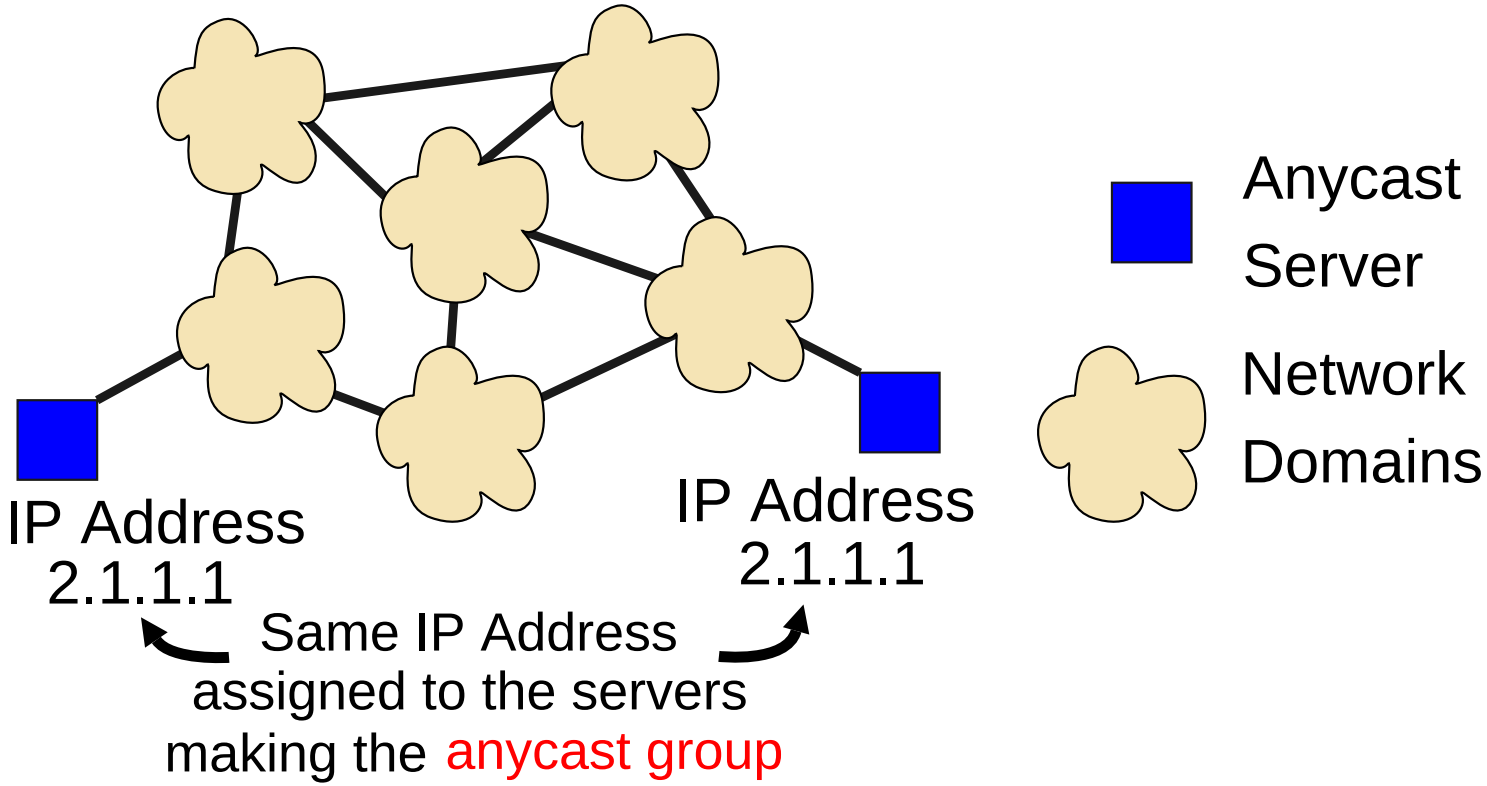
A Measurement-based Deployment Proposal for IP Anycast

Hitesh Ballani (Cornell University)
Paul Francis (Cornell University)
Sylvia Ratnasamy (Intel-Research)

IMC 2006

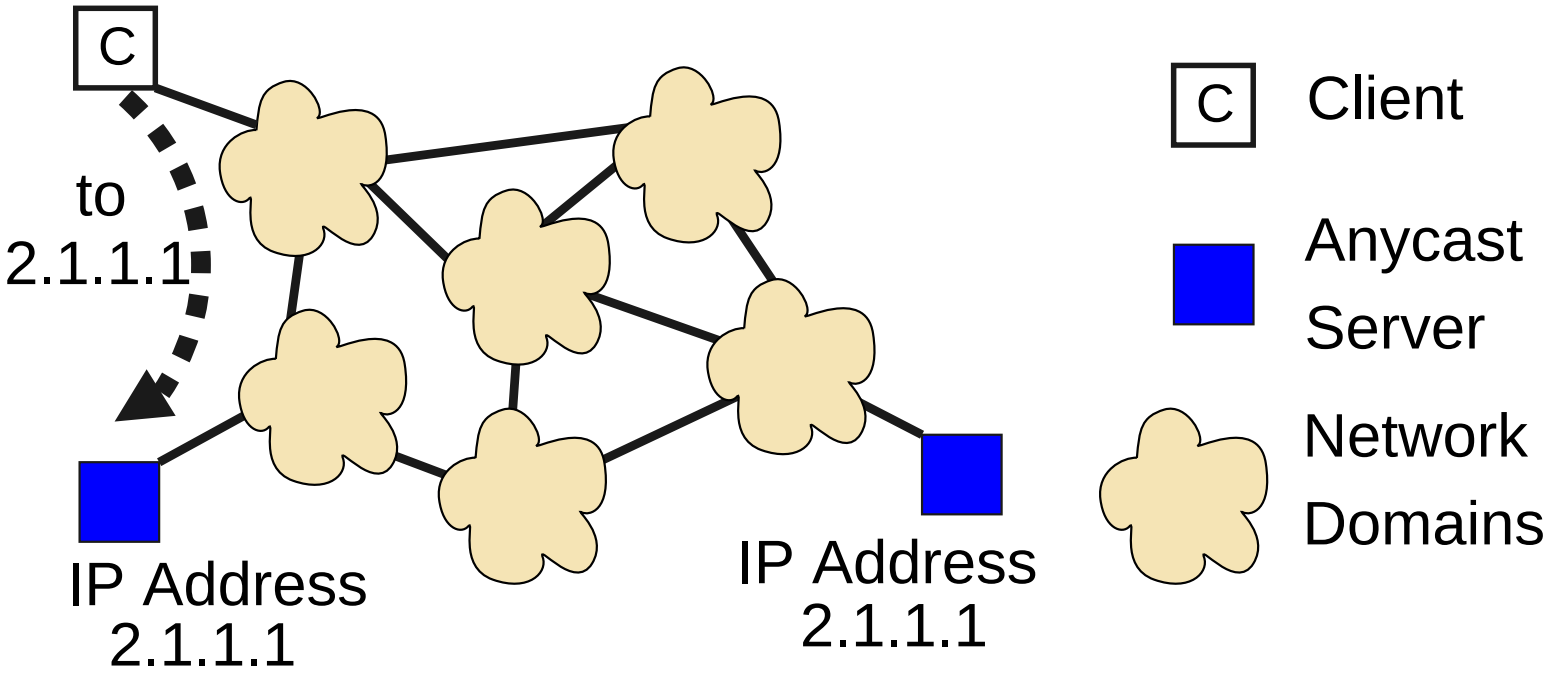
What is IP Anycast?



One-to-Any communication

with no changes to Internet routing and clients

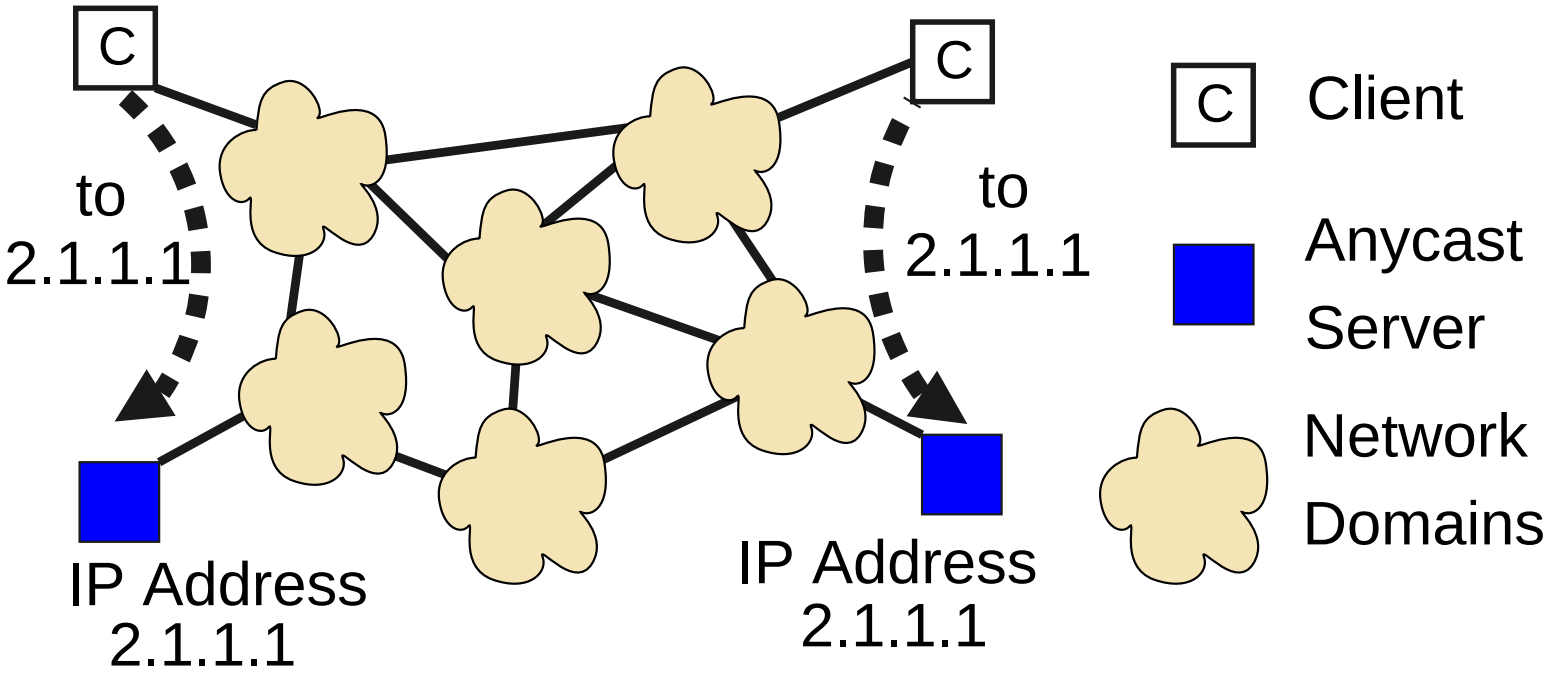
What is IP Anycast?



One-to-Any communication

with no changes to Internet routing and clients

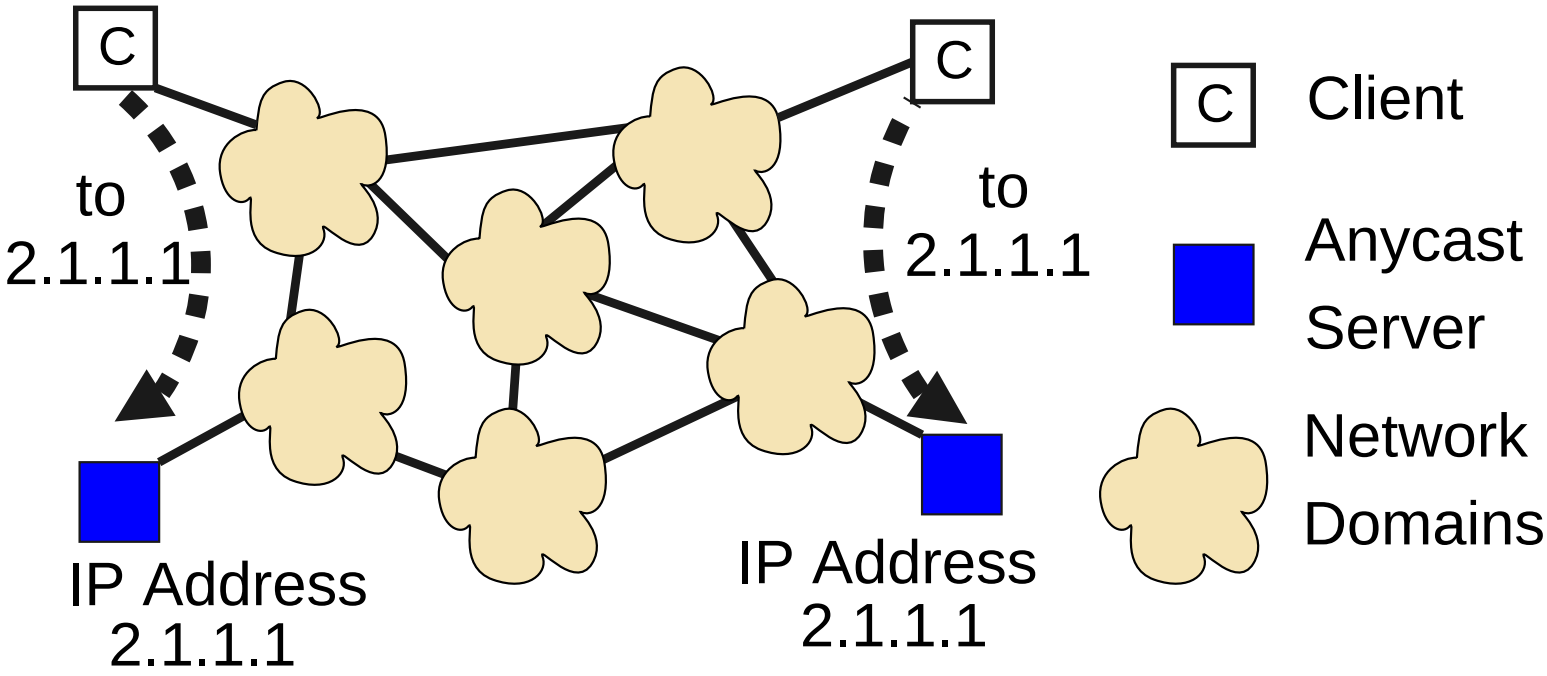
What is IP Anycast?



One-to-Any communication

with no changes to Internet routing and clients

What is IP Anycast?



IP Anycast as a Service Discovery Primitive

- Distributes client load across servers
- Reduces access latency for clients
- Offers network-level resilience to DDoS attacks

IP Anycast Usage

- Anycasting of six of the thirteen root-servers (C-Root, F-Root, I-Root, J-Root, K-Root, M-Root)
- IPv4-to-IPv6 transition [RFC 3068]
- Rendezvous discovery for IP multicast [RFC 3446]
- Other usage scenarios
  - AS112 Project [http://as112.net]
  - Commercial CDNs [http://cachefly.net]
  - DDos sinkholes [Greene et. al., NANOG'03]

IP Anycast Usage

- Anycasting of six of the thirteen root-servers (C-Root, F-Root, I-Root, J-Root, K-Root, M-Root)
- IPv4-to-IPv6 transition [RFC 3068]
- Rendezvous discovery for IP multicast [RFC 3446]
- Other usage scenarios
  - AS112 Project [http://as112.net]
  - Commercial CDNs [http://cachefly.net]
  - DDos sinkholes [Greene et. al., NANOG'03]

IP Anycast Usage

- Anycasting of six of the thirteen root-servers (C-Root, F-Root, I-Root, J-Root, K-Root, M-Root)
- IPv4-to-IPv6 transition [RFC 3068]
- Rendezvous discovery for IP multicast [RFC 3446]
- Other usage scenarios
  - AS112 Project [http://as112.net]
  - Commercial CDNs [http://cachefly.net]
  - DDos sinkholes [Greene et. al., NANOG'03]

IP Anycast Usage

- Anycasting of six of the thirteen root-servers (C-Root, F-Root, I-Root, J-Root, K-Root, M-Root)
- IPv4-to-IPv6 transition [RFC 3068]
- Rendezvous discovery for IP multicast [RFC 3446]
- Other usage scenarios
  - AS112 Project [http://as112.net]
  - Commercial CDNs [http://cachefly.net]
  - DDos sinkholes [Greene et. al., NANOG'03]

In spite of growing usage, IP Anycast and its interaction with IP Routing is not well understood!

IP Anycast is not well understood

Are clients routed to close-by anycast servers?

What is the impact of the failure of an anycast server?

Is the client load across the anycast server sites balanced?

Are subsequent packets from a client routed to the same anycast site?

IP Anycast is not well understood

Are clients routed to close-by anycast servers?

- **Proximity** offered by IP Anycast

What is the impact of the failure of an anycast server?

- **Failover** properties of IP Anycast

Is the client load across the anycast server sites balanced?

- **Load-distribution** across deployments

Are subsequent packets from a client routed to the same anycast site?

- **Affinity** offered by IP Anycast

A sneak peek at the study's results

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment				✓
Planned Deployment				✓

A sneak peek at the study's results

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗	✗	✗	✓
Planned Deployment	✓	✓	✗	✓

A sneak peek at the study's results

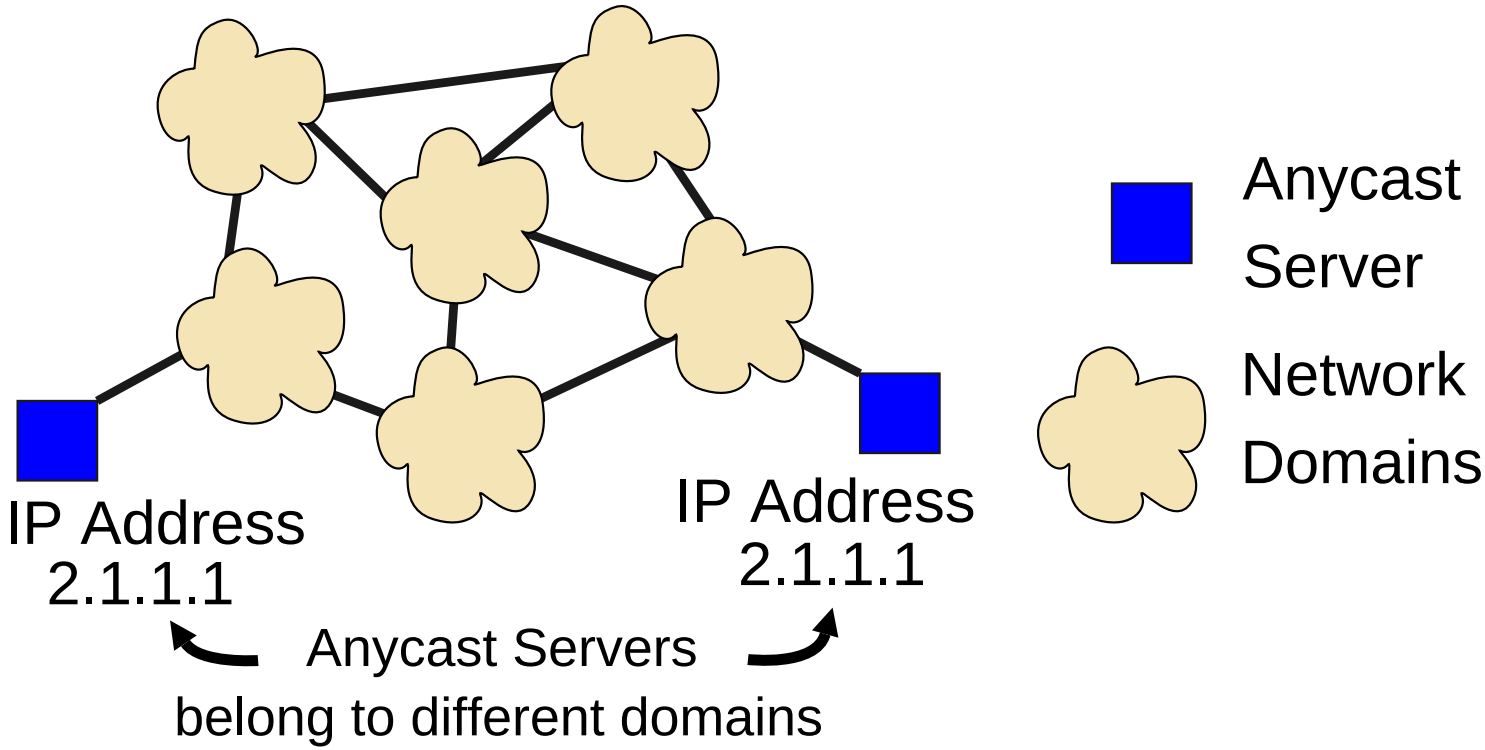
Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗	✗	✗	✓
Planned Deployment	✓	✓	✗	✓

Take Home Message

Talk Outline

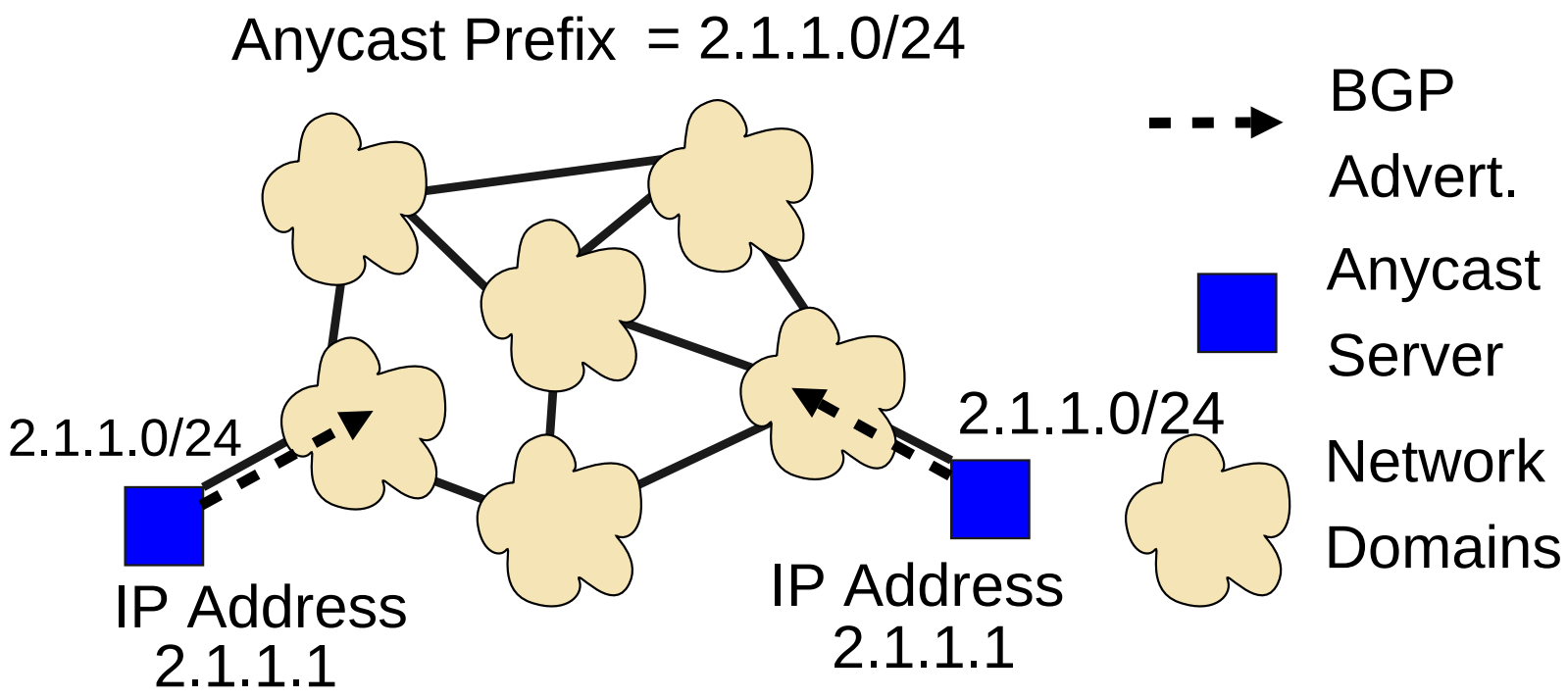
- Introduction
- Terminology
- Deployments Measured
- Methodology
- Measurements
  - Proximity
  - Failover
  - Load-distribution
  - Affinity
- Conclusions

Terminology



Study focusses on Inter-domain IP Anycast

Terminology



Anycast Servers advertise the **Anycast Prefix** into BGP through their **Upstream Provider**

