Internet Anycast: Performance, Problems and Potential

Zhihao Li, Dave Levin, Neil Spring, Bobby Bhattacharjee

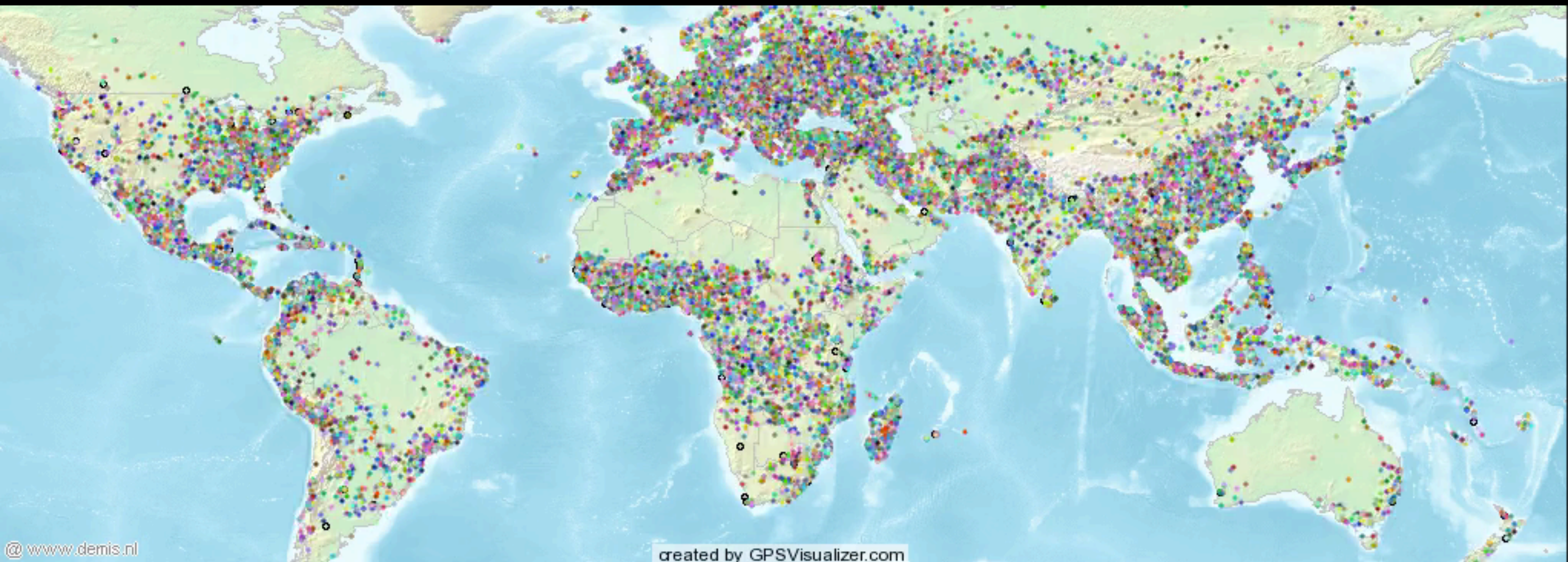
University of Maryland

Anycast is increasingly used

- DNS root servers:
  - All 13 DNS root servers
- Open DNS resolvers:
  - Google, Cloudflare, Quad9, OpenDNS, etc.
- Content Delivery Networks:
  - Cloudflare CDNs (Stack Overflow, Yelp, etc.)

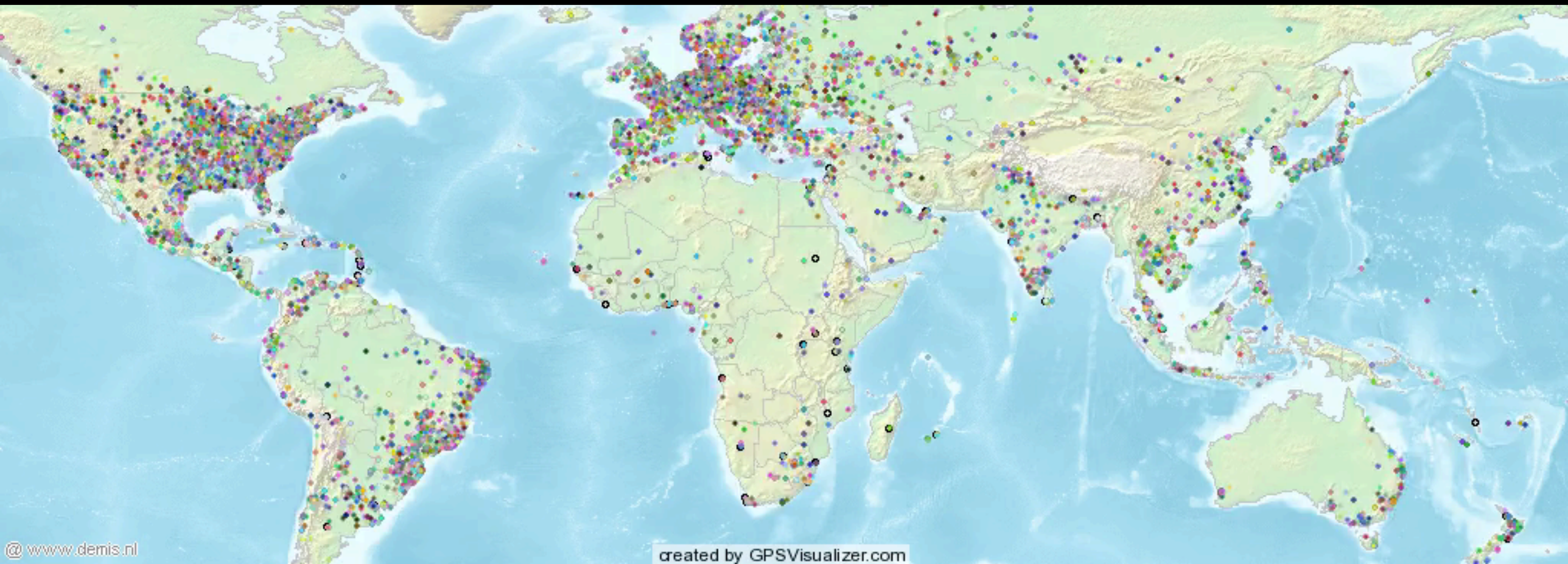
Mental model for anycast

- Packets sent to an anycast address *should* travel to a nearby site, subject to global/local constraints
- More sites *should* mean lower latency, better distribution, reliability against DoS attacks



Anycast often chooses poorly

- Actual distribution:
  - Many queries go to distant sites
- Video below used queries from D-root on Dec. 1st 2016



Outline

Performance — Do queries go to nearby anycast sites? Does adding anycast sites help?

Problems — Why not?

Potential — What can we do to fix it?

This is our data

- Trace of DNS queries at D-root servers
  - ~20% of DNS queries to D-root at each site
  - Obtain approximate location of query source using MaxMind
- Measurements from RIPE Altas probes
  - Over 9000 probes in ~180 countries and 3587 ASes
  - DNS CHAOS queries + traceroute
  - Measurements towards 9 out of 13 DNS root servers

Do queries go to nearby sites?

