Shortcuts through Colocation Facilities

Vasileios Kotronis[1], George Nomikos[1], Lefteris Manassakis[1], Dimitris Mavrommatis[1] and Xenofontas Dimitropoulos[1,2]

[1]Foundation for Research and Technology - Hellas (FORTH), Greece
[2]University of Crete, Greece

Latency matters….

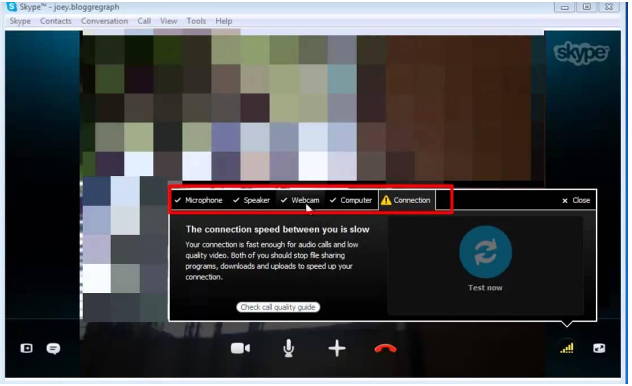
For Internet organizations...

“every **100ms** of latency cost 1% in sales”

“an extra **.5s** in search page generation time dropped traffic by 20%”

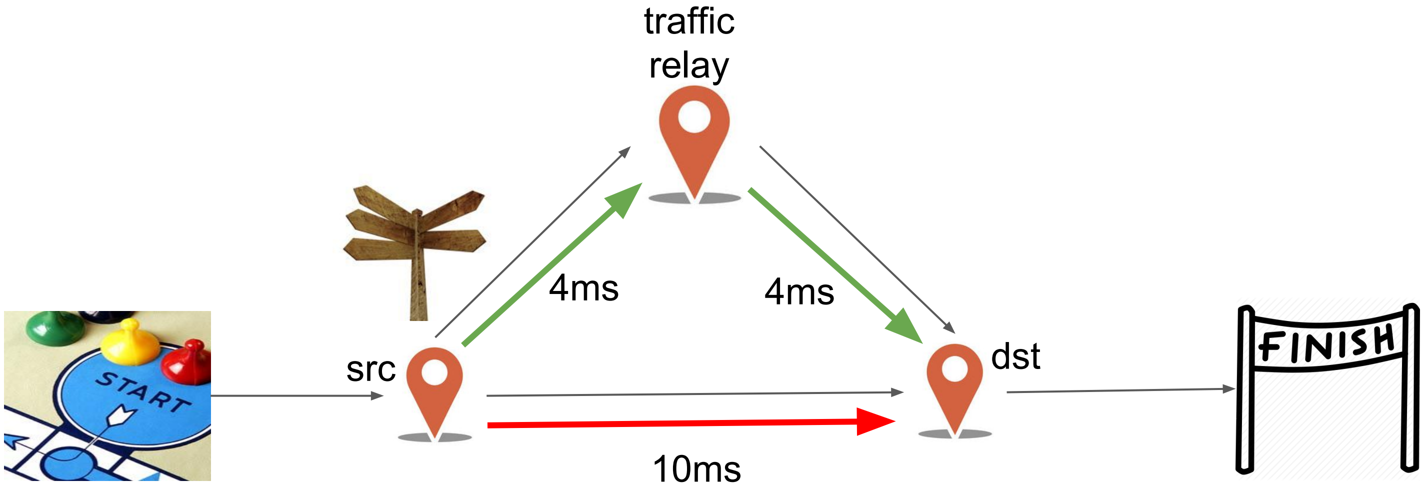
“A broker could lose $4 million/ms, if the electronic trading platform lags **5ms** behind competition”

...and end-users!



One way to reduce Internet latency:

Overlay networks exploiting TIVs



(**TIV** = **T**riangle **I**nequality **V**iolation)

Questions!

1) What are the **best locations** to place overlay TIV relays, to improve **performance** or **resiliency**?

Questions!

1) What are the best locations to place overlay TIV relays, to improve performance or resiliency?

2) **What** and **how much** benefit do these relays offer?

Who cares to answer them and Why?

- ➔ End-users and their overlay applications have much to gain
  - ◆ No need for strict SLAs or expensive networking setups
  - ◆ Cheap latency reductions using minimal numbers of relays

➔ Focus on → **Overlay-based Latency Improvement**

for → **Eyeball Networks** (access ISPs serving users at last mile)

investigating → **Colocation Facilities** (Colos) as potential relays

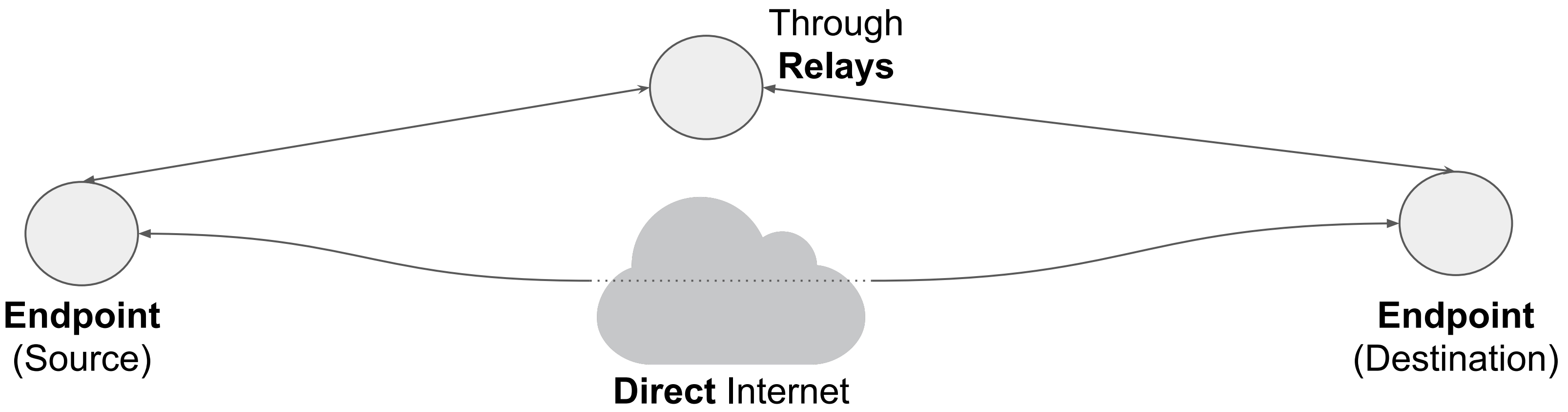
Why relays in Colocation facilities (Colos)?

- Space, power, cooling, physical security
- Usually host layer 2/3 interconnections
- Bring Internet organizations closer to:
  - Transit networks and eyeball ISPs
  - Content providers
  - Small/medium/large cloud providers
    → offer colocated VMs to third parties

⇒ Role of Colos as candidate TIV relays not explored!

