My 1st RIPE Meeting

RACI thanks!

Grant Agreement n. 318627

Where are the anycasters, reloaded

Dario Rossi
Professor
dario.rossi@telecom-paristech.fr
http://www.telecom-paristech.fr/~drossi/anycast

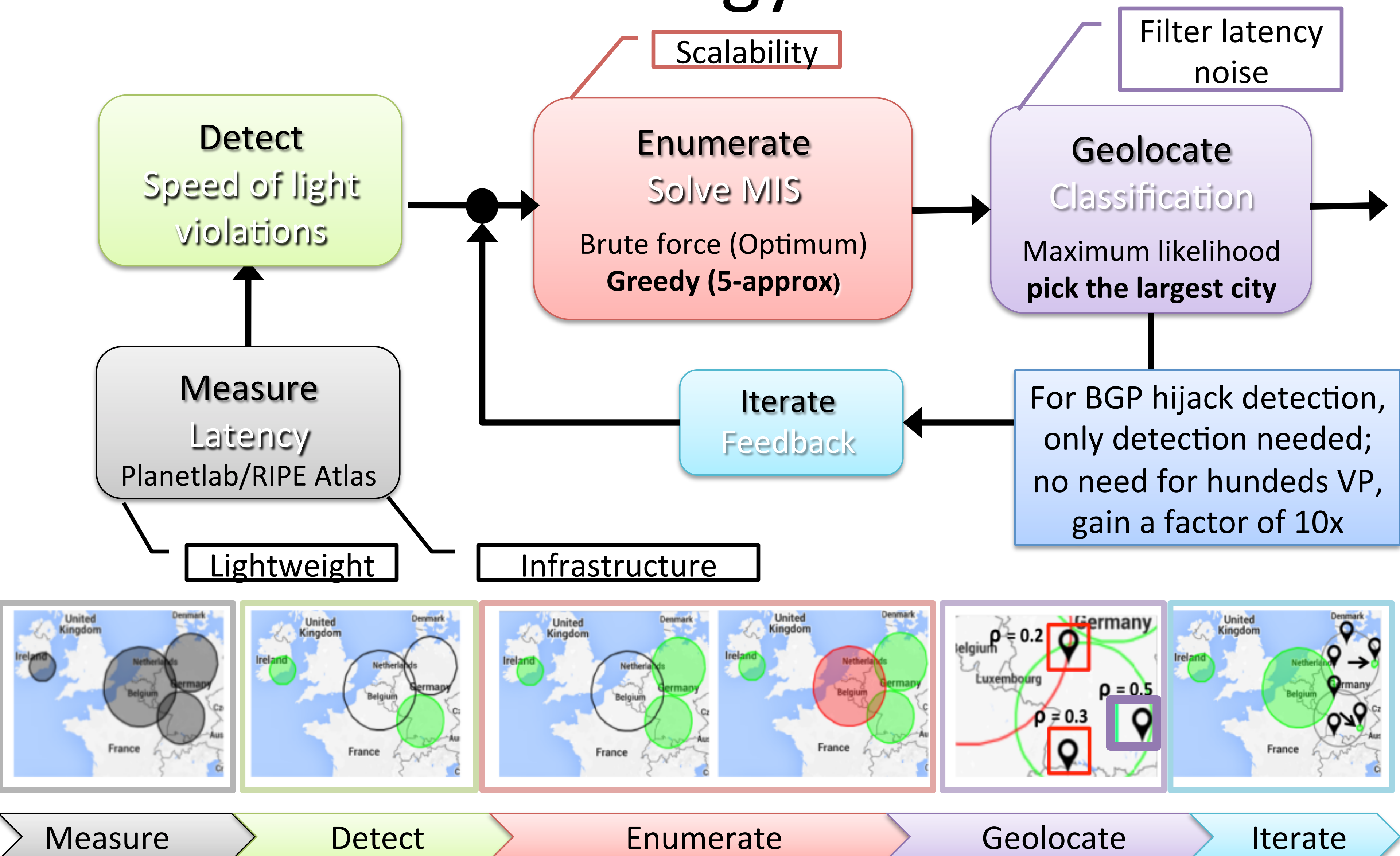
Joint work with Danilo Cicalese, Diana Joumblatt, Jordan Auge and Timur Friedman

MAT WG Talk

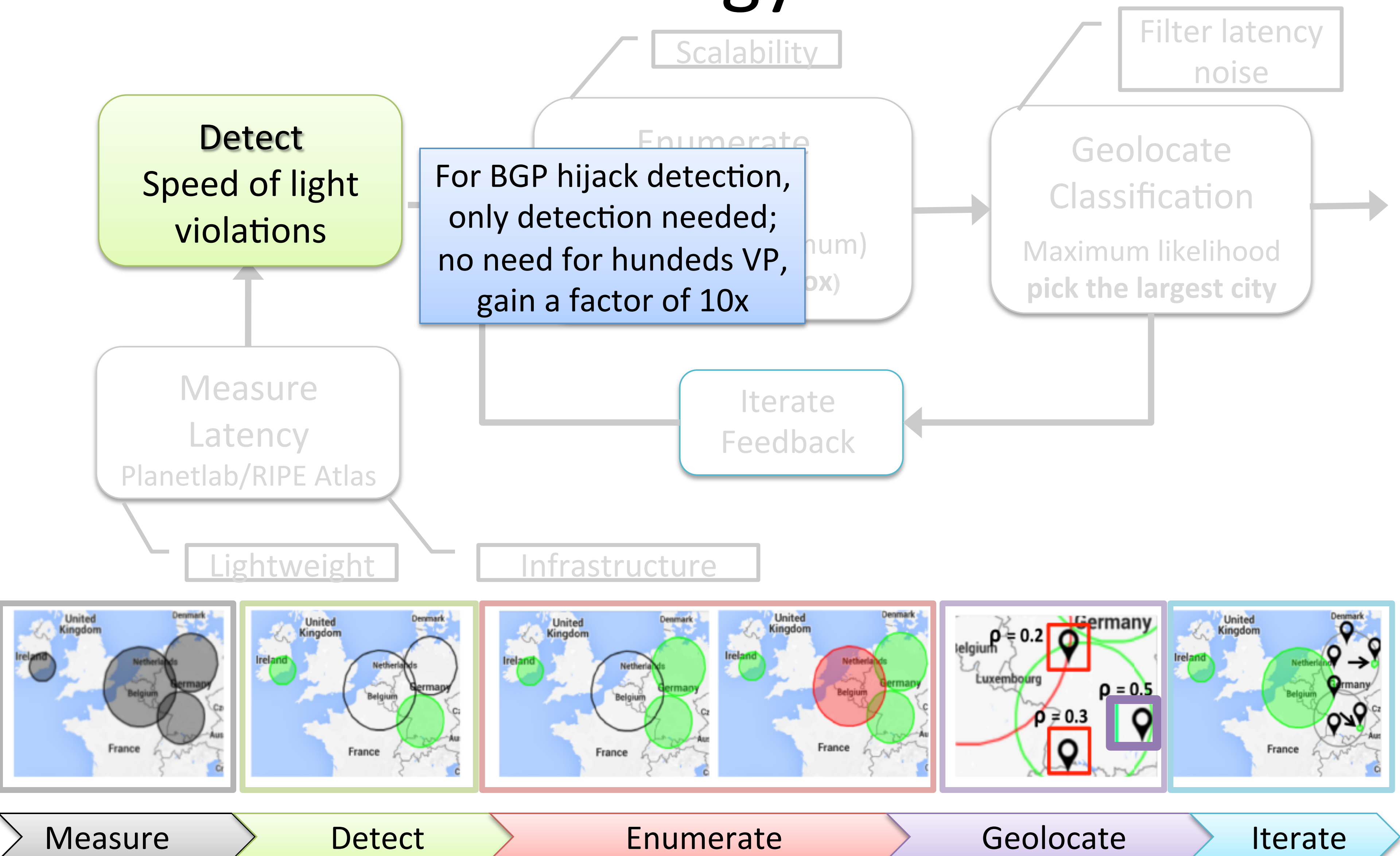
Agenda

- Aim
  - Informal talk; Keep this interactive
- Metodology refresh
- Results, reloaded
  - O(1) deployments (with ground truth): Details about geolocalization technique
  - O(10^7) deployments: The dark side of the census (from an Internet scan in few *hours* to few *minutes*)
- Note
  - Many thinks (software, web interface, etc.) already available, we of course have more (ask if interested) http://www. telecom-paristech.fr/~drossi/anycast

