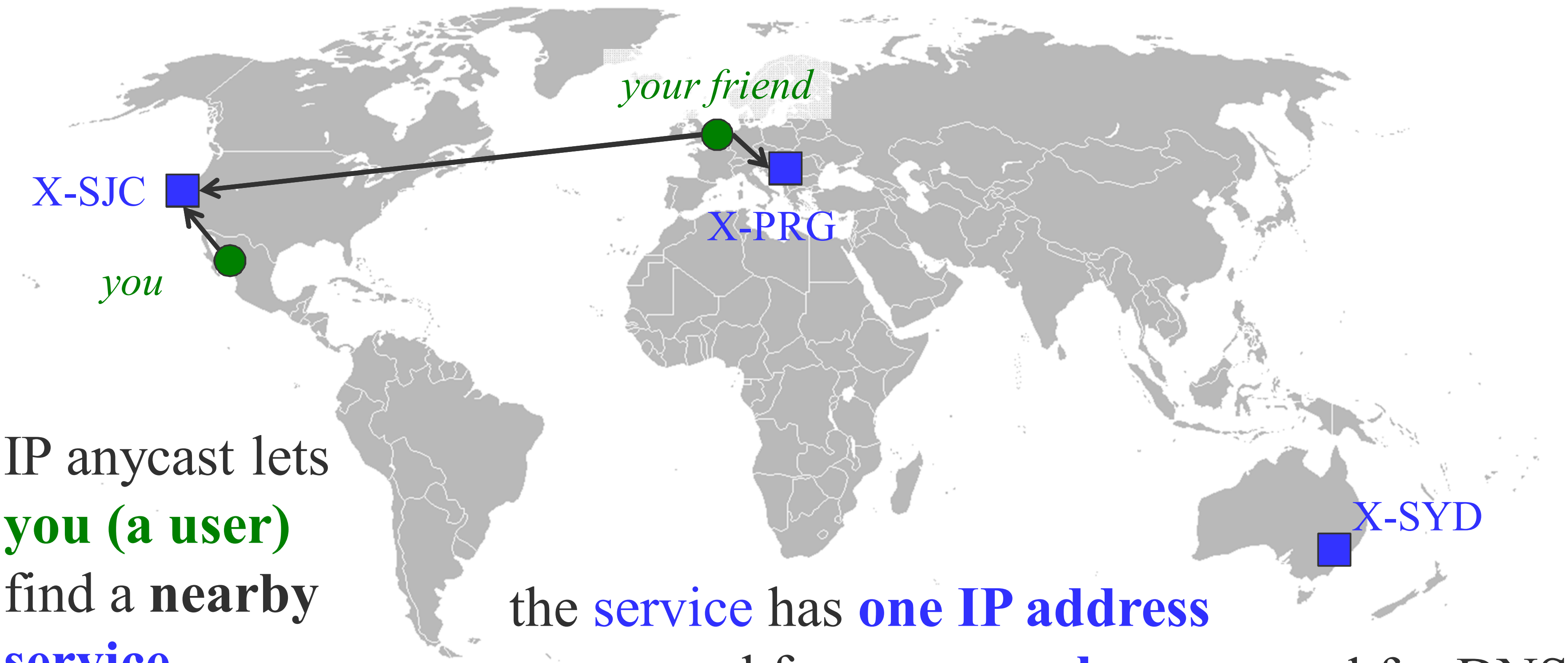
Anycast Latency: How Many Sites are Enough?

Ricardo de O. Schmidt[1], Jan Harm Kuipers[1], and *John Heidemann*[2]

[1]University of Twente [2]USC/ISI

DNS-OARC, Dallas, 2016-10-16

IP Anycast: Why?

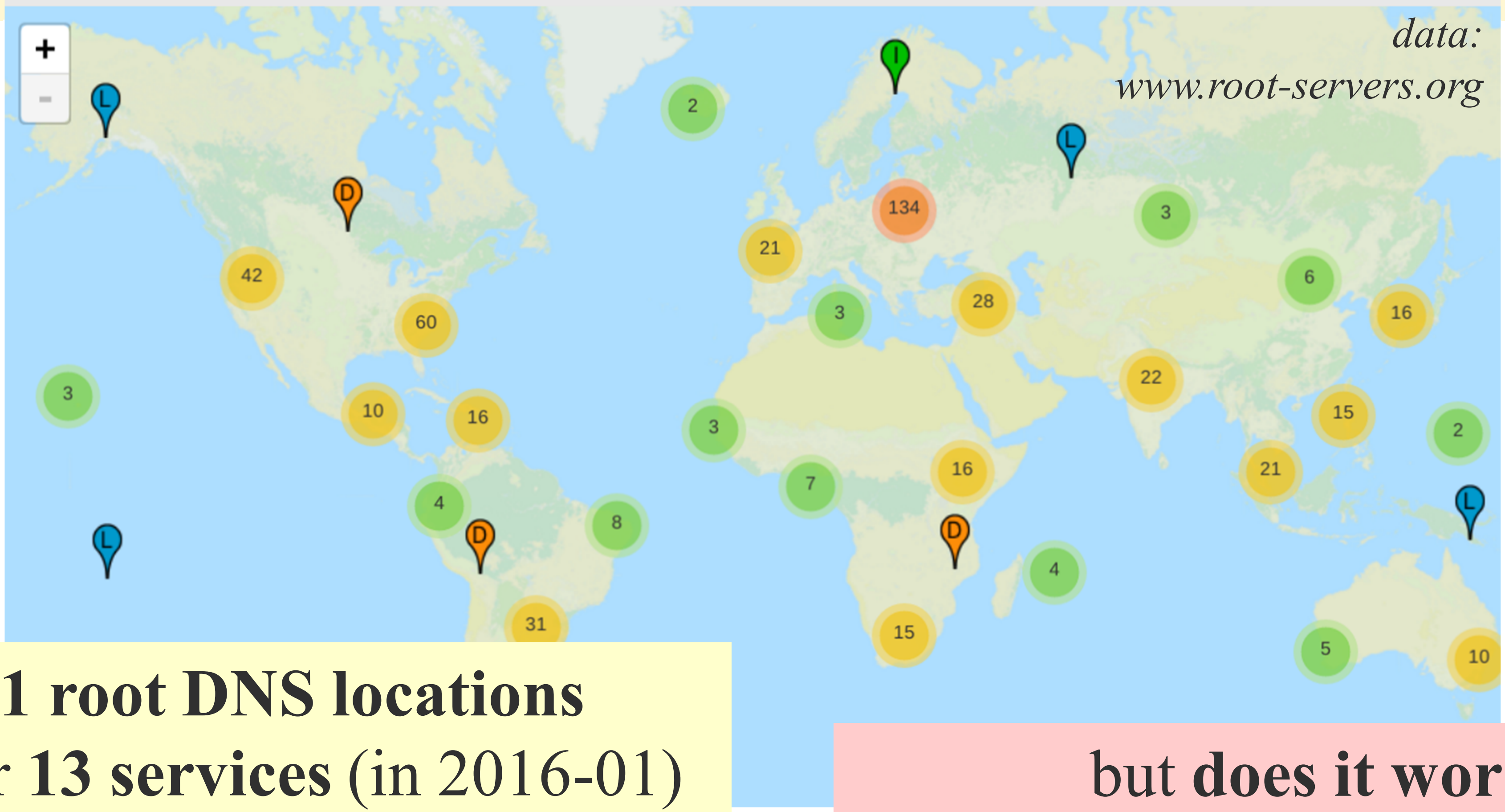


IP anycast lets **you (a user)** find a **nearby** **service**

the service has **one IP address** announced from **many places**; **BGP** routing finds “nearby”

used for DNS and some CDNs

Does Anycast Work?