Deployments Measured

External Deployments

F-Root	:	27 servers
J-Root	:	13 servers
AS112	:	20 servers

Deployments Measured

External Deployments

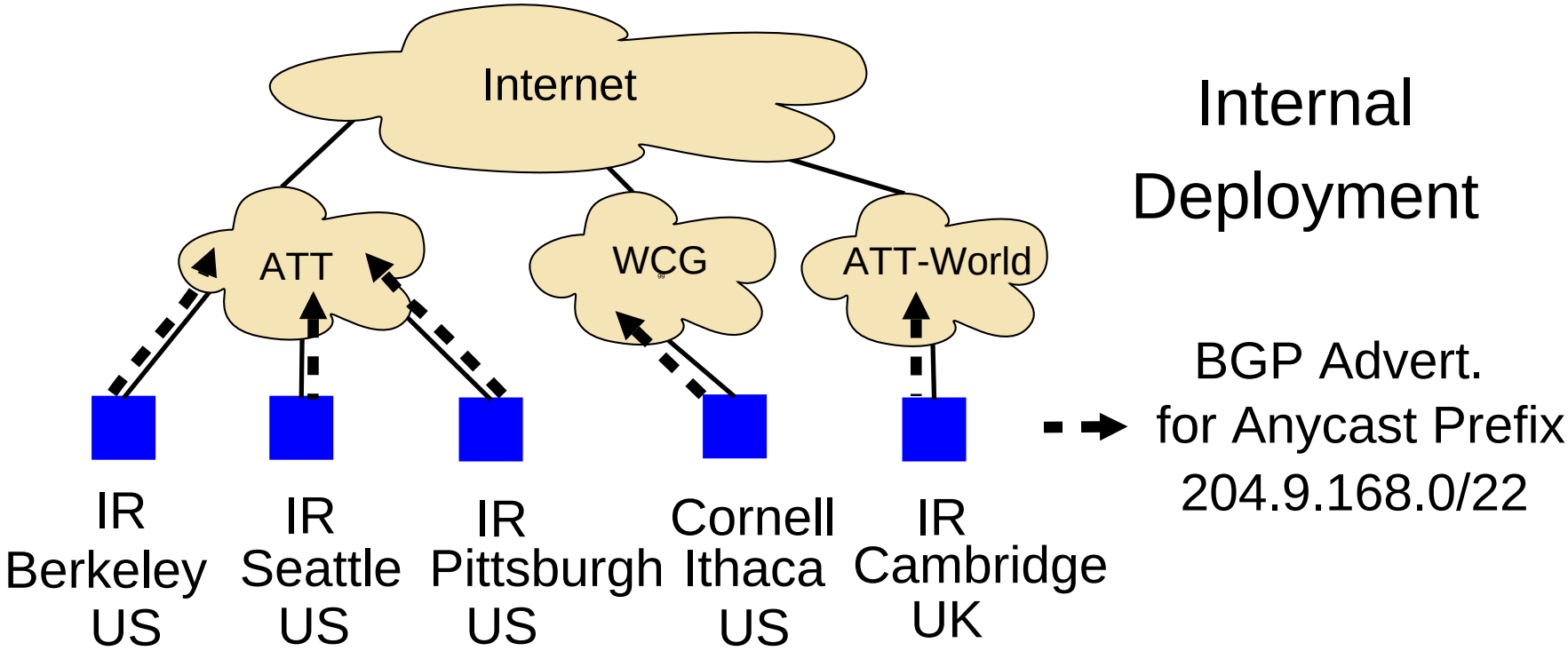
F-Root : 27 servers

J-Root : 13 servers

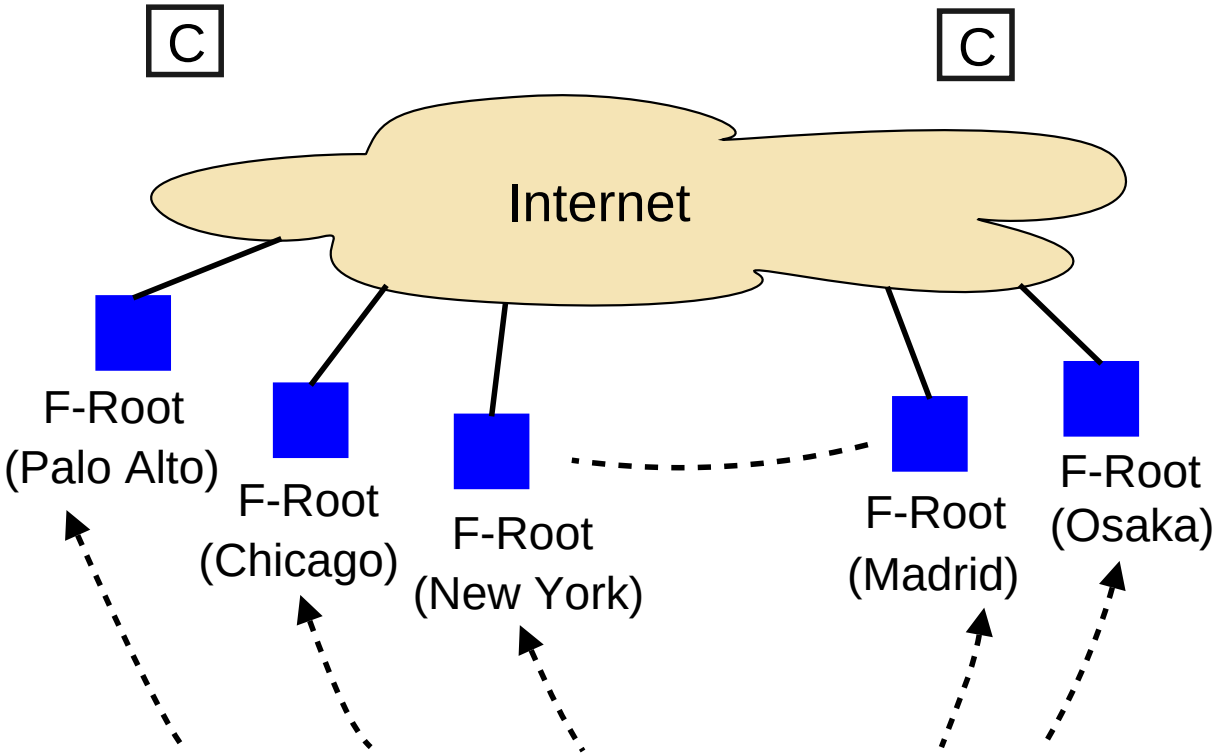
AS112 : 20 servers

Internal Deployment

Internal : 5 servers

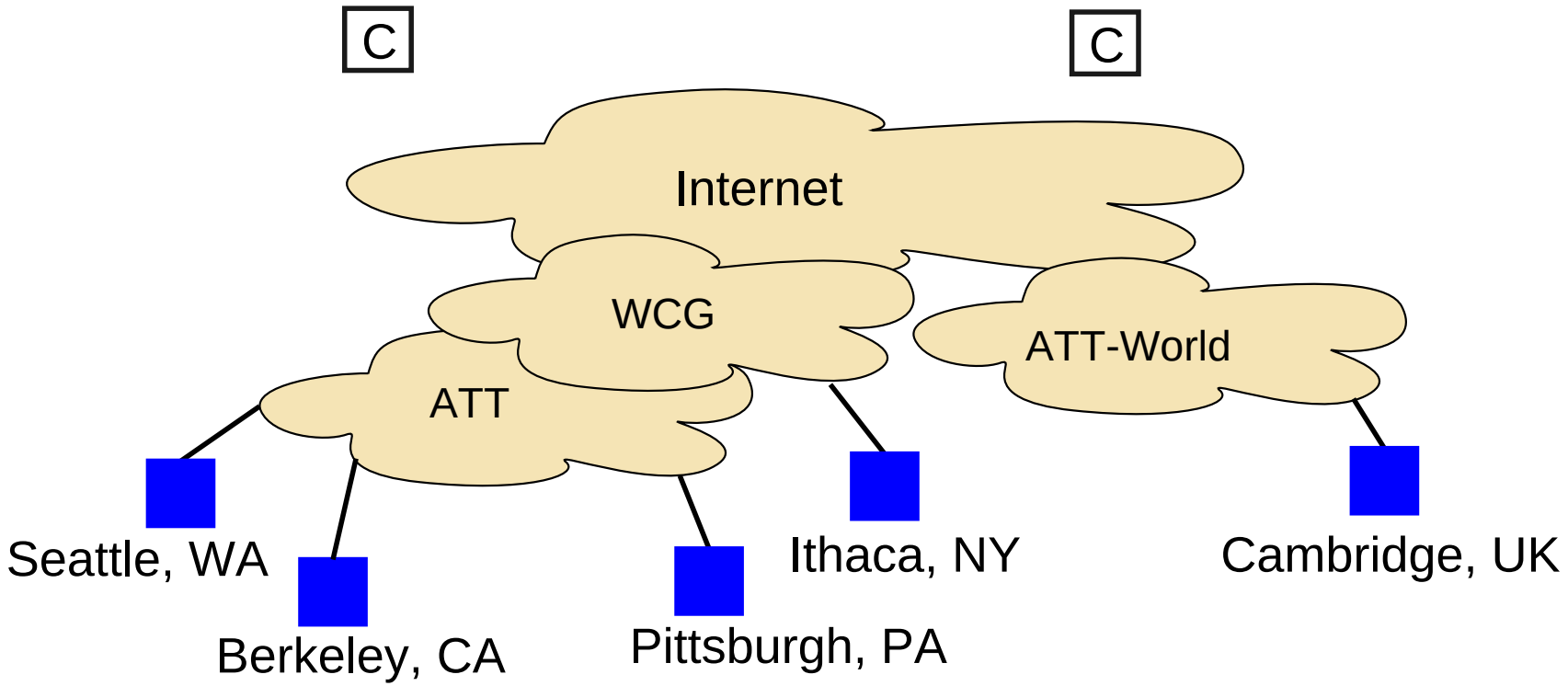


Probing Methodology



F-Root Deployment: Anycast Servers are DNS Servers

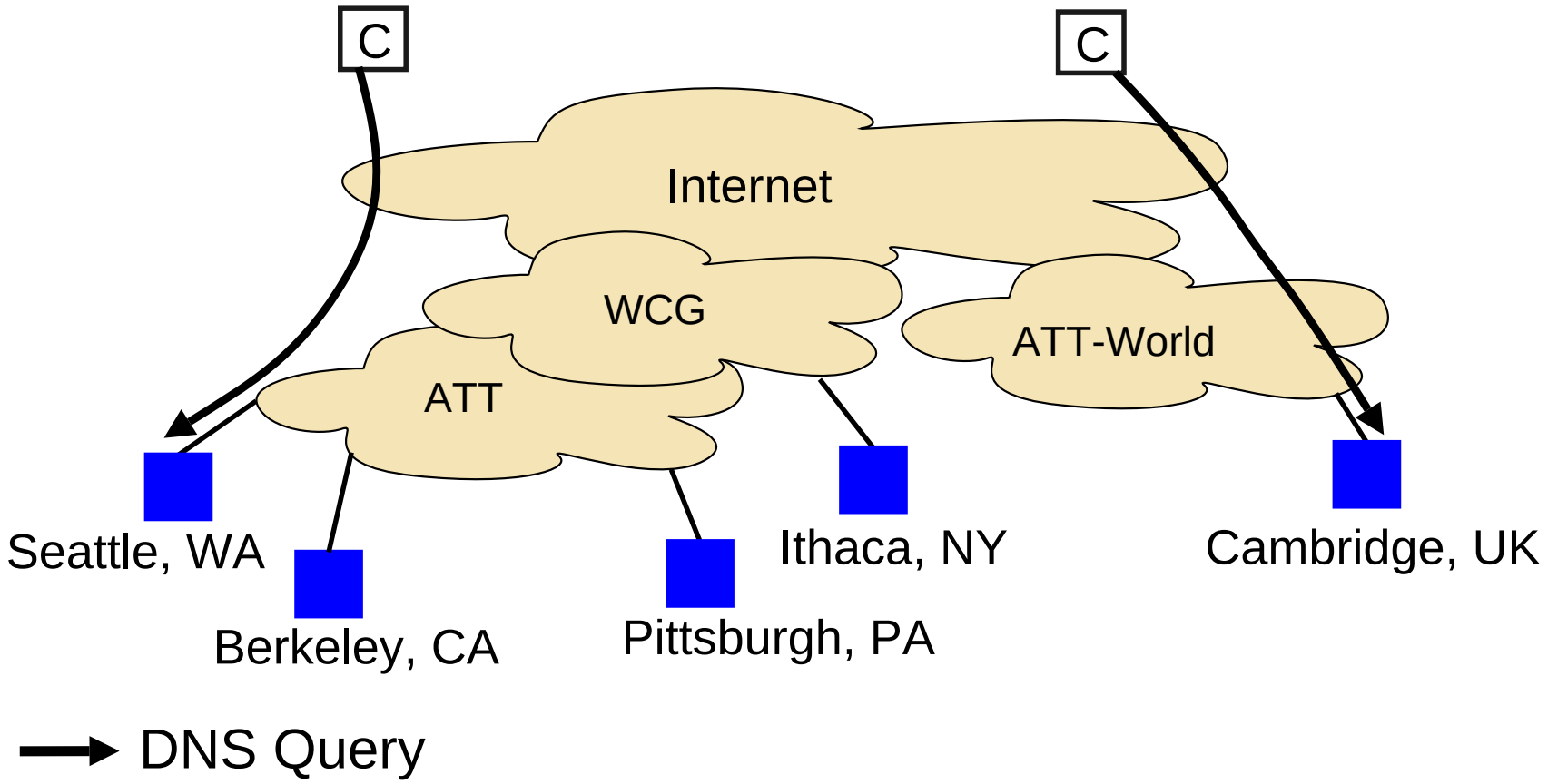
Probing Methodology



Internal Deployment: Anycast Servers can run DNS Servers

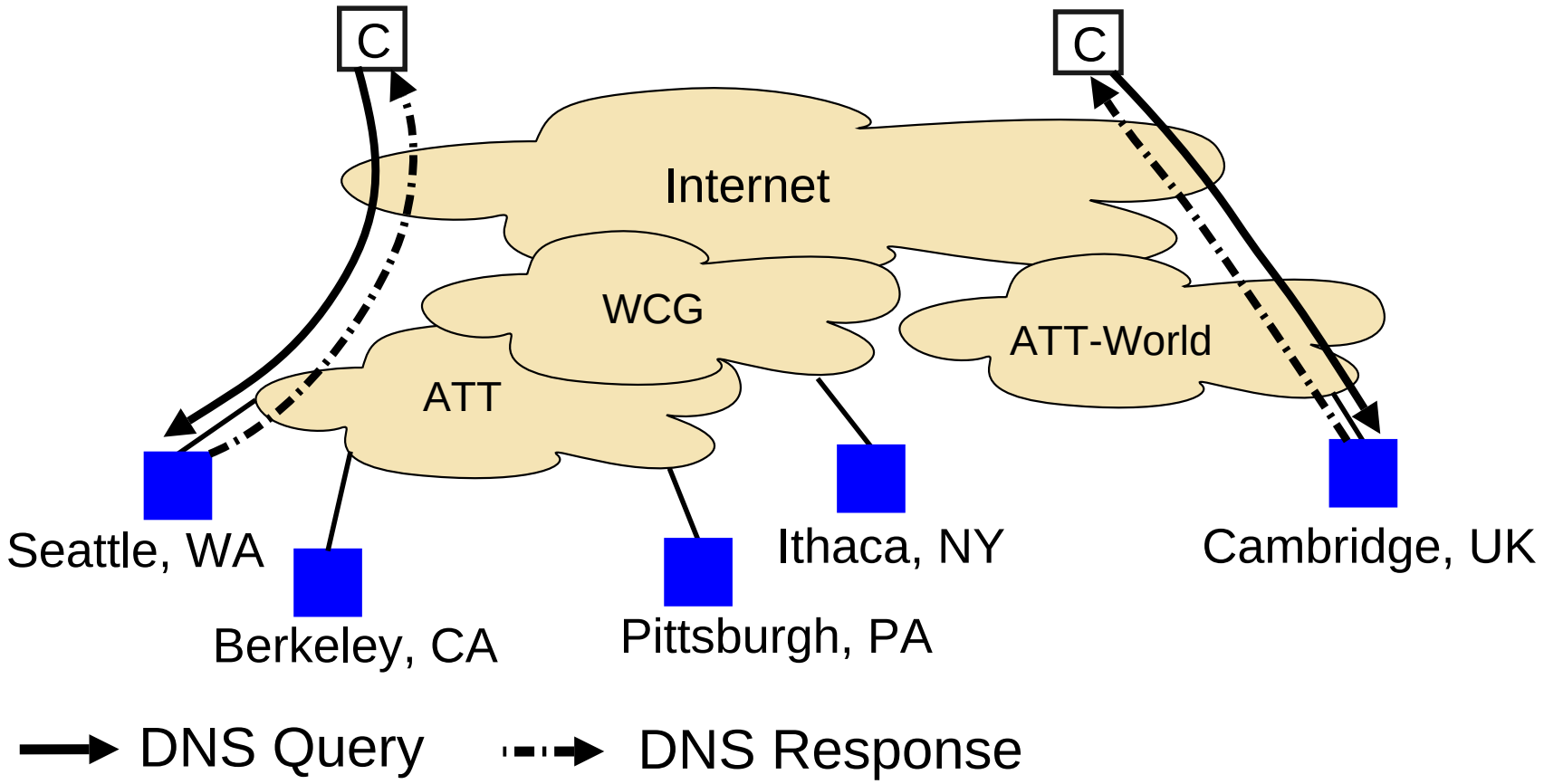
Probing Methodology

Anycast Servers can be probed
using DNS queries



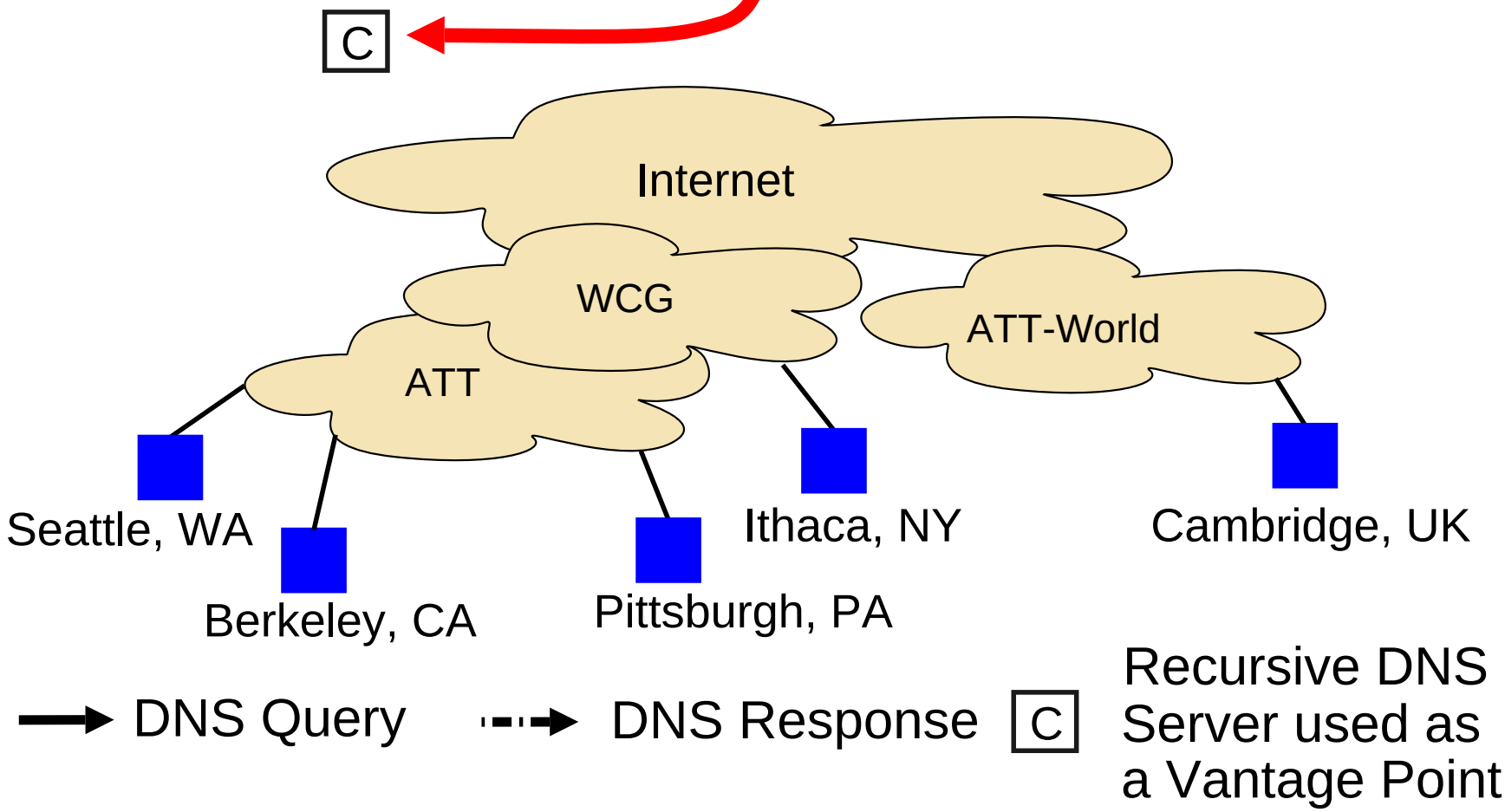
Probing Methodology

Anycast Servers can be probed
using DNS queries



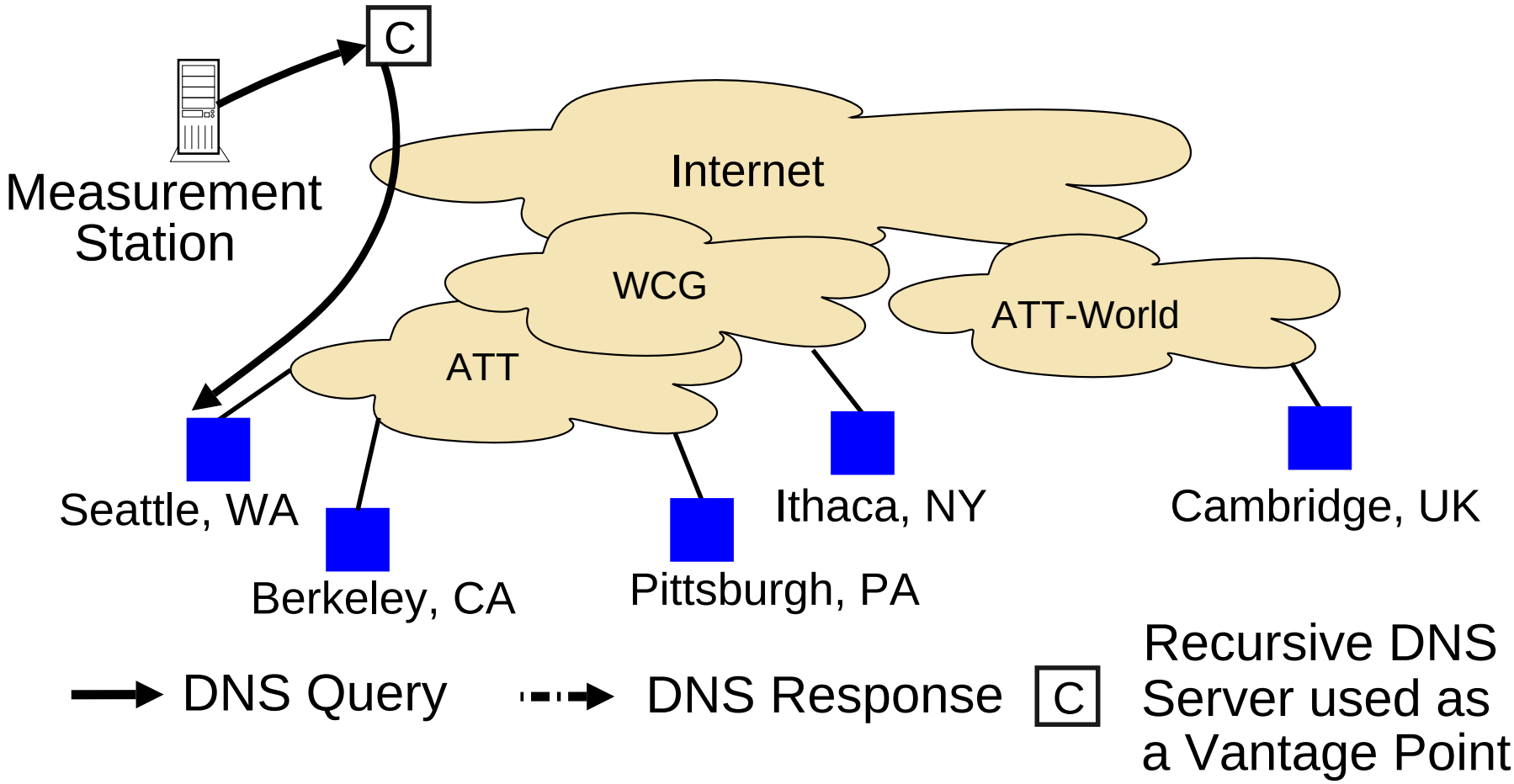
Probing Methodology

Recursive DNS Servers can be used as Vantage Points [KING, IMW'02]



