

DSCOPE: A Cloud-Native Internet Telescope

[Eric Pauley](#), Paul Barford, Patrick McDaniel

University of Wisconsin–Madison

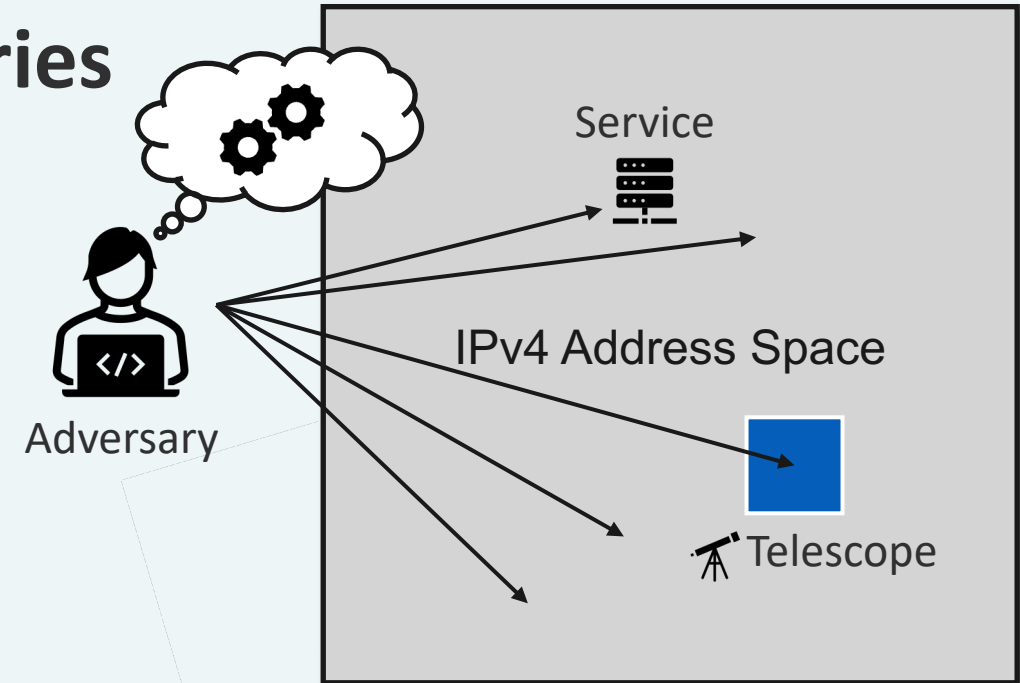
Background: Internet Scanning & Internet Telescopes

Goal: representative data on adversaries

Apparatus: Internet Telescopes

Routed regions of IPv4 address space

Receive but (usually) don't respond to traffic



The Changing Internet (Measurement) Landscape



Rise of Public Clouds

Adversaries target valuable IP ranges



Semantics Moving up Protocol Stack

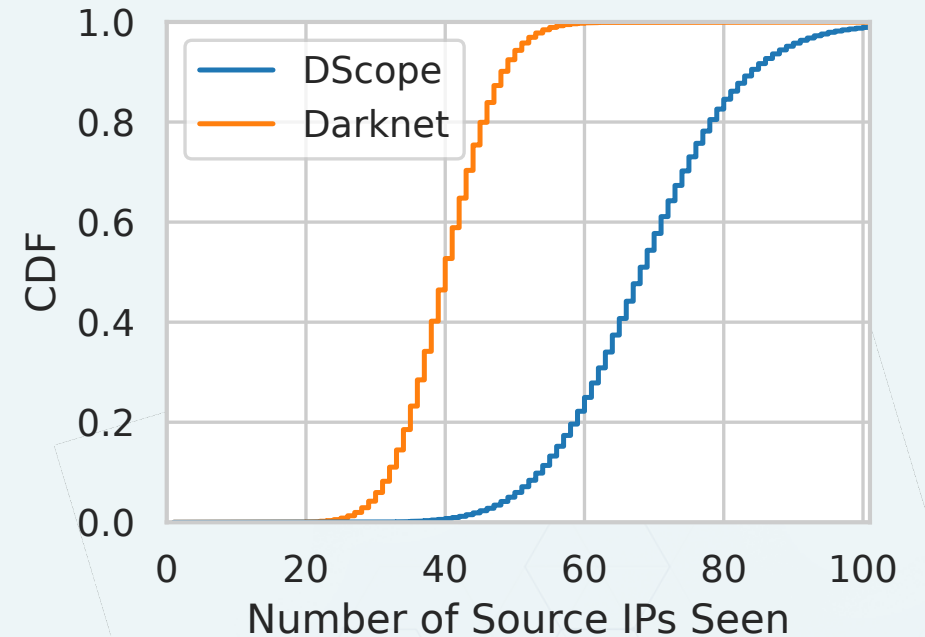
Passive measurement is incomplete



Sophisticated & Distributed Adversaries

Fixed footprints miss adversarial response

Darknet telescopes have incomplete coverage:



An Internet Telescope for the Modern Internet



Representative Traffic

Deployed to targeted cloud IP address ranges



Interactivity

Collects application-layer banner information

Elicits deeper adversarial behavior



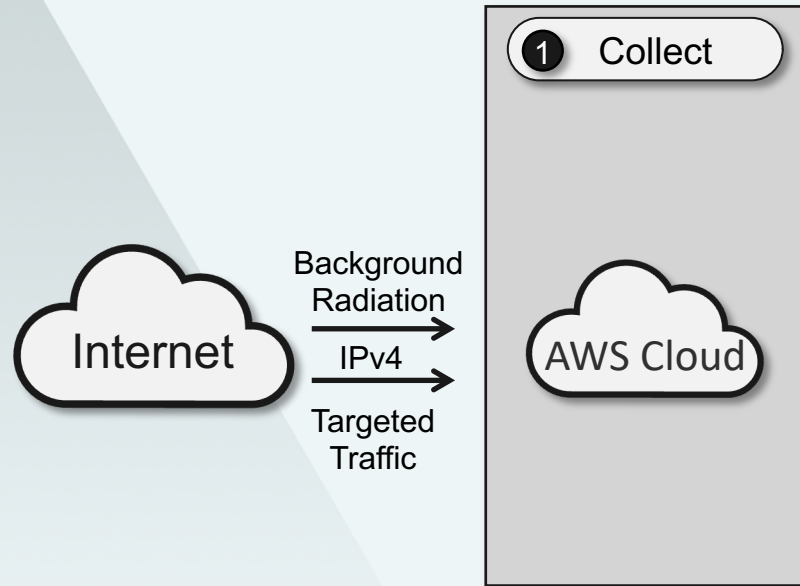
Agile through the IP address space

IP footprint varies over time

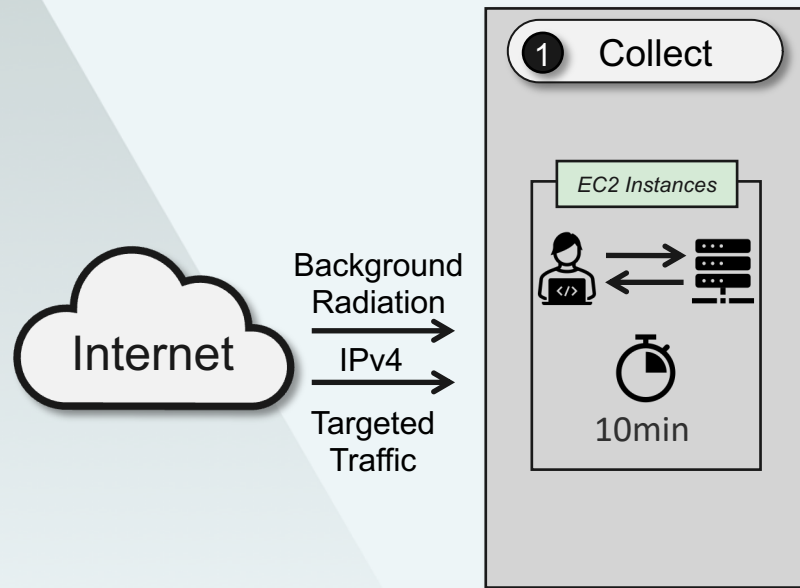
DSCOPE: A Global, Dynamic, Interactive Cloud Telescope



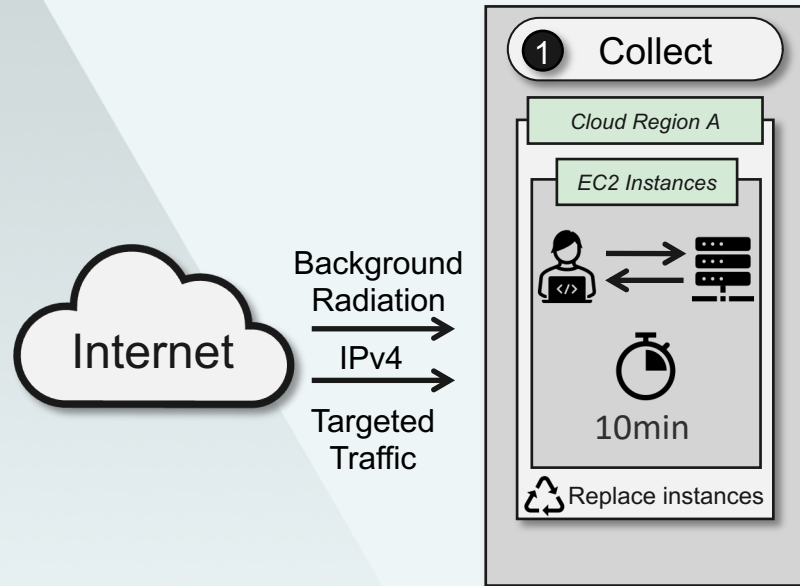
DSCOPE: A Global, Dynamic, Interactive **Cloud** Telescope



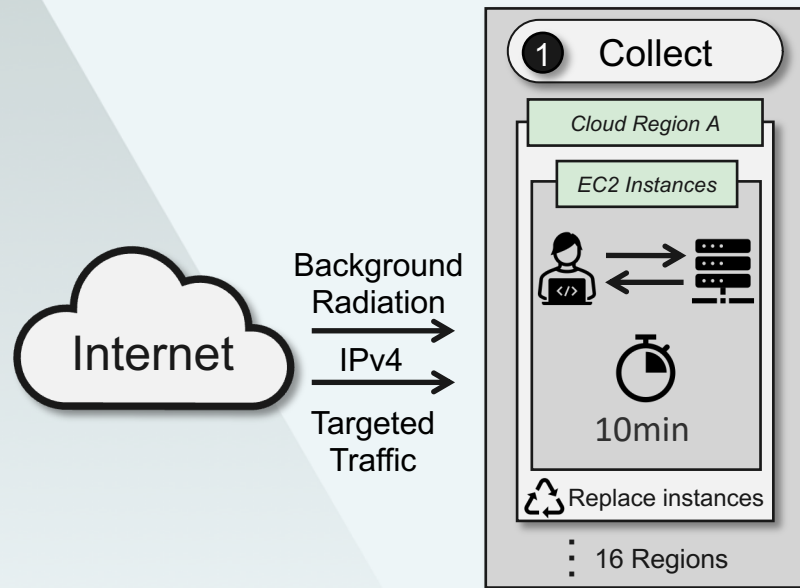
DSCOPE: A Global, Dynamic, **Interactive** Cloud Telescope



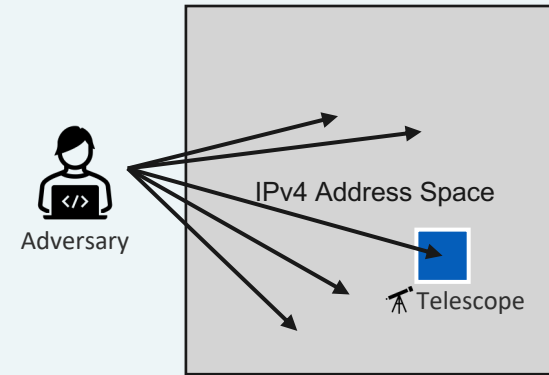
DSCOPE: A Global, **Dynamic**, Interactive Cloud Telescope