Methodology refresh

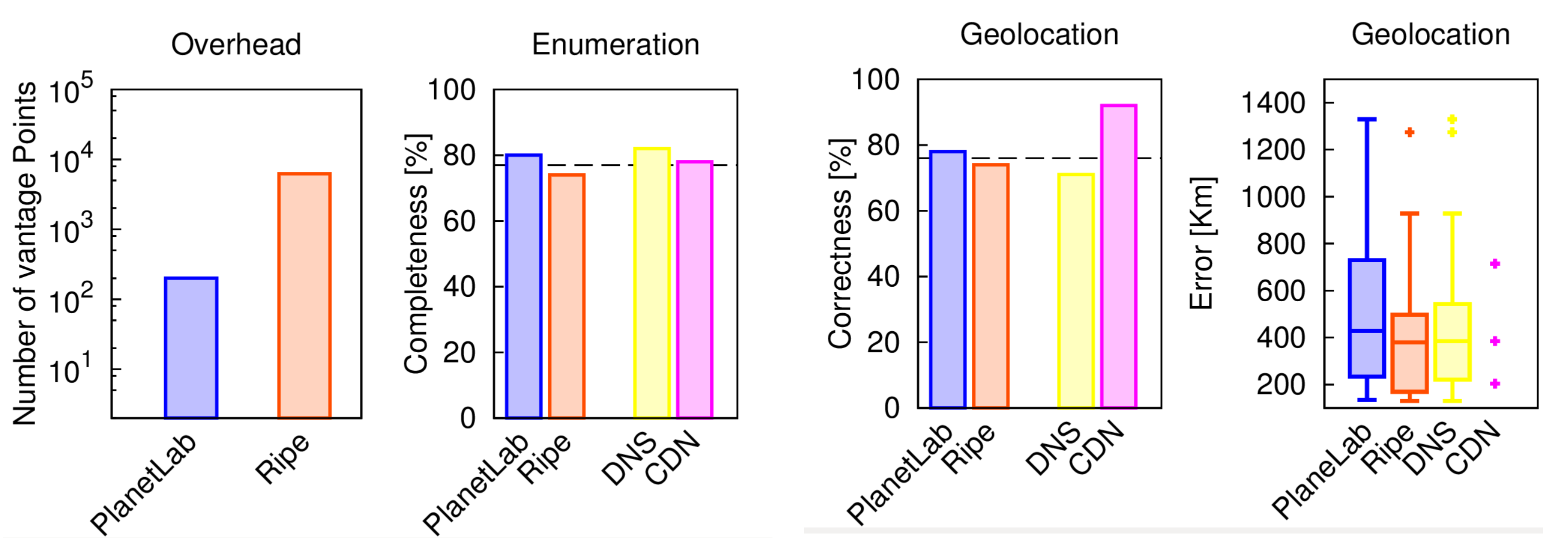


Methodology refresh

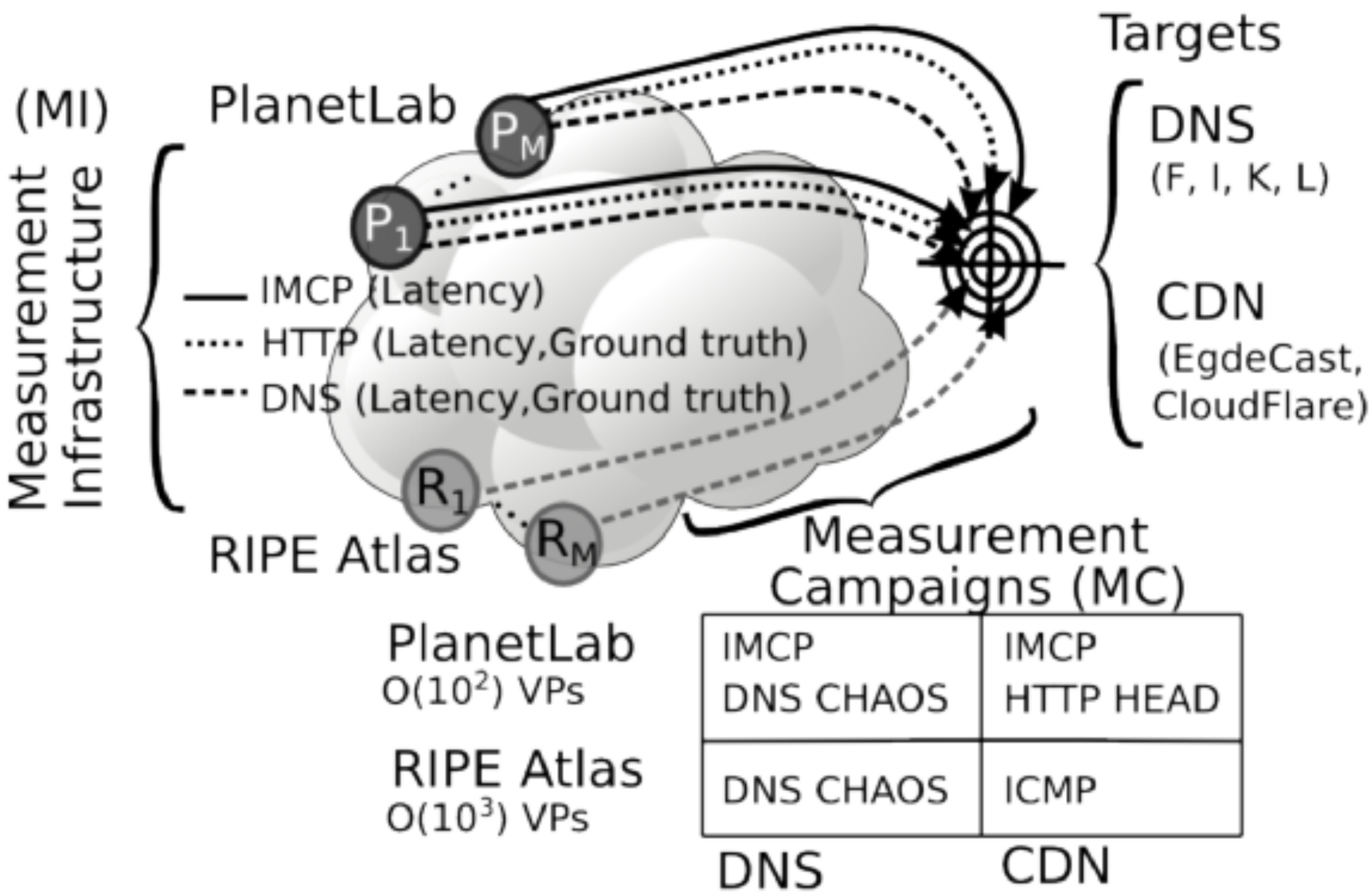


iGreedy performance

- **Accurate enumeration** over 75% recall
- **Precise geolocation** over 75% true positives
- **Protocol agnostic** DSN and CDN, etc.
- **Lightweight** 100x less probes than previous work
- **Consistent** across measurement infrastructure
- **Robust** in spite of very noisy latency measurements! (and some odd VP geolocation)

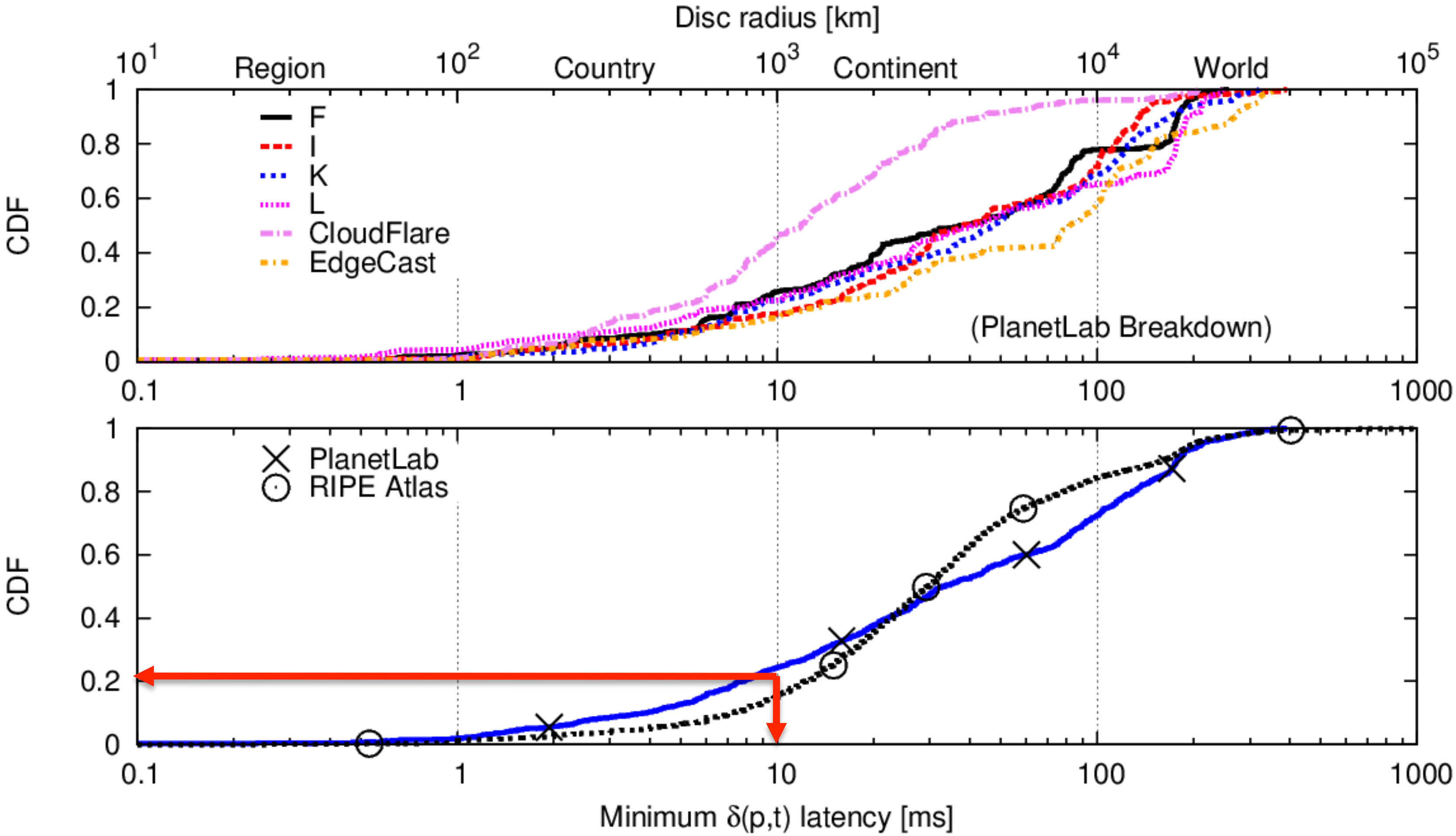


Measurement campaigns



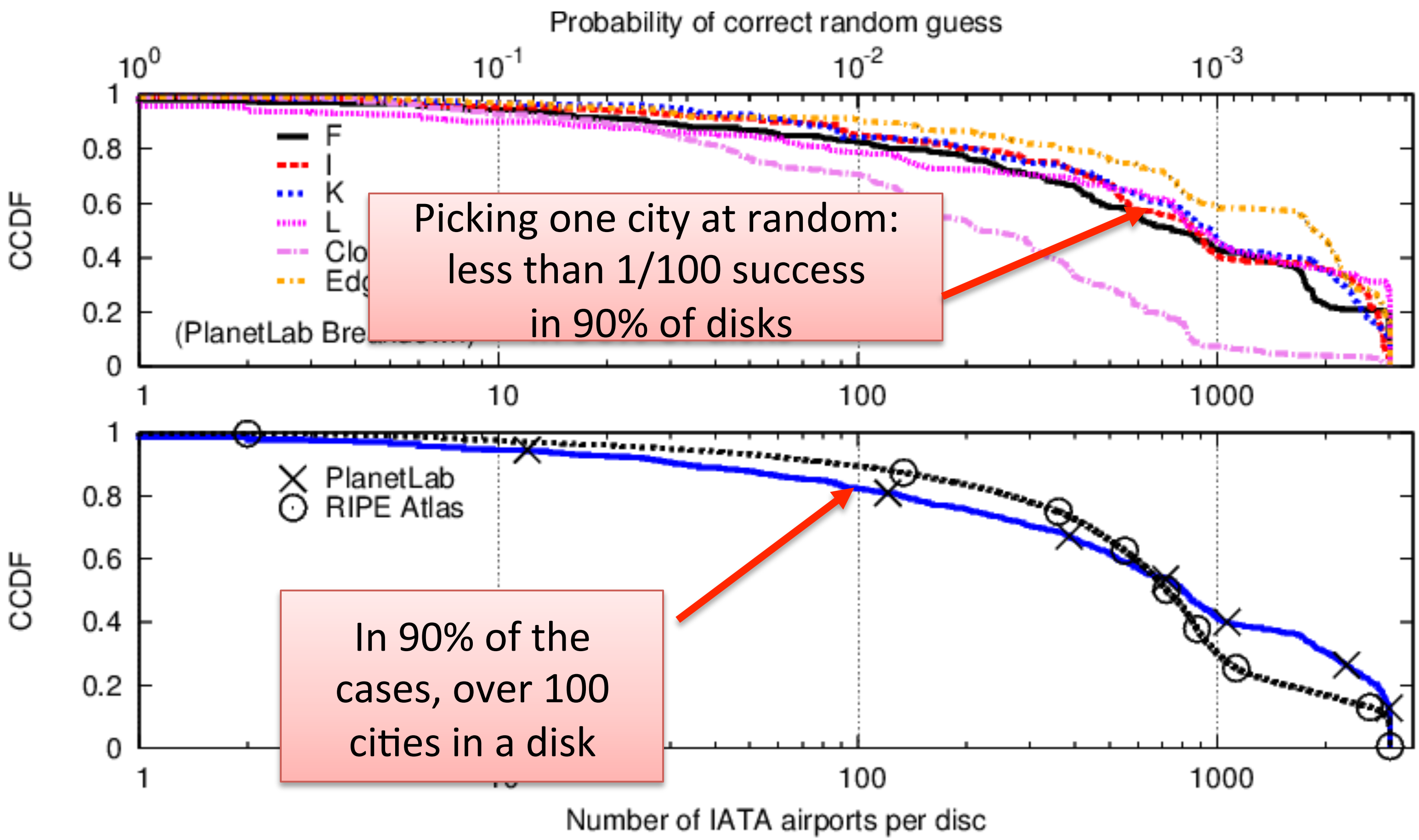
Experimental results in [4] gathered with open source software and dataset
Avail in the igreedy-v1.0 software package at [3]

Measurement campaign (1/2)



Very noisy delay measurements. Only 10% of disks are smaller than 1000km !!

Measurement campaigns (2/2)



Delay information: useful for enumeration, bad for geolocation !!

