



ReOpt: Near-Optimal Region Division for Low-Latency Regional Anycast

Yimeng Xu

with Minyuan Zhou, Congying Wang Jiaqi Zheng, Shuai Hao, Guihai Chen, Jiewu

State Key Laboratory for Novel Software Technology, Nanjing University

Old Dominion University

China Telecom Cloud Computing Research Institute & Temple University

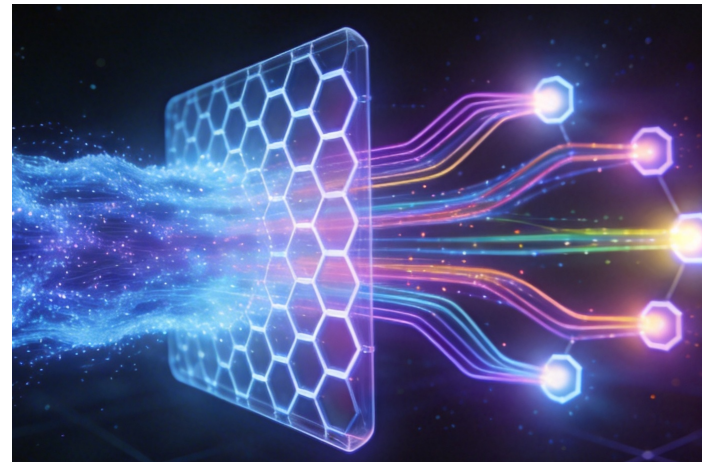


Advantages of anycast



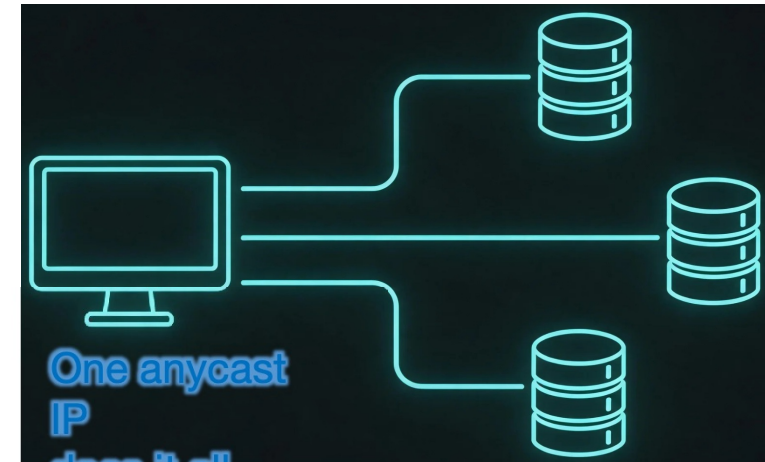
Robustness

Traffic is automatically rerouted once upon failure



Load Balancing

Prevent any single site from being overwhelmed



Configuration Simplification

One anycast IP address
Multiple servers

Regional anycast can make further improvement





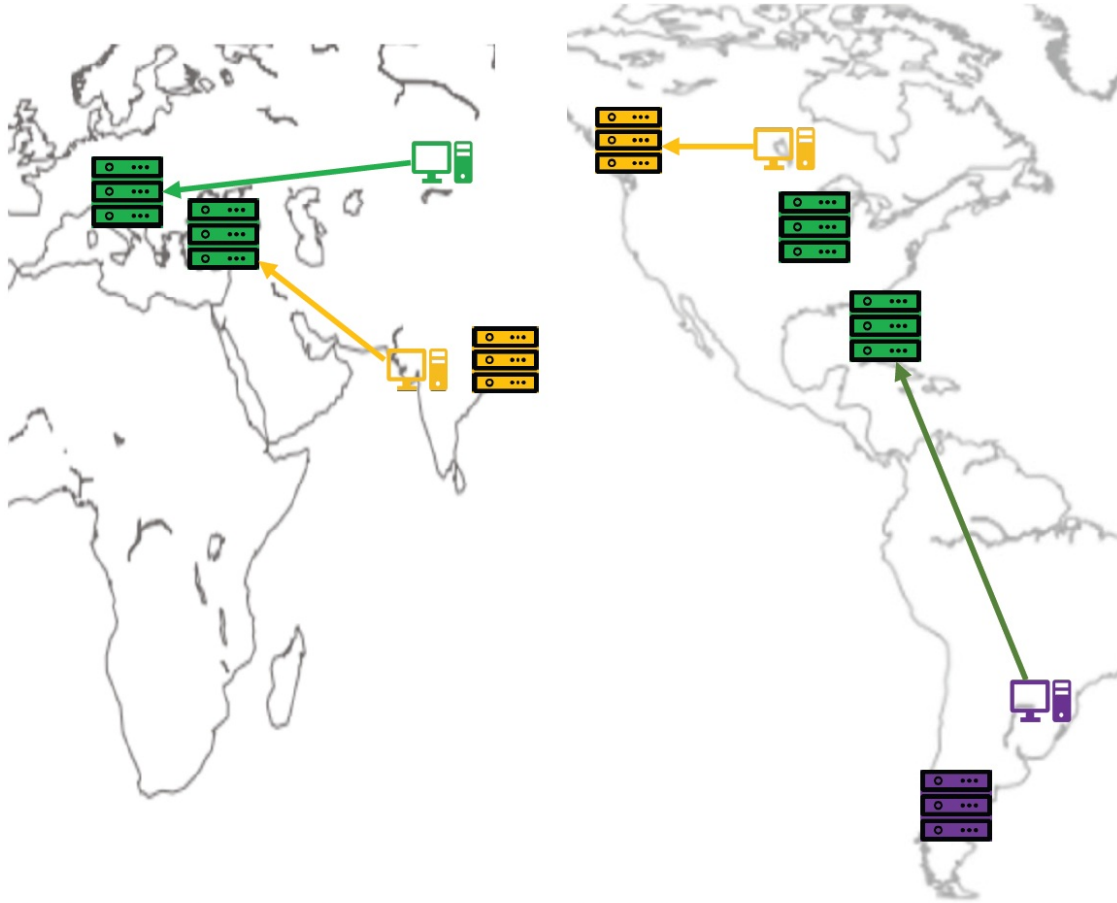
**Intuitive region partition is
geographical boundaries.**

Is it enough?

