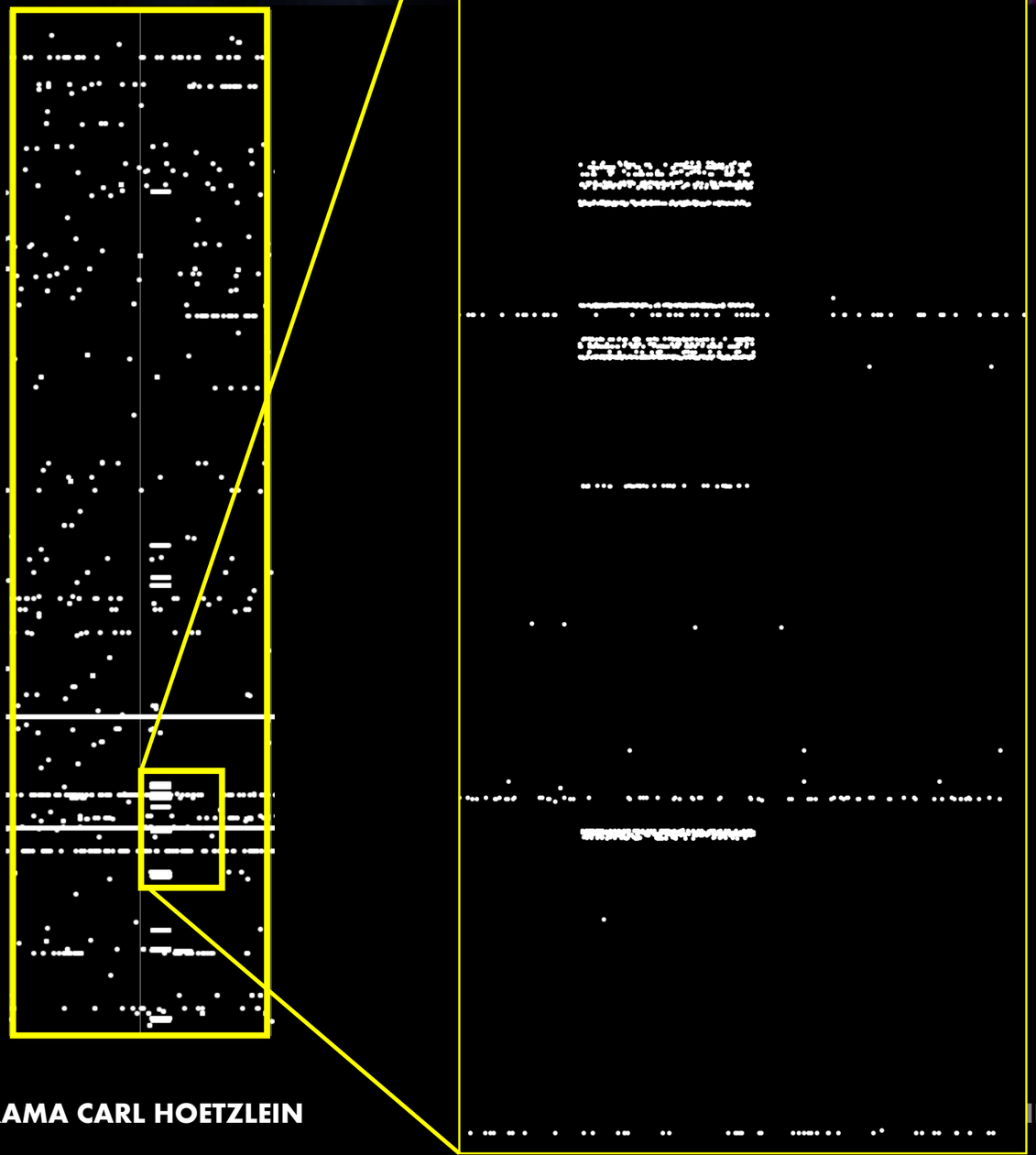


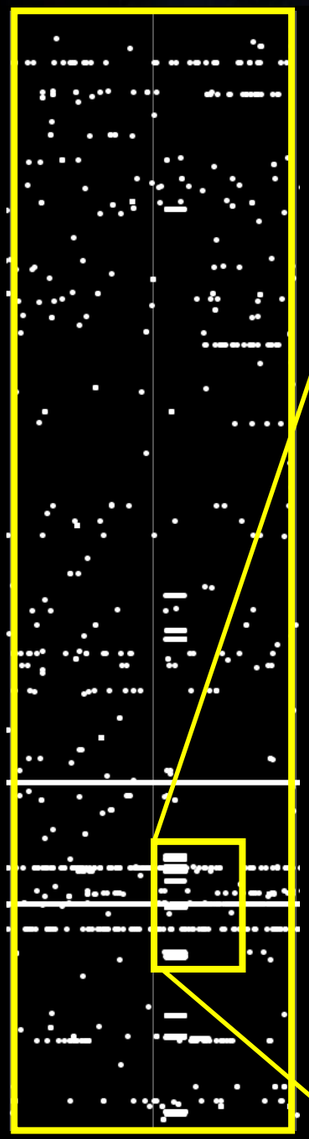
## Cumulative Filtered Results

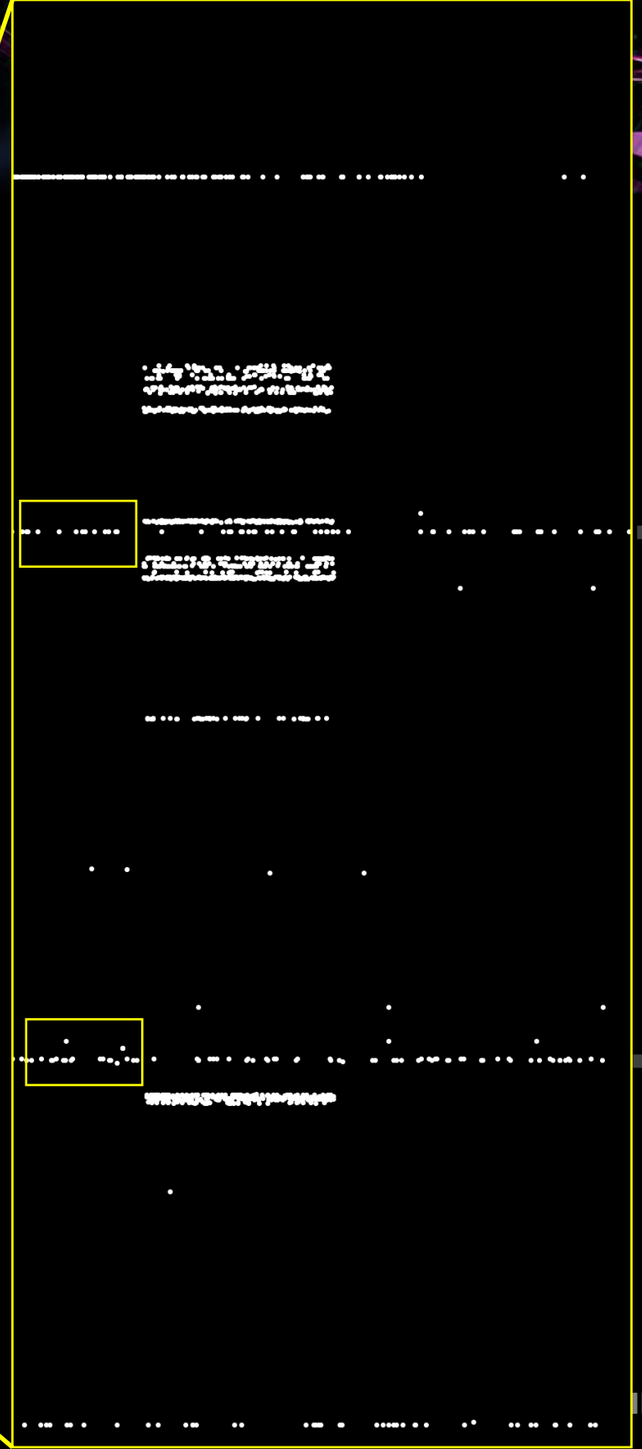


## Cumulative Filtered Results









Single IP

Multiple IPs

Group of machines within the same Class C subnet requesting multiple pages around the same time.

