Scanning the Internet for Liveness

Shehar Bano (*)
Facebook

Philipp Richter
Akamai Technologies
MIT

Mobin Javed
LUMS, Pakistan

Srikanth Sundaresan
Facebook

Zakir Durumeric
Stanford

Steven Murdoch
University College London

Richard Mortier
University of Cambridge

Vern Paxson
UC Berkeley, ICSI

Presented by: Shehar Bano
Research Scientist (Novi, Facebook)

bano@fb.com @thatBano

Paper pdf:
ACM SIGCOMM CCR 2018

(*) This work was carried out while working at UCL and Cambridge University

Internet Scanning

- A key technique to measure the Internet at scale
- Diverse applications:
  - Address space utilization
  - Host reachability
  - Topology
  - Service availability
  - Security vulnerabilities
  - Service discrimination

IP Liveness

- Does a target IP address respond to a probe packet?

IP Liveness

- Does a target IP address respond to a probe packet?

- Key Questions:
  - What type of probe packets should we send if we, for example, want to maximize the responding host population?

IP Liveness

- Does a target IP address respond to a probe packet?

- Key Questions:
  - What type of probe packets should we send if we, for example, want to maximize the responding host population?
  - What type of responses can we expect and which factors determine such responses?

IP Liveness

- Does a target IP address respond to a probe packet?

- Key Questions:
  - What type of probe packets should we send if we, for example, want to maximize the responding host population?
  - What type of responses can we expect and which factors determine such responses?
  - What degree of consistency can we expect when probing the same host with different probe packets?

Challenges

- Missing a systematic framework that allows us to understand IP liveness and, how it manifests in the form of host replies to active probing

Challenges

- Missing a systematic framework that allows us to understand IP liveness and, how it manifests in the form of host replies to active probing

- Depends on multiple factors:
  - How the scan was conducted

Challenges

- Missing a systematic framework that allows us to understand IP liveness and, how it manifests in the form of host replies to active probing

- Depends on multiple factors:
  - How the scan was conducted
  - How different protocols interact

Challenges

- Missing a systematic framework that allows us to understand IP liveness and, how it manifests in the form of host replies to active probing

- Depends on multiple factors:
  - How the scan was conducted
  - How different protocols interact
  - Filtering policies near the target IP

Challenges

- Missing a systematic framework that allows us to understand IP liveness and, how it manifests in the form of host replies to active probing

- Depends on multiple factors:
  - How the scan was conducted
  - How different protocols interact
  - Filtering policies near the target IP
  - Temporal churn in IP responsiveness

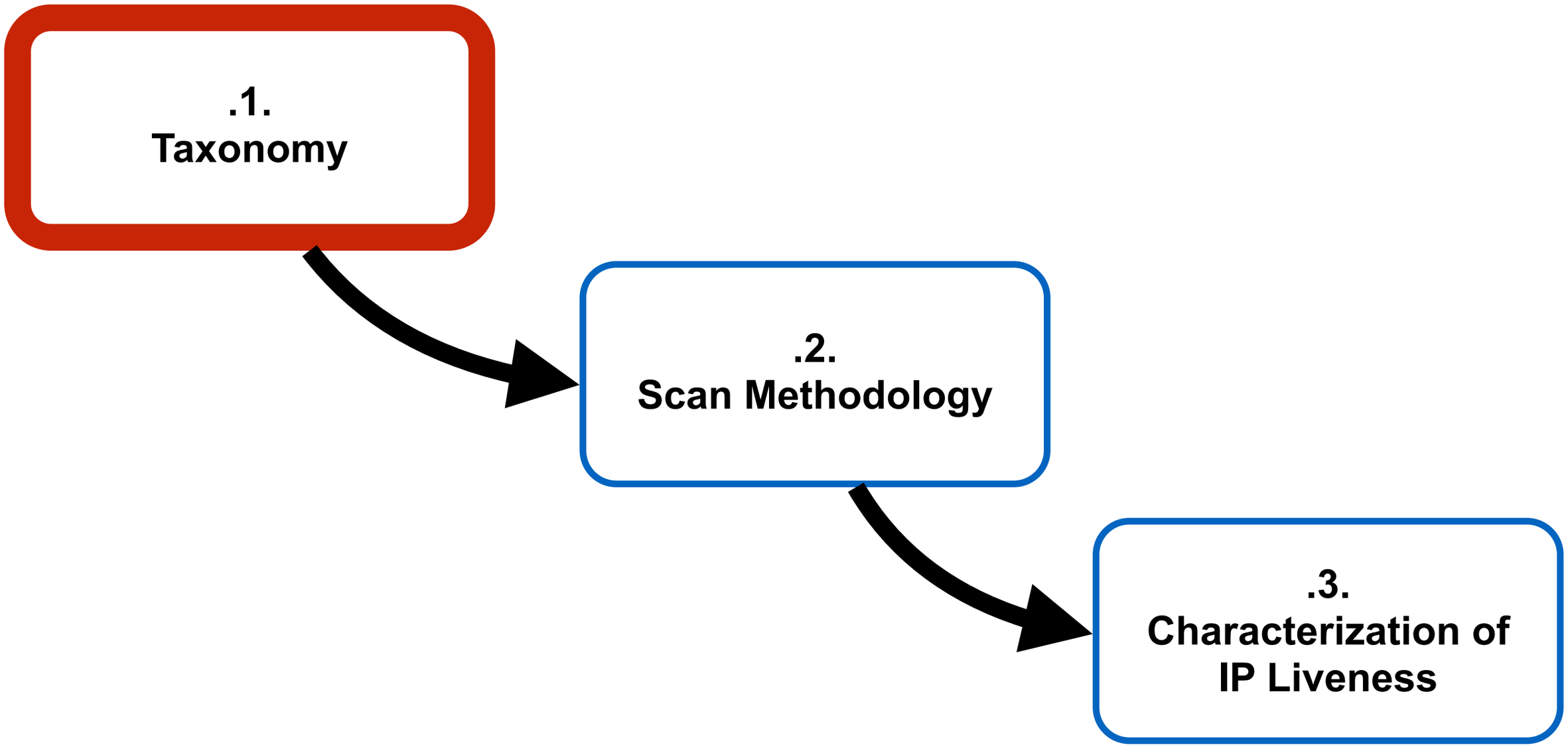
Contributions

- Taxonomy of Liveness

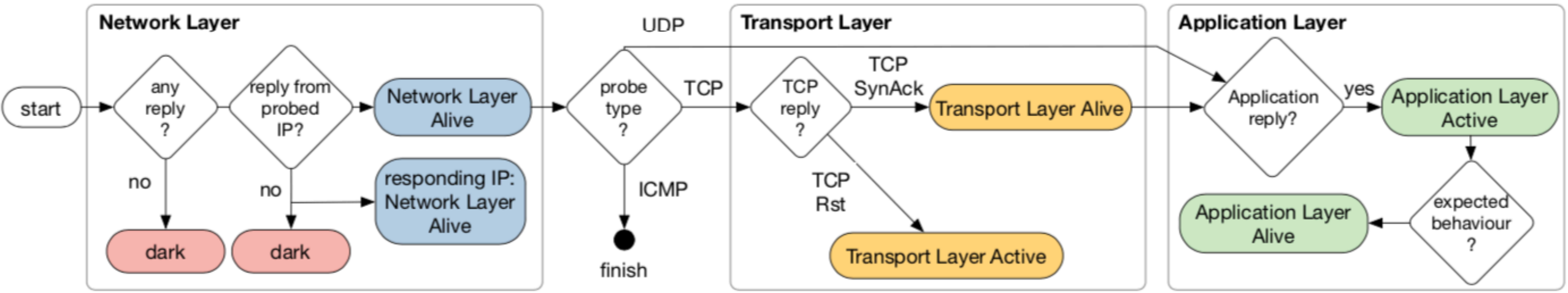
- Methodology for systematically inferring IP liveness by performing Internet-wide scans concurrently across a set of different protocols at various layers (ICMP, TCP, UDP)

- Analysis of gathered data to present an in-depth view of liveness

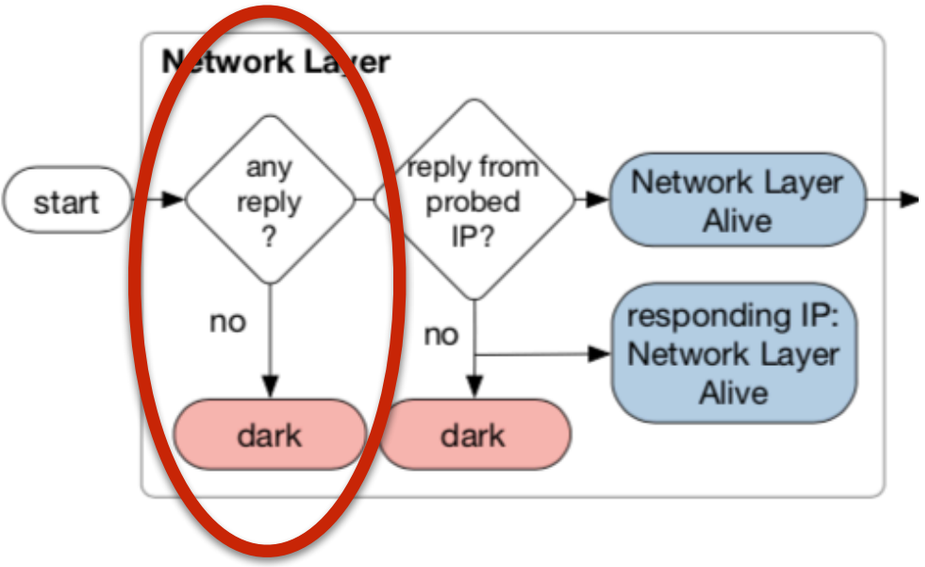
Roadmap


.1.
Taxonomy
.2.
Scan Methodology
.3.
Characterization of
IP Liveness

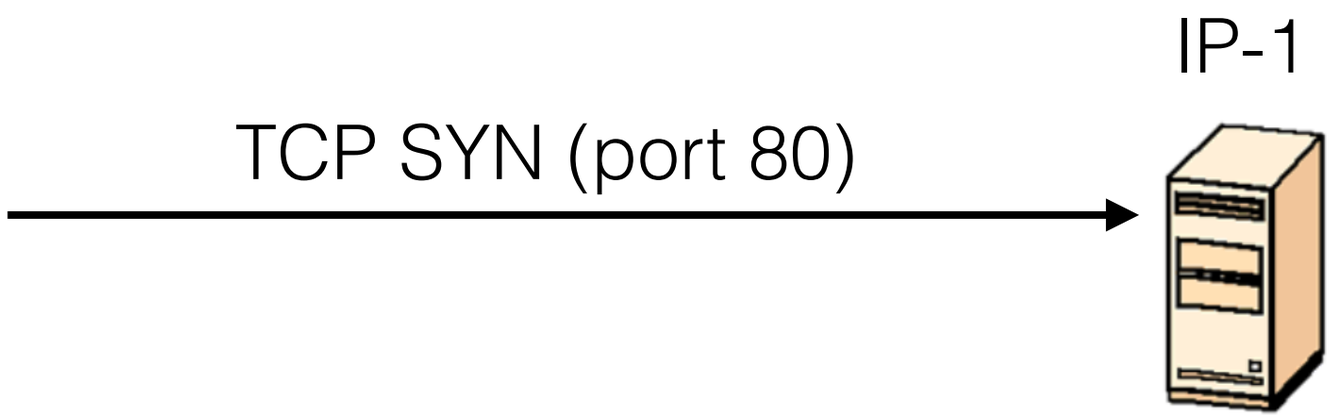
Taxonomy


Network Layer
start
any reply ?
reply from probed IP?
Network Layer Alive
no
no
dark
dark
responding IP: Network Layer Alive
UDP
probe type ?
TCP
ICMP
finish
Transport Layer
TCP reply ?
TCP SynAck
Transport Layer Alive
TCP Rst
Transport Layer Active
Application Layer
Application reply?
yes
Application Layer Active
expected behaviour ?
Application Layer Alive