No, it is not enough.



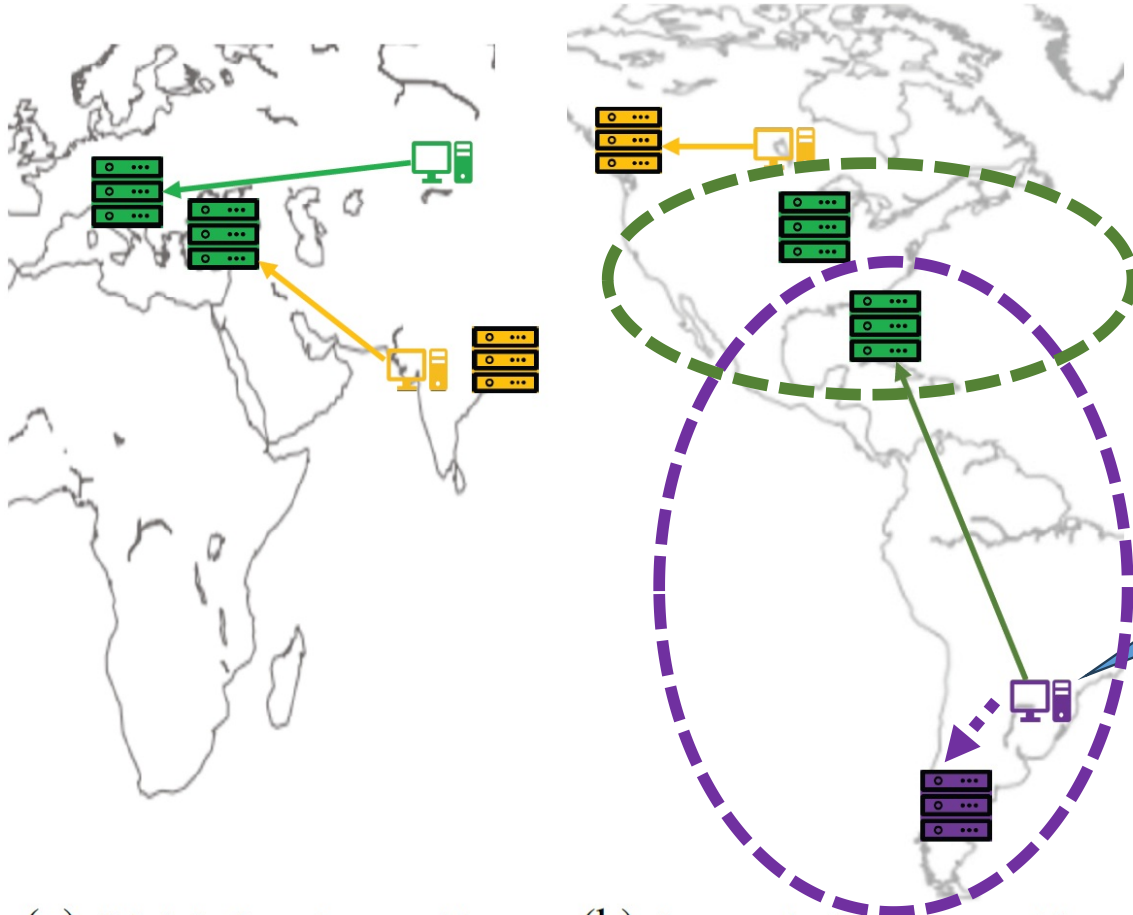
Unexpected path inflation



(a) *Edgio's 3-region partition.* (b) *Imperva's 6-region partition.*

Rigid geographical boundaries
still result in
suboptimal routing decisions

Policy-driven vs Latency-minimized

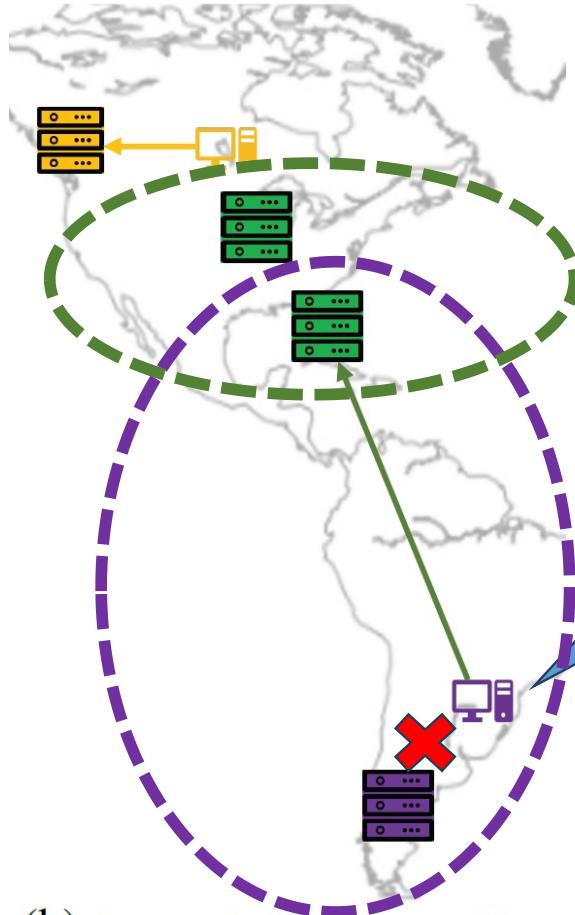


(a) *Edgio's 3-region partition.*

(b) *Imperva's 6-region partition.*

I want to achieve the latency-minimized site

Policy-driven vs Latency-minimized