- We compute the extra distance each query travelled over their geographically closest site
- Of course, we don't expect clients to find the *closest* site
- But most queries ***should*** have short extra distance

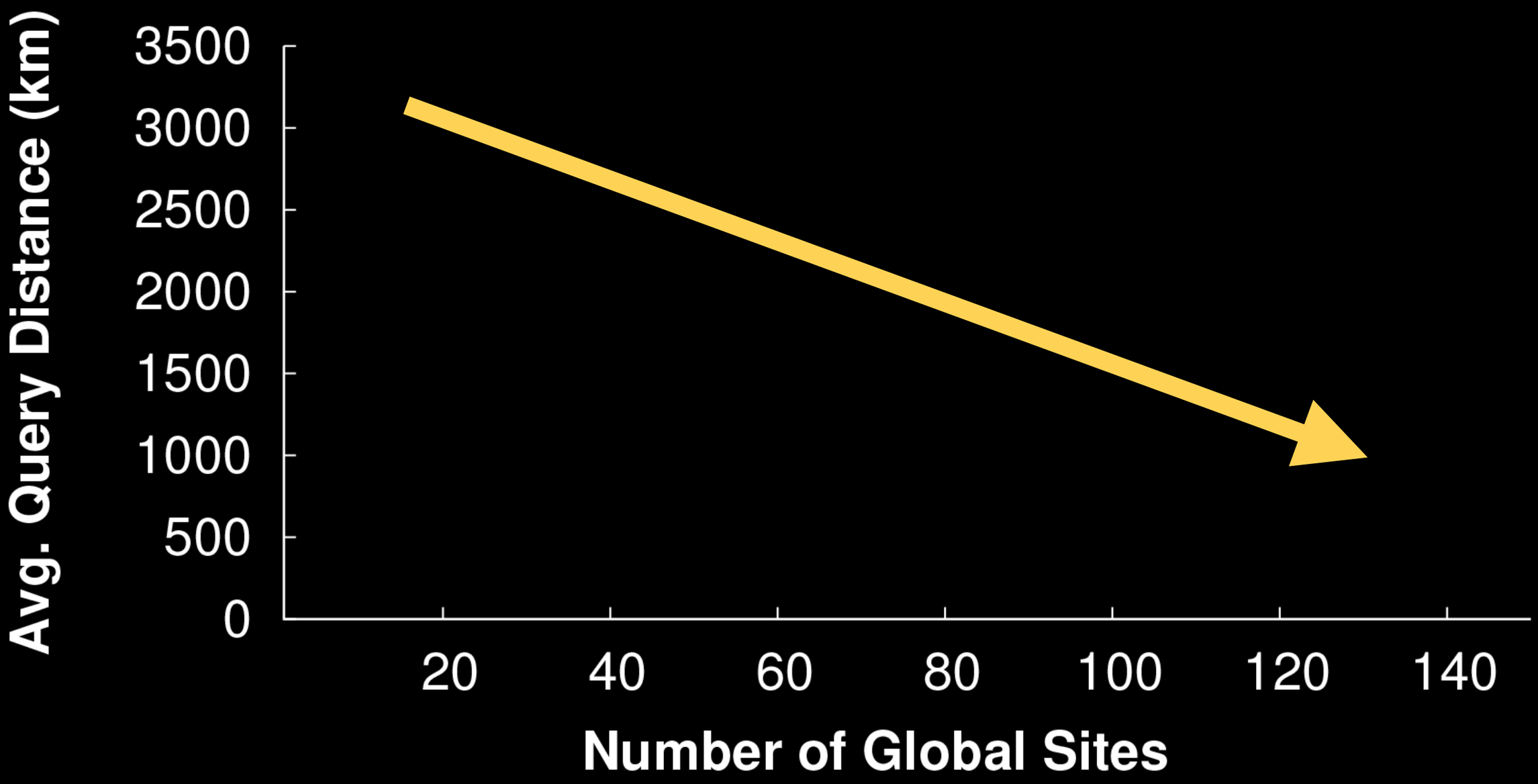
Over 1/3 queries traveled more than 1000km extra distance

The performance problem in D-root is representative of many current anycast deployments.

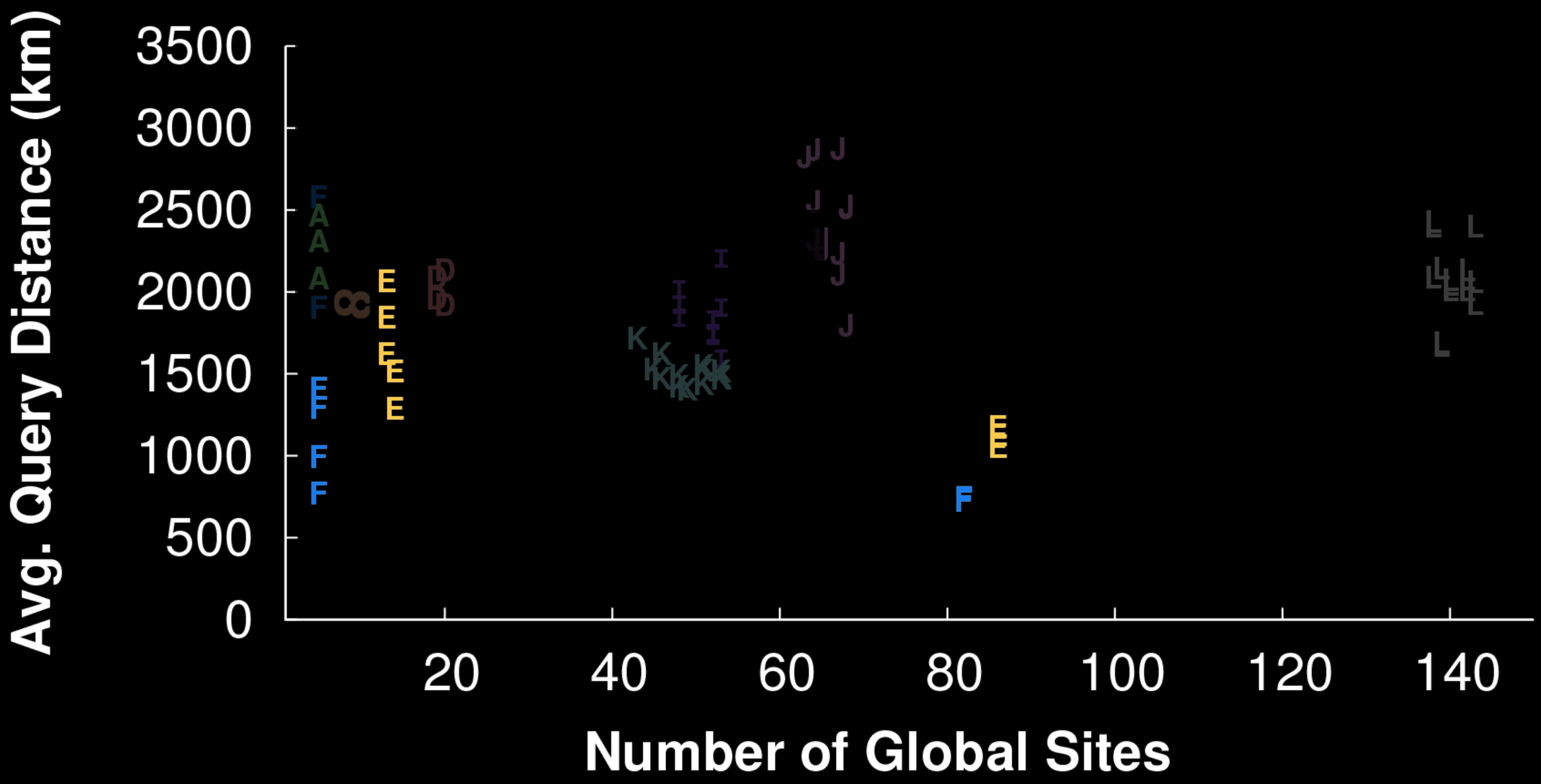
Does adding anycast (global) sites help?