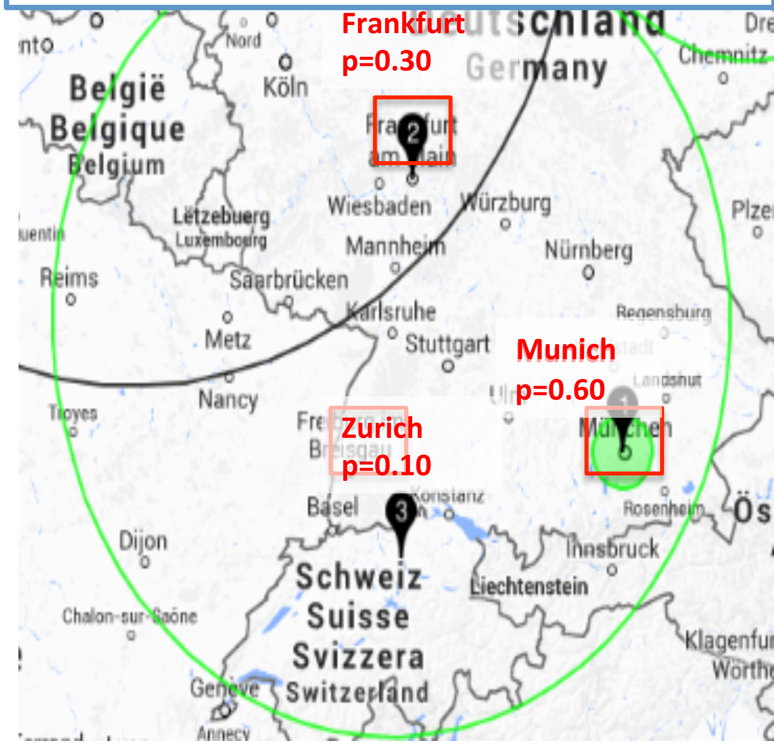
Geolocation

- Classification task
  - Map each disk Dp to most likely city
  - Compute likelihood (p) of each city in disk based on:
    - ci: Population of city *i*
    - Ai: Location of ATA airport of city *i*
    - d(x,y): Geodesic distance
    - α: city vs distance weighting

$$p_i = \alpha \frac{c_i}{\sum_j c_j} + (1-\alpha) \frac{d(p,t) - d(p,A_i)}{\sum_j d(p,t) - d(p,A_j)}$$

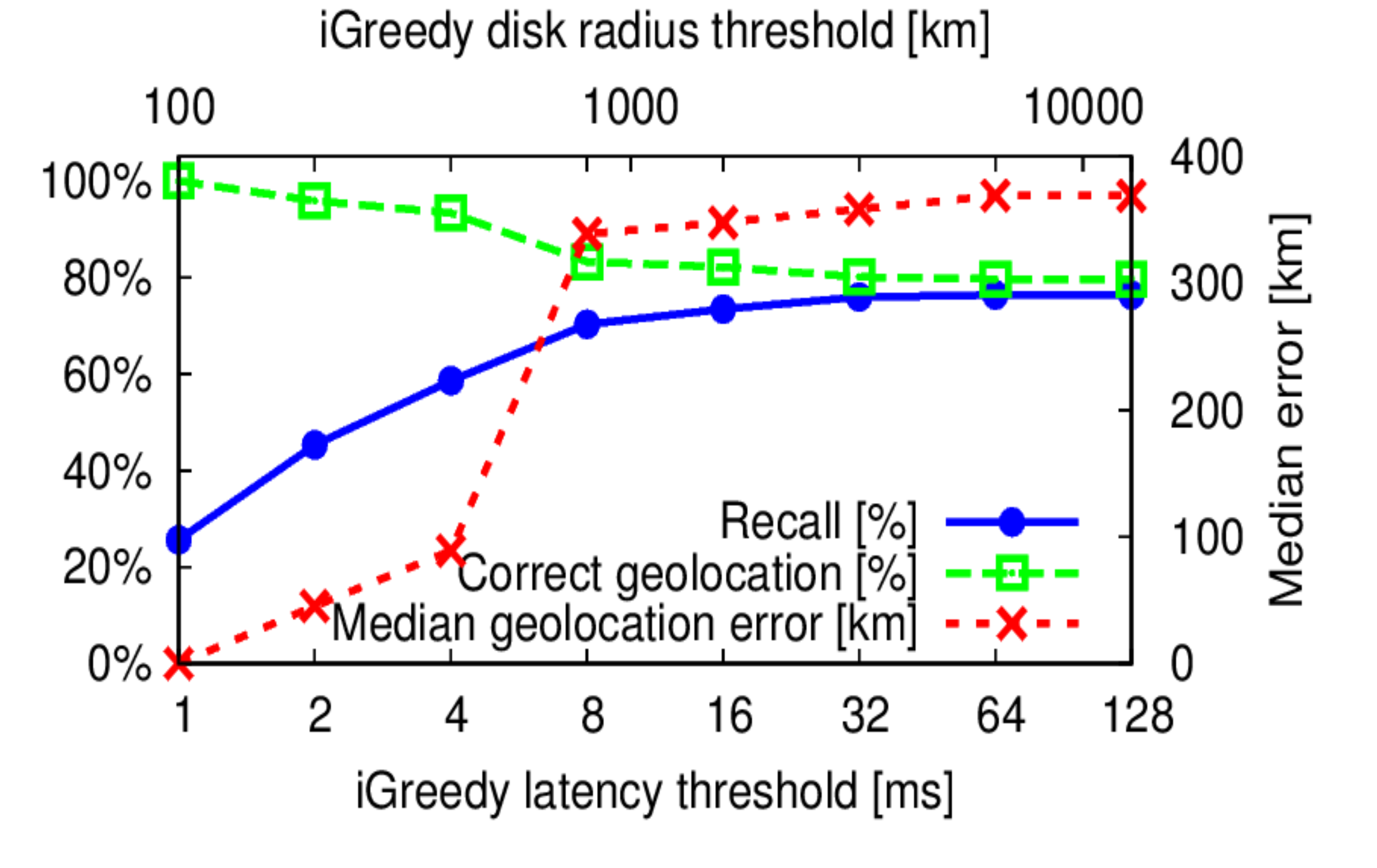
- Output policy
  - **Proportional**: Return all cities in D_p with respective likelihoods
  - **Argmax**: Pick city with highest likelihood

rationale: users lives in densely populated area; to serve users, servers are placed close to cities



In practice, pick the largest city is best ! (Argmax with α=1)

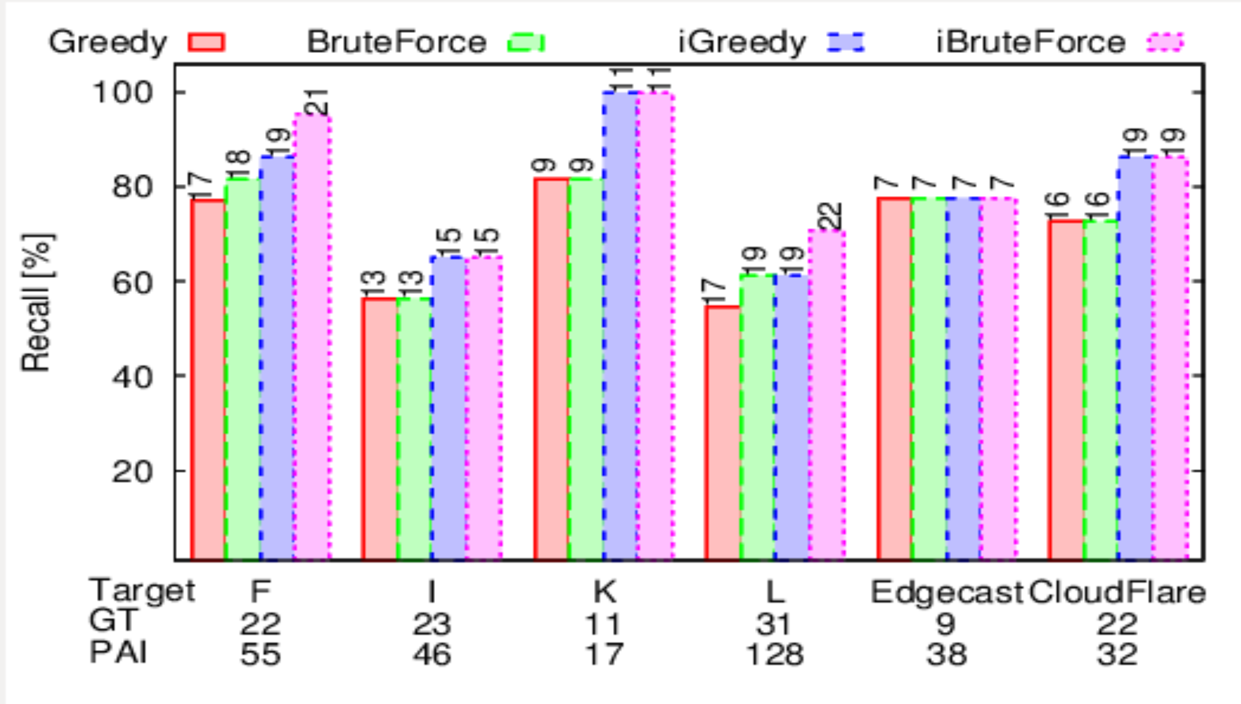
iGreedy performance: robustness



Not even need for filtering large disks, as iGreedy sorts disk
by increasing size, bad disks implicitly filtered out in the solution!!

iGreedy performance: MIS solver

- MIS performance:
  - In theory: Greedy = 5x-approximation of global optimum
  - In practice: Greedy solution ≈ Brute Force solution
  - Iteration introduces a significant benefit
  - O(100ms) greedy vs O(1000sec) brute force (for ~300 nodes)



In practice,
greedy
is good enough

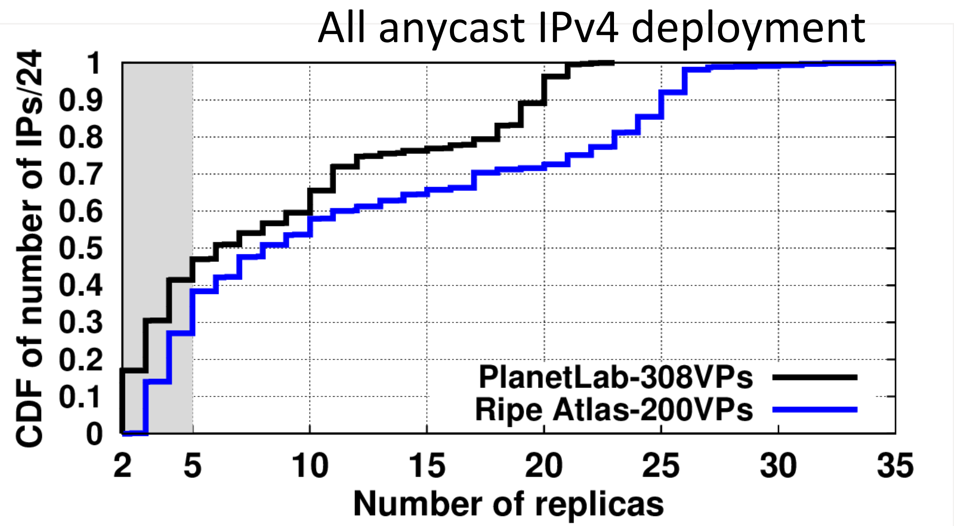
Vantage points impact (1/3)

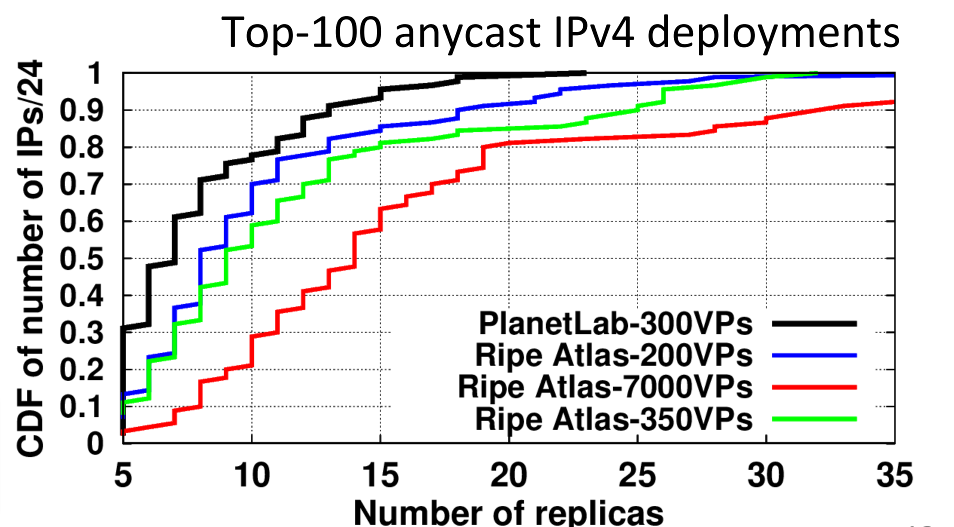
-

Footprint	VPs	ASes	Country
RIPE Atlas (all)	7k	2k	150
RIPE Atlas (subset)	200	139	83
PlanetLab	~300	180	30

- **RIPE7k** = full (at time of experiments)
  - Greater coverage, but artifacts due to geolocation inaccuracy
- **RIPE200** = selection of 200 VPs at least 100km far apart
  - Better than PlaneLab for fewer VPs
- **RIPE350** = 350VPs that have geolocation tags + those VP that yielded true positive for CDN/DNS
  - Best for RIPE
- **Union** of RIPE and PlanetLab
  - Even better

Union makes the force!





