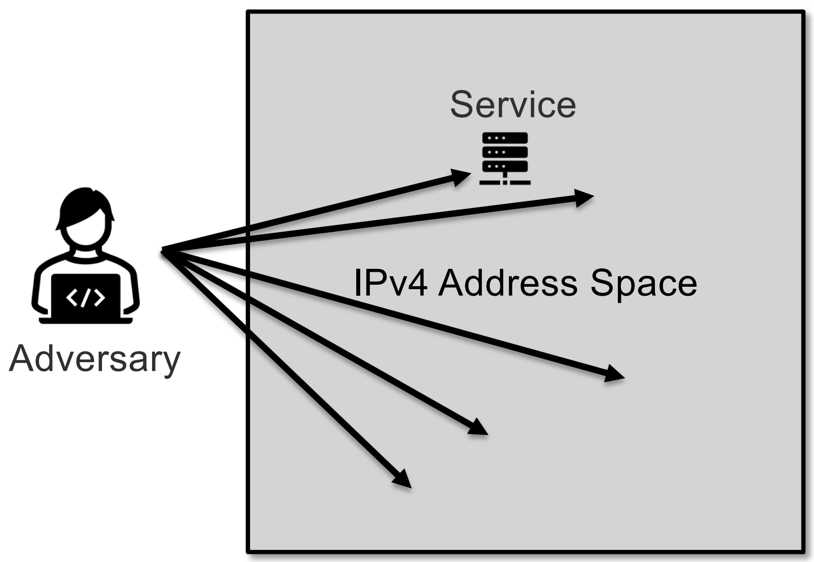
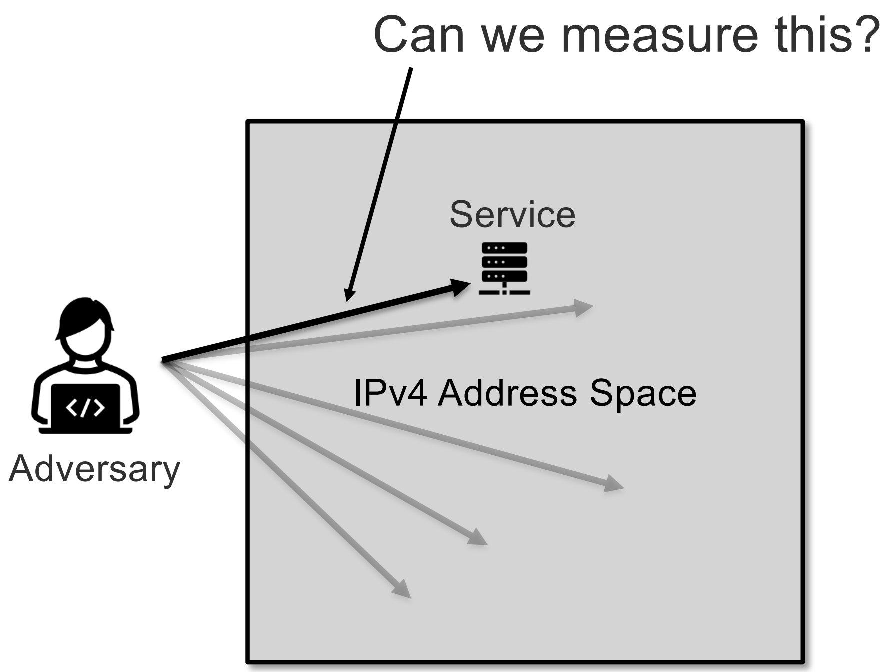
DScope: A Cloud-Native Internet Telescope

Eric Pauley, Paul Barford, Patrick McDaniel
University of Wisconsin–Madison

Why Measure the Internet?

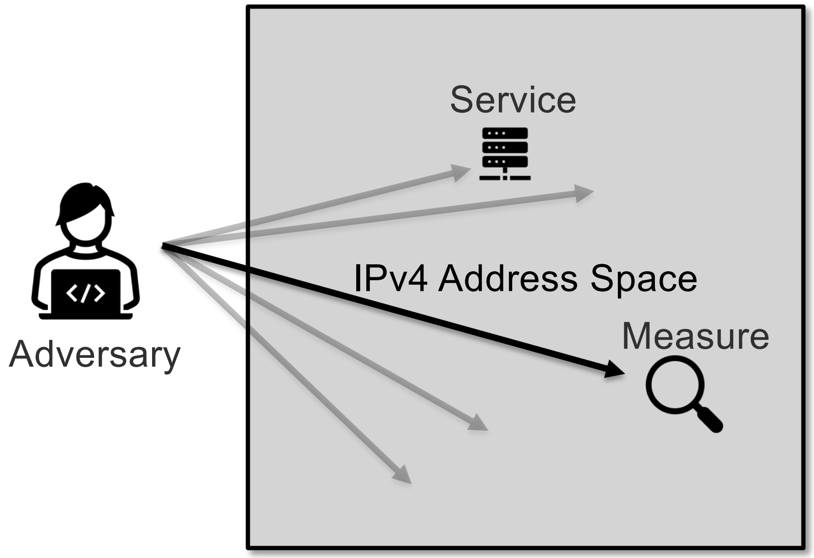


Why Measure the Internet?



NANOG™

Why Measure the Internet?

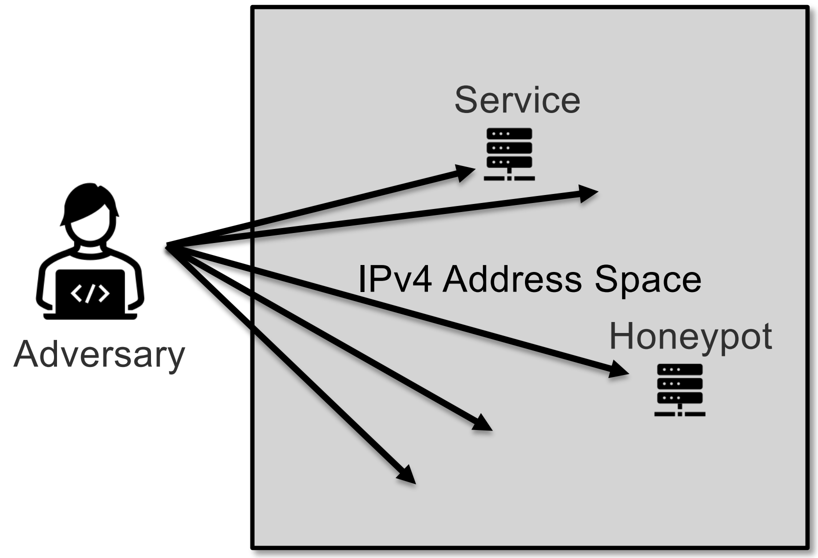


Honeypots: emulating vulnerable services (1970s-)

Idea: pose as vulnerable service

Pro: interactivity

Con: limited coverage (one IP)



NANOG™

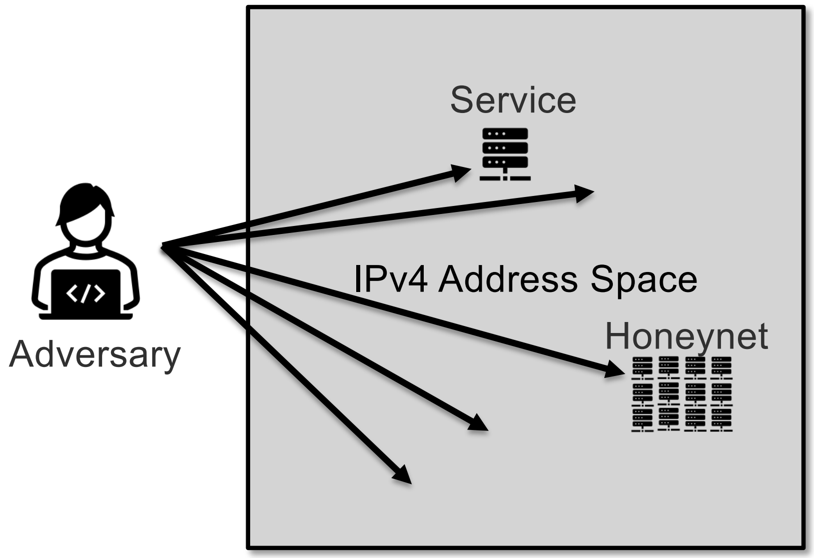
Honeynets: networks of honeypots (1999)

Deploy many honeypot IPs

Bonus: virtualize routing

Pro: interactivity and coverage!