Measurement methodology

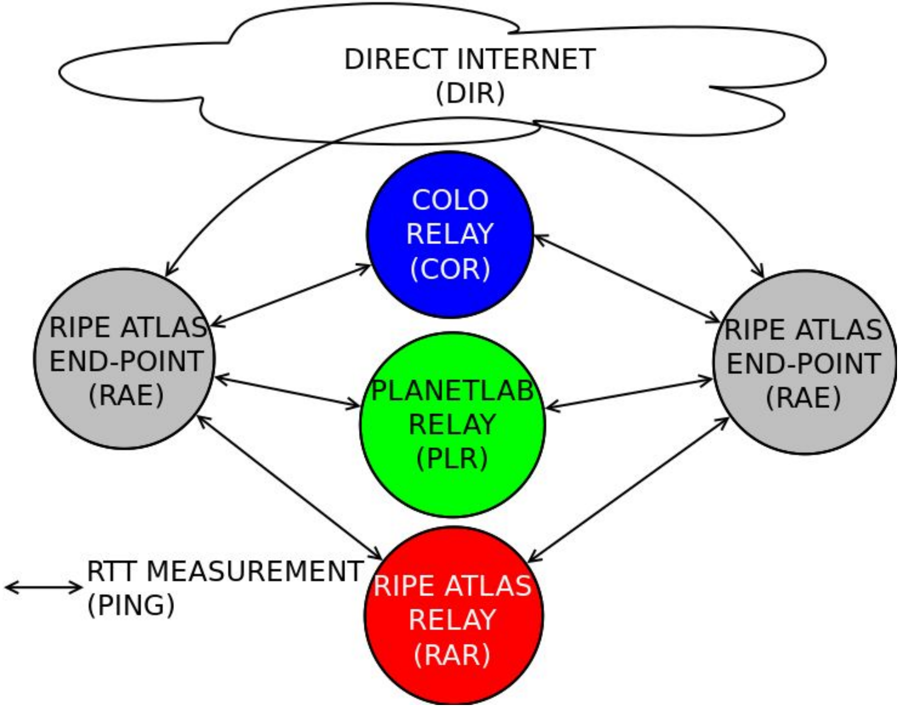
1. Pick a set of **endpoint** nodes (as source, destination)
2. For each source-dest pair measure the RTT of the **direct** path
3. Select a set of **feasible Relays** based on RTT
4. **Measure and stitch** the median RTT between source-relay and destination-relay on the relayed path



Measurement framework

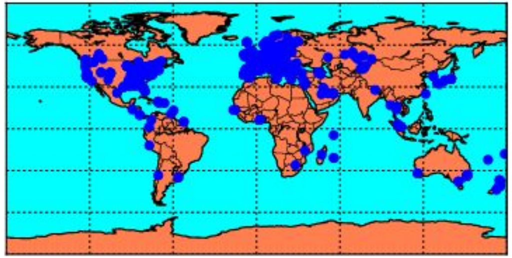
1. **Endpoints**
   - RIPE Atlas nodes (**RAE**) in Eyeballs

2. **Relays**
   - Colocation facilities (**COR**)
   - RIPE Atlas nodes (**RAR**)
     1. In eyeballs (RAR_eye)
     2. In other networks (RAR_other)
   - PlanetLab nodes (**PLR**)



Selecting RIPE Atlas Endpoints (**RAE**) in eyeballs

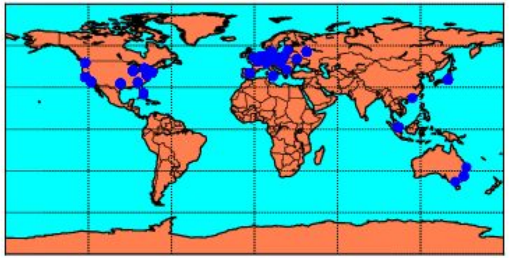
- End-users primarily reside in eyeballs
- We pick eyeball networks based on APNIC's dataset [1]
  - **223/225** countries host at least **1** AS serving **>10%** country's user population
  - **494 manually verified AS eyeball networks**
- We select RIPE Atlas nodes as endpoints within these networks
  - ~1.2K working probes/anchors
  - at 142 ASes
  - at 82 countries
  - ~82 RAE sampled per round (1/country)



[1] APNIC. "*IPv6 Measurement Campaign Dataset*". https://stats.labs.apnic.net/v6pop. Dataset collected on 31.03.2017.

Selecting Colo Relays (COR)

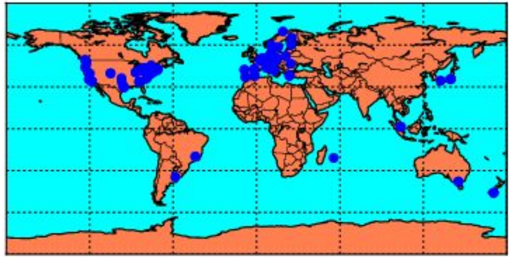
- Use publicly available dataset (router interface IPs → Colos) [1]
- Apply sequence of rules to exclude stale information
  - E.g., pingability, PeeringDB presence, RTT-based geolocation, etc.
- We select pingable IPs residing at Colos as relays
  - ~356 IPs
  - at 58 facilities
  - at 36 cities
  - ~129 COR sampled per round (1-3/facility)



[1] Giotsas, V., Smaragdakis, G., Huffaker, B., Luckie, M., et al. “*Mapping Peering Interconnections to a Facility*”. In Proc. of ACM CoNEXT, 2015.

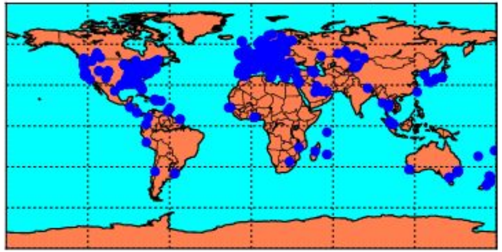
Selecting PlanetLab Relays (**PLR**)

- Hosts located (mostly) at research and academic institutions
- Allocated ~500 nodes at 62 PlanetLab sites
- Choose consistently accessible and pingable nodes
- ~60 PLR sampled per round (1-2/site)

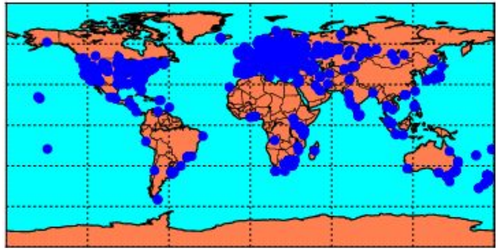


Selecting RIPE Atlas Relays (**RAR**)