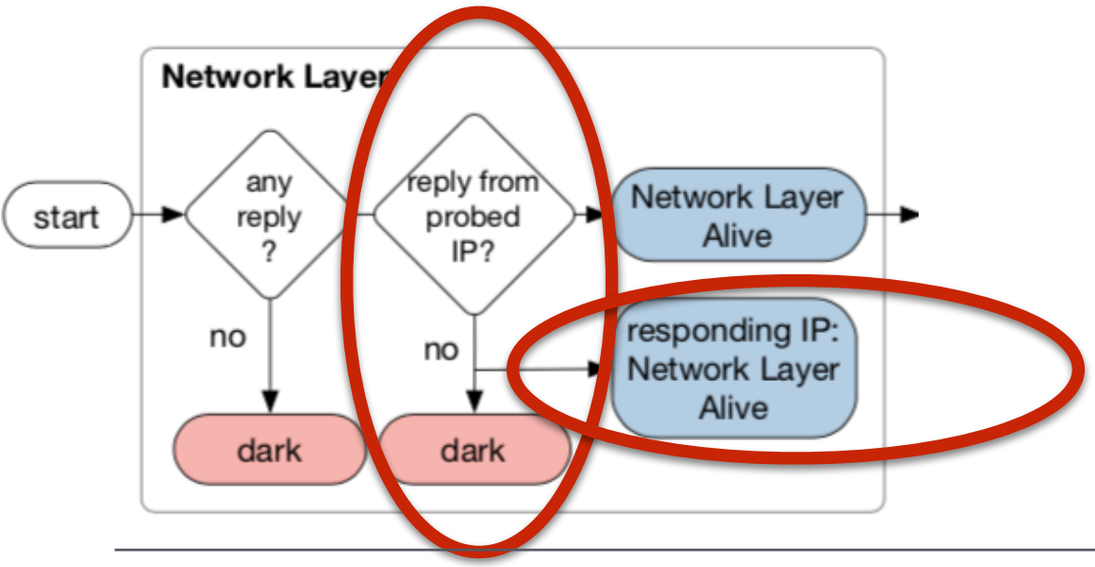
Taxonomy: Network Layer



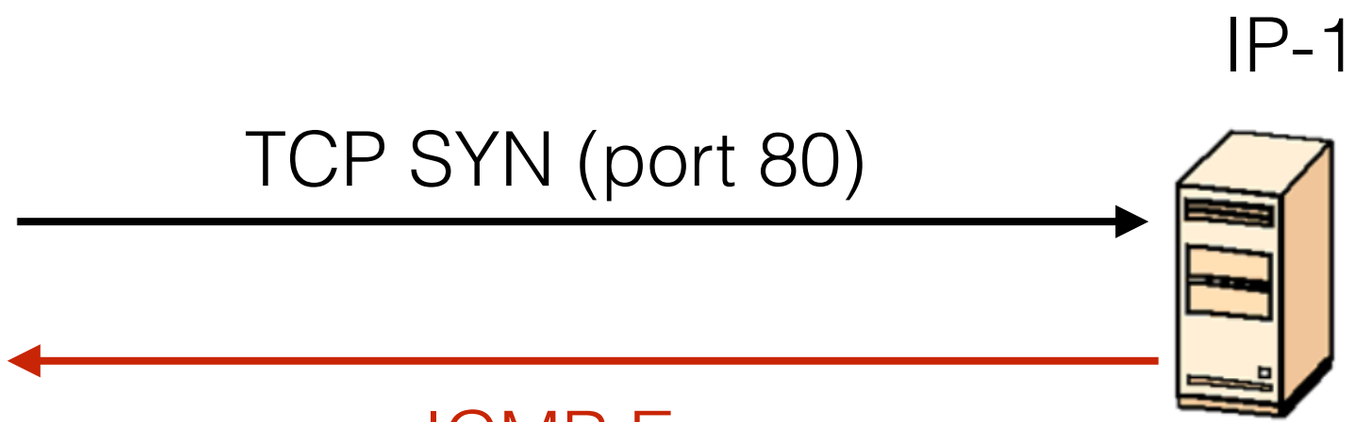
Network Layer Dark



Taxonomy: Network Layer

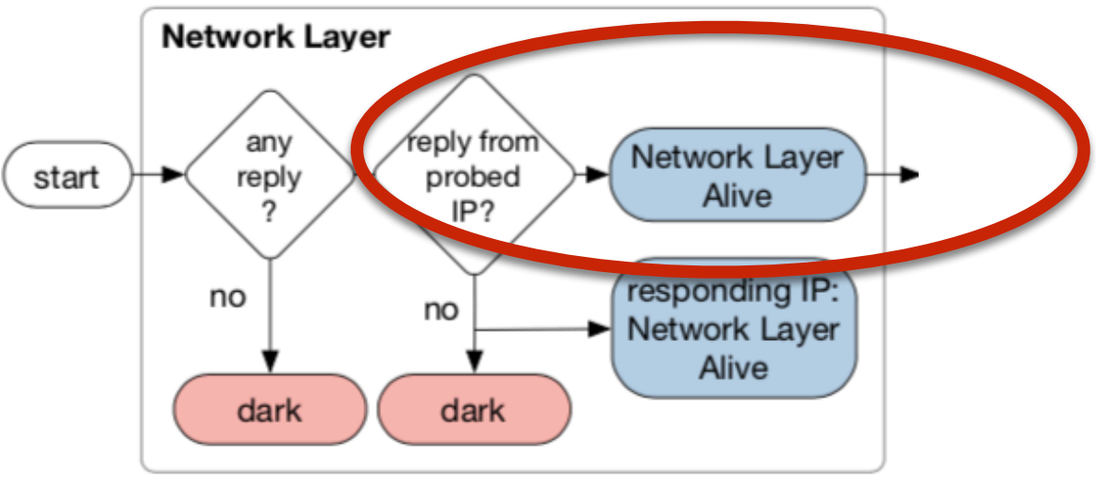


Network Layer Dark



Network Layer Alive

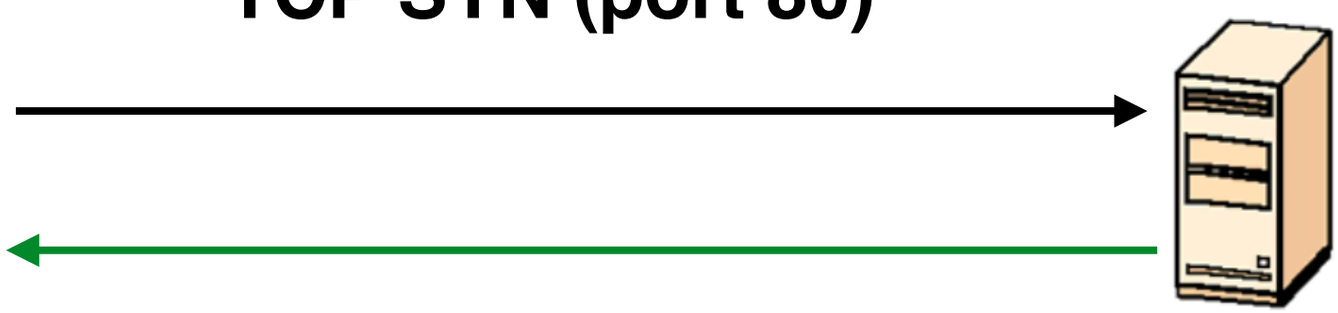
Taxonomy: Network Layer



Network Layer Alive

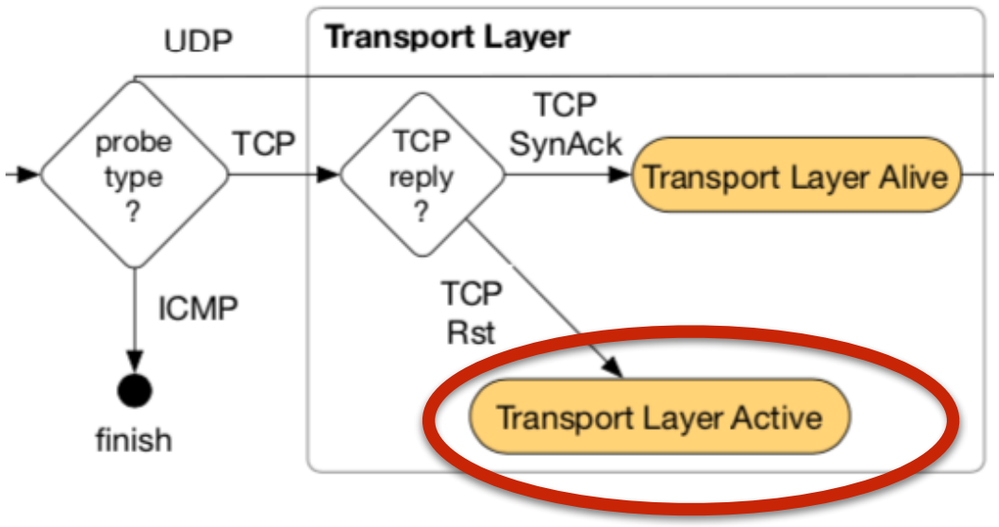
IP-1

TCP SYN (port 80)

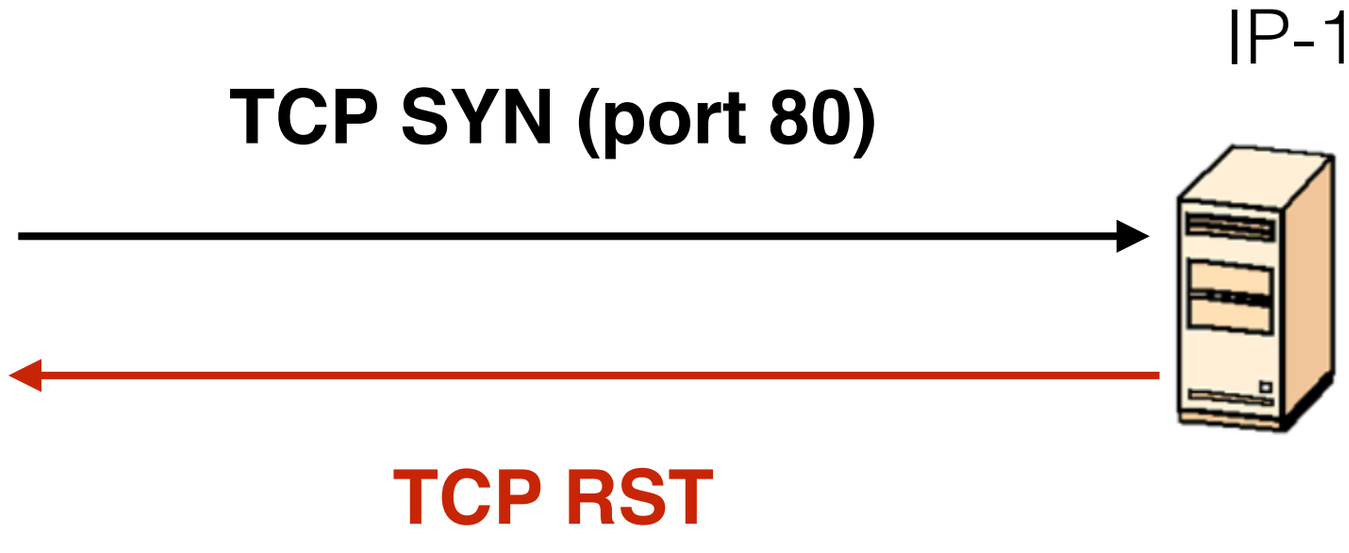


TCP SYN-ACK

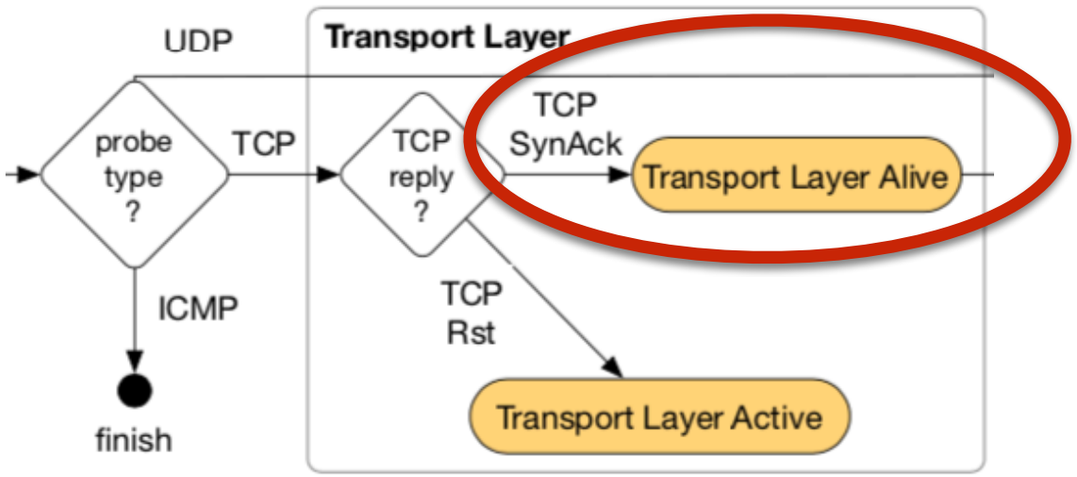
Taxonomy: Transport Layer (TCP)



Transport Layer Active



Taxonomy: Transport Layer (TCP)