Con: still limited footprint



Telescopes: Large-scale measurement (2001)

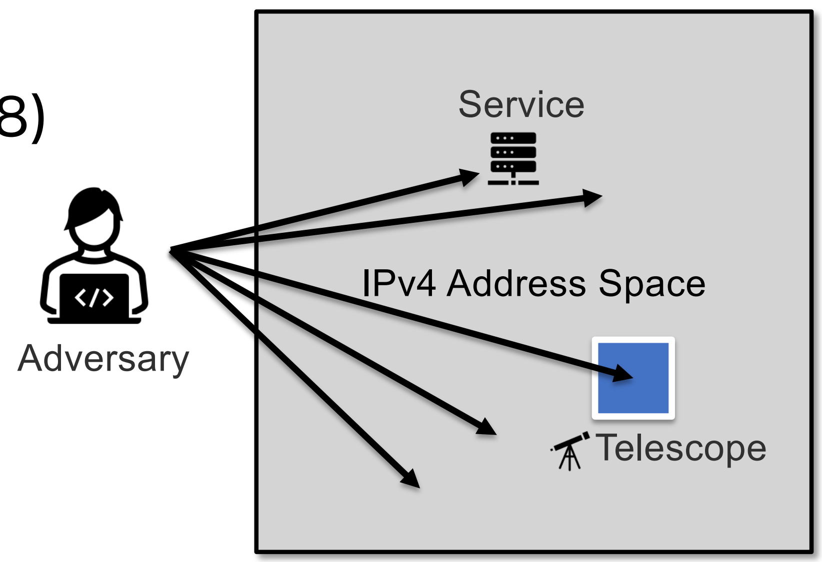
Passively measure large IP blocks (/8)

E.g., UCSD-NT, Merit

Pro: Massive footprints

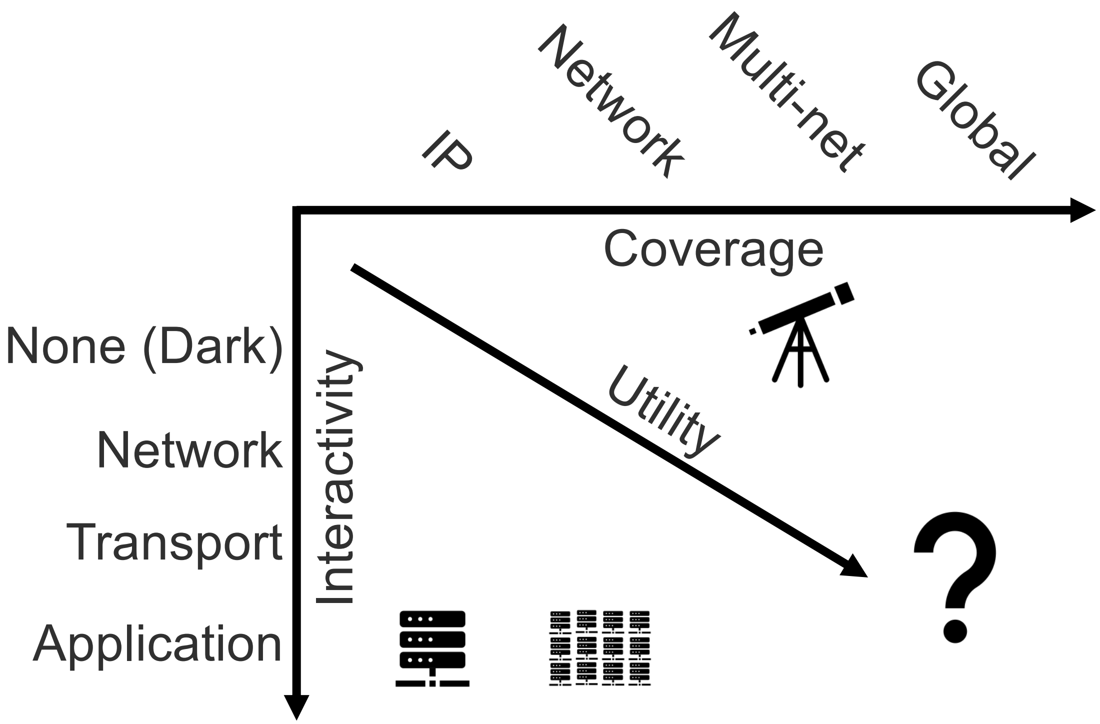
Cons:

- limited interactivity
- homogeneous IP Space



NANOG™

The space of (inbound) Internet Measurement



- Emergent Threats
- Botnets
- Backscatter
- Routing
- Misconfigurations

The Changing Internet (Measurement) Landscape

Rise of Public Clouds

Adversaries target valuable IP ranges

Semantics Moving up Protocol Stack

Passive measurement is incomplete

Sophisticated & Distributed Adversaries

Fixed footprints miss adversarial response

NANOG™

An Internet Telescope for the Modern Internet

Representative Traffic
- Deployed to targeted cloud IP address ranges globally

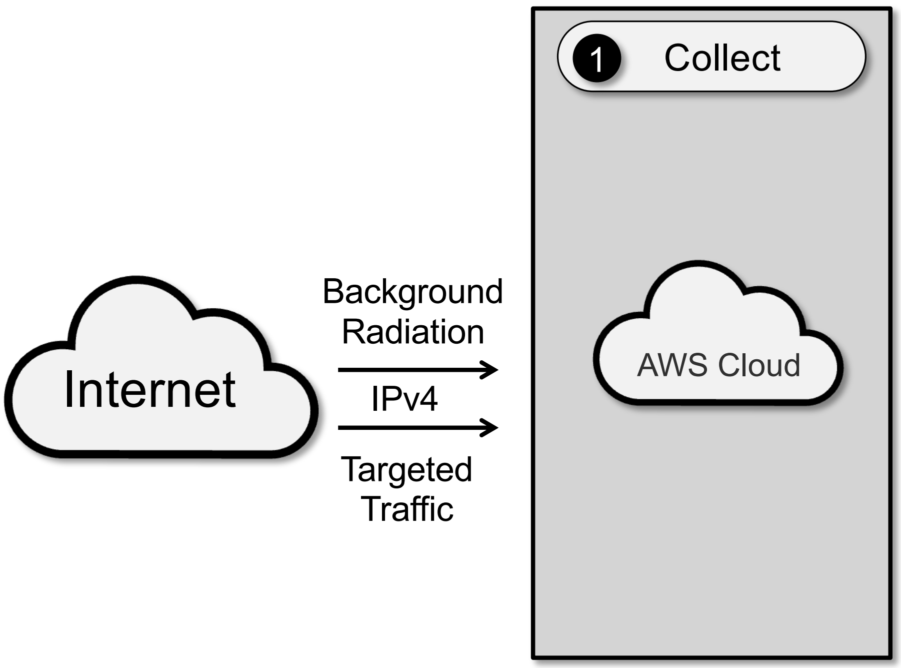
Interactivity
- Collects application-layer banner information
- Elicits deeper adversarial behavior

Agile through the IP address space
- IP footprint varies over time

DScope: A Global, Dynamic, Interactive Cloud Telescope

NANOG™

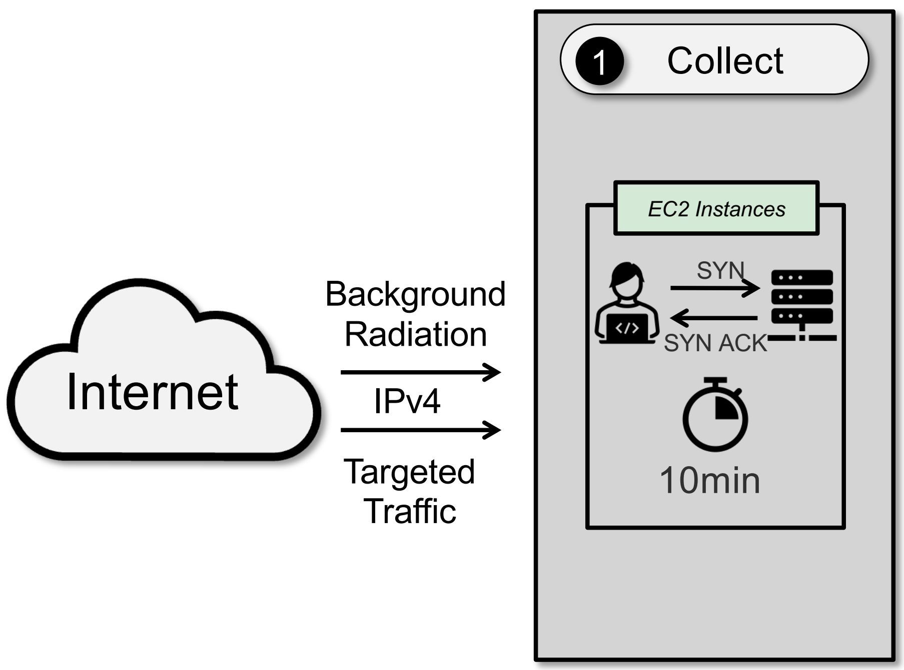
DScope: A Global, Dynamic, Interactive Cloud Telescope



Cloud provider IP footprints and costs:

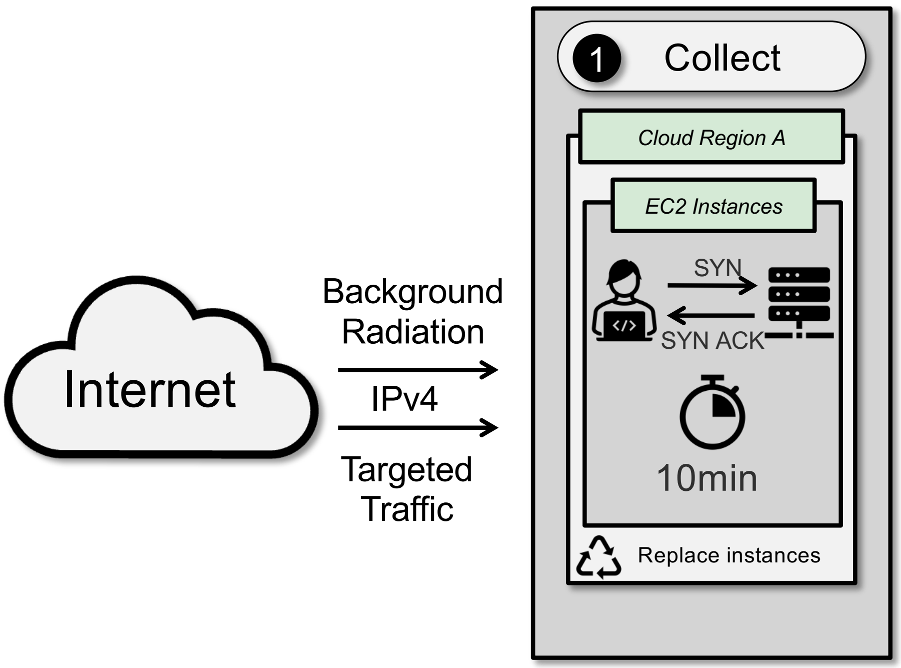
Provider	IPs	# /8s	Cost (USD/IP-Hr)
GCP [15]	11.5 M	34	0.005
Azure [3]	35.7 M	13	0.044
AWS [2]	134 M	82	0.0016

DScope: A Global, Dynamic, **Interactive** Cloud **Telescope**



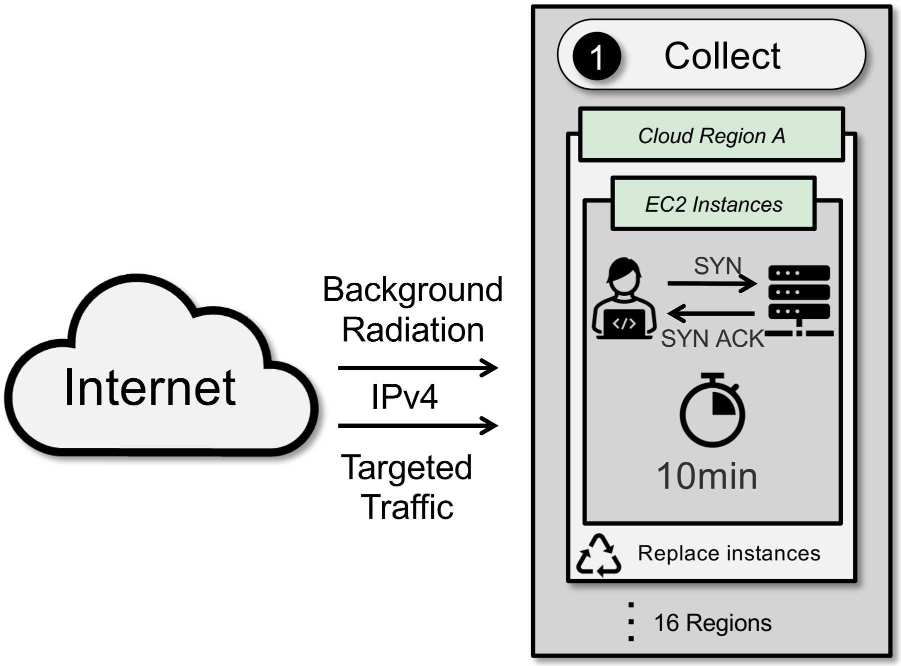
`iptables -j DNAT`

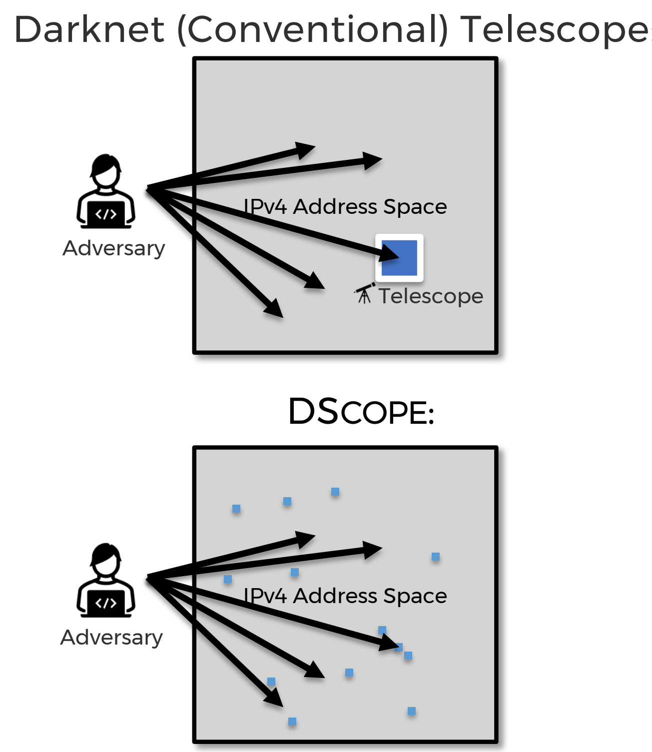
DSCOPE: A Global, Dynamic, Interactive Cloud Telescope



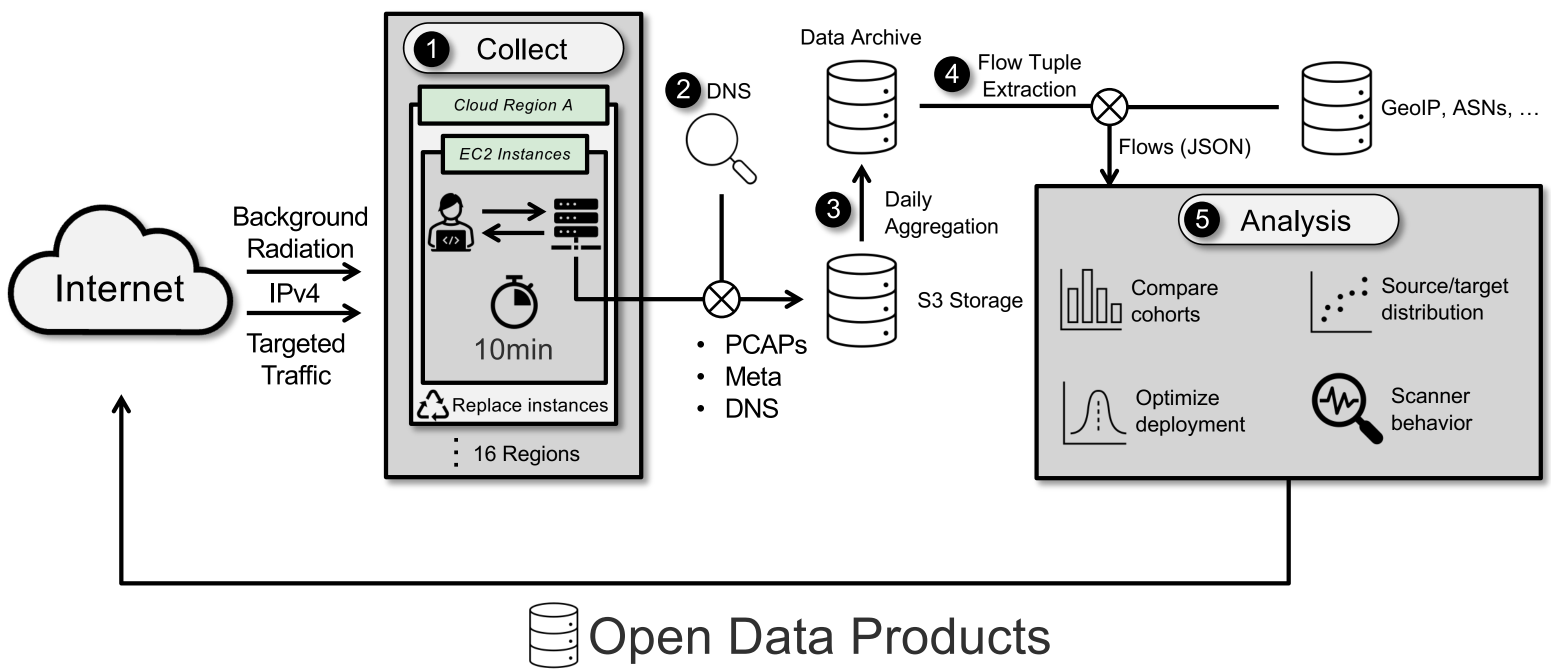
NANOG™

DSCOPE: A Global, Dynamic, Interactive Cloud Telescope





DScope: A Global, Dynamic, Interactive Cloud Telescope and Analysis Platform!