data: www.root-servers.org

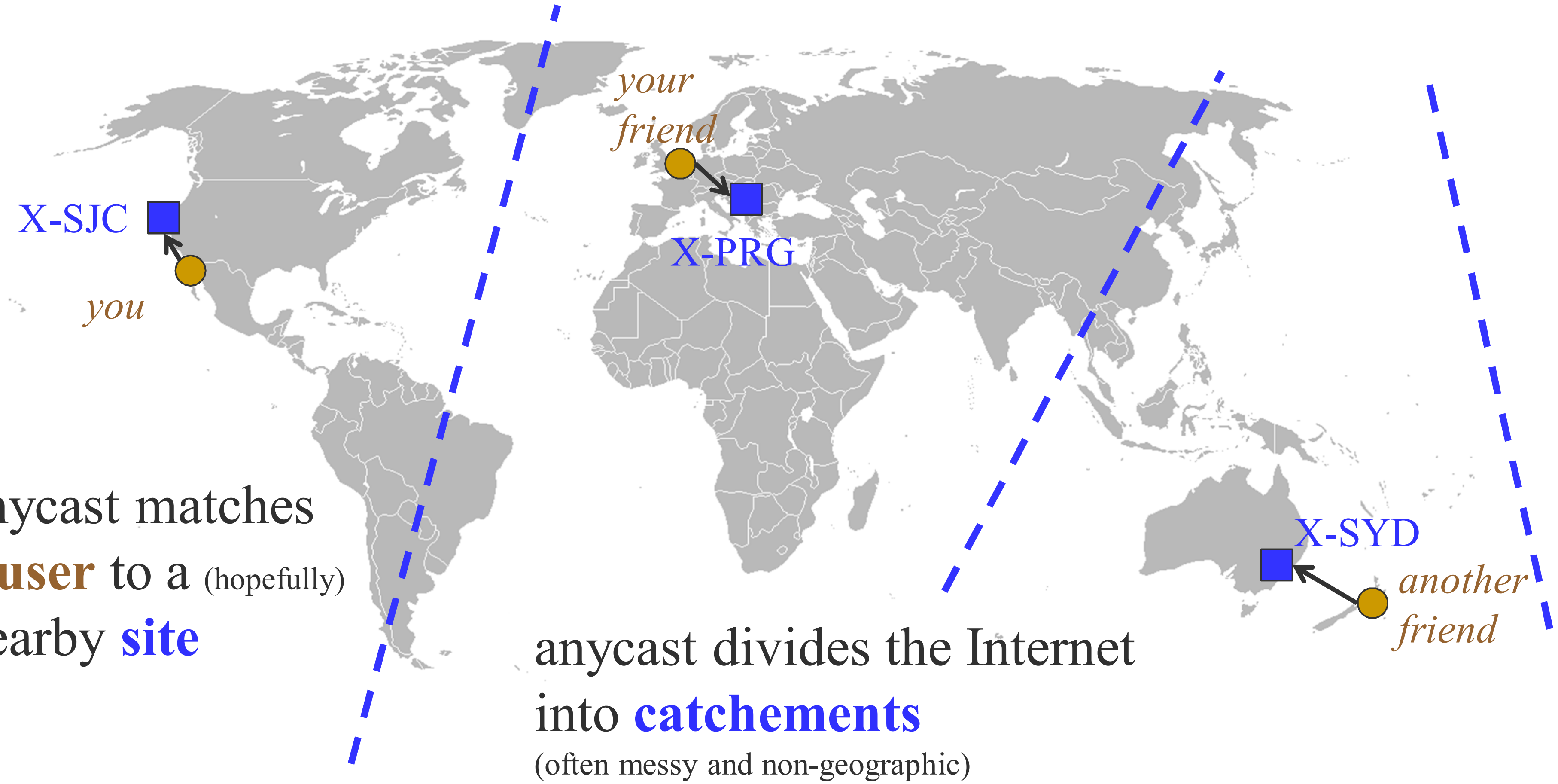
561 root DNS locations
for **13 services** (in 2016-01)
large capex and opx

but **does it work?**
what does "work" mean?

What does "Anycast Works" Mean?

- here: latency from user to service
  - (other reasons: DDoS defense, collaboration)
- questions
  - does anycast give good *absolute* latency?
  - does *instance location* matter?
  - does *BGP routing policy* interfere?
  - what about *the tail?*
  - *how many instances* do you need?
- contribution: answering these with measurement

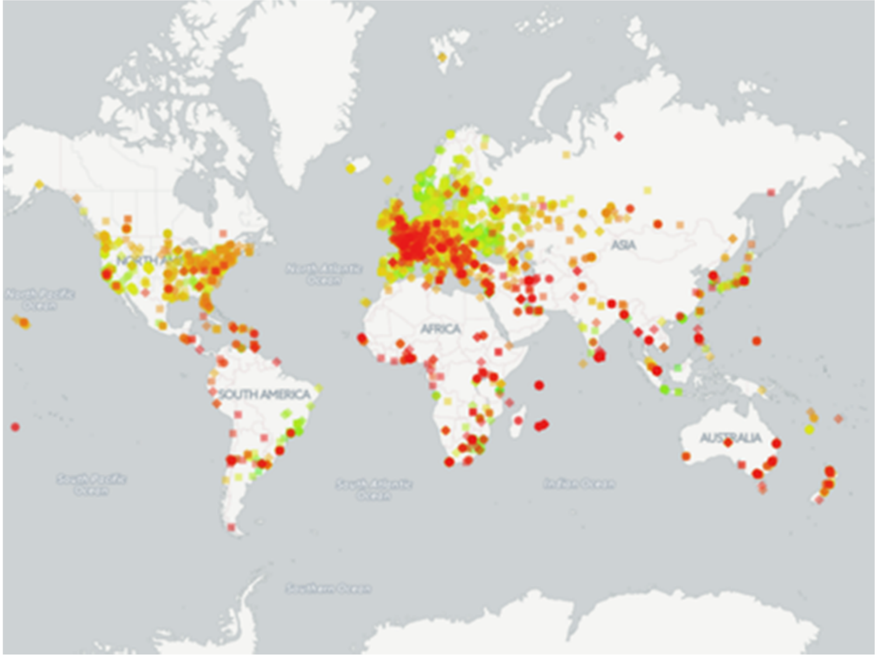
The Anycast Ideal



anycast matches a **user** to a (hopefully) nearby **site**

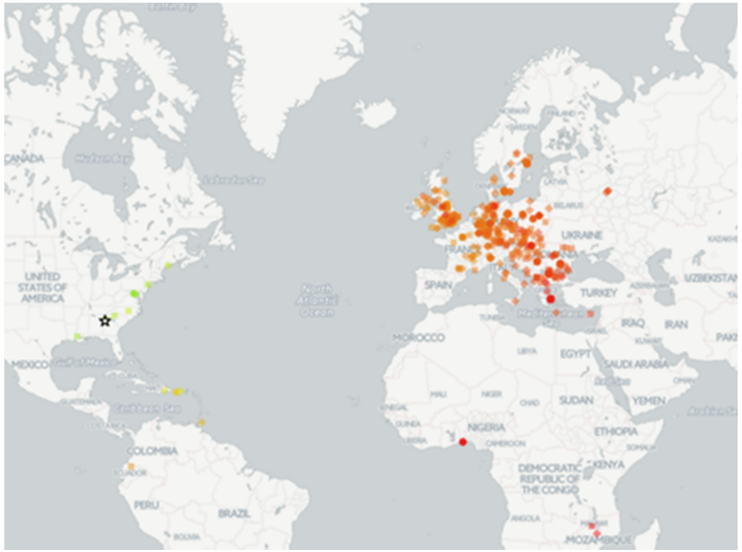
anycast divides the Internet into **catchements**
(often messy and non-geographic)

Anycast’s Reality



latency is often high (red: >200ms)

(Data: Ray Bellis, mid-2015 F-Root, from https://labs.ripe.net/Members/ray_bellis/researching-f-root-anycast-placement-using-ripe-atlas)