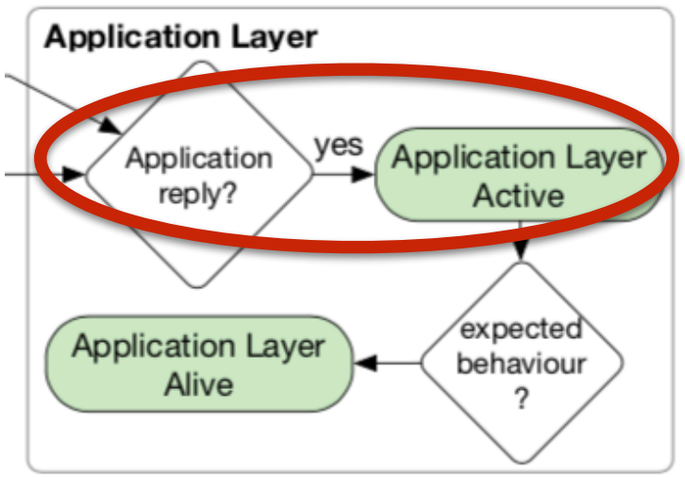
Transport Layer Alive

IP-1

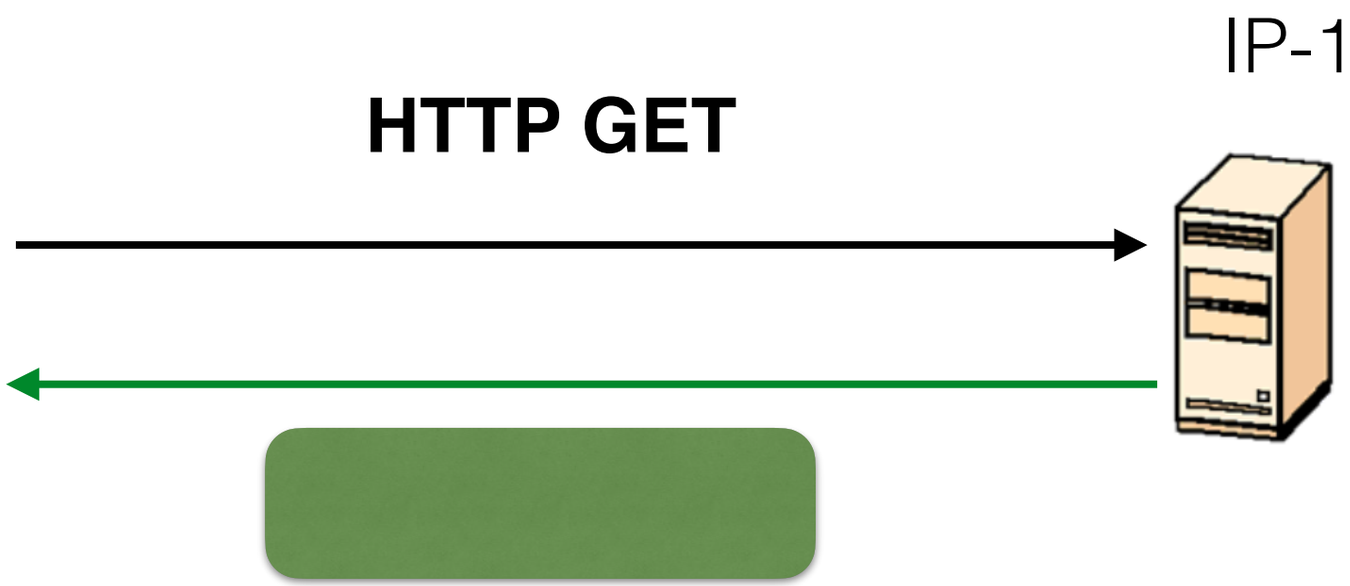
TCP SYN (port 80)

TCP SYN-ACK

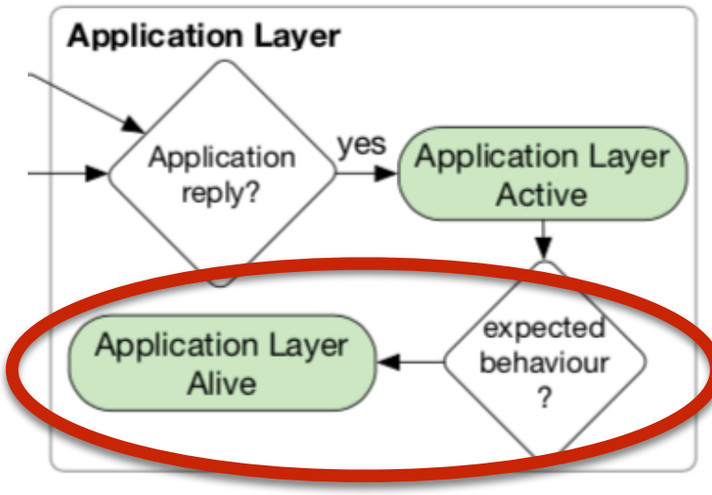
Taxonomy: Application Layer



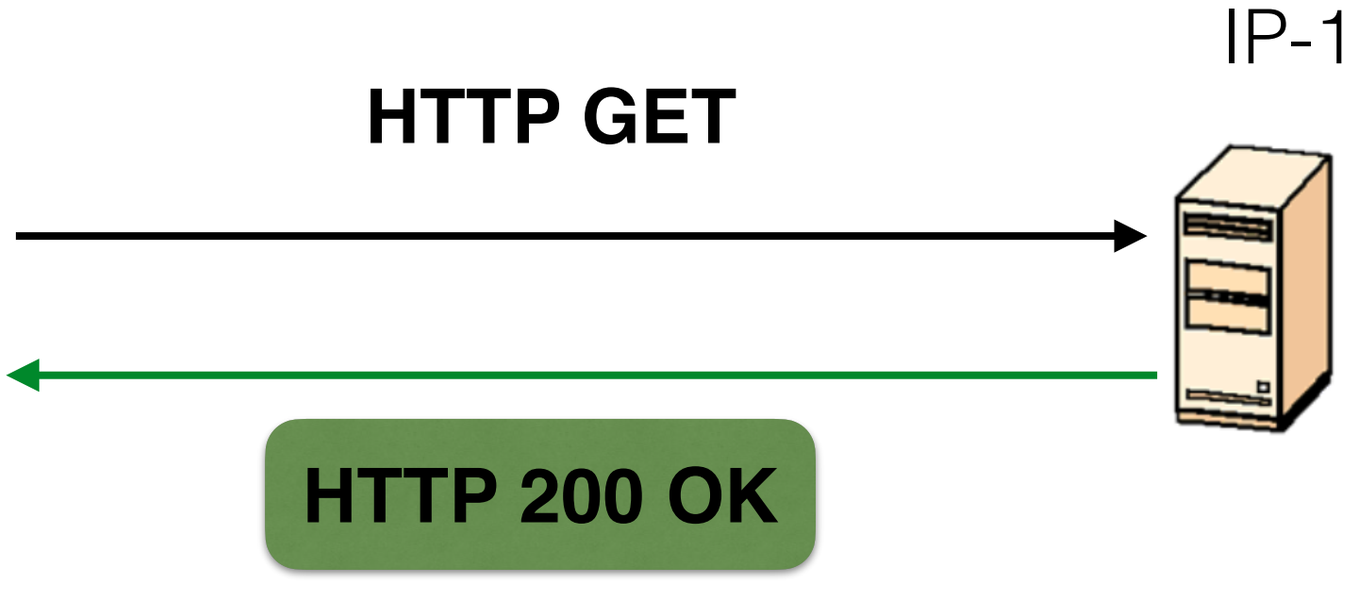
Application Layer Active



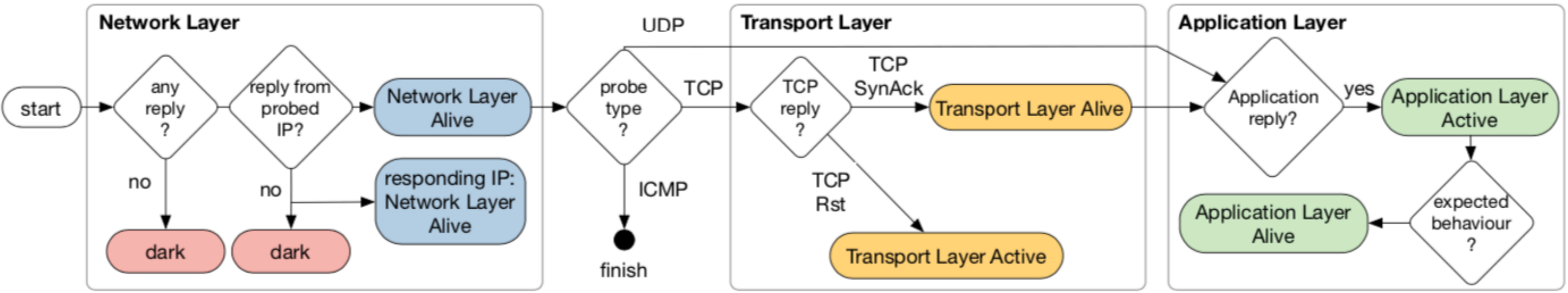
Taxonomy: Application Layer



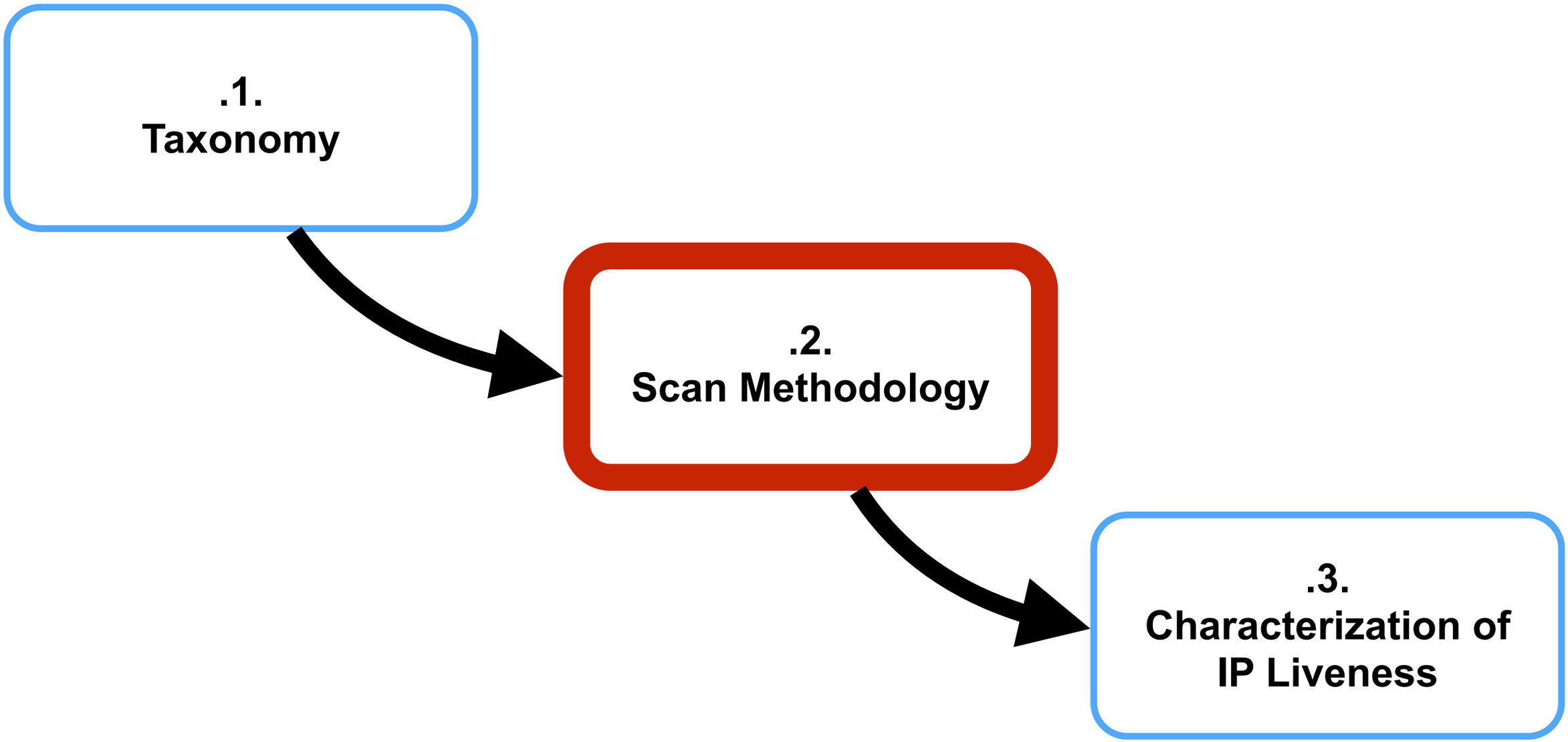
Application Layer Alive



Taxonomy


Network Layer
start
any reply ?
reply from probed IP?
Network Layer Alive
no
no
dark
dark
responding IP: Network Layer Alive
UDP
probe type ?
TCP
ICMP
finish
Transport Layer
TCP reply ?
TCP SynAck
Transport Layer Alive
TCP Rst
Transport Layer Active
Application Layer
Application reply?
yes
Application Layer Active
expected behaviour ?
Application Layer Alive

Roadmap


.1.
Taxonomy
.2.
Scan Methodology
.3.
Characterization of
IP Liveness

Scan Methodology

- 8 concurrent scans:
  - ICMP Echo scan
  - TCP Syn scans: Port 22 (SSH), 23 (Telnet), 80 (HTTP), 443 (HTTPS), and 7547 (CPE WAN Management Protocol, CWMP)
  - UDP-based applications: DNS and NTP

- ZMap (scan), SiLK (data analysis)

Scan Methodology: Considerations

- Temporal churn: Simultaneous scans

Scan Methodology: Considerations

- Temporal churn: Simultaneous scans
- Reply capture completeness: Record both positive and negative replies

