

Reconstructing the Internet History Mining through Path Stitching

DK Lee, Keon Jang, Changhyun Lee, Sue Moon, Gianluca Iannaccone*

FISC Student Workshop
August 27, 2009

Division of Computer Science, KAIST
Intel Research, Berkeley*

“Measurement Data”

- Internet performance measurement data useful to
 - Internet scientists, engineers or operators
 - Network application developers
 - End users
- Traditional active measurements:
 - Define estimation methodologies for delay, path, loss.
 - Carefully construct an active probing strategy
 - Instrument end-systems to collect measurement

“Measurement Data” Retrieval

- ***Problem statements***

Given two arbitrary points x and y in the Internet,
We estimate Internet forwarding $path(x, y)$, and
retrieve queried measurement data on $path(x, y)$
without additional active measurements.

- Our vision is to offer measurements retrieval
as “DNS-like” Internet service: Internet SIBILLA

Talk Outline

- Path Stitching algorithm
 - Constructing the path segment repository
 - Approximation and preference rules
 - Sources of errors
 - Evaluation

- Reconstructing the past history of the Internet
 - Pairs of points to examine
 - Trend detection

Part I. “Path Stitching”

*A light-weight algorithm for
Internet-wide path and delay estimation
using existing measurements*

“Path Stitching”

- Path and delay estimation between any pair of hosts
- Key assumption:
 - “Many good measurement data are available already.”
- Decoupling the data collection phase from the data analysis

Key ideas behind *path stitching*

Internet separates *inter-* and *intra-domain* routing;