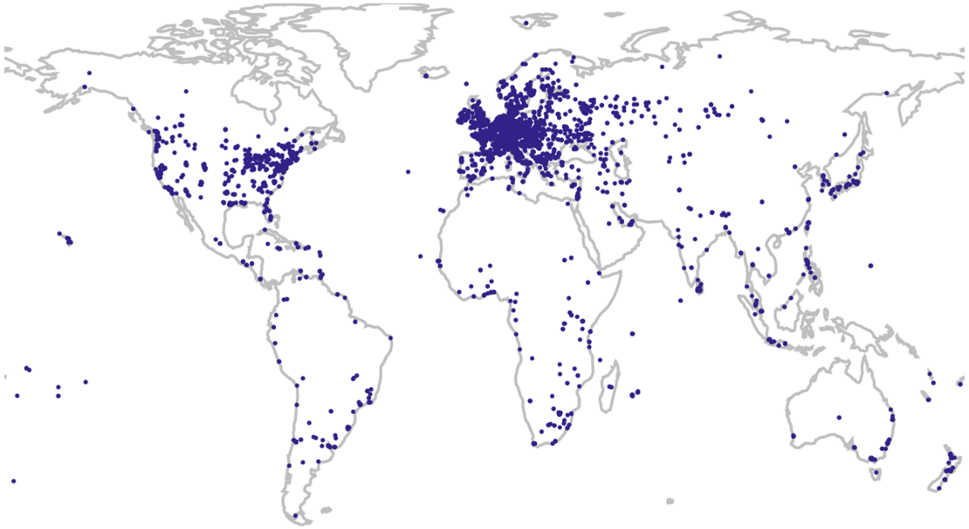


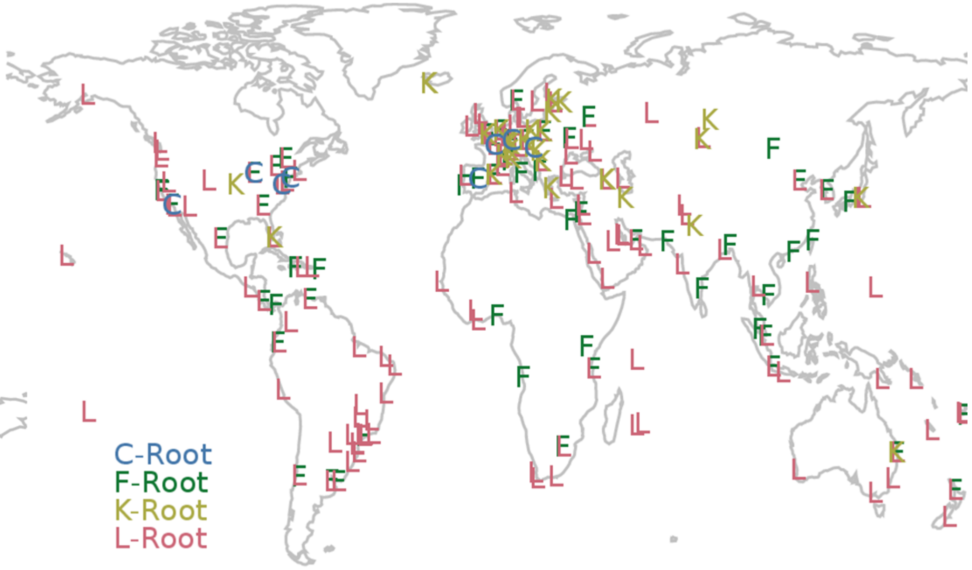
because routing is hard

(*correct* is easy,
but *performance* bugs
are obscure and
easy to overlook)

What We Measure

Sources:
~7000 RIPE Atlas instances sending pings and traceroutes to estimate latency

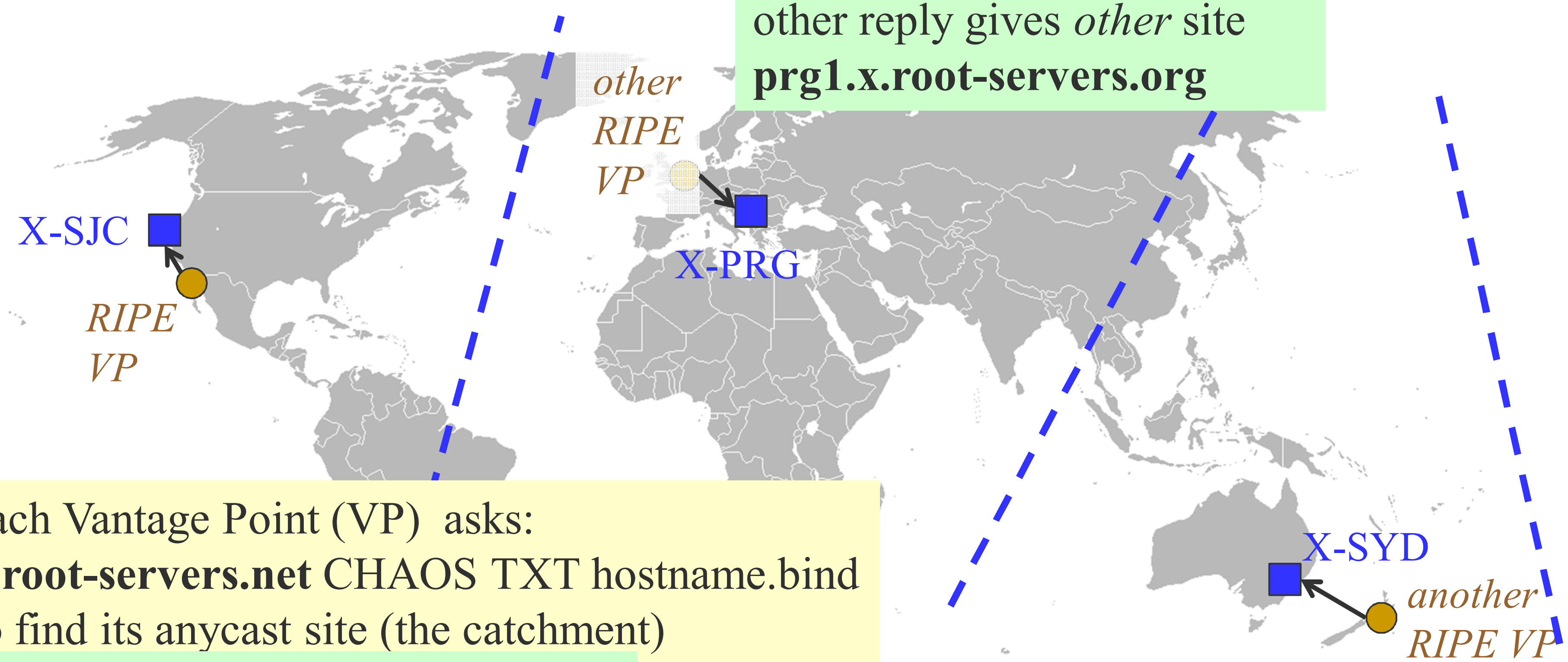




Destinations: 4 DNS Root Letters

letter	sites (local)	date
C	8 (0)	2015-09
F	58 (53)	2015-12
K	33 (14)	2015-11
NK*	36 (1)	2016-04
L	144 (0)	2015-12

Measurement Step 1: Your *Current* Site



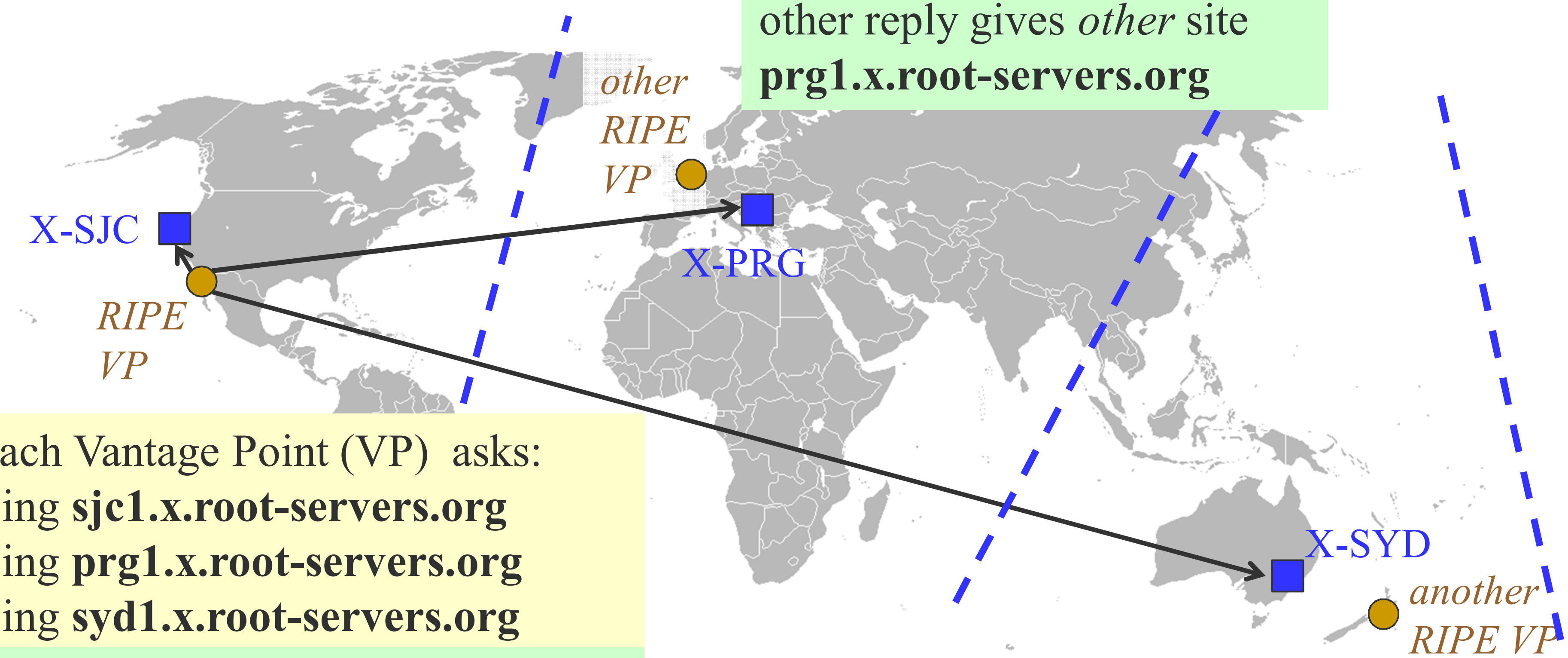
other reply gives *other* site
prg1.x.root-servers.org