Scan Methodology: Considerations

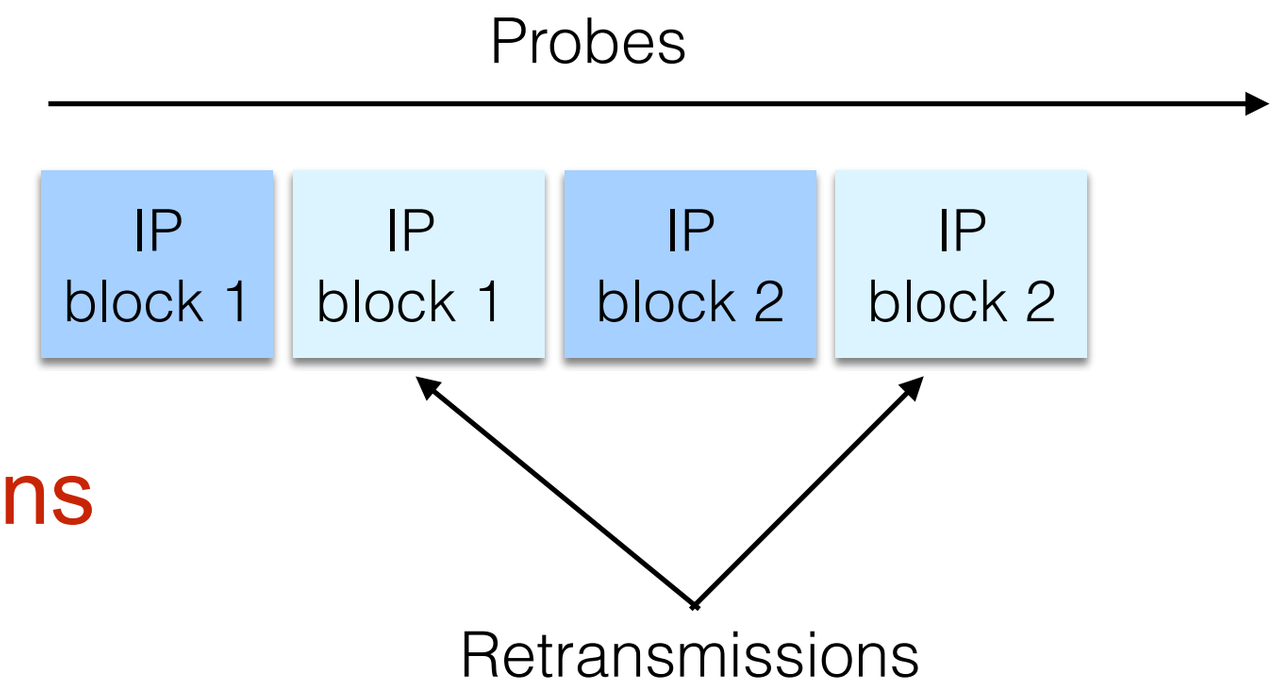
- Temporal churn: Simultaneous scans
- Reply capture completeness: Record both positive and negative replies

- Packet loss mitigation:
  - Redundant probes

Scan Methodology: Considerations

- Temporal churn: Simultaneous scans
- Reply capture completeness: Record both positive and negative replies

- Packet loss mitigation:
  - Redundant probes
  - Delayed retransmissions

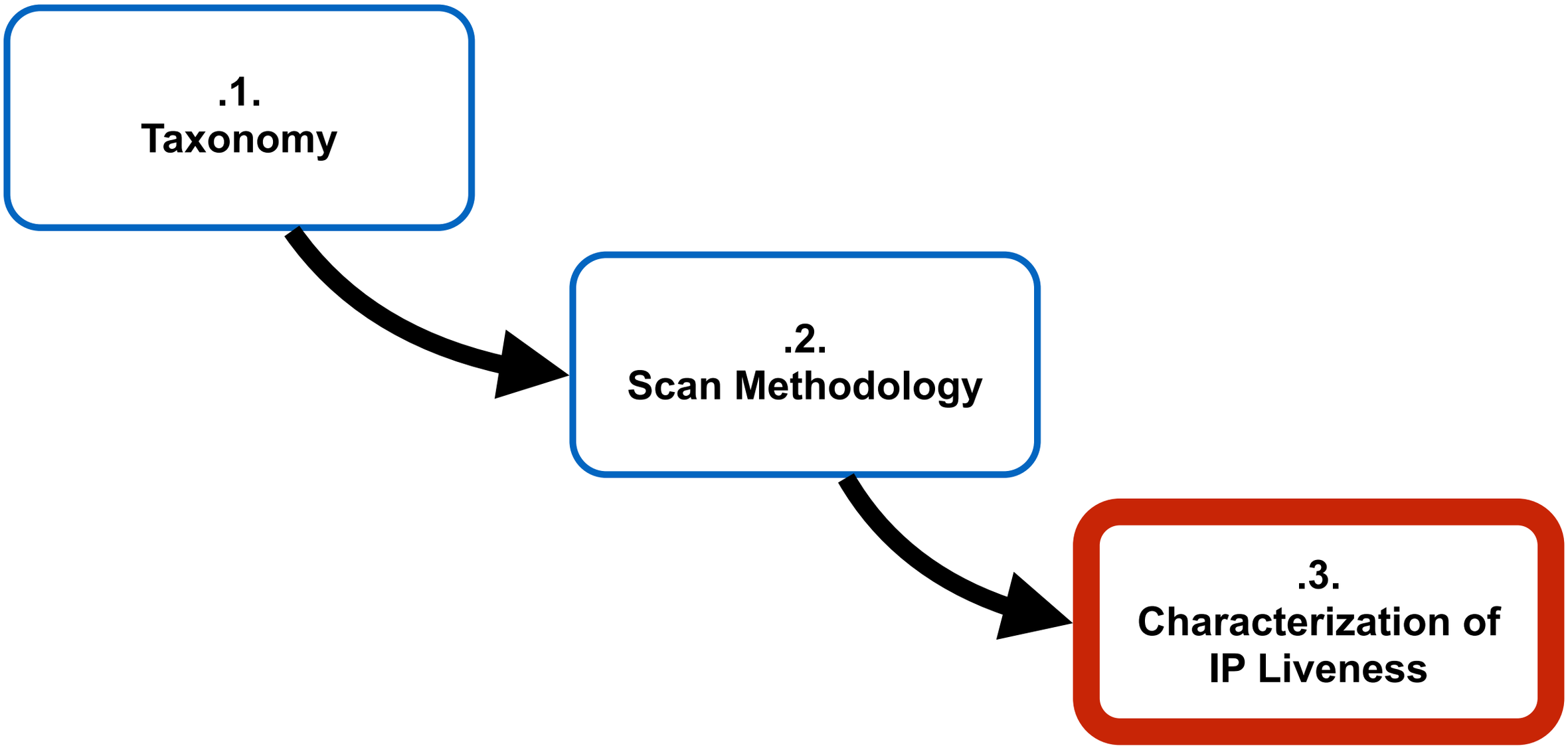


Scan Methodology

- 8 concurrent scans: ~24 hours, 2.3 TB data

- Overall, our scans recorded 487M network alive IPs out of 3.6B probed (*IP_all*)

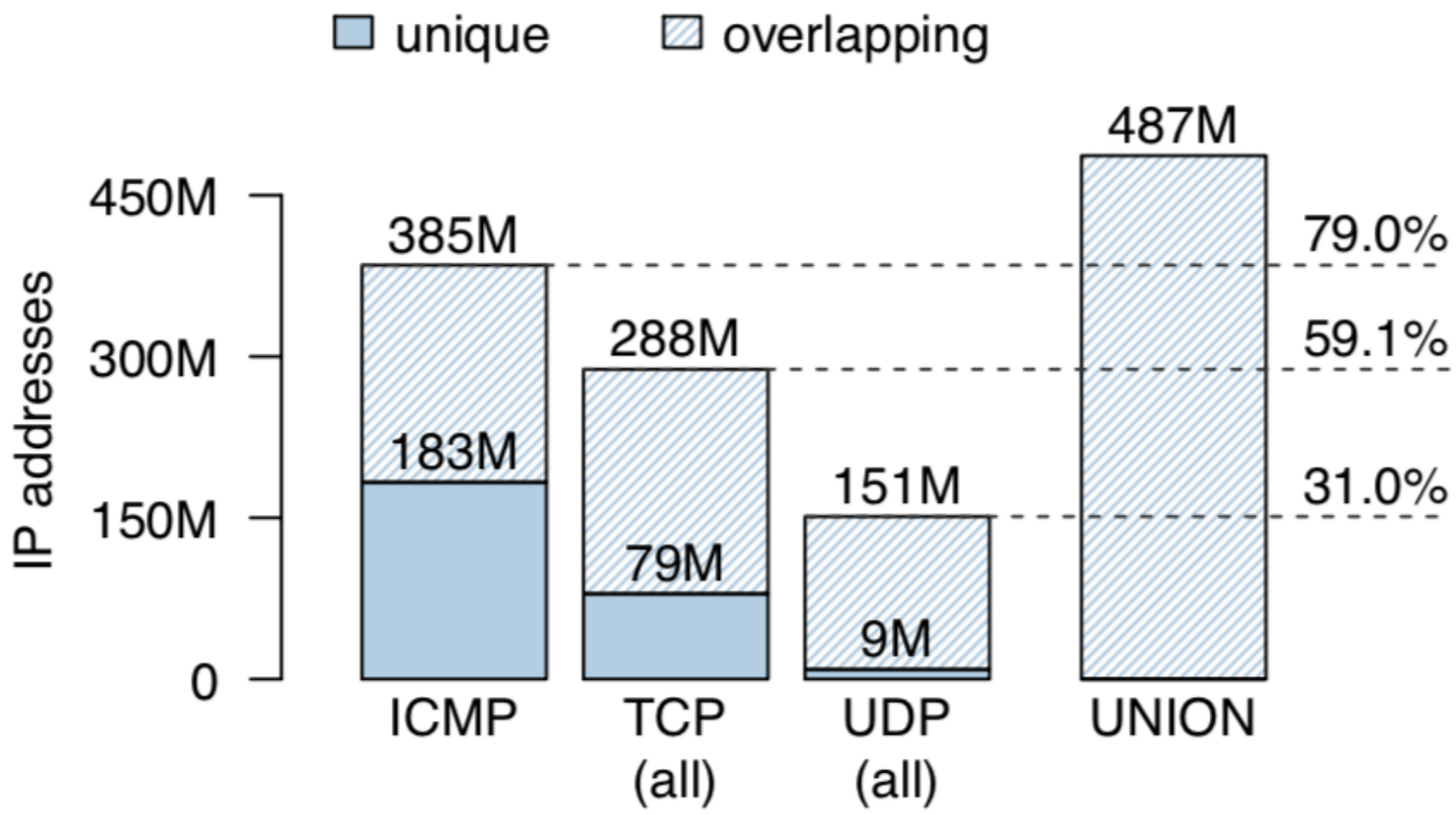
Roadmap


.1.
Taxonomy
.2.
Scan Methodology
.3.
Characterization of
IP Liveness

Characterizing IP Liveness

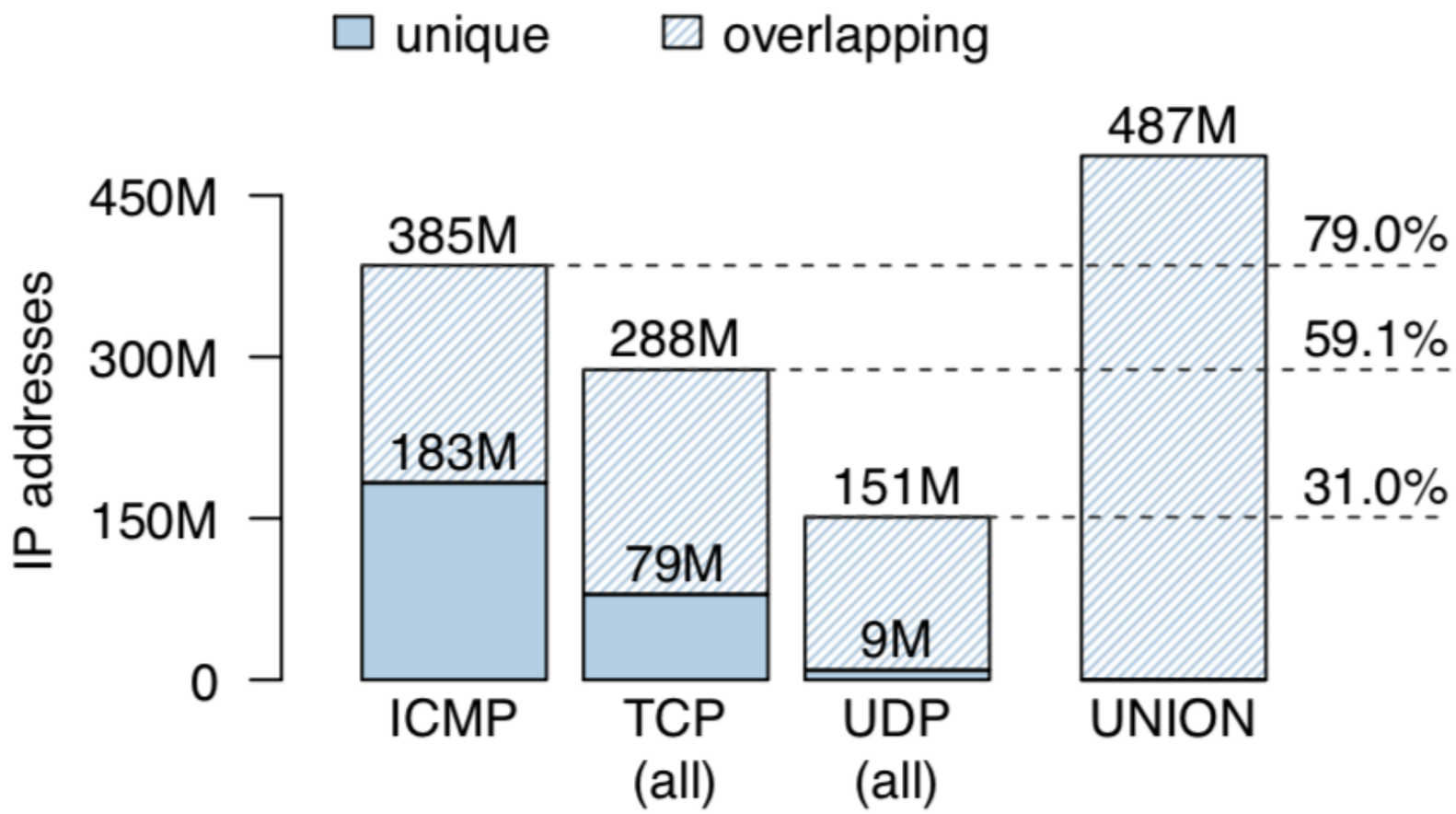
Network Layer

What is the coverage of different probe types?