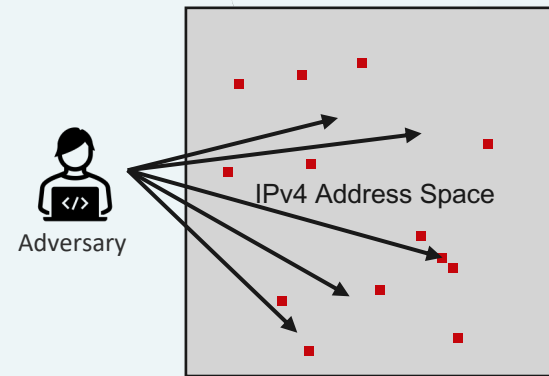
DSCOPE: A **Global**, Dynamic, Interactive Cloud Telescope



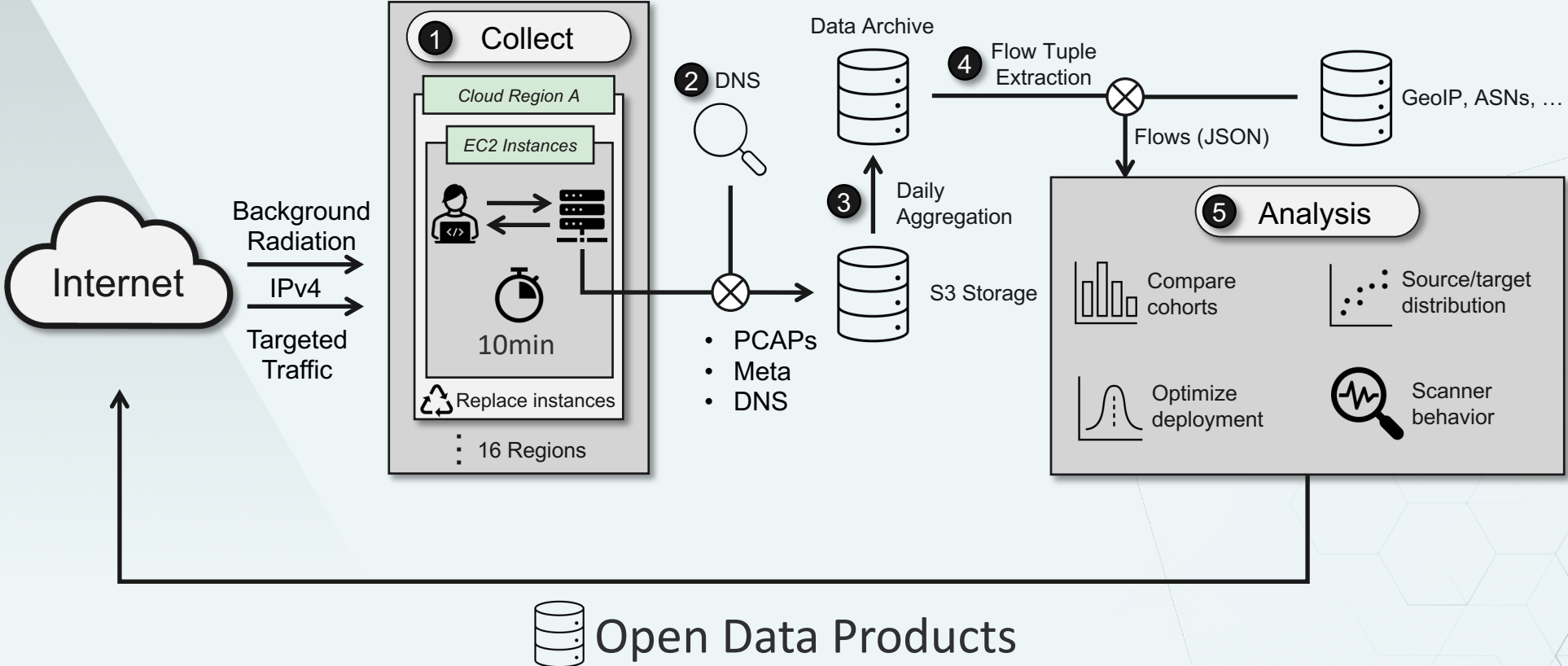
Darknet (Conventional) Telescope:



DSCOPE:



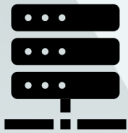
DSCOPE: A Global, Dynamic, Interactive Cloud Telescope and Analysis Platform!



DSCOPE by the numbers:



2+ years of collected traffic



6.3M IPv4s

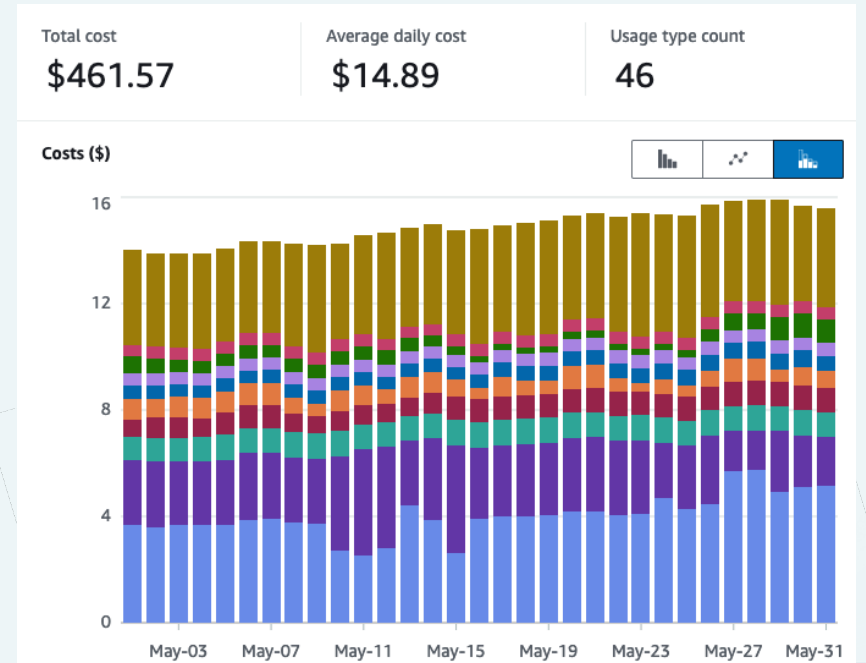


110k /24 networks

More than any other telescope



>15M source IPs measured



Results: 18 findings on cloud-based Internet measurement

Finding		Metric
<i>Cloud Targeting (Section 4)</i>		
(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
<i>Interactivity & Service Lifecycle (Section 5)</i>		
(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 ($p < 10^{-4}$)
<i>Intra-cloud Targeting (Section 6)</i>		
(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× 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
<i>Optimizing Collection (Section 7)</i>		
(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