To predict a new path, path stitching

» *Splits* paths into AS-path segments, and

» *Stitches* path segments together

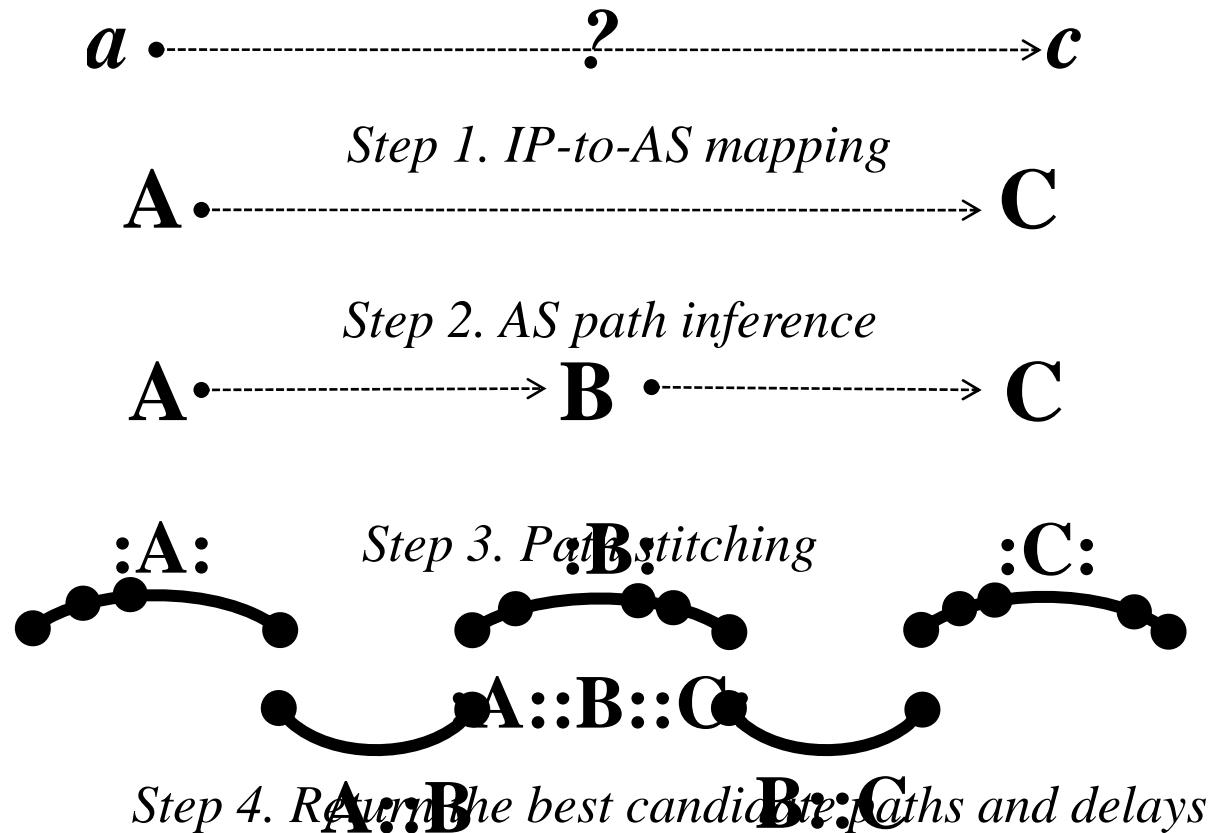
» *Using* BGP routing information

Data Sets

- *CAIDA Ark's traceroutes*
 - 1 round of *traceroute* outputs from 18 sources to every /24 prefix
 - 14 millions of *traceroute* outputs
- BGP routing tables
 - University of Oregon, *RouteViews'* BGP listener
 - *RIPE RIS'* 14 monitoring points (rrc00 ~ rrc07, rrc10 ~ rrc15)

Overview of "Path Stitching"

- What's the *router-level paths* and *latency estimates* between two arbitrary Internet hosts *a* and *c*?



Addressing Sources of Error – (1)

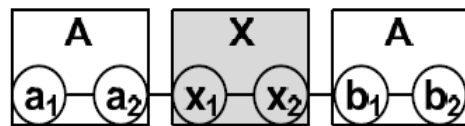
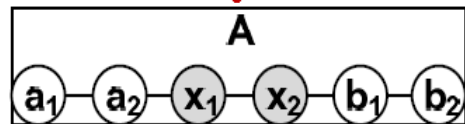
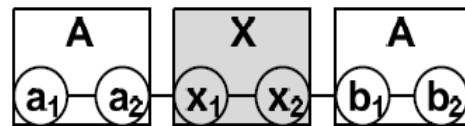
- IP-to-AS Mapping

- Single and multiple origin AS mismatches
- Incorporate connectivity between ASes despite the mapping problem

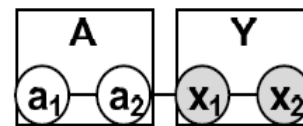
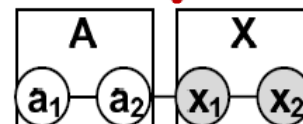
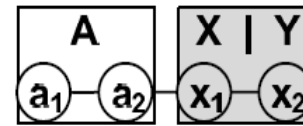
IP-to-AS mapping errors



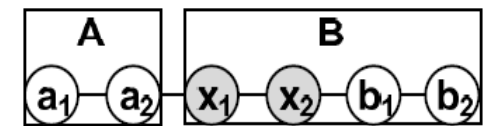
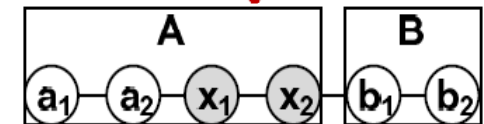
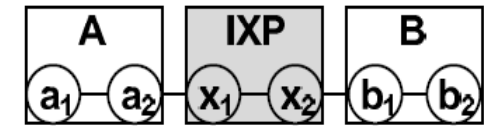
Build indexes for all possible combinations



(a) Two-hop AS loops



(b) MOASes



(c) IXPs

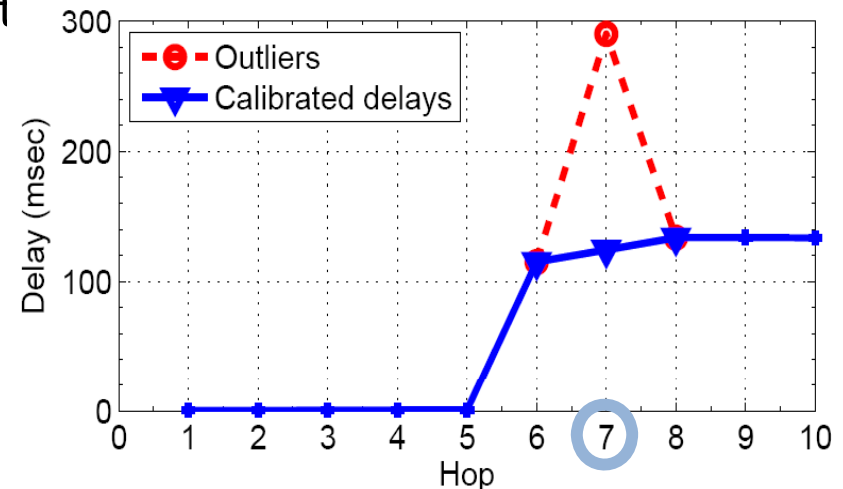
Addressing Sources of Error – (2)