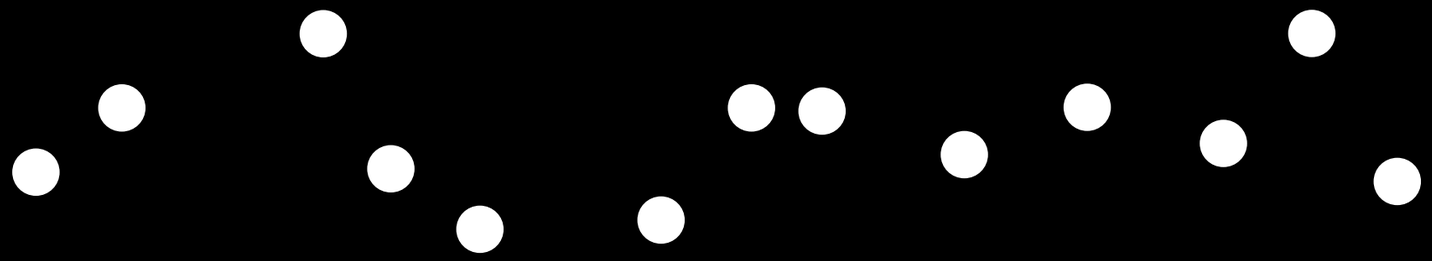
| | IP | # pages | unique pages | unique ratio | # days | max consecutive | daily hits min | daily hits max | daily range min hrs | daily range max hrs | frequency min ppm | frequency max ppm | page |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 38.123.121.143 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | / |
| 4 | 38.153.133.35 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/2982 |
| 5 | 38.170.161.142 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/101 |
| 6 | 38.170.169.92 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/928 |
| 7 | 38.170.189.155 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/63 |
| 8 | 38.170.190.30 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/2981 |
| 9 | 38.202.3.106 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /queries |
| 10 | 39.107.87.112 | 3 | 3 | 1 | 0 | 1 | 3 | 3 | 0 | 0 | 0 | 0 | /api |
| 11 | 40.76.163.23 | 33 | 19 | 0.58 | 0 | 1 | 33 | 33 | 0.09 | 0.09 | 5.106 | 5.106 | / |
| 12 | 40.77.167.0 | 26 | 25 | 0.96 | 11.9 | 12 | 1 | 5 | 0.00 | 5.99 | 0.004 | 0.863 | /events/1523 |
| 13 | 40.77.167.1 | 23 | 23 | 1 | 18.6 | 3 | 1 | 4 | 0.06 | 0.89 | 0.002 | 0.010 | /events/586 |
| 14 | 40.77.167.2 | 7 | 7 | 1 | 5.2 | 2 | 1 | 3 | 0.07 | 0.07 | 0.002 | 0.002 | /grids/64 |
| 15 | 40.77.167.3 | 30 | 29 | 0.97 | 18.8 | 8 | 1 | 4 | 1.48 | 4.59 | 0.003 | 0.013 | /events/567 |
| 16 | 40.77.167.4 | 30 | 30 | 1 | 19.2 | 10 | 1 | 4 | 2.17 | 7.81 | 0.002 | 0.006 | /events/1175 |
| 17 | 40.77.167.5 | 4 | 4 | 1 | 2.6 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | /events/517 |
| 18 | 40.77.167.6 | 29 | 28 | 0.97 | 12.3 | 7 | 1 | 4 | 0.06 | 3.90 | 0.002 | 0.008 | /events/1411 |
| 19 | 40.77.167.8 | 27 | 25 | 0.93 | 13.1 | 6 | 1 | 6 | 0.19 | 3.84 | 0.002 | 0.008 | /about |
| 20 | 40.77.167.9 | 7 | 7 | 1 | 4.0 | 2 | 1 | 4 | 0.07 | 0.07 | 0.002 | 0.002 | /events/2138 |
| 21 | 40.77.167.10 | 6 | 6 | 1 | 2.2 | 2 | 1 | 4 | 5.90 | 5.90 | 0.004 | 0.004 | /events/919 |
| 22 | 40.77.167.11 | 39 | 39 | 1 | 18.3 | 5 | 1 | 5 | 0.08 | 4.48 | 0.002 | 0.047 | /events/1021 |
| 23 | 40.77.167.12 | 1 | 1 | 1 | 0.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/684 |
| 24 | 40.77.167.13 | 40 | 40 | 1 | 19.6 | 4 | 1 | 7 | 0.00 | 11.72 | 0.003 | 0.008 | /events/3232 |
| 25 | 40.77.167.14 | 25 | 24 | 0.96 | 17.4 | 7 | 1 | 3 | 0.04 | 1.49 | 0.002 | 0.004 | /hab_events/81 |
| 26 | 40.77.167.15 | 4 | 4 | 1 | 2.0 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | /events/380 |
| 27 | 40.77.167.16 | 12 | 12 | 1 | 9.7 | 3 | 1 | 5 | 4.75 | 4.75 | 0.006 | 0.006 | /hab_events/413 |
| 28 | 40.77.167.17 | 38 | 37 | 0.97 | 18.3 | 5 | 1 | 6 | 0.12 | 2.72 | 0.002 | 0.135 | /monitoringlocations/469 |
| 29 | 40.77.167.19 | 29 | 29 | 1 | 12.8 | 6 | 1 | 5 | 0.12 | 11.84 | 0.003 | 0.033 | /hab |
| 30 | 40.77.167.20 | 31 | 30 | 0.97 | 15.0 | 6 | 1 | 7 | 0.01 | 10.31 | 0.003 | 0.397 | /events/789 |
| 31 | 45.83.65.46 | 1 | 1 | 1 | 0.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | / |
| 32 | 45.83.65.56 | 1 | 1 | 1 | 0.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | / |
| 33 | 45.83.65.158 | 1 | 1 | 1 | 0.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | / |
| 34 | 45.83.66.37 | 1 | 1 | 1 | 0.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | / |
| 35 | 45.89.148.2 | 3 | 3 | 1 | 14.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/1778 |
| 36 | 45.89.148.3 | 7 | 7 | 1 | 16.2 | 1 | 1 | 4 | 2.66 | 2.66 | 0.009 | 0.009 | /queries?action=index&controll |
| 37 | 45.89.148.4 | 4 | 3 | 0.75 | 6.8 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | /queries/new?q%5Bs%5D=valu |
| 38 | 45.89.148.5 | 6 | 6 | 1 | 5.5 | 2 | 1 | 2 | 0 | 0 | 0 | 0 | /queries?action=index&controll |
| 39 | 45.89.148.6 | 3 | 3 | 1 | 17.0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /queries/new?q%5Bs%5D=valu |
| 40 | 45.89.148.7 | 4 | 4 | 1 | 14.5 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/1901 |
| 41 | 45.89.148.8 | 2 | 2 | 1 | 12.7 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /queries/new?q%5Bs%5D=valu |
| 42 | 45.89.148.9 | 3 | 3 | 1 | 6.7 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /monitoringlocations/443 |
| 43 | 45.89.148.10 | 2 | 2 | 1 | 1.3 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /events/1478 |
| 44 | 45.89.148.11 | 4 | 4 | 1 | 14.7 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /testing_events/246 |
| 45 | 45.89.148.12 | 3 | 3 | 1 | 9.7 | 1 | 1 | 2 | 0 | 0 | 0 | 0 | /queries?action=index&controll |
| 46 | 45.89.148.13 | 4 | 4 | 1 | 13.5 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | /queries?action=index&controll |

# Subnet Hashing

Aggregate all page hits across a subnet and *then* perform scoring metrics.

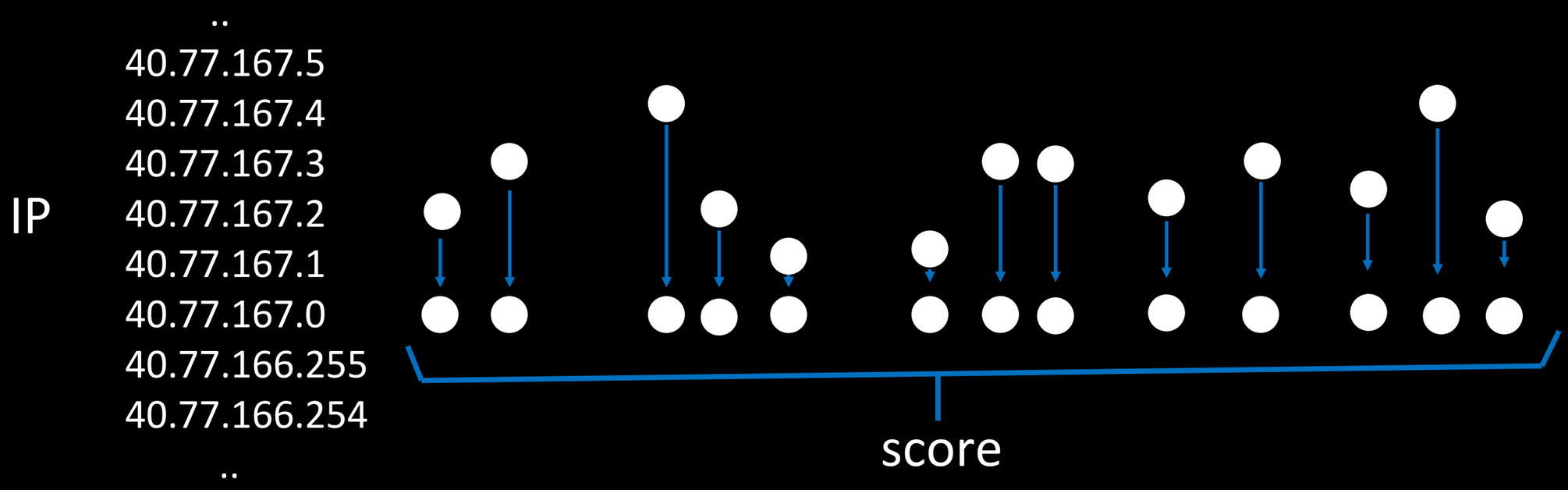
IP

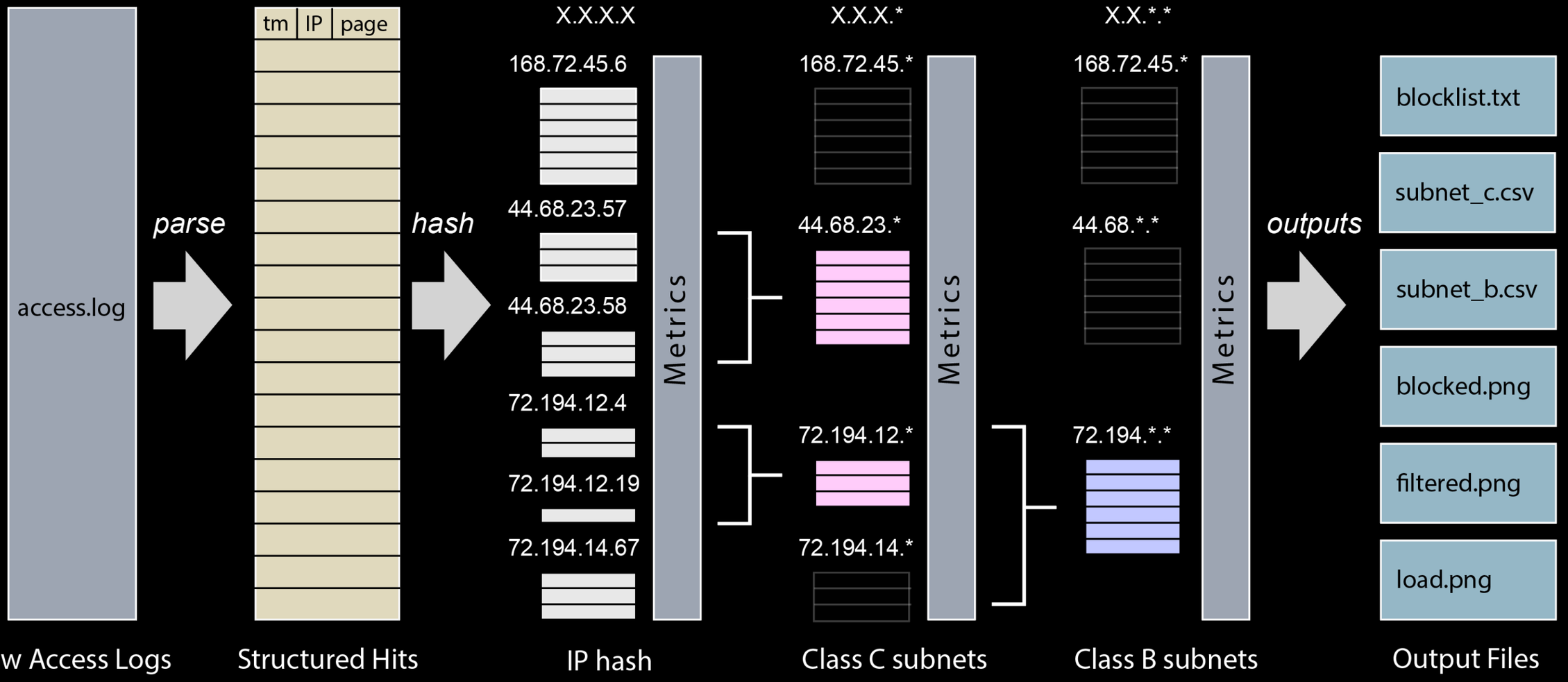
..
40.77.167.5
40.77.167.4
40.77.167.3
40.77.167.2
40.77.167.1
40.77.167.0
40.77.166.255
40.77.166.254
..