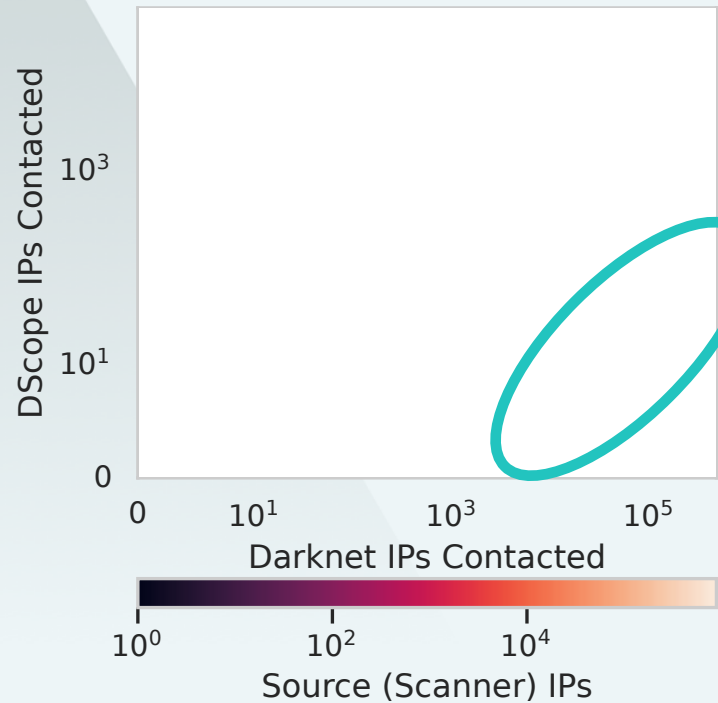
Results: 18 findings on cloud-based Internet measurement

	Finding	Metric
	<i>Cloud Targeting (Section 4)</i>	
Coverage 	(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
	<i>Interactivity & Service Lifecycle (Section 5)</i>	
Interactivity 	(F6) Some scanner IPs demonstrate clearly non-random behavior.	1.7% of traffic ($p < 10^{-4}$)
	(F8) Delayed scanners are not seen by existing darknet telescopes.	90% telescope avoidance ($p < 10^{-4}$)
	<i>Intra-cloud Targeting (Section 6)</i>	
Validity 	(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× 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
	<i>Optimizing Collection (Section 7)</i>	
Cost 	(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.	< 10min for max yield
	(F18) Extended measurement on a given IP is not necessary to achieve high coverage.	90% IP coverage at 72 minutes

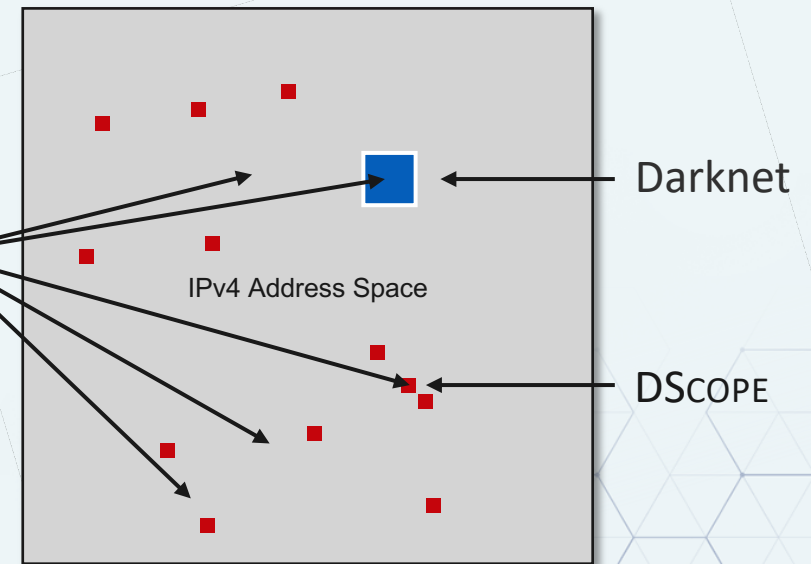


Coverage: Is Internet Scanning Random?