each Vantage Point (VP) asks:
x.root-servers.net CHAOS TXT hostname.bind
to find its anycast site (the catchment)

reply gives *that VP's* anycast site:
sjc1.x.root-servers.org

or *another*
syd1.x.root-servers.org

Measurement Step 2: Distance to *All Possible* Sites



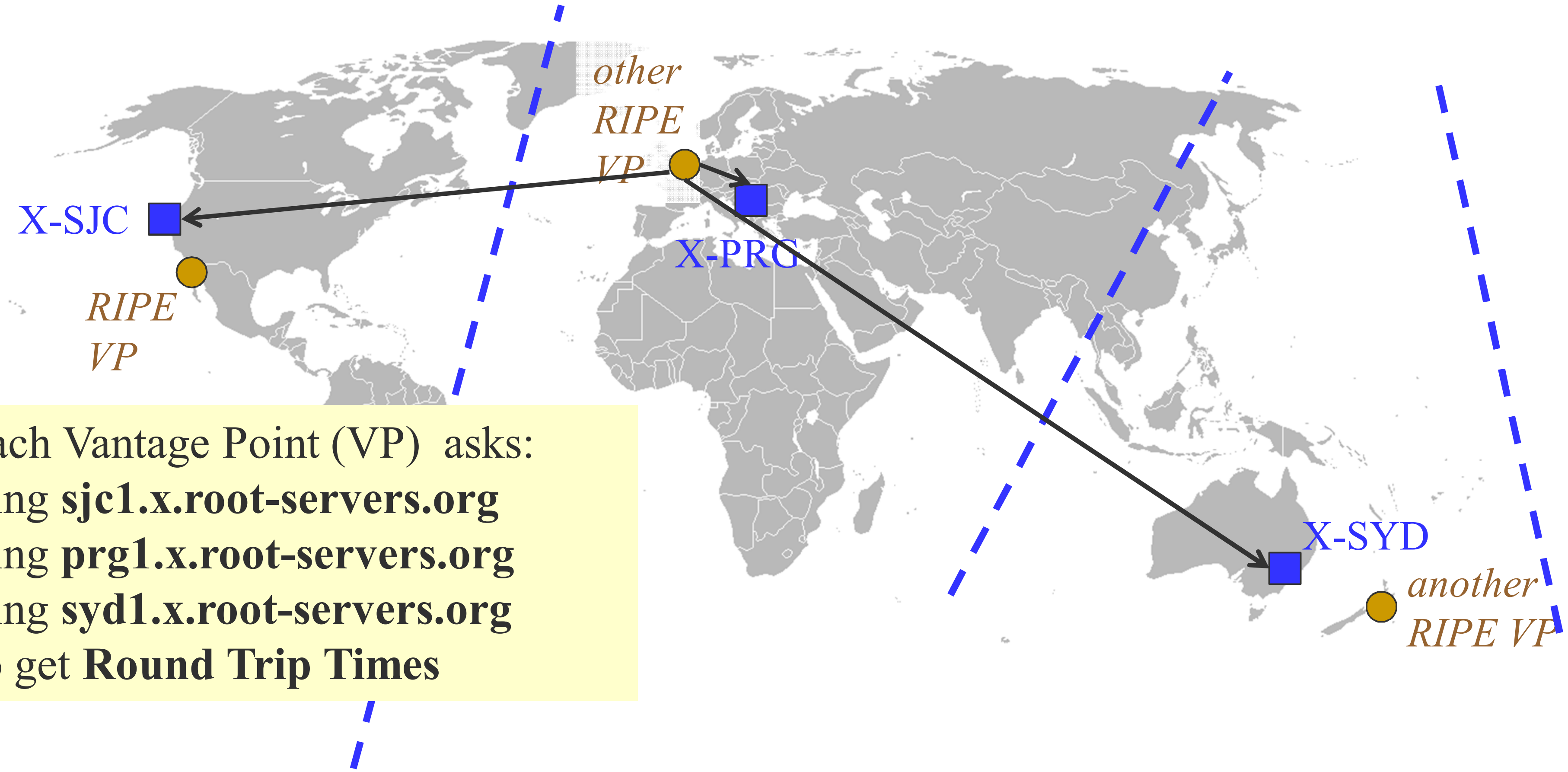
other reply gives *other* site
prg1.x.root-servers.org

each Vantage Point (VP) asks:
ping **sjc1.x.root-servers.org**
ping **prg1.x.root-servers.org**
ping **syd1.x.root-servers.org**