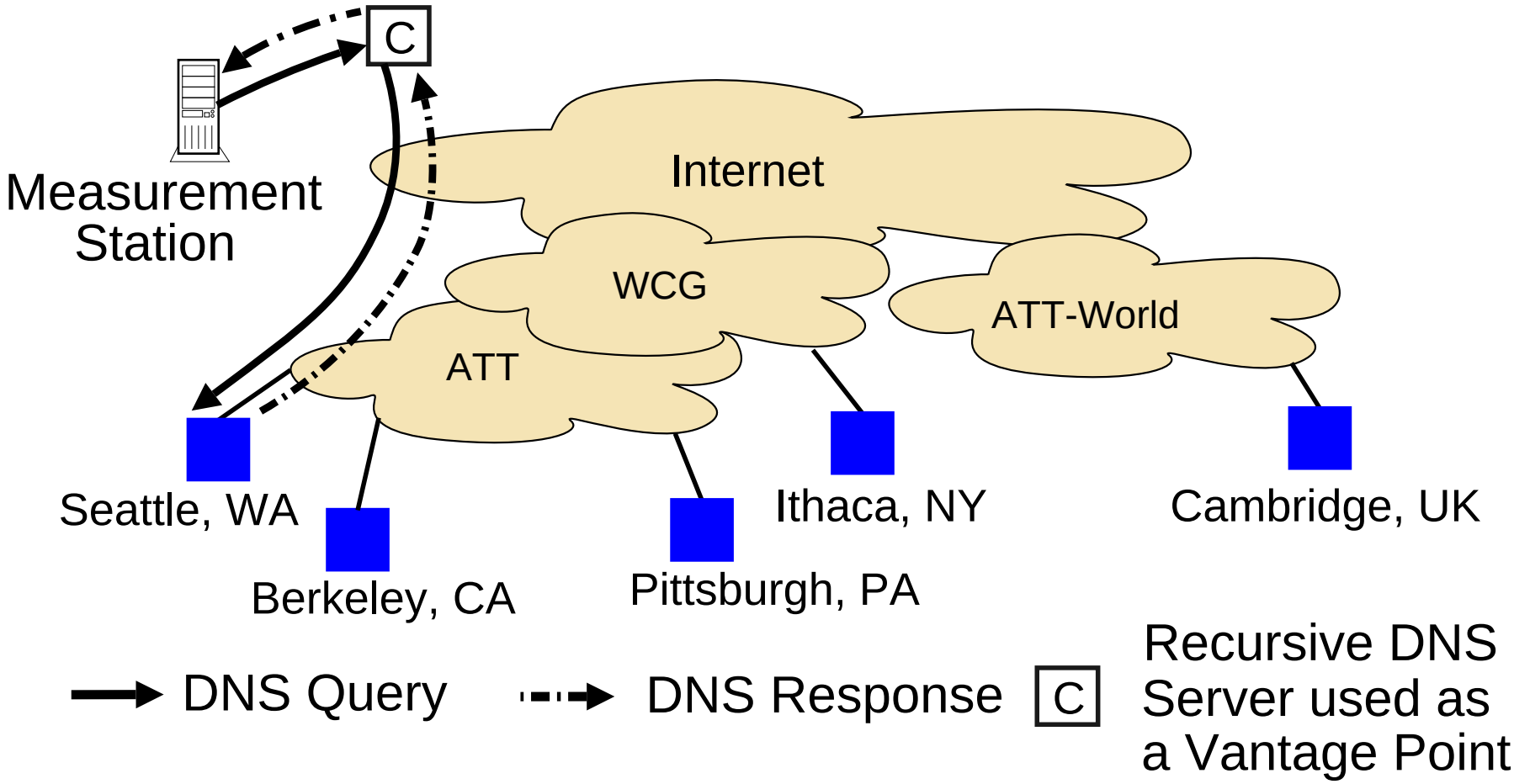
Probing Methodology

Query the Recursive DNS Server
such that it queries the Anycast Server



Probing Methodology

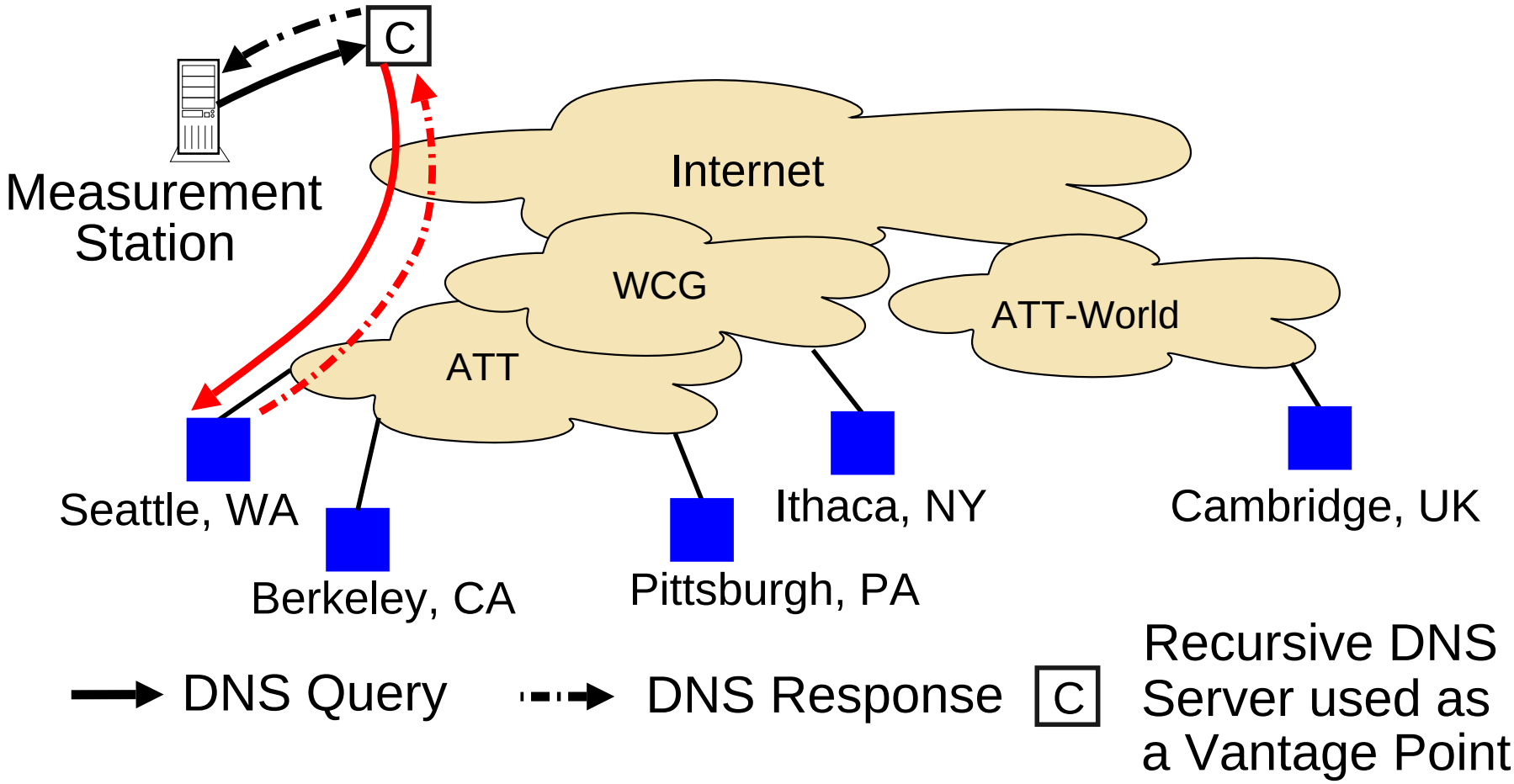
Response from the Anycast Server is forwarded onto the Measurement Host



Probing Methodology

Measurements: 1). Anycast Server being accessed by C

2). Latency from C to the Anycast Server



Probing Methodology

23,858 Recursive DNS Servers used as Vantage Points

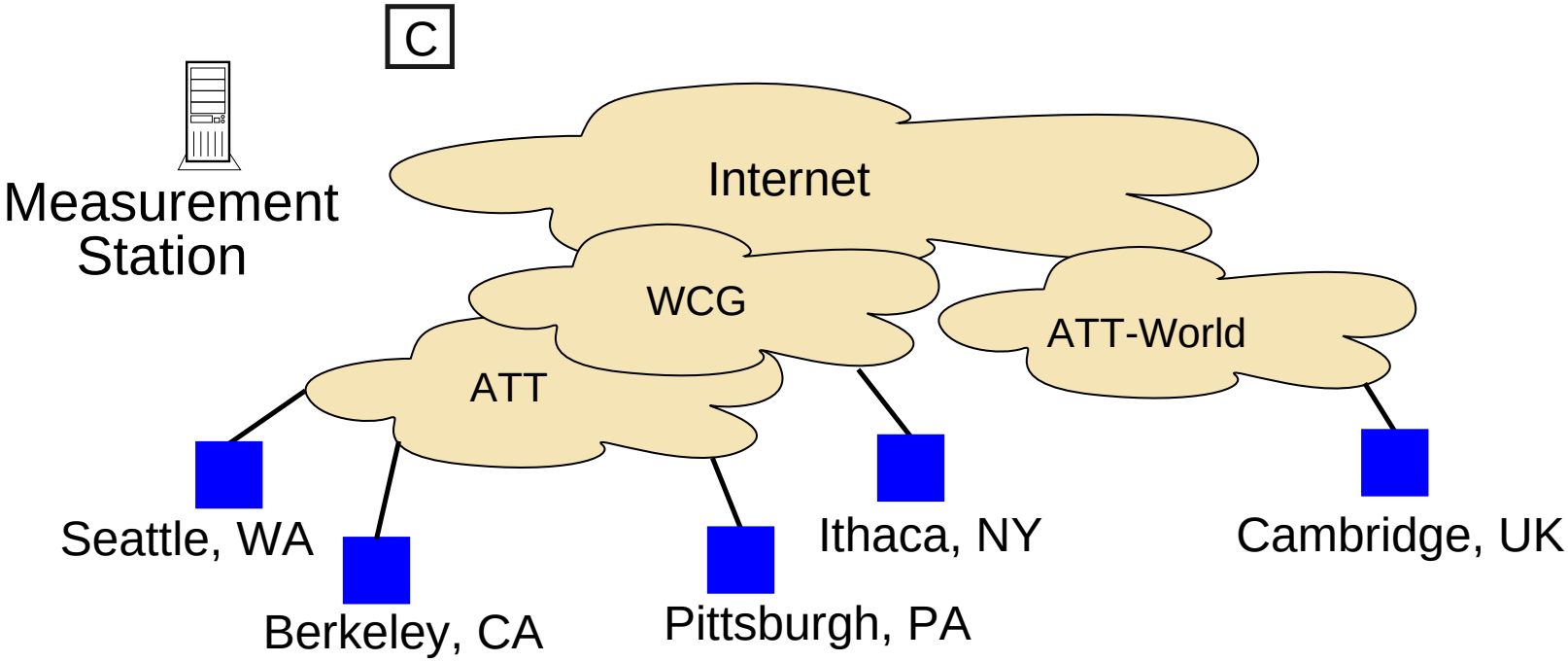
Region	No. of clients	% of Total
North America	12931	54.827
Central America	317	1.344
South America	461	1.954
Europe	5585	23.680
Asia	2402	10.184
S.E. Asia	566	2.400
Oceania	1196	5.071
Africa	187	0.792
Arctic Region	9	0.038
Unknown	204	0.864
Total	23858	100.000

Talk Outline

- Introduction
- Terminology
- Deployments Measured
- Methodology
- Measurements
  - Proximity
  - Failover
  - Load-distribution
  - Affinity
- Conclusions

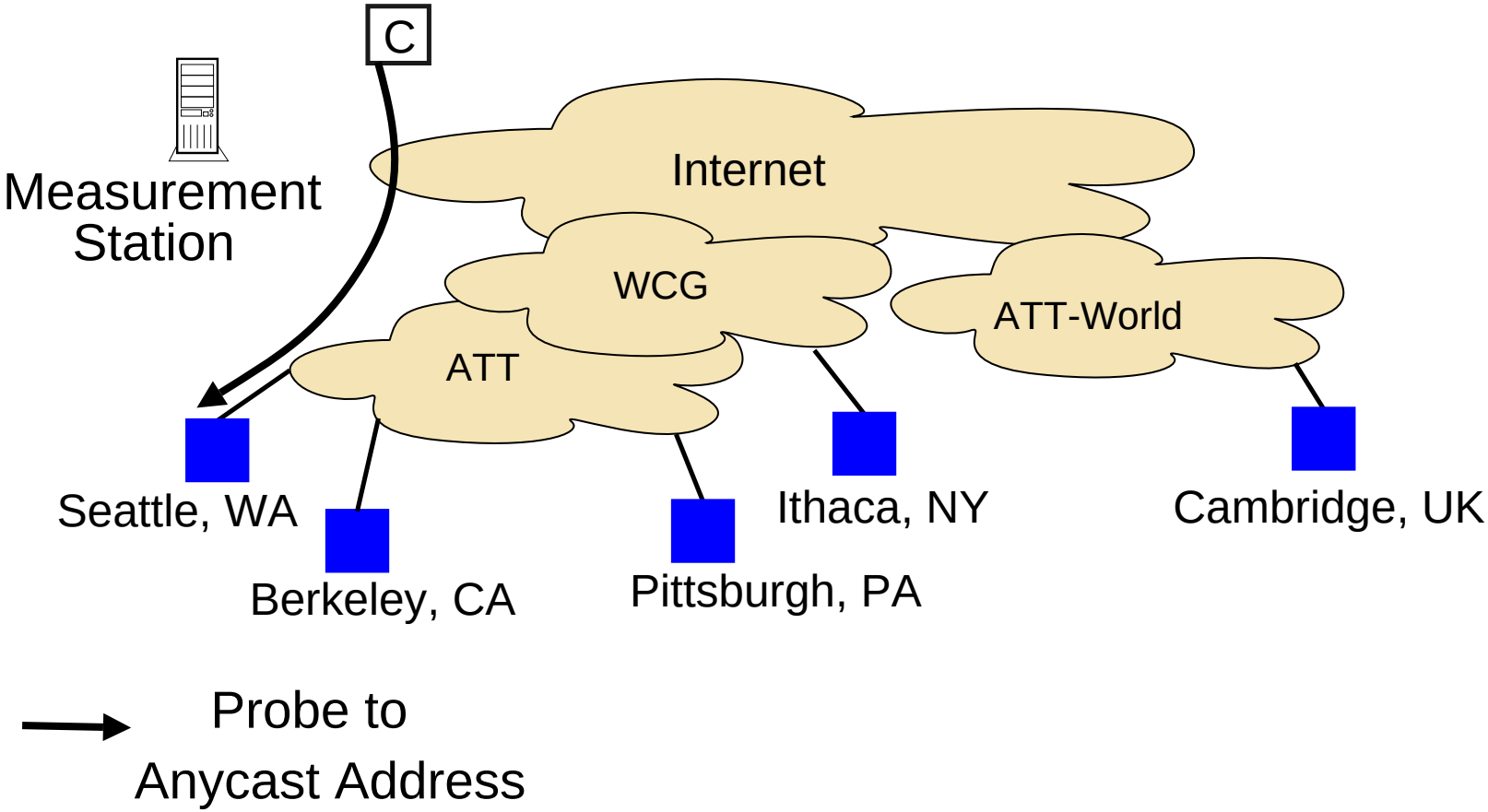
Proximity

Anycast Server is chosen by Inter-Domain Routing



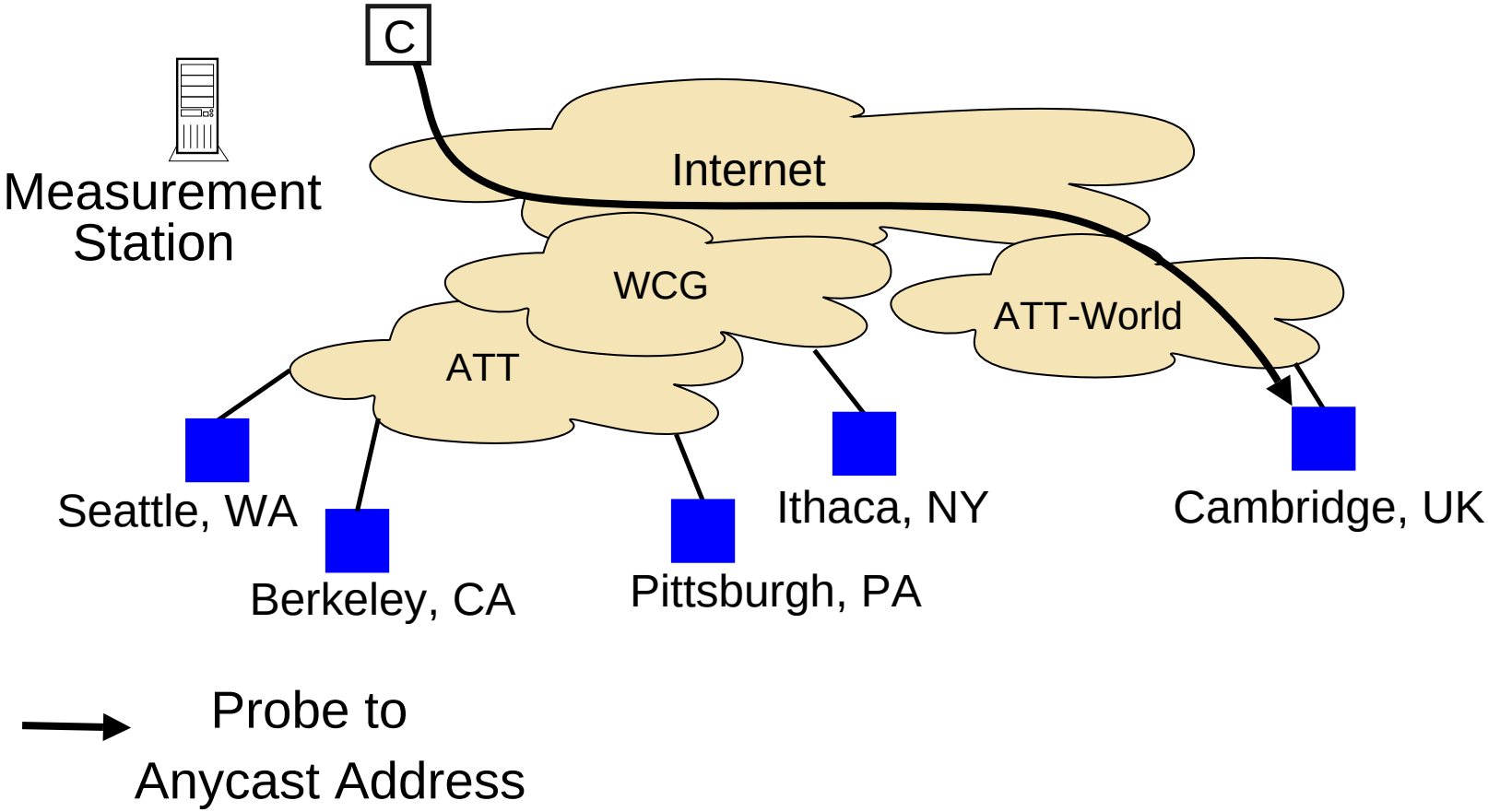
Proximity

Ideally, Clients would be routed
to a close-by Anycast Server



Proximity

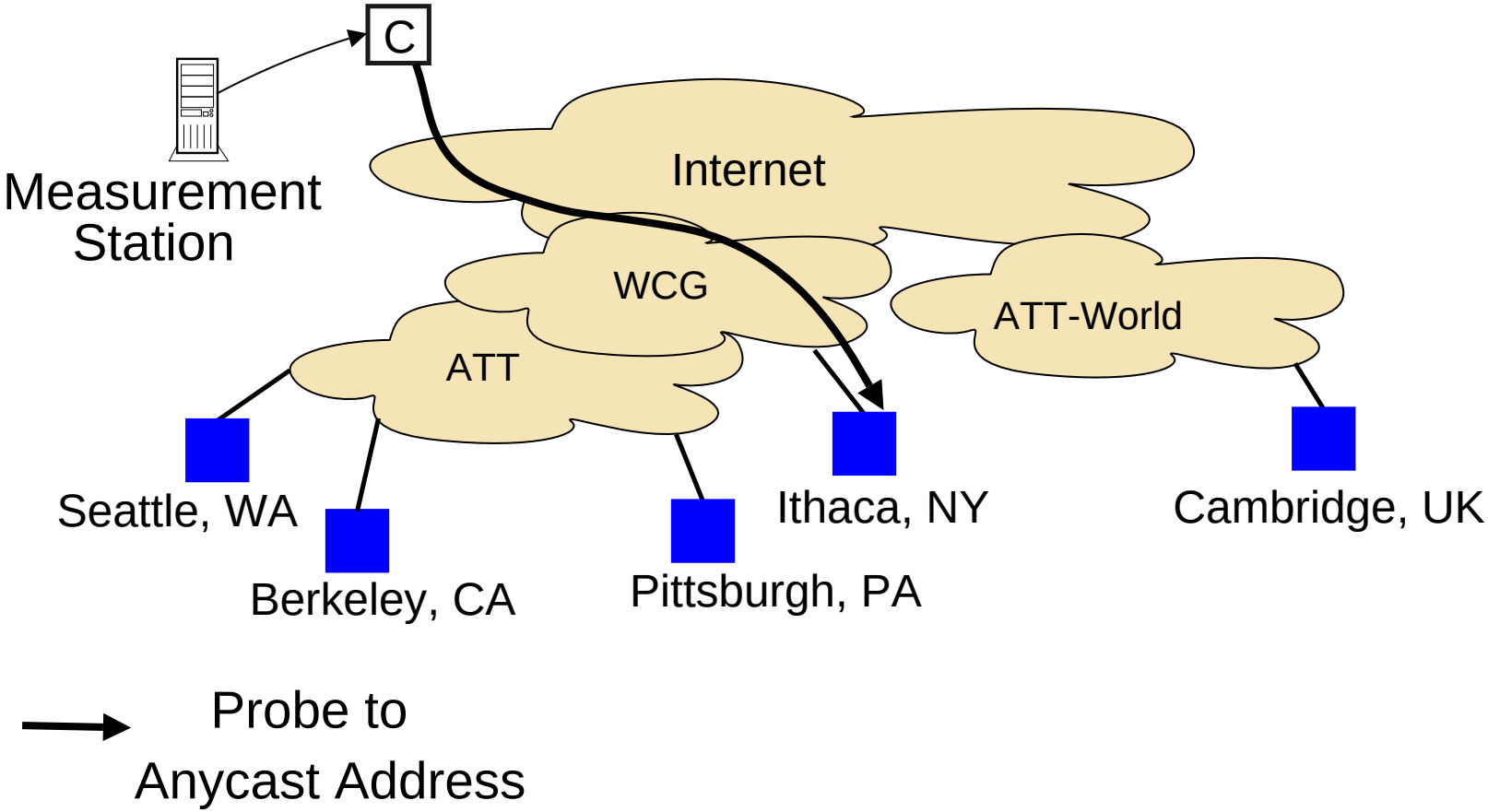
Poor choice of Anycast Server is possible!