Recall: Null-Hypothesis of Random Scanning



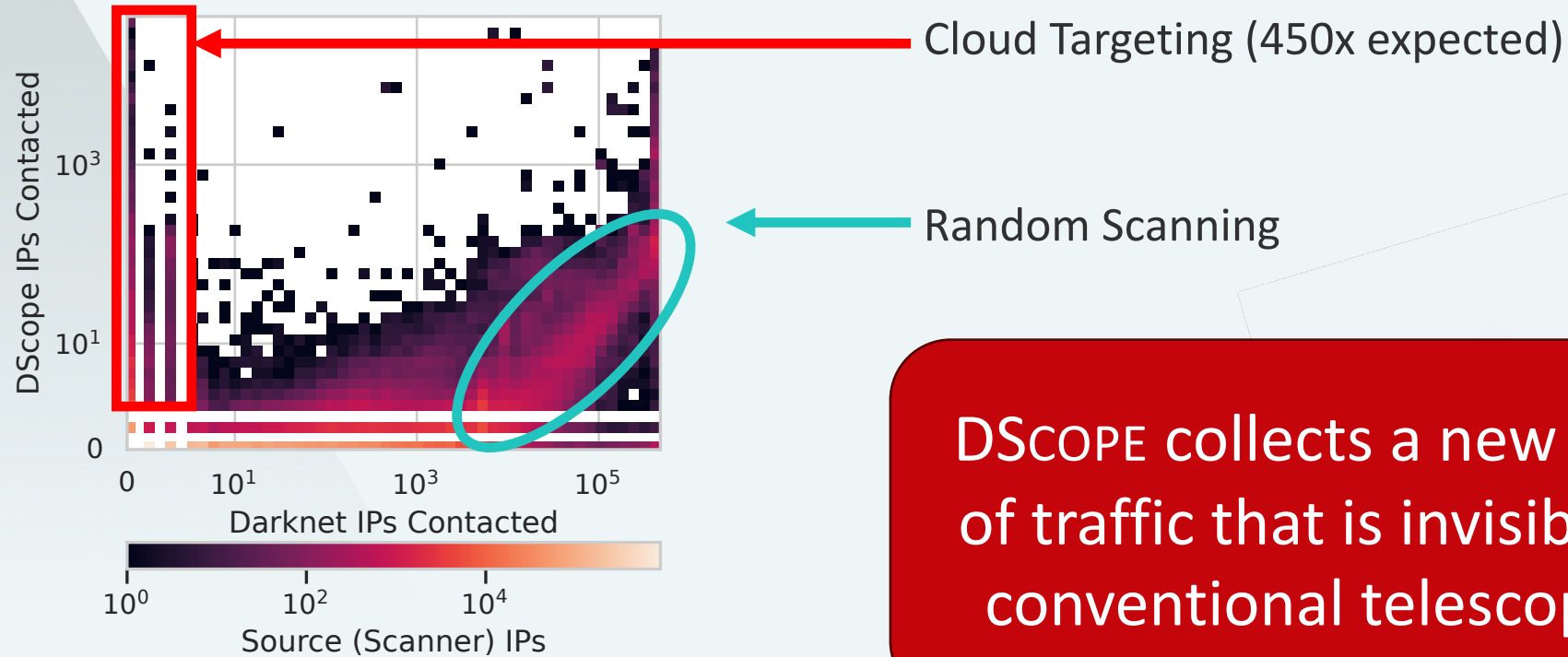
← Random Scanning





Coverage: Is Internet Scanning Random?

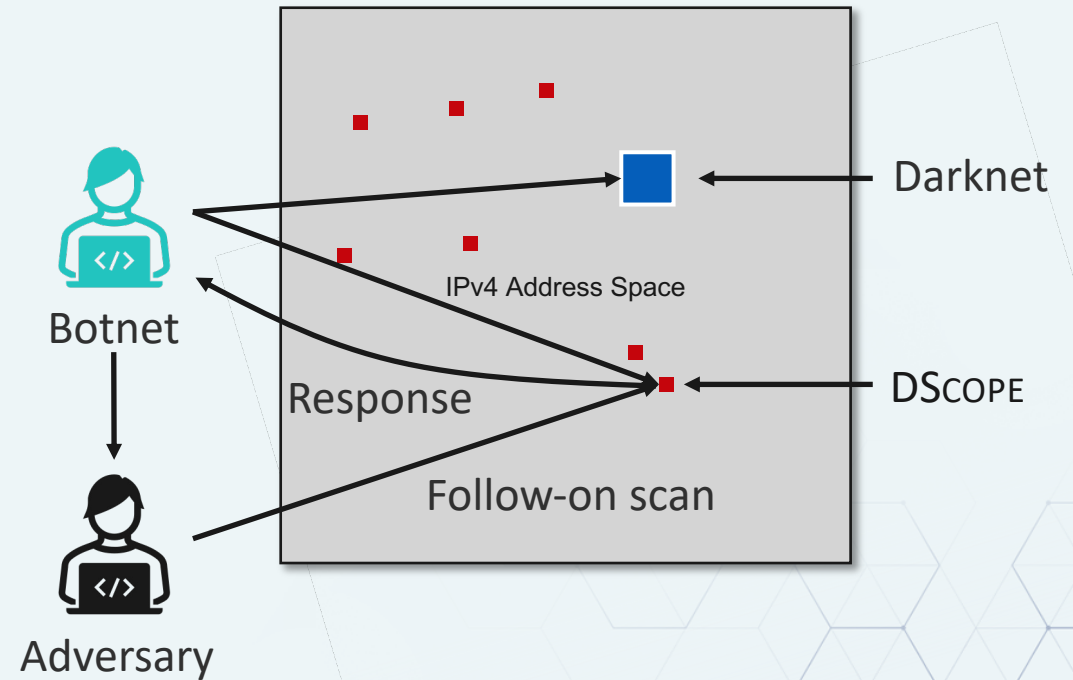
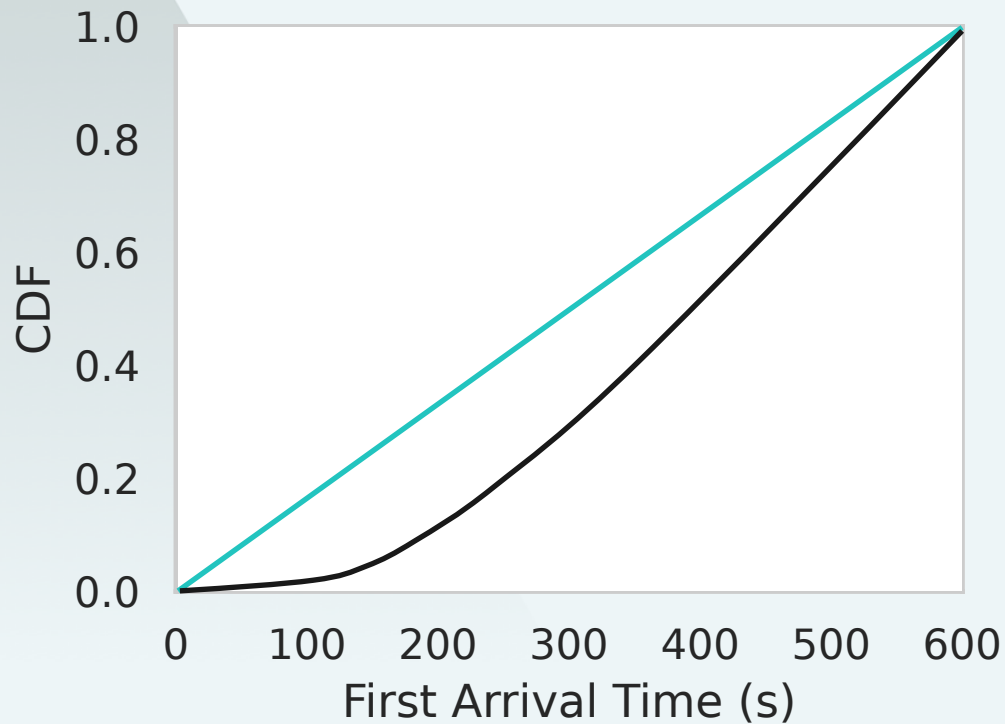
Recall: Null-Hypothesis of Random Scanning