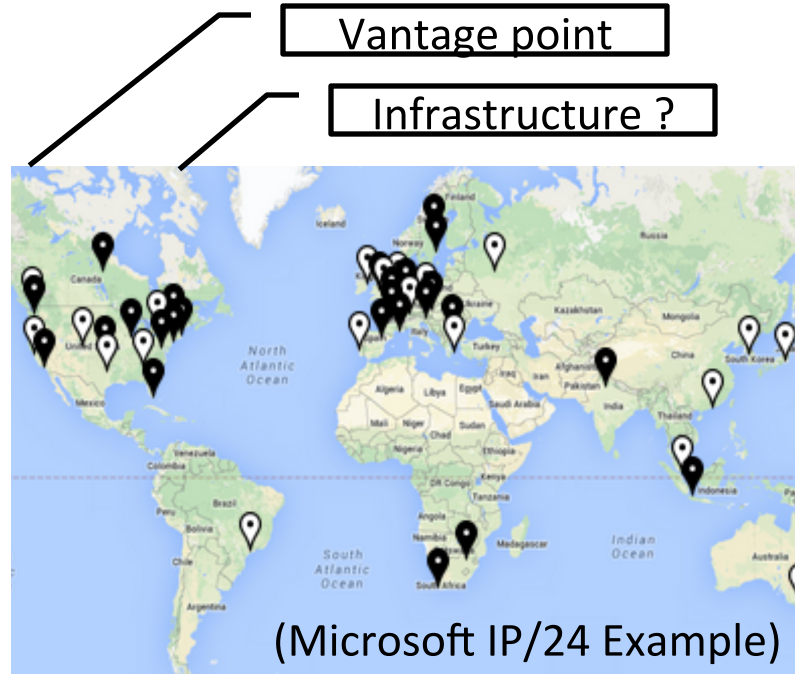
Vantage points impact (2/3)

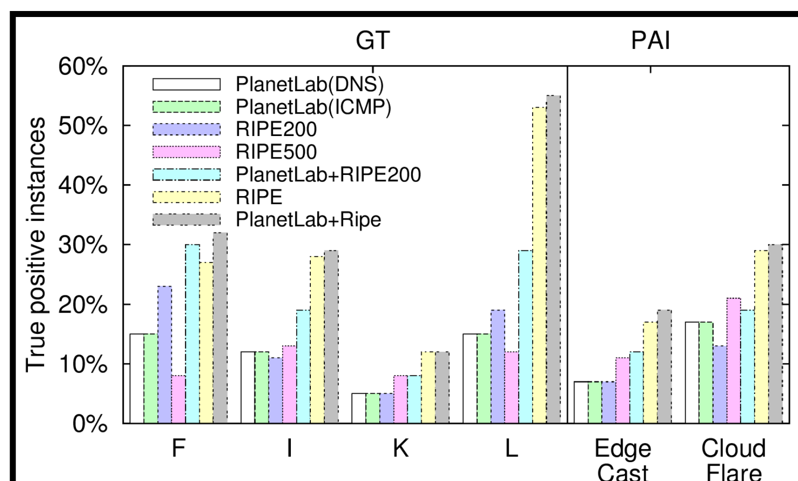
-

Footprint	VPs	ASes	Country
RIPE Atlas (all)	7k	2k	150
RIPE Atlas (subset)	200	139	83
PlanetLab	~300	180	30

- **RIPE7k** = full (at time of experiments)
  - Greater coverage, but artifacts due to geolocation inaccuracy
- **RIPE200** = selection of 200 VPs at least 100km far apart
  - Better than PlaneLab for fewer VPs
- **RIPE350** = 350VPs that have geolocation tags + those VP that yielded true positive for CDN/DNS
  - Best for RIPE
- **Union** of RIPE and PlanetLab
  - Even better

Not always :(



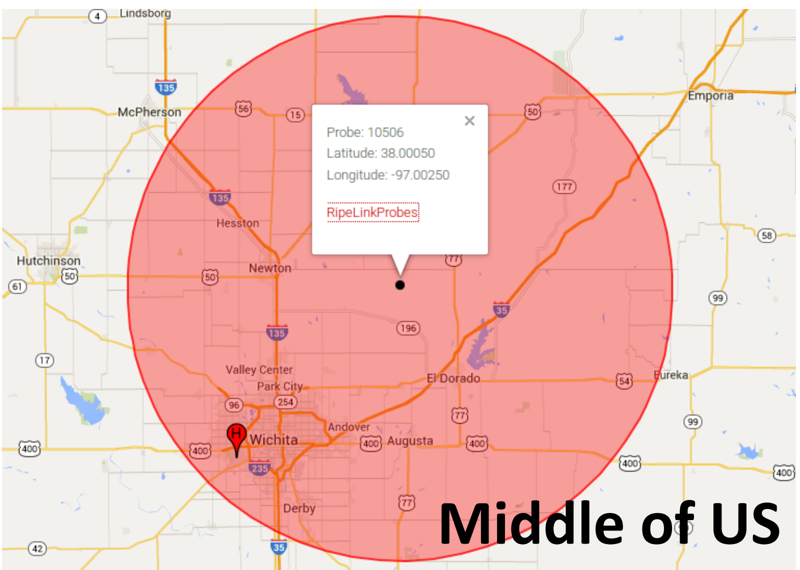


Vantage points impact (3/3)

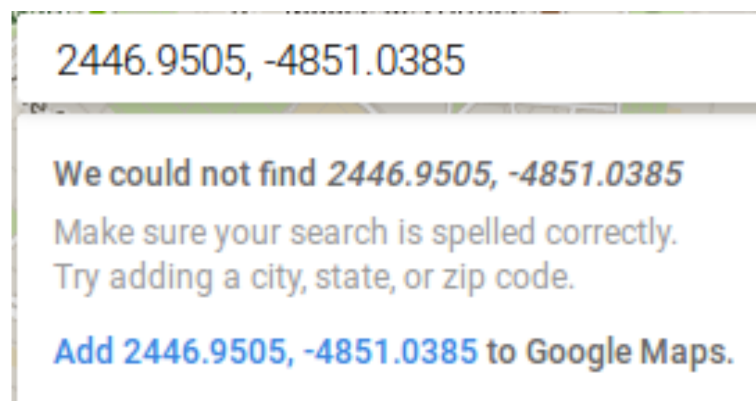
- The owner of the VP sets the geolocation
- 500 VP have a tag *system-auto-geoip-city*
- Can we trust it?
- Only 350 are 100 km apart

Pretty messy. We also have picteures of PlanetLab nodes in Navajo reserves or swimming in the ocean

Romanian VP gone hiking



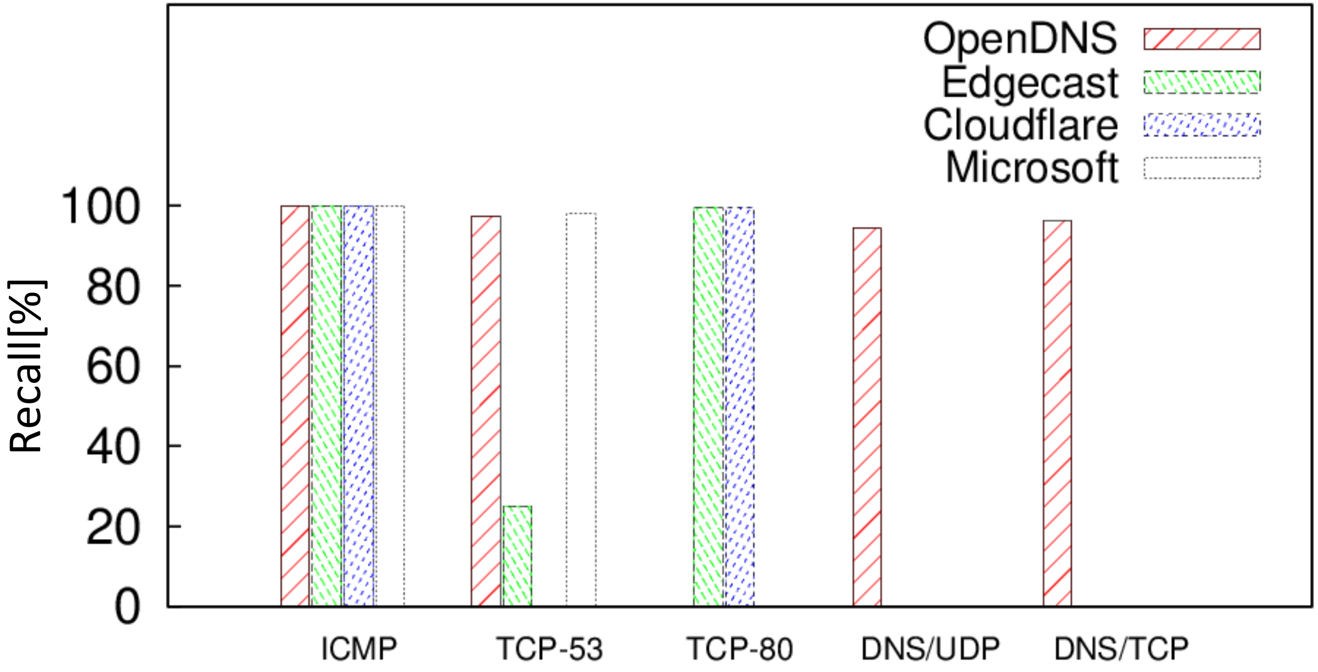
Middle of US



Middle of Nowhere

Protocol impact

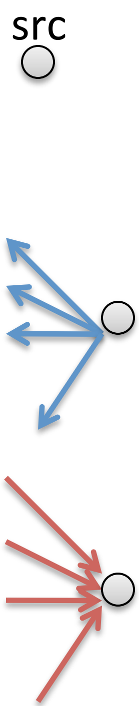
- If multiple protocols answer, no noticeable difference in the output, however



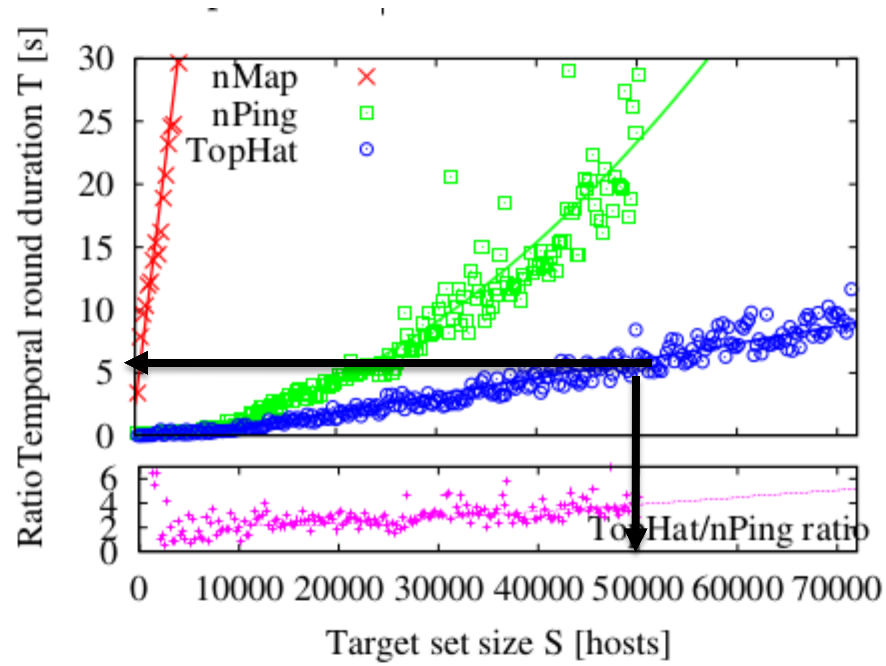
ICMP service agnostic, maximizes (*) reply

(*) CloudFlare stopped replying our pings :(

Probing rate



- Our tool can scan more than 10k hosts/sec per src
- We spread destinations with a linear shift register so to maximize interarrival time between two ICMP echo request at destination



- However, ICMP echo reply aggregate at the source, that receives about 10k replies/sec
- Some ingress firewalls in PlanetLab nodes think this is an attack and rate limits
- No problem with 1k requests/sec

For BGP hijack detection, removing ICMP filters yields a 10x speedup !