- We analyze longitudinal RIPE Atlas measurements
- Evaluate performance of 9 DNS roots in 2017
- Compute average query distance in each week

Site counts hardly matter



Site counts hardly matter



- Each letter represents a performance value from one-week's data
  - Vertical displacement of letters show performance variations under the same number of site

Outline

Performance — No! Over 1/3 queries traveled 1000+ km more than necessary; Adding sites hardly improves

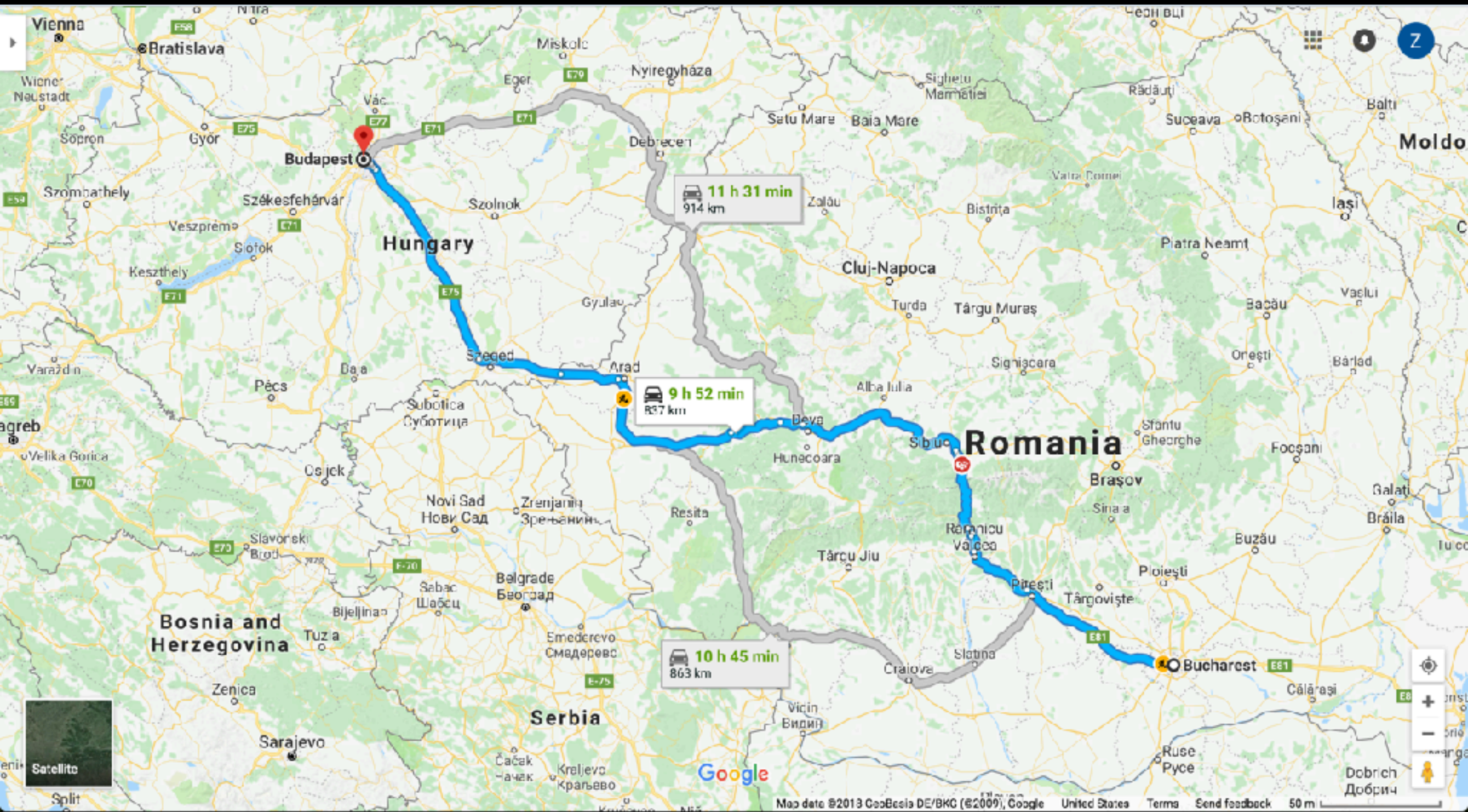
Problems — Why not?

Potential — What can we do to fix it?

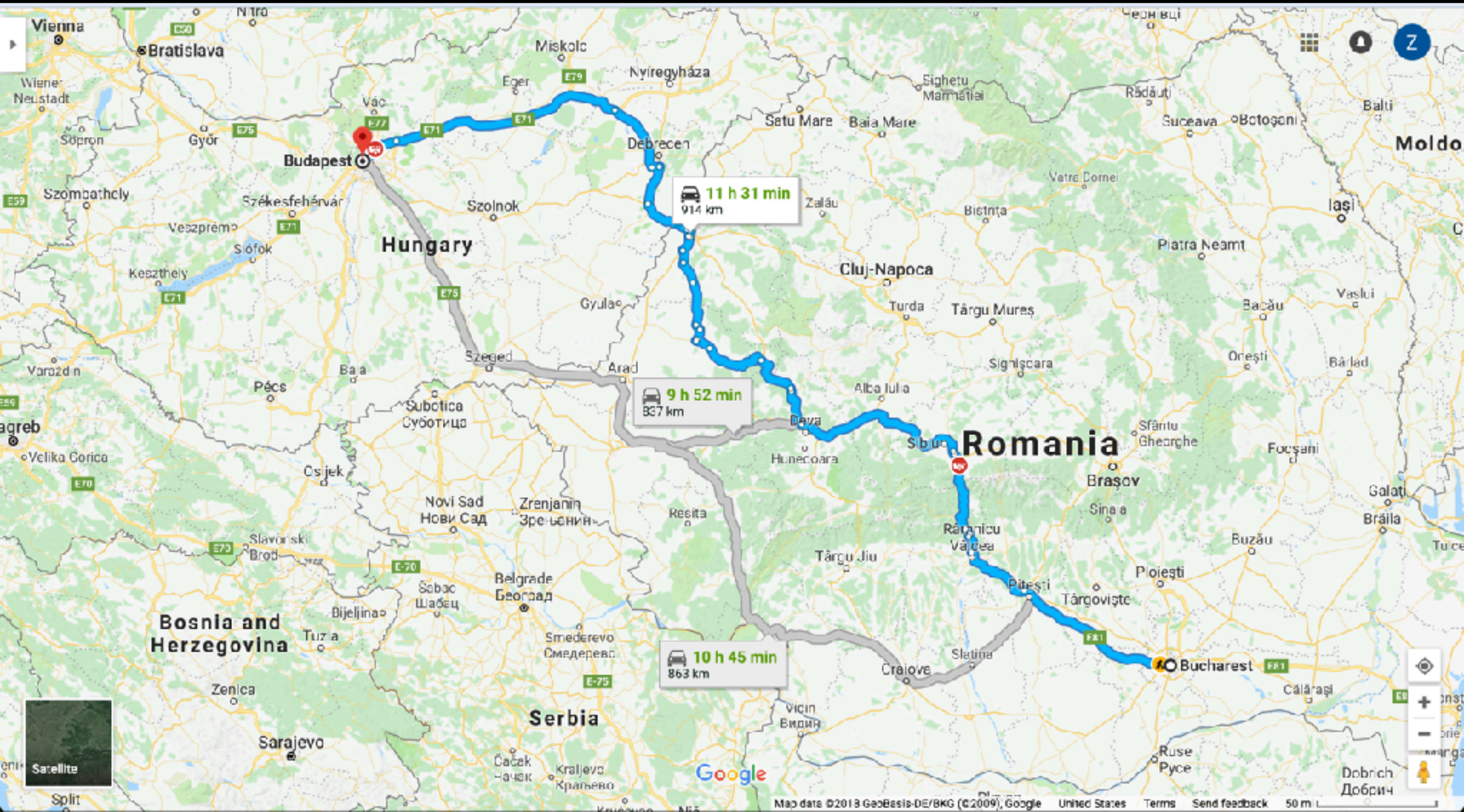
Why doesn't anycast work as expected?