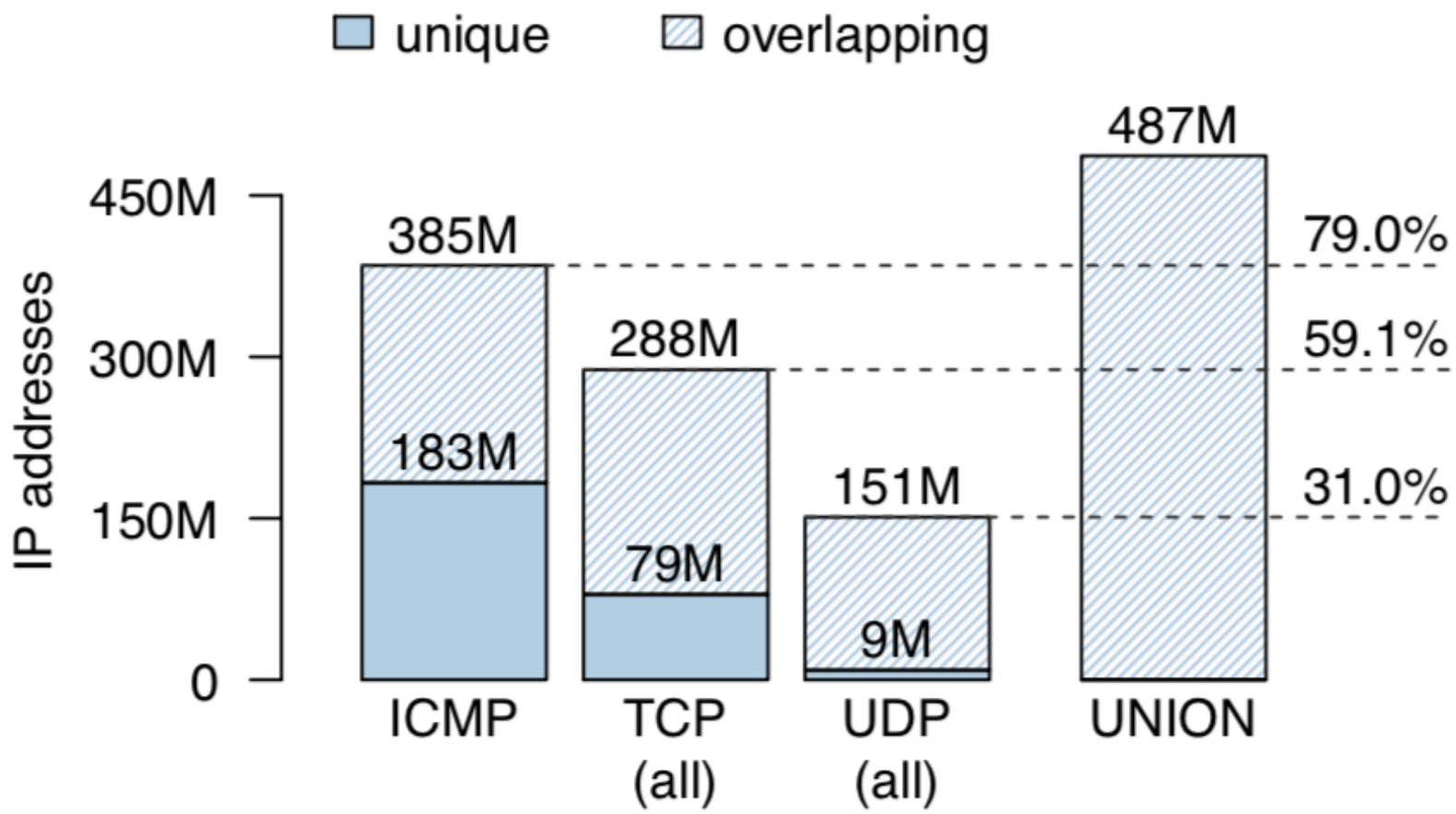
(a) Network layer alive IP addresses.

What is the coverage of different probe types?



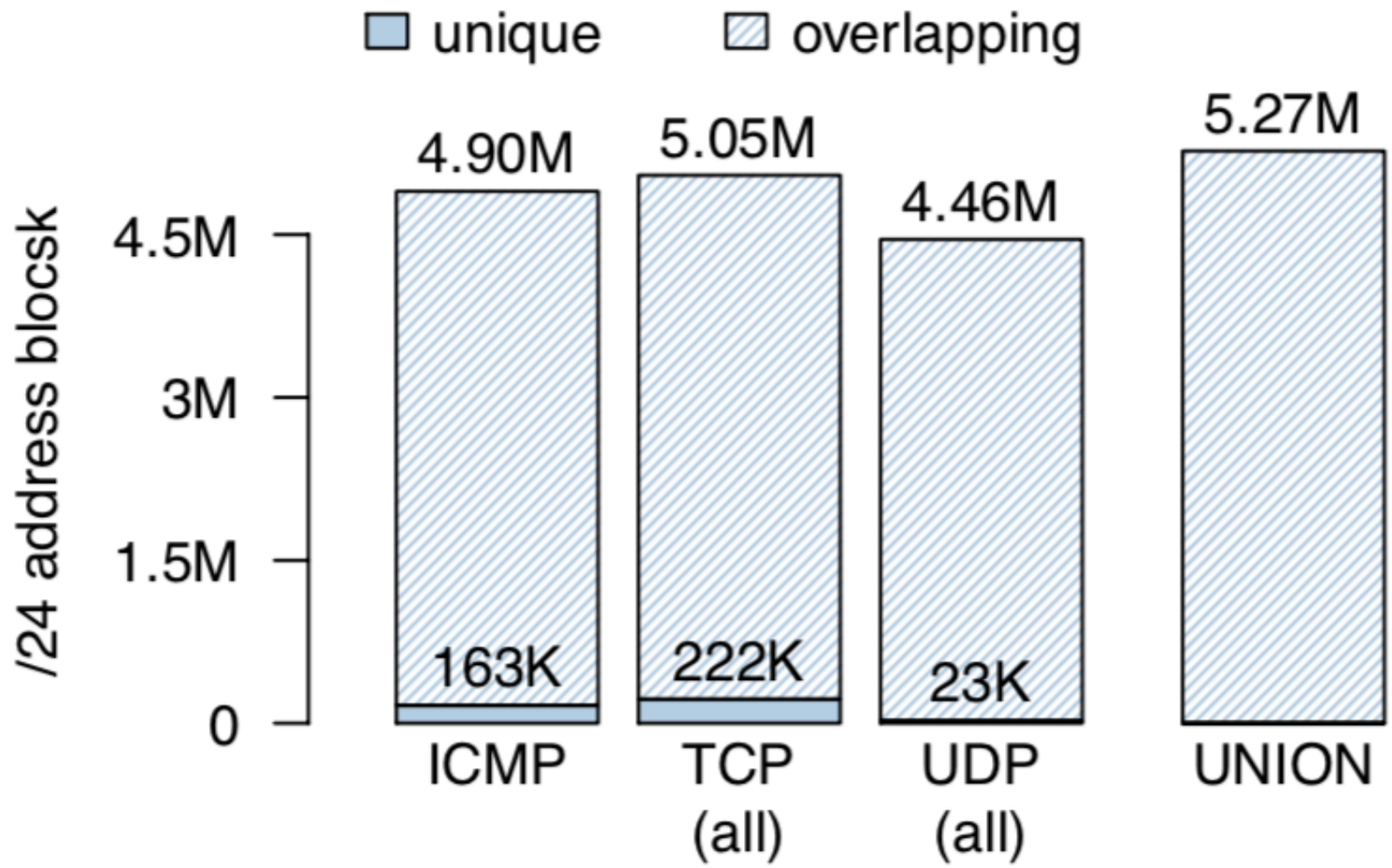
ICMP Echo probes are most effective in discovering network active IPs (79% of IP_all), followed by TCP probes

What is the coverage of different probe types?



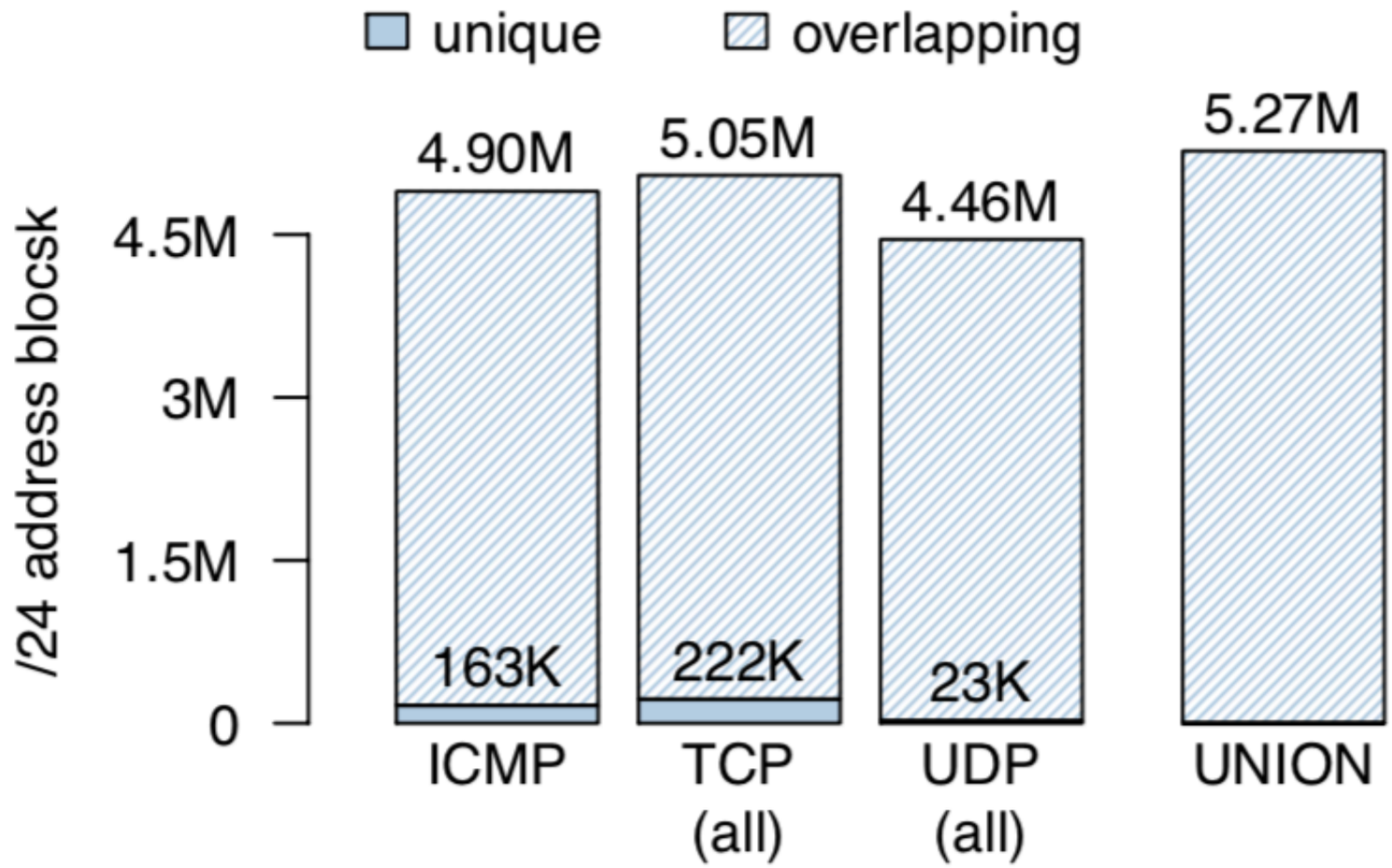
16% of IP_all can only exclusively be discovered via TCP,
2% can only be discovered via UDP probes

What is the coverage of different probe types?