Proximity

Probe the Deployment's Anycast Address from the client

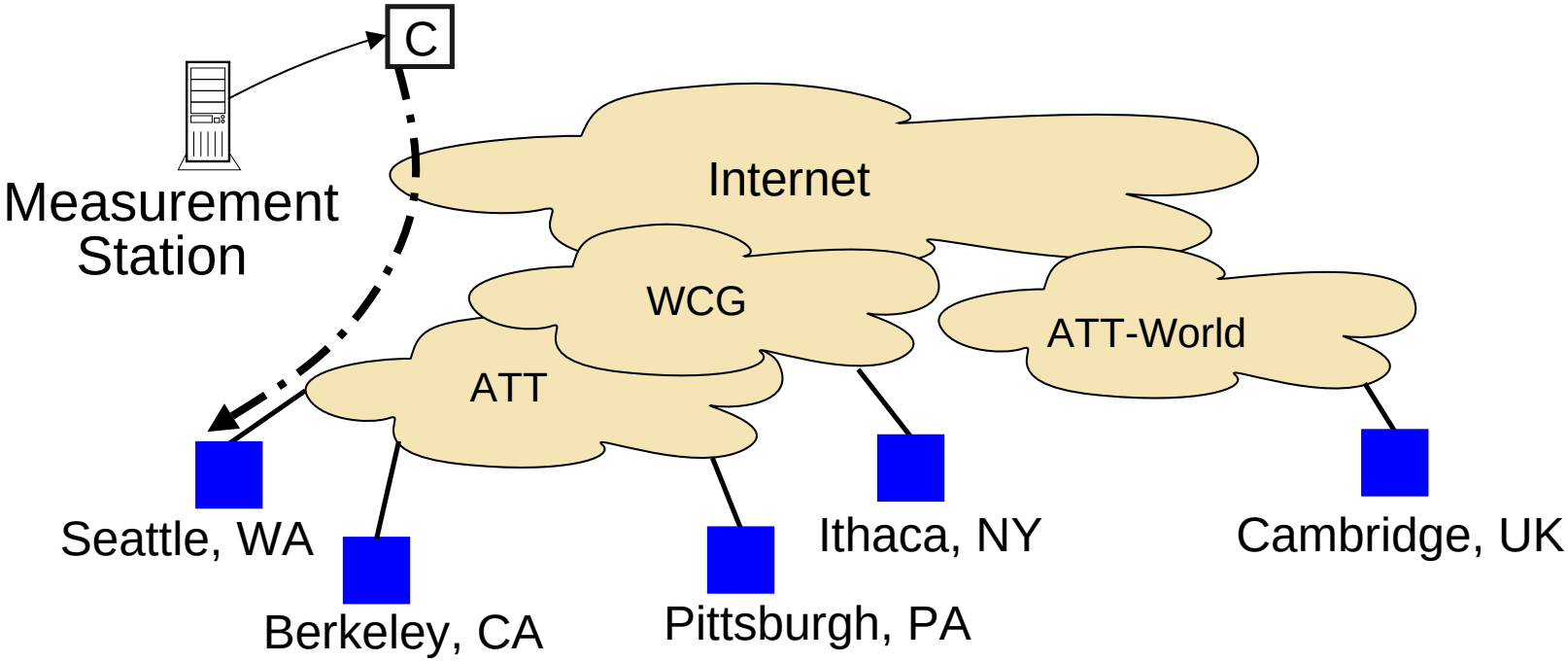
Probe Latency = Anycast Latency



Proximity

Client probes individual Anycast Servers

Latency to closest Anycast Server = Min. Unicast Latency

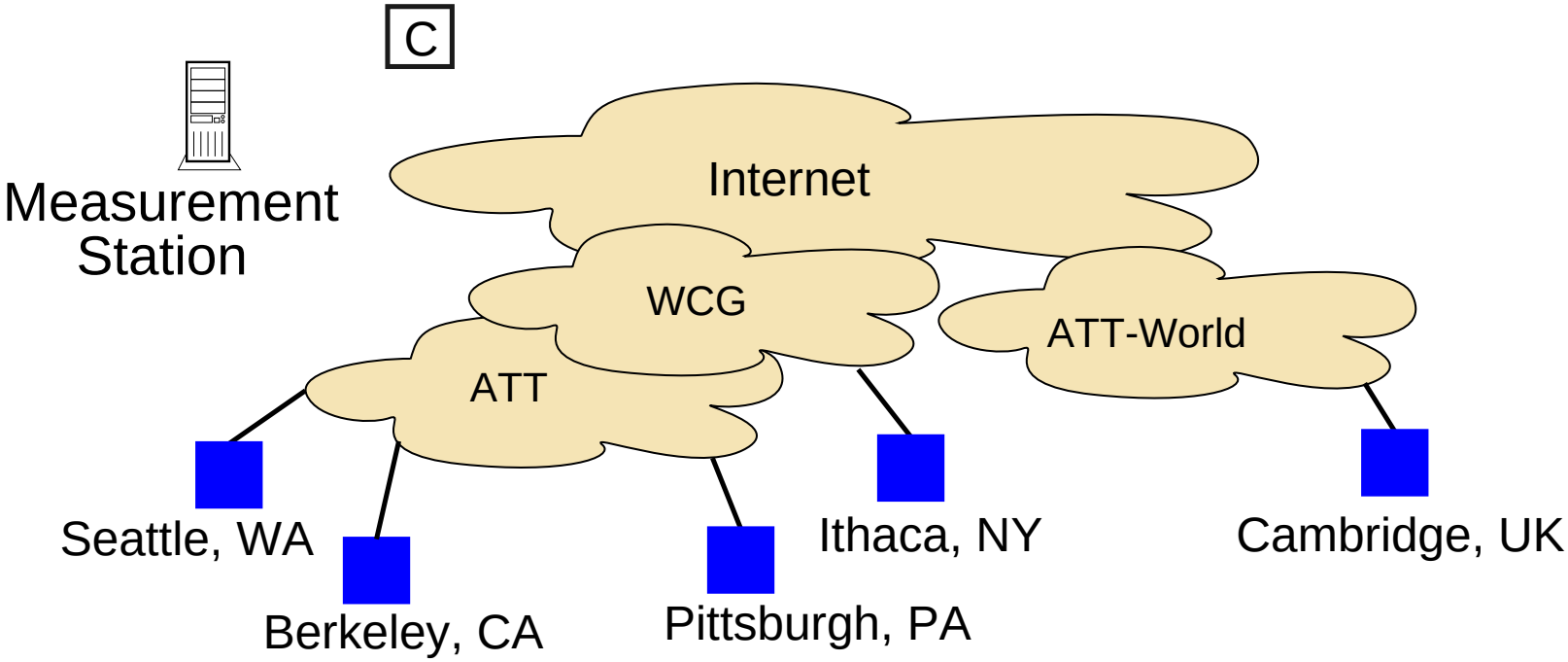


Probe to Anycast Address

Probe to Unicast Address

Proximity

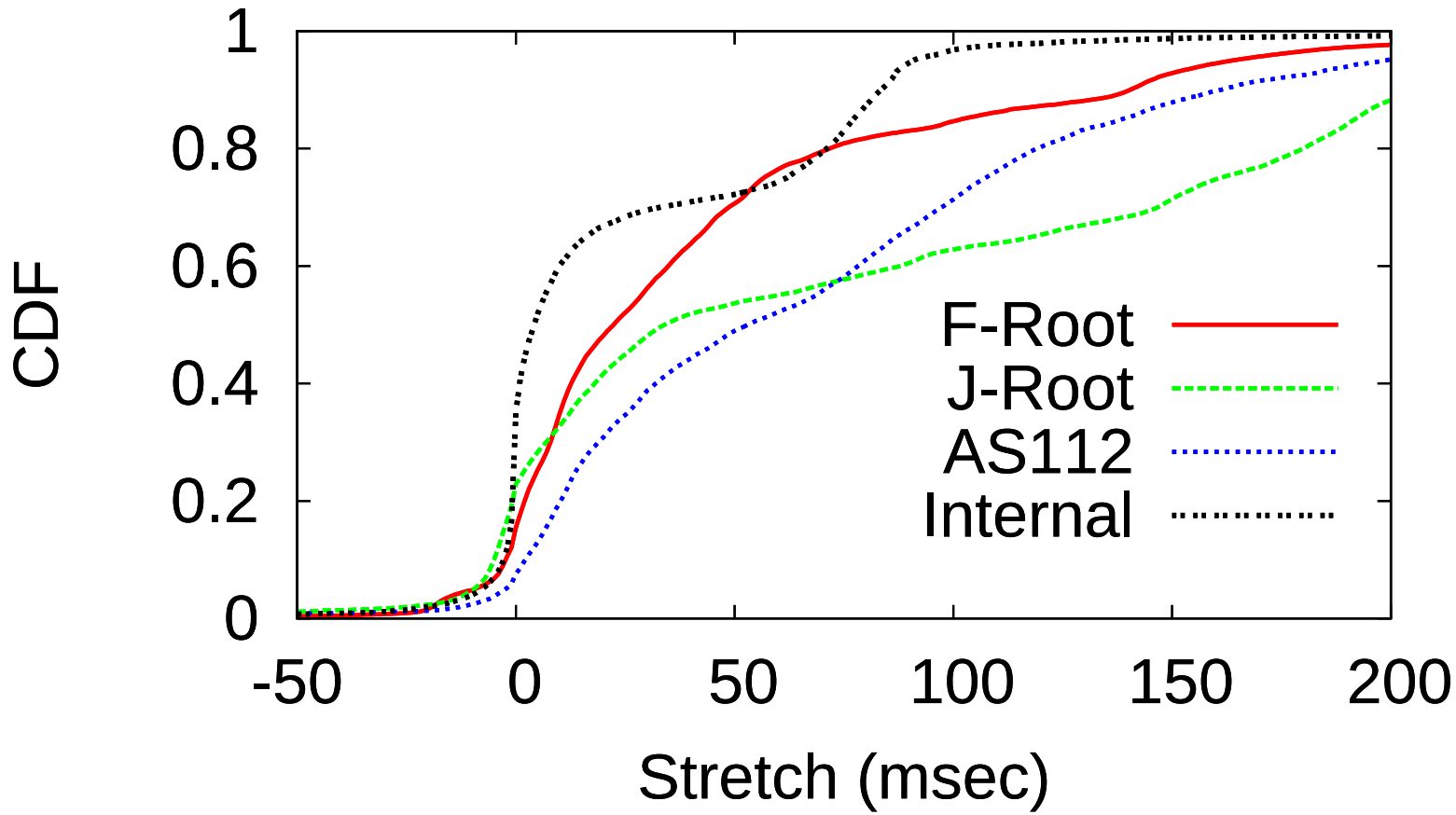
STRETCH = (Anycast Latency
- Minimum Unicast Latency)



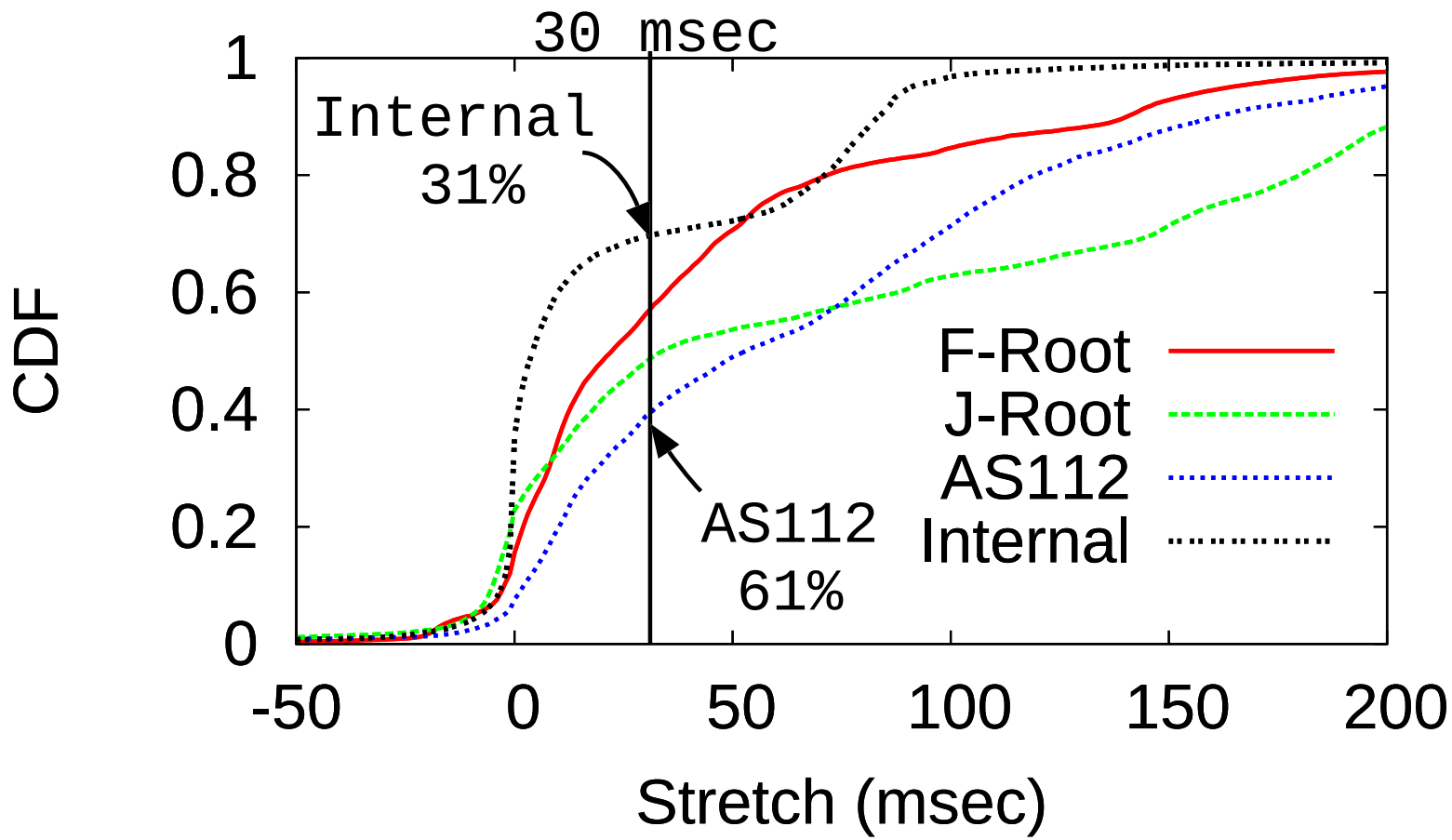
Probe to Anycast Address

Probe to Unicast Address

Proximity


1
0.8
0.6
0.4
0.2
0
CDF
-50
0
50
100
150
200
Stretch (msec)
F-Root
J-Root
AS112
Internal

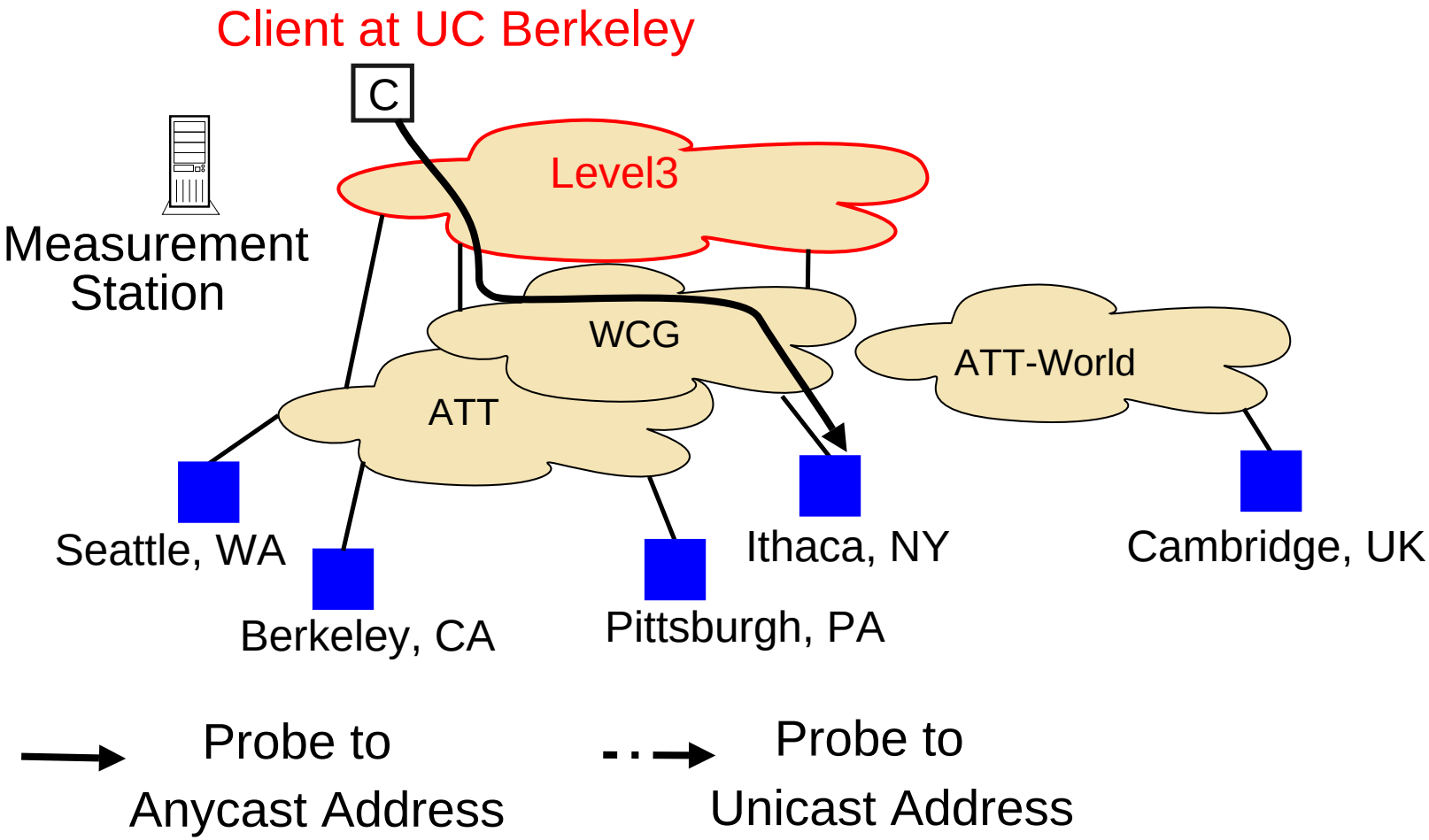
Proximity



All four deployments offer poor Proximity

Investigating Poor Proximity

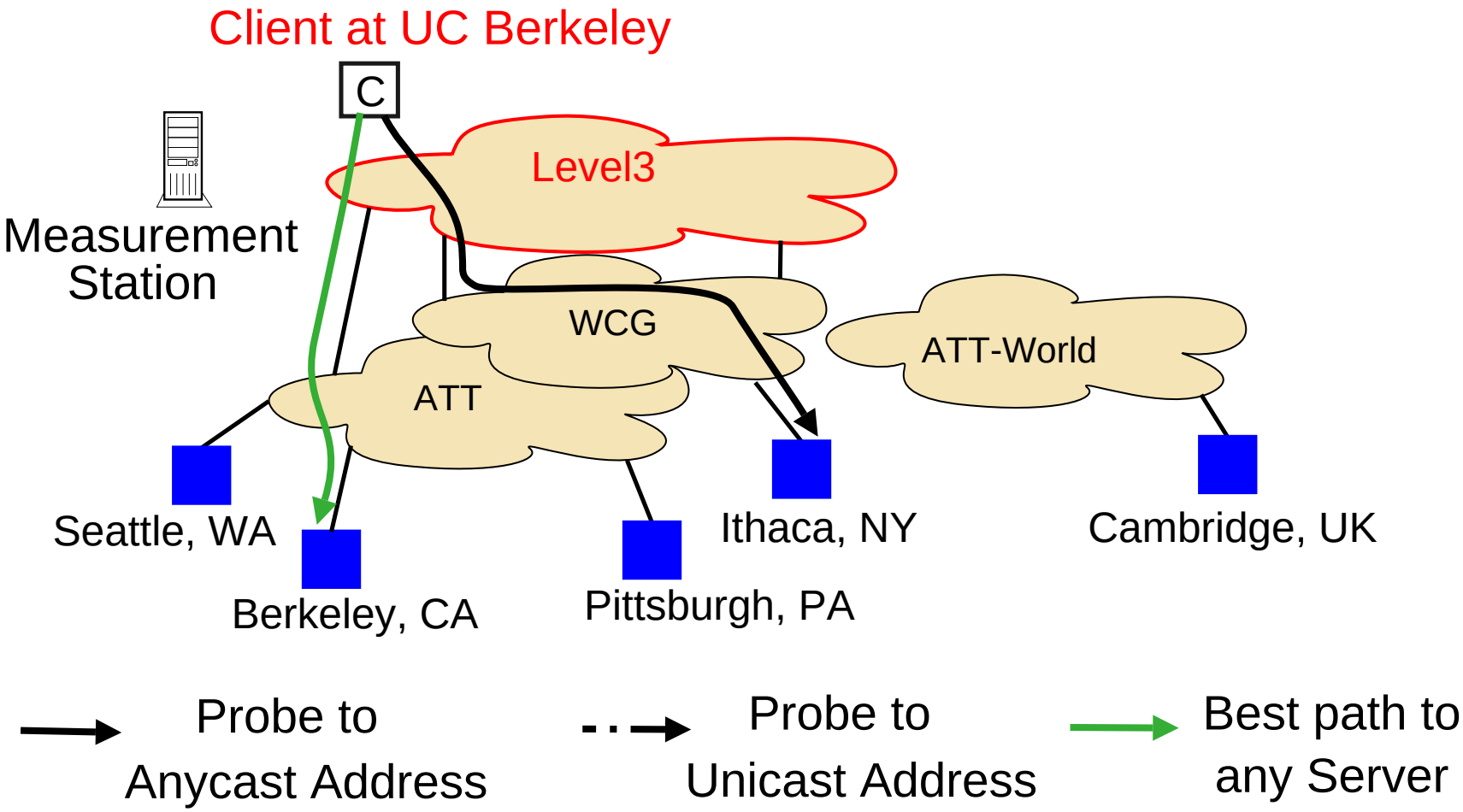
Client probes Anycast address of Internal Deployment



Investigating Poor Proximity

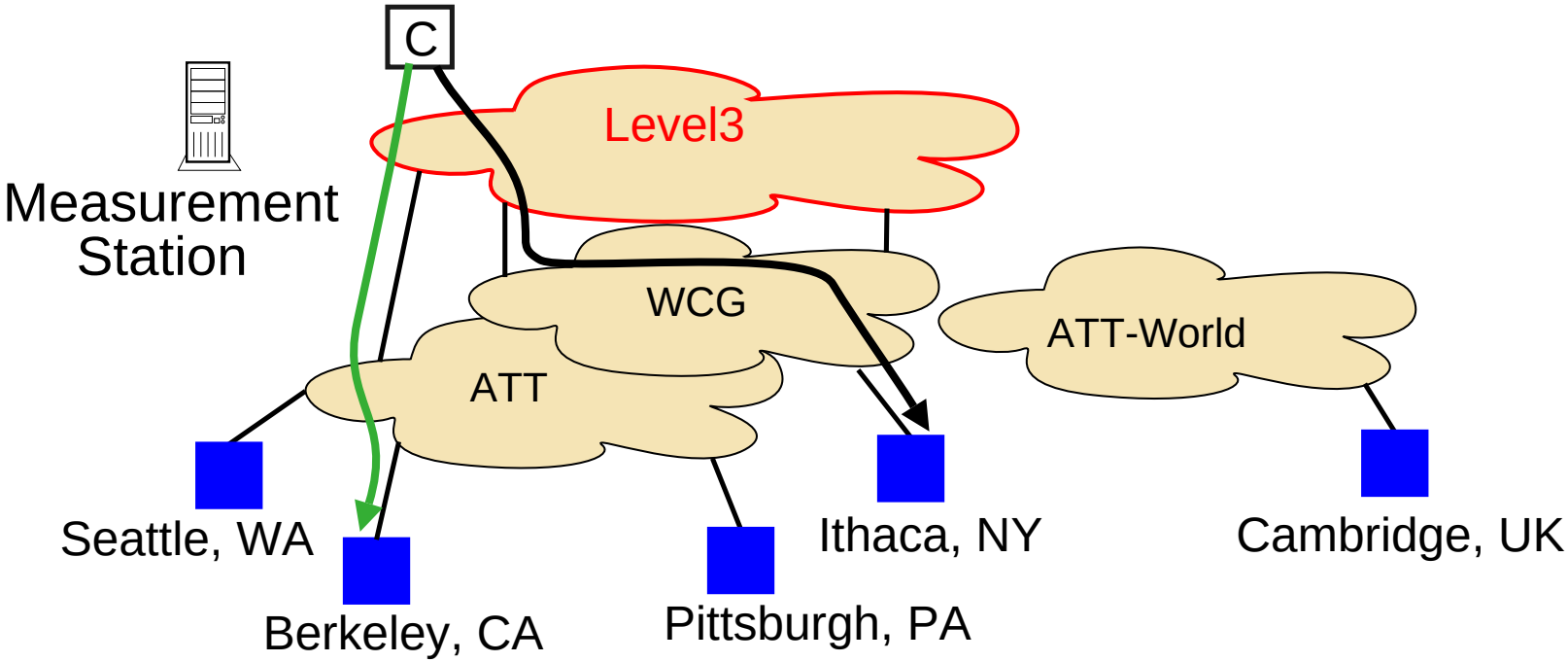
Client probes Anycast address of Internal Deployment

Routed to Ithaca (NY) instead of Berkeley (CA)