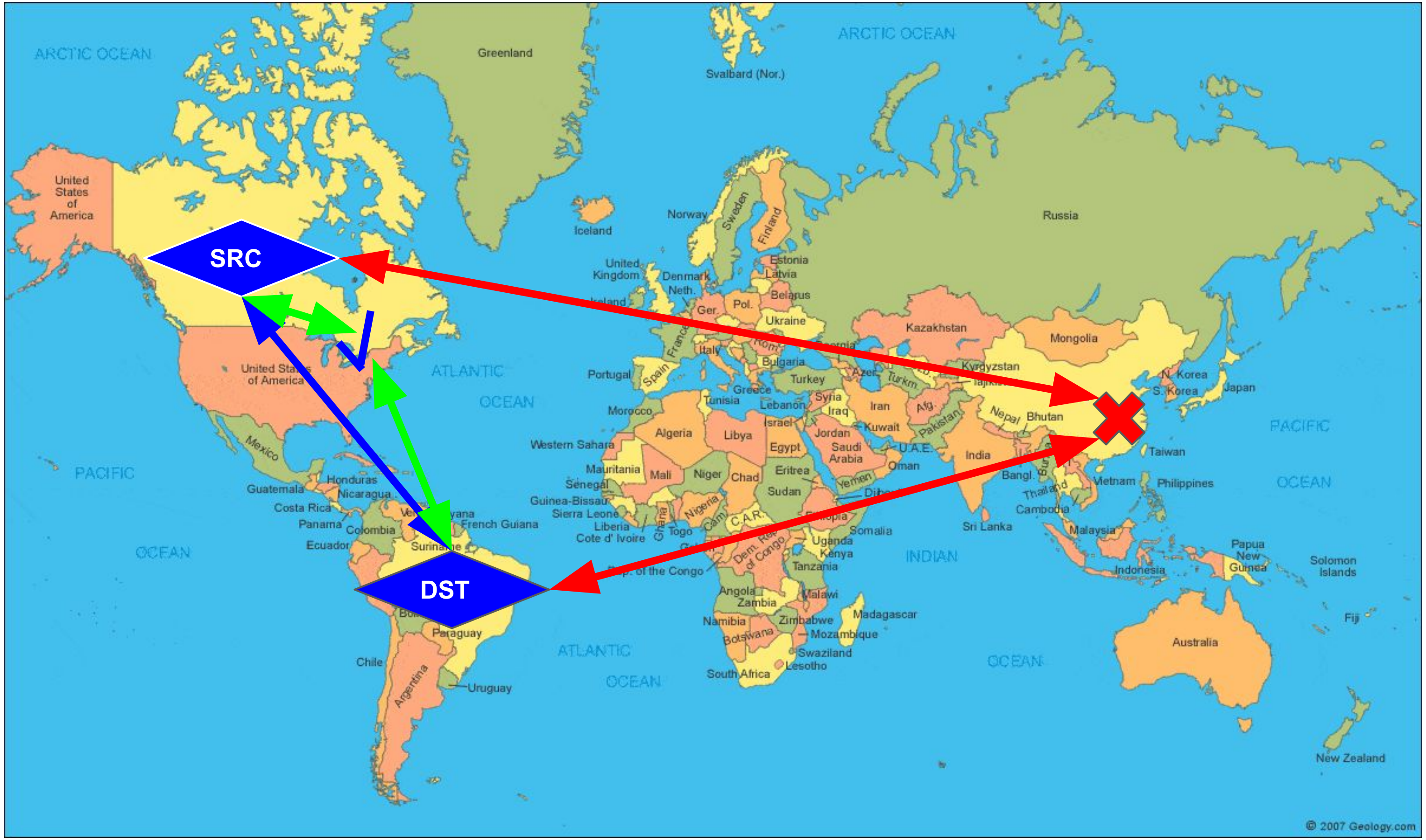
- At eyeballs (*RAR_eye*)
  - ~1.2K working probes/anchors
  - at 142 ASes
  - at 82 countries
  - ~82 RAR_eye sampled per round (1/country)



- At other networks (*RAR_other*)
  - ~2.5K remaining working probes/anchors
  - at 102 countries
  - ~102 RAR_other sampled per round (1/country)



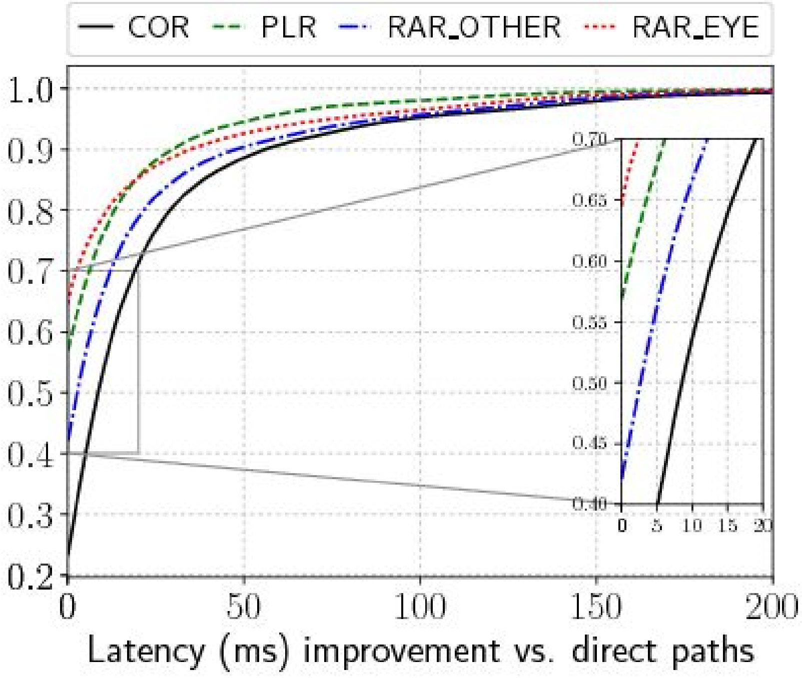
Which of the relays are feasible?



Size of measurement campaign

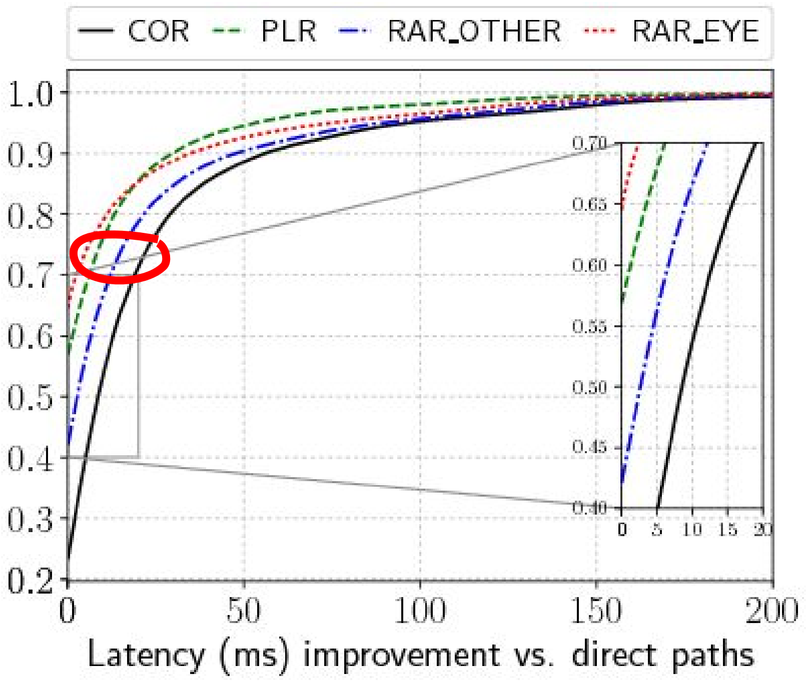
- One month measurement of **45** rounds (20 Apr - 17 May 2017)
- Utilized **~4.5K** relays and **~1K** endpoints in total
- Gathered **~8.7** million pings
- Studied **~29** million relayed paths

Latency improvements* per relay type



*Improvements between 1-200 ms are shown (83% of total cases)

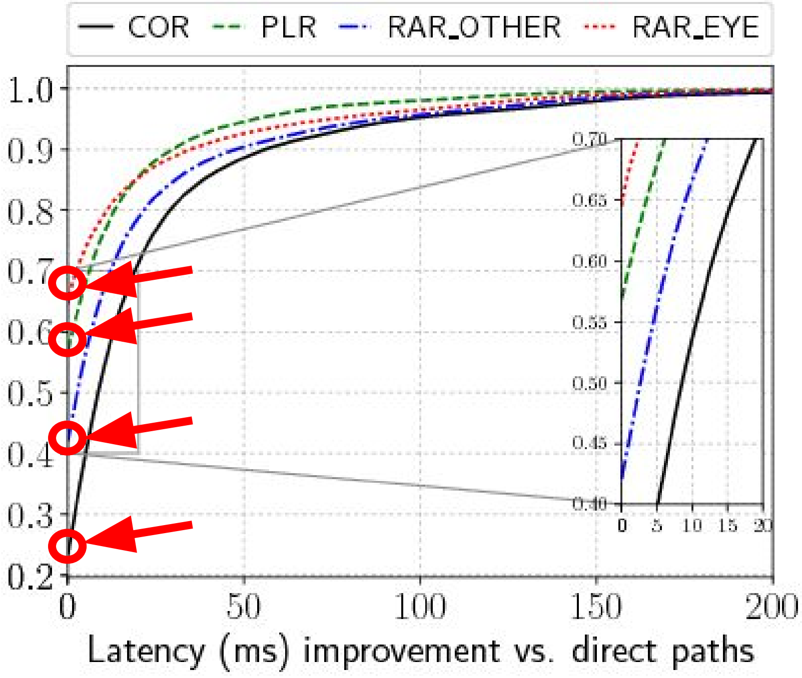
Latency improvements* per relay type



- Median reduction ~**12-14** ms

*Improvements between 1-200 ms are shown (83% of total cases)