NANOG™

DScope by the numbers

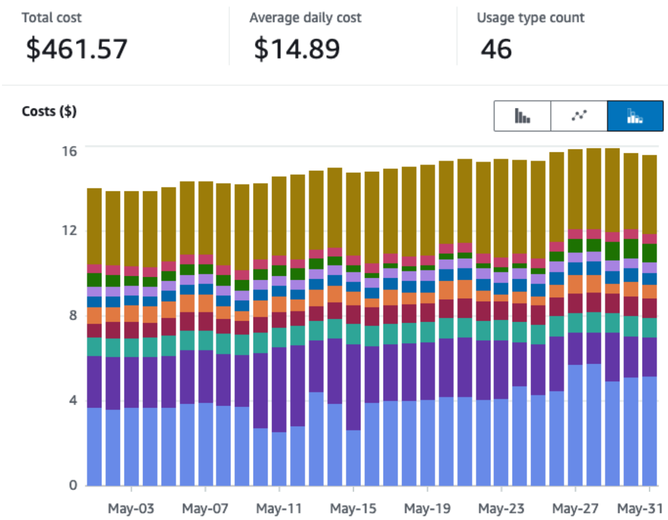
2+ years of collected traffic

6.3M IPv4s

110k /24 networks

More than any other telescope

>15M source IPs measured



NANOG™

Results: 18 findings on cloud-based Internet measurement

	Finding	Metric
Cloud Targeting (Section 4)		
(F1)	An interactive cloud telescope receives traffic from substantially more IP addresses.	73% more traffic
(F2)	Cloud IP traffic is more variable than darknets.	95% higher σ_{IP}
(F3)	Scanners target cloud IP ranges or avoid telescopes.	450$\times$ higher than expected under H_0
(F4)	Scanners that are seen by both darknet/cloud telescopes are largely untargeted.	N/A
(F5)	Scans targeting existing telescopes are primarily random.	N/A
Interactivity & Service Lifecycle (Section 5)		
(F6)	Some scanner IPs demonstrate clearly non-random behavior.	1.7% of traffic ($p < 10^{-4}$)
(F7)	Delayed scanners leverage information from other sources to target responsive IPs.	$>$ 90% discernible source
(F8)	Delayed scanners are not seen by existing darknet telescopes.	90% telescope avoidance ($p < 10^{-4}$)
Intra-cloud Targeting (Section 6)		
(F9)	Quantity of scanners differs across cloud regions, but intra-region variance dominates.	$\pm 0.3\sigma$ variation between regions
(F10)	Source IP variance differs between regions.	6$\times$ variation in σ
(F11)	Scanners target cloud IP addresses based on outdated data.	21% fewer scanners to 2021 AWS IPs
(F12)	Traffic to individual regions is largely consistent with untargeted scanning.	$<$ 10% regional targeting
(F13)	Some sophisticated scanners precisely target physical regions within cloud IP blocks.	4$\times$ background rate for region/port
(F14)	Scanners show minimal preference to groups of regions in similar geographies.	0.02 lower overlap in same-geography
Optimizing Collection (Section 7)		
(F15)	Observed traffic increases over time after instance deployment, but only to a point.	67% increase
(F16)	Scanners targeting ORION are less likely to be reactive.	34% increase
(F17)	Short-lived use of IP addresses maximizes economical yield of new behavior.	$<$ 10 min for max yield
(F18)	Extended measurement on a given IP is not necessary to achieve high coverage.	90% IP coverage at 72 minutes

NANOG™

Results: 18 findings on cloud-based Internet measurement

Coverage

Interactivity

Validity

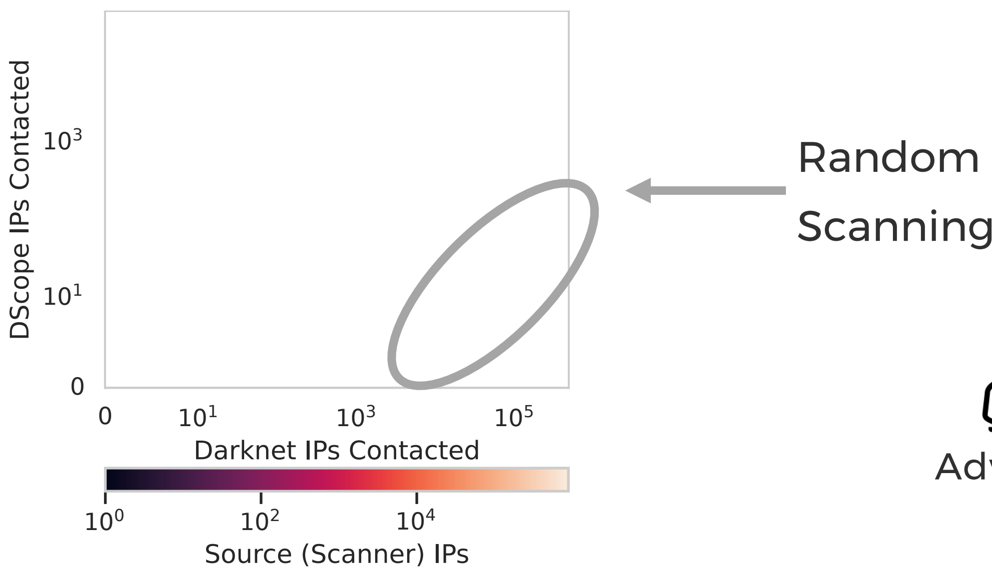
Cost

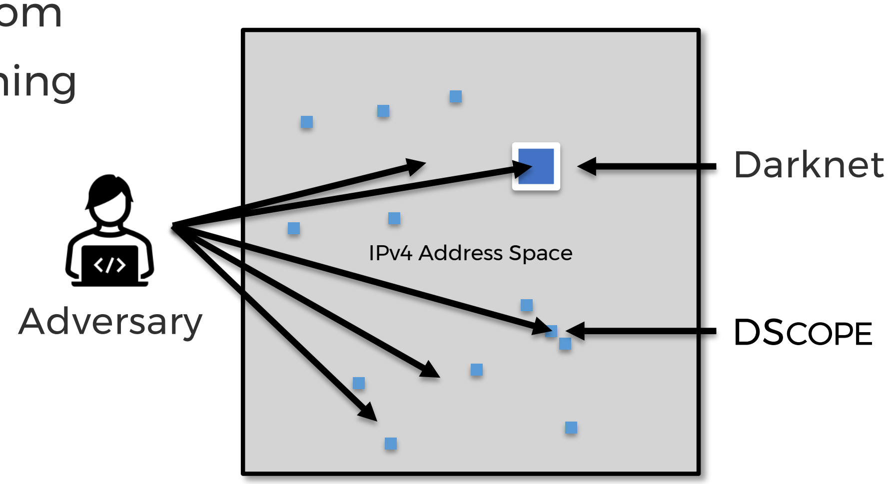
	Finding	Metric
Cloud Targeting (Section 4)		
(F1)	An interactive cloud telescope receives traffic from substantially more IP addresses.	73% more traffic
(F2)	Cloud IP traffic is more variable than darknets.	95% higher σ_{IP}
(F3)	**Scanners target cloud IP ranges or avoid telescopes.**	**450× higher than expected under H_0**
(F4)	Scanners that are seen by both darknet/cloud telescopes are largely untargeted.	N/A
(F5)	Scans targeting existing telescopes are primarily random.	N/A
Interactivity & Service Lifecycle (Section 5)		
(F6)	Some scanner IPs demonstrate clearly non-random behavior.	1.7% of traffic ($p < 10^{-4}$)
(F7)	Delayed scanners leverage information from other sources to target responsive IPs.	> 90% discernible source
(F8)	**Delayed scanners are not seen by existing darknet telescopes.**	**90% telescope avoidance ($p < 10^{-4}$)**
Intra-cloud Targeting (Section 6)		
(F9)	Quantity of scanners differs across cloud regions, but intra-region variance dominates.	±0.3σ variation between regions
(F10)	Source IP variance differs between regions.	6× variation in σ
(F11)	**Scanners target cloud IP addresses based on outdated data.**	**21% fewer scanners to 2021 AWS IPs**
(F12)	Traffic to individual regions is largely consistent with untargeted scanning.	< 10% regional targeting
(F13)	Some sophisticated scanners precisely target physical regions within cloud IP blocks.	4× background rate for region/port
(F14)	Scanners show minimal preference to groups of regions in similar geographies.	0.02 lower overlap in same-geography
Optimizing Collection (Section 7)		
(F15)	Observed traffic increases over time after instance deployment, but only to a point.	67% increase
(F16)	Scanners targeting ORION are less likely to be reactive.	34% increase
(F17)	**Short-lived use of IP addresses maximizes economical yield of new behavior.**	**< 10 min for max yield**
(F18)	Extended measurement on a given IP is not necessary to achieve high coverage.	90% IP coverage at 72 minutes

NANOG™

Coverage: Is Internet Scanning Random?