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

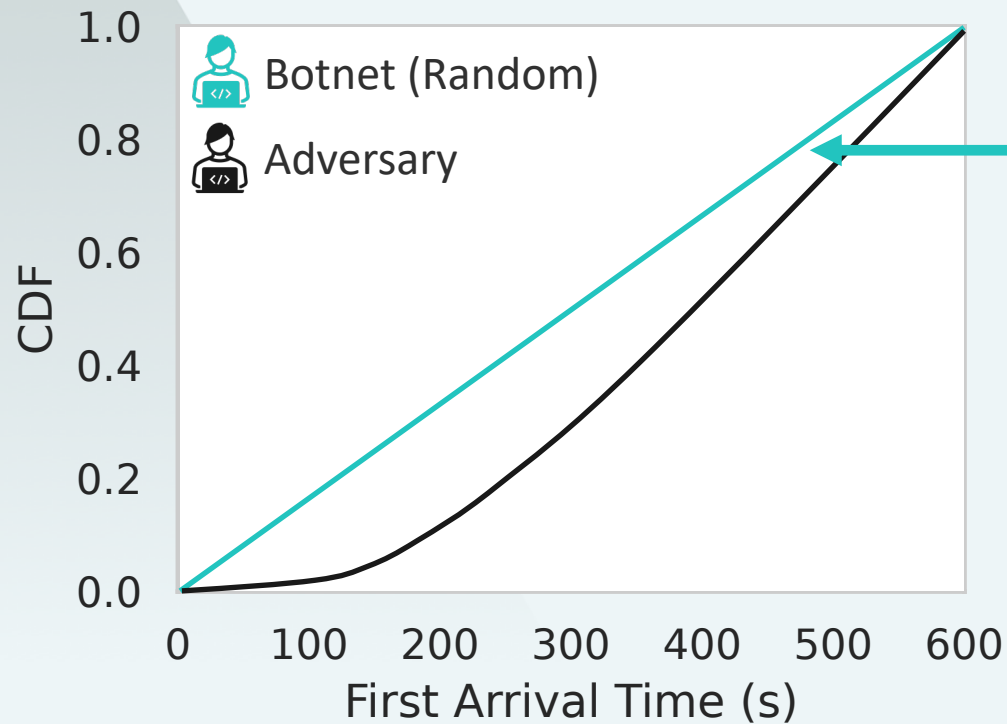
Interactivity: Service Lifecycle and follow-on scans

Question: does **interactivity** induce adversarial response?



Interactivity: Service Lifecycle and follow-on scans

Question: does **interactivity** induce adversarial response?



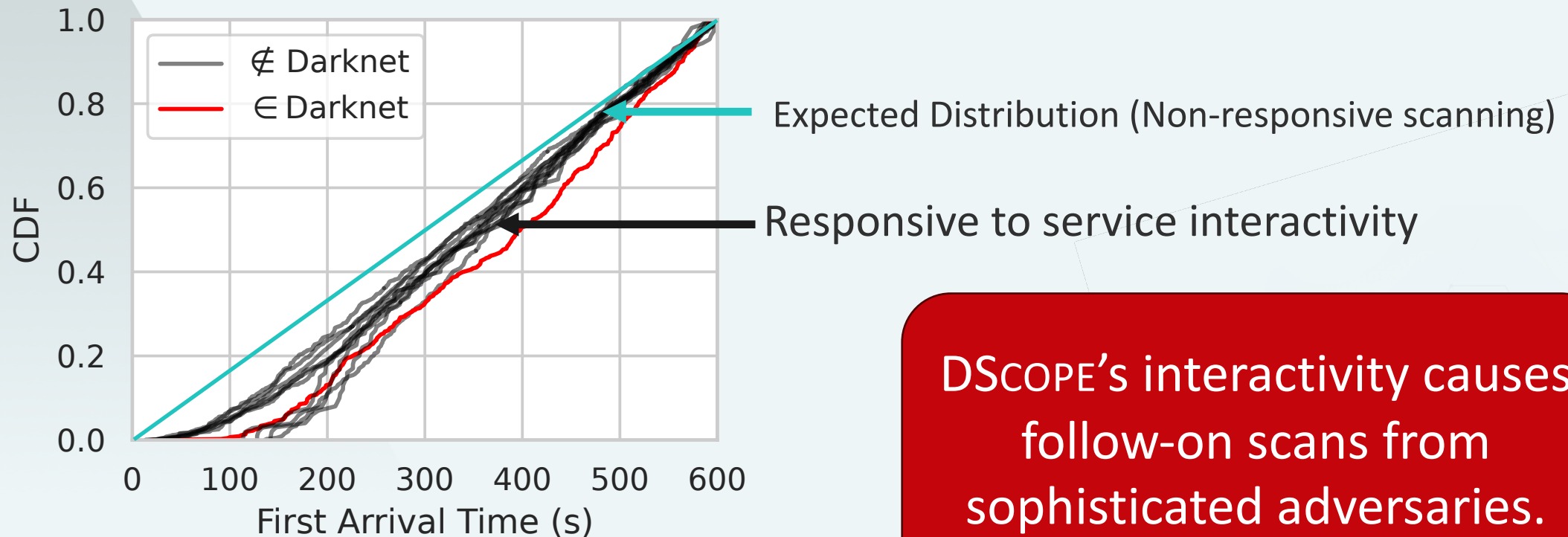
Expected Distribution (Non-responsive scanning):

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

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

Interactivity: Service Lifecycle and follow-on scans

Question: does **interactivity** induce adversarial response?