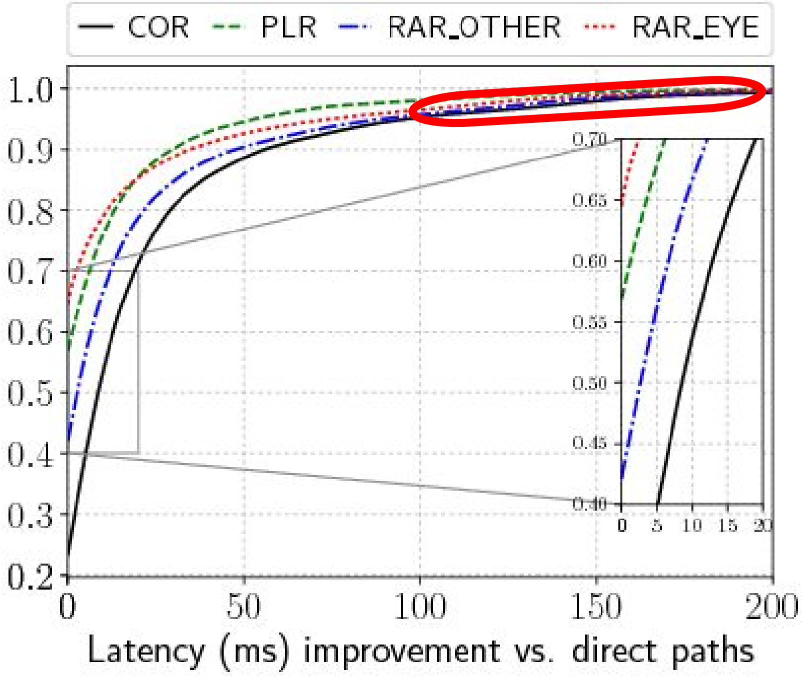
Latency improvements* per relay type



- Median reduction ~**12-14** ms
- Better than direct % of total cases:
  - COR: **76%**
  - RAR_other: **58%**
  - PLR: **43%**
  - RAR_eye: **35%**

*Improvements between 1-200 ms are shown (83% of total cases)

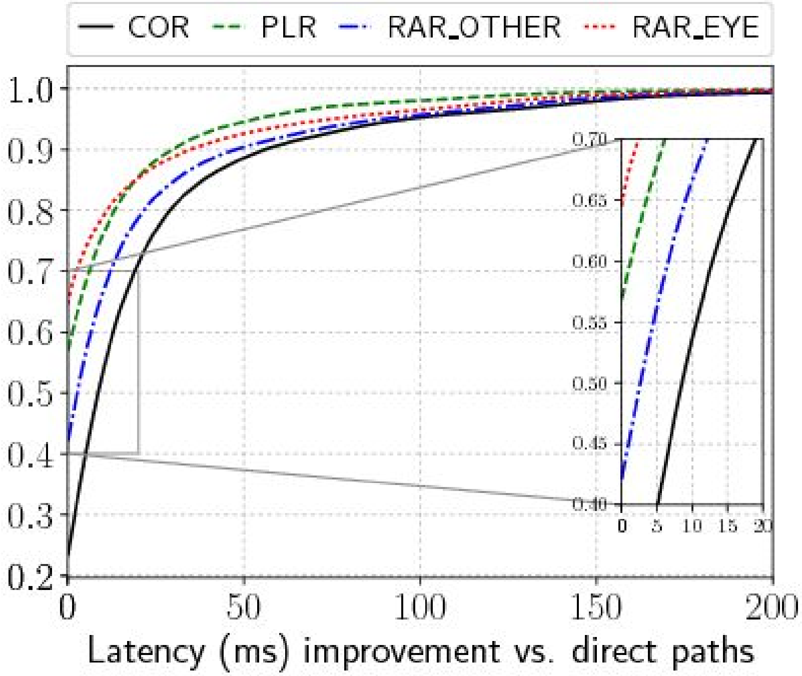
Latency improvements* per relay type



- Median reduction ~**12-14** ms
- Better than direct % of total cases:
  - COR: **76%**
  - RAR_other: **58%**
  - PLR: **43%**
  - RAR_eye: **35%**
- Reductions **>100ms** in **5%** of total cases (COR, RAR_other)

*Improvements between 1-200 ms are shown (83% of total cases)