Recall: Null-Hypothesis of Random Scanning

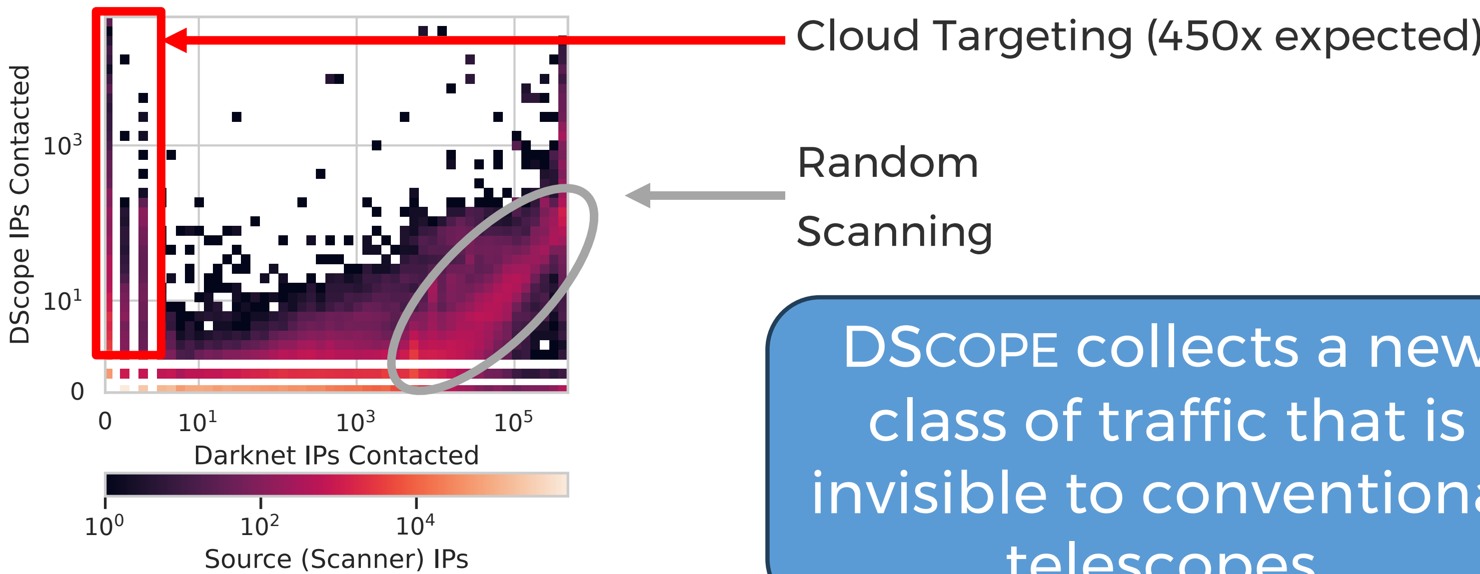




NANOG

Coverage: Is Internet Scanning Random?

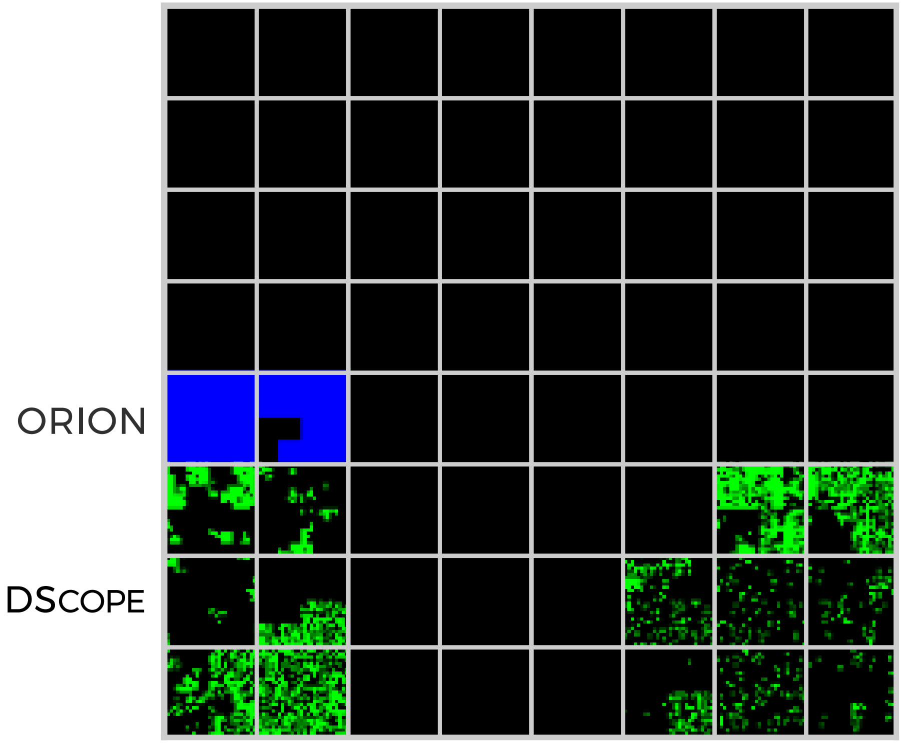
Recall: Null-Hypothesis of Random Scanning



DScope collects a new class of traffic that is invisible to conventional telescopes.

Coverage: Is Internet Scanning Sequential?

IPv4 /8 around Merit's ORION telescope:

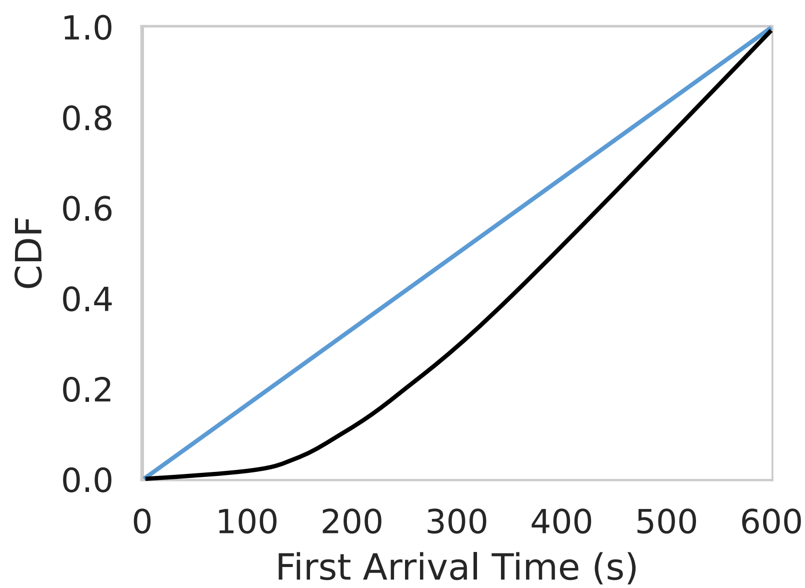


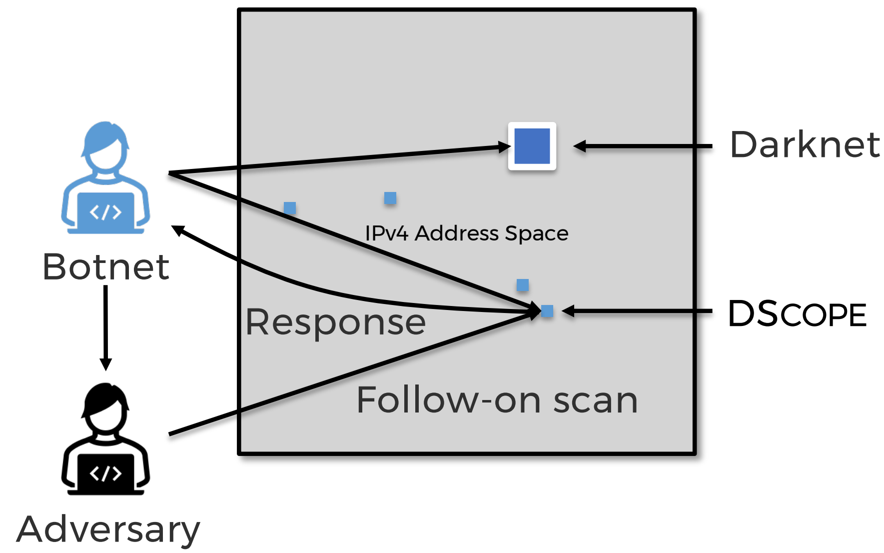
Question: Are IPs near ORION more likely to share traffic?

Answer: No difference (not sequential)

Interactivity: Service Lifecycle and follow-on scans

Does interactivity induce adversarial response?