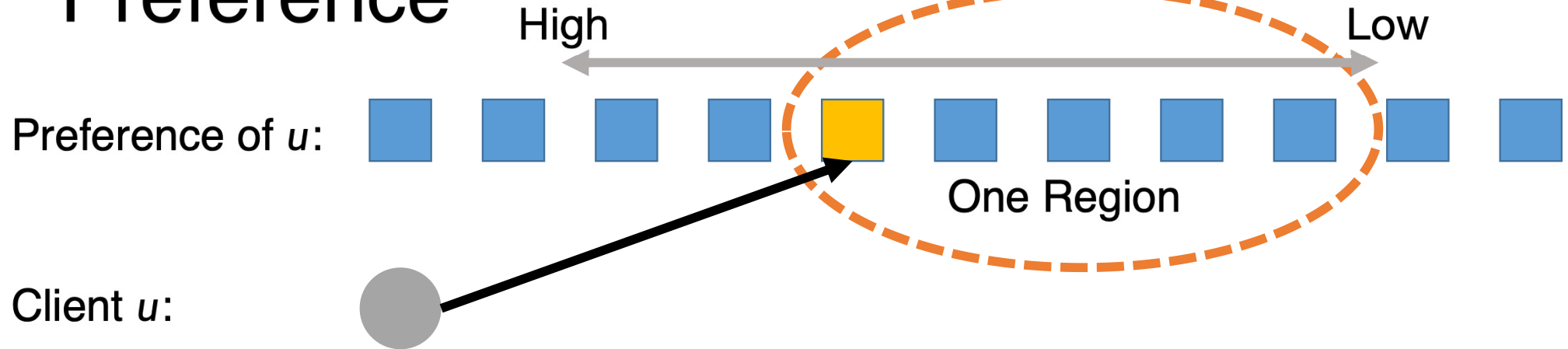
No, you don't.

Policy-driven
Routing Protocol

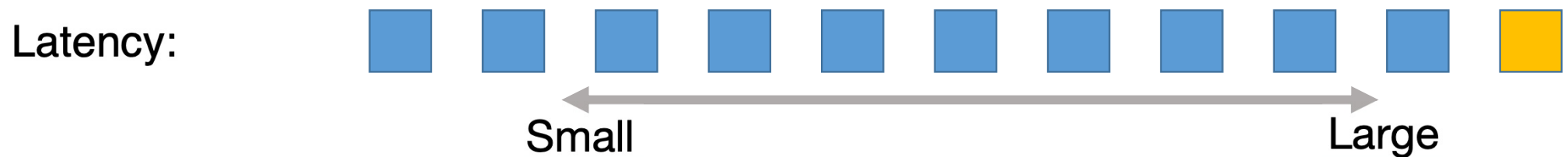
(a) Edgio's 3-region partition.

(b) Imperva's 6-region partition.

Preference



This Inconsistency Makes Optimization Tough

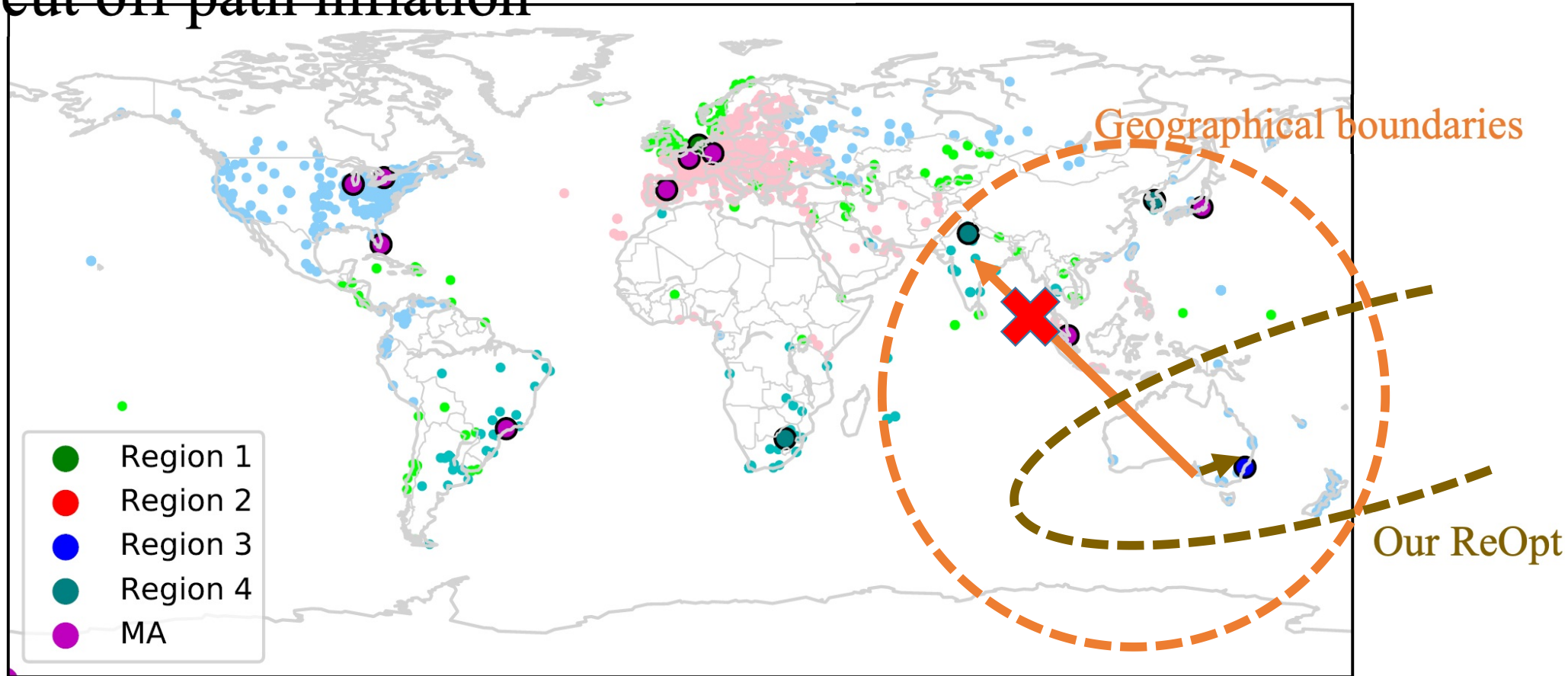




**Any solutions to overcome
this inconsistency?**



A carefully-designed region partition strategy can cut off path inflation





How to formalize this problem?