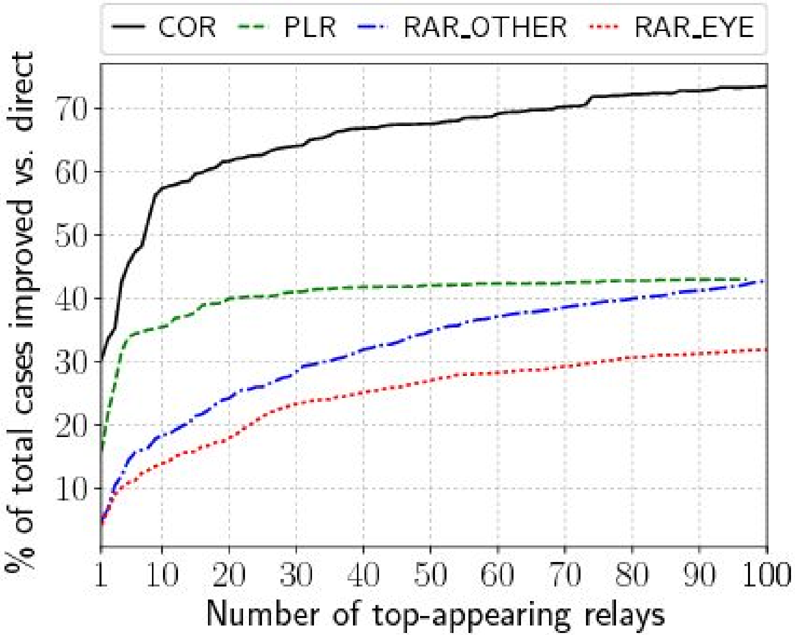
Latency improvements* per relay type



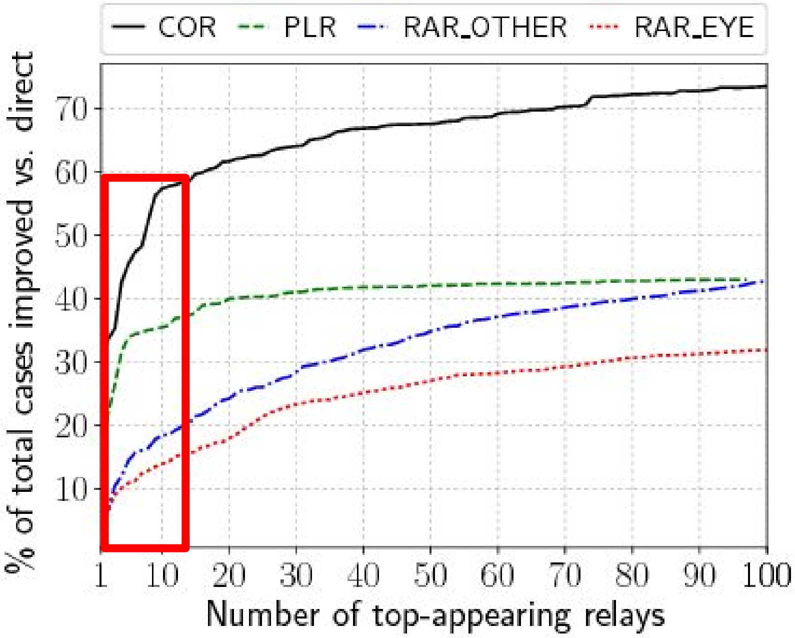
- Median reduction ~**12-14** ms
- Better than direct % of total cases:
  - COR: **76%**
  - RAR_other: **58%**
  - PLR: **43%**
  - RAR_eye: **35%**
- Reductions **>100ms** in **5%** of total cases (COR, RAR_other)
- **8** COR relays yield reductions/pair

*Improvements between 1-200 ms are shown (83% of total cases)

How many relays are enough?

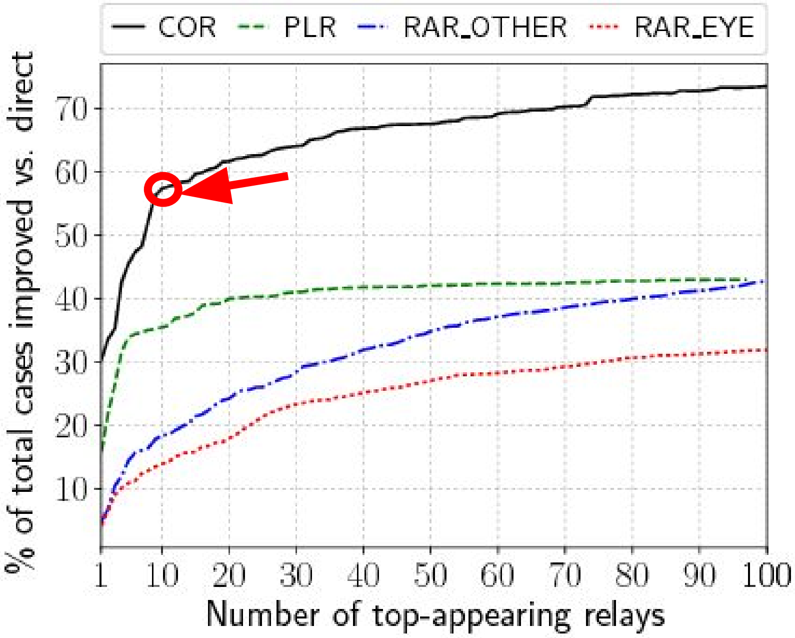


How many relays are enough?



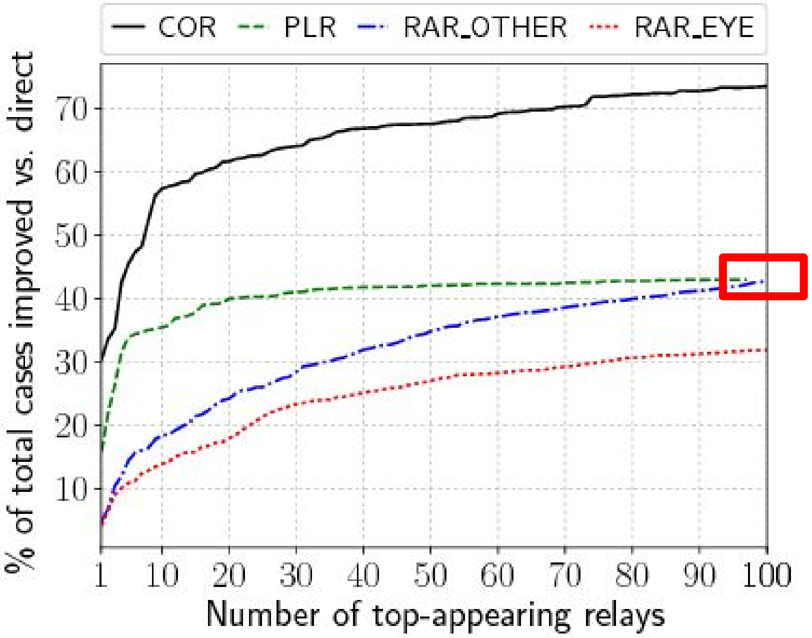
- Improved pairs ↑ rapidly with few COR, PLR relays

How many relays are enough?



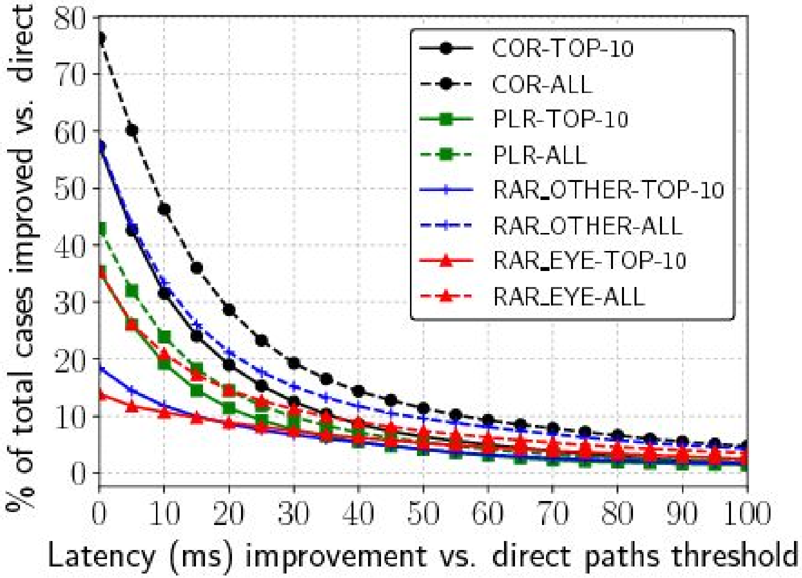
- Improved pairs ↑ rapidly with few COR, PLR relays
- **10** COR at **6** Colos improve ~ **58%** of total cases

How many relays are enough?

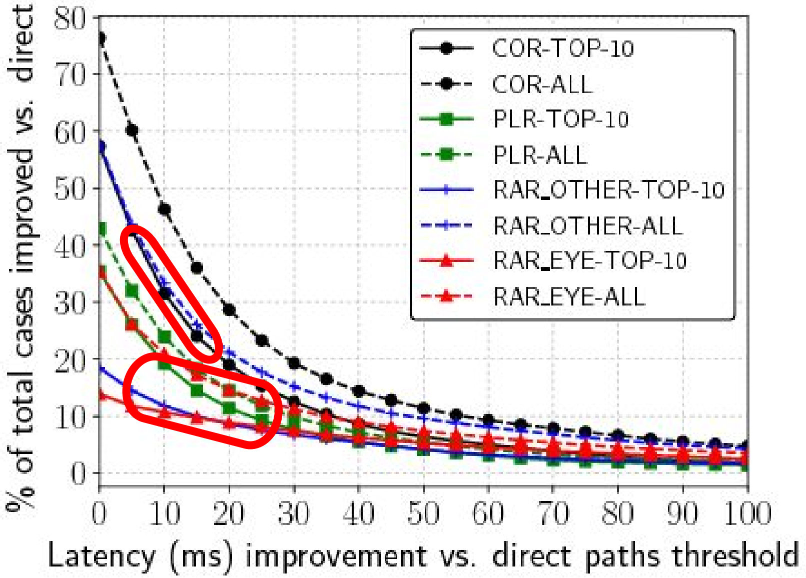


- Improved pairs ↑ rapidly with few COR, PLR relays
- **10** COR at **6** Colos improve ~ **58%** of total cases
- RAR_other **2nd** best, **but >>100** relays

How many relays are enough?