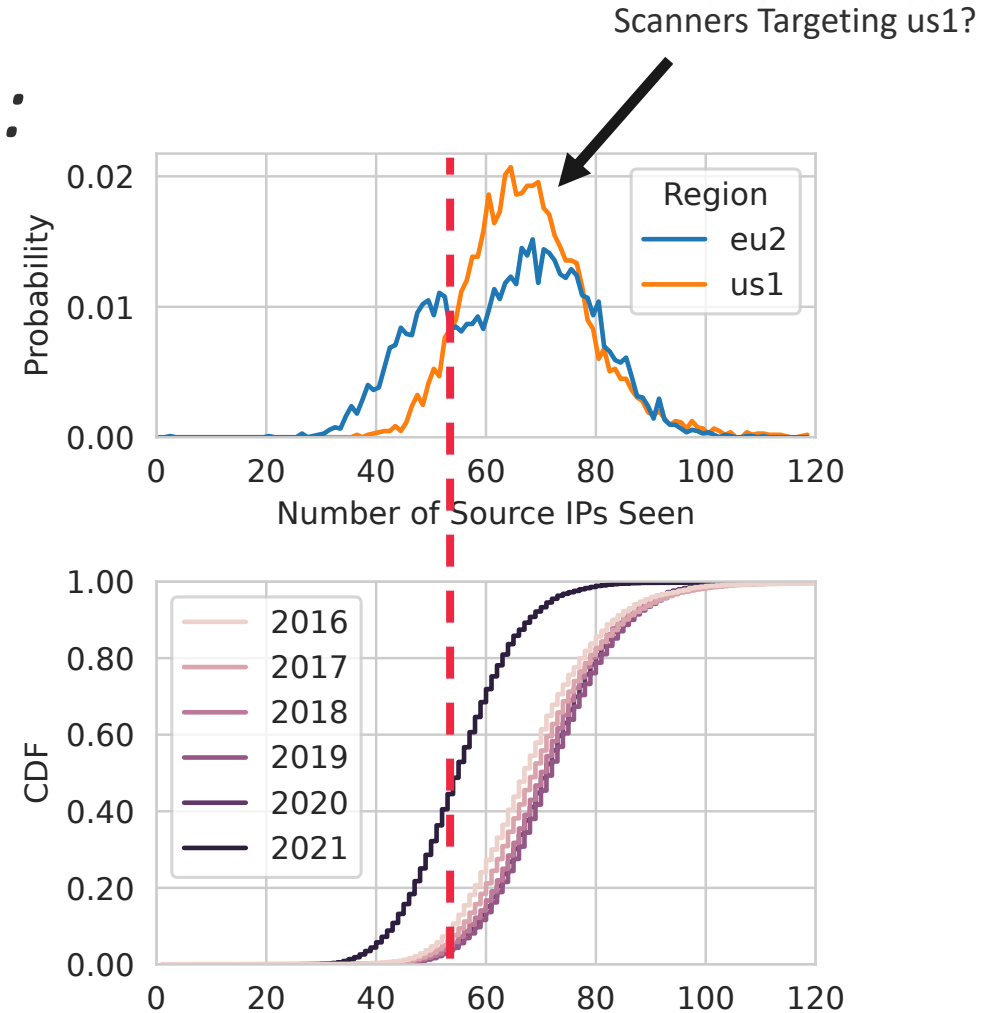
DSCOPE's interactivity causes follow-on scans from sophisticated adversaries.

Cloud Traffic Distributions & Statistical Validity

Challenge: Every cloud IP is *unique*:

- IP address history
- Latent configuration

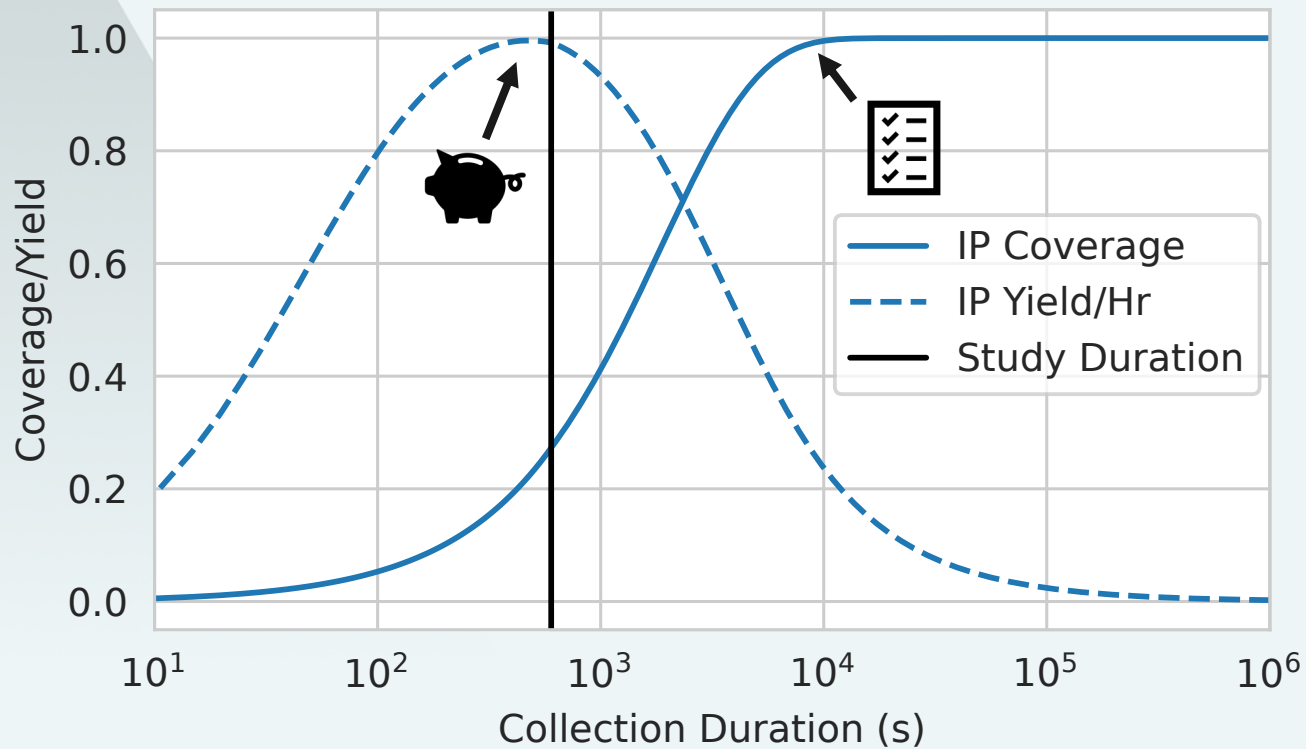
DSCOPE's large footprint allows for elimination of confounding factors.