LMARP:

A **L**atency-**M**inimized
Anycast **R**egion **P**artition
algorithm





Our formulation

- **Objective Function:** the summation of RTTs between clients and target sites.
- **Assignment Constraint:** a client is assigned to exactly one site.
- **Multi-announcement Constraint:** a site can announce multiple IP prefixes, so that it can join multiple regions.
- **Country-aware Constraint:** the clients in the same country shares the same IP prefix, so that they should be assigned to the same region.
- **Preference Constraint:** for a client u , we use \succ_u to describe the preference.

NP-hard!

Detailed mathematical expressions can be found in our paper.





NP-hardness

- **Objective Function:** the summation of RTTs between clients and target sites.
- **Assignment Constraint:** a client is assigned to exactly one site.
- **Multi-announcement Constraint:** a site can announcement multiple IP prefixes, so that it can join multiple regions.

- **Country-aware Constraint:** the clients in the same country shares the same

~~IP prefixes so that they are assigned to the same region.~~ Simplified-EMARP is still an NP-hard problem

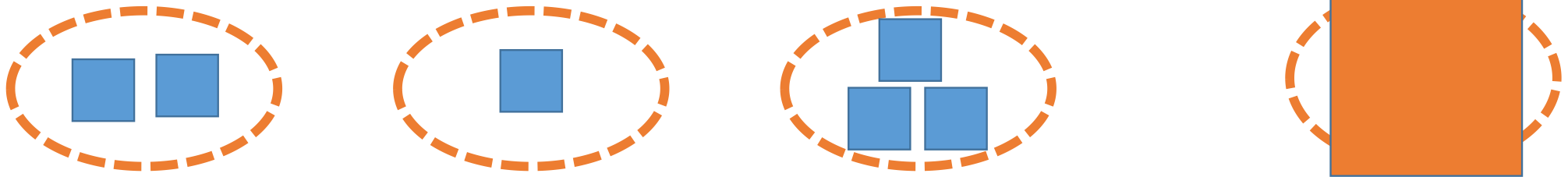
- **Preference Constraint:** for a client u , we use \succ_u to describe the preference.

Detailed proof can be found in our paper.



NP-hardness

- Take use of the inconsistency: sites with higher preference have higher latency
- Set a virtual site: globally highest preference and infinity latency
- Obtain optimal solution with specific form: one-element region



- Finish the reduction: get a solution for an ℓ -median problem

Detailed proof can be found in our paper.