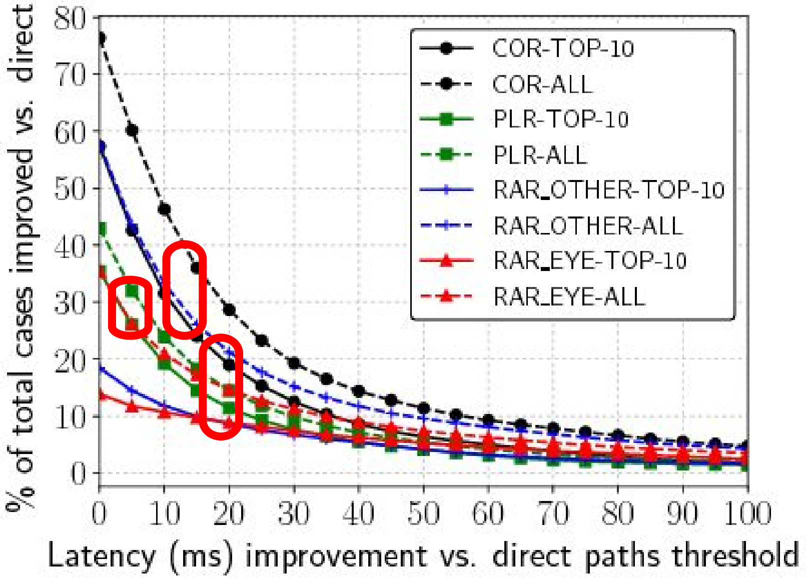


How many relays are enough?



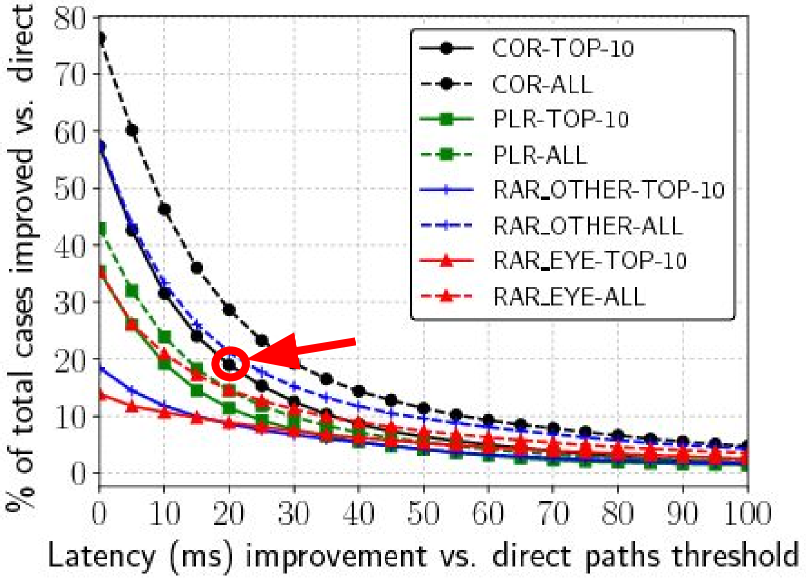
- top-10 COR > top-10 {PLR, RAR}

How many relays are enough?



- top-10 COR > top-10 {PLR, RAR}
- Different gaps between top-10 and all

How many relays are enough?



- top-10 COR > top-10 {PLR, RAR}
- Different gaps between top-10 and all
- **20%** of all pairs > **20ms** with top-10 COR

Top-10 facilities*

Facility Name (PDB ID)	% of Improved Cases	City (Country)	#Nets	#IXPs	Cloud Services	PDB top-10
1) Telehouse North (34)	47	London (GB)	361	6	✓	✓
2) Equinix-AM7 (62)	46	Amsterdam (NL)	184	4	✓	✓
3) Nikhef (18)	34	Amsterdam (NL)	151	6	✓	✗
4) Equinix-FR5 (60)	30	Frankfurt (DE)	235	11	✓	✓
5) Telehouse West (835)	29	London (GB)	89	5	✓	✗
6) Digital Realty Telx (125)	29	Atlanta (US)	125	2	✓	✗
7) Incolocate (105)	29	Hamburg (DE)	22	3	✓	✗
8) Interxion (68)	27	Brussels (BE)	58	3	✓	✗
9) Digital Realty Telx (10)	22	New York (US)	112	5	✓	✗
10) Equinix-LD8 (45)	21	London (GB)	208	4	✓	✓

* Facilities of top-20 Colo relays (ranked according to their frequency of presence in improved paths), and their location and connectivity characteristics.

Top-10 facilities*

Facility Name (PDB ID)	% of Improved Cases	City (Country)	#Nets	#IXPs	Cloud Services	PDB top-10
1) Telehouse North (34)	47	London (GB)	361	6	✓	✓
2) Equinix-AM7 (62)	46	Amsterdam (NL)	184	4	✓	✓
3) Nikhef (18)	34	Amsterdam (NL)	151	6	✓	✗
4) Equinix-FR5 (60)	30	Frankfurt (DE)	235	11	✓	✓
5) Telehouse West (835)	29	London (GB)	89	5	✓	✗
6) Digital Realty Telx (125)	29	Atlanta (US)	125	2	✓	✗
7) Incolocate (105)	29	Hamburg (DE)	22	3	✓	✗
8) Interxion (68)	27	Brussels (BE)	58	3	✓	✗
9) Digital Realty Telx (10)	22	New York (US)	112	5	✓	✗
10) Equinix-LD8 (45)	21	London (GB)	208	4	✓	✓

* Facilities of top-20 Colo relays (ranked according to their frequency of presence in improved paths), and their location and connectivity characteristics.

Top-10 facilities*

Facility Name (PDB ID)	% of Improved Cases	City (Country)	#Nets	#IXPs	Cloud Services	PDB top-10
1) Telehouse North (34)	47	London (GB)	361	6	✓	✓
2) Equinix-AM7 (62)	46	Amsterdam (NL)	184	4	✓	✓
3) Nikhef (18)	34	Amsterdam (NL)	151	6	✓	✗
4) Equinix-FR5 (60)	30	Frankfurt (DE)	235	11	✓	✓
5) Telehouse West (835)	29	London (GB)	89	5	✓	✗
6) Digital Realty Telx (125)	29	Atlanta (US)	125	2	✓	✗
7) Incolocate (105)	29	Hamburg (DE)	22	3	✓	✗
8) Interxion (68)	27	Brussels (BE)	58	3	✓	✗
9) Digital Realty Telx (10)	22	New York (US)	112	5	✓	✗
10) Equinix-LD8 (45)	21	London (GB)	208	4	✓	✓

* Facilities of top-20 Colo relays (ranked according to their frequency of presence in improved paths), and their location and connectivity characteristics.