Cost Optimization: How long should DSCOPE hold IPs?

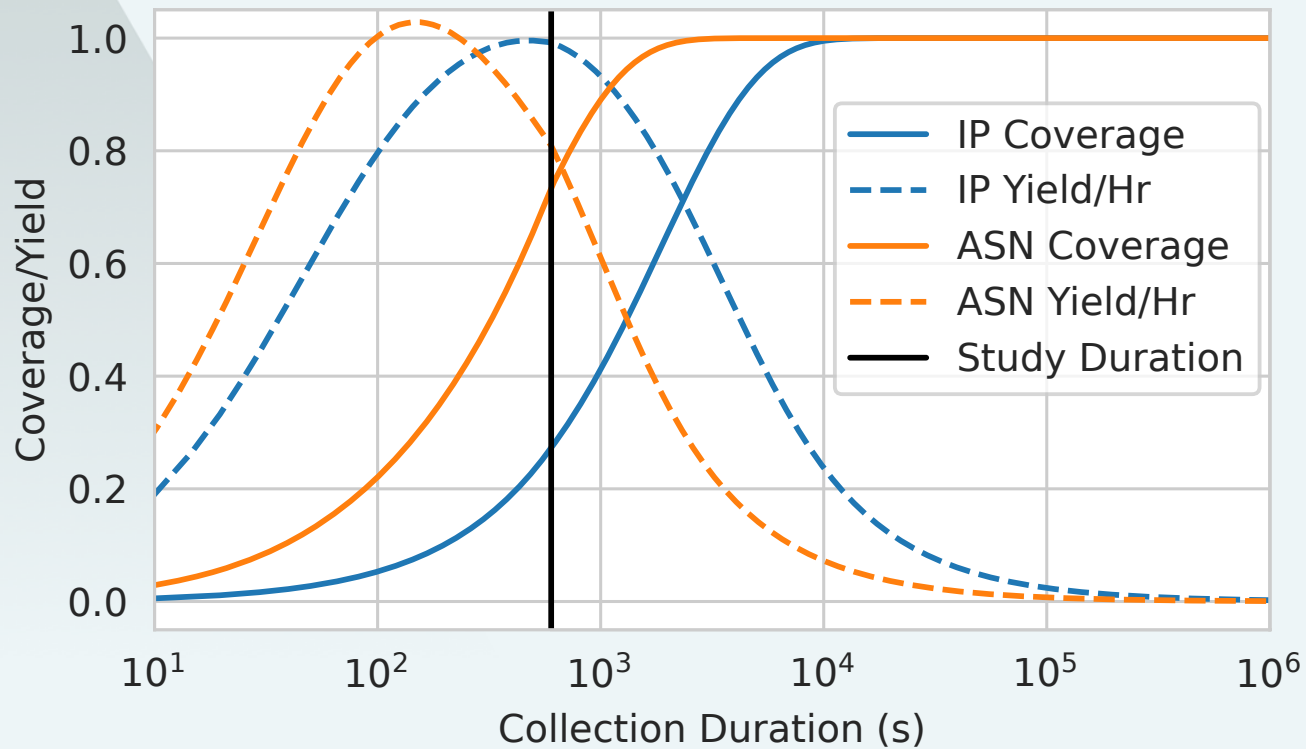
Goal: Max coverage with min cost (IP-hours)





Cost Optimization: How long should DSCOPE hold IPs?

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



DSCOPE's deployment can be optimized for high coverage or economical yield of Internet phenomena.

Takeaways – DSCOPE achieves:



Representative Traffic and Global Coverage



Interactivity & Service Lifecycle



Agility through IP Space



Price Performance



One more thing...





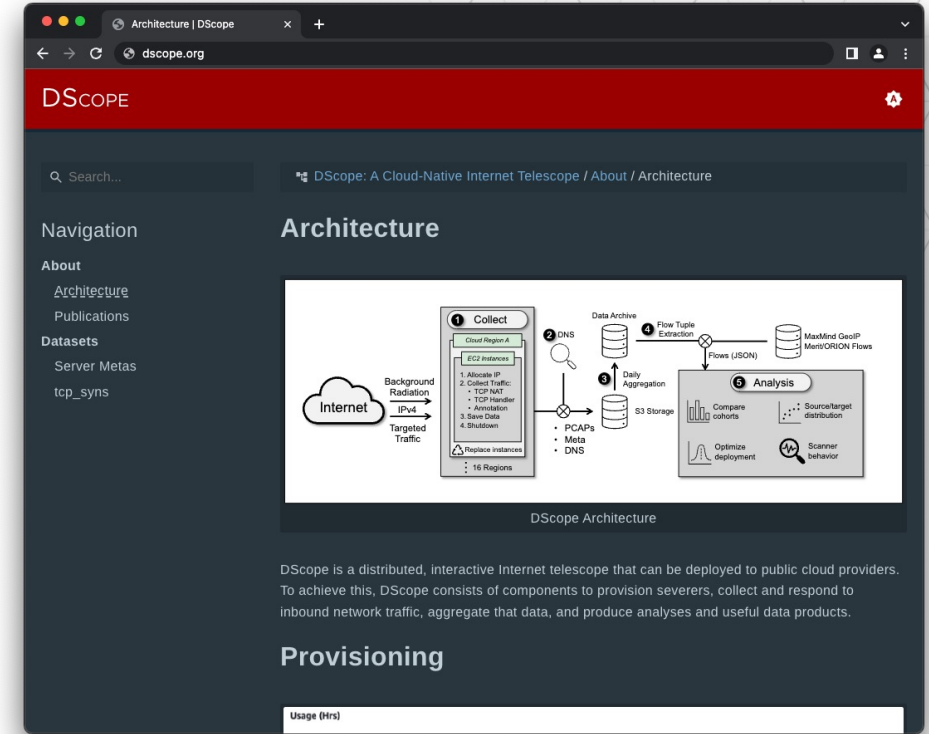
DSCOPE.ORG and Open Data

Data Products

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

Interactive Visualizations

- Emergent Threats
- Cloud Scanning
- Deployment Health



epauley@cs.wisc.edu



dscope.org/sec23