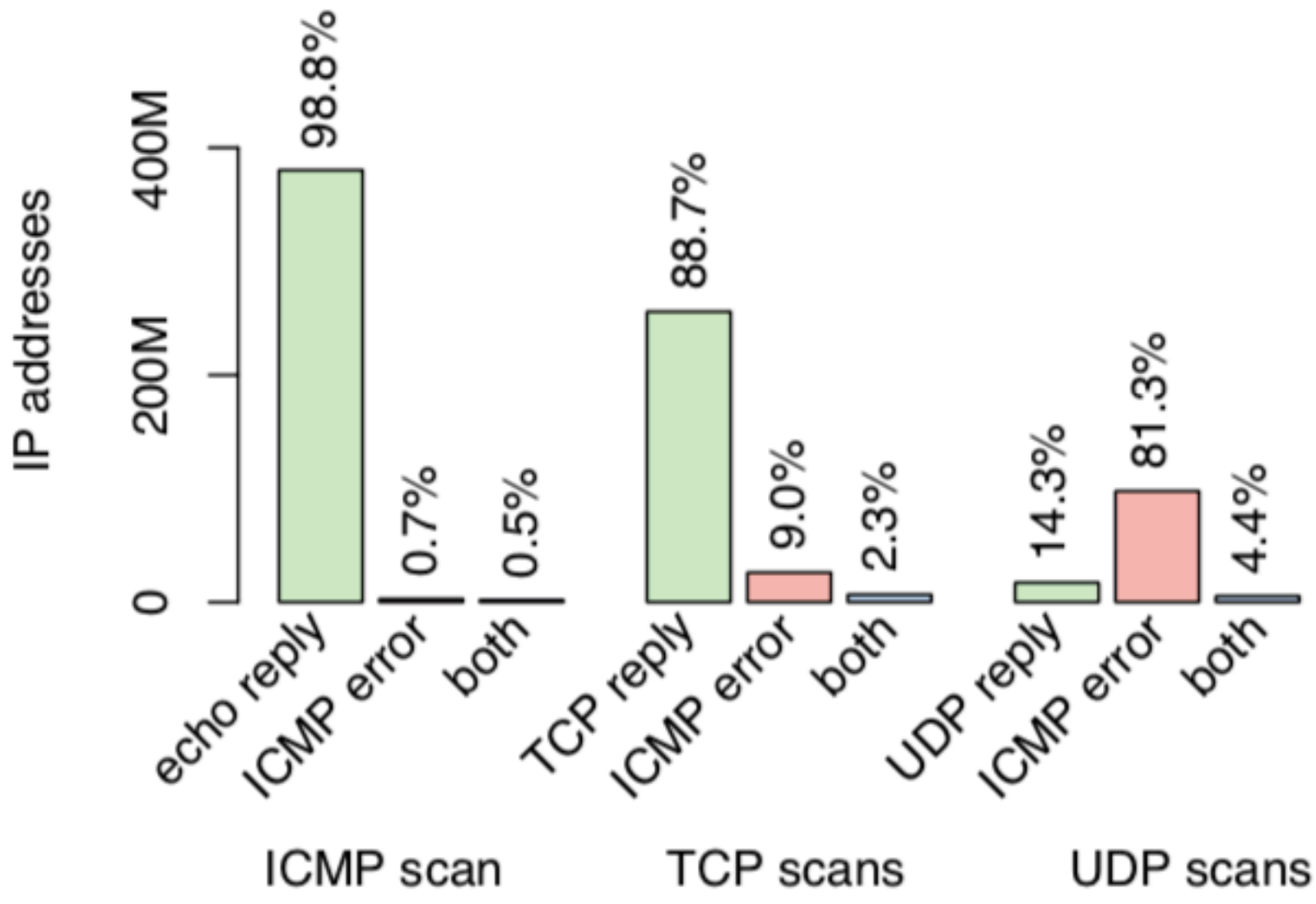
(b) Network layer alive /24 blocks.

What is the coverage of different probe types?



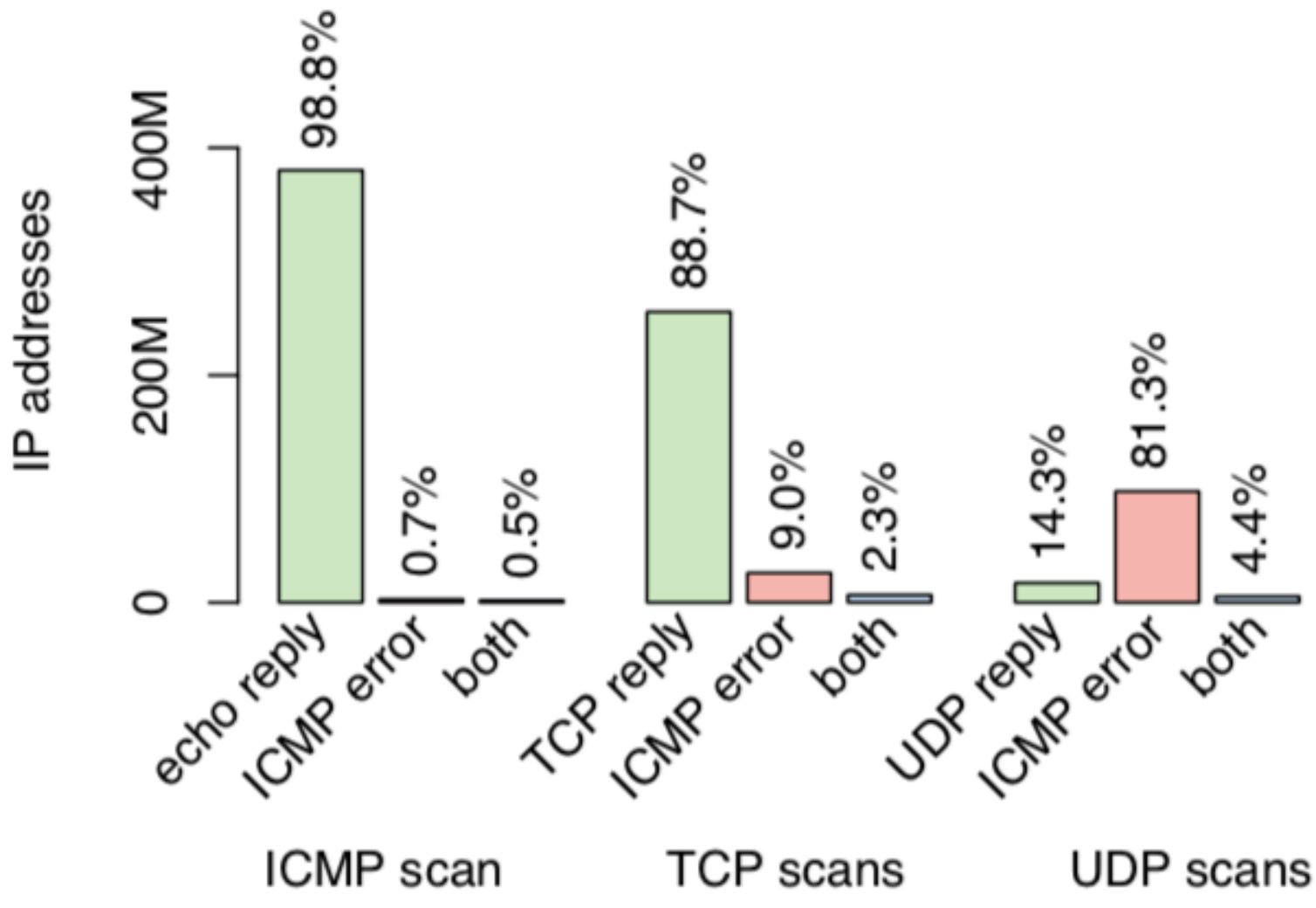
TCP scans show the highest coverage, discovering some 5M active /24 blocks, slightly more (≈ 3%) than ICMP Echo

What is the coverage of different probe responses?



(a) Breakdown of responses to scan types.

What is the coverage of different probe responses?

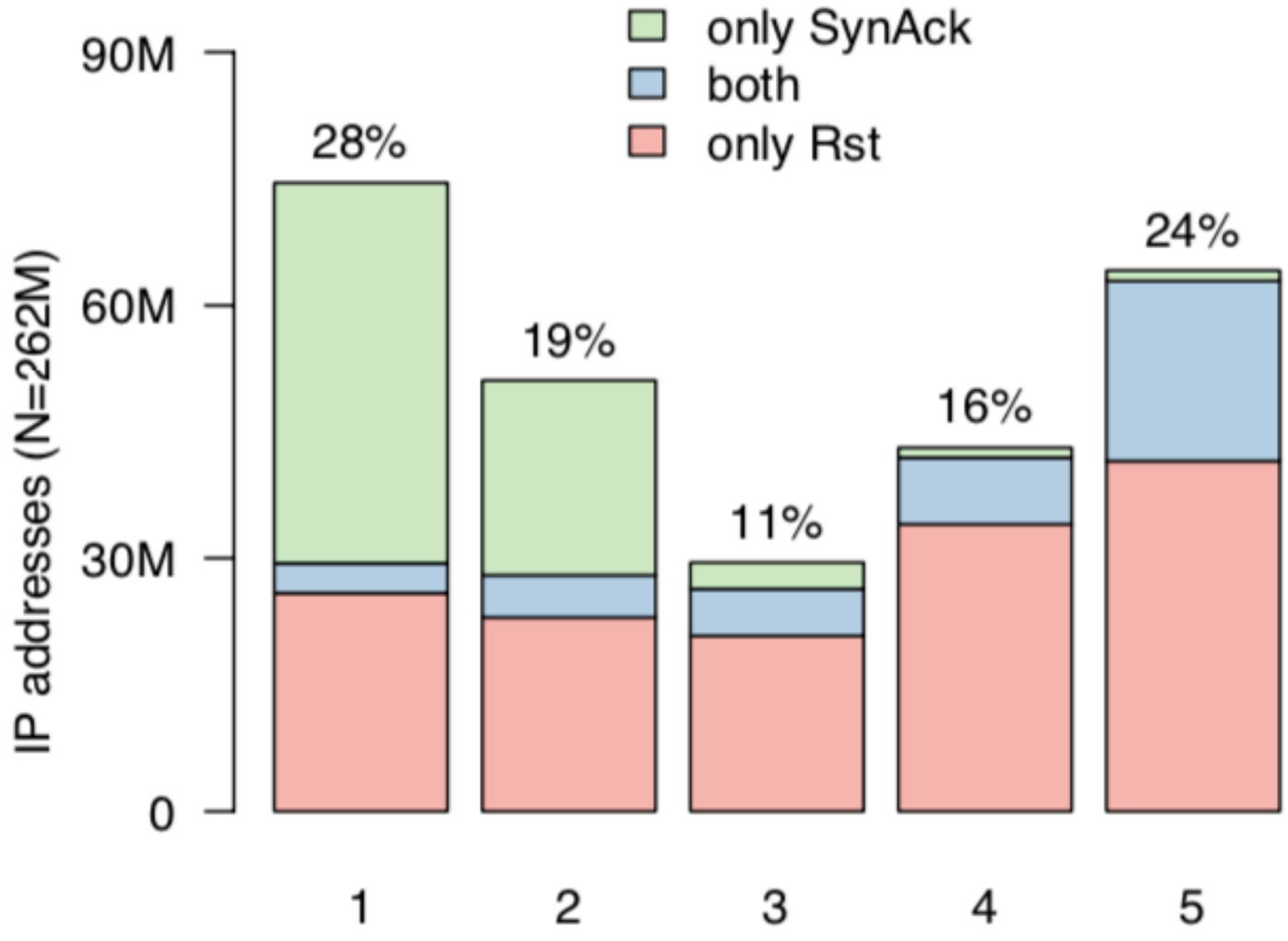


2.3% of IP_all are discoverable only through ICMP Error responses

Characterizing IP Liveness

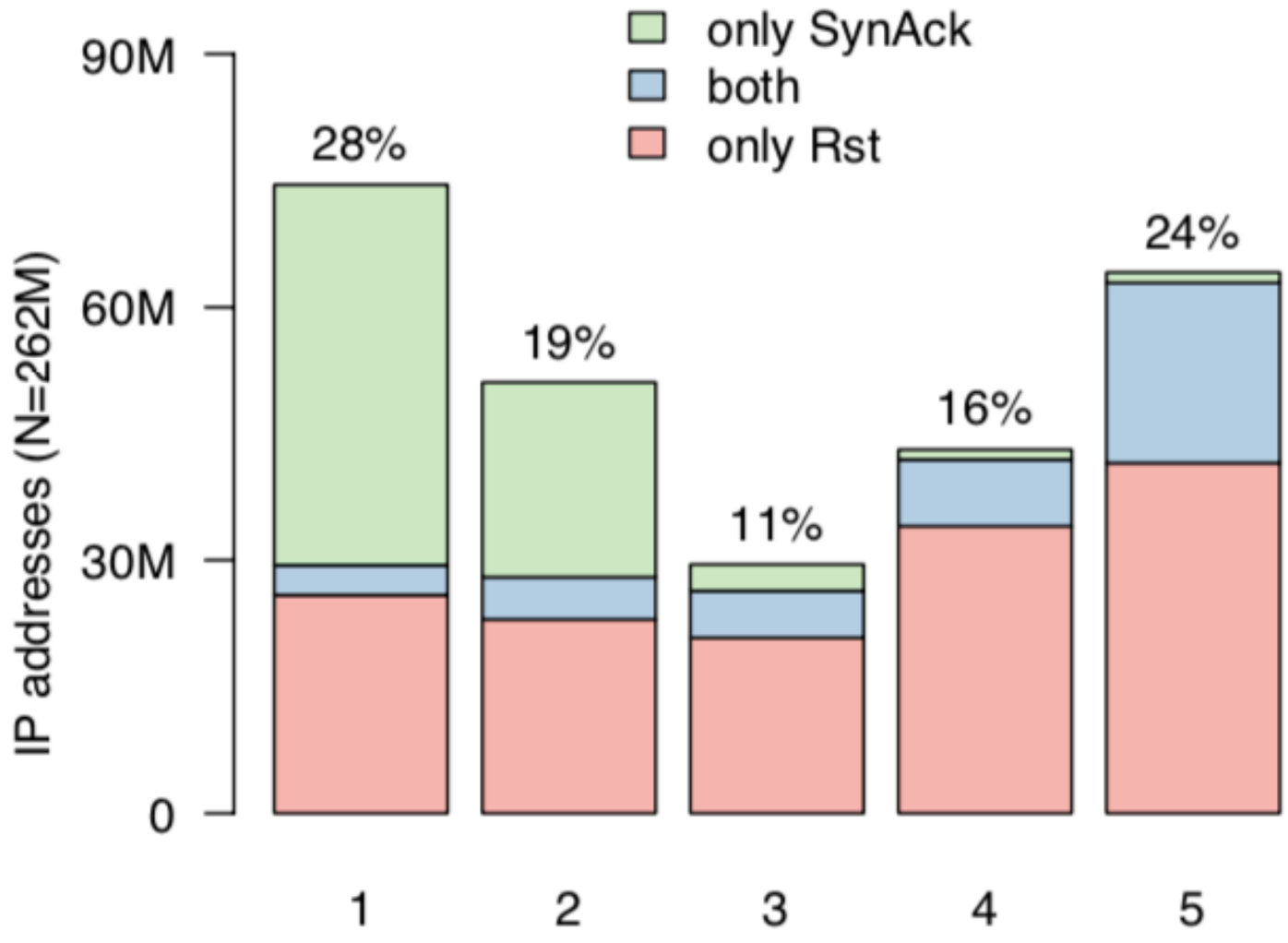
Transport Layer

How does the probed port affect the responsive population?