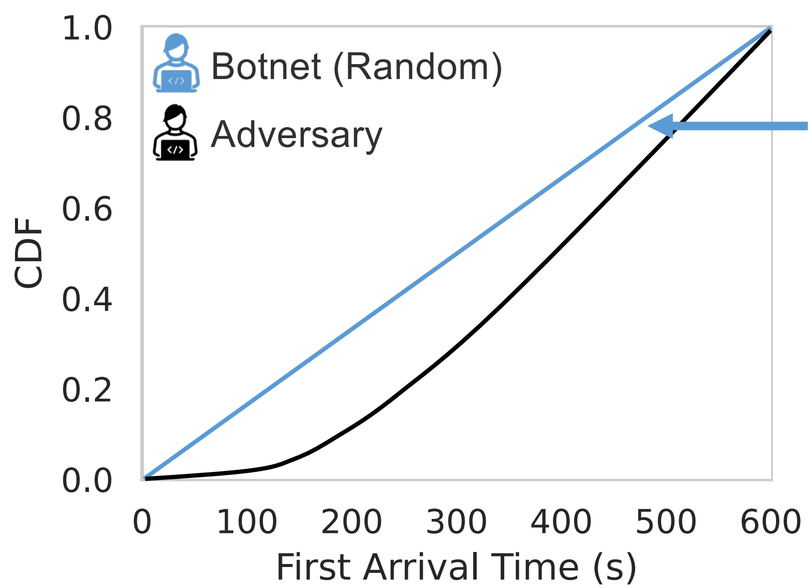




NANOG™

Interactivity: Service Lifecycle and follow-on scans

Does interactivity induce adversarial response?



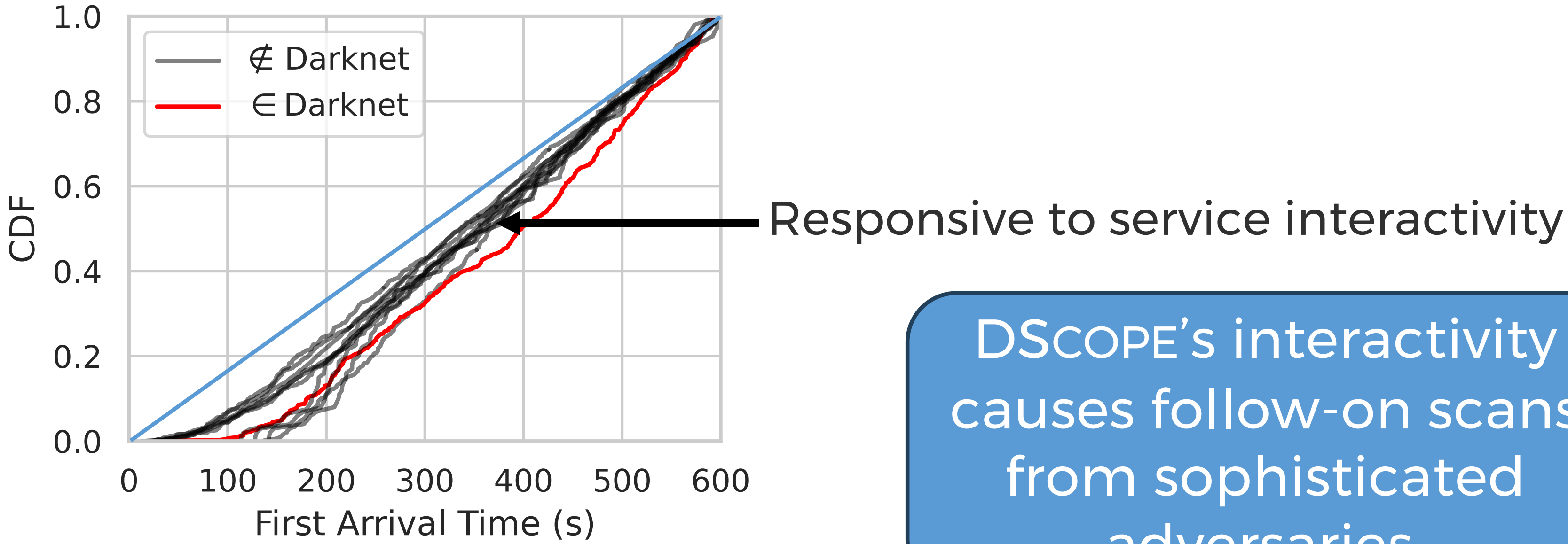
Expected Distribution (Non-responsive scanning):

$$f_T(t) = \frac{\lambda e^{-\lambda t}}{1 - e^{-\lambda m}} \qquad (0 <= t <= m)$$

Approach: Goodness-of-fit (K-S) test

Interactivity: Service Lifecycle and follow-on scans

Does interactivity induce adversarial response?



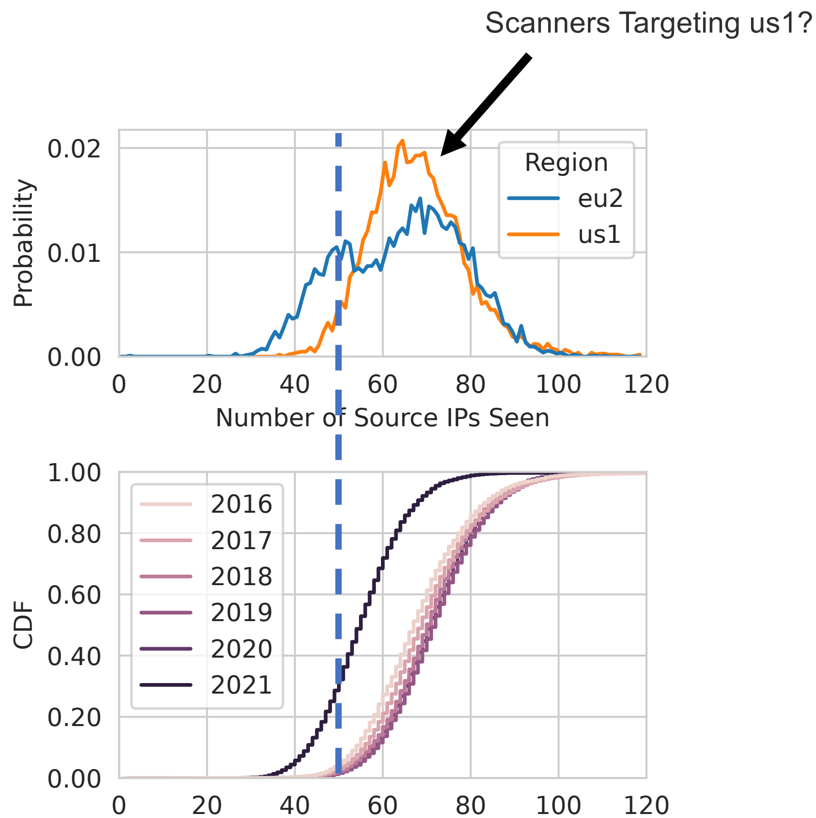
DSCOPE's interactivity causes follow-on scans from sophisticated adversaries.

Cloud Traffic Distributions & Statistical Validity

Challenge: Every cloud IP is *unique*:

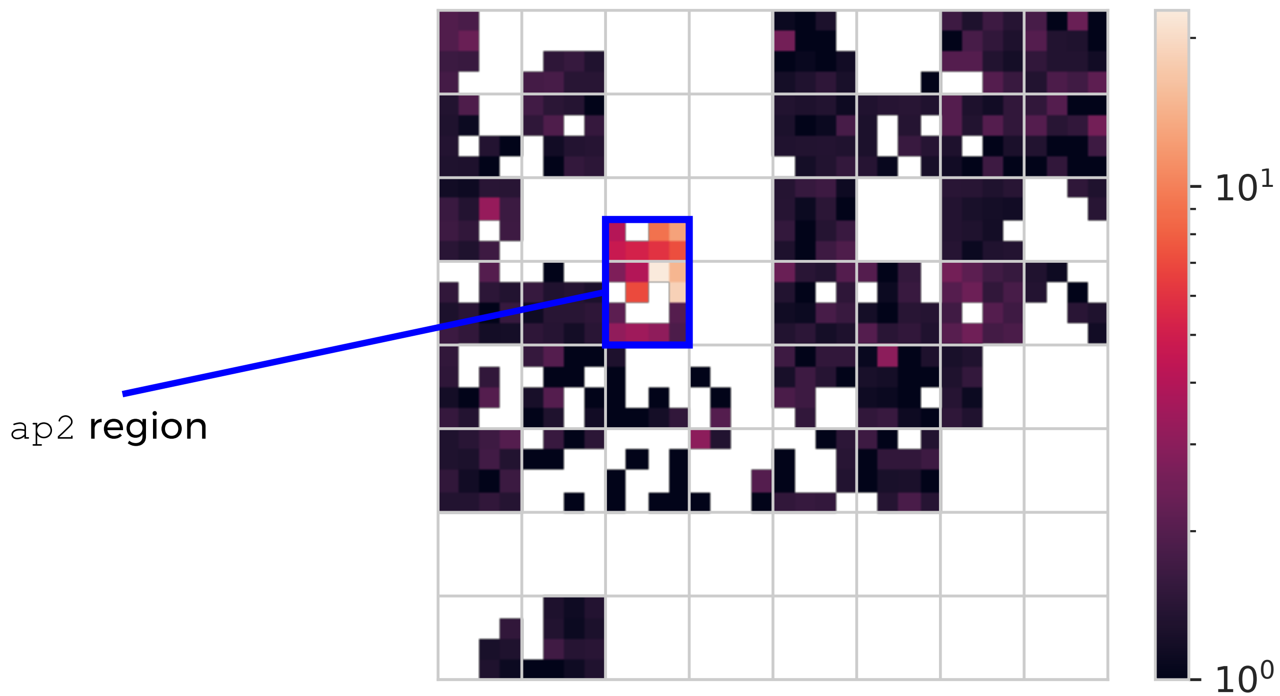
- IP address history
- Latent configuration

DScope's large footprint allows for elimination of confounding factors.