- Routing and topology constraints introduce path inflation
  - The actual path is longer than the shortest
- However, path inflation in anycast is different…
  - Directs packets away from the closest anycast site

Path inflation in unicast is like taking a detour



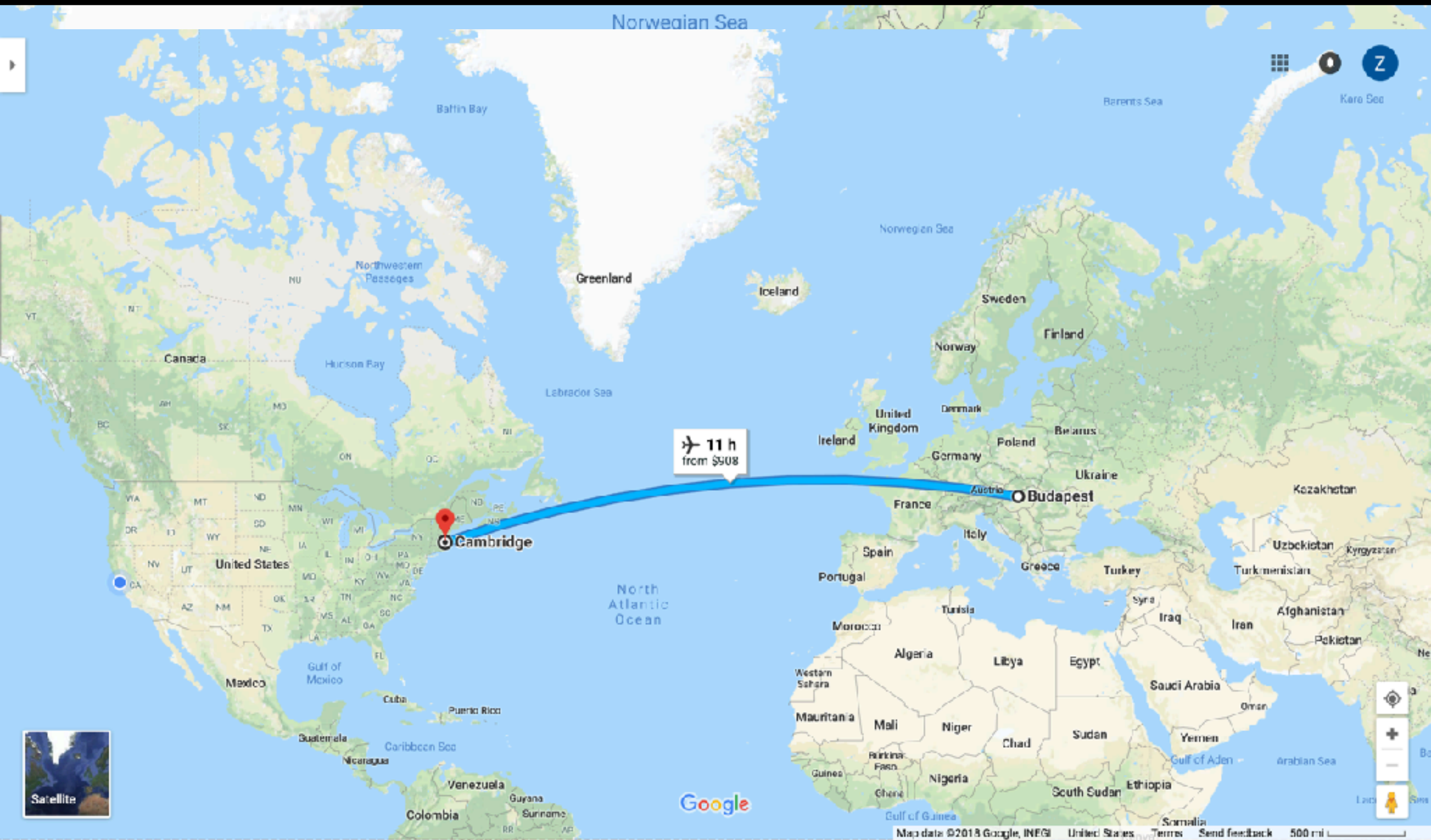
Path inflation in unicast is like taking a detour