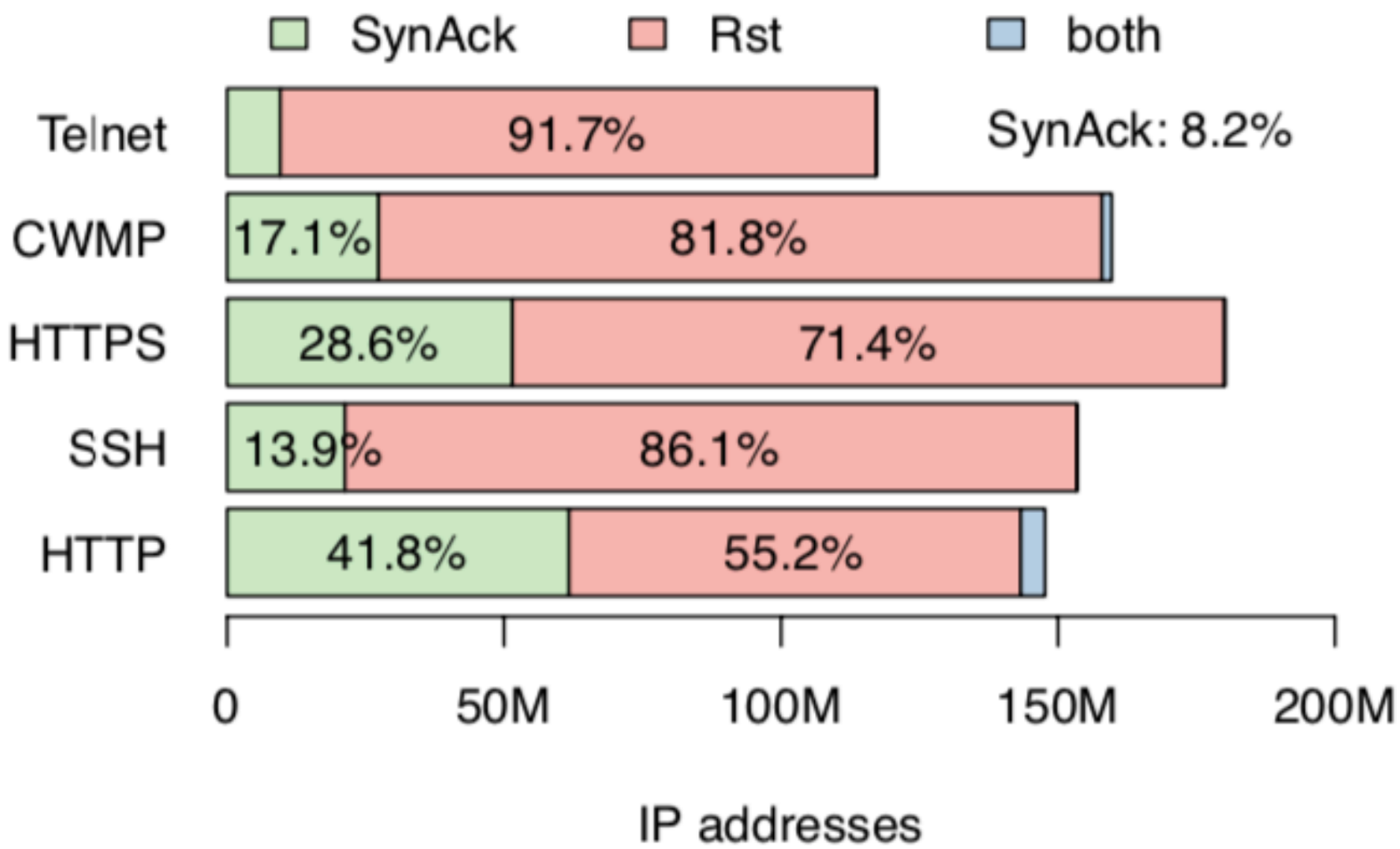
(a) TCP stack completeness/consistency.

How does the probed port affect the responsive population?



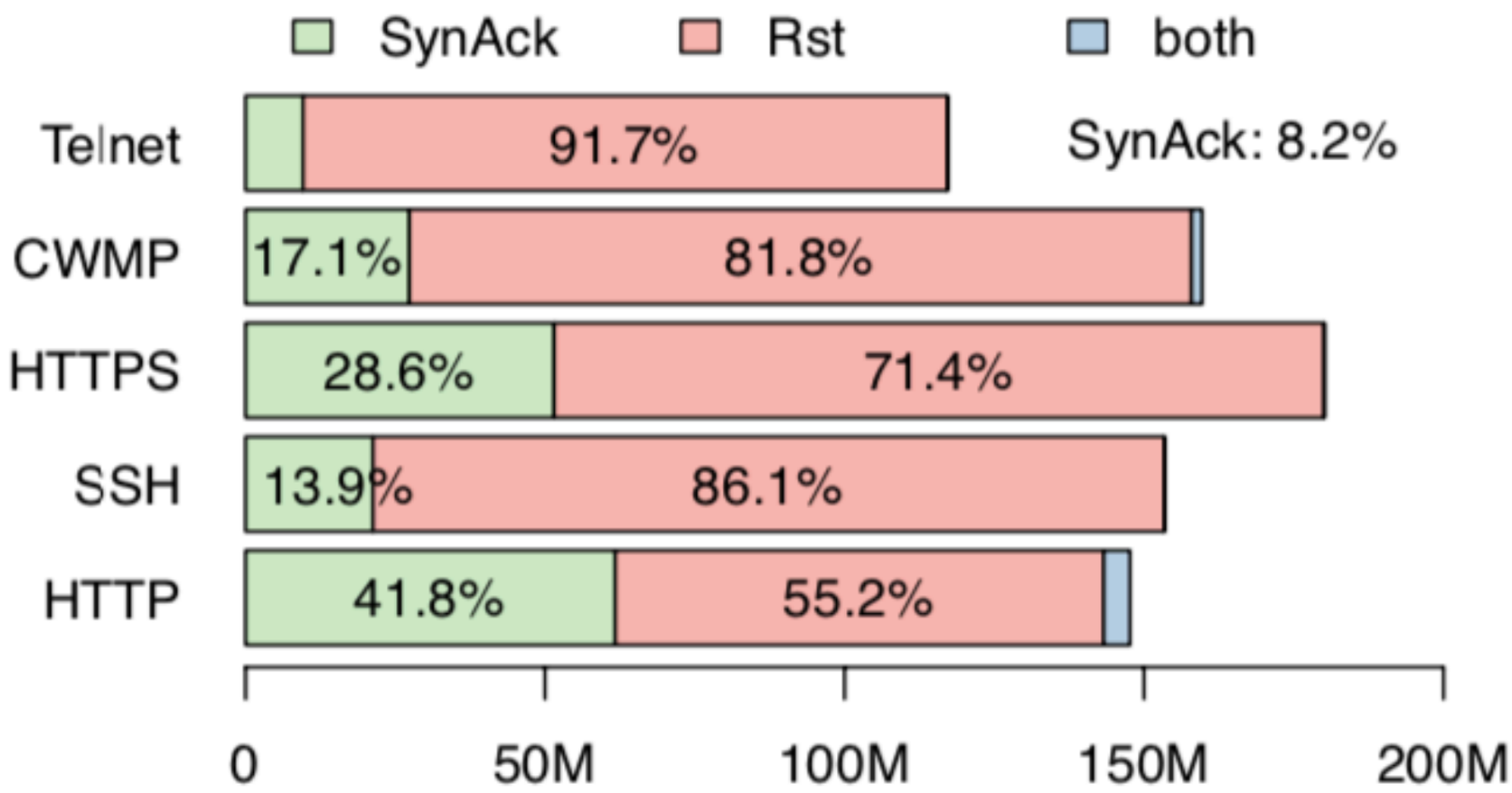
Only 24% of active hosts respond to probe packets on all five ports (potentially due to firewalling and/or filtering)

What is the coverage by probe response type?



(b) Breakdown of transport layer responses.

What is the coverage by probe response type?

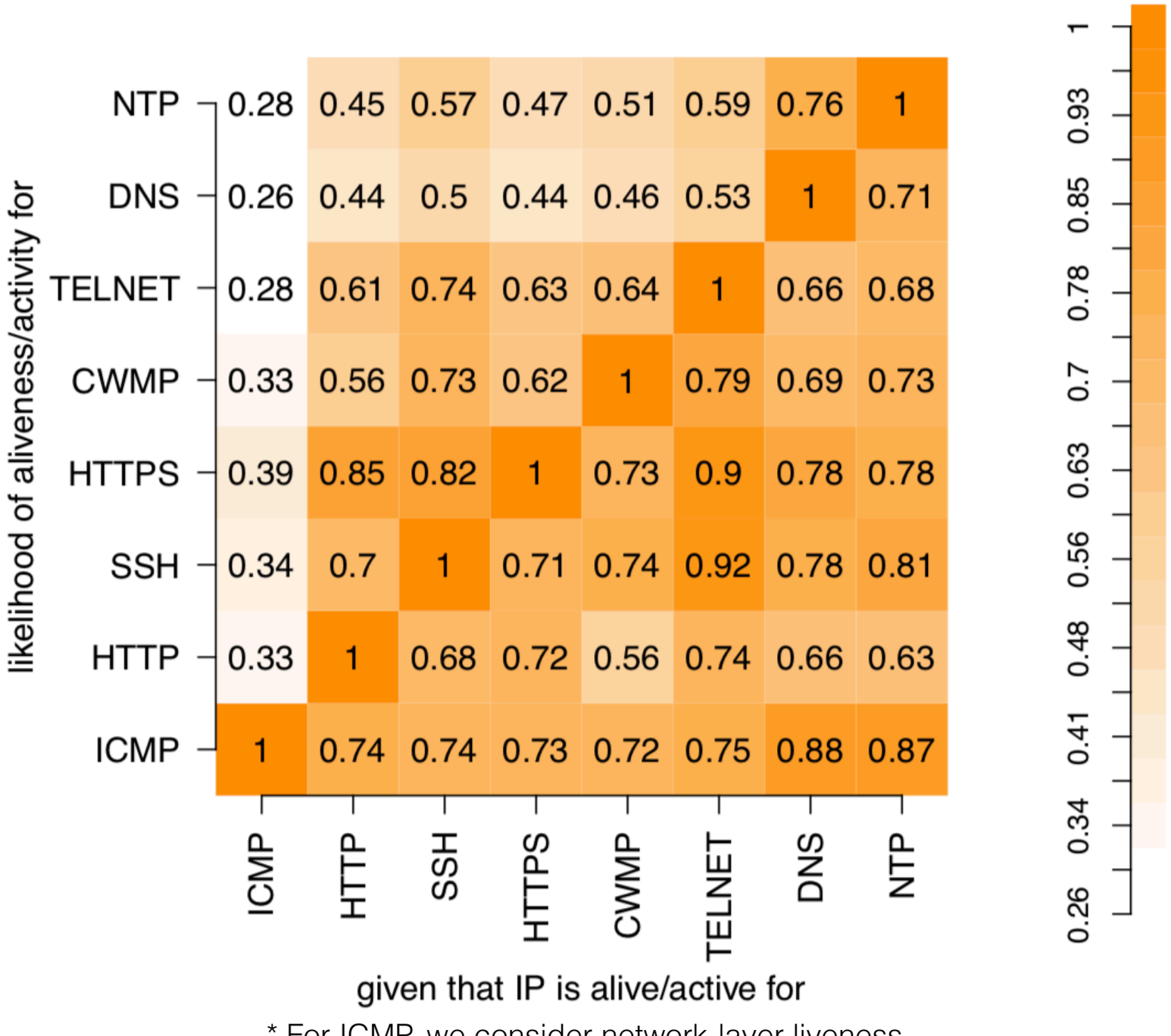


11.5% of all TCP activity can exclusively be found via CWMP. SSH, HTTP, and HTTPS provide unique coverage of 3–6% of active IPs.