reply gives *that VP's* anycast site:
sjc1.x.root-servers.org

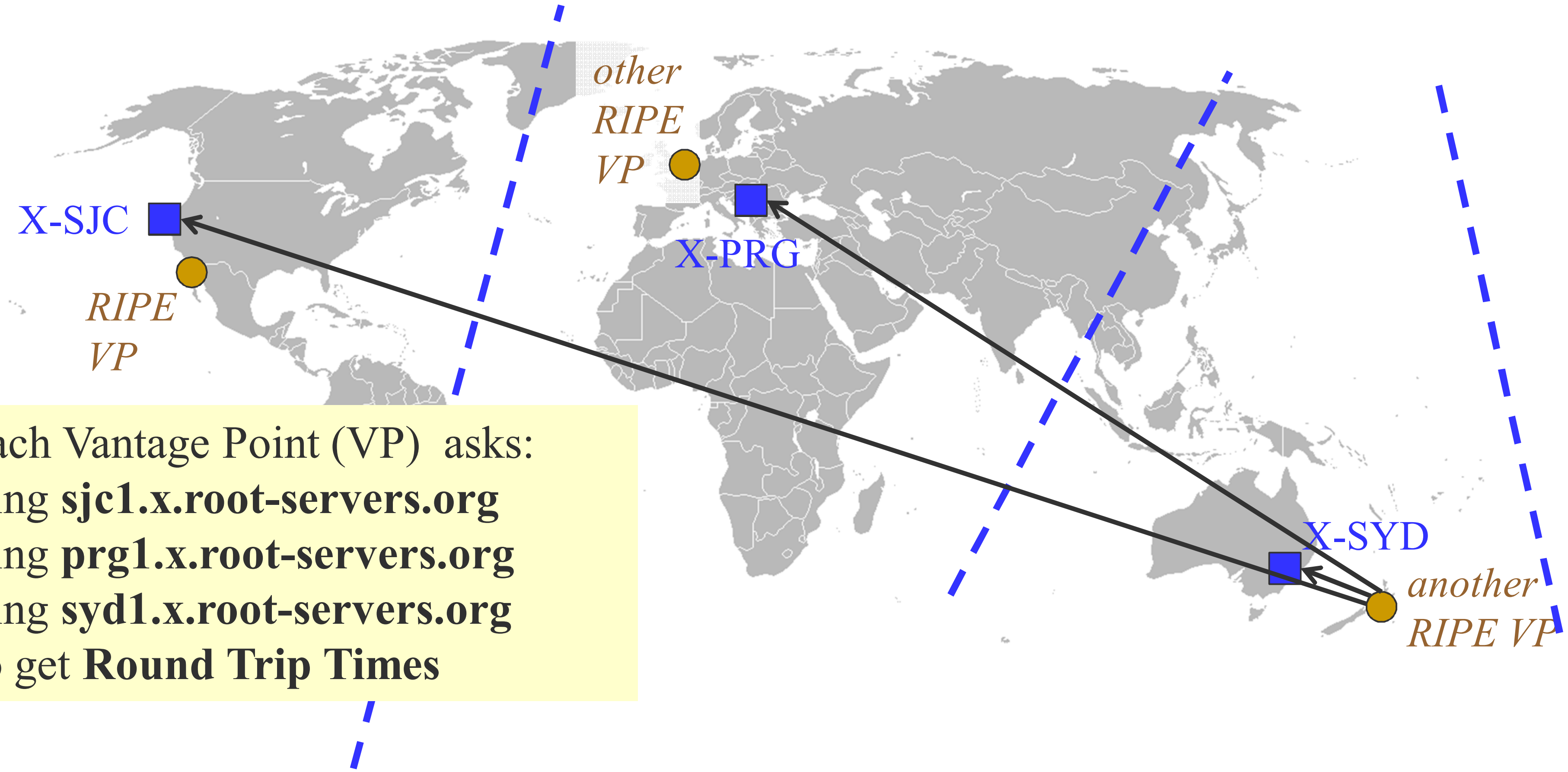
or *another*
syd1.x.root-servers.org

Measurement Step 2:
Distance to *All Possible* Sites



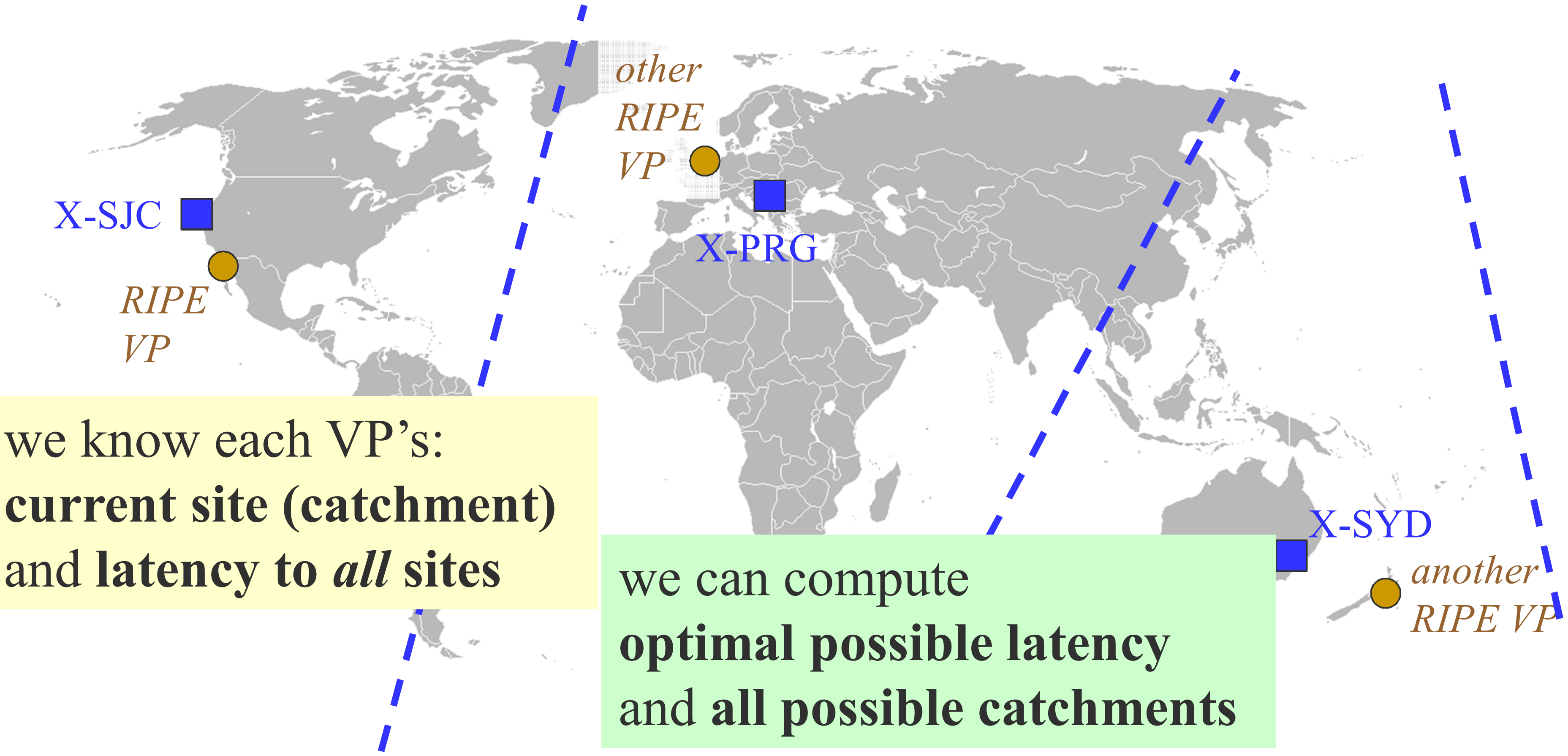
each Vantage Point (VP) asks:
ping **sjc1.x.root-servers.org**
ping **prg1.x.root-servers.org**
ping **syd1.x.root-servers.org**
to get **Round Trip Times**

Measurement Step 2: Distance to *All Possible* Sites



each Vantage Point (VP) asks:
ping **sjc1.x.root-servers.org**
ping **prg1.x.root-servers.org**
ping **syd1.x.root-servers.org**
to get **Round Trip Times**

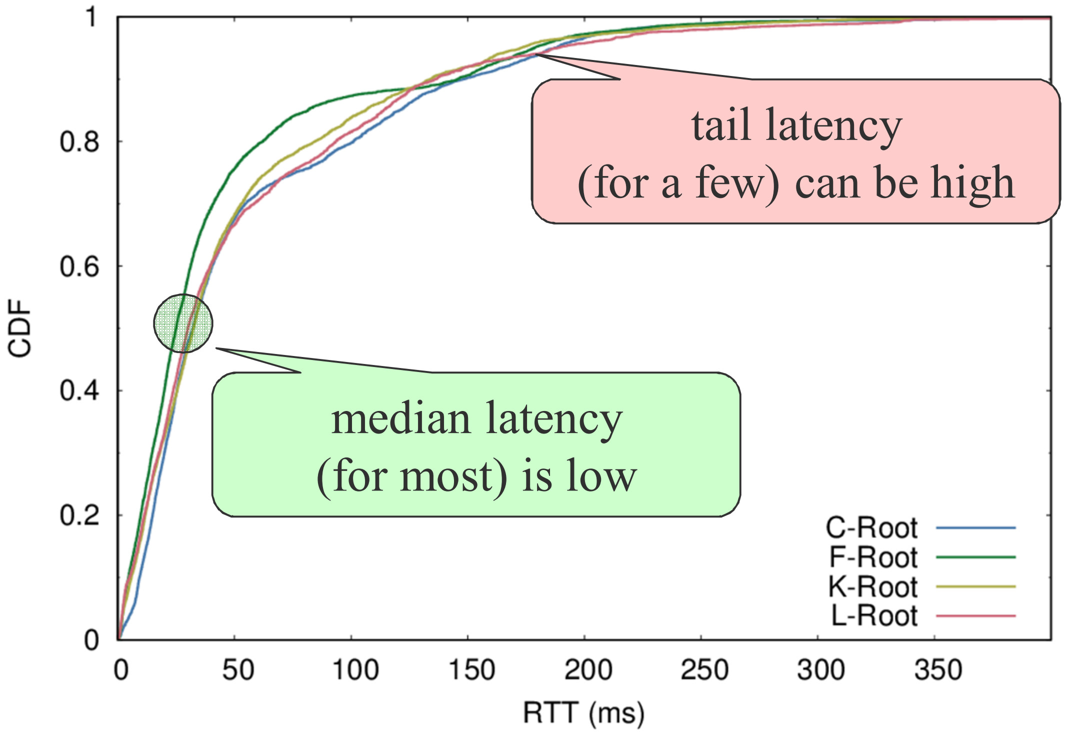
Measurement Outcome



we know each VP's:
current site (catchment)
and **latency to *all* sites**

we can compute
optimal possible latency
and **all possible catchments**

Is Absolute Latency Good?



