DSCOPE: A Cloud-Native Internet Telescope

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



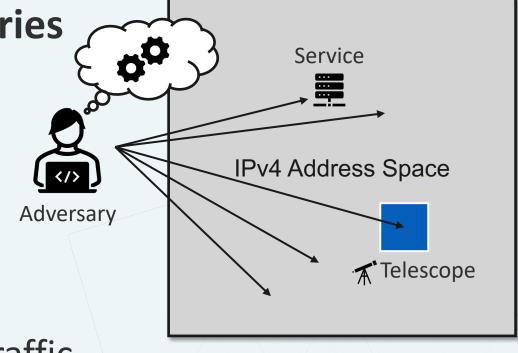
Background: Internet Scanning & Internet Telescopes

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

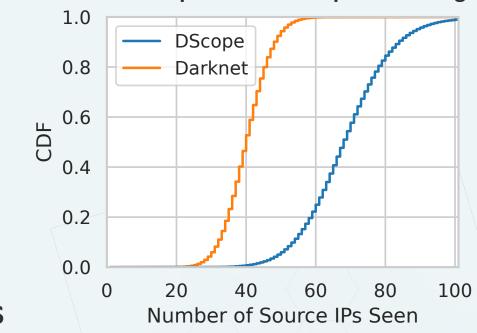
Passive measurement is incomplete



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

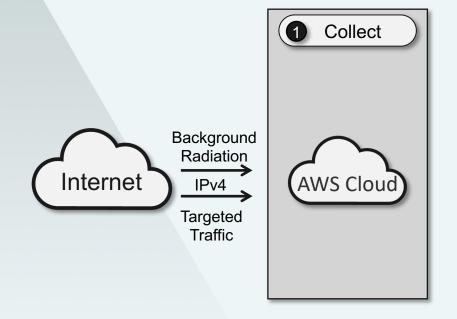


Agile through the IP address space

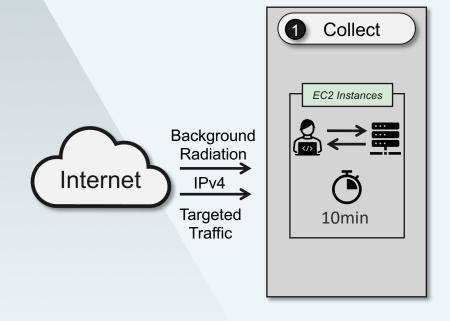
IP footprint varies over time



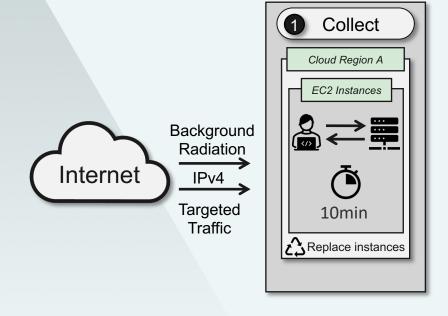




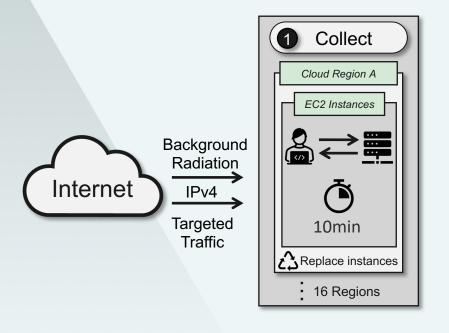




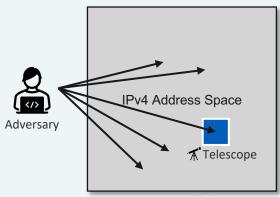




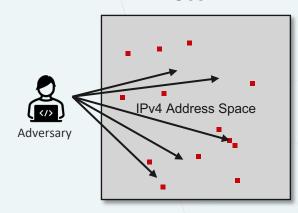




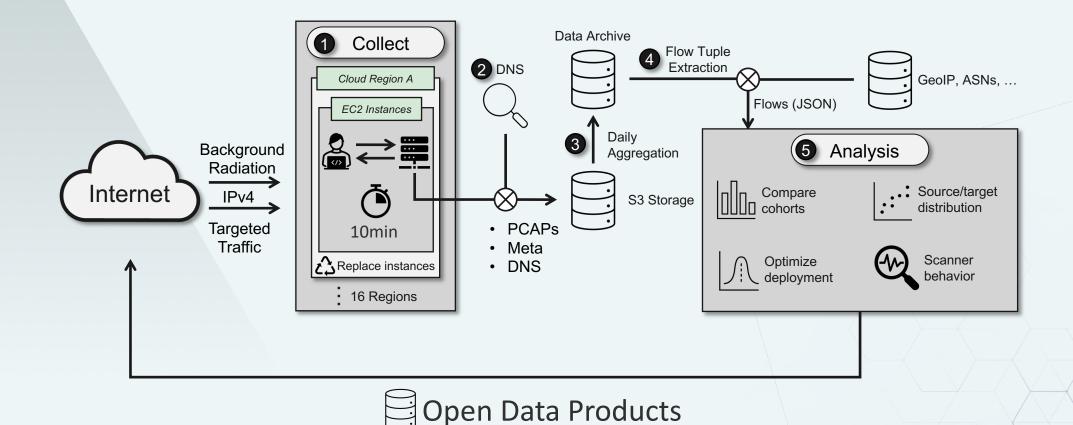
Darknet (Conventional) Telescope:



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		
Cloud Targ	geting (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		
Interactivii	ty & 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.	$6 \times$ 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× background rate for region/port		
(F14)	Scanners show minimal preference to groups of regions in similar geographies.	0.02 lower overlap in same-geography		
	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		