Investigating Poor Proximity

See this example at
http://pias.gforge.cis.cornell.edu/trace.html



Probe to Anycast Address

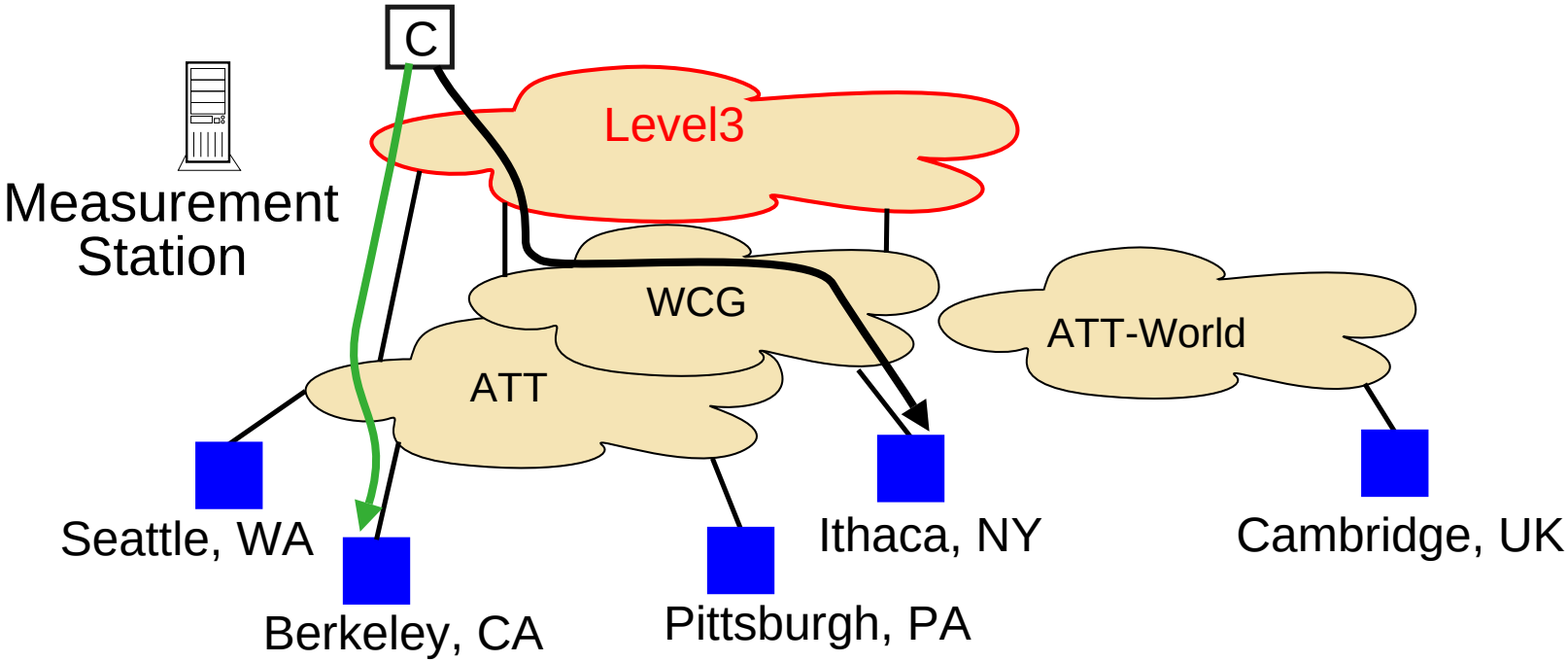
Probe to Unicast Address

Best path to any Server

Investigating Poor Proximity

Level3 has two paths of 2 AS-hops: through ATT and WCG

Level3 does not realize that these lead to different locations



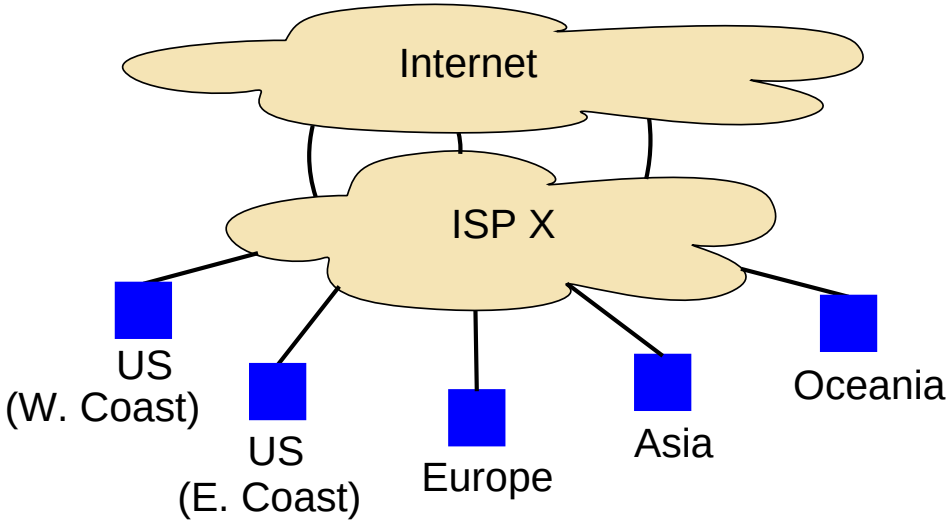
Probe to Anycast Address

Probe to Unicast Address

Best path to any Server

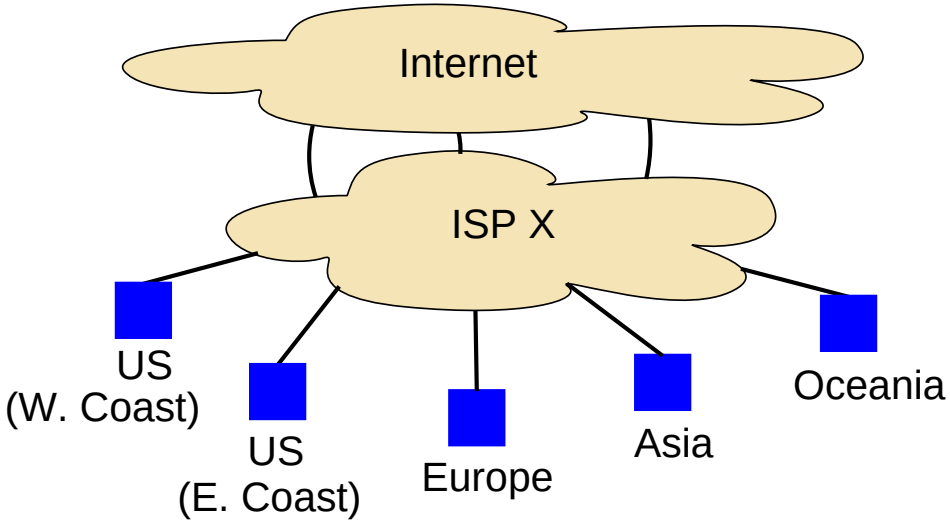
Alleviating Poor Proximity

Anycast Servers should have the same Upstream ISP

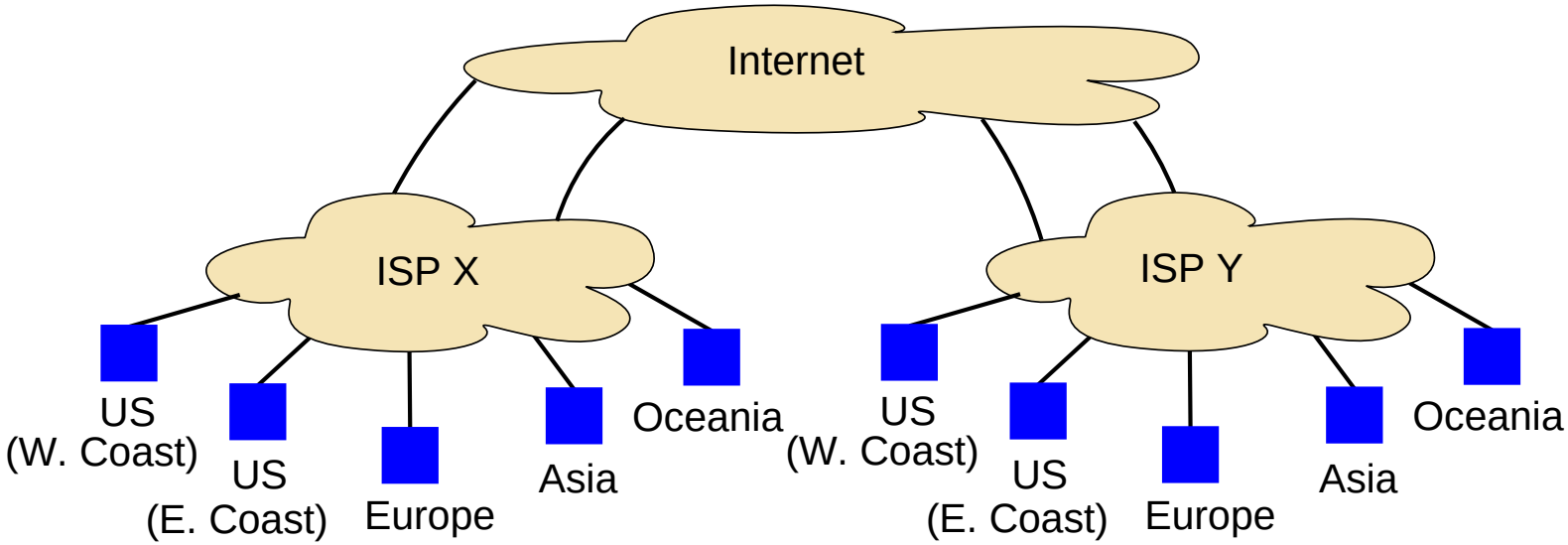


Alleviating Poor Proximity

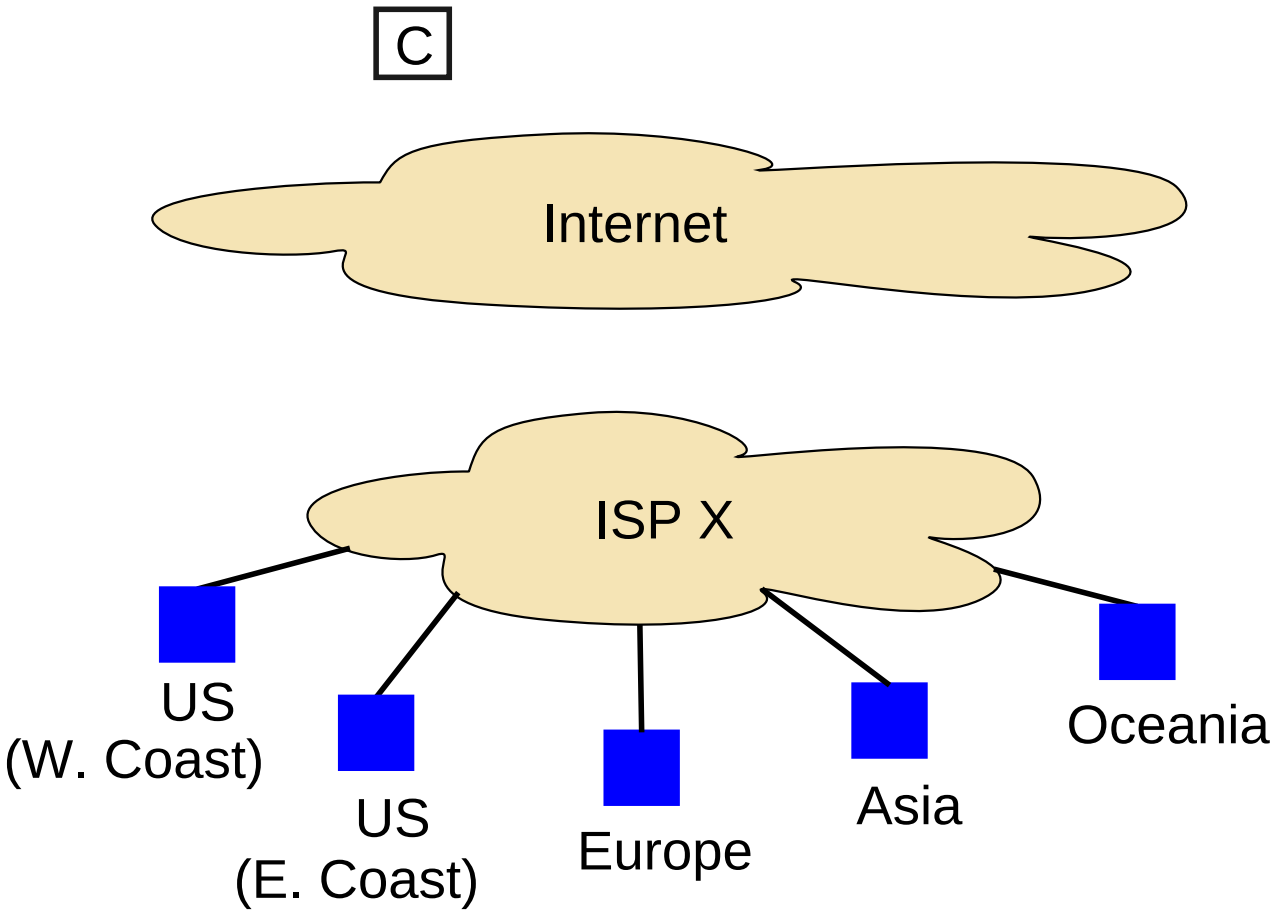
Anycast Servers should have the same Upstream ISP