``Double'' ping

- (In lack of a better name)
- Idea
  - try to discover something about penultimate hop without performing a full traceroute (be fast and simple)
- Means
  - on reception of an ICMP echo reply X with TTLx
  - issue one ICMP echo request Y limiting the TTLy to the closest power of 2 to X – TTLx -1
  - (i.e., expected number of hops minus one)
- Question
  - Do you expect any relevant info about the penultimate AS in the path, that we can leverage for BGP hijack detection?

Opportunistic & limited traceroute

Conclusions

- iGreedy novel technique to investigate and especially geolocate anycast deployment
  - ✓ Practical lightweight, fast and protocol agnostic
  - ✓ Ready open-source software to issue, analyze and display RIPE Atlas measurement (using your credits!)
  - ✓ Useful Web interface to (significant subset of) census results already available

Interested ? Drop an email dario.rossi@enst.fr !
(but cc danilo.cicalese@enst.fr to get a timely reply)

References

Plenary talk:

[1] D. Cicalese , D. Joumblatt, D. Rossi, J, Auge, M.O Buob, T. Friedman. **A Fistful of Pings: Accurate and Lightweight Anycast Enumeration and Geolocation** , IEEE INFOCOM, 2015

[2] D. Cicalese , J. Auge, D. Joumblatt, T. Friedman, D. Rossi, **Characterizing IPv4 Anycast Adoption and Deployment** , ACM CoNEXT, Dec 2015

[3] http://www.telecom-paristech.fr/~drossi/anycast

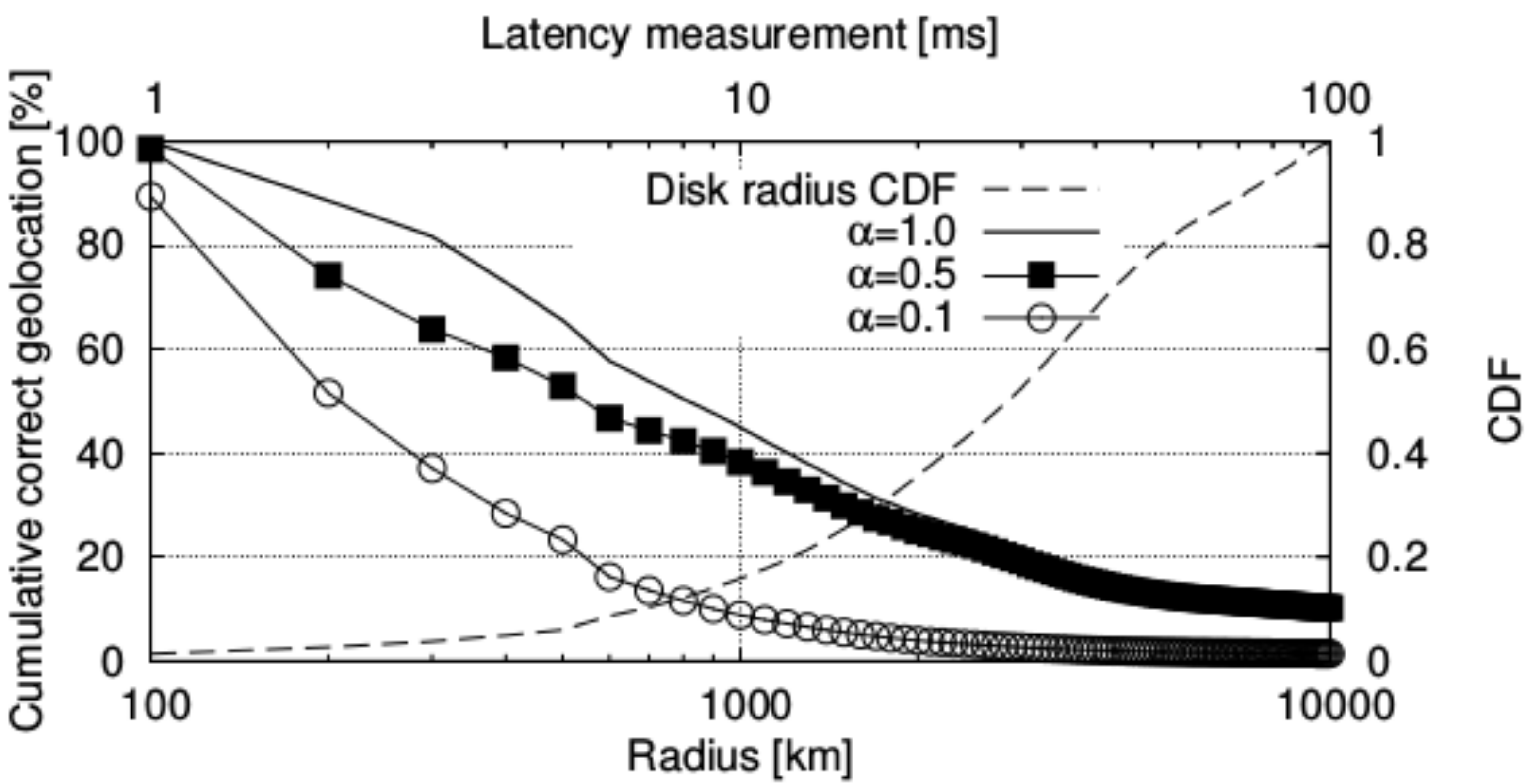
[4] Cicalese, Danilo, Joumblatt, Diana , Rossi, Dario, Buob, Marc-Olivier , Auge, Jordan and Friedman, Timur , **Latency-Based Anycast Geolocalization: Algorithms, Software and Datasets** . In *Tech. Rep.*, 2015.

?? || //

!! After this slides, too many backup slides

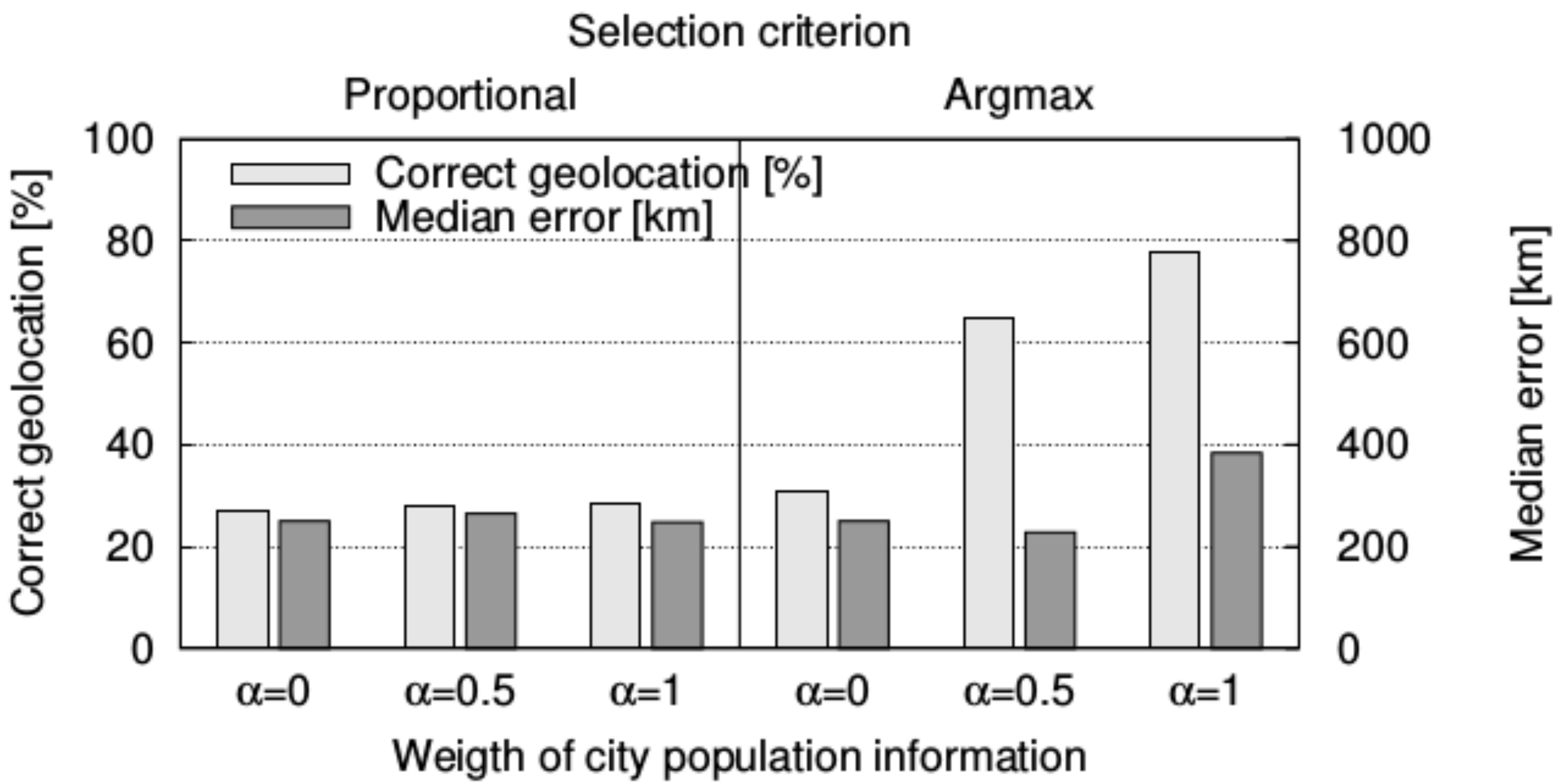
Reverse cronological order:
•[4] Tech Rep: Why the largest city ?
•[2] CoNEXT: Why ICMP ? What about CDN? Duration?
•[1] INFOCOM: Methodology details amd comparison

TechRep/ Why largest city



(a) Geolocalization over individual disks (RIPE, PlanetLab)

TechRep/ Why largest city



(b) iGreedy sensitivity (PlanetLab)

CoNEXT/ CDN Performance

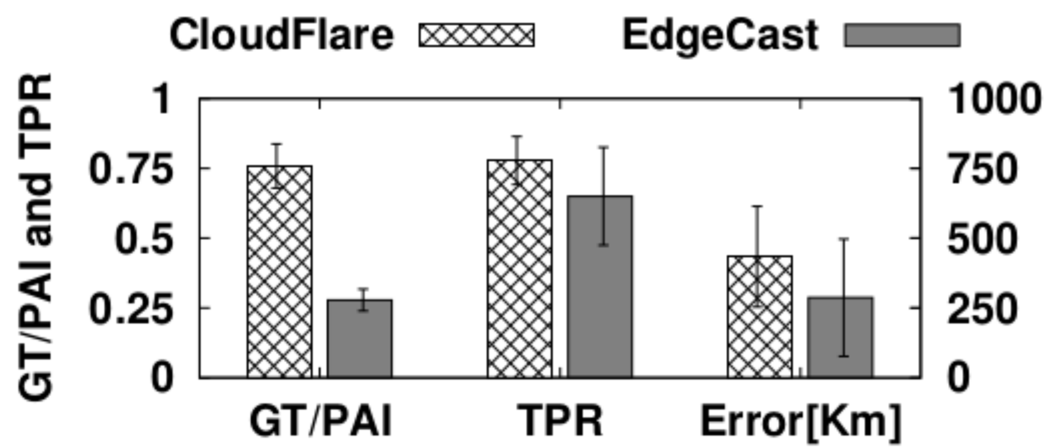


Figure 7: Validation with CloudFlare and Ed
Cast ASes. Bars represent standard deviati
among IP/24 of the AS.