Results: 18 findings on cloud-based Internet measurement

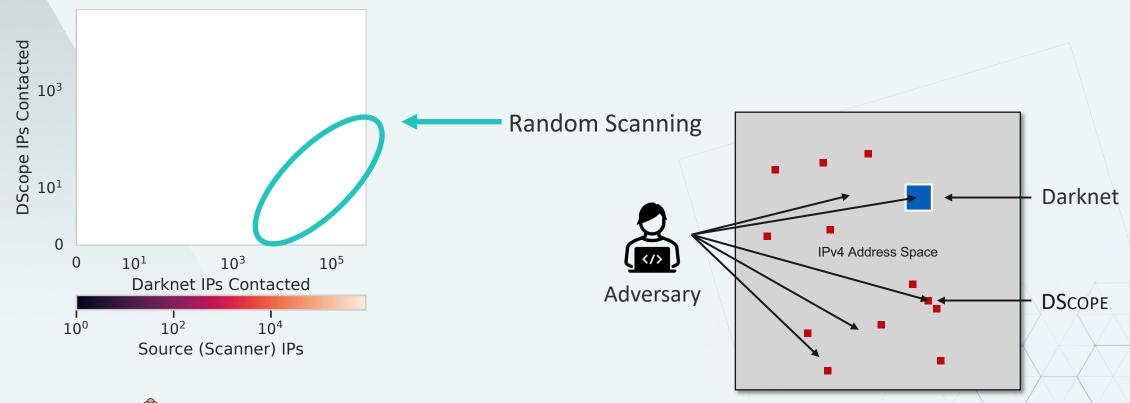
	Finding	Metric	
Cloud Tar	geting (Section 4)		
(F1) (F2)	An interactive cloud telescope receives traffic from substantially more IP addresses. Cloud IP traffic is more variable than darknets.	73% more traffic	
(F3)	Scanners target cloud IP ranges or avoid telescopes.	450× higher than expected	under H_0
(F4) (F5)	Scanners that are seen by both darknet/cloud telescopes are largely untargeted. Scans targeting existing telescopes are primarily random.	N/A N/A	
Interactivi	ity & Service Lifecycle (Section 5)		
(F6) (F7)	Some scanner IPs demonstrate clearly non-random behavior. Delayed scanners leverage information from other sources to target responsive IPs	1.7% of traffic $(p < 10^{-4})$ > 90% discernible source	
(F8)	Delayed scanners are not seen by existing darknet telescopes.	90% telescope avoidance (p	$< 10^{-4}$)
Intra-clou	d Targeting (Section 6)		
(F9) (F10)		$0.\pm0.3\sigma$ variation between respect to 0.3σ	gions
(F11)	Scanners target cloud IP addresses based on outdated data.	21% fewer scanners to 2021	AWS IPs
(F12) (F13) (F14)	Traffic to individual regions is largely consistent with untargeted scanning. Some sophisticated scanners precisely target physical regions within cloud IP blocks. Scanners show minimal preference to groups of regions in similar geographies.		_
Ontimizin	a Callection (Section 7)		
ODUMUZ.M	E Collection (Section /)		
(F15) (F16)	Observed traffic increases over time after instance deployment, but only to a point. Scanners targeting ORION are less likely to be reactive.	67% increase	
(F15)	Observed traffic increases over time after instance deployment, but only to a point.		
	(F1) (F2) (F3) (F4) (F5) Interactive (F6) (F7) (F8) Intra-cloud (F9) (F10) (F11) (F12) (F13) (F14)	Cloud Targeting (Section 4) (F1) An interactive cloud telescope receives traffic from substantially more IP addresses. (F2) Cloud IP traffic is more variable than darknets. (F3) Scanners target cloud IP ranges or avoid telescopes. (F4) Scanners that are seen by both darknet/cloud telescopes are largely untargeted. (F5) Scans targeting existing telescopes are primarily random. Interactivity & Service Lifecycle (Section 