- AS Path Inference

- Multi-homing is one of the main obstacles to the accurate AS path inference [Mao *et al.* SIGMETRICS 2005]
- We extract **first-hop information** from the Ark traceroute data. (We garner first hop information for 5,387 ASes)

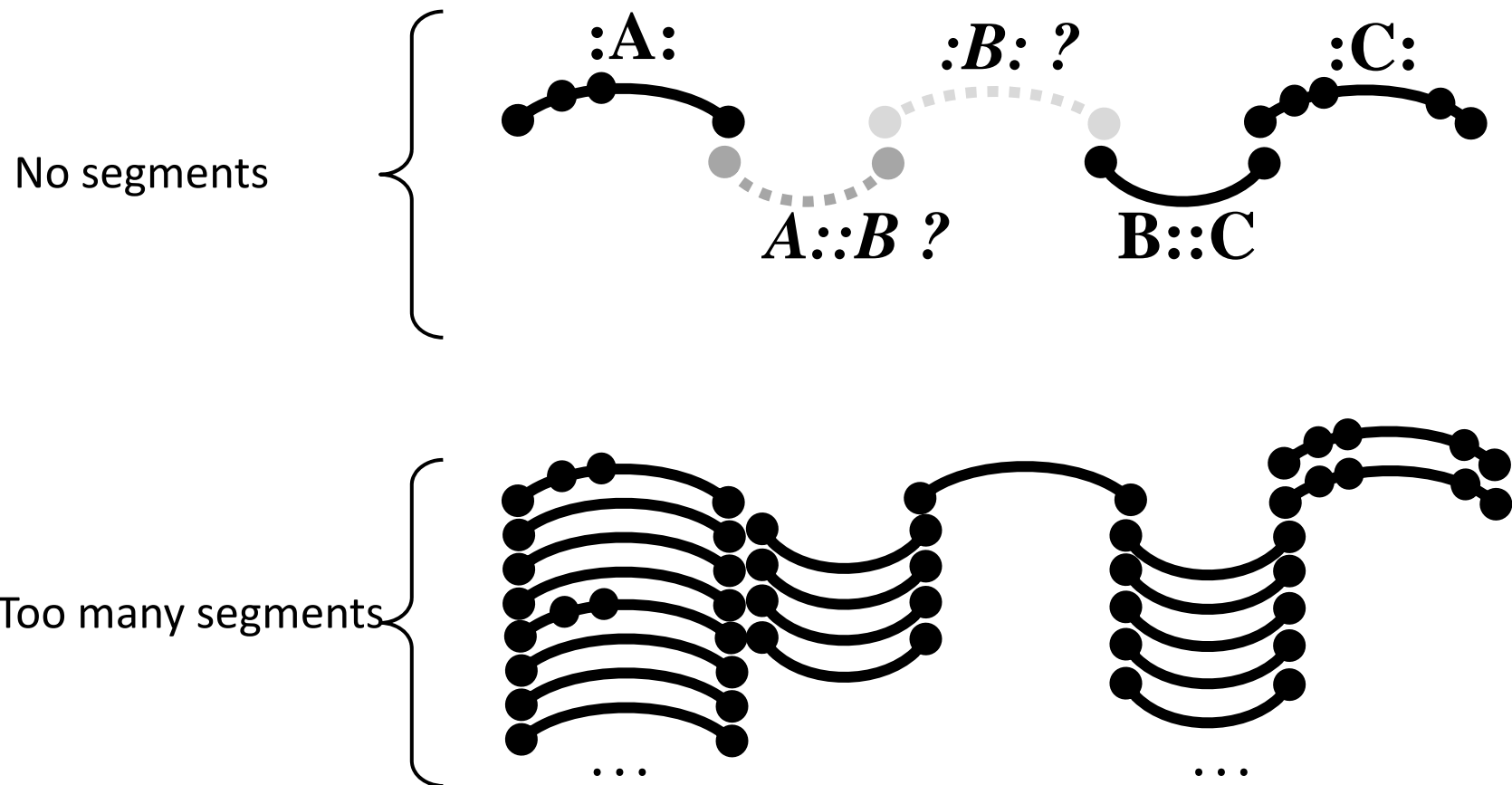
- Traceroutes

- Internet dynamics captured by traceroute provide both the **median** and **the most recent** measurements.
- **Nondecreasing delay principle** ----->



Too Few or Too Many Path Segments

- In Step 3 of our algorithm,



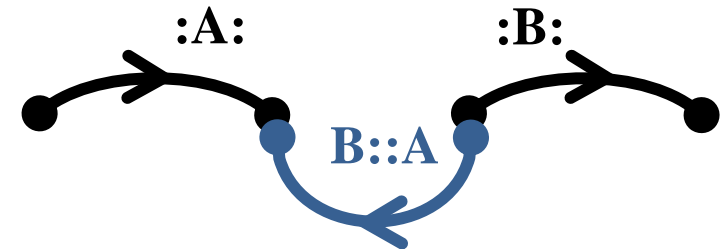
No Segments: Approximations

(i) Missing AS

- » No solutions (other than collecting more measurements.)