NANOG

Geographic Targeting: An Example

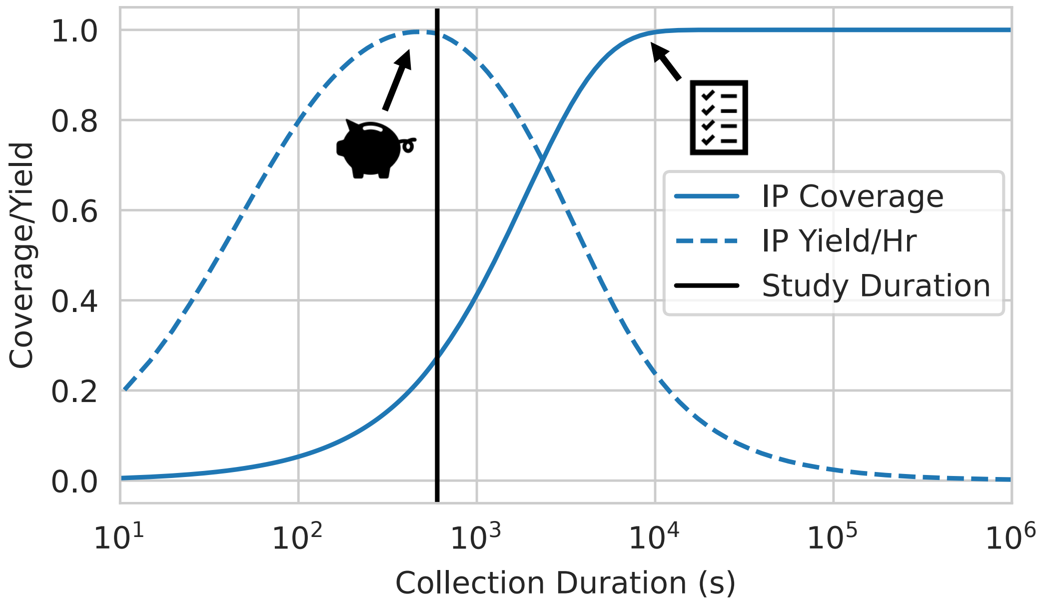


Hilbert Diagram of port 445 traffic seen by 3.0.0.0/8 IP addresses

NANOG

Cost Optimization: How long should DScope hold IPs?

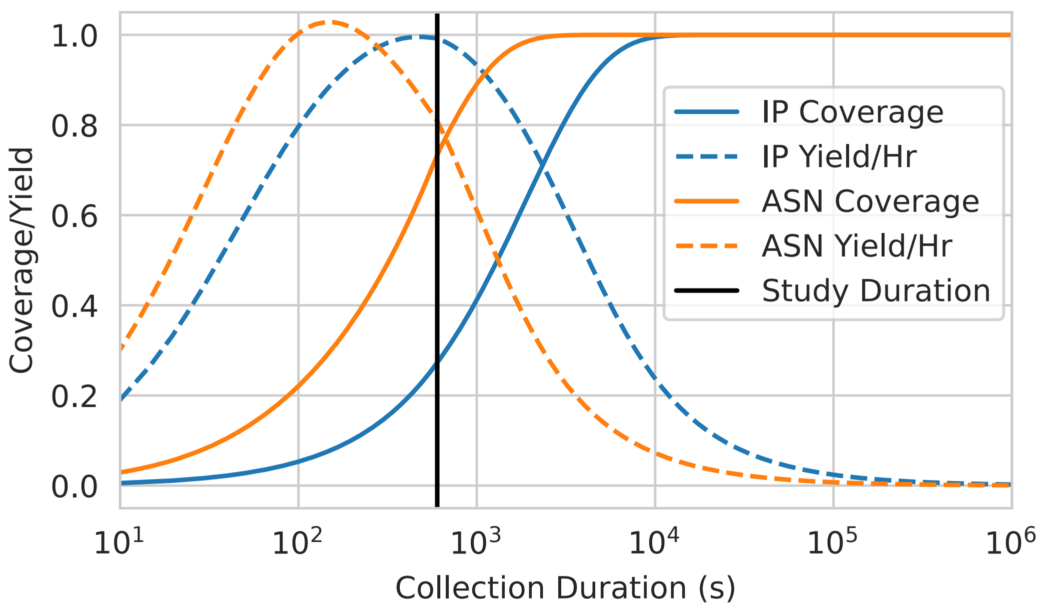
Goal: Max coverage with min cost (IP-hours)



NANOG

Cost Optimization: How long should DScope hold IPs?

Goal: Max coverage with min cost (IP-hours)



DScope's deployment can optimize for coverage or yield of Internet phenomena.

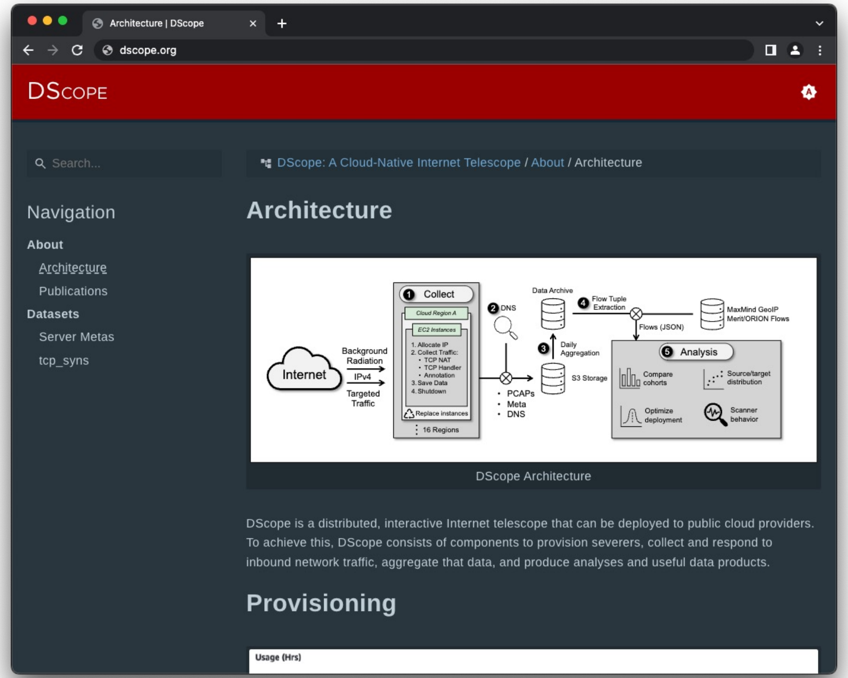
DScope achieves:

- Representative Traffic and Global Coverage
- Interactivity & Service Lifecycle
- Agility through IP Space
- Price Performance
- Useful data...?

DSCOPE.ORG and Open Data

- Data Products
  - Standard formats (JSON, PCAP)
  - 2+ years of data (more daily)
  - Data sharing agreements WIP

- Interactive Visualizations
  - Emergent Threats
  - Cloud Scanning
  - Deployment Health



NANOG™

What data does DScope provide?

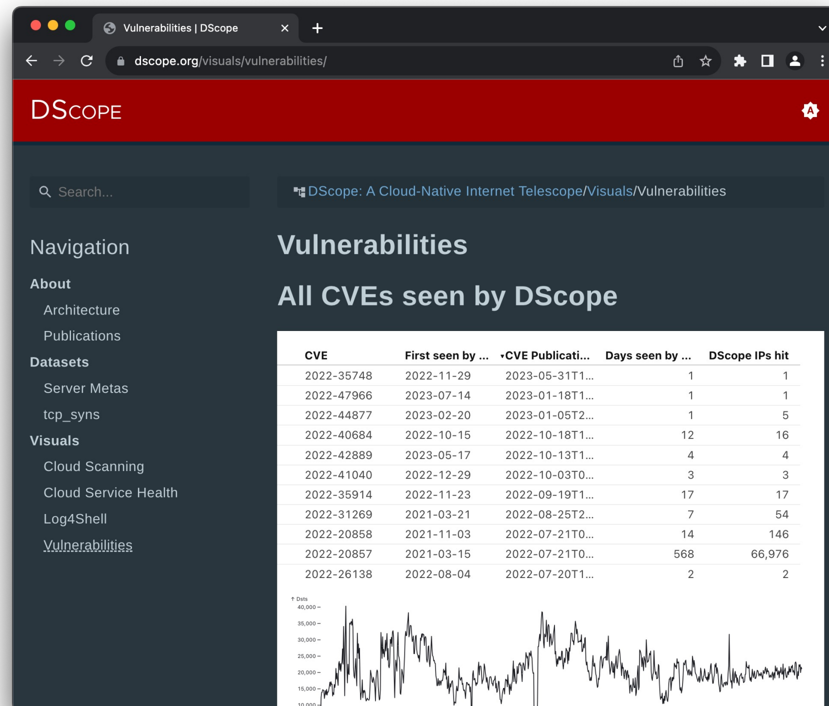
- Broad Application-layer traffic
- Cloud-targeted phenomena
- General-purpose telescope data