CoNEXT/Why ICMP

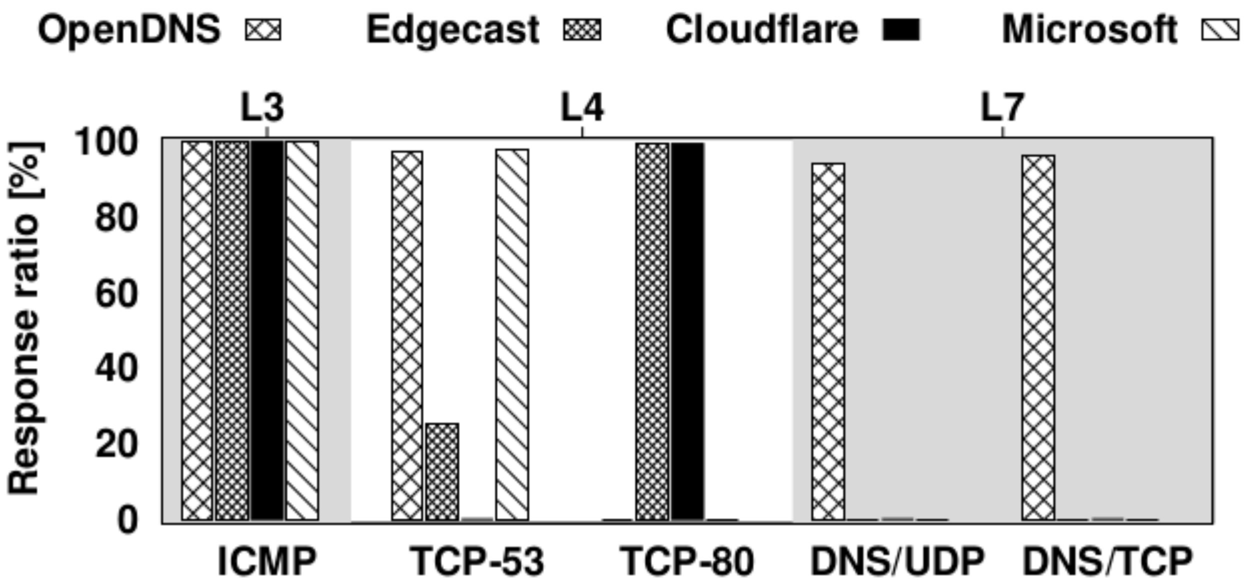


Figure 6: Response rates seen by heterogeneous protocols across different targets.

CoNEXT/Duration

Table 1: Textual (0) vs binary (1-4) censuses

Census ID	Format	Size (host,total)	Analysis
0	csv	(270M, 79G)	>3 days
1-4	binary	(21M, 6G)	3 hr

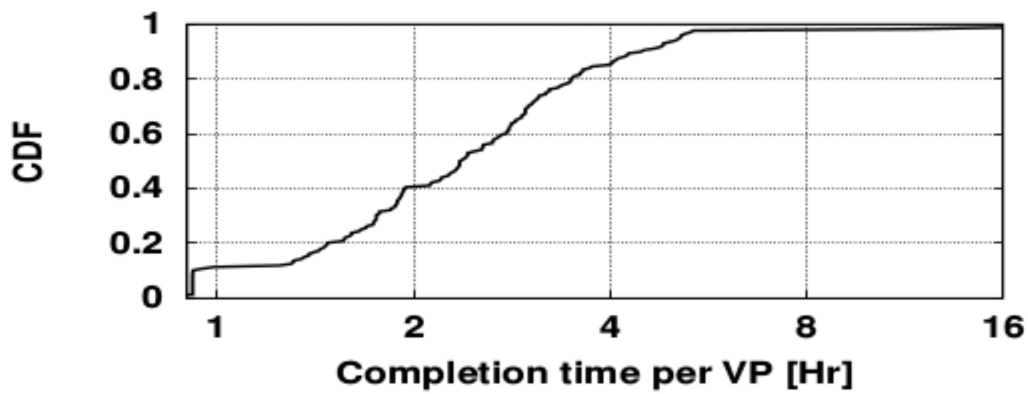
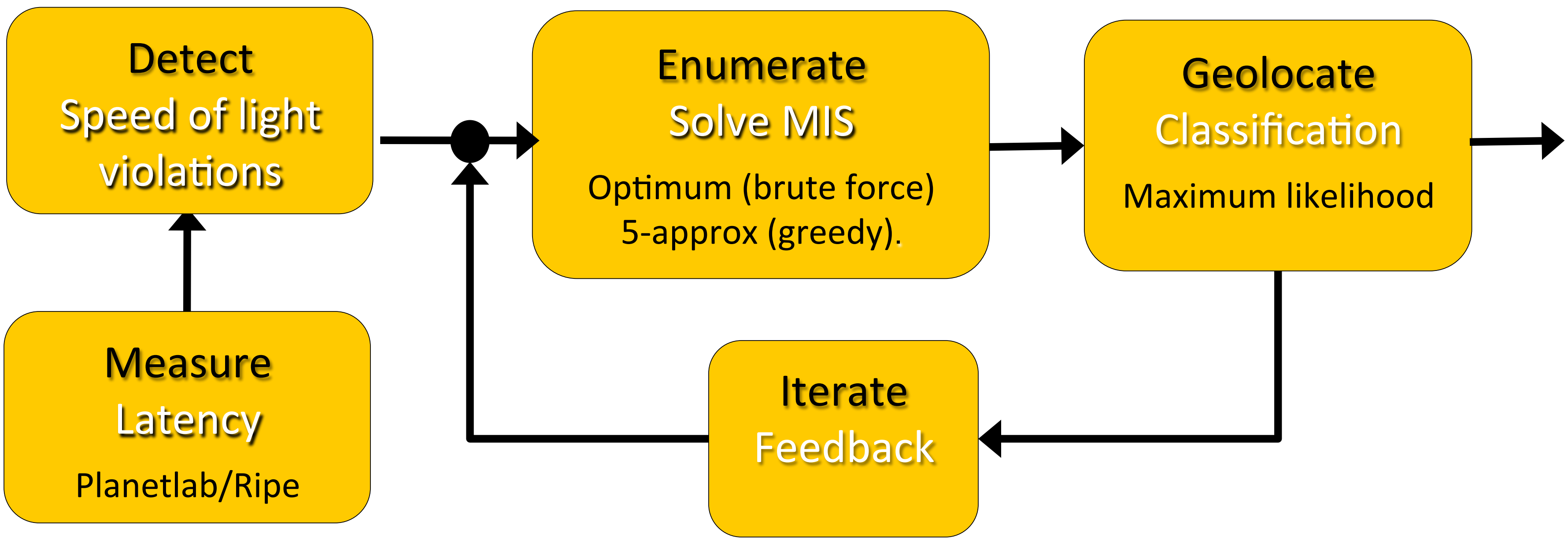


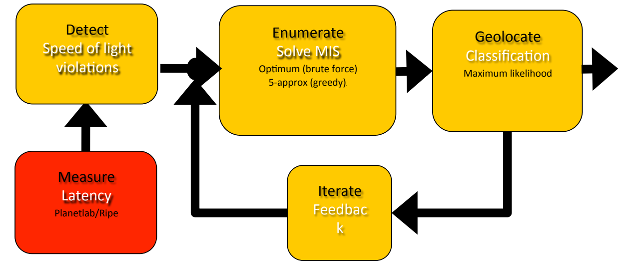
Figure 8: CDF of per-vantage point completion time, over all censuses

INFOCOM

Methodology overview



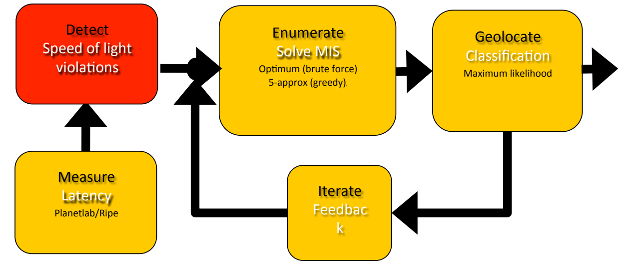
Measure

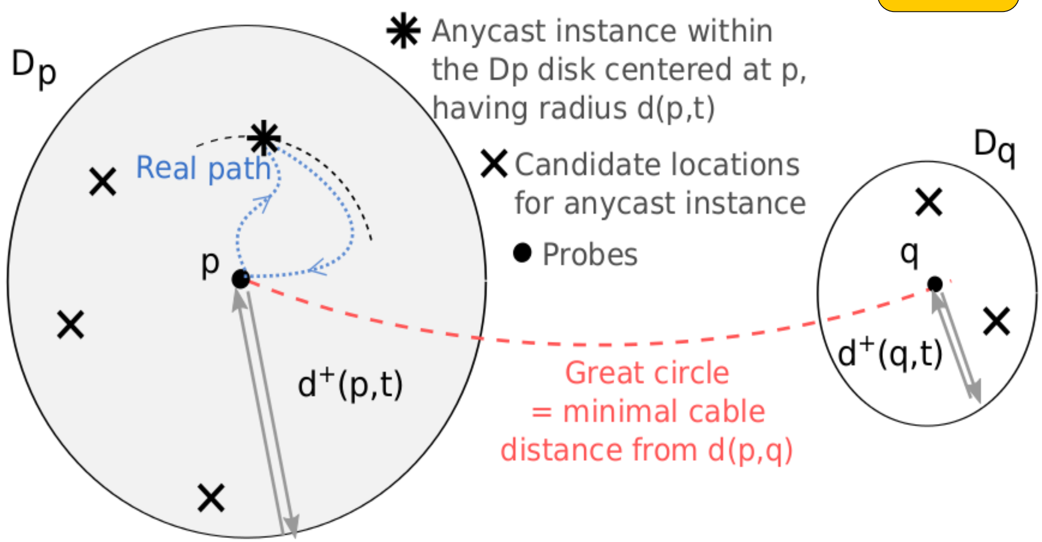


- PlanetLab
  - 300 vantage points
  - Geolocated with Spotter (ok for unicast)
  - Freedom in type of measurement
    ICMP, DNS, TCP-3way delay, etc
- RIPE
  - 6000 vantage points
  - Geolocated with MaxMind (ok for unicast)
  - More constrained (ICMP, traceroute)

In this talk: min over 10 ICMP samples

Detect



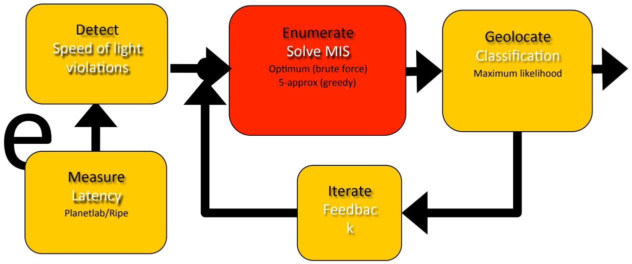


The vantage points **p** and **q** are referring to two different instances if:

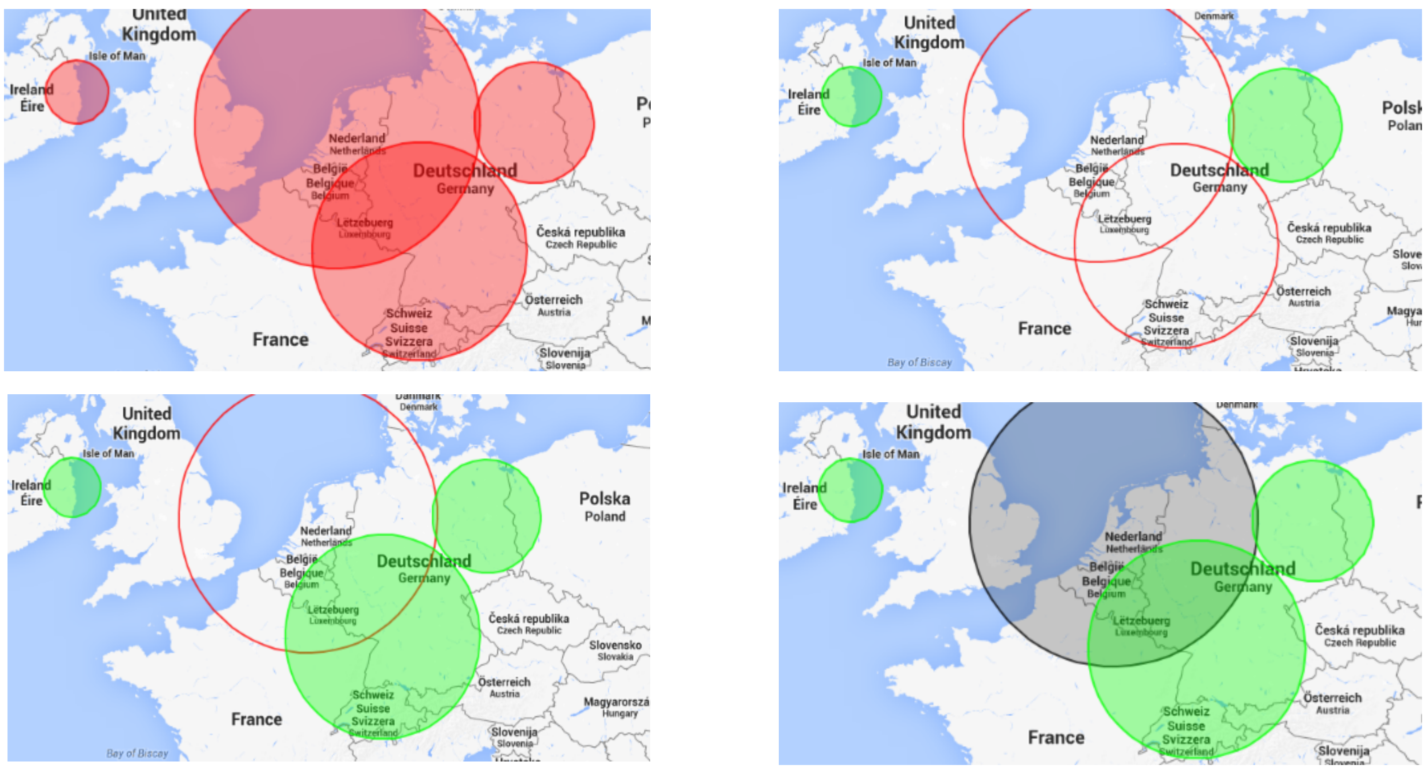
$$d(p;q) > d(p;t) + d(q;t)$$

Packets cannot travel faster than the speed of light

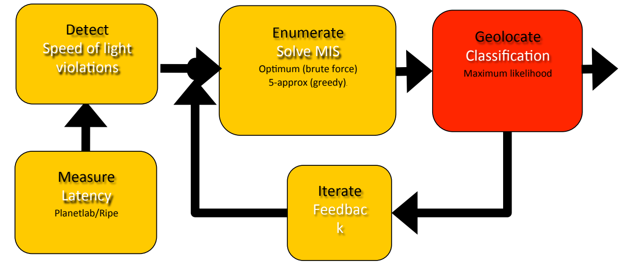
Enumerate



- Find a maximum independent set $\mathcal{E}$
  - of discs such that: $\forall \mathcal{D}_p, \mathcal{D}_q \in \mathcal{E}, \mathcal{D}_p \cap \mathcal{D}_q = \emptyset$
  - Brute force (optimum) vs Greedily from smallest (5-approximation)

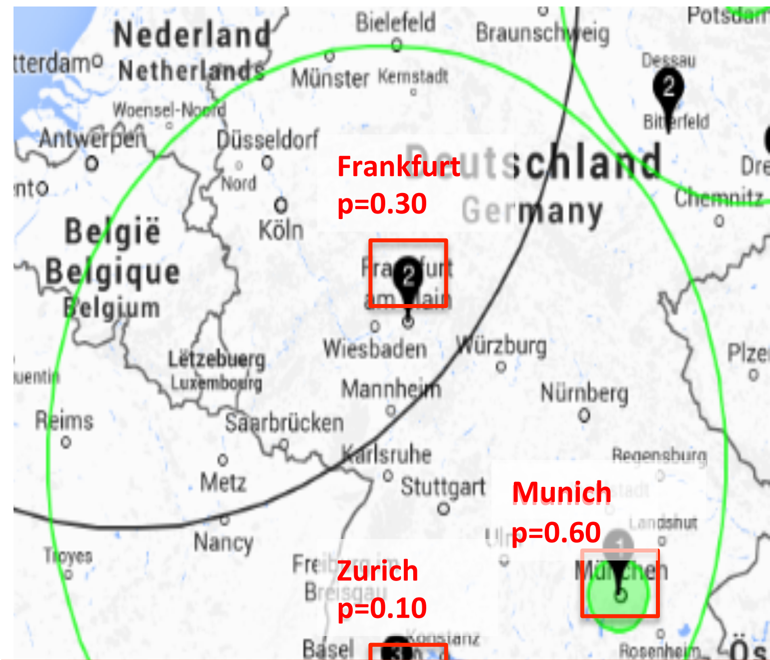


Geolocate



- Classification task
  - Map each disk Dp to most likely city
  - Compute likelihood (p) of each city in disk based on:
    - ci: Population of city *i*
    - Ai: Location of ATA airport of city *i*
    - d(x,y): Geodesic distance
    - α: city vs distance weighting

$$p_i = \alpha \frac{c_i}{\sum_j c_j} + (1-\alpha) \frac{d(p,t) - d(p,A_i)}{\sum_j d(p,t) - d(p,A_j)}$$

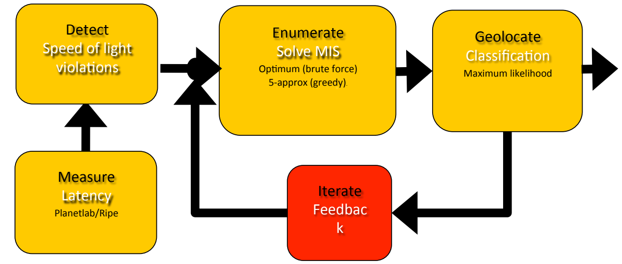


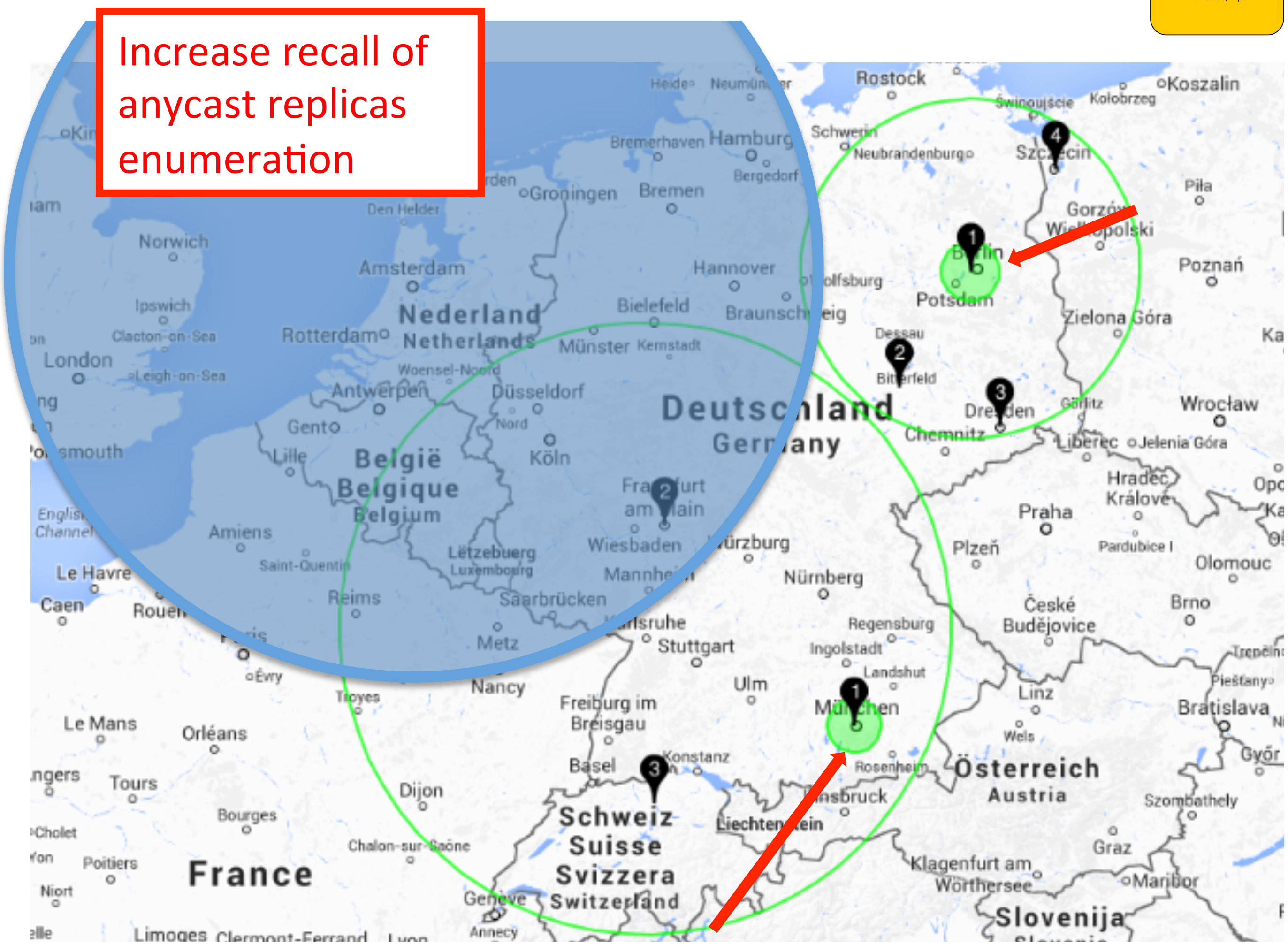
- Output policy
  - **Proportional**: Return all cities in D_p with respective likelihoods
  - **Argmax**: Pick city with highest likelihood

rationale: users lives in densely populated area; to serve users, servers are placed close to cities

airports: simplifies validation against ground truth (DNS)

Iterate





- Collapse
  - Geolocated disks to city area
- Rerun
  - Enumeration on modified input set

Validation

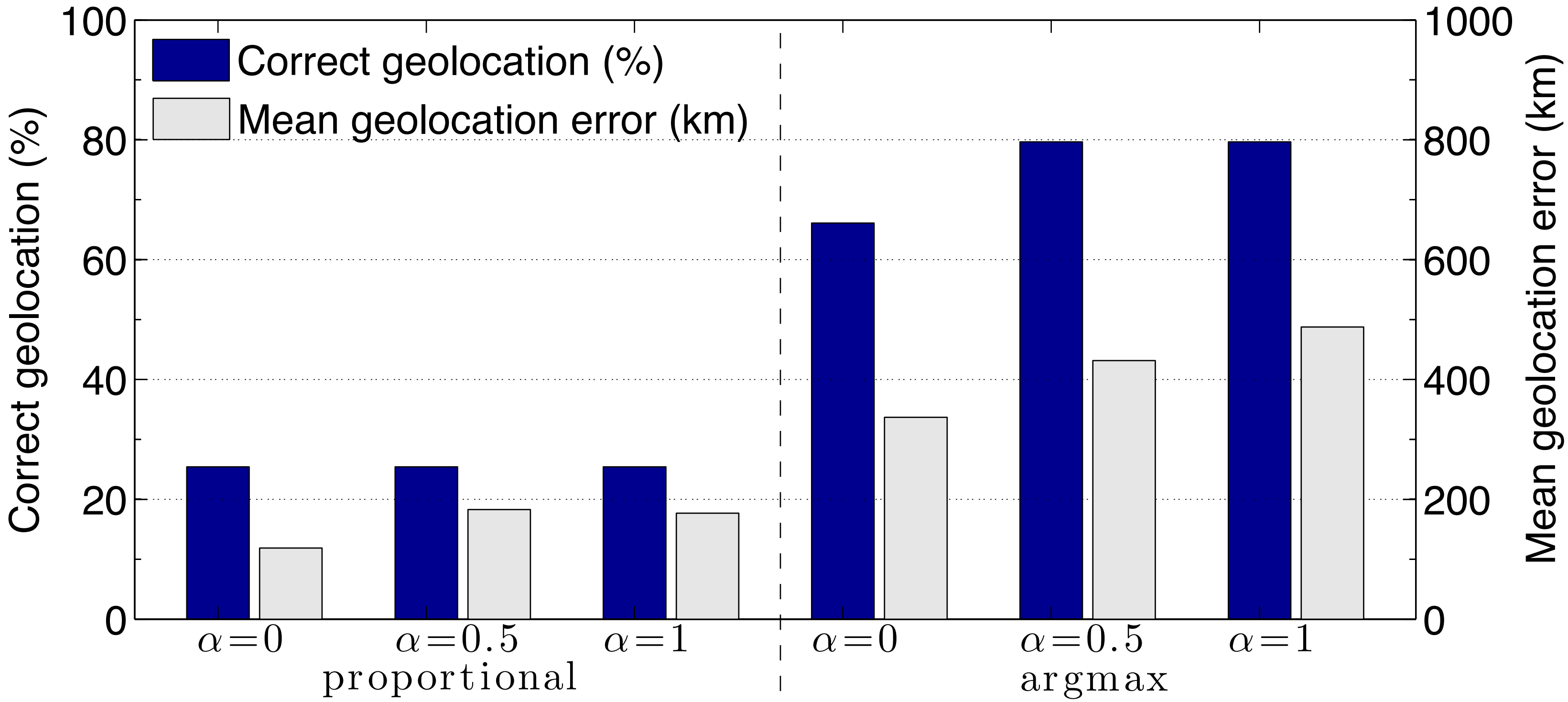
- Validation dataset
  - 200 VPs from PlanetLab
  - DNS CHAOS queries
  - DNS root servers F, K, I, L
  - City-level ground truth about sever location available
    - E.g., IATA code, IXPs short names
- Enumeration
  - Impact of solver (Brute force vs. greedy, Iteration)
- Geolocation
  - Impact of geolocation parameters (city vs distance weight, policy)