(ii) Missing inter-domain segment

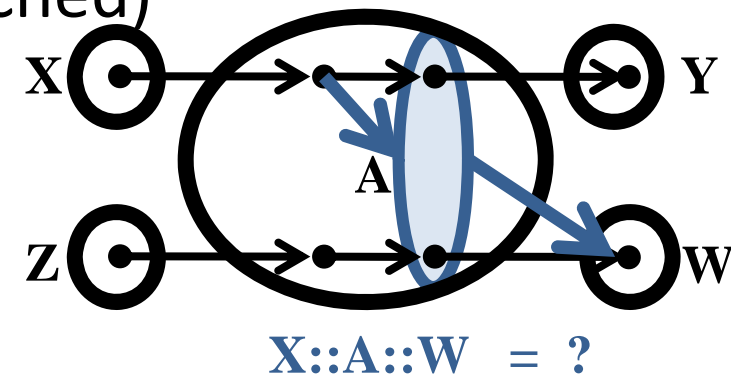
- » Search for reverse path segments.
(i.e., if we cannot find **A::B**, use **B::A** instead)






(iii) Path segments do not rendezvous at the same address

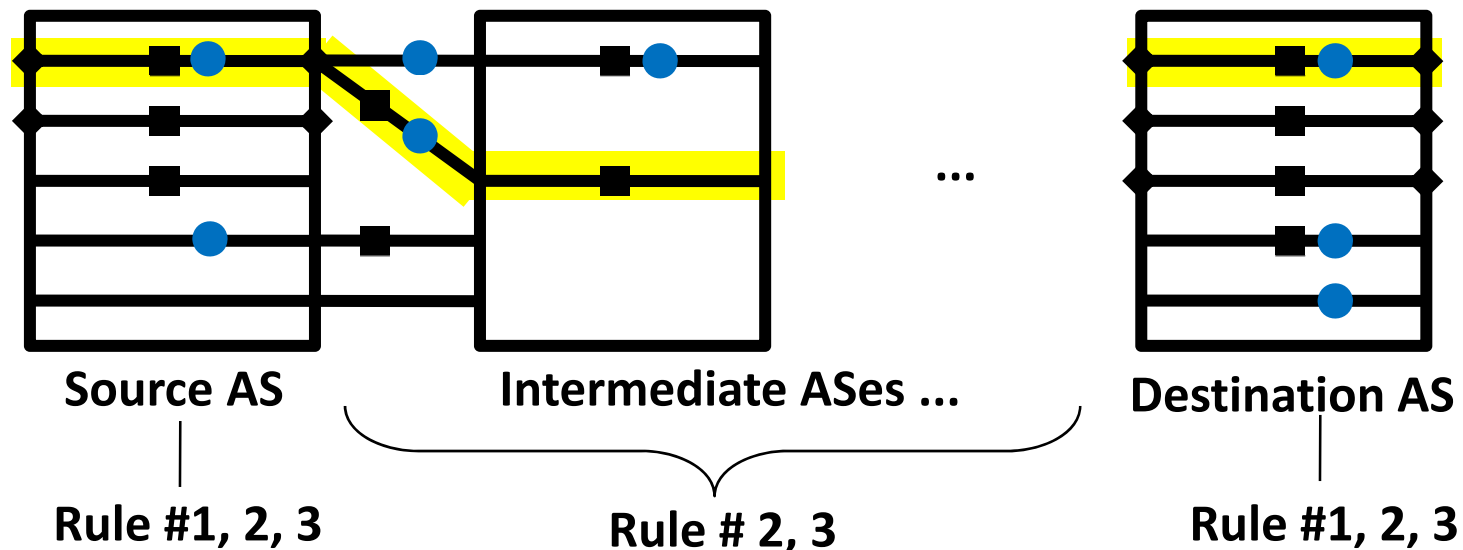
(i.e., the segment cannot be stitched)