Multiple Providers: Geographically cover all providers

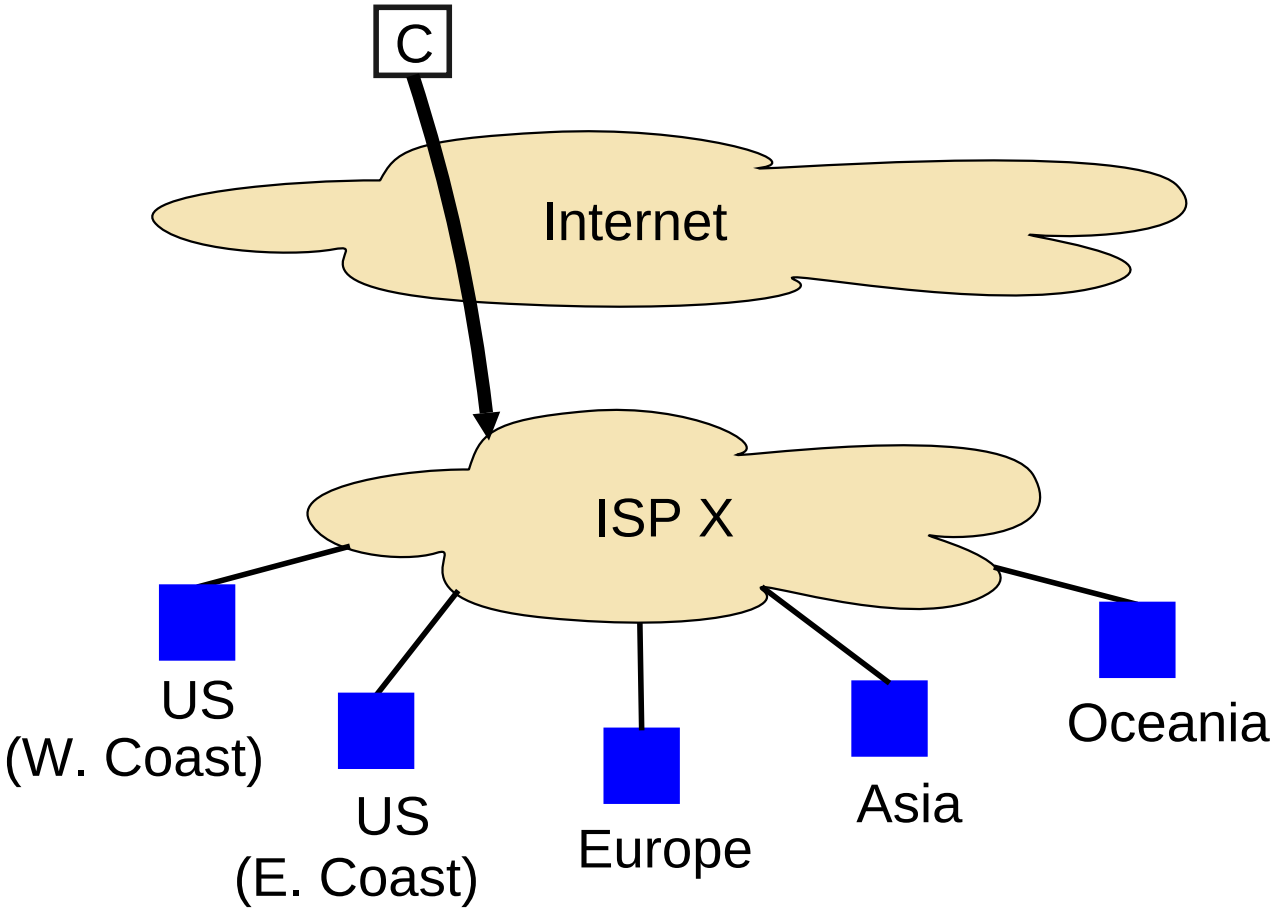


Alleviating Poor Proximity



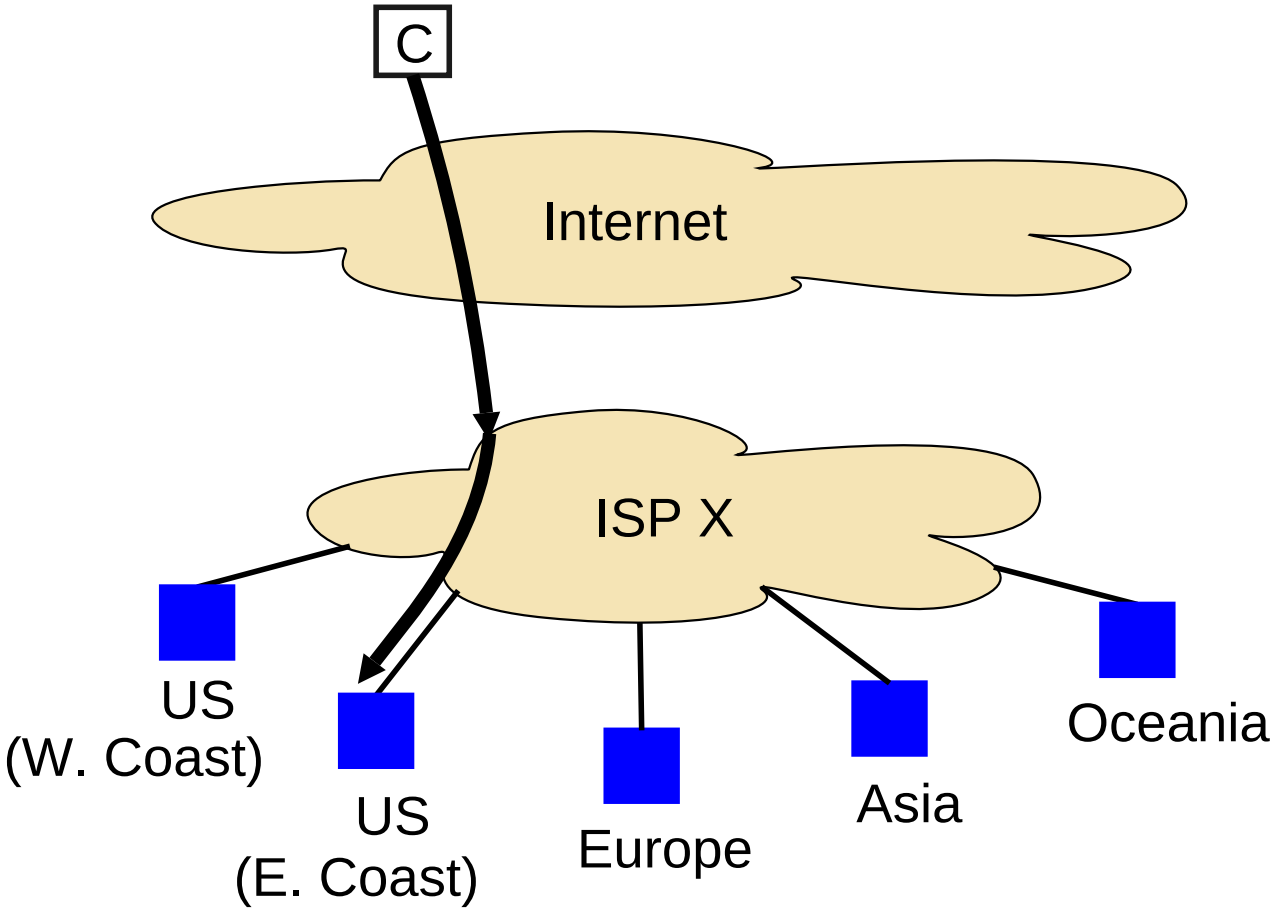
Alleviating Poor Proximity

Anycast probes from client(s) routed to ISP X



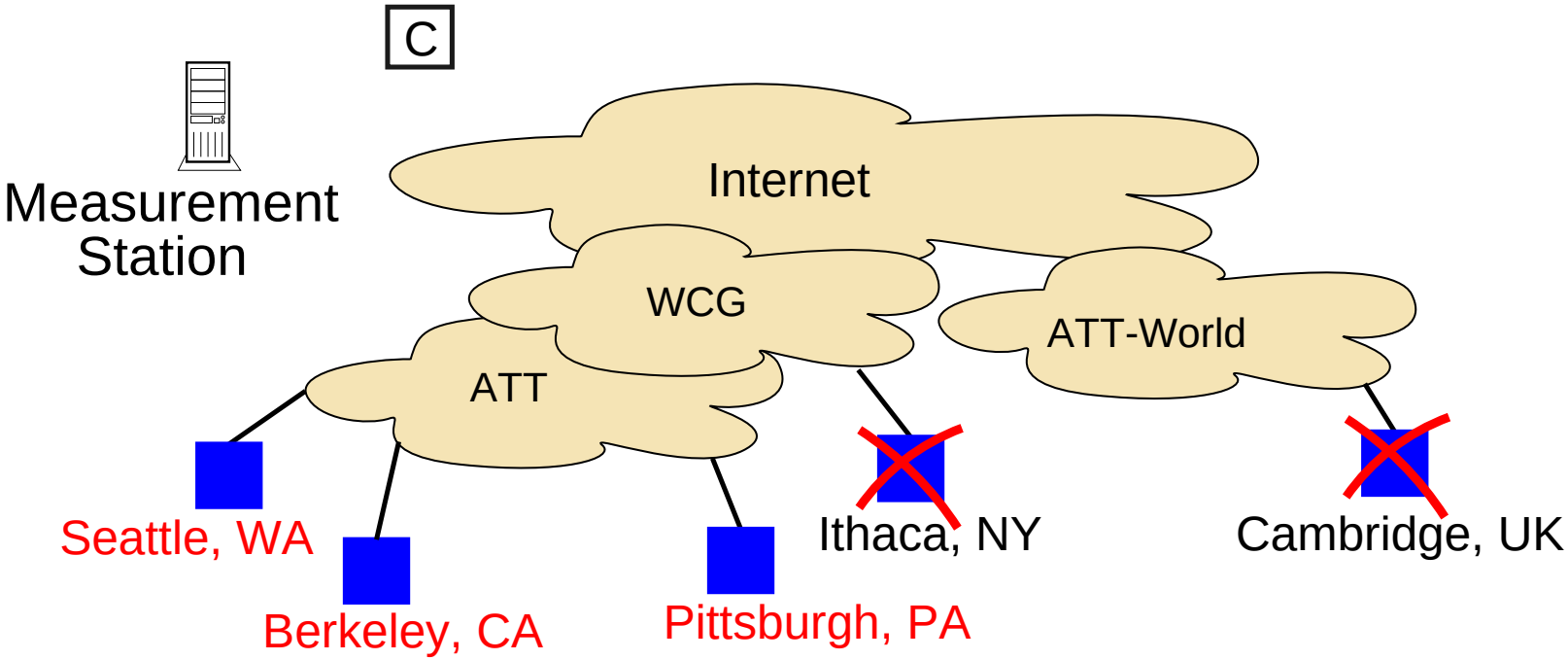
Alleviating Poor Proximity

Server chosen is based on X's intra-domain routing

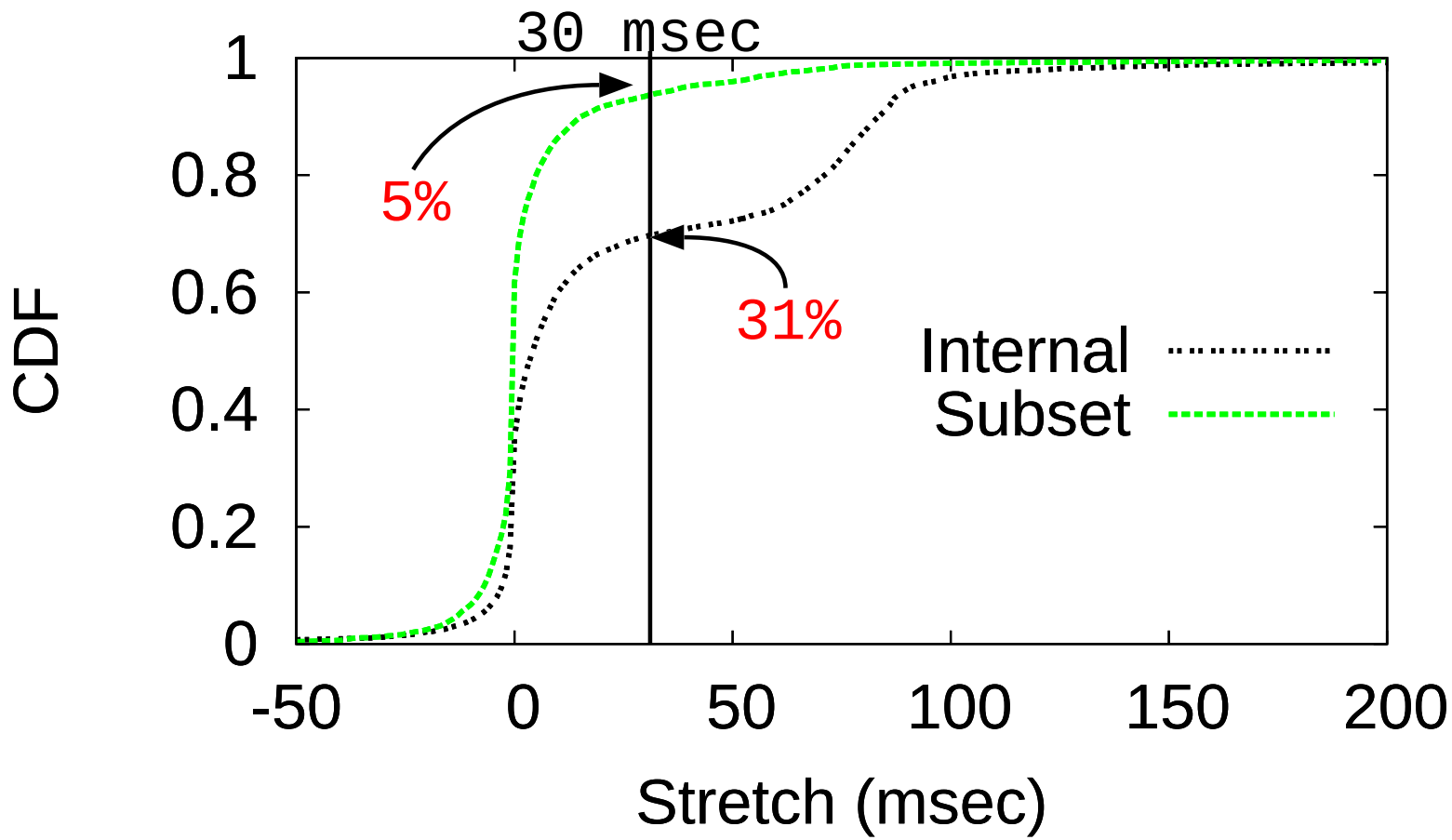


Verifying our hypothesis

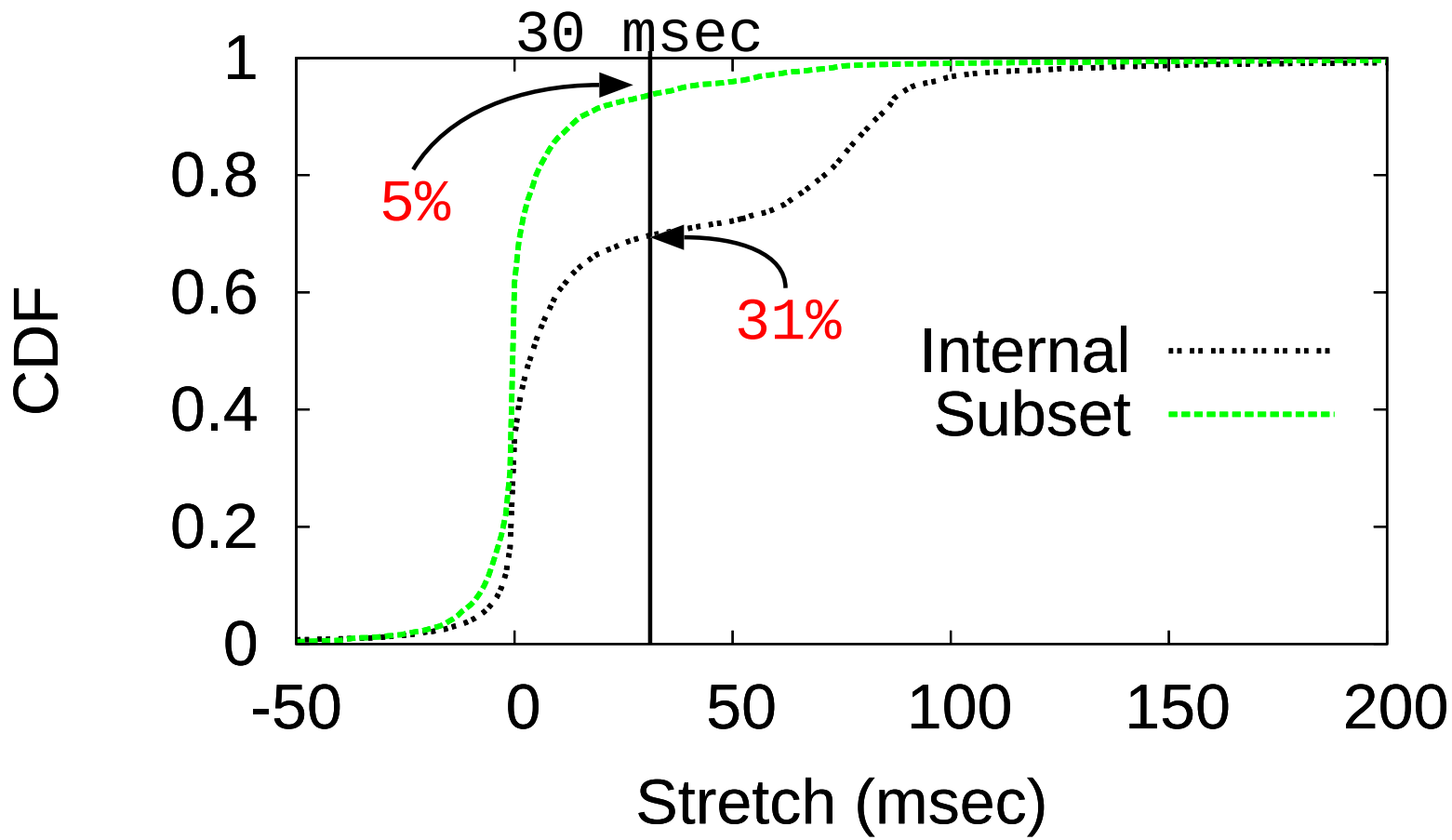
A subset of the Internal Deployment



Verifying our hypothesis

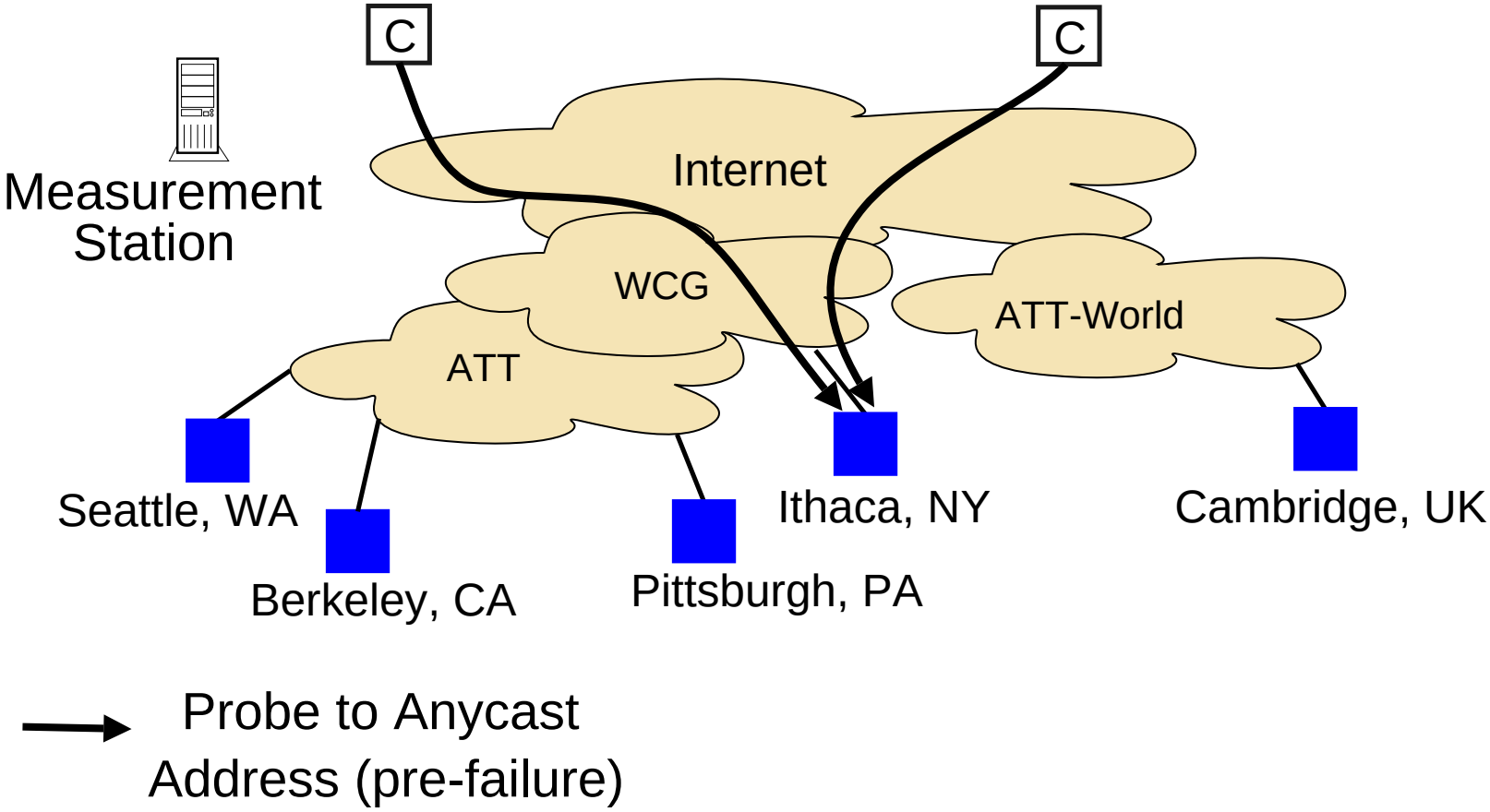


Verifying our hypothesis



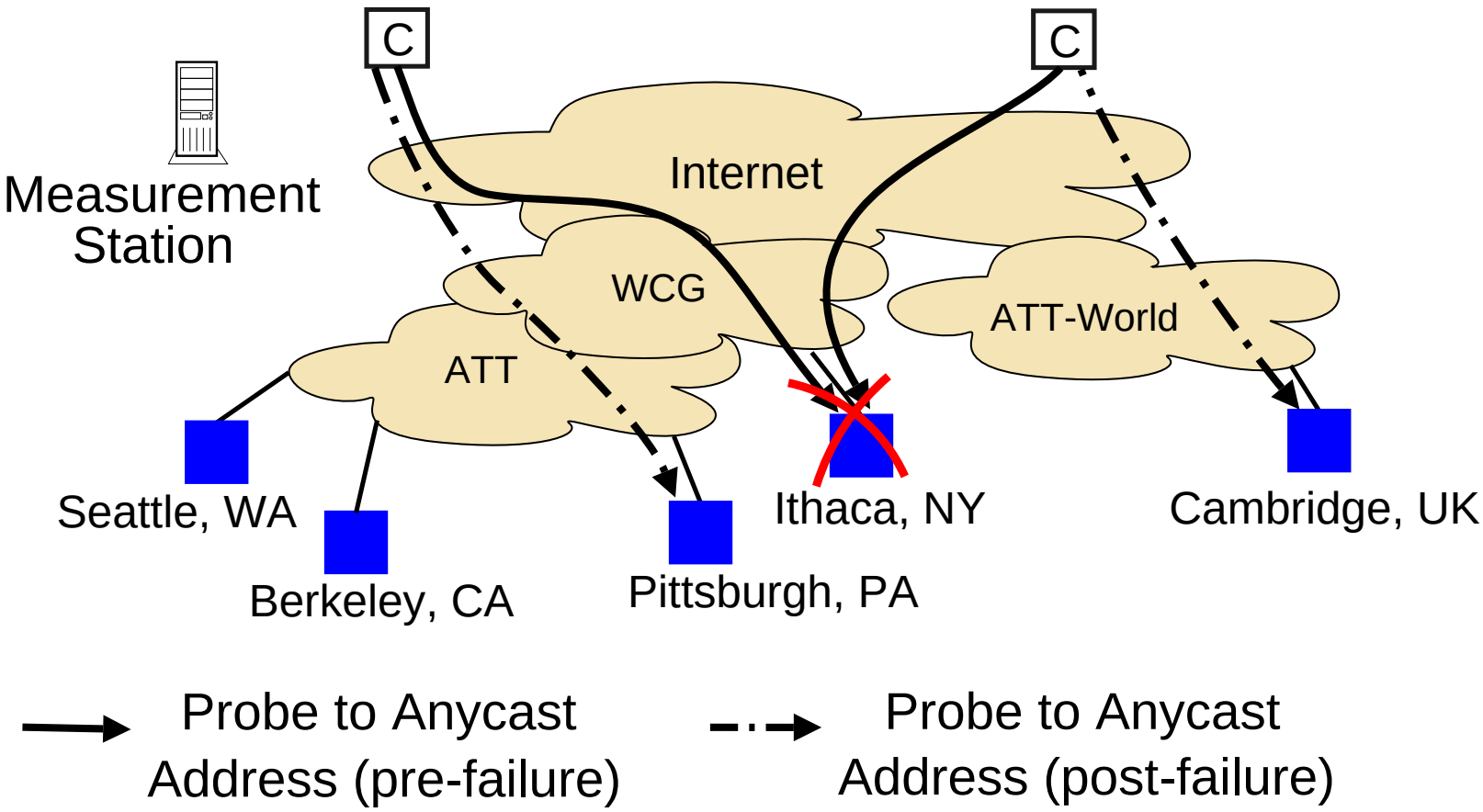
Planned Anycast Deployment $\Rightarrow$ good Proximity

Failover


C
C
Measurement Station
Internet
WCG
ATT-World
ATT
Seattle, WA
Berkeley, CA
Pittsburgh, PA
Ithaca, NY
Cambridge, UK
Probe to Anycast Address (pre-failure)

Failover

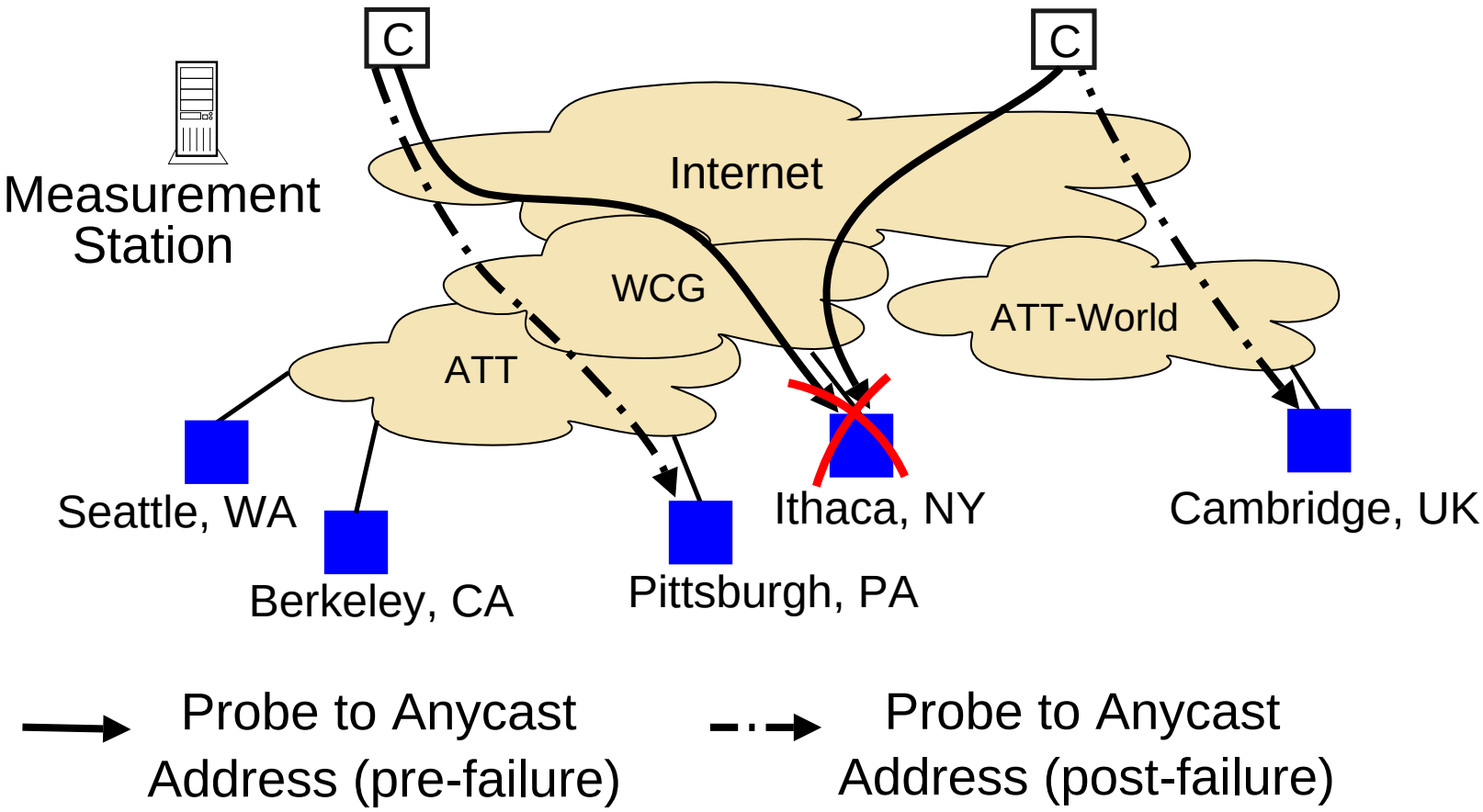
Clients are re-routed to a different Anycast Server



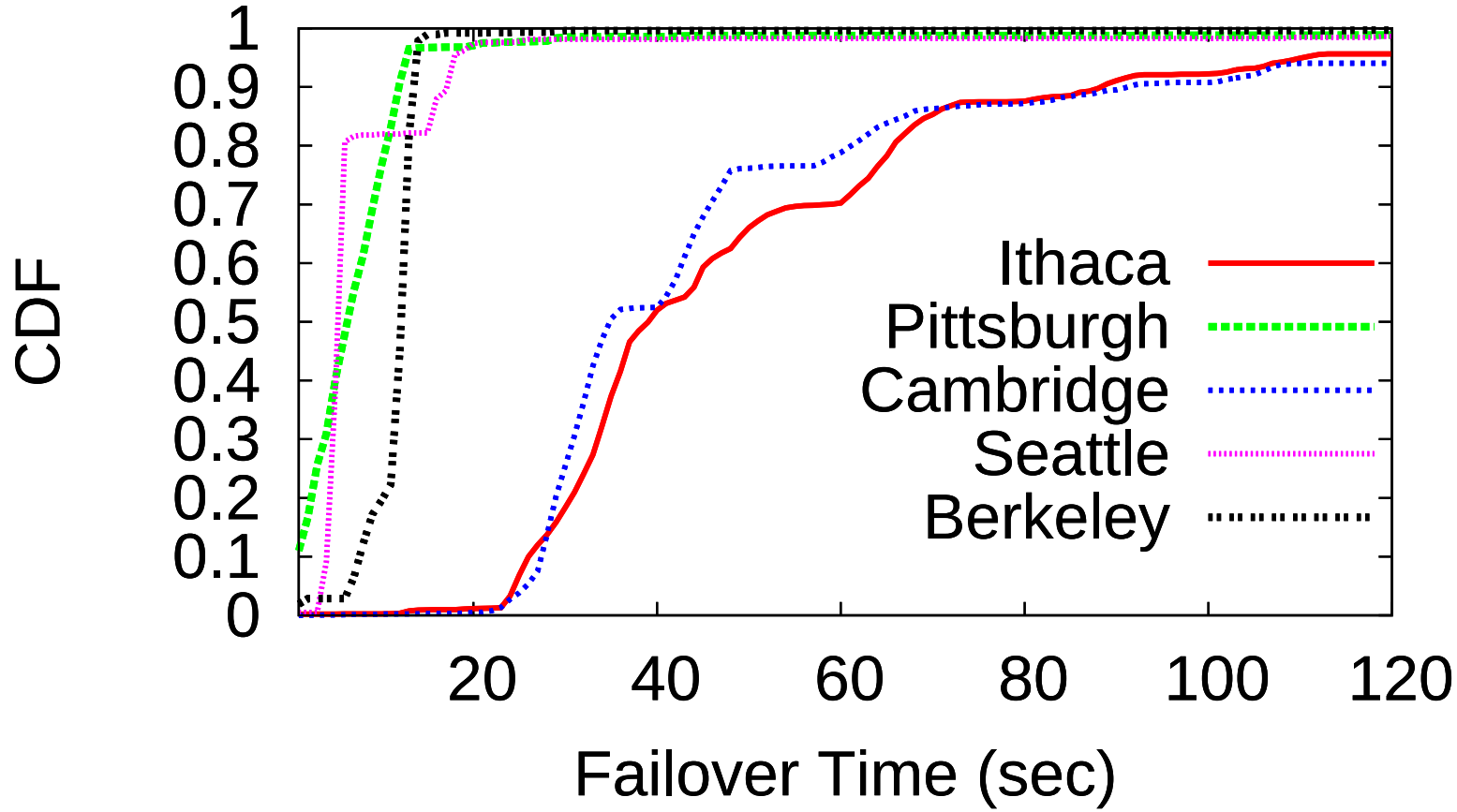
Failover

Clients are re-routed to a different Anycast Server

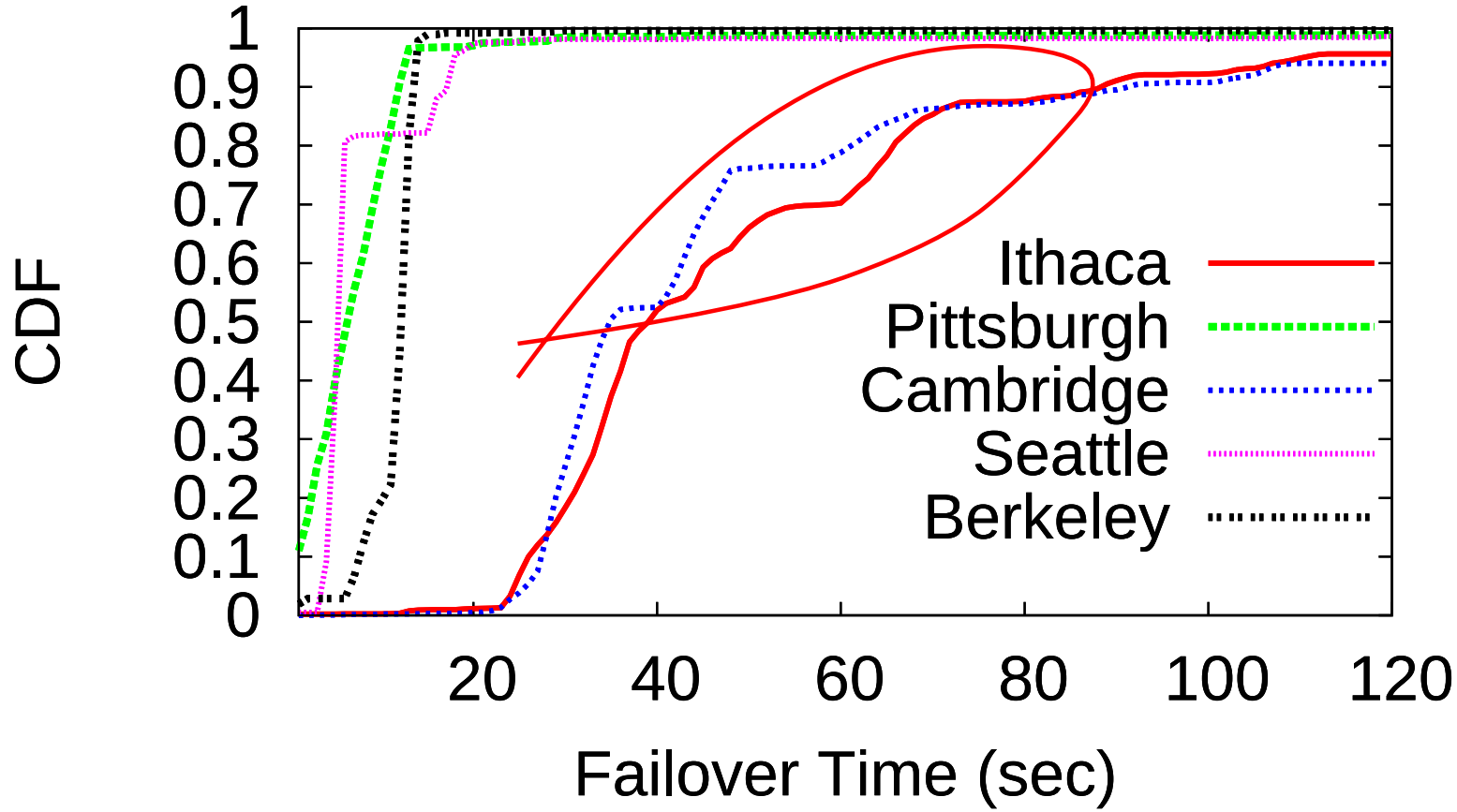
What is the failover time?



Failover

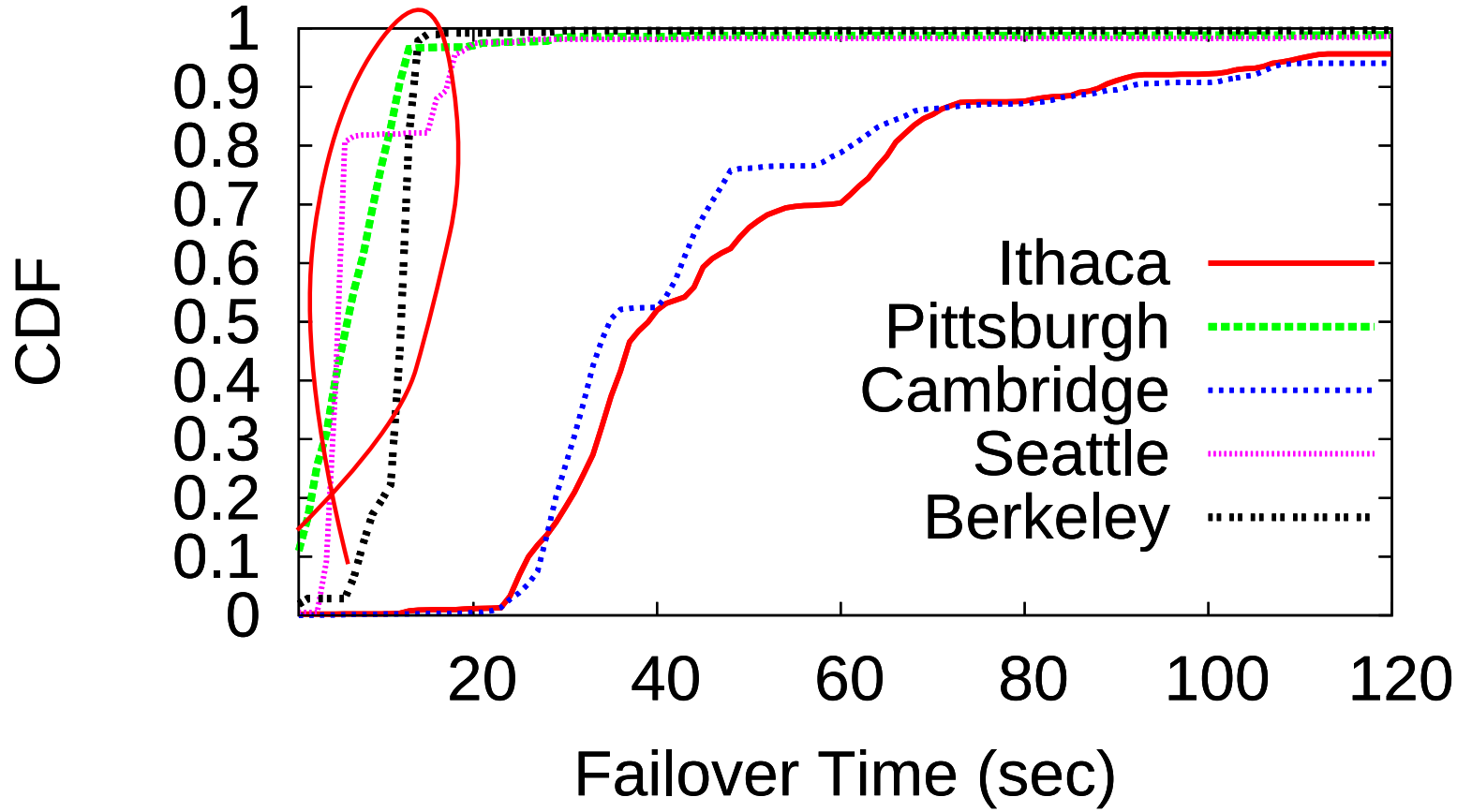

CDF
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
20
40
60
80
100
120
Failover Time (sec)
Ithaca
Pittsburgh
Cambridge
Seattle
Berkeley