# Subnet Hashing

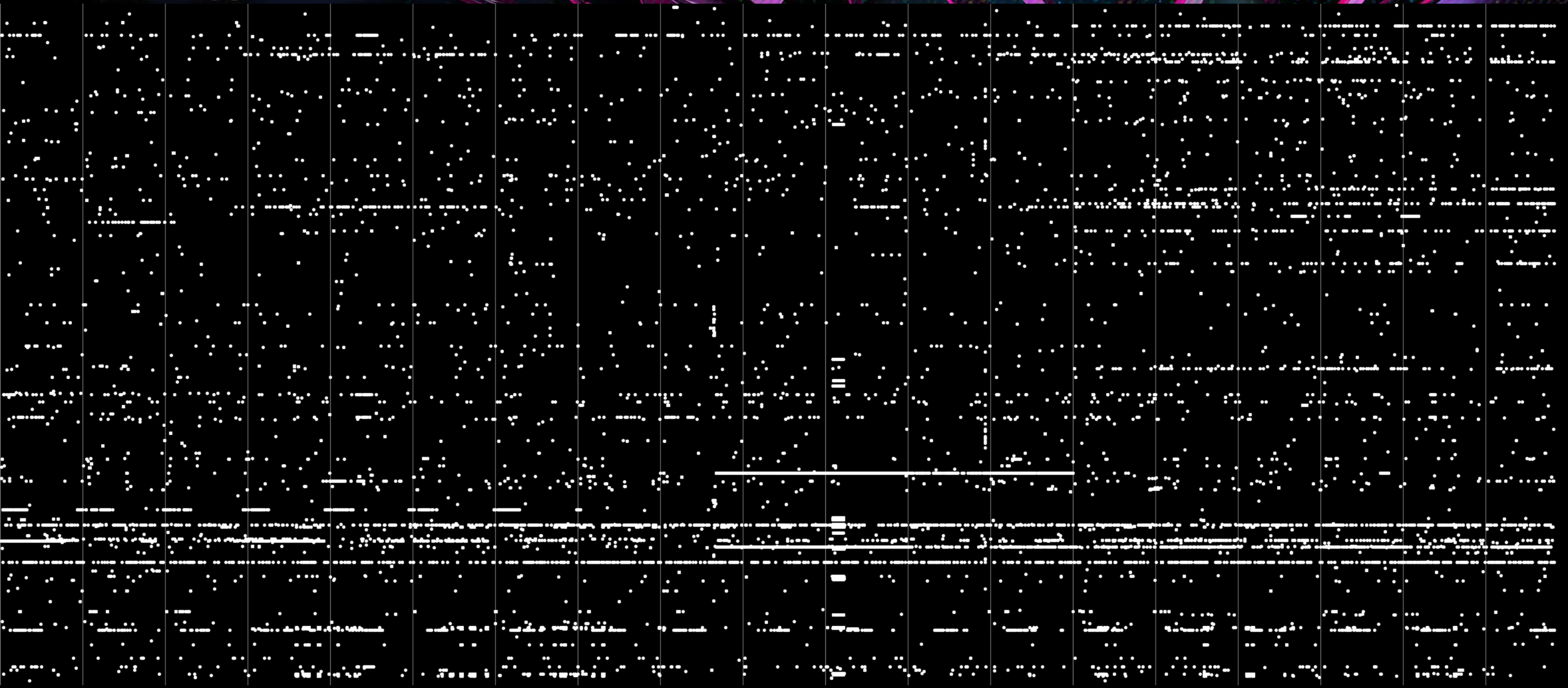
Aggregate all page hits across a subnet and *then* perform scoring metrics.