- » Use clustering heuristics:
Clustering by *Router or PoP*
Clustering by the *IP prefix*



Too Many Segments: Preference Rules

- Rule #1: Proximity 
Preference to the paths closest to the queried source and destination addresses
- Rule #2: Destination-bound path segments 
Preference to the segments from traceroutes with the same destination prefix
- Rule #3: Most recent path segment 
Preference to the most recent path segment



Evaluation

- Additional data set for comparison:

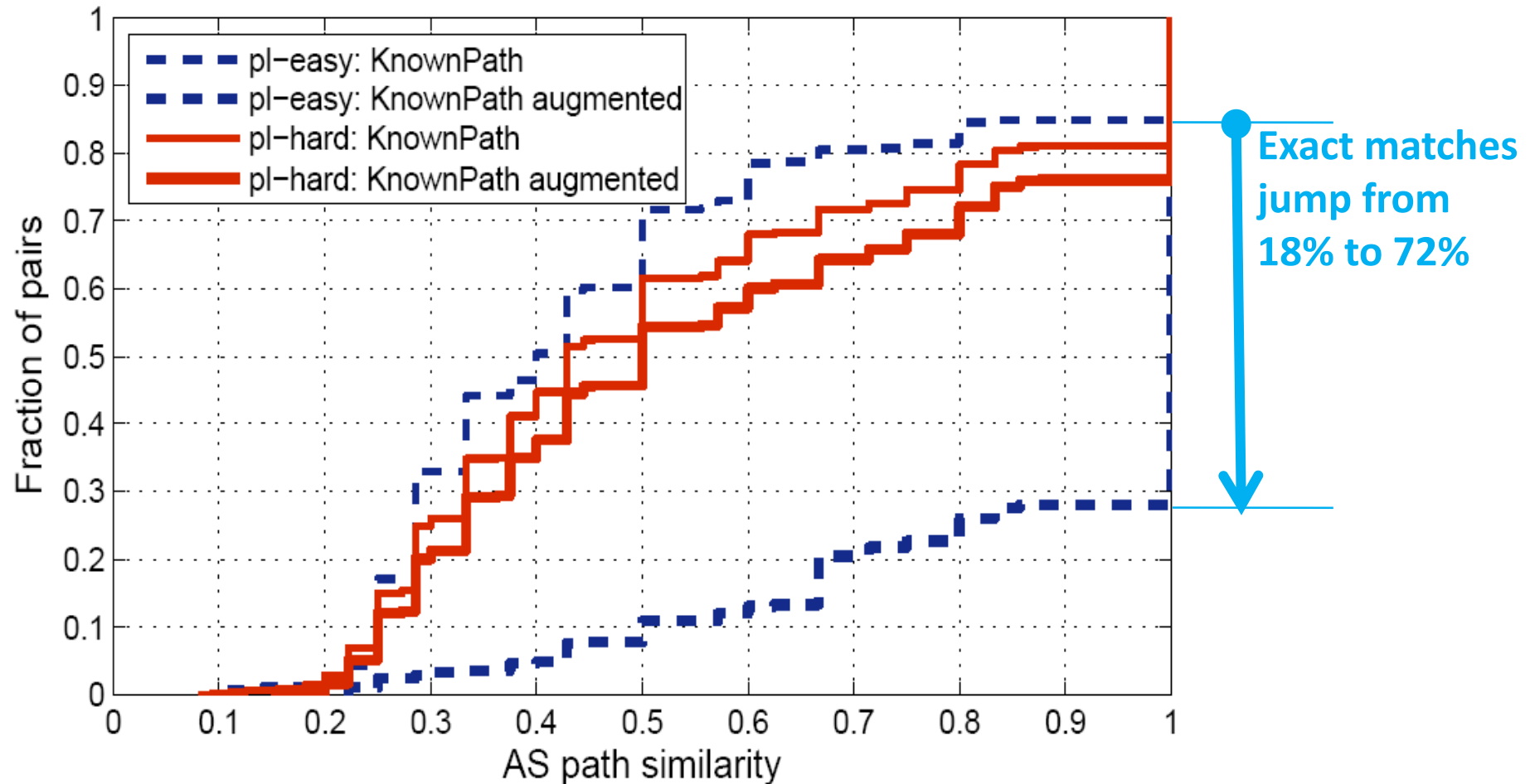
- Perform traceroute 50 times a day between 184 PL-nodes (real measurements)
- **462 *pl-easy*** pairs and **10,077 *pl-hard*** pairs
- For every pair, estimate path and delays using path stitching.