Top-10 facilities*

Facility Name (PDB ID)	% of Improved Cases	City (Country)	#Nets	#IXPs	Cloud Services	PDB top-10
1) Telehouse North (34)	47	London (GB)	361	6	✓	✓
2) Equinix-AM7 (62)	46	Amsterdam (NL)	184	4	✓	✓
3) Nikhef (18)	34	Amsterdam (NL)	151	6	✓	✗
4) Equinix-FR5 (60)	30	Frankfurt (DE)	235	11	✓	✓
5) Telehouse West (835)	29	London (GB)	89	5	✓	✗
6) Digital Realty Telx (125)	29	Atlanta (US)	125	2	✓	✗
7) Incolocate (105)	29	Hamburg (DE)	22	3	✓	✗
8) Interxion (68)	27	Brussels (BE)	58	3	✓	✗
9) Digital Realty Telx (10)	22	New York (US)	112	5	✓	✗
10) Equinix-LD8 (45)	21	London (GB)	208	4	✓	✓

* Facilities of top-20 Colo relays (ranked according to their frequency of presence in improved paths), and their location and connectivity characteristics.

Conclusions

- Colos are “**core**” locations for relays ⇒ **low-latency** TIV paths
- **10** COR-relays in **6** Colos yield better-than-direct overlay paths in **~58%** of the total cases
- Other overlays require orders of magnitude more relays
- Code and datasets available online ⇒ http://inspire.edu.gr/shortcuts_colocation_facilities/

Conclusions

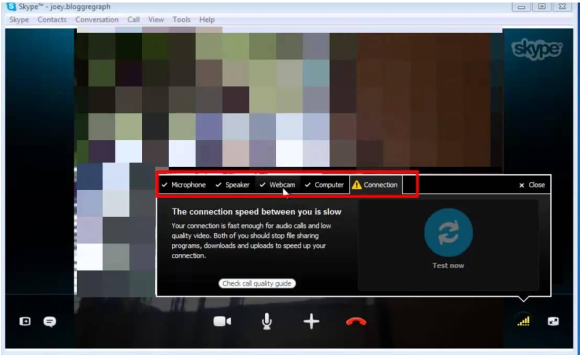
- Colos are “**core**” locations for relays ⇒ **low-latency** TIV paths
- **10** COR-relays in **6** Colos yield better-than-direct overlay paths in **~58%** of the total cases
- Other overlays require orders of magnitude more relays
- Code and datasets available online
  ⇒ http://inspire.edu.gr/shortcuts_colocation_facilities/
- **Future work**:
  → root cause(s) for COR performance
  → correlation with regional effects (e.g., country-level)

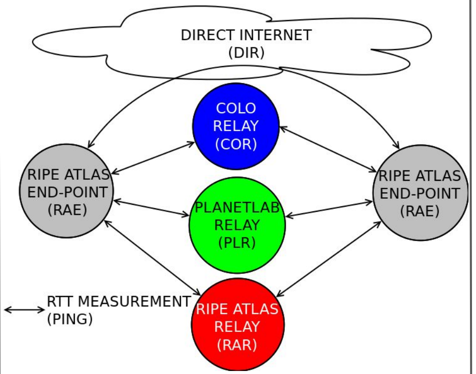
FORTH
FOUNDATION FOR RESEARCH AND TECHNOLOGY - HELLAS
INSTITUTE OF COMPUTER SCIENCE

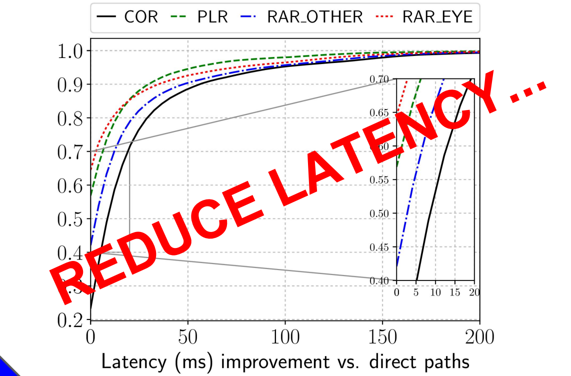
www.inspire.edu.gr

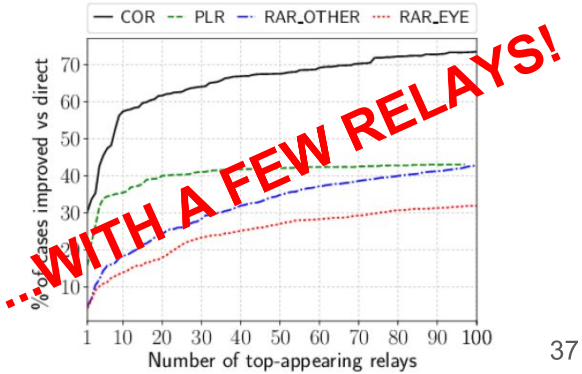
Thank you! Questions?

INternet Security & Privacy Intelligence REsearch Group









vkotronis@ics.forth.gr

BACKUP

More on RIPE Atlas node selection

- Running latest firmware version (system-v3)
  - Avoid msm interference artifacts affecting older versions [1]
- Publicly available (is-public = True)
- Connected and pingable (status = 1, system-ipv4-works)
- Tagged with their geolocation coordinates (geometry)
- Stable, connectivity-wise, during the last month (system-ipv4-stable-30d)

BACKUP

[1] Holterbach, T., Pelsser, C., Bush, R., and Vanbever, L. *“Quantifying interference between measurements on the RIPE Atlas platform”*. In Proceedings of the Internet Measurement Conference (2015), ACM, pp. 437–443.

Verification of IP → facility mappings

1. **Single-facility & active PeeringDB presence** (*1008/2675 IPs*)
2. **Pingability** (*764/1008 IPs*)
3. **Same IP-ownership** (IP2AS, no MOAS) (*725/764 IPs*)
4. **Active facility presence of ASN** (*725/725 IPs*)
5. **RTT-based geolocation using Periscope LGs** (*356/725 IPs*)

Biases - Limitations

- RIPE Atlas deployment bias
- 1/country RAE endpoint selection
  - Country-level diversity (not complete geographical/population-level)
  - But e.g., US is treated similarly as smaller European countries
- Unexpected measurement artifacts
  - E.g., nodes getting offline due to transient problems during msm

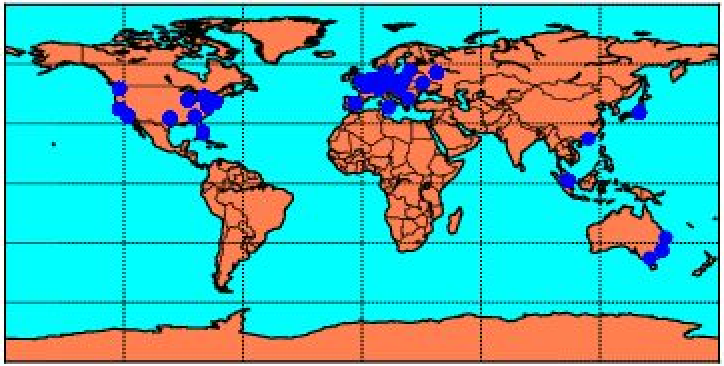
BACKUP

⇒ May affect the facility ranking

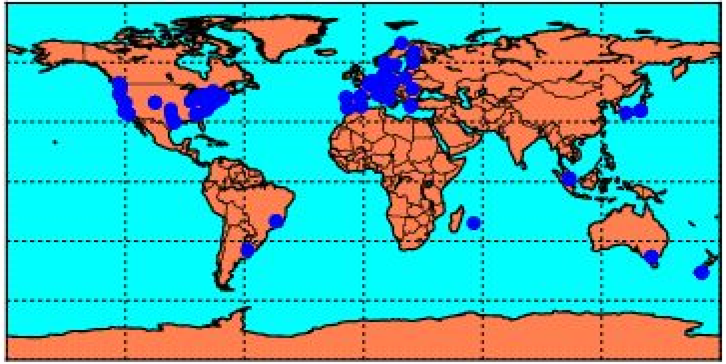
⇒ Does not affect insights on the contribution of Colos as relays

Where on earth are all these relays?