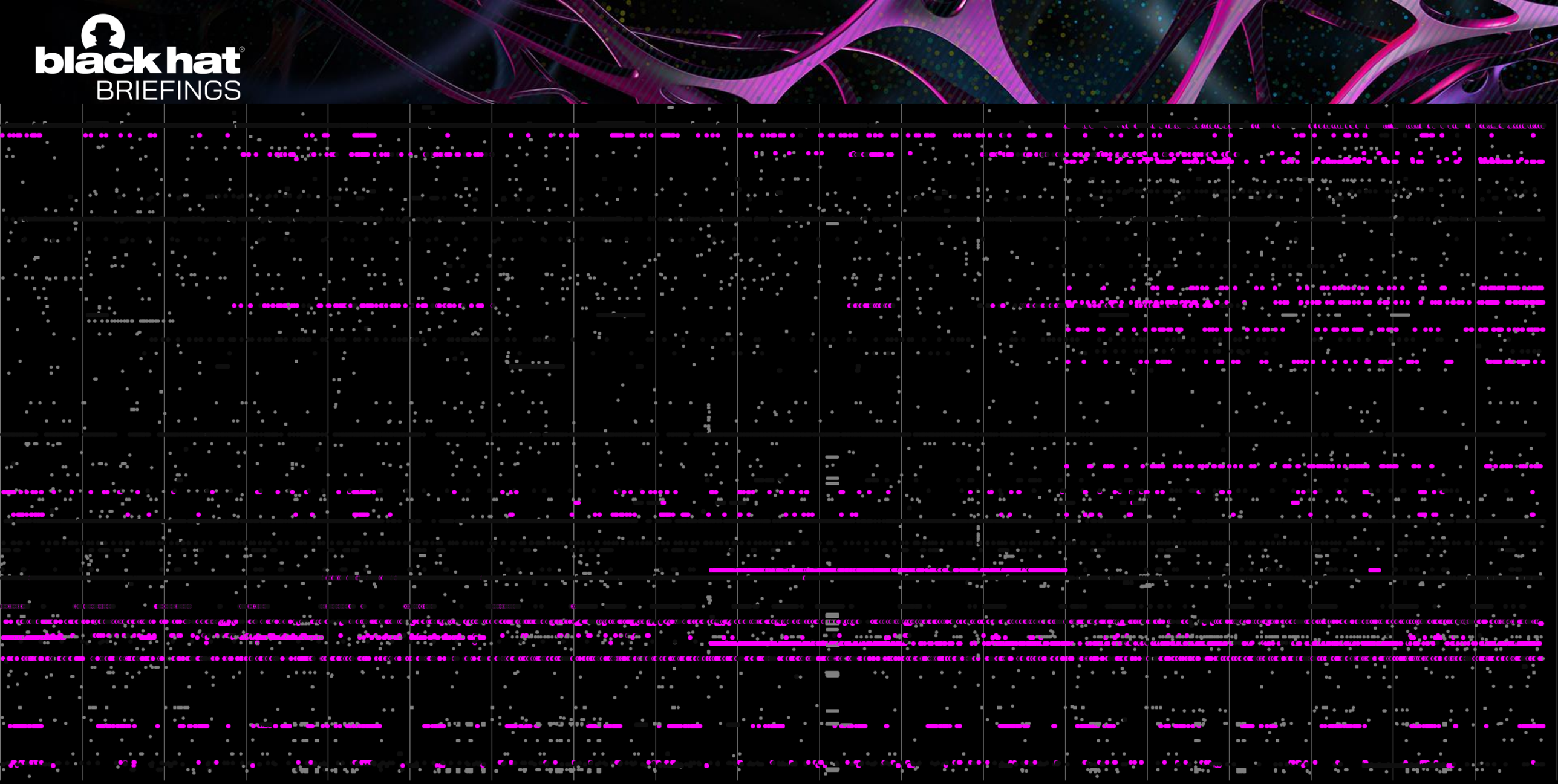



# Hierarchical IP Hashing with Metric Scoring

# Final Results



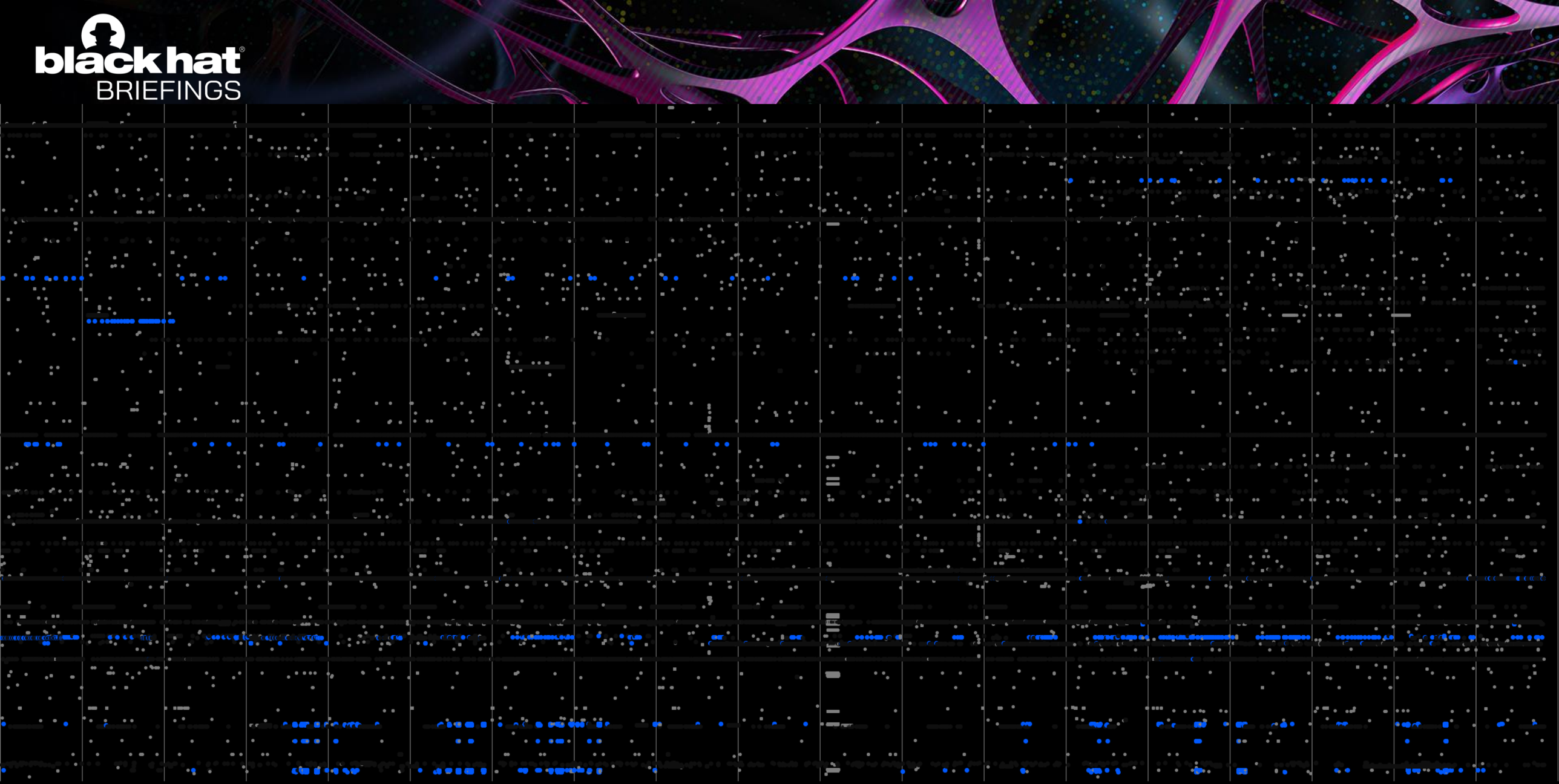
## Filtered Result – Prior to Subnet Hashing