tail latency
(for a few) can be high

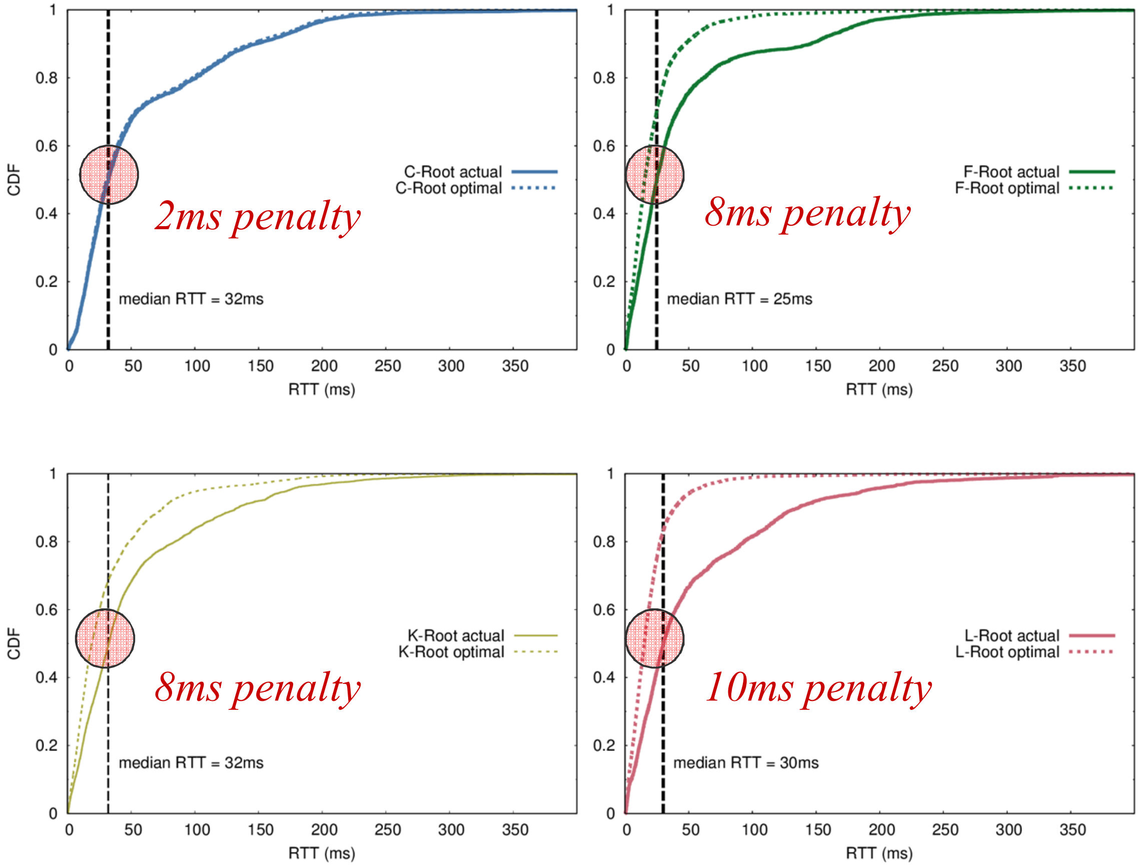
median latency
(for most) is low

yes (usually)

median latency is low

F: 25ms
L: 30ms
C and K: 32 ms

Is Absolute Latency *Optimal*?



2ms penalty

8ms penalty

8ms penalty

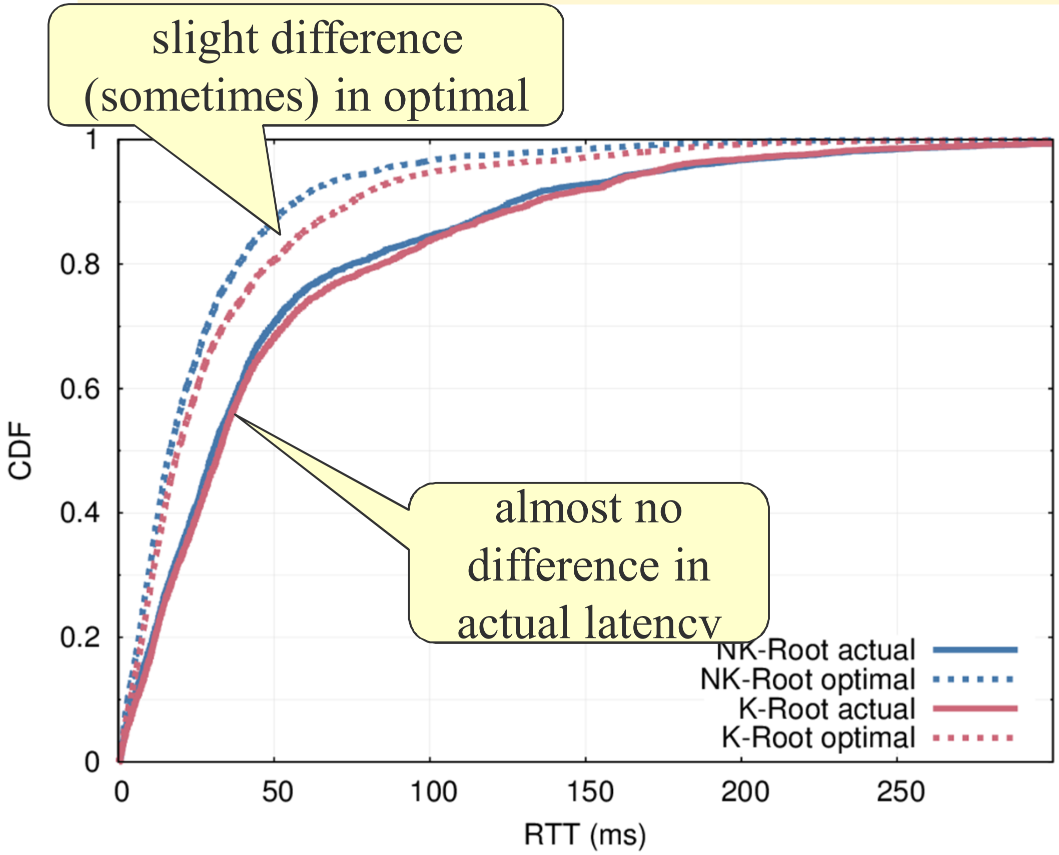
10ms penalty

yes (nearly)

actual latency within 2-10ms of *best possible* (at median)

(compare solid vs. dotted lines of same color)

Does Routing Policy Interfere?



slight difference (sometimes) in optimal

almost no difference in actual latency

local policy sites only serve *their* AS

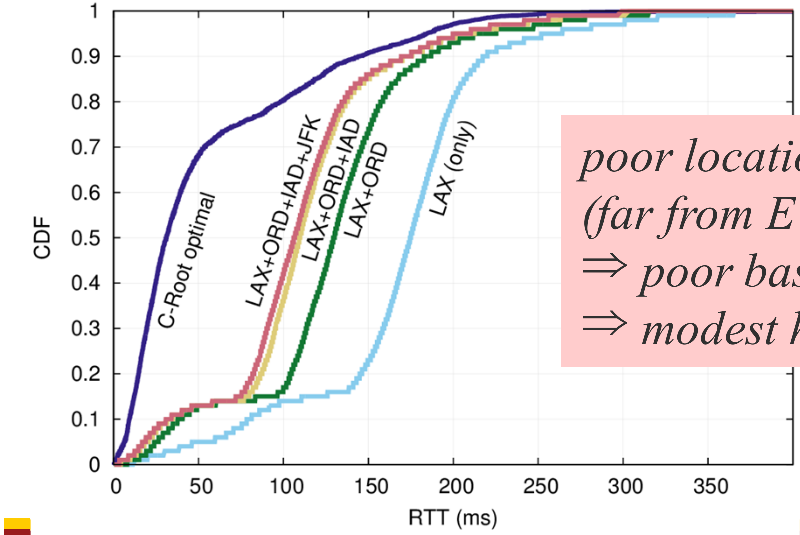
not much penalty

(compare
pink: K in 2015, with half local nodes vs.
blue: NK in 2016 with all but one global)