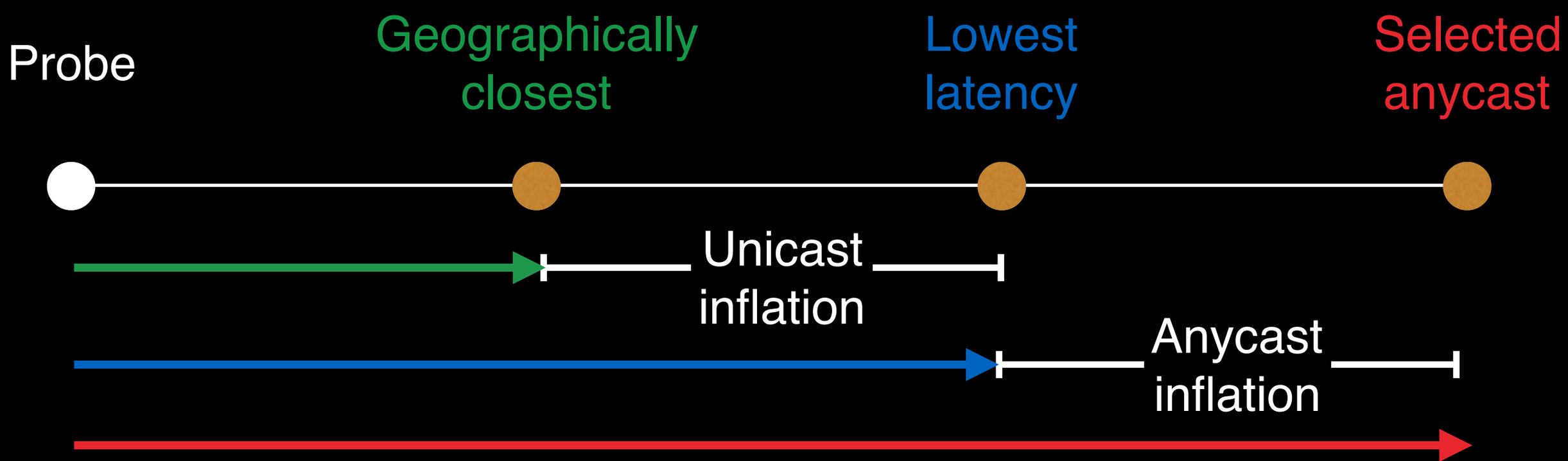
Path inflation in anycast is different

Say we jump on a flight to “Cambridge”…

Path inflation in anycast is like flying to a different Cambridge



Anycast & unicast path inflation

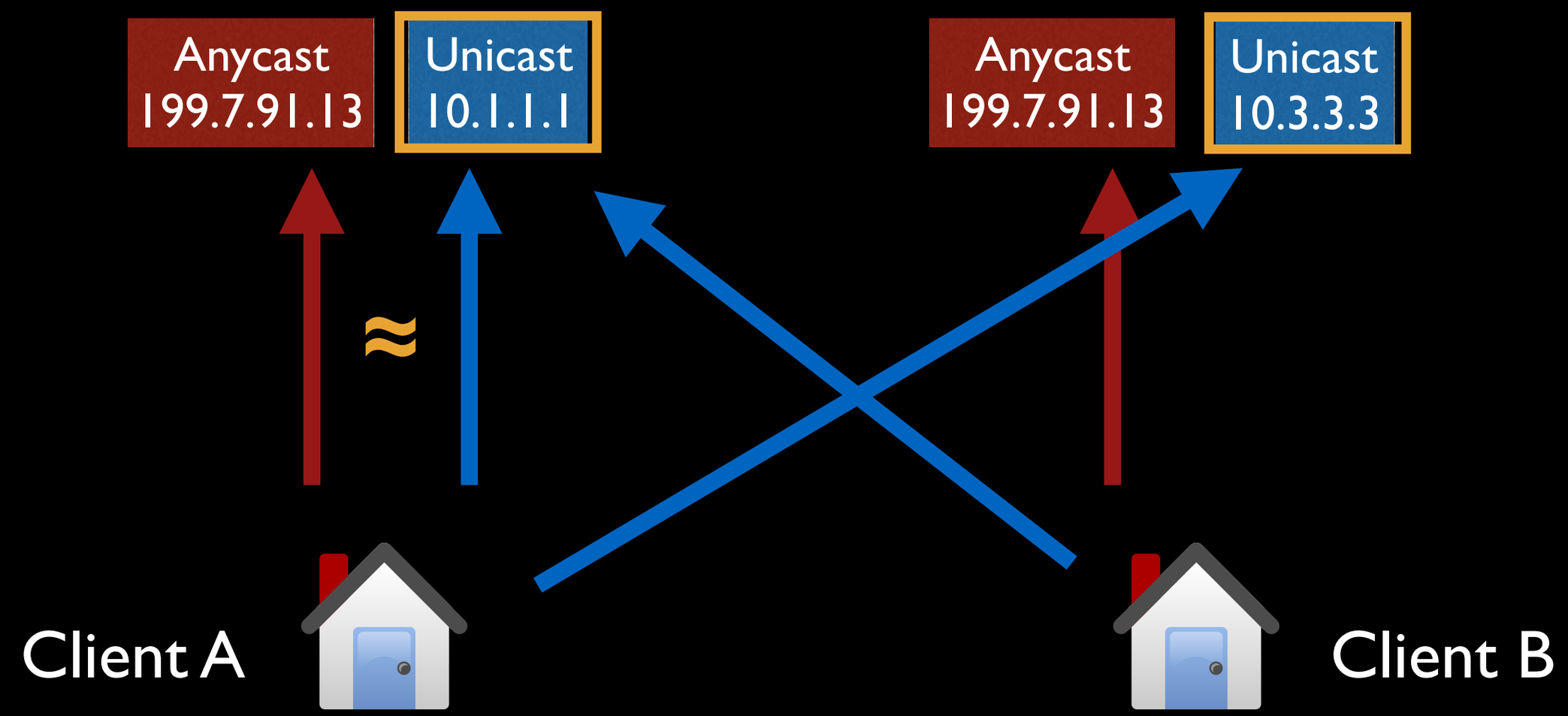


Unicast inflation: Difference between the predicted latency[1] to geographically closest site, and latency to lowest latency site.

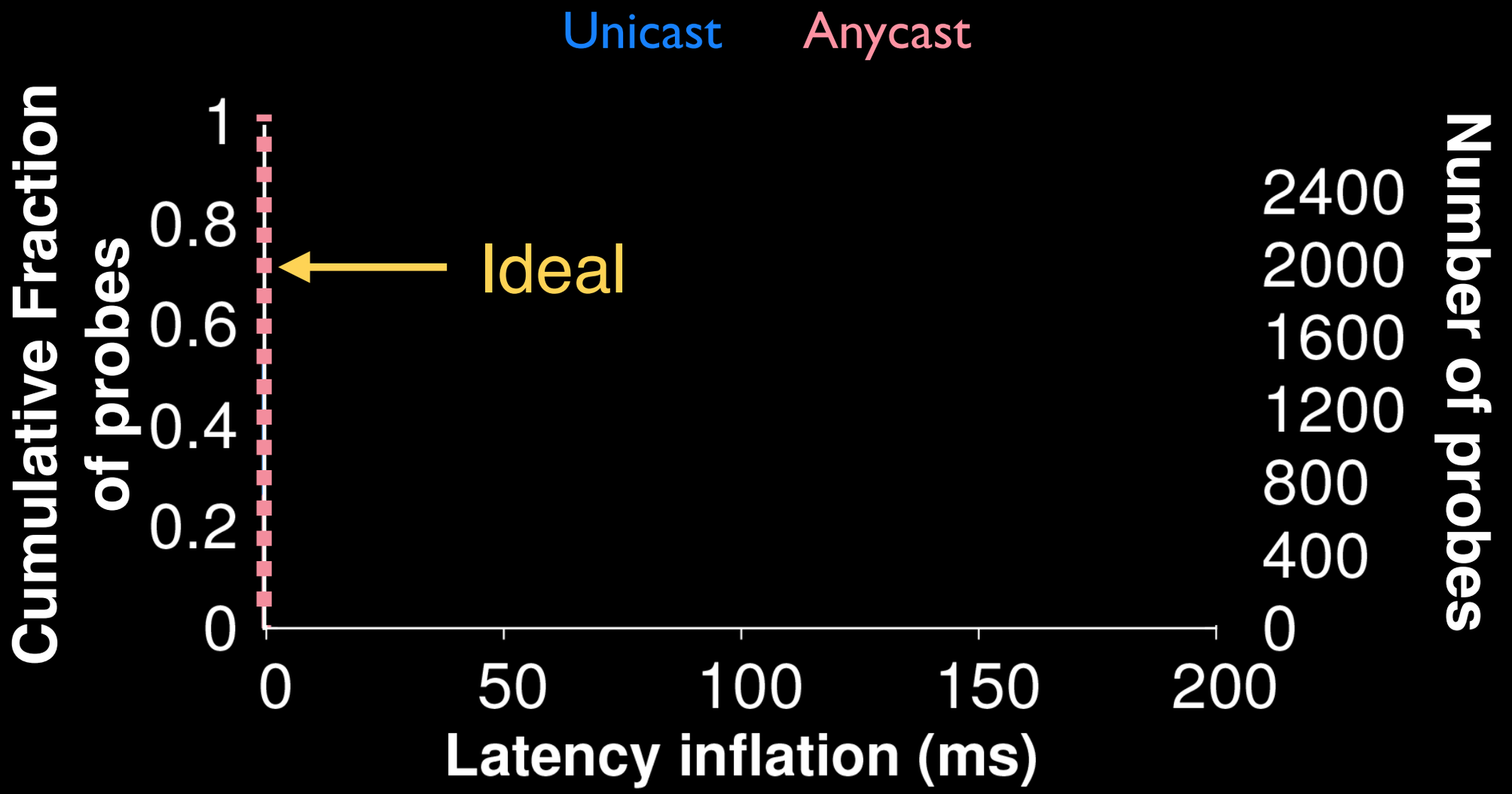
Anycast inflation: Difference between latency to lowest latency site, and latency to selected site.