5) (F6) Some scanner IPs demonstrate clearly non-random behavior. (F7) Delayed scanners leverage information from other sources to target responsive IPs (F8) Delayed scanners are not seen by existing darknet telescopes. Intra-cloud Targeting (Section 6) (F9) Quantity of scanners differs across cloud regions, but intra-region variance dominates (F10) Source IP variance differs between regions (F11) Scanners target cloud IP addresses based on outdated data. (F12) Traffic to individual regions is largely consistent with untargeted scanning. (F13) Some sophisticated scanners precisely target physical regions within cloud IP blocks.	Cloud Targeting (Section 4) (F1) An interactive cloud telescope receives traffic from substantially more IP addresses. (F3) Scanners target cloud IP ranges or avoid telescopes. (F4) Scanners that are seen by both darknet/cloud telescopes are largely untargeted. (F5) Scans targeting existing telescopes are primarily random. Interactivity & Service Lifecycle (Section 5) (F6) Some scanner IPs demonstrate clearly non-random behavior. (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. 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 (F11) Scanners target cloud IP addresses based on outdated data. (F12) Traffic to individual regions is largely consistent with untargeted scanning. (F13) Some sophisticated scanners precisely target physical regions within cloud IP blocks. (F14) Scanners show minimal preference to groups of regions in similar geographies. (F15) Outcome IP variance differs here are primarily random. (F16) Source IP variance differs here are primarily random. (F17) Intraffic to individual regions is largely consistent with untargeted scanning. (F17) Intraffic to individual regions is largely consistent with untargeted scanning. (F17) Scanners show minimal preference to groups of regions in similar geographies. (F18) Outcome IP addresses are are primarily random. (F19) Note traffic (p < 10 ⁻⁴) (F19) Power traffic (p < 10 ⁻⁴) (F10) Power traffic (p < 10 ⁻⁴) (F10) Power traffic (p < 10 ⁻⁴) (F10)





Coverage: Is Internet Scanning Random?

Recall: Null-Hypothesis of Random Scanning

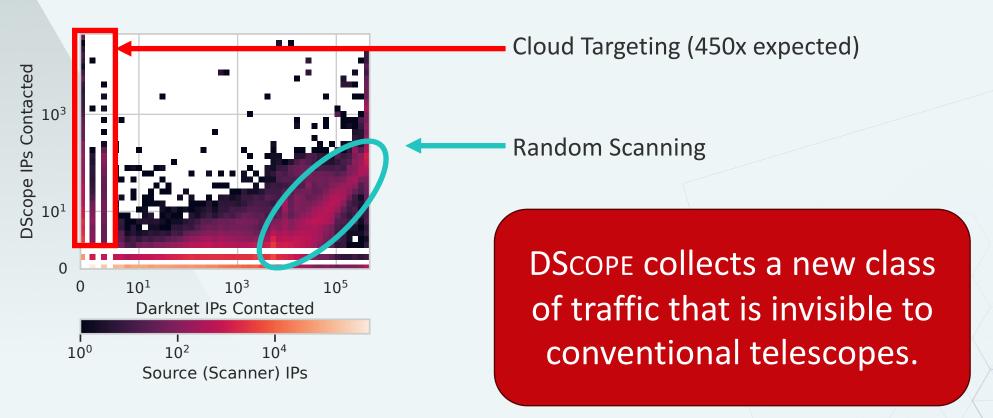






Coverage: Is Internet Scanning Random?