App-layer Data: Vulnerabilities

Data: Traffic matches against IDS rulesets

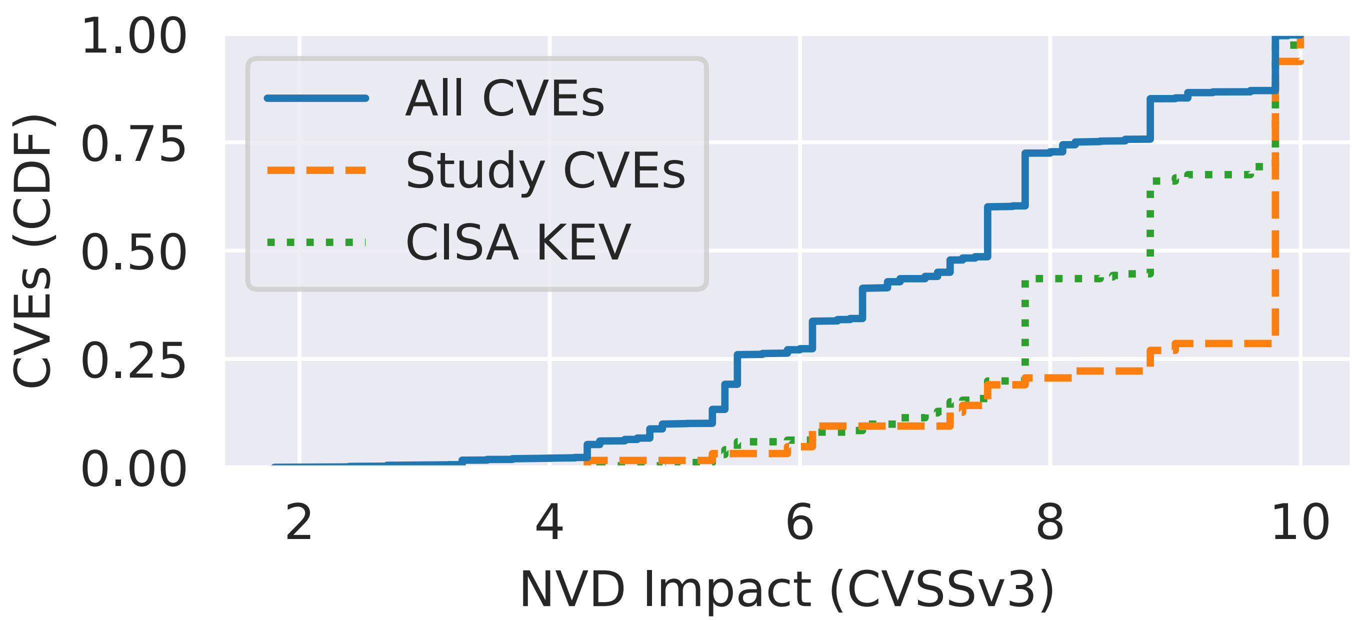
Analyses:

- CVE trends
- Exploit Sources

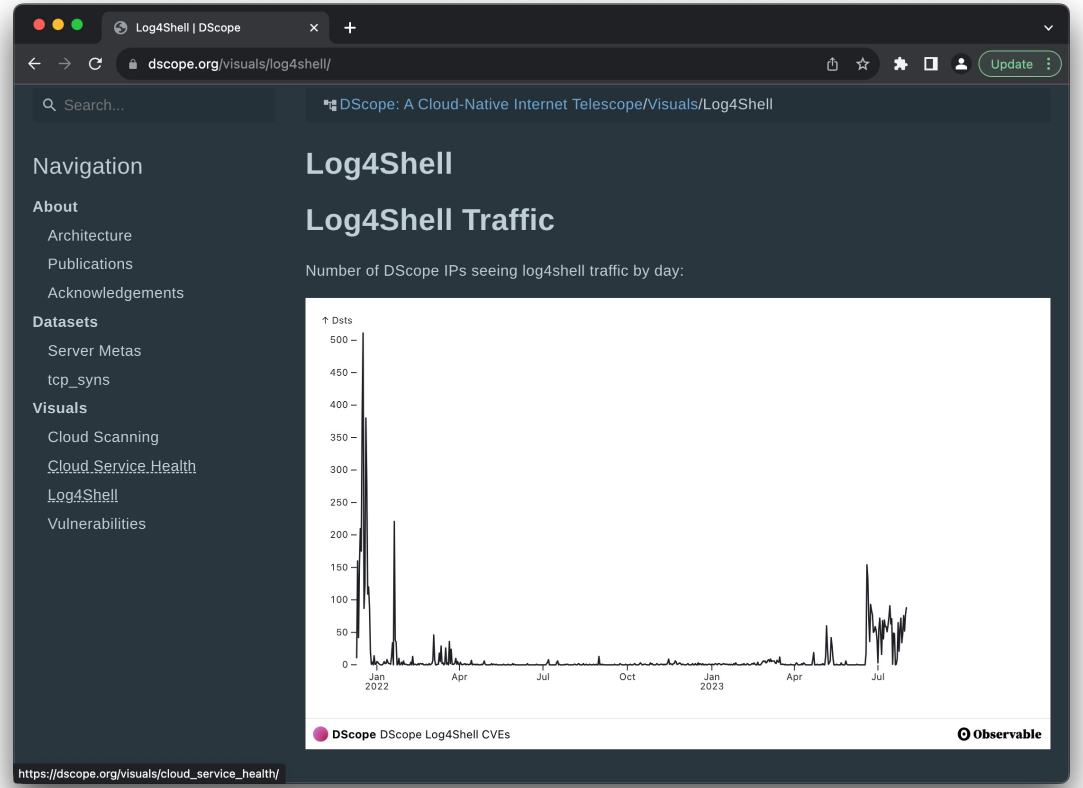


NANOG

App-Layer Data: Is DScope representative?

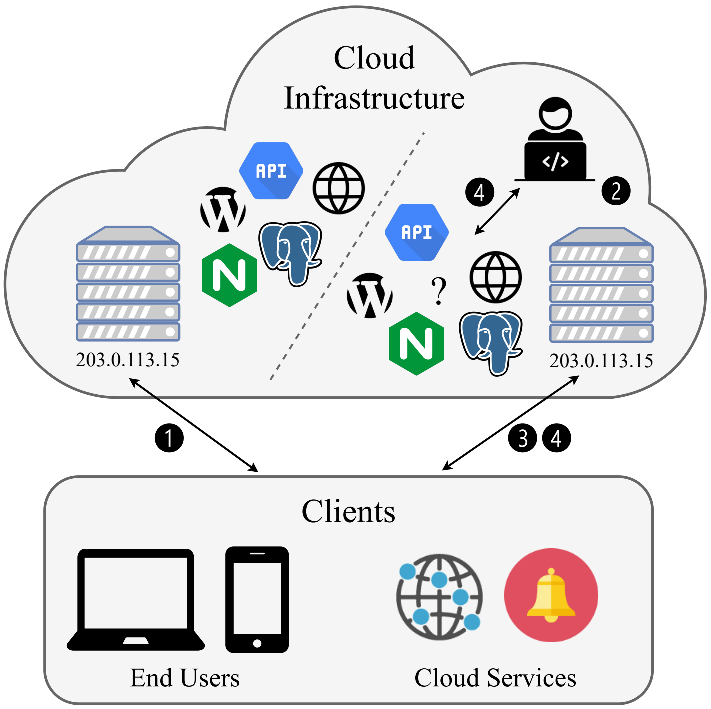


Example: Log4Shell



NANOG™

Measuring Cloud Squatting



- Idea: Cloud IPs receive traffic intended for previous tenants

- Measurement: Identify vulnerable configurations through traffic analysis

NANOG™

Cloud Squatting: Vulnerability at Scale

<u>Cloud Services</u>

>5M messages

4 cloud services

<u>Third-Party Services</u>

- >3M messages
- Numerous Services

<u>DNS</u>

- 5400 Websites
- 23 top-1000

<u>Example Sensitive Data Received</u>

Financial

Personal

Location

Remote Code Execution

Passwords

Images

General-Purpose Telescope Data

- Raw PCAPs
  - Application layer or synthetic-darknet
  - Limited to TCP traffic
- Scanning Events
  - Caveats: non-linear address space

Building Future Vantage Points

Goal: Quality > Quantity

- DScope achieves quality by using diverse cloud IPs
- Fewer IPs yield more representative phenomena
- *What* are we trying to gain coverage of?

Approach: Increase footprint diversity

- Spread across operators, geographies, services
- Collaborations with industry to instrument networks
- Get in touch for more details!

Thanks!

DScope.org

epauley@cs.wisc.edu