[1] Agarwal et al. Matchmaking for Online Games and Other Latency-Sensitive P2P Systems. SIGCOMM'09.

Use unicast representatives to estimate alternate site performance

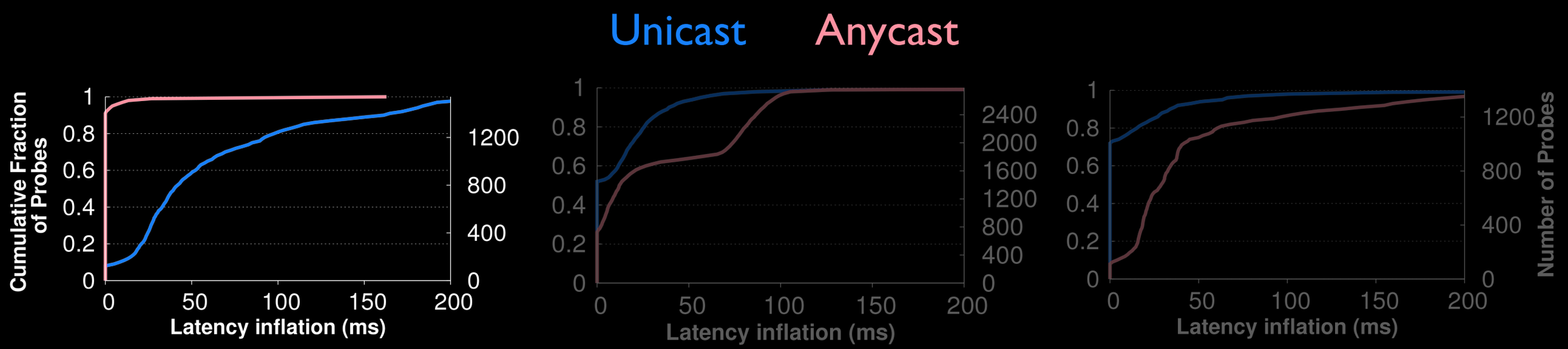


Anycast path inflation is larger than unicast path inflation



D-root

Anycast path inflation is larger than unicast path inflation



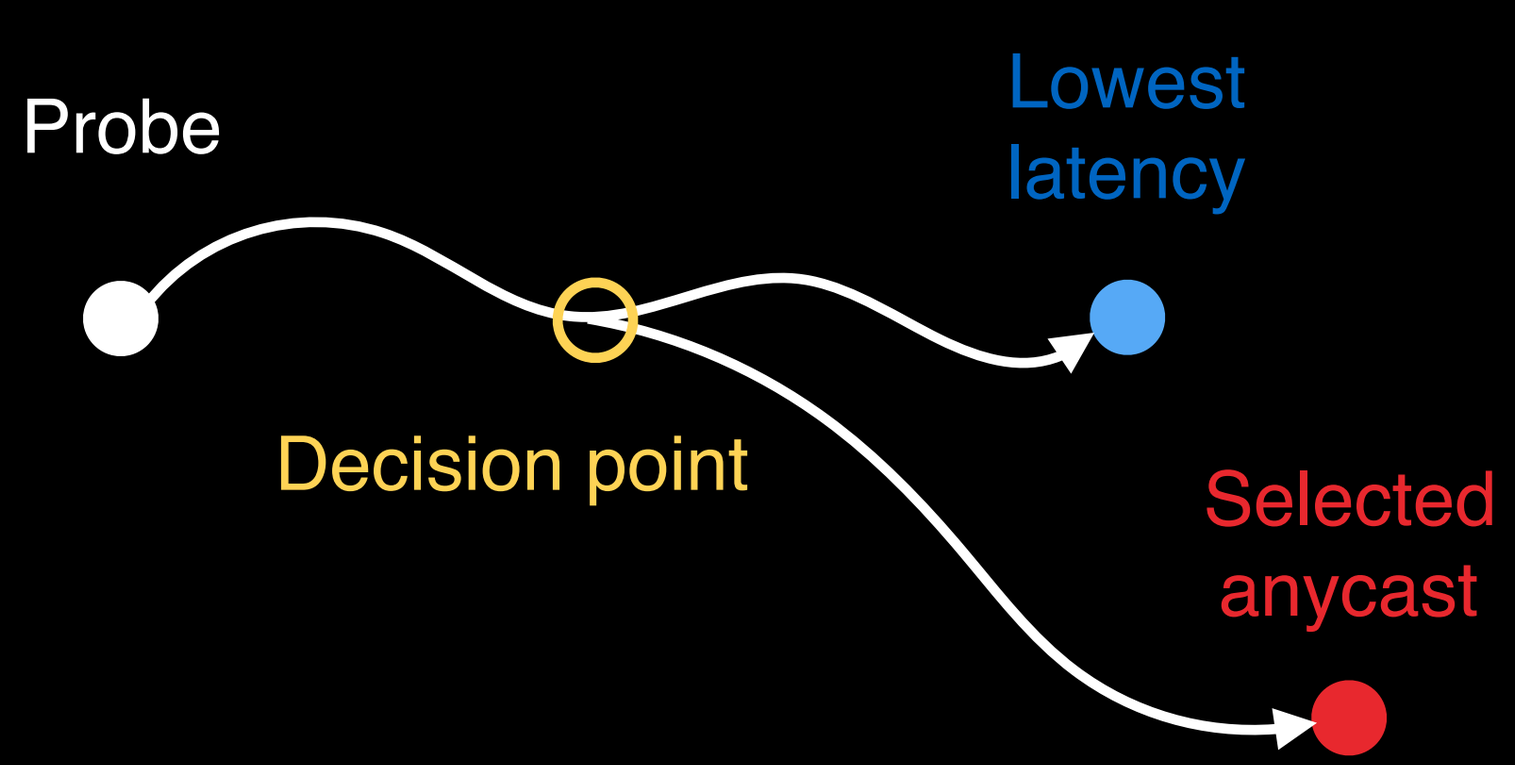
C-root **D-root** **K-root**

All C-root sites share the same provider

All other roots use multiple providers

Would the best routes comply with typical routing policies?

- Extract AS-level paths[2] from traceroutes
  - to selected site and to lowest-latency site
- Find the 'decision point' where they diverge