→ Source PL-nodes co-locate with Ark monitors
(namely, *amw-us*, *cbg-uk*, *cjj-kr*, *dub-ie*, *gig-br*)

- Evaluation of

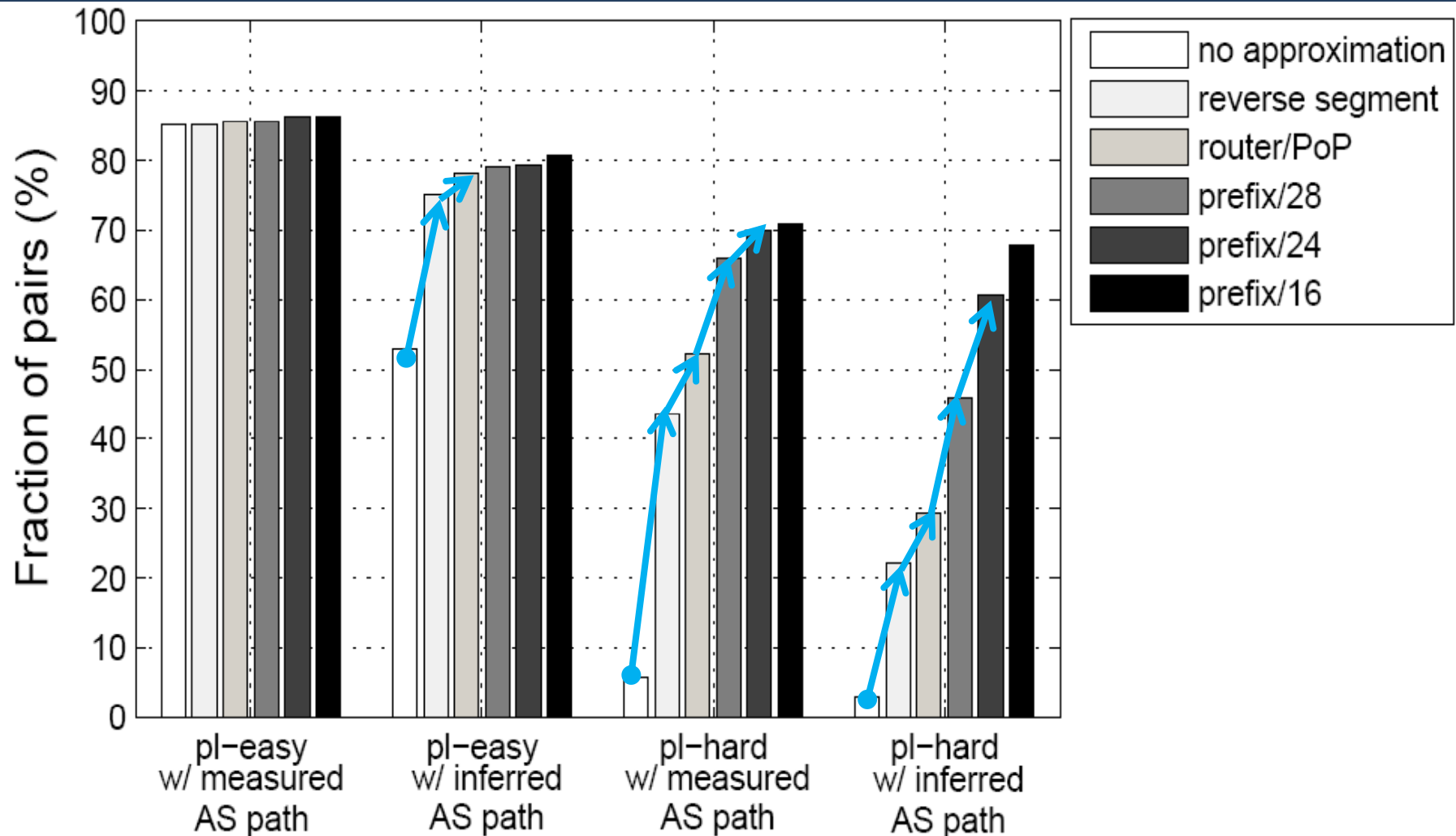
- Quality of Inferred AS Path
- Approximation methods
- Preference rules
- Accuracy in comparison with *iPlane* [Madhyastha *et al*, OSDI 2006]

AS Path Accuracy



Improvement in *pl-easy* pairs shows
the potential value of the additional information