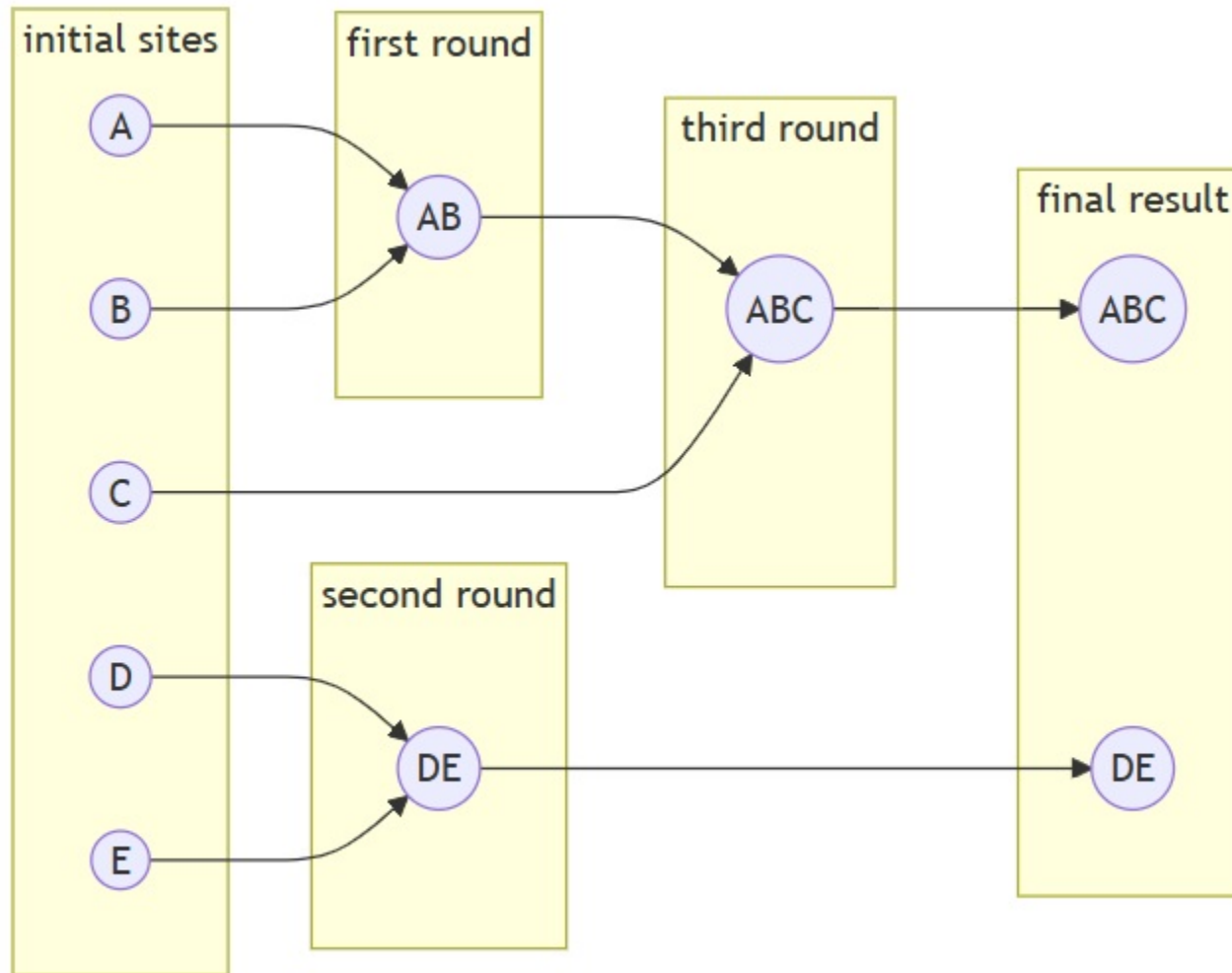


Our algorithms

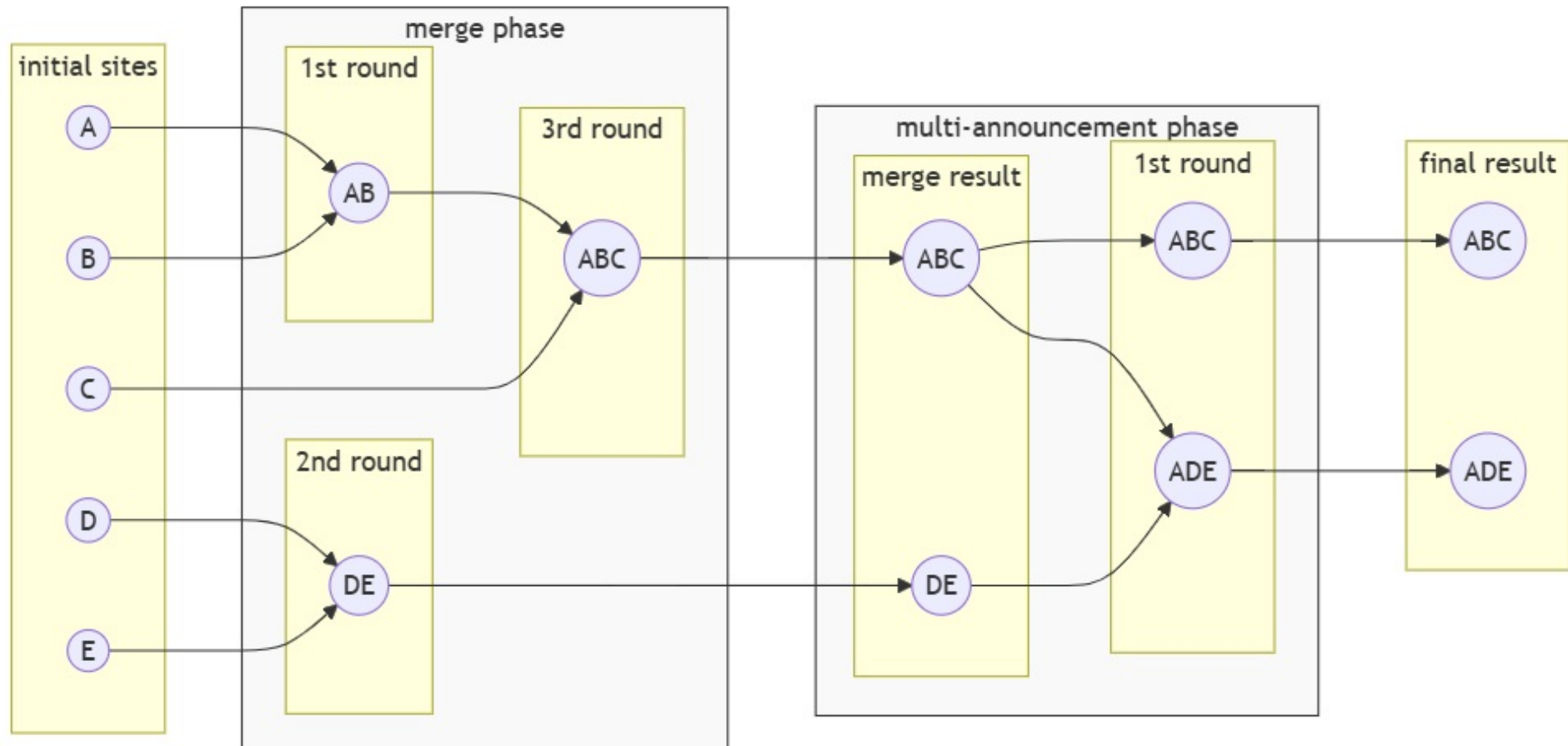
- RGA: A reverse greedy algorithm (RGA)
- RGA-MA: RGA with Multi-announcement constraints
- RGA-MA-C: RGA-MA with Country-aware constraints