[2] Mao et al. Towards an accurate AS-level traceroute tool. SIGCOMM'03

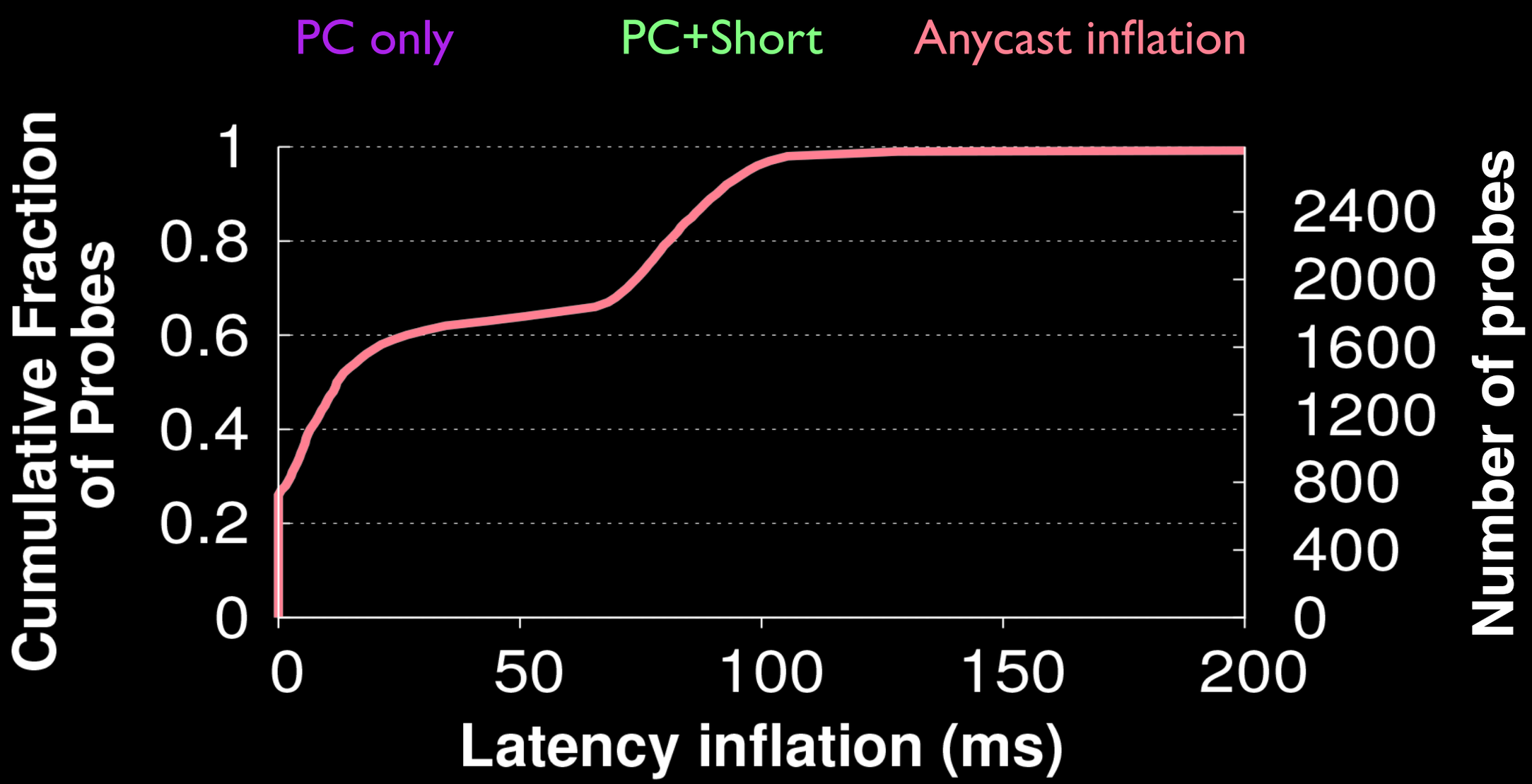
Would the best routes comply with typical routing policies?

- Extract AS-level paths[2] from traceroutes
  - to selected site and to lowest-latency site
- Find the 'decision point' where they diverge
- Examine if the better route is not selected due to:

Prefer-Customer (PC)	AS prefers routes through its customer ASes over the peer ASes, over its provider ASes
Shortest AS path (Short)	AS prefers routes with shortest AS path length

[2] Mao et al. Towards an accurate AS-level traceroute tool. SIGCOMM'03

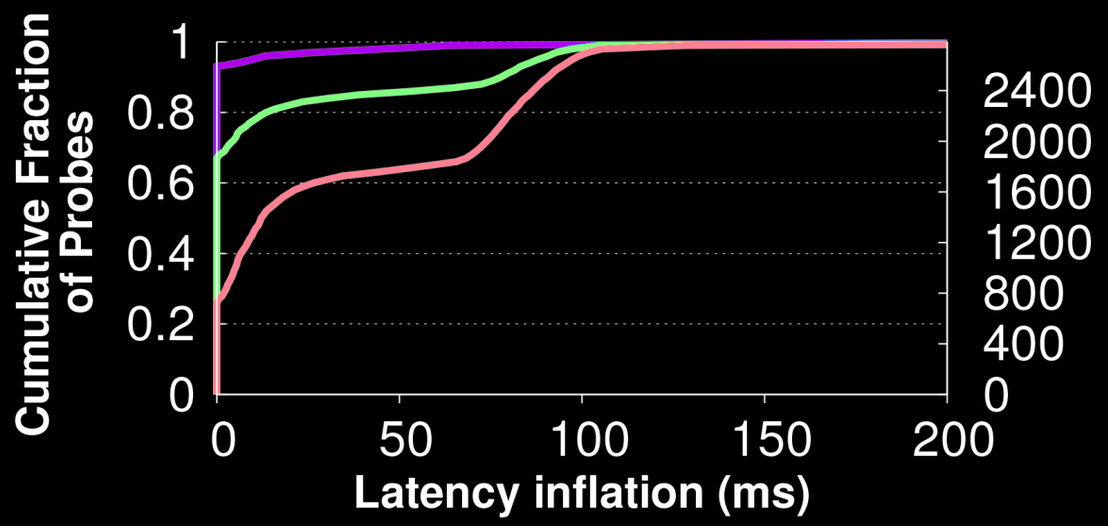
Can correct while following typical routing policies



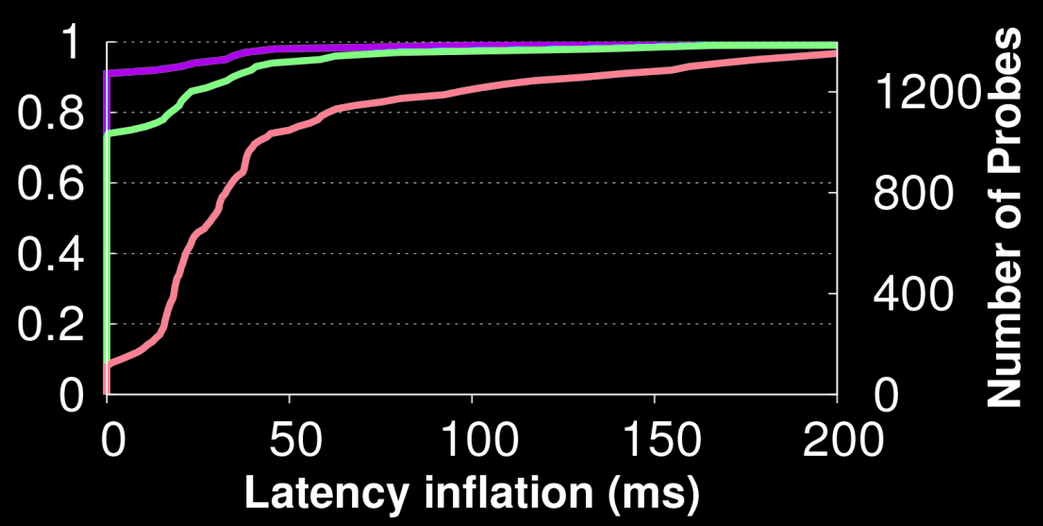
D-root

Can correct while following typical routing policies

PC only PC+Short Anycast inflation



D-root



K-root

Outline

Performance — No! Over 1/3 queries traveled 1000+ km more than necessary; Adding sites hardly improves

Problems — Given routes with equal preference to different sites, BGP usually chooses poorly

Potential — What can we do to fix it?