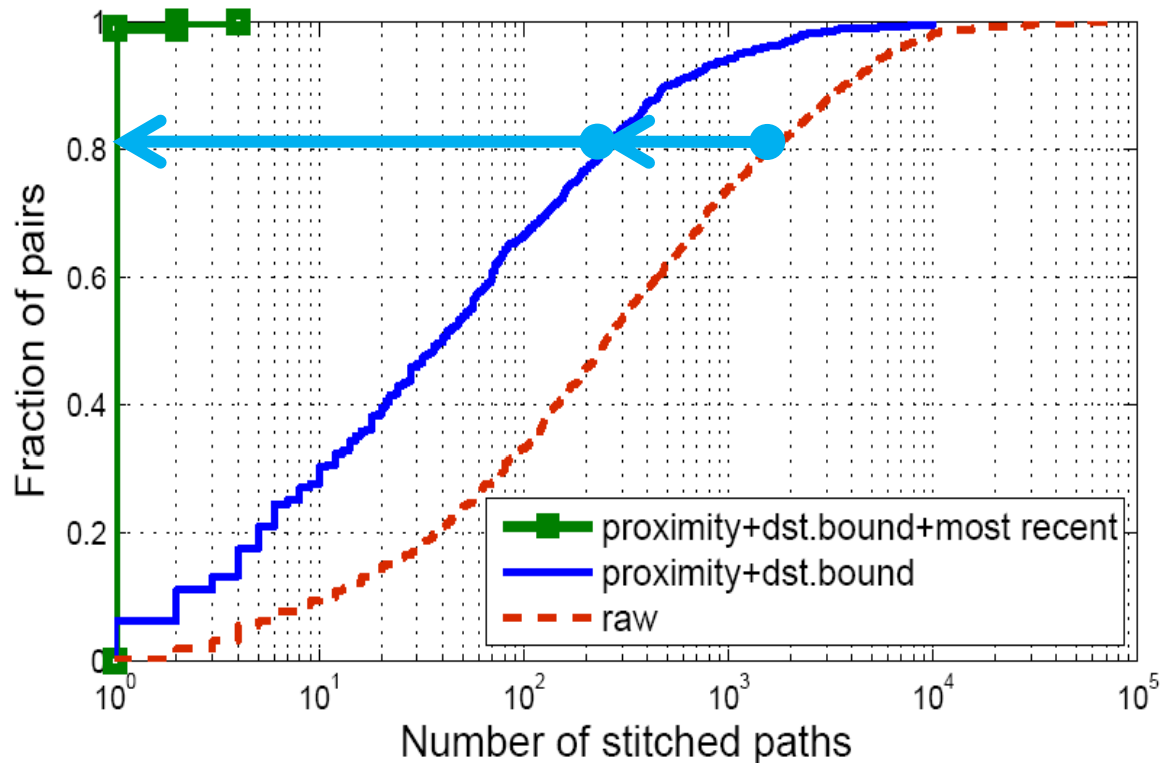
Approximations



As predicted, we show incremental improvement in the fraction of pairs with stitched paths