# Blocking Class C Subnets

RAMA CARL HOETZLEIN

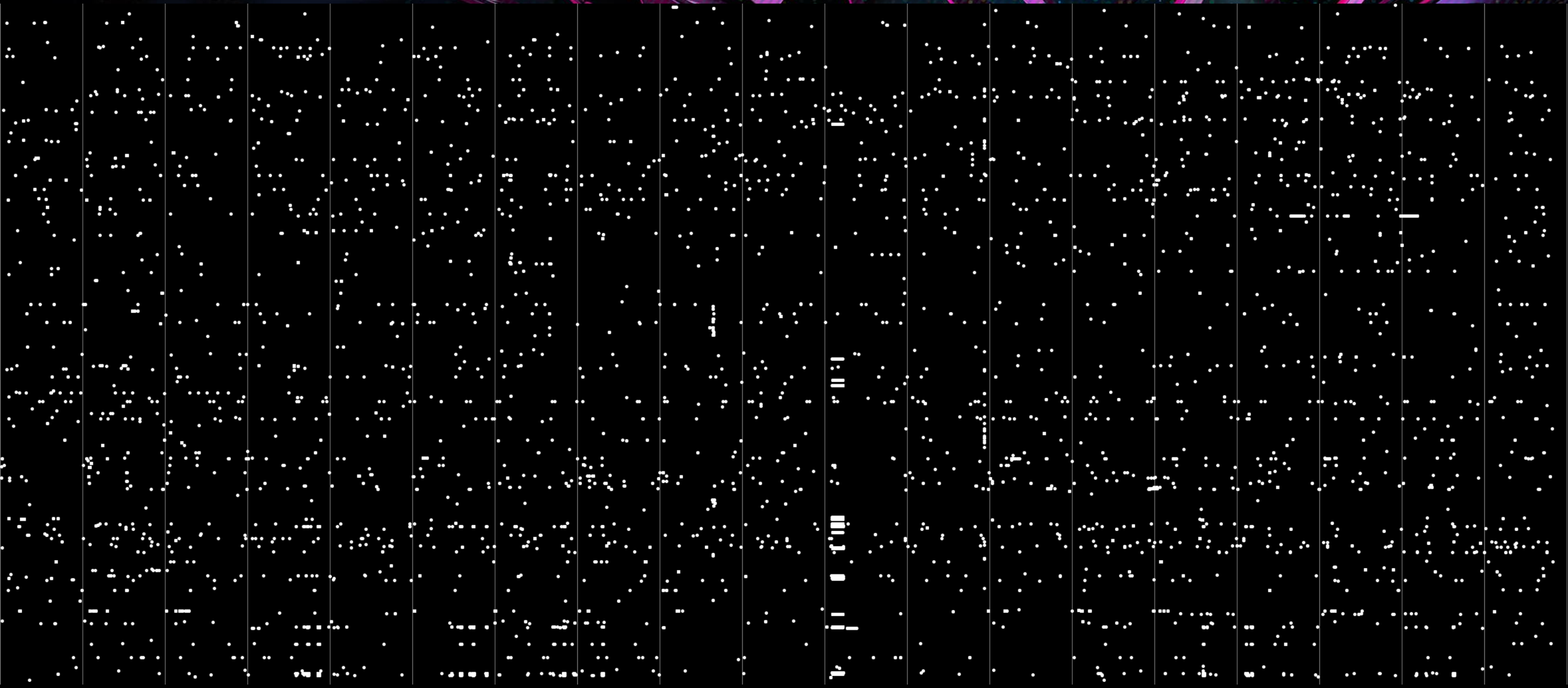
PROTECTING SMALL ORGANIZATIONS IN THE ERA OF AI BOTS



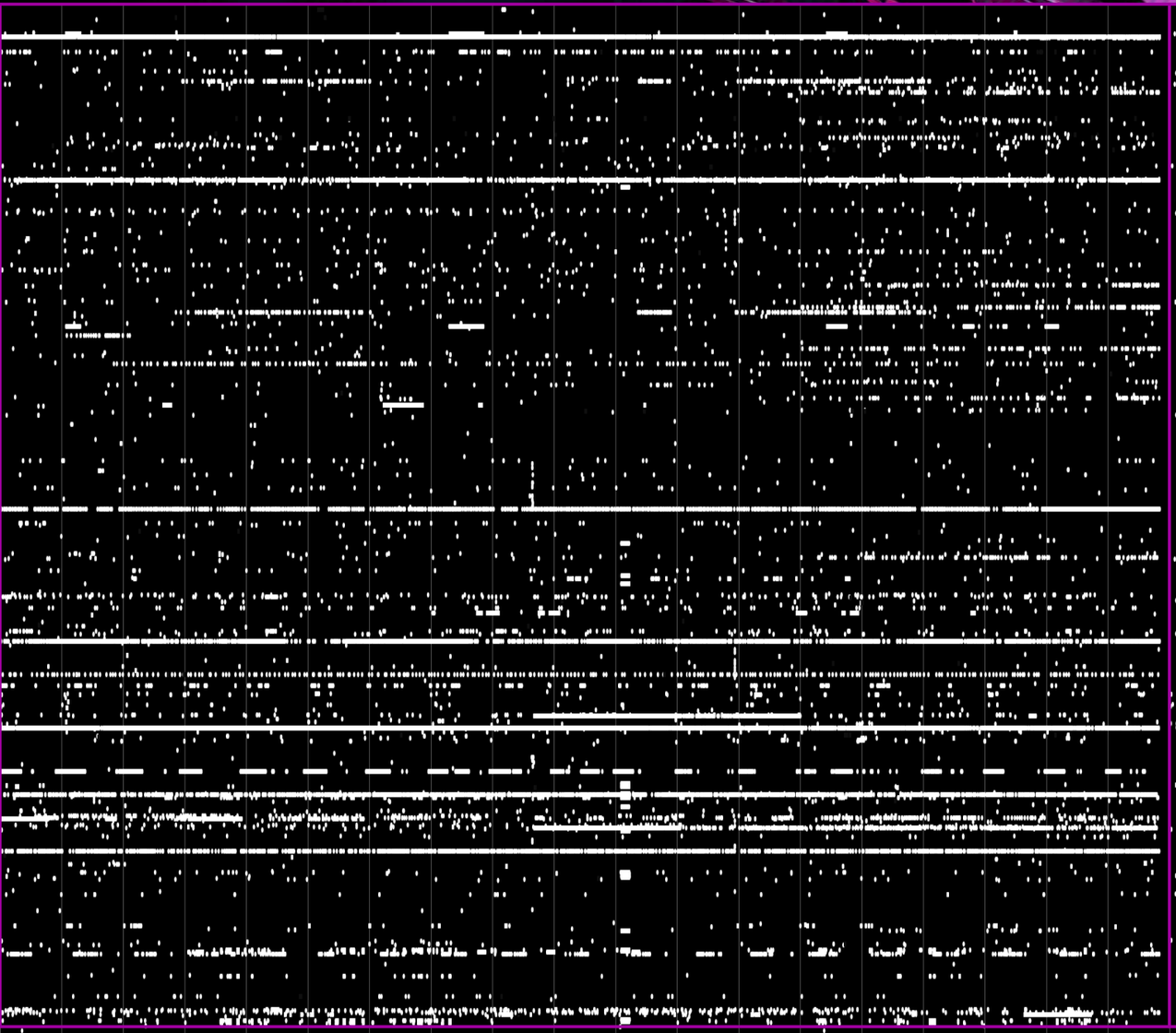
# Blocking Class B Subnets

RAMA CARL HOETZLEIN

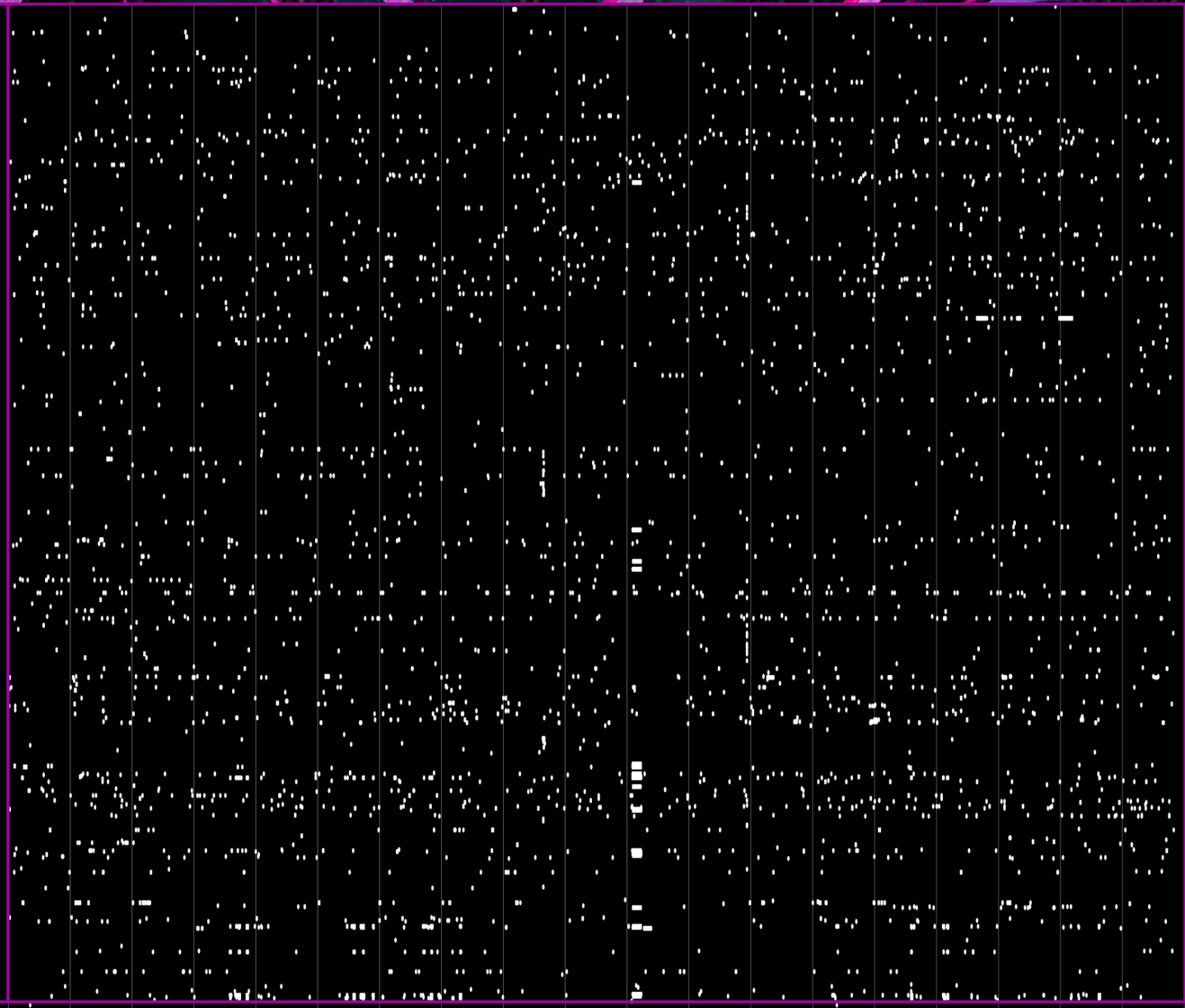
PROTECTING SMALL ORGANIZATIONS IN THE ERA OF AI BOTS



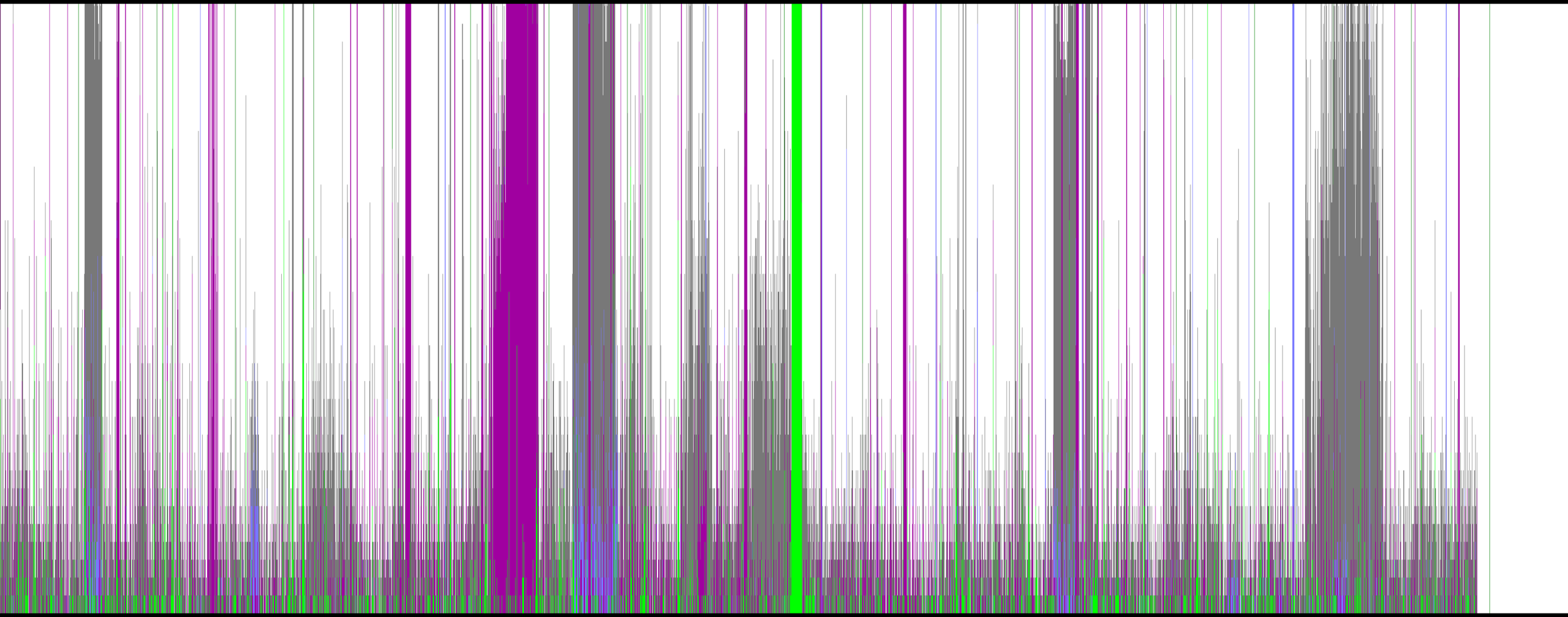
## Final Result



Original Traffic



Final Result



## Estimated Load Analysis

Original

C Filtering

B Filtering

Final Server Load

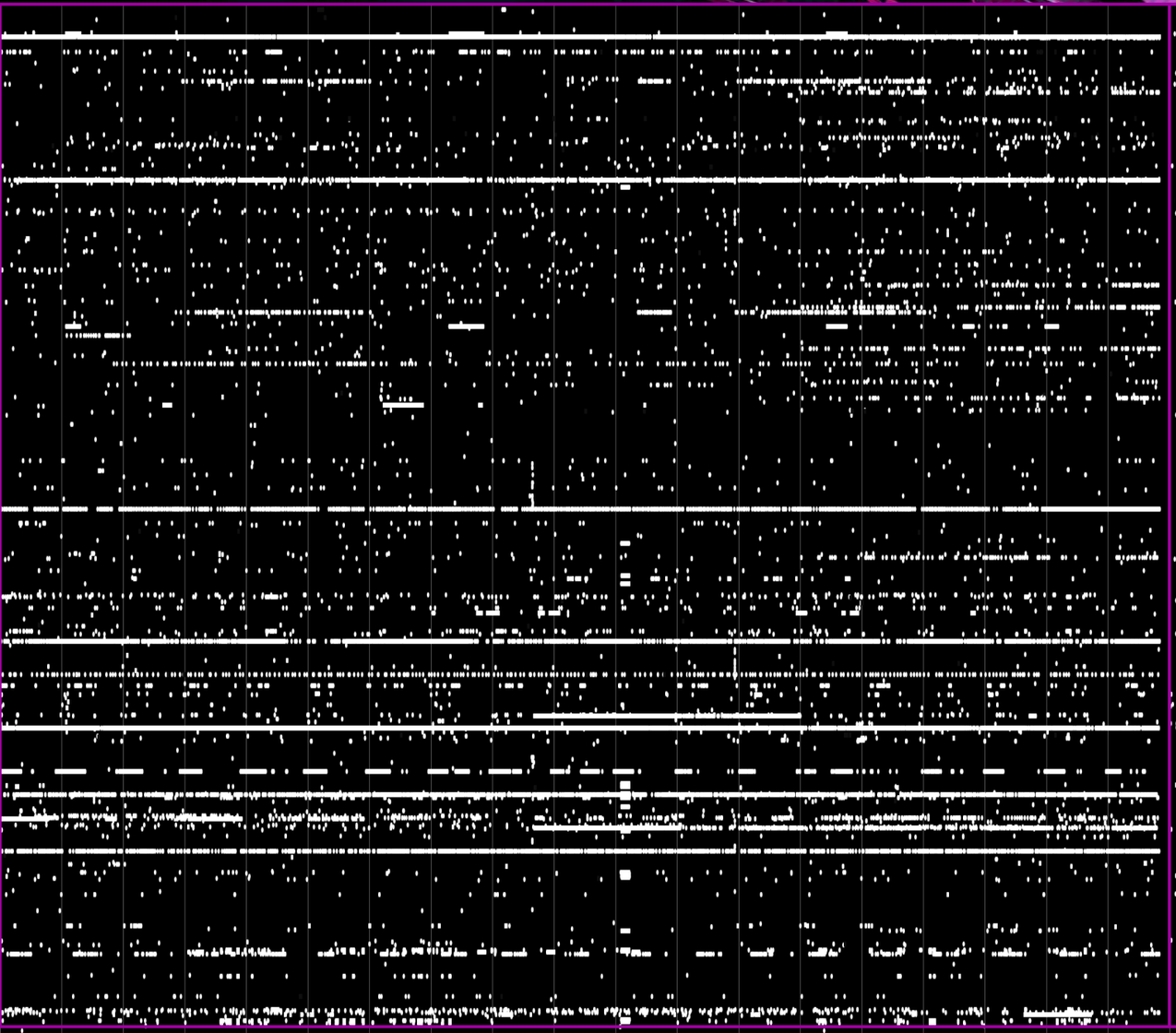
# Results

94% reduction in traffic

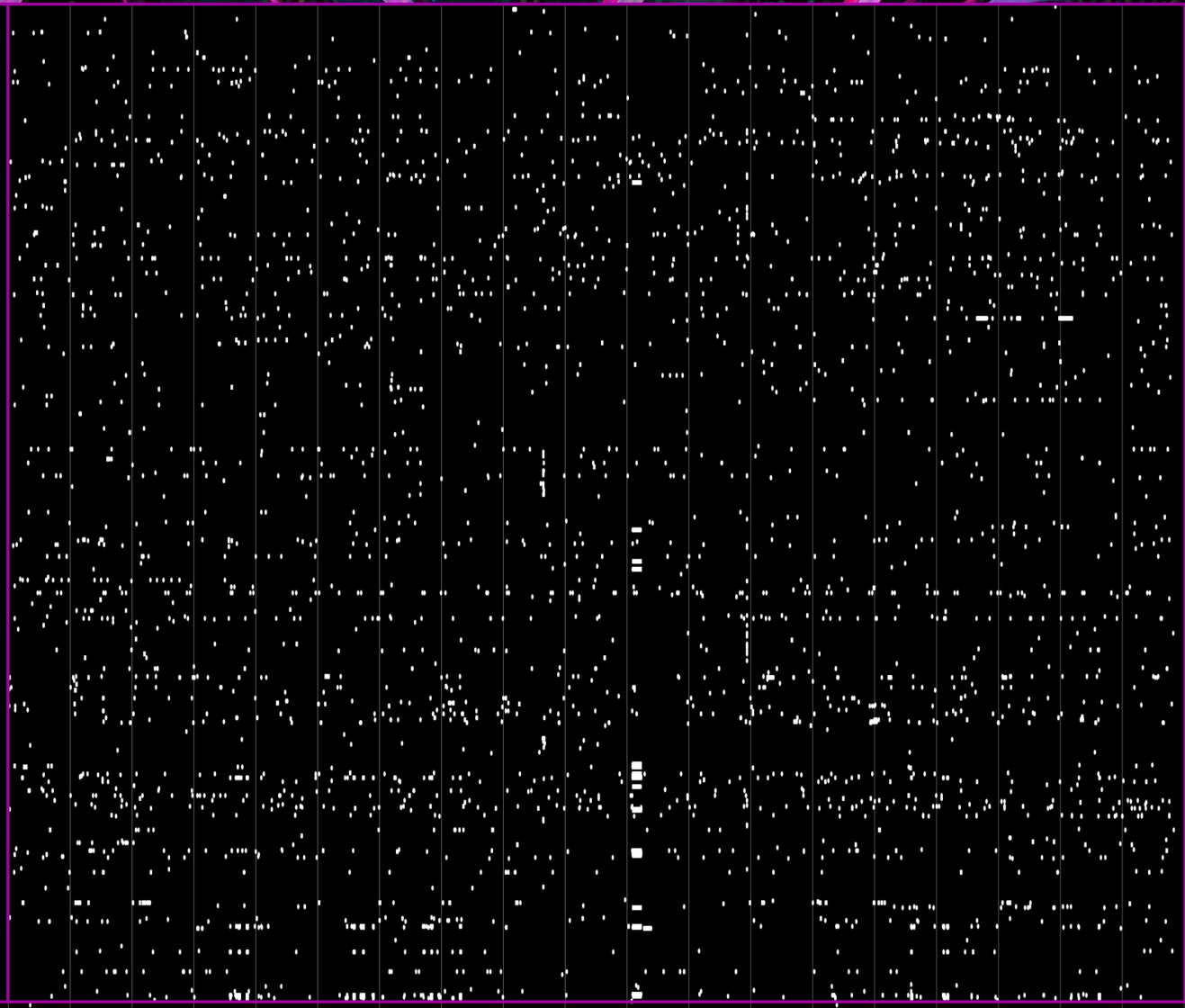
# Results

## 94% reduction in traffic

| Filter | Workload (ave. requests per min) | Stage Reduction % | Cumulative Reduction % |
|---|---|---|---|
| None | 10.7 | | |
| Throttling | 7.1 | 33% | 33% |
| Consecutive | 6.2 | 9% | 42% |
| Daily range | 5.2 | 9% | 51% |
| Daily max | 4.9 | 3% | 54% |
| C Subnet | 3.4 | 14% | 68% |
| B Subnet | 0.6 | 26% | 94% |



Original Traffic



Final Result

# Protecting Small Organizations

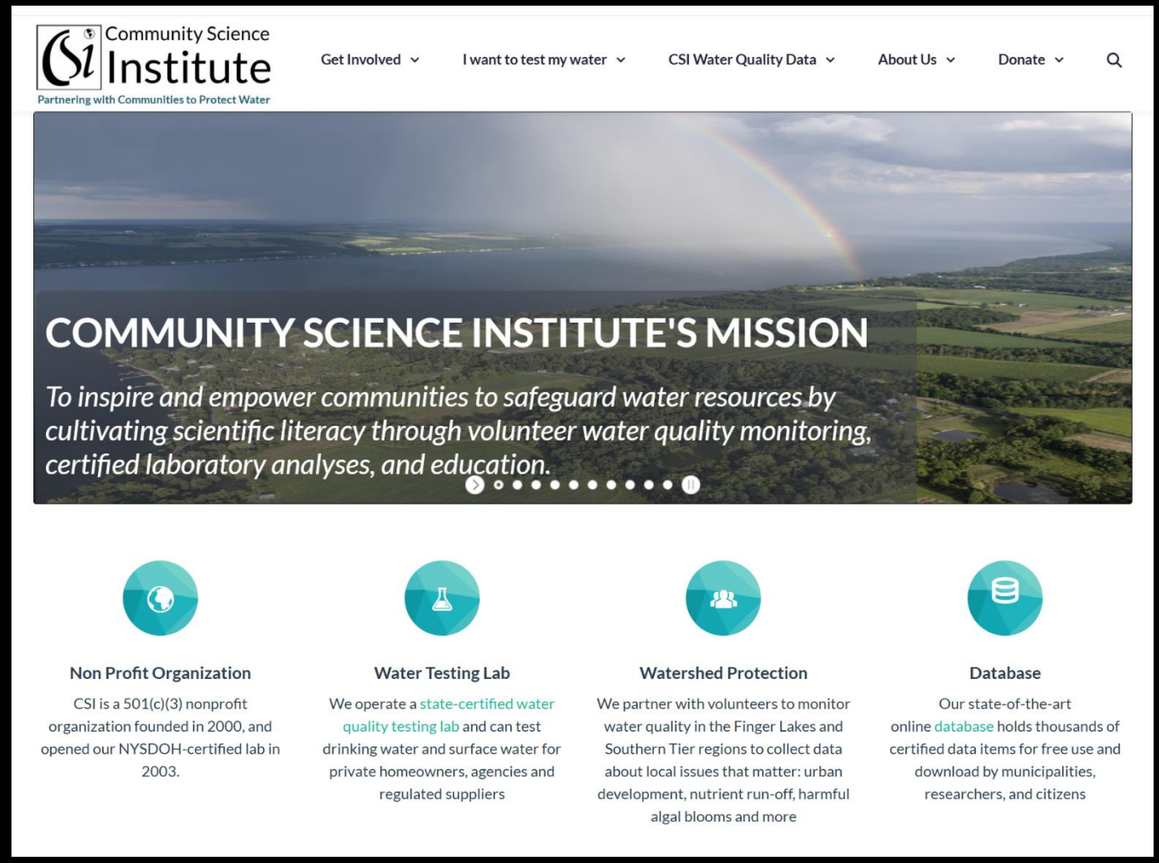
We found that - even when well behaved and observing rate limits - the sheer volume of AI bot requests can overwhelm the servers of small organizations.

# Protecting Small Organizations

## Policy

"Our water quality data is available to the public for free.

We prefer to have a human-in-the-loop, and discourage AI crawlers so that our servers remain responsive to our human users."



# Protecting Small Organizations

Grants for non-profits and small orgs often depend on viewership statistics for new or renewed funding.

# Protecting Small Organizations

Grants for non-profits and small orgs often depend on viewership statistics for new or renewed funding.

**LOGRIP** provides an upper bound on real human views, with blocked/permitted stats per day, at least better than raw traffic stats.

| Date | All | Blocked | Allowed | Reduction |
|---|---|---|---|---|
| 7/16/2025 | 11359 | 10807 | 552 | 95.1% |
| 7/17/2025 | 13476 | 12965 | 512 | 96.2% |

# Conclusions

- Understand the extent of AI crawler & bot activity
- Defend small organizations (single machines)
  from large organizations (many machines in data centers)!
- Be able to specify defense policy
- Know (to the extent possible) the implications
  of those policies
- Do all of this easily, cheaply and open source

**LOGRIP**

A simple, lightweight, open source tool for generating blocklists *and* policy visualizations based on access logs.