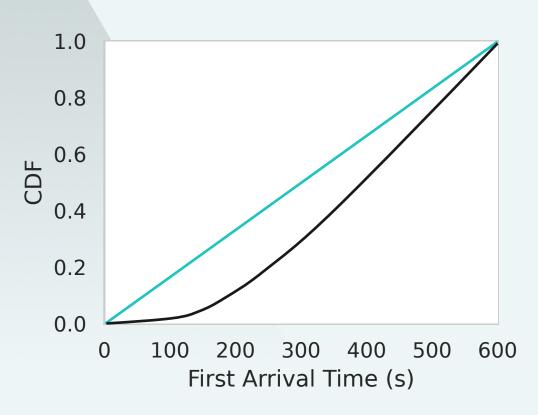
Recall: Null-Hypothesis of Random Scanning

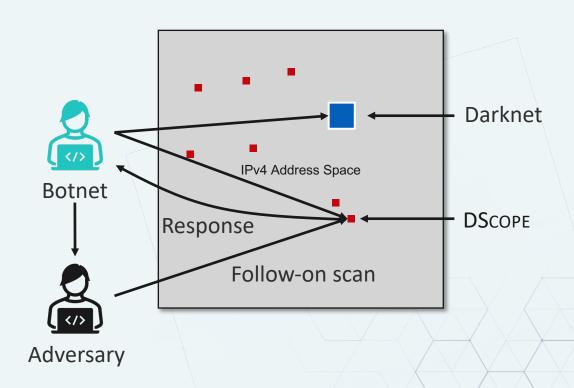




[日] Interactivity: Service Lifecycle and follow-on scans

Question: does interactivity induce adversarial response?

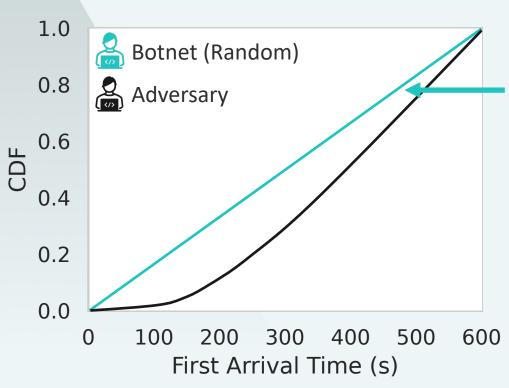






(目Interactivity: Service Lifecycle and follow-on scans

Question: does interactivity induce adversarial response?



Expected Distribution (Non-responsive scanning):

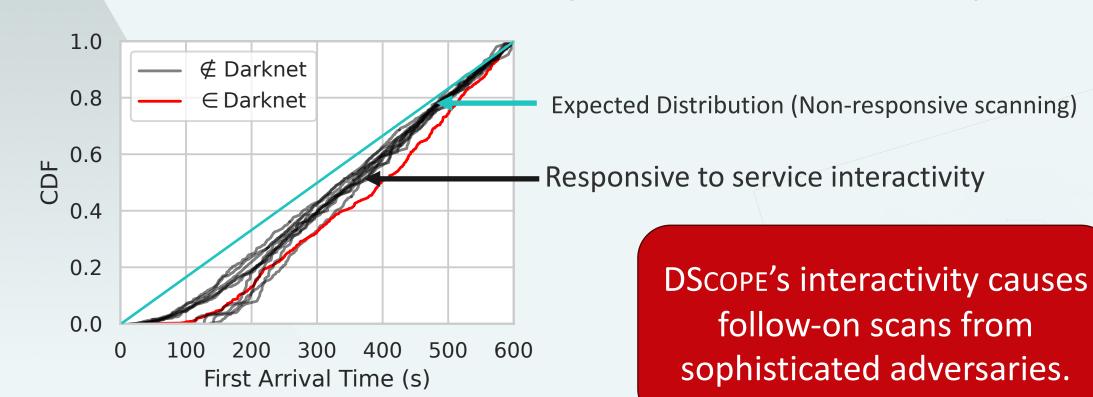
$$f_T(t) = \frac{\lambda e^{-\lambda t}}{1 - e^{-\lambda m}} \qquad (0 \le t \le m)$$

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



(目Interactivity: Service Lifecycle and follow-on scans

Question: does interactivity induce adversarial response?