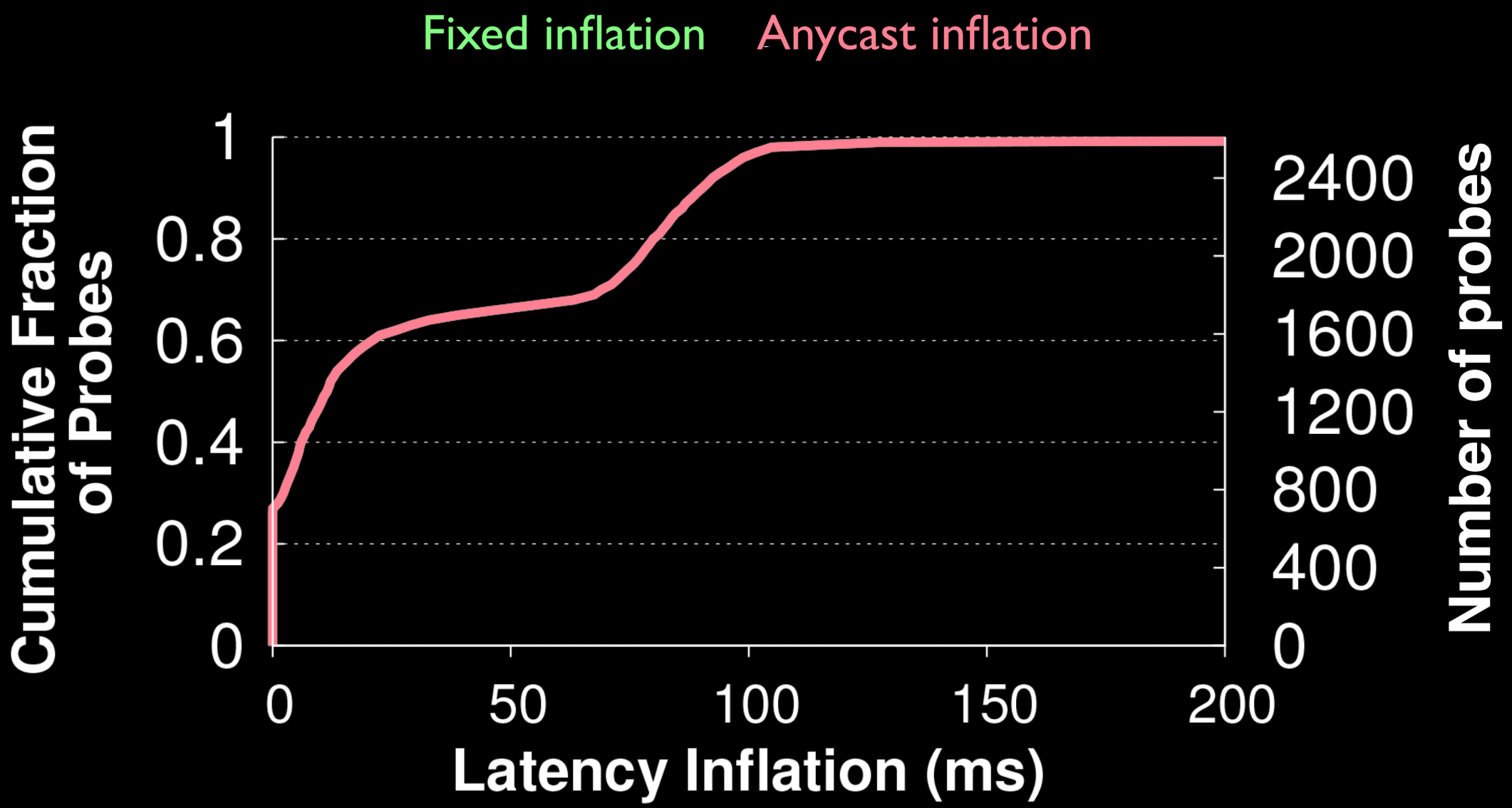
Embedding geographic information in BGP announcements

- Embed geographic location of the anycast sites reachable through the announcement
- Use BGP community attributes
  - Two 16-bit values X:Y
  - X represents the AS number that sets the community
  - Y encodes the latitude and longitude

Evaluate the geographic hint through simulation with real network traces

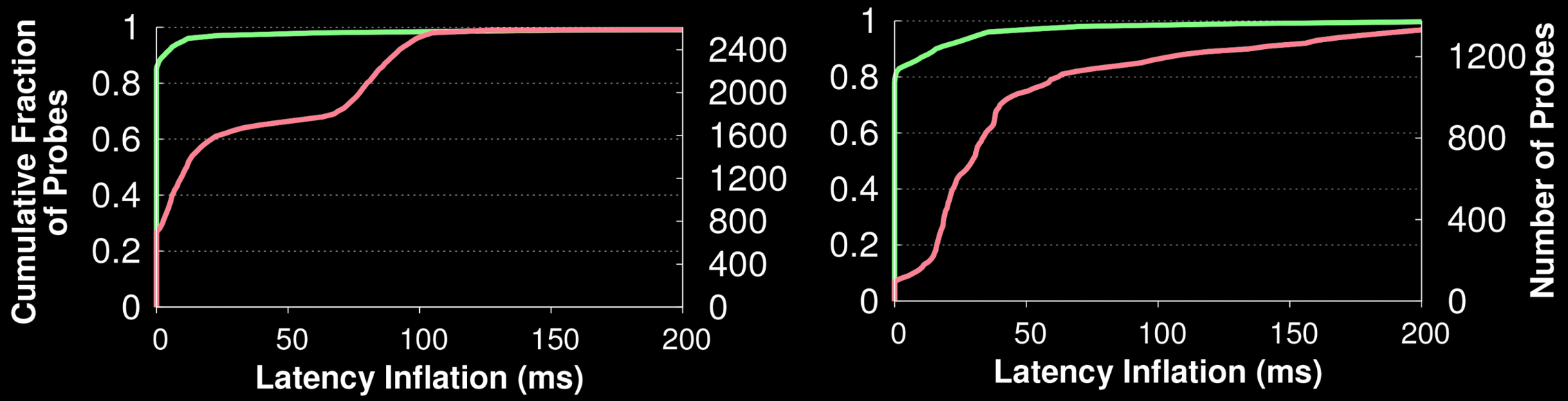
- Using the traceroutes from each RIPE Atlas probe
  - to the selected site and to lowest-latency site
- Identify the ‘decision point’
- Identify the geo-closest site to the ‘decision point’, refer to as geo-hinted site
- Benefits of geo-hint is the difference between latency to geo-hinted site and to selected site

A simple geographic hint provides large improvement on anycast performance



D-root

A simple geographic hint provides large improvement on anycast performance



D-root K-root

Outline

Performance	No! Over 1/3 queries traveled 1000+ km more than necessary; Adding sites hardly improves
Problems	Given routes with equal preference to different sites, BGP usually chooses poorly
Potential	A simple geo-hint in BGP communities can recover much of the performance deficit

Other results

- Anycast does not distribute traffic in a balanced manner
- Performance problems in anycast are common across deployments
- RIPE-Atlas measurements overrepresent Europe, but this effect only over-estimates anycast performance
- Unicast management addresses for anycast sites are good representatives
- Customized community attributes are more effective than expected, given the default configuration in most routers

Anycast doesn't work as well as it should, but it can

- Inefficiencies in anycast are excessive
  - Queries to most DNS roots travel to distant sites
  - Adding sites hardly improves anycast performance

- Poor route selection in BGP causes larger path inflation in anycast than in unicast
  - There exist equal-preference routes to closer sites
  - But no mechanism to choose the best among them

- Incrementally deployable "geo-hints" in BGP can recover most of performance deficit

Data available at:
cs.umd.edu/projects/droot/