Validation: Enumeration

Algorithm	F	I	K	L
Greedy	17	13	9	20
BruteForce	18 +6%	13 -	9 -	20 -
iGreedy	18 +6%	15 +15%	10 +11%	22 +10%
iBruteForce	21 +23%	15 +15%	10 +11%	22 +10%
Dataset CHAOS UB	22	23	11	33
Published GT	55	46	17	128

Recall of iterative greedy solver as good as brute force
iGreedy solver O(100ms) faster than brute force O(1000s)

Validation: Geolocation



Argmax better than proportional
Distance from disc border and city population have similar weights

Measurement campaign

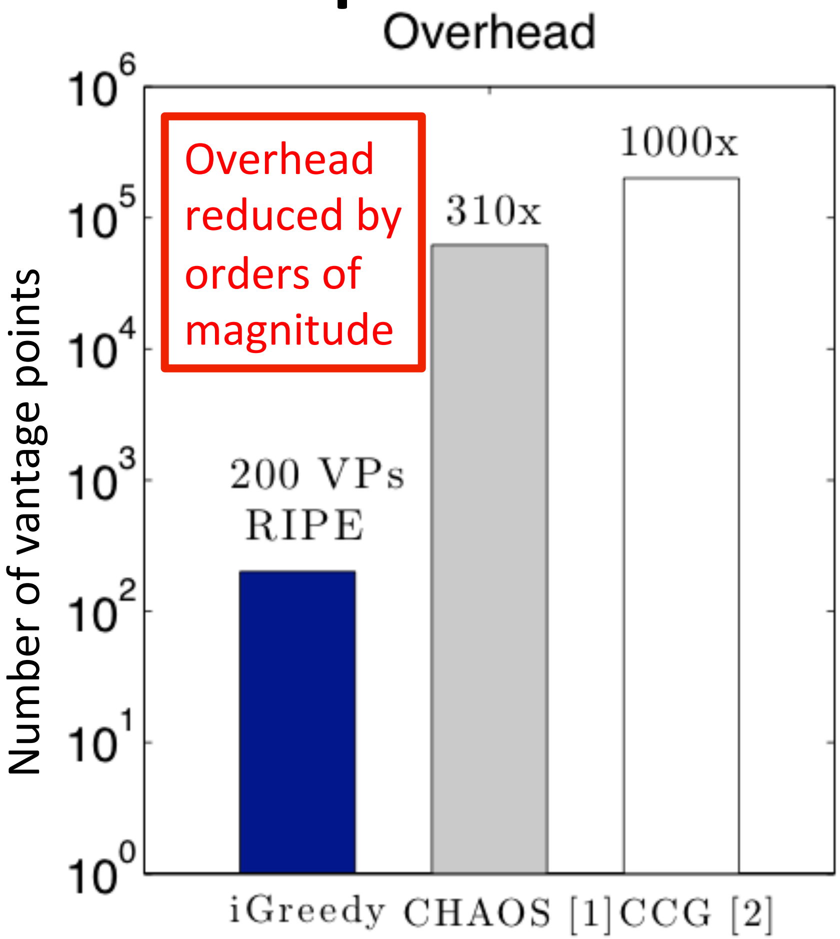
- Infrastructures
  - 200 PlanetLab VPs from 26 countries and105 AS
  - 6000 RIPE Atlas VPs from 122 countries and 2168 AS.

- Additionally
  - 500 RIPE Atlas VPs selected uniformly at random
  - 200 RIPE Atlas VPs stratified selection(>100km from each other)
  - Same size as PlanetLab, possibly larger diversity (200/6000)

- Focus
  - Comparison of iGreedy with state of the art
  - Robustness with respect to vantage point selection

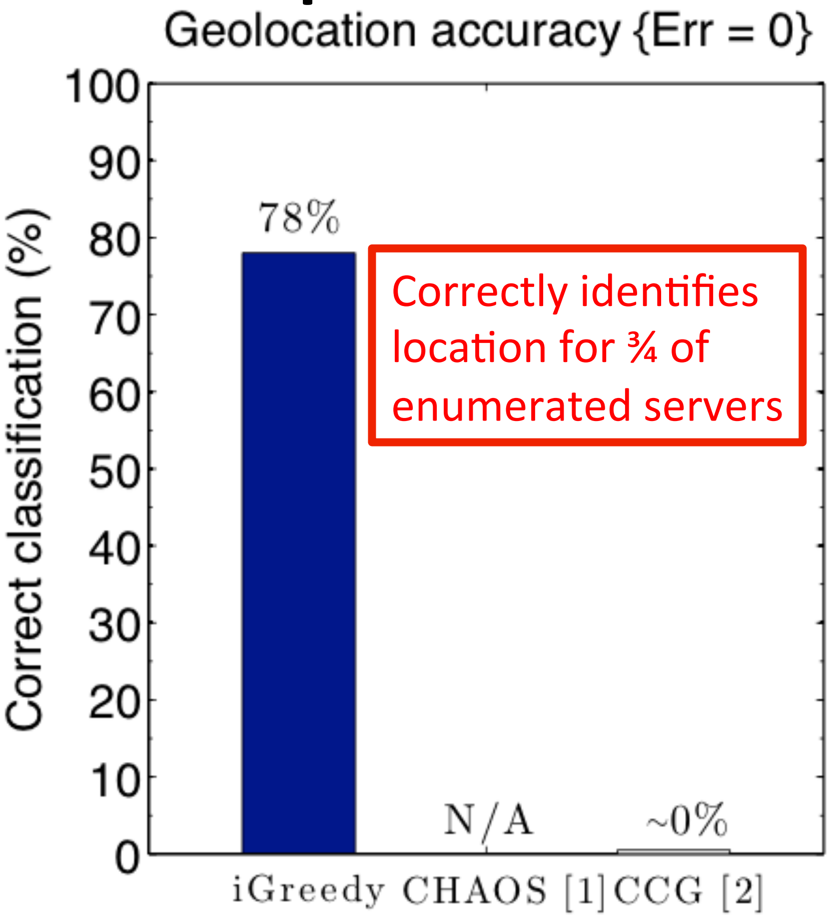
L root server in Europe

Comparison with state of the art



[1] X. Fan, J. Heidemann and R. Govindan, “Evaluating anycast in the Domain Name System” in Proc. IEEE INFOCOM, 2013.
[2] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann and R. Govindan, “Mapping the expansion of Google’s serving infrastructure” in Proc. ACM IMC, 2013.

Comparison with state of the art



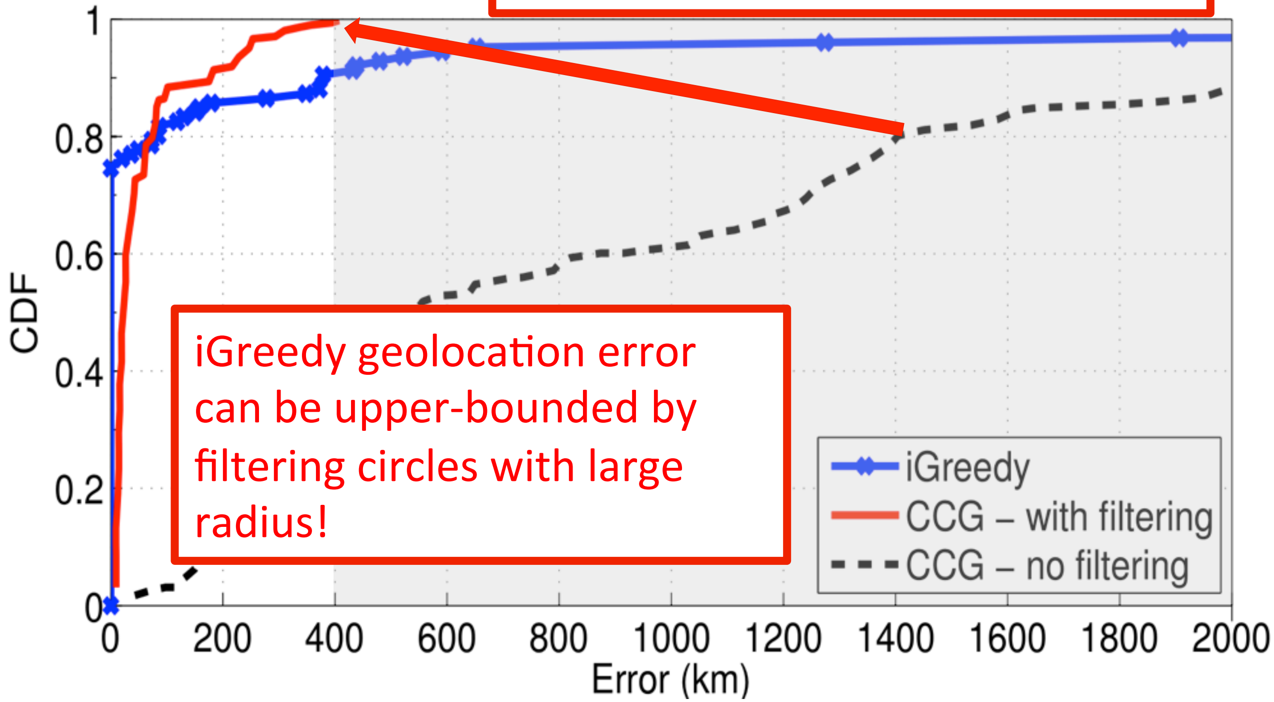
[1] X. Fan, J. Heidemann and R. Govindan, "Evaluating anycast in the Domain Name System" in Proc. IEEE INFOCOM, 2013.
[2] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann and R. Govindan, "Mapping the expansion of Google's serving infrastructure" in Proc. ACM IMC, 2013.

Qualitative comparison with CCG

- Big, Big, Big caveat
  - These numbers are not directly comparable !
  - Dataset and aim differ (Google infrastructure)

- iGreedy performance
  - No error for 78% of replicas
  - 271 km median error for misclassified replicas

CCG filtering wastes 20% of measurements a posteriori



iGreedy geolocation error can be upper-bounded by filtering circles with large radius!

[2] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann and R. Govindan, "Mapping the expansion of Google's serving infrastructure" in Proc. ACM IMC, 2013.

Robustness to vantage point selection

	RIPE Atlas			PlanetLab
Dataset	**Full 6000**	**Unif 500**	**Strat 200**	**Full 200**
iGreedy / CHAOS UB	76%	52%	73%	73%
CHAOS UB / GT	80%	54%	72%	36%
Geolocated	76%	63%	78%	74%
Mean geolocation error (km)	333	569	361	162

Number of VPs has little impact on recall!
Strategic selection of VPs is promising

Summary of achievements

- Detect, enumerate, & geolocate anycast replicas
  - Protocol agnostic and lightweight
    - Based on a handful of delay measurement O(100) VPs
    - 1000x fewer VPs than state of the art
  - Enumeration
    - iGreedy use75% of the probes (25% discarded due to overlap)
    - Overall 50% recall (depends on VPs; stratification is promising)
  - Geolocation
    - Correct geolocation for 78% of enumerated replicas
    - 361 km mean geolocation error for all enumerated replicas (271 km median for erroneous classification)