# New Tool

# Running **LOGRIP**

https://github.com/quantasci/logrip

**Input:**
access log
config file (log format, policy)

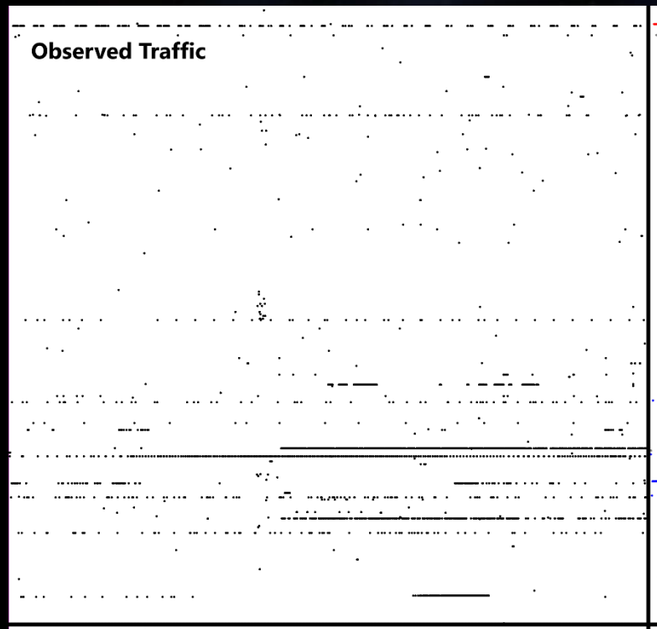
**Output:**
blocklist
B-subnet list
C-subnet list
full IP list
policy visualizations
load estimation

```
rama@Precision-Tower-3620:/diska/codes/build/logrip$ ./logrip example_log.txt
LOGRIP
Copyright (c) 2024-2025, Quanta Sciences, Rama Hoetzlein
Apache 2.0 License

Loading config: /diska/codes/logrip/assets/ruby.conf
 Using format: * Started {GET} "{PAGE}" for {X.X.X.X} at {YYYY-MM-DD} {HH:MM:SS}

Reading log: /diska/codes/logrip/assets/example_log.txt
 5%. 236 read, 0 skipped.
 10%. 554 read, 1 skipped.
 15%. 876 read, 1 skipped.
 20%. 1183 read, 2 skipped.
 25%. 1481 read, 2 skipped.
 30%. 1786 read, 2 skipped.
 35%. 2076 read, 2 skipped.
 40%. 2386 read, 4 skipped.
 45%. 2723 read, 5 skipped.
 50%. 3020 read, 5 skipped.
 55%. 3310 read, 5 skipped.
 60%. 3606 read, 5 skipped.
 65%. 3918 read, 5 skipped.
 70%. 4234 read, 5 skipped.
 75%. 4554 read, 6 skipped.
 80%. 4843 read, 23 skipped.
 85%. 5143 read, 25 skipped.
 90%. 5447 read, 26 skipped.
 95%. 5747 read, 28 skipped.
 100%. 6032 read, 28 skipped.

Construct IP Hash.
Preparing Days.
Start date: 2025-01-23 00:00:00
End date:   2025-01-24 23:59:59
Total days: 2
Processing IPs.
Constructing C-Subnets.
Constructing B-Subnets.
Constructing A-Subnets.
Processing IPs. C-Subnets.
Processing IPs. B-Subnets.
Computing Blocklist.
Writing Blocklist.
Writing IPs (B-Subnets)... 189 ips.
Writing IPs (C-Subnets)... 233 ips.
Writing IPs (All Mach)... 1555 ips.
Writing Pages.
Writing Hits.
Writing Visualizations.
Writing Loads.
Done.
rama@Precision-Tower-3620:/diska/codes/build/logrip$
```

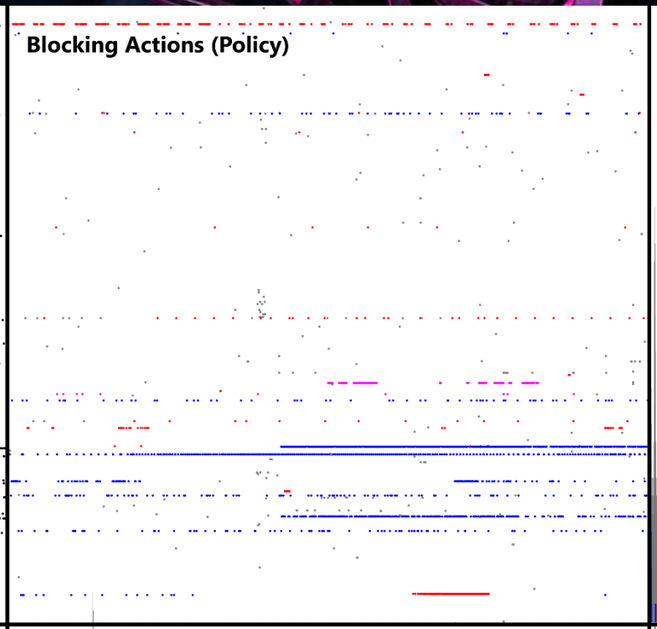
Features:

- Open source
- Cmd line based
- Read any log format
- Config policy settings
- Fast.
  150k log in 10 sec

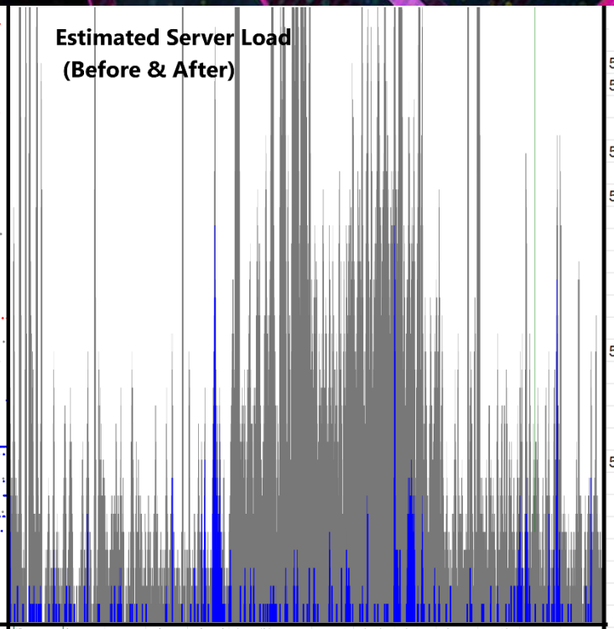
## Observed Traffic



## Blocking Actions (Policy)



## Estimated Server Load (Before & After)



## Filtered Traffic

# LOGRIP

## All Output Products

## Metrics by IP

| IP | ip_cnt | page_c | uniq_c | uniq_r | elapse | max_c | num_r | min_hi | min_h | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.136.111.218 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | | | | |
| 4.151.218.216 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /owa/auth/logon.aspx |
| 4.227.36.31 | 1 | 422 | 414 | 0.98 | 0.06 | 1 | 0 | 422 | 1.168 | 4.658 | 422 | 1.168 | 4.658 /bmi/monitoringlocations/629 |
| 4.227.36.50 | 1 | 41 | 41 | 1 | 0 | 1 | 0 | 41 | 0.026 | 20.51 | 41 | 0.026 | 20.51 /queries/?page=5&q[s]=date%2 |
| 4.227.36.122 | 1 | 6 | 6 | 1 | 0 | 1 | 0 | 6 | 0.003 | 25 | 6 | 0.003 | 25 /queries/new?q%5Bs%5D=dat |
| 5.102.173.71 | 1 | 12 | 12 | 1 | 0.93 | 2 | 0 | 1 | 12.9 | 0.009 | 11 | 12.9 | 0.009 /events/3085 |
| 5.181.190.248 | 1 | 11 | 1 | 0.09 | 0.99 | 2 | 0 | 1 | 14.89 | 0.008 | 10 | 14.89 | 0.008 / |
| 8.48.71.250 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringsets/7 |
| 8.211.42.174 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /dns-query?dns=pHkBAAABAAA |
| 17.241.75.55 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/842 |
| 17.241.75.92 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /sites/117 |
| 17.241.75.106 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/499 |
| 17.241.75.110 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/344 |
| 17.241.75.127 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/1519 |
| 17.241.219.9 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/1989 |
| 17.241.219.12 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /hab_events/701 |
| 17.241.219.24 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/120 |
| 17.241.219.44 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringlocations/382 |
| 17.241.219.52 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /hab_events/688 |
| 17.241.219.114 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /hab_events/655 |
| 17.241.219.145 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringlocations/512 |
| 17.241.219.172 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringlocations/685 |
| 17.241.219.182 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /hab_events/169 |
| 17.241.227.19 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/1662 |
| 17.241.227.65 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/3107 |
| 17.241.227.124 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /hab_events/667 |
| 17.241.227.154 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringlocations/562 |
| 17.241.227.167 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/2872 |
| 17.241.227.235 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /events/2572 |
| 18.97.9.169 | 1 | 175 | 175 | 1 | 0.12 | 1 | 0 | 175 | 2.738 | 1.024 | 175 | 2.738 | 1.024 /hab_events/92 |
| 20.49.136.28 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringsets/7 |
| 20.159.64.138 | 1 | 4 | 3 | 0.75 | 0 | 1 | 0 | 4 | 3E-04 | 5.294 | 4 | 3E-04 | 5.294 /hab |
| 23.146.184.101 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 / |
| 24.59.56.143 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 /monitoringlocations/504 |
| 27.150.86.197 | 1 | 2 | 2 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 /queries/new?q%5Bs%5D=ana |
| 31.13.224.222 | 1 | 2 | 1 | 0.5 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 /.env |