Failover



Ithaca and Cambridge servers have slow failover

Failover

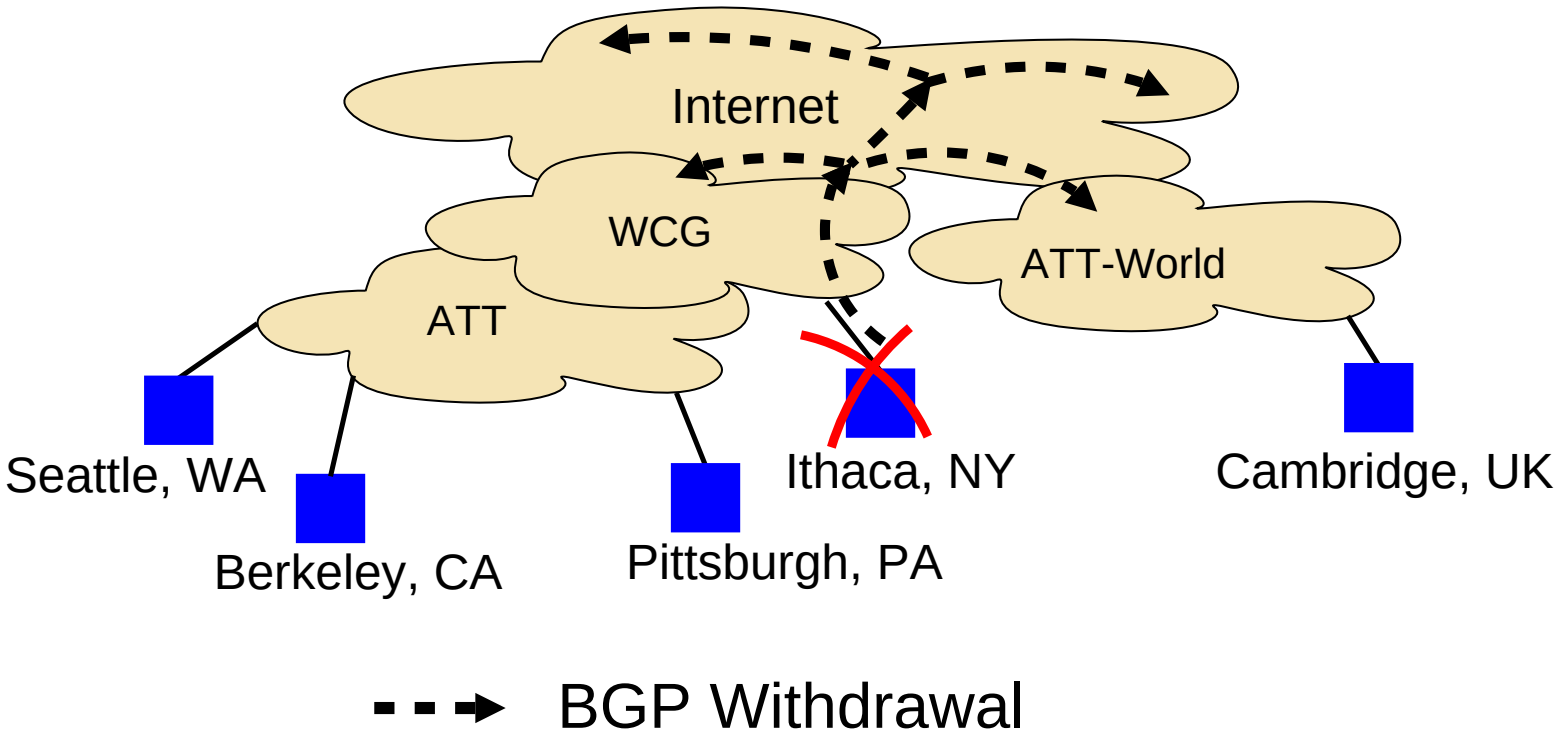


Other servers have faster failover

Failover

BGP Withdrawal propagated beyond WCG

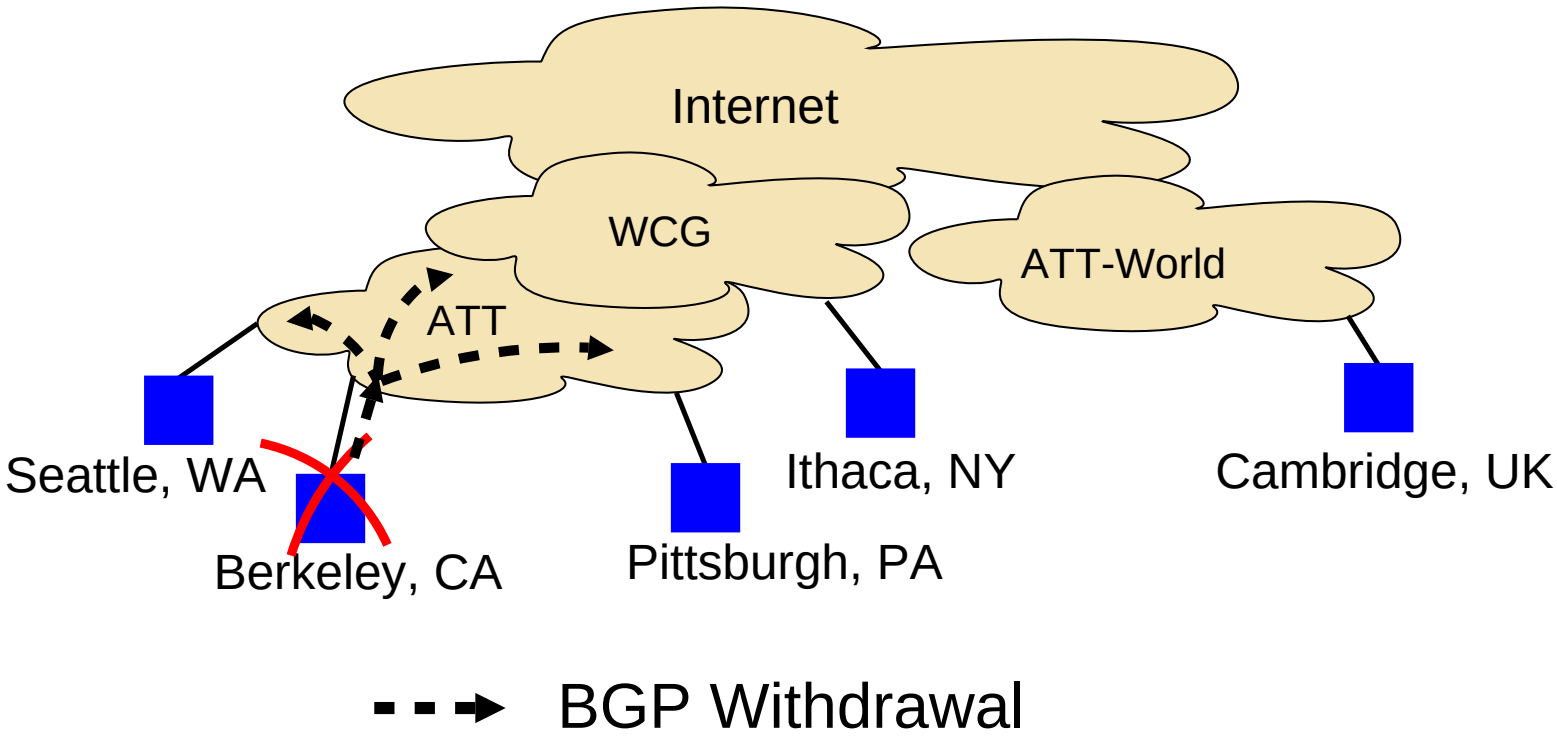
Global Routing Event → Slow convergence



Failover

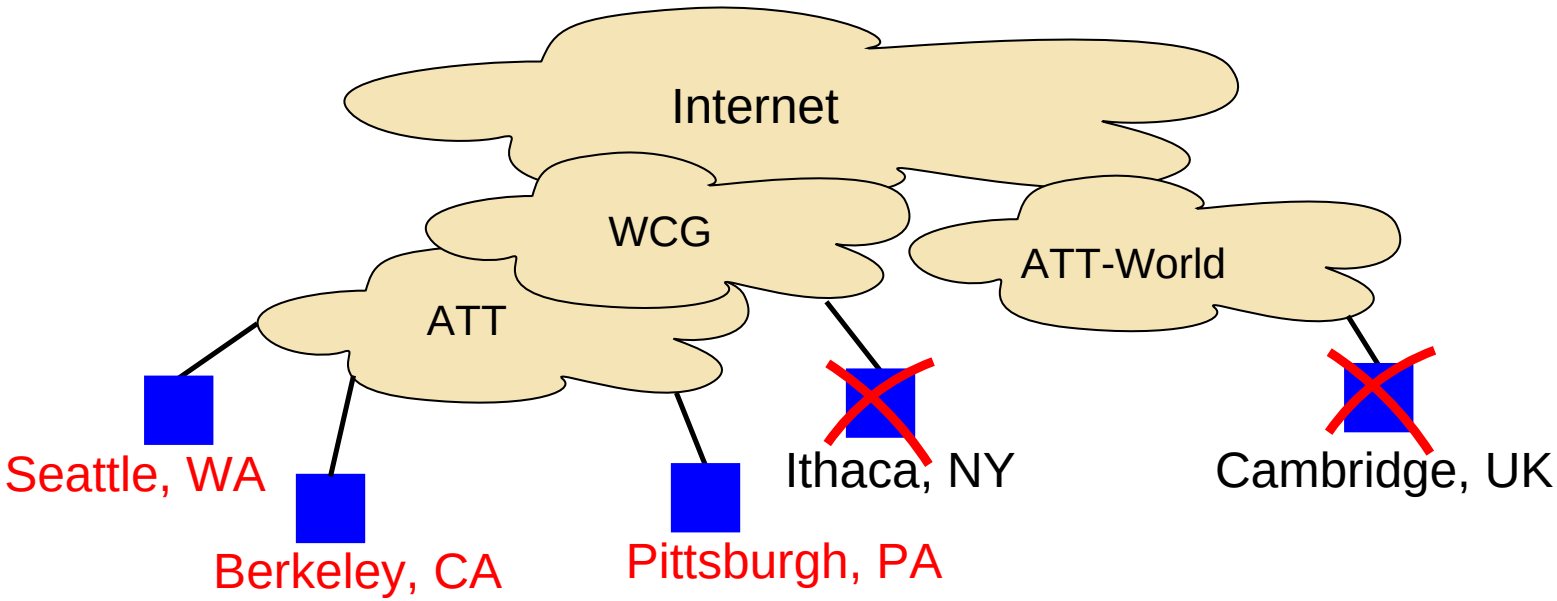
BGP Withdrawal restricted to ATT

Local Routing Event → Faster convergence



Failover

Planned Deployment → Fast Failover



Load Distribution

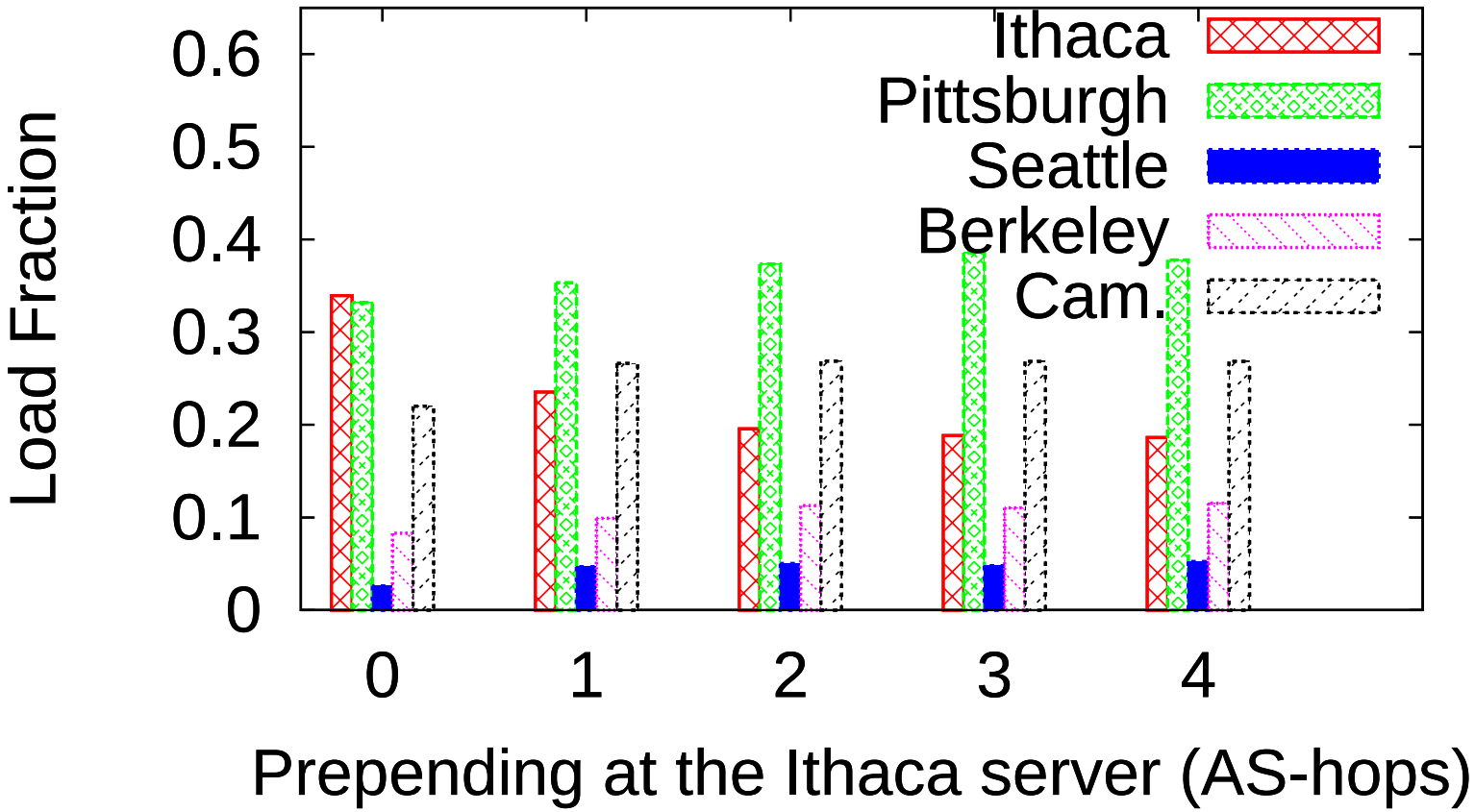
Distribution of client-load across Anycast Servers

- Can operators control this load?
- Used AS-Path Prepending for controlling load

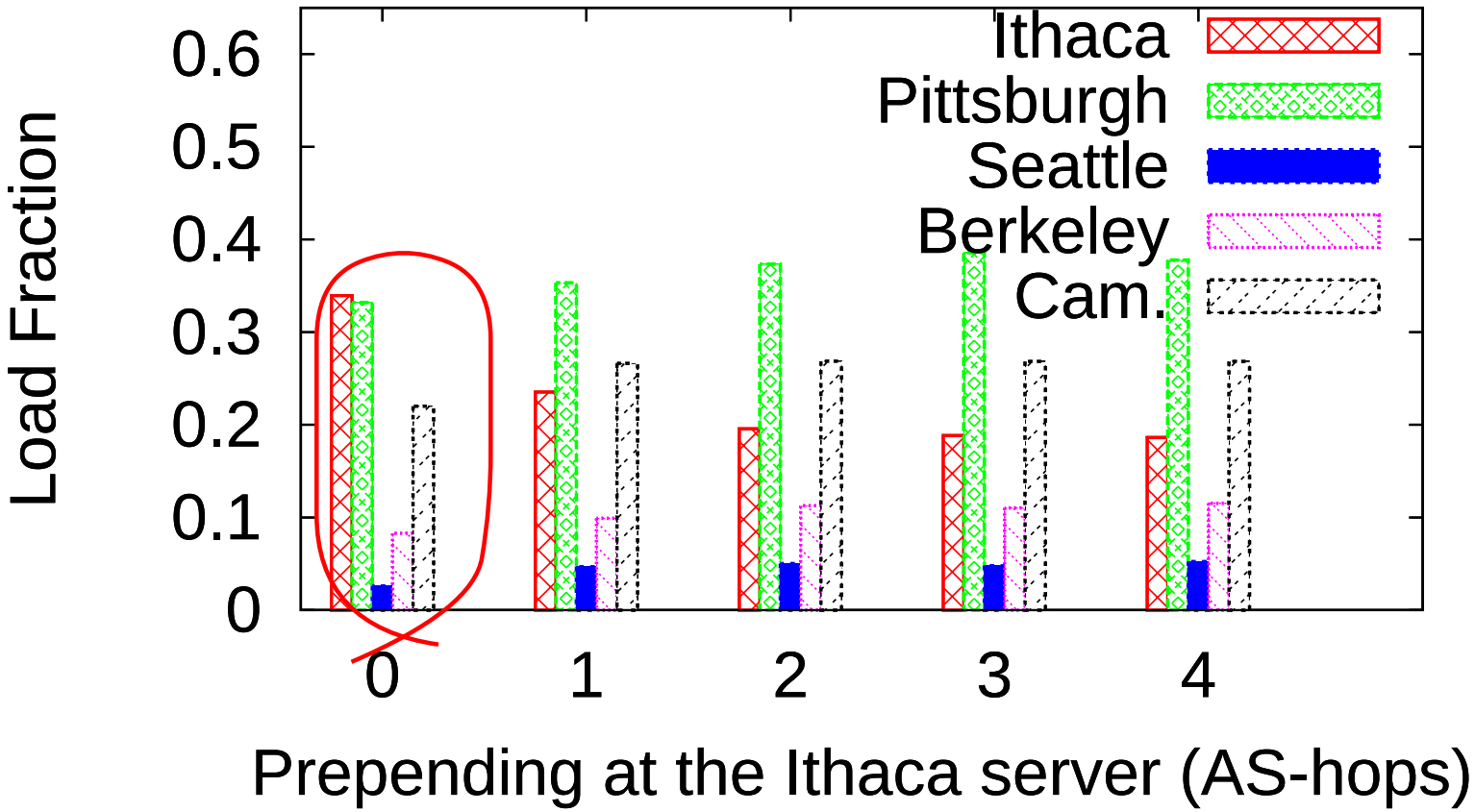
AS-Path Prepending a BGP Advertisement

- Changing the advertisement's AS-Path length
- n-hop Prepending: Add n ASs to the AS-Path