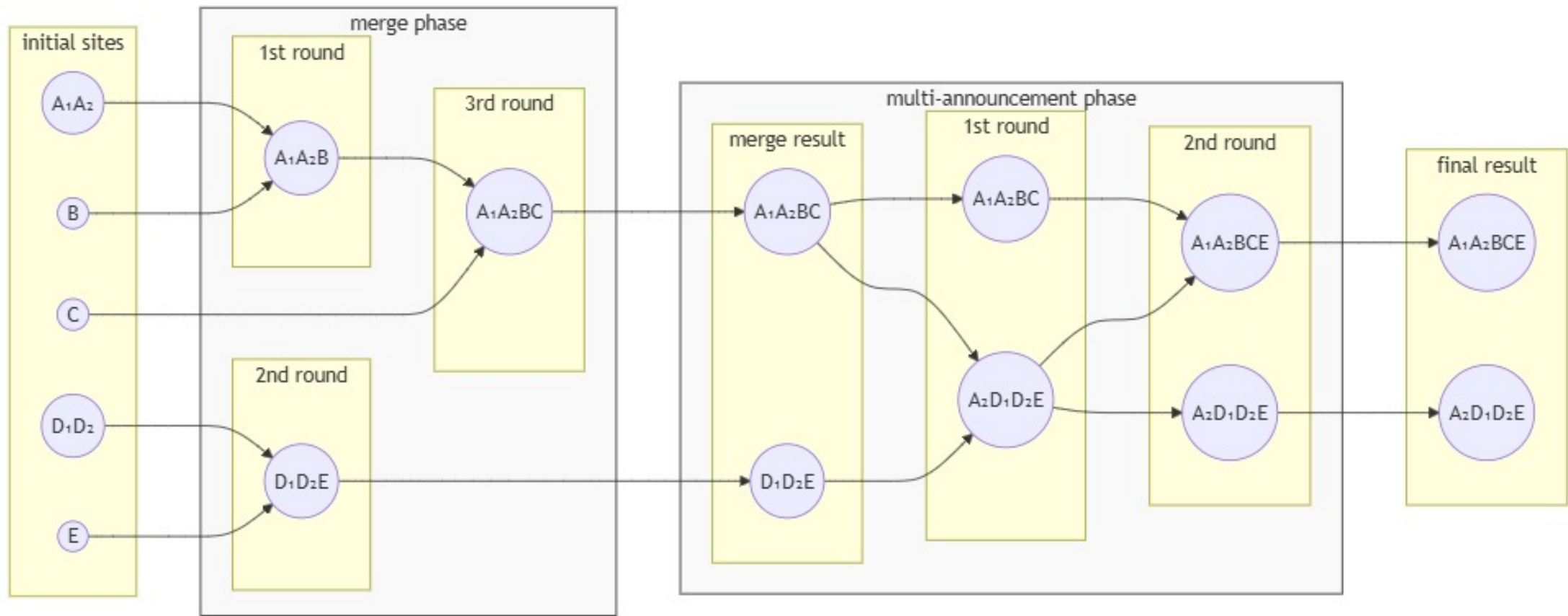
RGGA



RGMA-MA



RGA-MA-C



Theoretical guarantee

Algorithm	Approximation Ratio
RGA	$O(\log(m - \ell))$ where m and ℓ is the amount of sites and target regions, respectively
RGA-MA	$O(\log(m - \ell))$
RGA-MA-C	$O(\log(p - \ell))$ where p is the amount of countries

Time complexity: polynomial-time

Detailed theoretical analysis can be found in our paper.

Setup: two public measurement platforms

- **Ripe Atlas:**
 - it is a global measurement infrastructure.
 - Our study utilizes all 5092 active probes spanning 137 countries.

Region	EMEA (Europe, Middle East, Africa)	NA (North America)	LatAm (Central America, South America)	APAC (Asia, Pacific)
#probes	3626	804	125	537

~~• We analyze results and present performance metrics separately for each region.~~

- **The PEERING testbed:**
 - It enables researchers to inject BGP announcements into the global Internet routing system.
 - It integrates Verfploeter* to measure catchment areas per anycast site and determine client preferences.

* Wouter B. de Vries, et. al.. 2017. Broad and load-aware anycast mapping with verfploeter. In Proceedings of the 2017 Internet Measurement Conference (IMC '17).



Evaluation: benchmark schemes

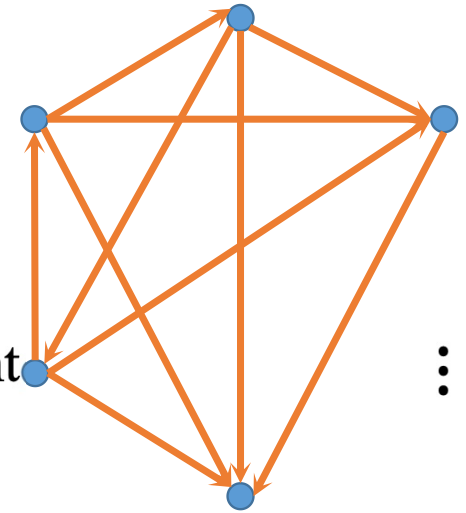
We evaluate the following schemes:

- **Edgio-3:** EG3 spans America (North and South), Asia-Pacific (APAC) and Europe-Middle East-Africa (EMEA).
- **Edgio-4:** EG4 covers North America (NA), South America (SA), APCA and EMEA.
- **Imperva-6:** IM6 encompasses United States, Canada, SA, APCA, EMEA, and Russia as distinct operational regions.
- **ReOpt:** Our proposed regularization-based algorithm.