Characterizing IP Liveness

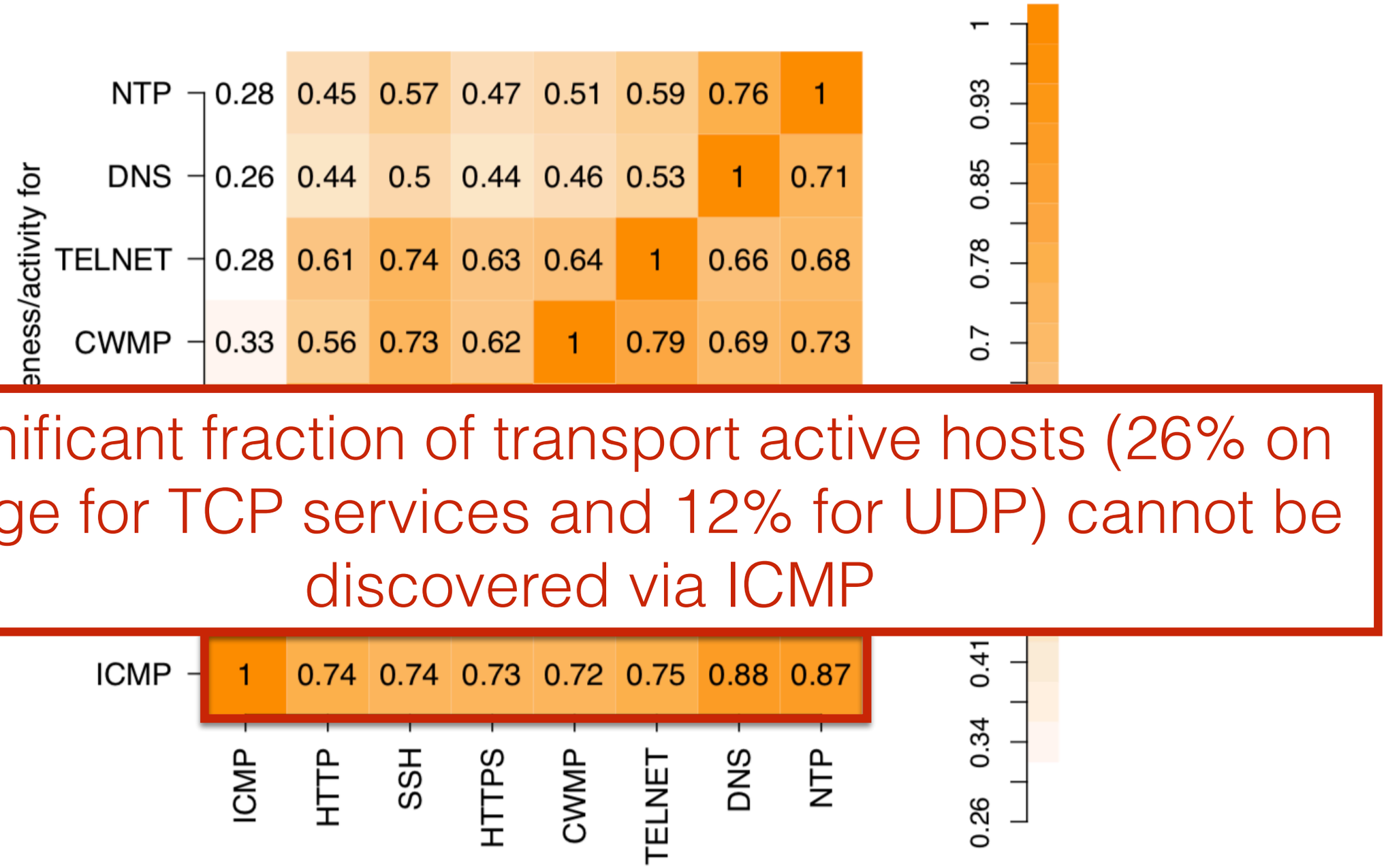
Cross-protocol

What is conditional activity per probe type?



* For ICMP, we consider network-layer liveness

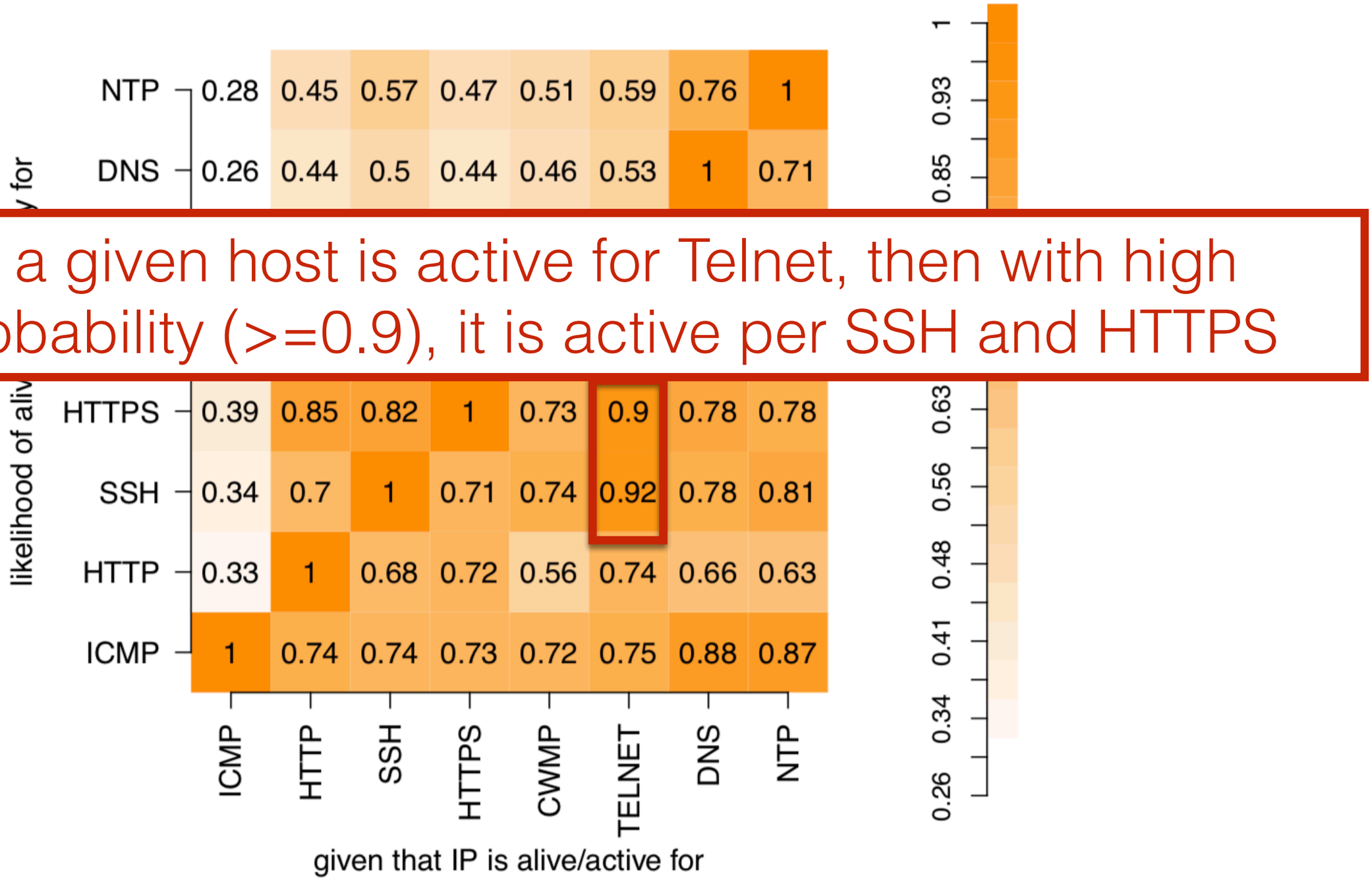
What is conditional activity per probe type?



A significant fraction of transport active hosts (26% on average for TCP services and 12% for UDP) cannot be discovered via ICMP

* For ICMP, we consider network-layer liveness

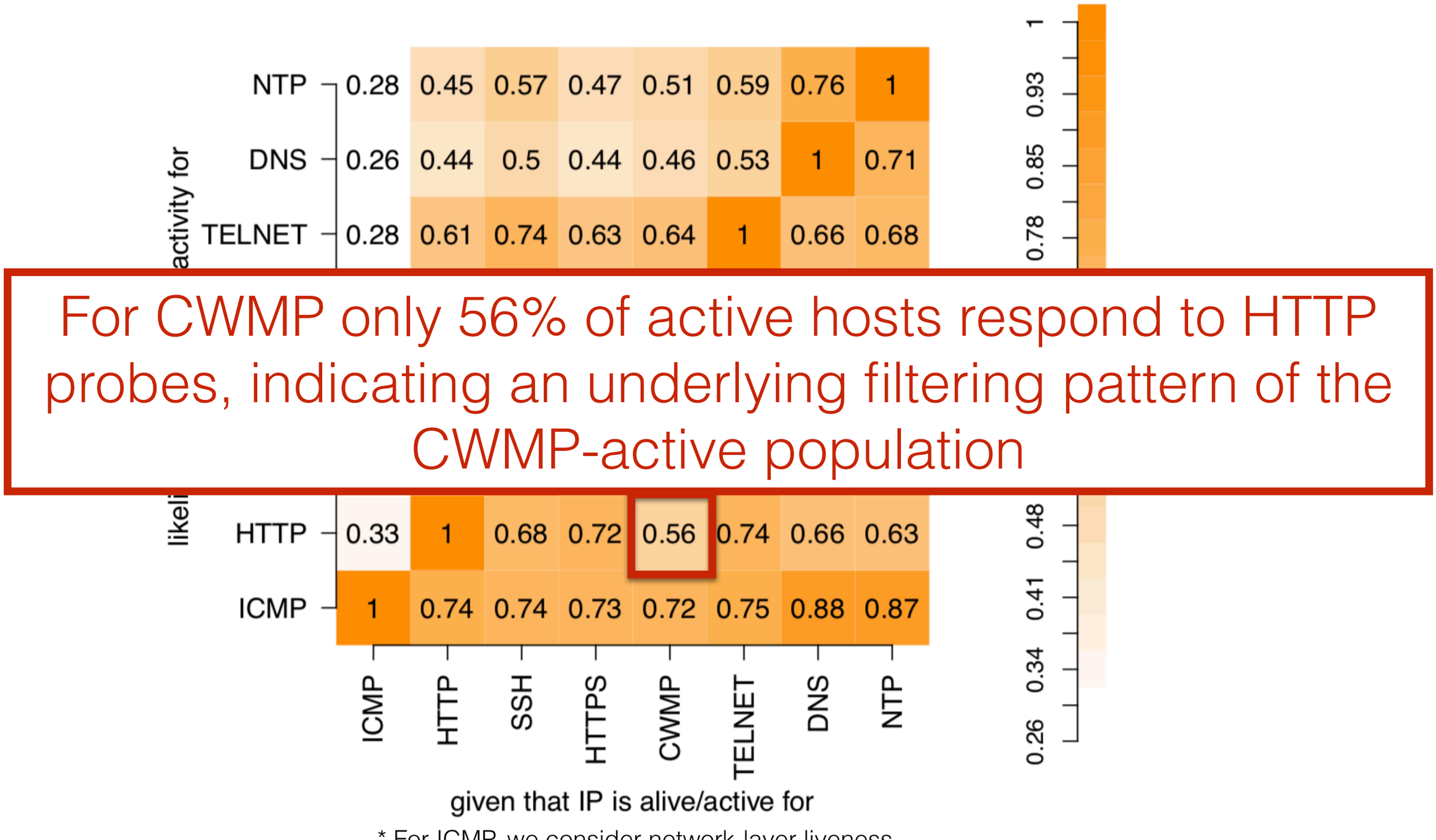
What is conditional activity per probe type?



If a given host is active for Telnet, then with high probability (>=0.9), it is active per SSH and HTTPS

* For ICMP, we consider network-layer liveness

What is conditional activity per probe type?



For CWMP only 56% of active hosts respond to HTTP probes, indicating an underlying filtering pattern of the CWMP-active population

* For ICMP, we consider network-layer liveness

Conclusion

- Comprehensive and least noisy picture of the state of Internet liveness

Conclusion

- Comprehensive and least noisy picture of the state of Internet liveness
- The taxonomy can serve as a basis for designing and executing future measurement studies, when it comes to decisions such as:
  - What type of probe packets should be employed?

Conclusion

- Comprehensive and least noisy picture of the state of Internet liveness
- The taxonomy can serve as a basis for designing and executing future measurement studies, when it comes to decisions such as:
  - What type of probe packets should be employed?
  - What type of responses should be captured?

Conclusion

- Comprehensive and least noisy picture of the state of Internet liveness
- The taxonomy can serve as a basis for designing and executing future measurement studies, when it comes to decisions such as:
  - What type of probe packets should be employed?
  - What type of responses should be captured?
  - How to interpret responses?

Conclusion

- Comprehensive and least noisy picture of the state of Internet liveness
- The taxonomy can serve as a basis for designing and executing future measurement studies, when it comes to decisions such as:
  - What type of probe packets should be employed?
  - What type of responses should be captured?
  - How to interpret responses?
  - Whether it is appropriate to use the output of one scan as input for subsequent measurements?

Thanks
Q & A

bano@fb.com
@thatBano

https://sheharbano.com

Paper pdf:
ACM SIGCOMM CCR 2018