Load Distribution

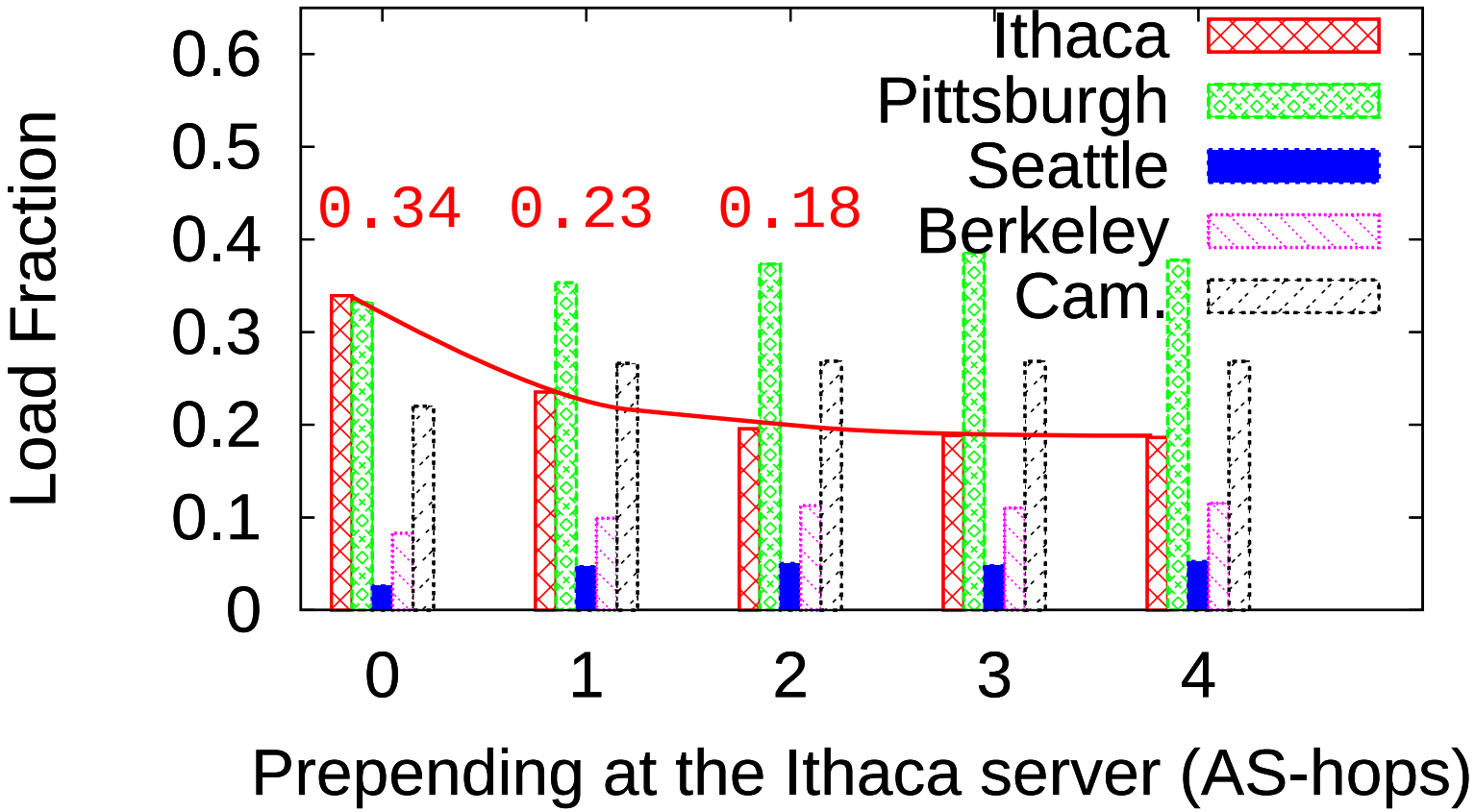


Load Distribution

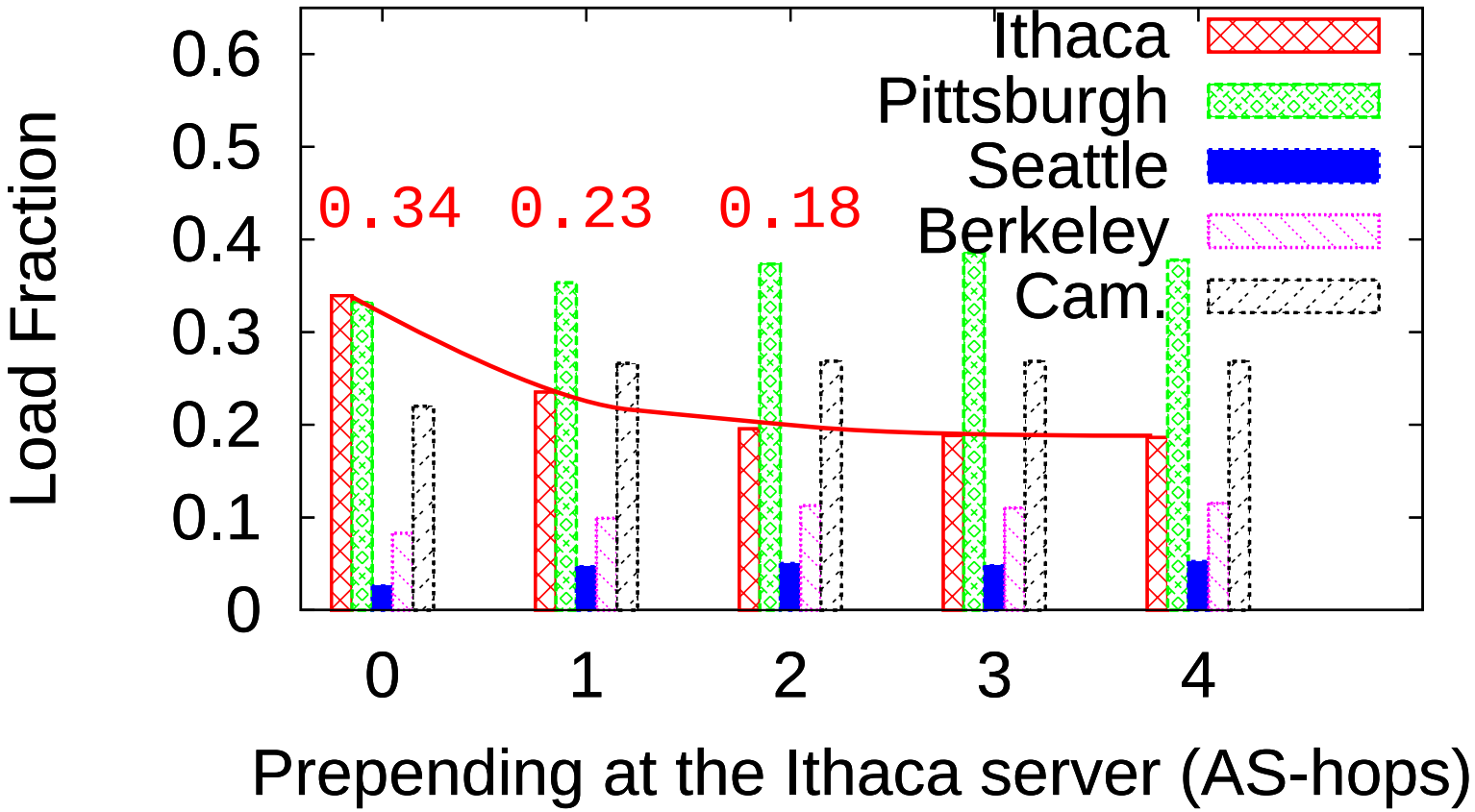


Skewed distribution of clients

Load Distribution



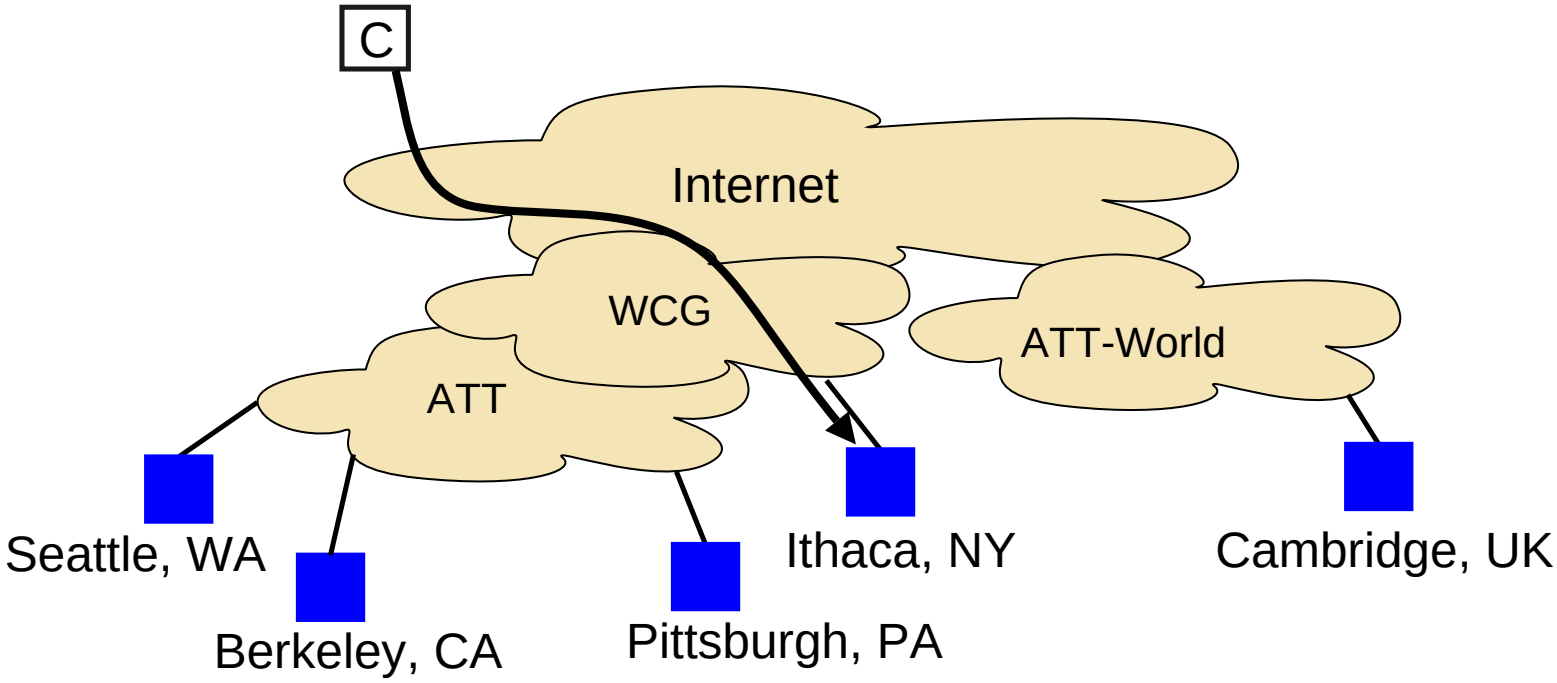
Load Distribution



AS-Path Prepending provides coarse-grained control over client load

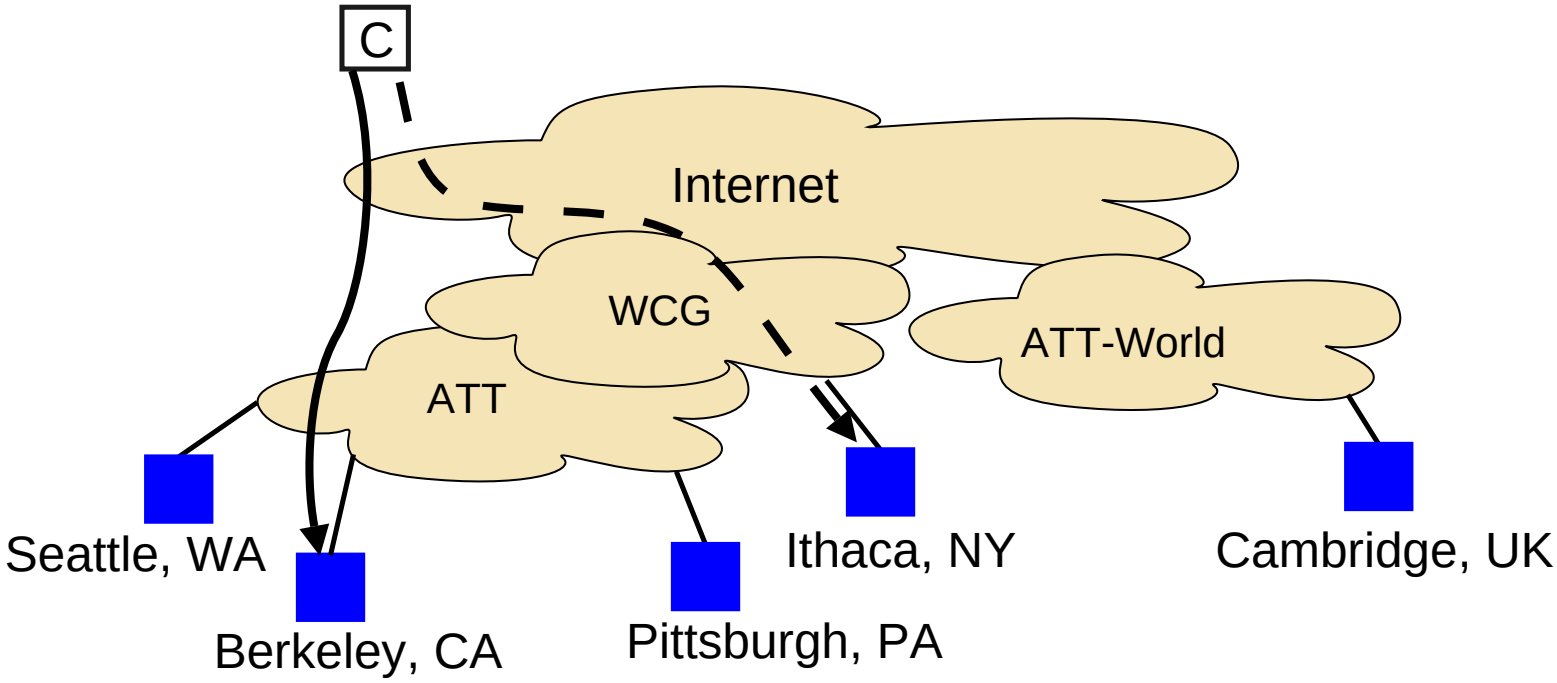
Affinity

IP Anycast is a network-layer service



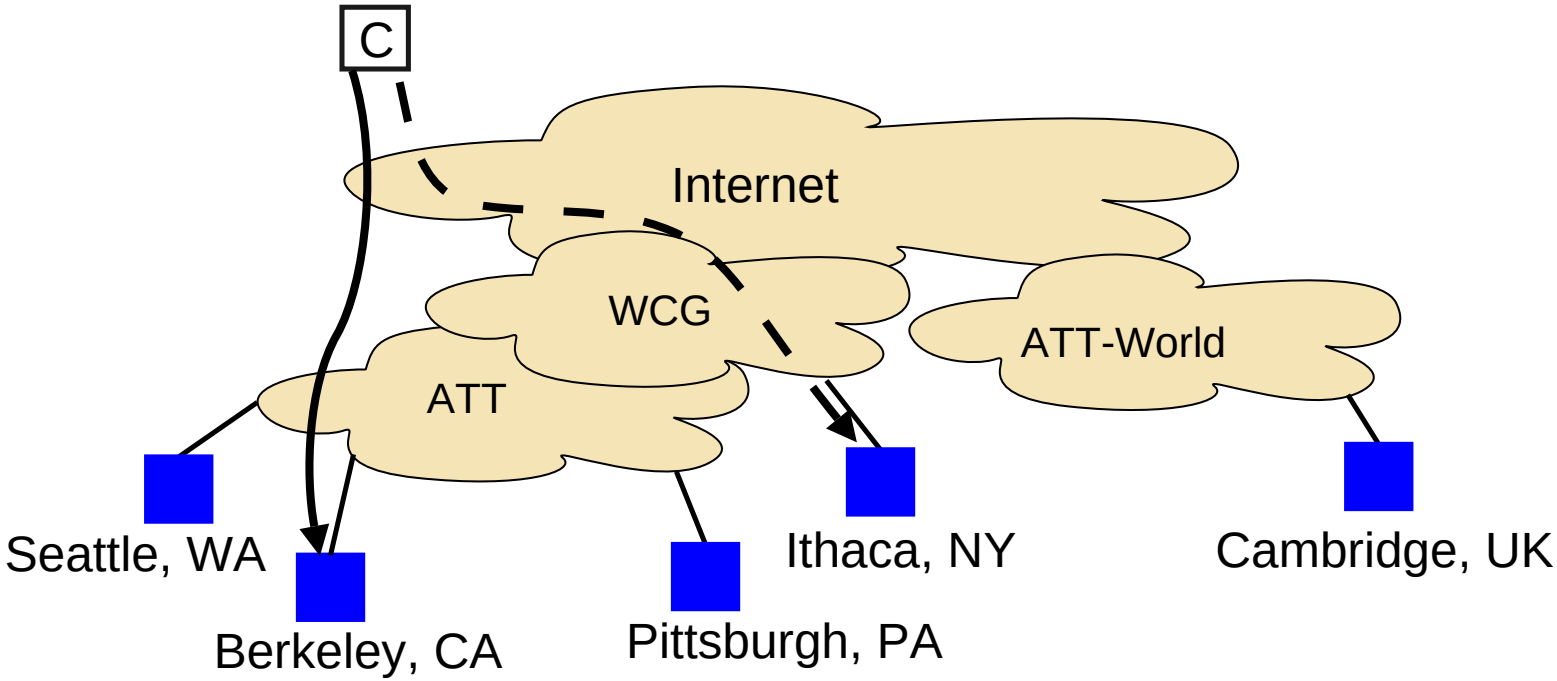
Affinity

Client can flap to a different Anycast Server

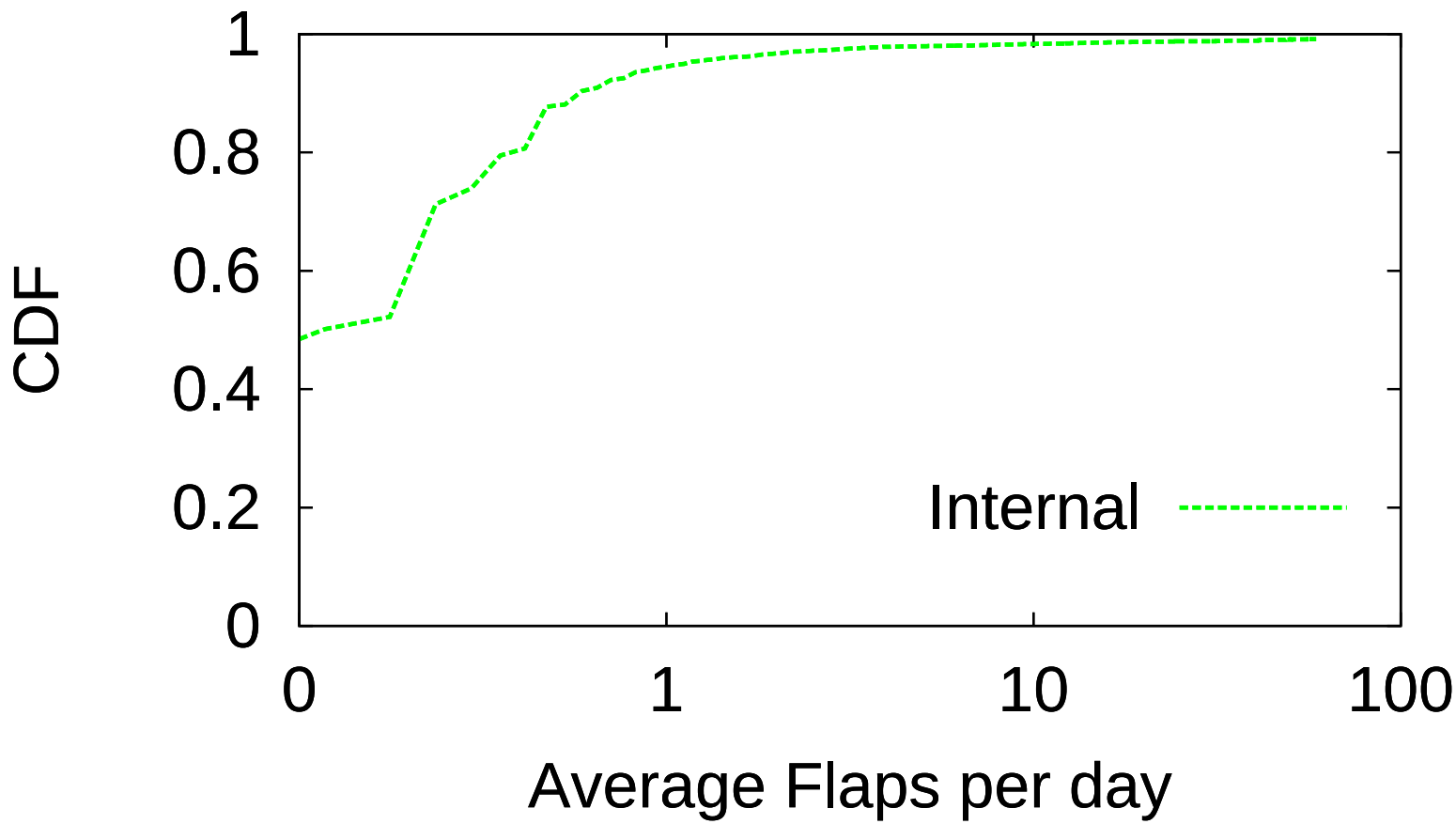


Affinity

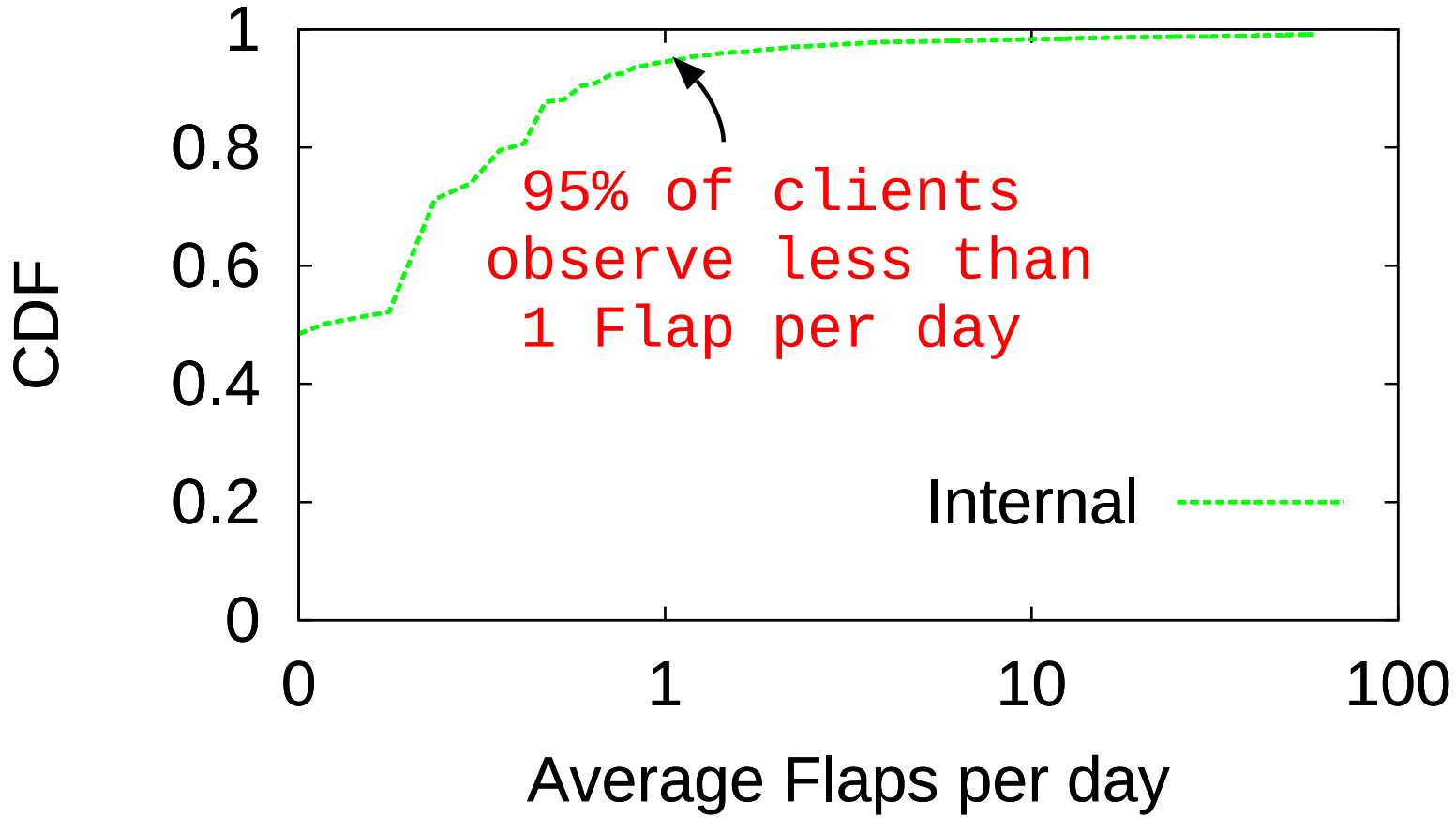
Client can flap to a different Anycast Server
What is the Affinity offered by IP Anycast?



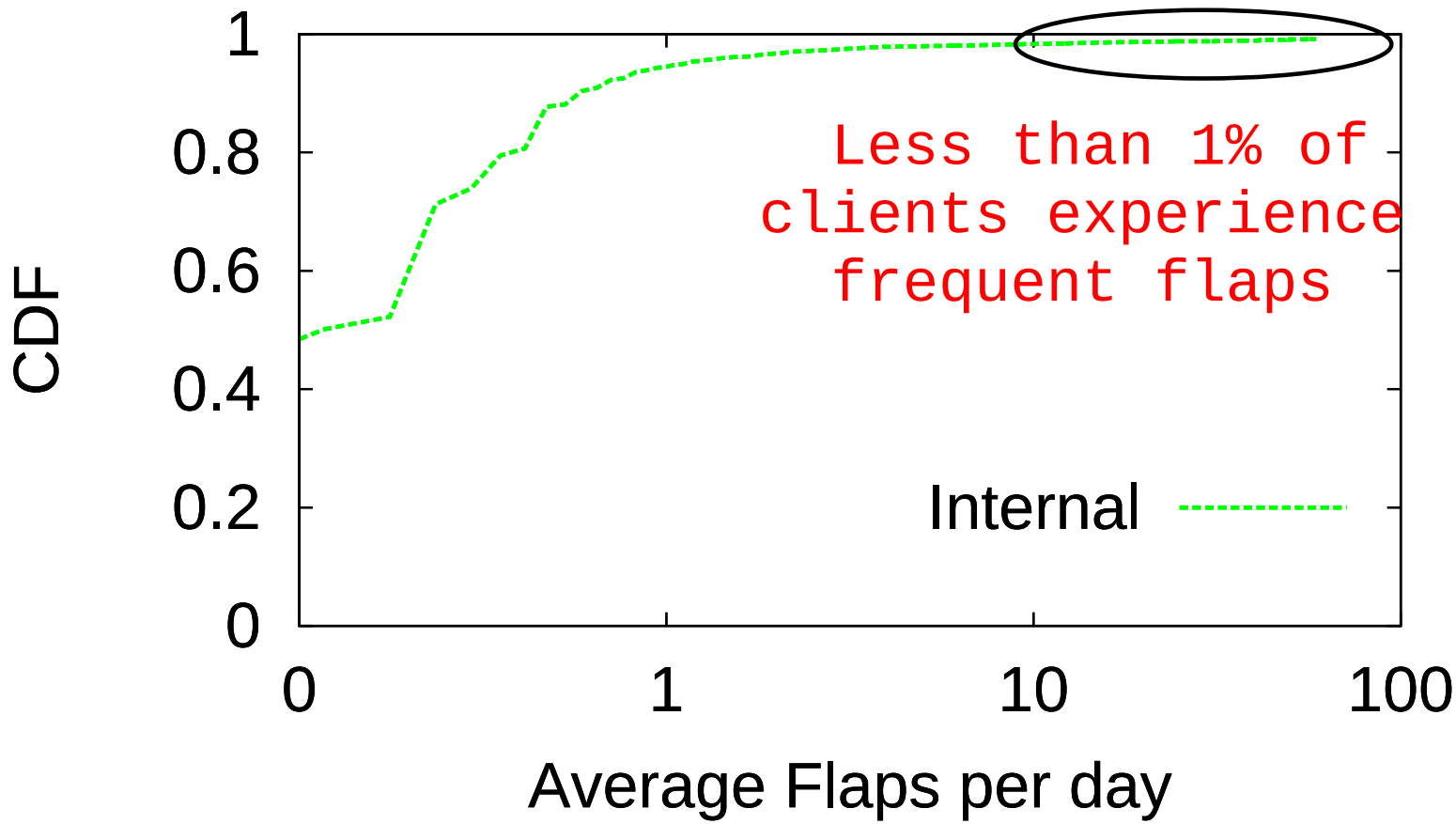
Affinity



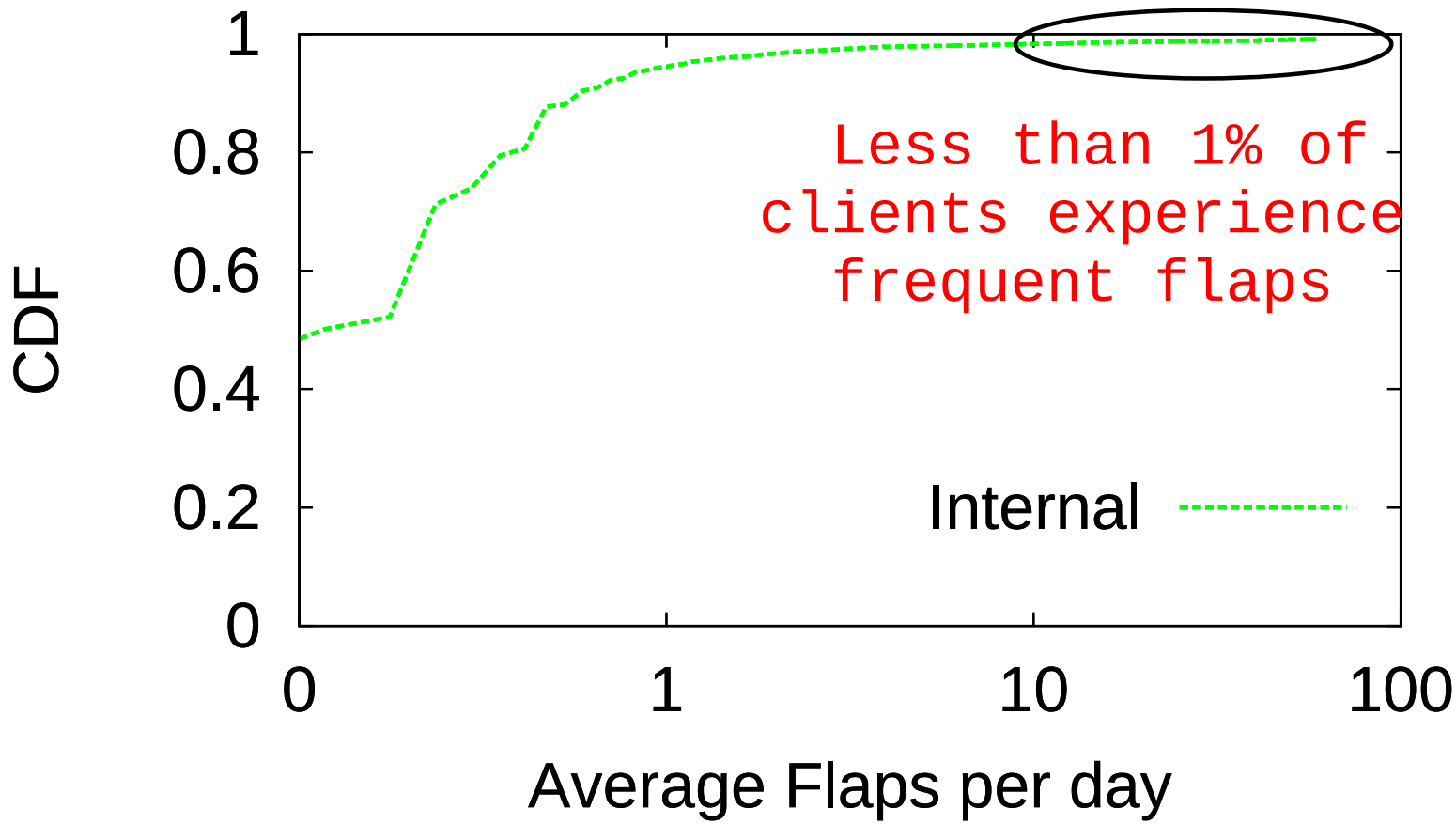
Affinity


1
0.8
0.6
0.4
0.2
0
CDF
0
1
10
100
Average Flaps per day
95% of clients
observe less than
1 Flap per day
Internal

Affinity


1
0.8
0.6
0.4
0.2
0
CDF
0
1
10
100
Average Flaps per day
Less than 1% of clients experience frequent flaps
Internal

Affinity



Frequent flaps can be attributed to load-balancing at the client site

Conclusions

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗ Poor	✗ Slow		

Conclusions

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗ Poor	✗ Slow		
Planned Deployment	✓ Good	✓ Fast		

Due to the planned deployment

Conclusions

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗ Poor	✗ Slow	✗ Skewed	
Planned Deployment	✓ Good	✓ Fast	✗ Manipulatable	

BGP Traffic Engineering techniques

Conclusions

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗ Poor	✗ Slow	✗ Skewed	✓ Good*
Planned Deployment	✓ Good	✓ Fast	✗ Manipulatable	✓ Good*

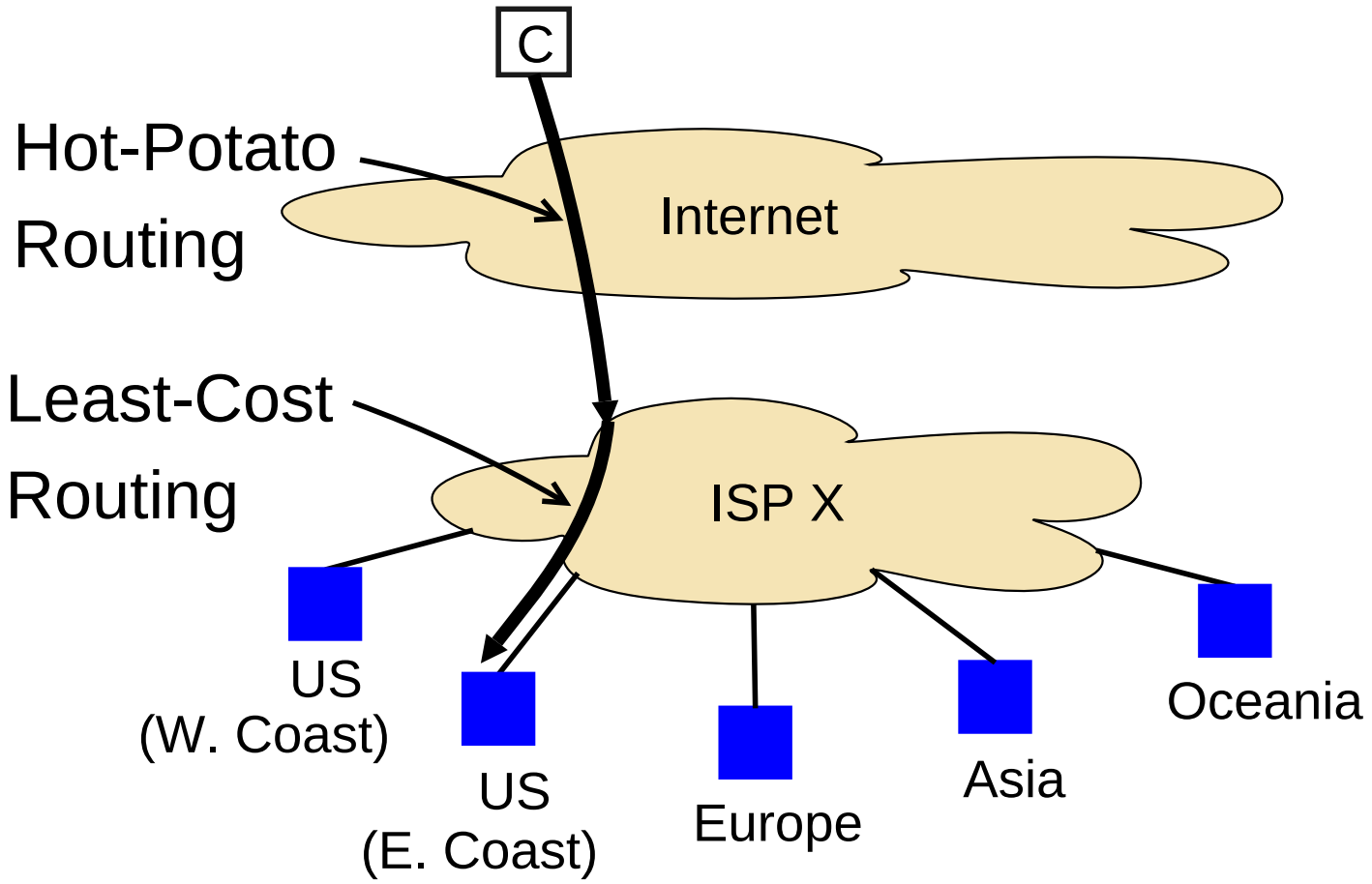
Conclusions

Property / IP Anycast	Proximity	Failover	Load	Affinity
Ad-Hoc Deployment	✗ Poor	✗ Slow	✗ Skewed	✓ Good*
Planned Deployment	✓ Good	✓ Fast	✗ Manipulatable	✓ Good*

Traces

http://pias.gforge.cis.cornell.edu/measure.php

Alleviating Poor Proximity



Only Proximity measurements for the External Deployments

Probes to External Deployments

Cannot determine the identity of the responding server