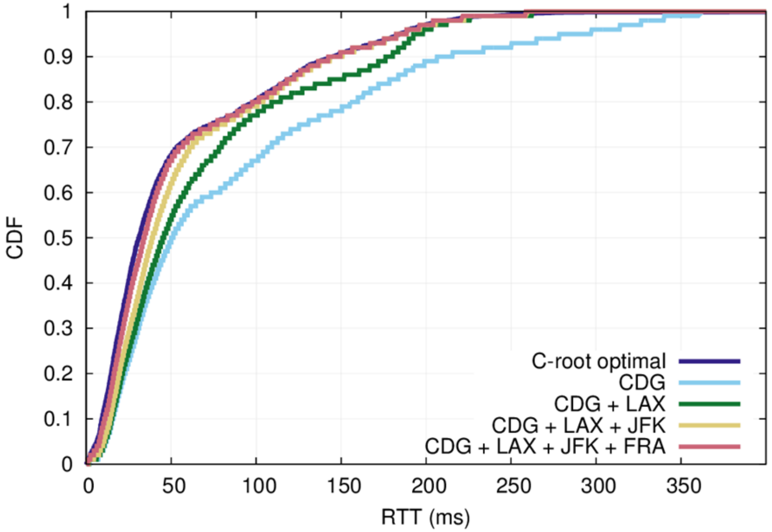
Does Location Matter?

yes (a lot!)

simulate 1-4-node anycast services
experimental measurements from C-Root
measure from RIPE Atlas (bias to Europe)



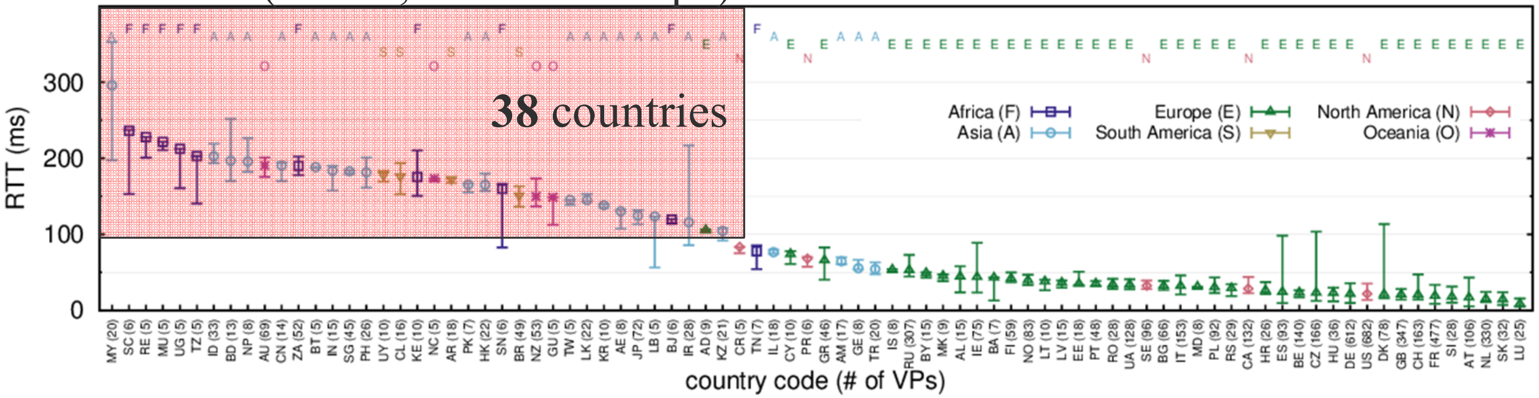
poor locations
(far from EU)
⇒ poor base latency w/1
⇒ modest help w/2-4



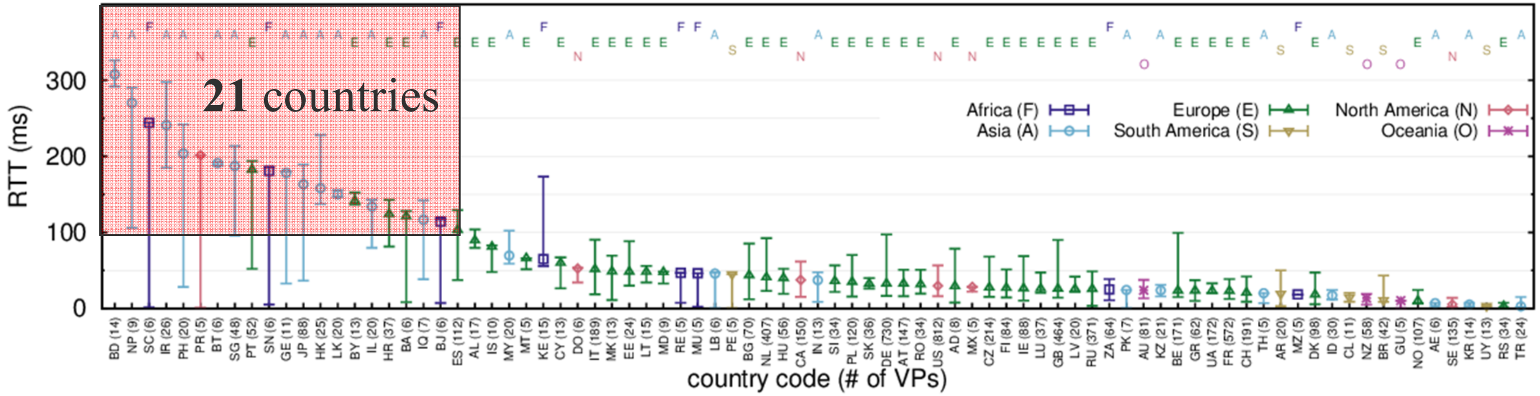
optimal locations
(geographically dispersed)
⇒ 1 is good
⇒ 2-4 help "pull in tail"

What About the Tail?

C-Root (6 sites, US and Europe)



L-Root (144 sites, everywhere)



show median and quartile RTTs per country

if median RTT >100ms is *high latency*:
more sites
=> fewer high latency countries

more sites helps the tail (21 vs. 38)

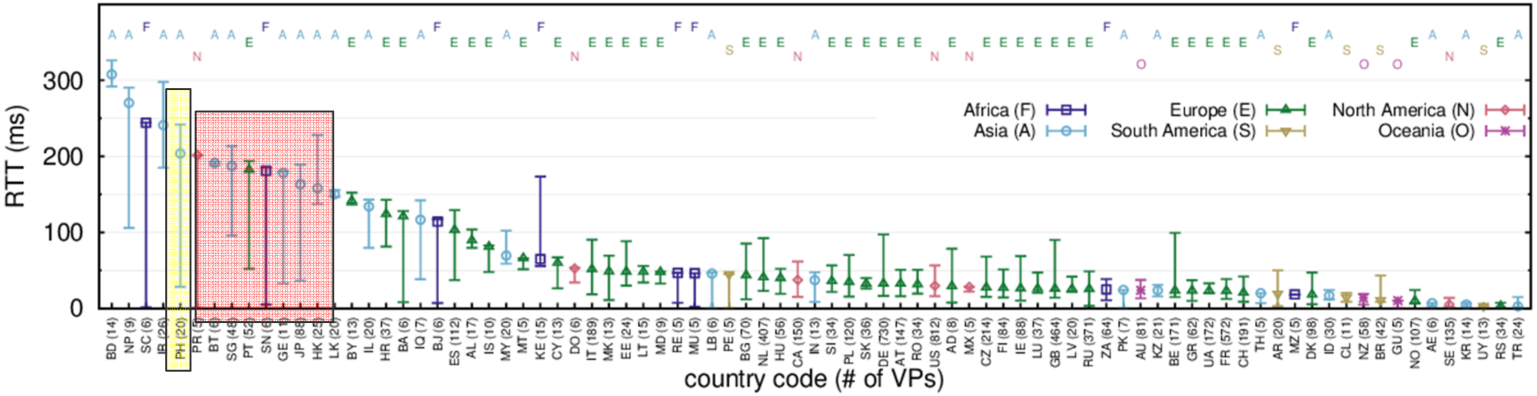
Routing in the Tail?

routing shows *large* variation (compare median vs. 25%ile)