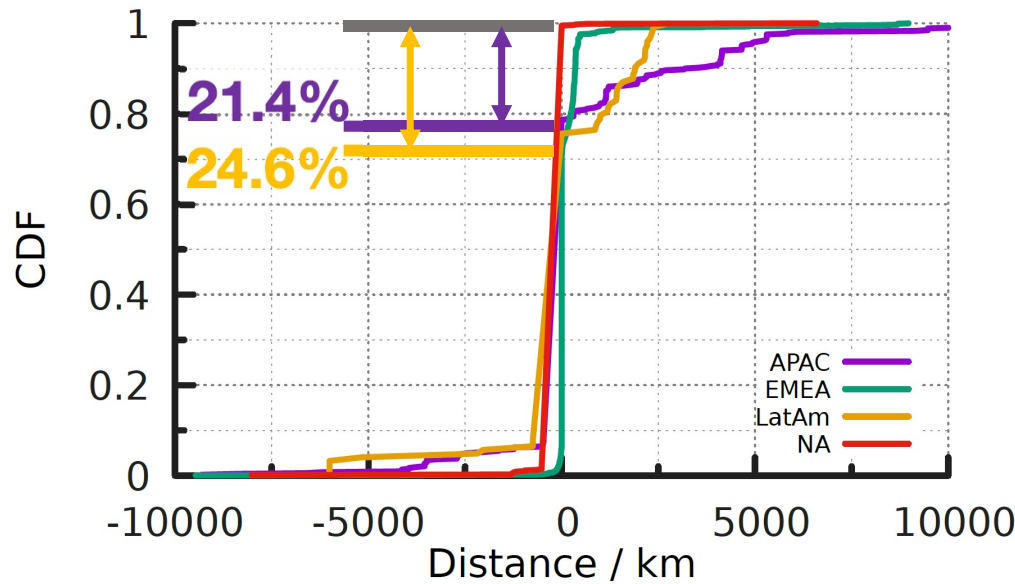
Data Imputation: Preference

- Make pair-wise comparison among all sites for each client
 - How can we do it *
- Construct a preference graph for each client complete or consistent total order across all sites across a given client.
- Sort preference like ranking single round robin competitors

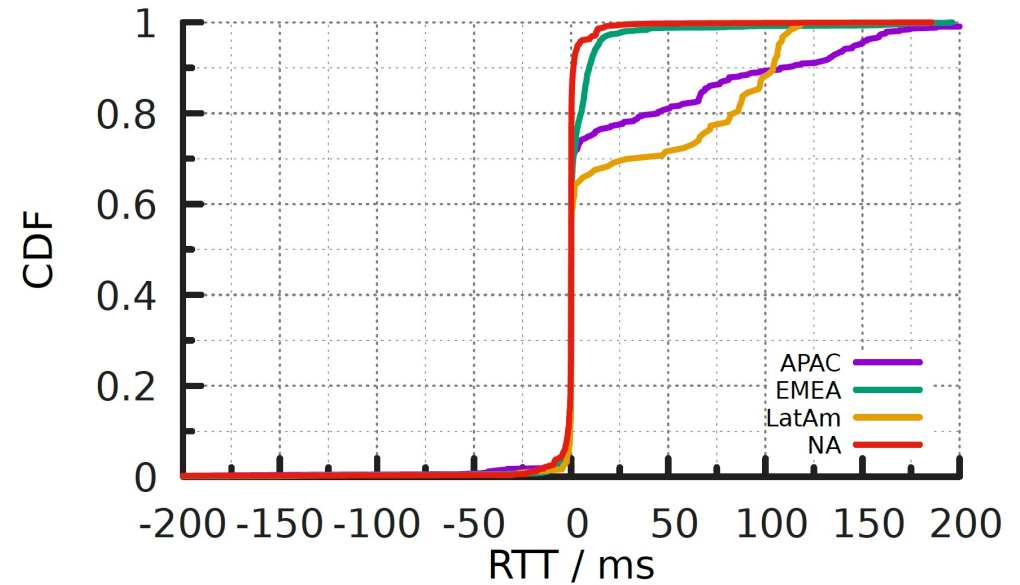


* Xiao Zhang, et. Al.. 2021. AnyOpt: predicting and optimizing IP Anycast performance. In Proceedings of the 2021 ACM SIGCOMM 2021 Conference (SIGCOMM '21).

Experiment results: ReOpt-6 deep dive to the improvements



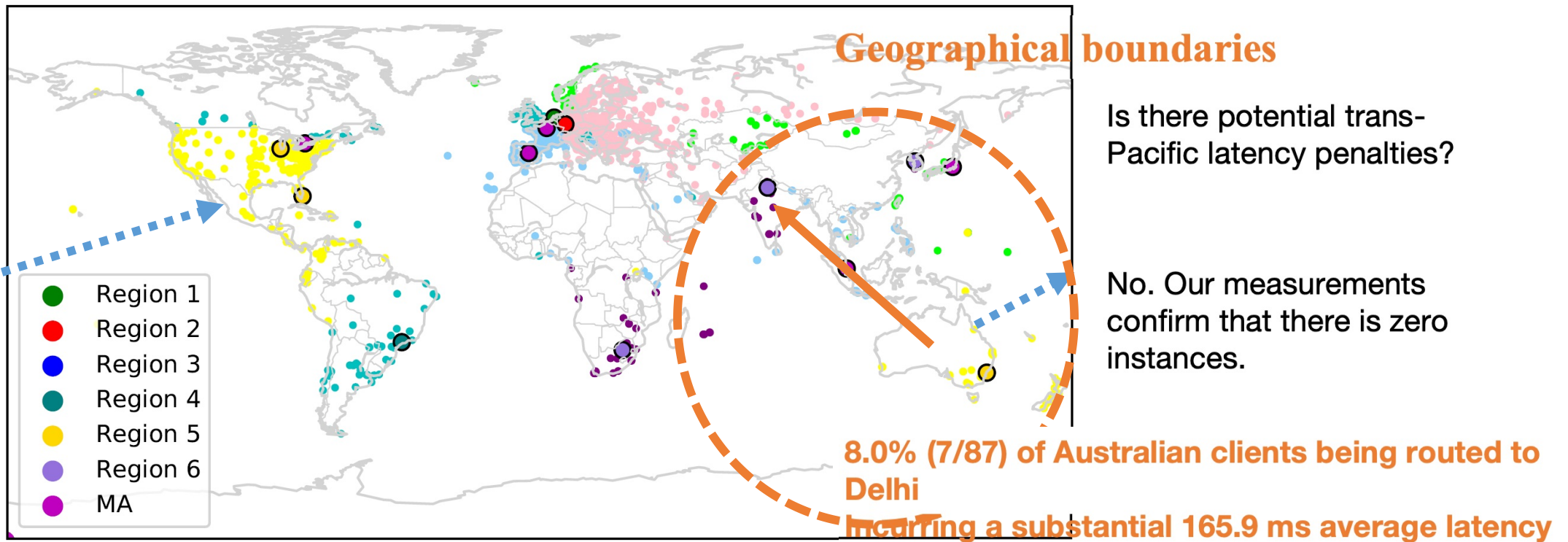
(b) *Distance difference.*



(a) *RTT difference.*

Translate to substantial latency improvements of average RTT.