## Metrics by B-Subnet

| IP | ip_cnt | page_cnt | uniq_cnt | uniq_ratio | elapsed(d | max_cons | num_r | | | | | | pm |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3.136.*.* | 1 | 1 | 1 | 1 | 0 | 1 | | | | | | | 0 |
| 4.151.*.* | 1 | 1 | 1 | 1 | 0 | 1 | | | | | | | 0 |
| 4.227.*.* | 3 | 469 | 461 | 0.98 | 0.37 | 2 | | | | | | | 338 |
| 5.102.*.* | 1 | 12 | 12 | 1 | 0.93 | 2 | 0 | 1 | 12.90278 | 0.009064 | 11 | 12.90278 | 0.009064 |
| 5.181.*.* | 1 | 11 | 1 | 0.09 | 0.99 | 2 | 0 | 1 | 14.89167 | 0.007656 | 10 | 14.89167 | 0.007656 |
| 8.48.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 8.211.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 17.241.*.* | 20 | 20 | 20 | 1 | 1.11 | 2 | 0 | 8 | 1.360278 | 0.007781 | 12 | 8.150278 | 0.048832 |
| 18.97.*.* | 1 | 175 | 175 | 1 | 0.12 | 1 | 0 | 175 | 2.738333 | 1.024232 | 175 | 2.738333 | 1.024232 |
| 20.49.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 20.159.*.* | 1 | 4 | 3 | 0.75 | 0 | 1 | 0 | 4 | 0.000278 | 5.294118 | 4 | 0.000278 | 5.294118 |
| 23.146.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 24.59.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 27.150.*.* | 1 | 2 | 2 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 31.13.*.* | 1 | 2 | 1 | 0.5 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 34.77.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 34.79.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 35.159.*.* | 1 | 5 | 3 | 0.6 | 0 | 1 | 0 | 5 | 0.017222 | 2.033899 | 5 | 0.017222 | 2.033899 |
| 40.77.*.* | 38 | 103 | 98 | 0.95 | 1.09 | 2 | 0 | 17 | 1.503055 | 0.061053 | 86 | 21.63861 | 0.108389 |
| 42.58.*.* | 1 | 2 | 2 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 42.179.*.* | 1 | 2 | 2 | 1 | 0 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 44.220.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 45.20.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 45.55.*.* | 1 | 8 | 8 | 1 | 0 | 1 | 0 | 8 | 0 | 0 | 8 | 0 | inf |
| 45.66.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 45.79.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 45.89.*.* | 132 | 197 | 166 | 0.84 | 0.7 | 1 | 0 | 197 | 16.17667 | 0.193874 | 197 | 16.17667 | 0.193874 |
| 45.93.*.* | 140 | 210 | 172 | 0.82 | 0.7 | 1 | 0 | 210 | 15.61083 | 0.208102 | 210 | 15.61083 | 0.208102 |
| 45.156.*.* | 5 | 5 | 2 | 0.4 | 0.44 | 1 | 0 | 5 | 4.940833 | 0.00638 | 5 | 4.940833 | 0.00638 |
| 45.200.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 46.19.*.* | 1 | 2 | 1 | 0.5 | 0.53 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 46.250.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 47.79.*.* | 2 | 2 | 2 | 1 | 0.51 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 47.82.*.* | 2 | 3 | 3 | 1 | 0 | 1 | 0 | 3 | 0.000278 | 1.578948 | 3 | 0.000278 | 1.578948 |
| 47.128.*.* | 8 | 8 | 8 | 1 | 0.1 | 1 | 0 | 8 | 1.209722 | 0.048382 | 8 | 1.209722 | 0.048382 |
| 47.134.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 47.245.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 48.217.*.* | 2 | 2 | 2 | 1 | 0.91 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 49.51.*.* | 2 | 2 | 1 | 0.5 | 0.43 | 1 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| 51.81.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 51.178.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 51.222.*.* | 6 | 7 | 7 | 1 | 0.16 | 1 | 0 | 7 | 1.903056 | 0.026651 | 7 | 1.903056 | 0.026651 |
| 52.4.*.* | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |

## Page Hits by IP

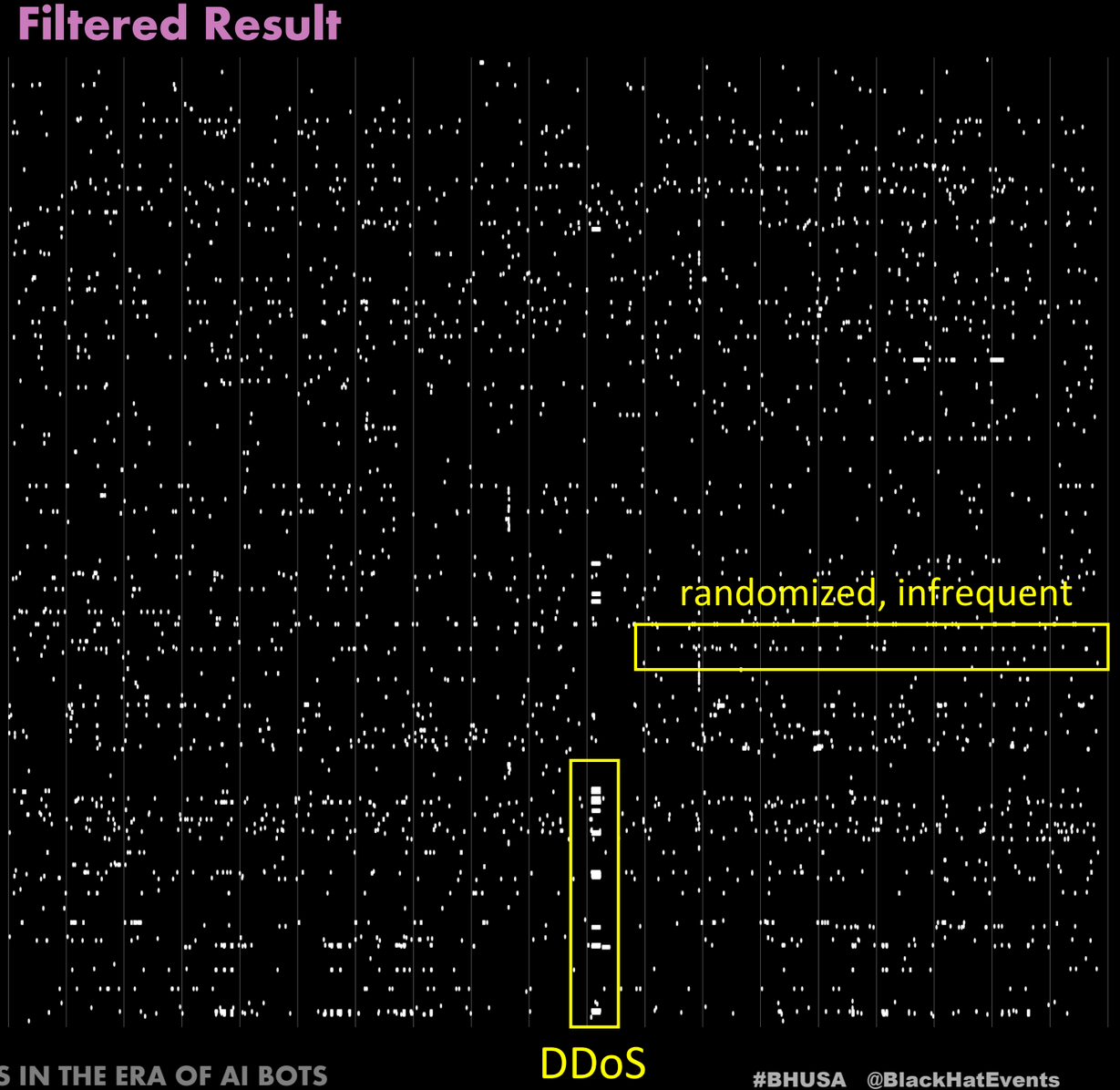
| | | |
|---|---|---|
| 57.141.7.14 | 1 | |
| 57.141.7.15 | 3 | |
| | | 1 /bmi/monitoringregions/4 |
| | | 1 /events/2731 |
| 57.141.7.16 | 2 | |
| | | 1 / |
| 57.141.7.17 | 7 | |
| | | 1 / |
| | | 1 /bmi_events/62 |
| | | 1 /events/2424 |
| | | 1 /events/835 |
| | | 1 /groundwater_queries?page=320 |
| | | 1 /groundwater_queries?page=432 |
| 57.141.7.18 | 5 | |
| | | 1 / |
| | | 1 /bmi_events/152 |
| | | 1 /events/2286 |
| | | 1 /queries/new?q%5Bs%5D=date+asc |
| 57.141.7.19 | 8 | |
| | | 1 /bmi_events/167 |
| | | 1 /events/1100 |
| | | 1 /events/2332 |
| | | 1 /events/2863 |
| | | 1 /queries/new?q%5Bs%5D=analyte_name+desc |
| | | 1 /queries/new?q%5Bs%5D=event_flow+asc |
| | | 1 /queries?page=97&q%5Bs%5D=event_flow+asc |
| 57.141.7.20 | 9 | |
| | | 1 /events/1778 |
| | | 1 /events/1912 |
| | | 1 /groundwater_queries?page=322 |
| | | 1 /monitoringlocations/530 |
| | | 1 /monitoringlocations/8 |
| | | 1 /monitoringsets/25 |
| | | 1 /queries?page=997&q%5Bs%5D=event_flow+asc |
| | | 1 /sites/158 |
| 57.141.7.21 | 6 | |
| | | 1 /events/259 |
| | | 1 /events/2747 |
| | | 1 /queries/new?q%5Bs%5D=monitoringlocation_name+asc |
| | | 1 /queries?page=6&q%5Bs%5D=event_flow+asc |
| | | 1 /sitemap |
| 57.141.7.22 | 10 | |
| | | 1 /bmi/monitoringlocations/382 |
| | | 1 /events/1301 |
| | | 1 /events/2218 |
| | | 1 /events/2260 |
| | | 1 /events/2467 |
| | | 1 /events/37 |
| | | 1 /monitoringlocations/684 |
| | | 1 /queries/new?q%5Bs%5D=analyte_name+asc |
| | | 1 /queries/new?q%5Bs%5D=event_flow+asc |

# Limitations

Cannot stop DDoS attacks
- acquire random IPs

Many AI crawlers still present
- well disguised, more random

At this point -
Human vs. Machine becomes harder to distinguish



# Future Goals

# Future Goals

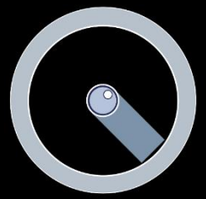
- Now in use. Measure post-blocking activity with client.

# Future Goals

- Now in use. Measure post-blocking activity with client.
- Ground truth data for human and non-human activity (both are difficult to replicate!)

# Future Goals

- Now in use. Measure post-blocking activity with client.
- Ground truth data for human and non-human activity (both are difficult to replicate!)
- Study policy parameter sensitivity and/or optimize



QUANTA
Quanta Sciences

https://github.com/quantasci

we are a knowledge systems, AI and data visualization startup

**LOGRIP**

https://github.com/quantasci/logrip

Open source, Apache 2.0 license

arXiv

https://arxiv.org/abs/2508.03130

RAMA KARL

https://ramakarl.com/

rama karl hoetzlein

Thank you!

Rama Karl Hoetzlein


black hat
BRIEFINGS