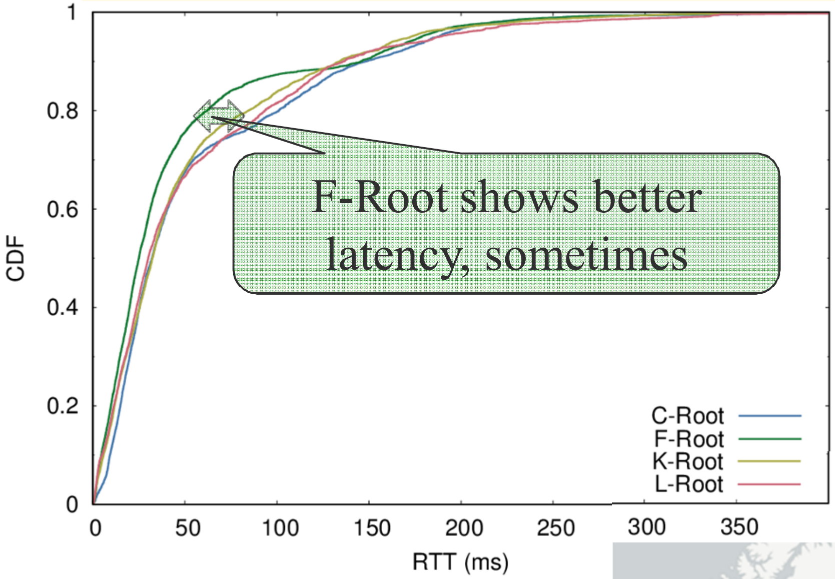
reason: site in country doesn’t reach all country ISPs

Phillipines (PH):
20 VPs
only 7 reach site in PH
13 to US and Australia

L-Root (144 sites, everywhere)



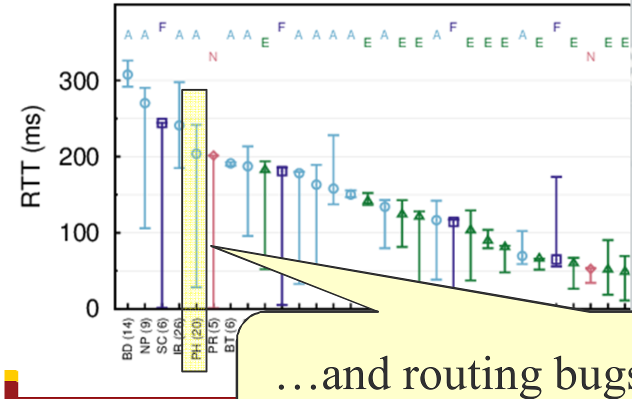
What Matters Most?

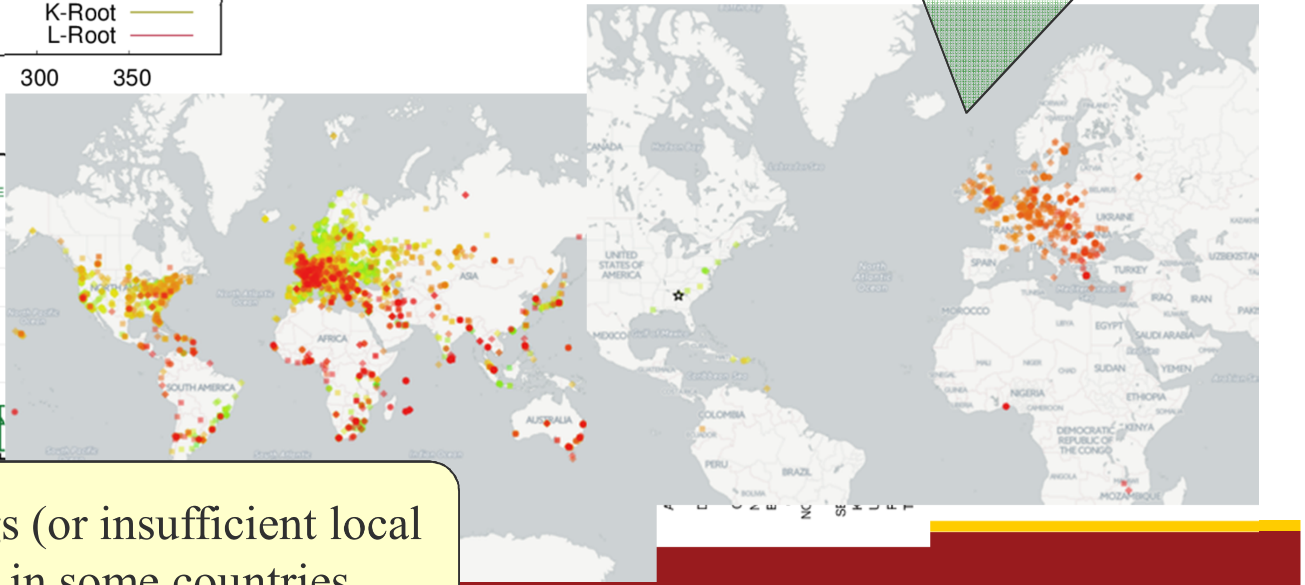


F-Root shows better latency, sometimes

fixing routing bugs!

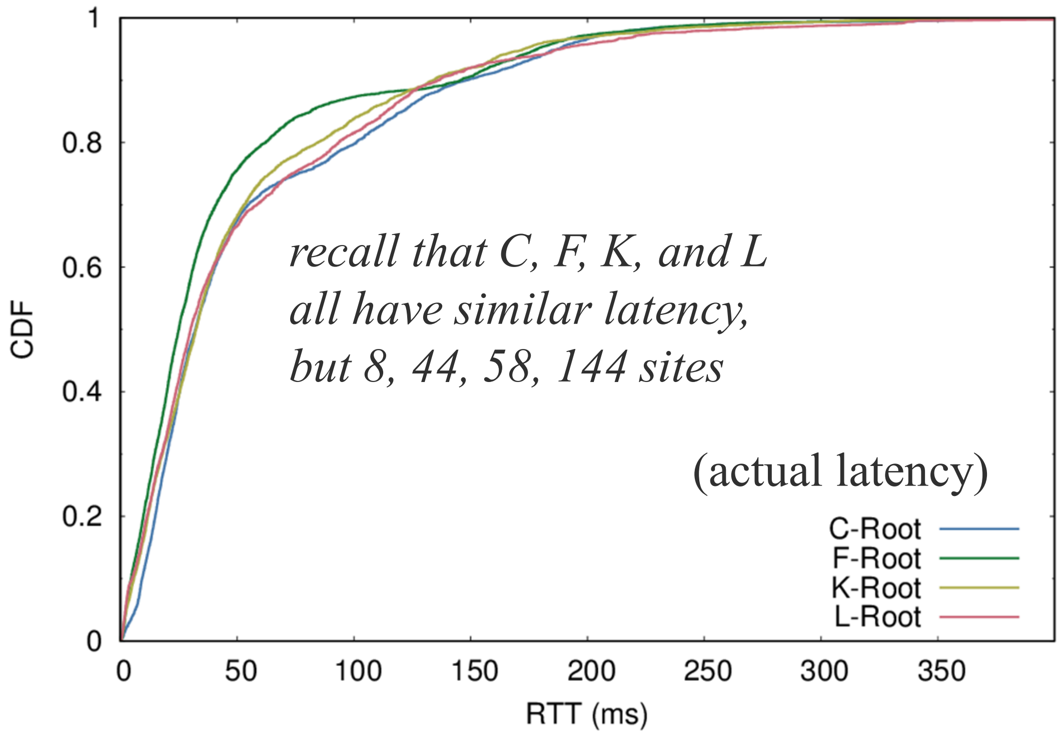
…they debugged routing with Ripe Atlas months before our measurements





…and routing bugs (or insufficient local interconnect?) in some countries

So How Many Instances?



recall that C, F, K, and L all have similar latency, but 8, 44, 58, 144 sites

(actual latency)

not very many!
12 would be good

for *an* anycast service
where caching works well
(not a CDN w/5 min TTL)

and only about **latency**
not DDoS resilience
nor policy concerns

and *location matters!*
and more help the tail

Conclusions

- first systematic study of anycast optimality
  - detailed paper: http://www.isi.edu/~johnh/Schmidt16a
- considerations for anycast design?

- data is available: https://ant.isi.edu/datasets/anycast/ and http://traces.simpleweb.org/

- thanks to RIPE Atlas and U. Twente