Experiment results: ReOpt-6 deep dive to the improvements



- **Cluster clients with their empirically observed lowest-latency sites**
- **Strategically aggregate countries only when doing so preserves routing quality for affected clients**

Experiment results: ReOpt-6 deep dive to the improvements

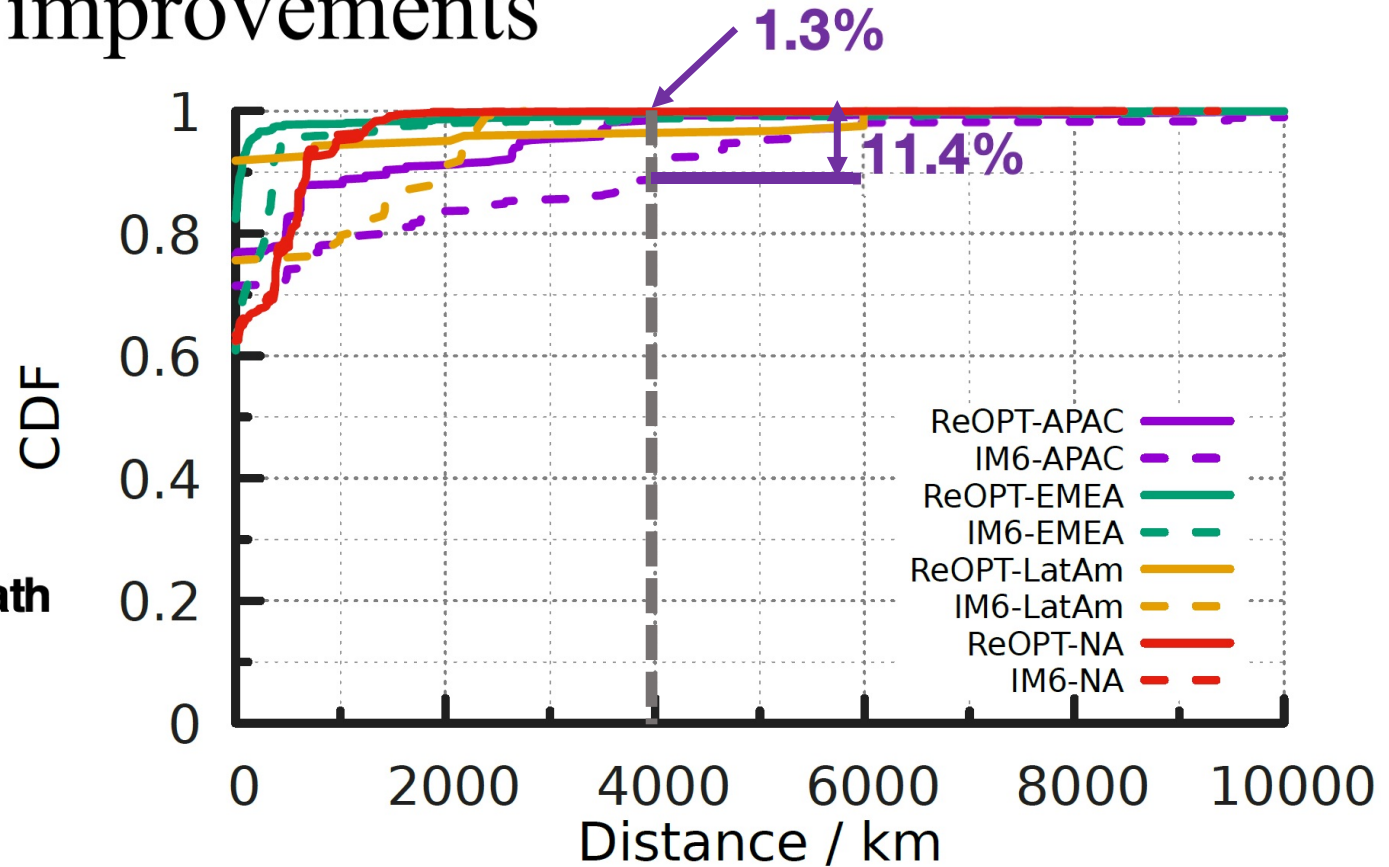


More clients connect
to their geo-closest
site



16.0% improvement

ReOpt reduces both **extreme path
inflation** and **tail latency**
through its intelligent client-site
matching strategy



More evaluations on the PEERING testbed can be found in
our paper.

(c) *Extra distance.*



Large-scale simulations

Table II: SIMULATION RESULTS

		(Number of Sites, Number of Clients)		
		(10, 200)	(30, 1000)	(200, 5000)
Total RTT	ReOpt	73.69 s	446.26 s	37.78 min
	OPT	70.88 s	493.24 s	Out of Memory
Running Time	ReOpt	0.70 s	9.10 min	8.49 h
	OPT	2.43 h	> 24 h	Out of Memory

- **Time Efficiency:** Achieve near-optimal objective function with much less running time
 $0.7s \ll 2.43h$

- **Scalability:** while problem size grows proportionally, runtime increases sub-linearly
 $\frac{200}{30} > \frac{5000}{1000} = \frac{500}{100}$ $\frac{8.49h}{9.10min} \ll \frac{9.10min}{0.70s}$

Hardware: Intel i5-12400F CPU with 32 GB RAM; Optimizer: Gurobi; Data: randomly