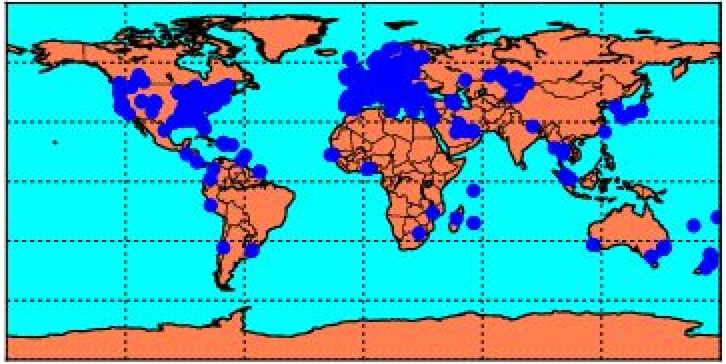
COR



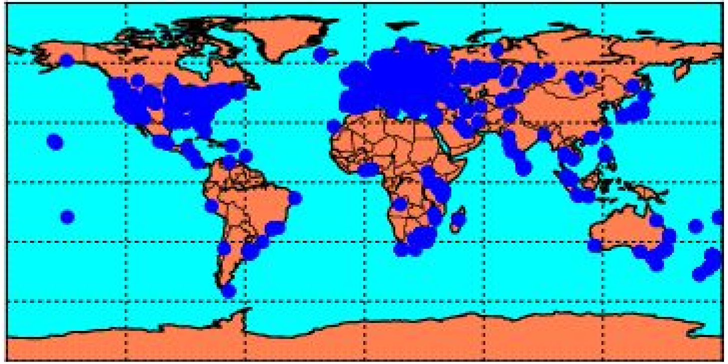
PLR



RAR_EYE



RAR_OTHER



BACKUP

Related work

- RON [1]: Resilient -and potentially faster than default BGP- paths
- VIA [2]: Overlay and prediction-based techniques for Internet telephony
- ARROW [3]: Secure e2e tunnels relayed via ISP waypoints
- MeTRO [4], CRONets [5]: Virtual routers in the cloud(s)
- Use of overlays ⇒ delicate balance between
    - overlay-based optimization, policy-driven TE (e.g., on the enterprise level)
- Tendency towards inter-domain overlay networks, using relays at:
    - data centers, ISPs, the last mile
- **The role of Colos not sufficiently explored at scale!**

BACKUP

[1] Andersen, D., et al. "*The Case for Resilient Overlay Networks*". In Proc. of IEEE HotOS, 2001.
[2] Jiang, J., et al. "*Via: Improving internet telephony call quality using predictive relay selection*". In Proc. of ACM SIGCOMM, 2016.
[3] Peter, S., et al. "*One Tunnel is (Often) Enough*". ACM SIGCOMM CCR 44, 4 (2015), 99–110.
[4] Makkes, M. X., et al. "*MeTRO: Low Latency Network Paths with Routers-on-Demand*". In Proc. of EU Conference on Parallel Processing, 2013.
[5] Cai, C. X., et al. "*CRONets: Cloud-Routed Overlay Networks*". In Proc. of IEEE ICDCS, 2016.

Future work

1. Root cause(s) for the performance of COR
    a. Initial hints: location, connectivity to IXPs, # colocated networks, etc.
2. Underlying reasons for the good performance of RAR_other
    a. RIPE Atlas deployment in commercial (core) networks?
    b. Investigate ASes where the nodes are present
3. Regional effects uncovered via traceroute measurements
    a. Correlations between latency and characteristics of traversed countries
    b. Correlations between the latency and proximity of endpoints/relays to submarine cable landing points [1]

BACKUP

[1] TeleGeography. "*Submarine Cable Map*". https://www.submarinecablemap.com/. Accessed: 11.09.2017.

Formulas related to the relay feasibility

Propagation delay between points n_1, n_2:

$$t(n_1, n_2) = \frac{d(n_1, n_2)}{c * \frac{2}{3}}$$

(*Speed of light in fiber*)

Feasible relays *f* must satisfy:

$$2 * [t(n_1, f) + t(f, n_2)] \leq RTT(n_1, n_2)$$

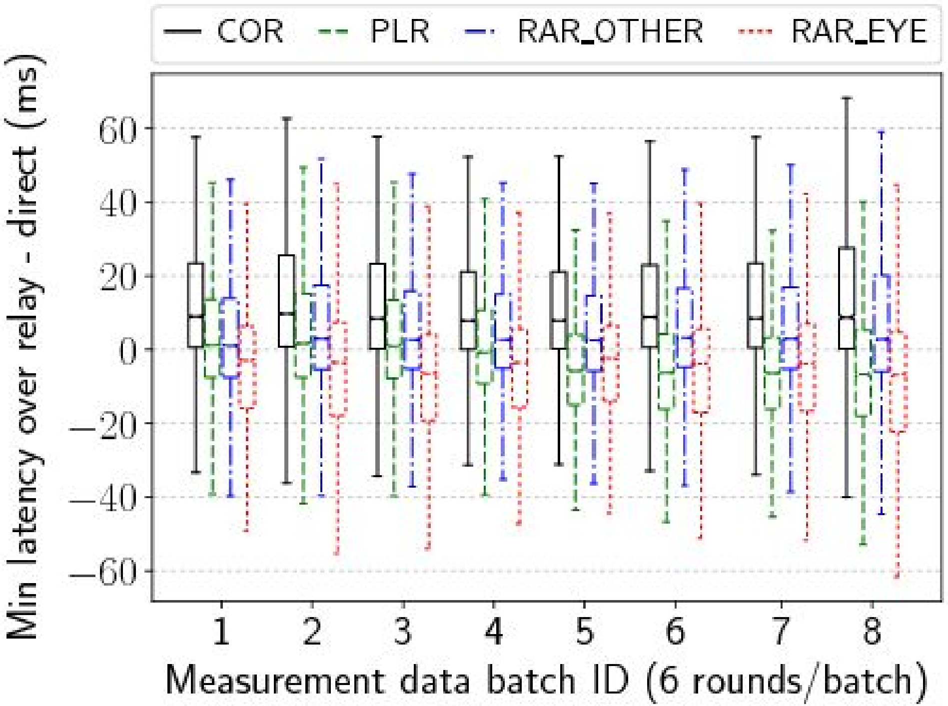
BACKUP

Changing countries and paths

- Path inflation can prevent relays close to endpoints, from using alternate low-latency paths
- 74% of studied paths → inter-continental (conducive to path inflation)
- The **latency** over *COR*-relayed paths is **lower** than direct paths:
  - in **75%** of the cases, when relays are in different countries than both endpoints
  - in **50%** of the cases, when relays are in the same country as one of the endpoints

BACKUP

Stability over time



- Consistent patterns for:
  >75 % (COR),
  >50% (RAR_other),
  <50% (PLR, RAR_eye)
  yielding lower-latency paths

BACKUP

- CV = SD of median RTTs of each pair (direct/relayed) divided by the pair's average RTT
- CV **< 10%** in **90%** of the cases
  ⇒ **stable** overlays