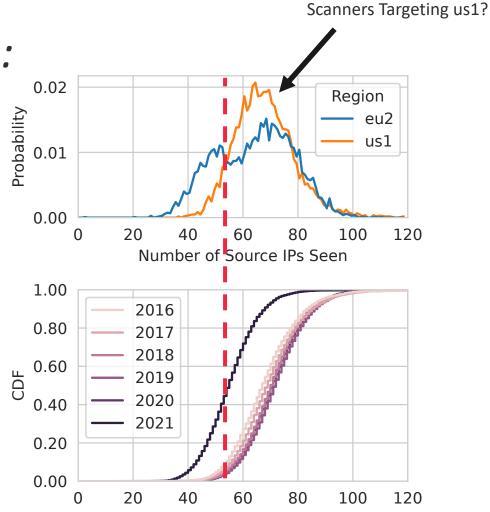


⚠ Cloud Traffic Distributions & Statistical Validity

Challenge: Every cloud IP is unique:

- IP address history
- Latent configuration

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

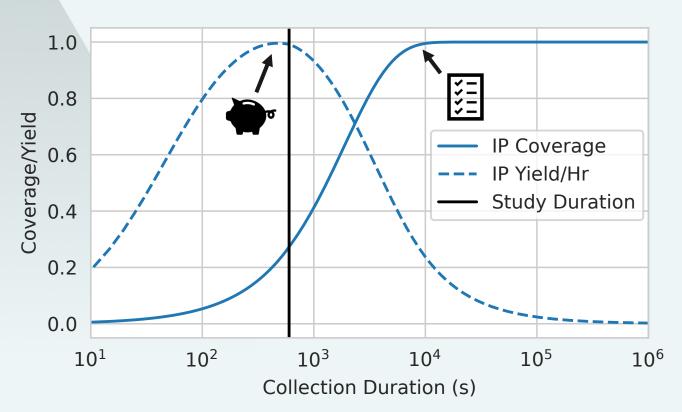






*Cost Optimization: How long should DScope hold IPs?

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

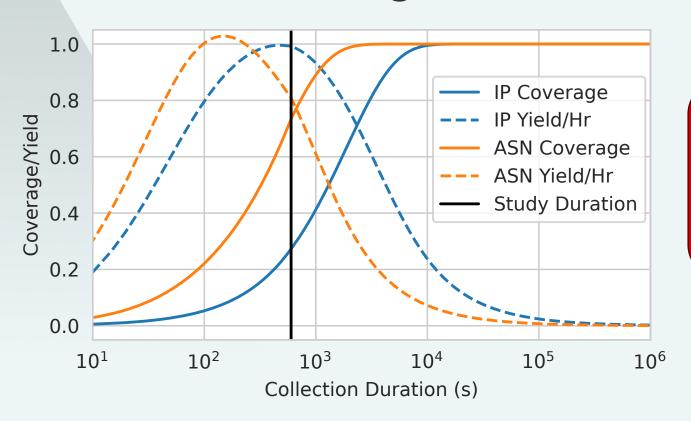






**Cost Optimization: How long should DScope hold IPs?

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

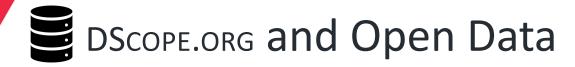


Price Performance



One more thing...



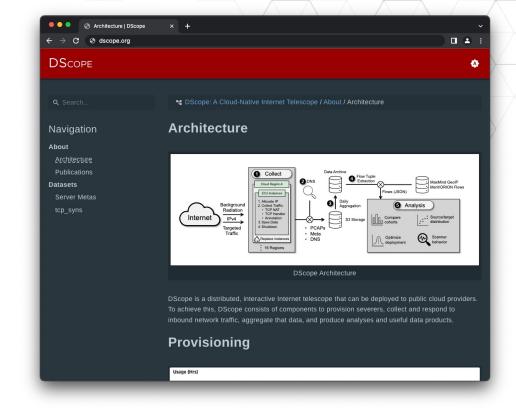


Data Products

- 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