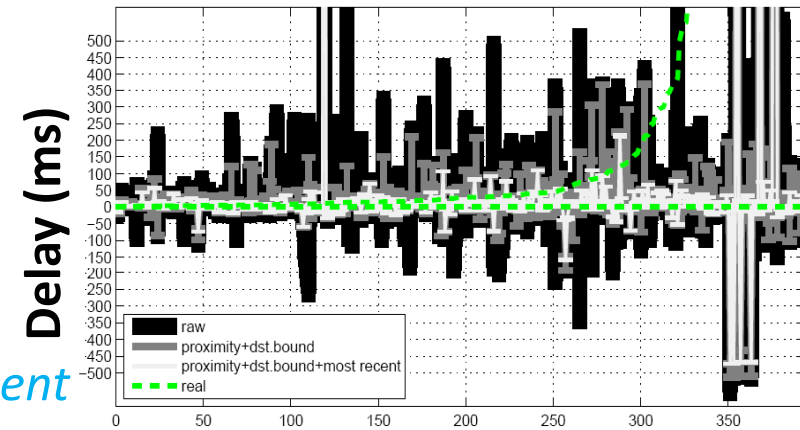
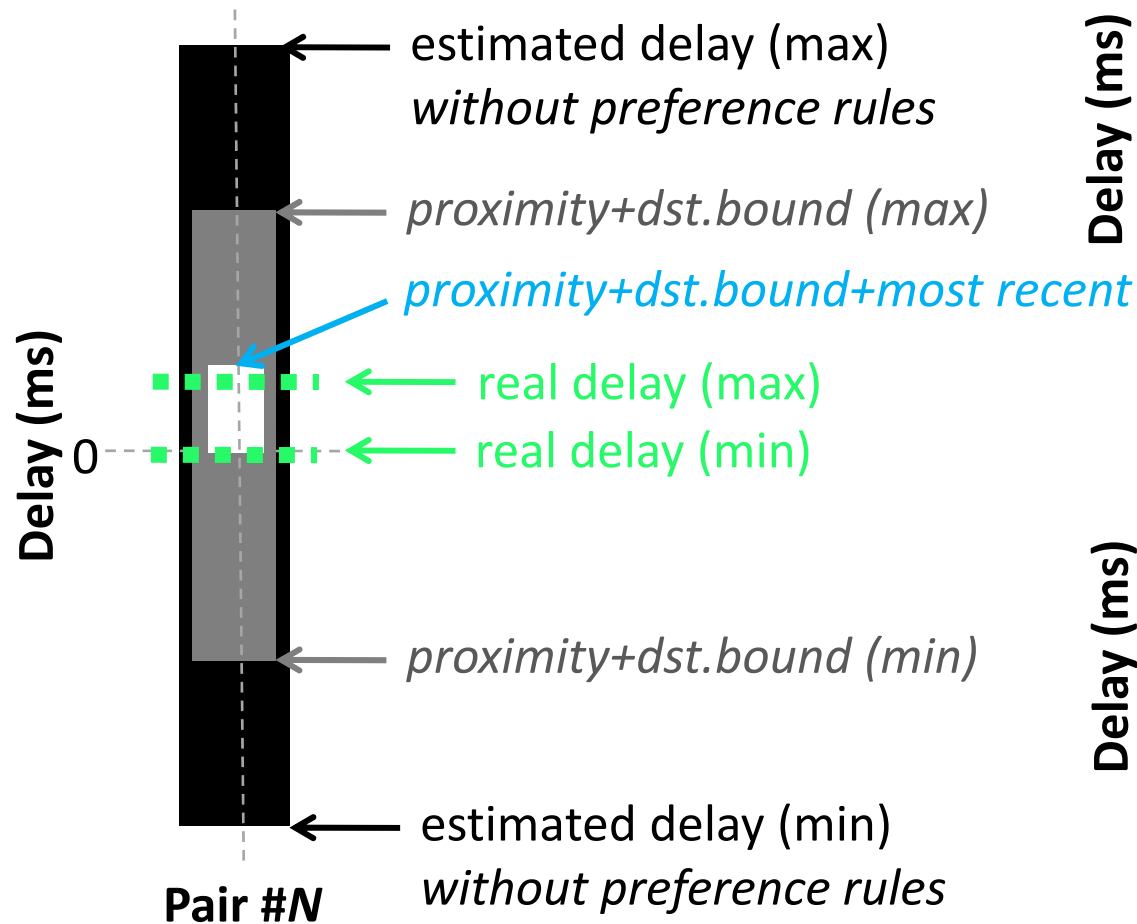
Preference Rules – (1)

- We consider only *pl-easy* and *pl-hard* pairs that find stitched paths without any approximation method.

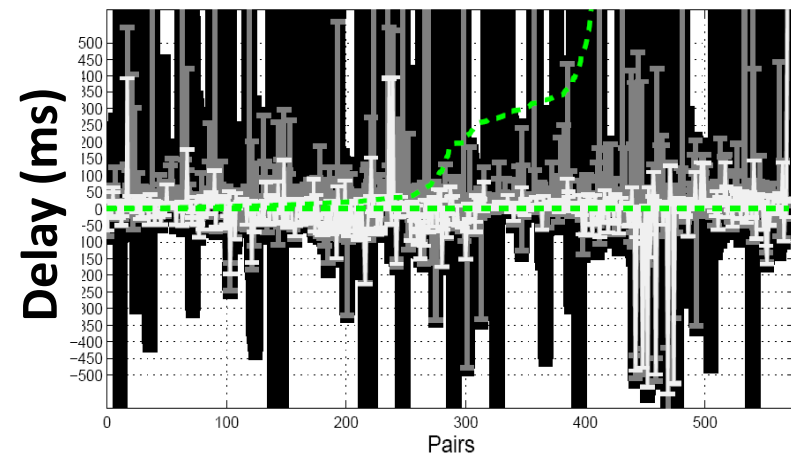


By applying preference rules, number of stitched paths decrease greatly.

Preference Rules – (2)



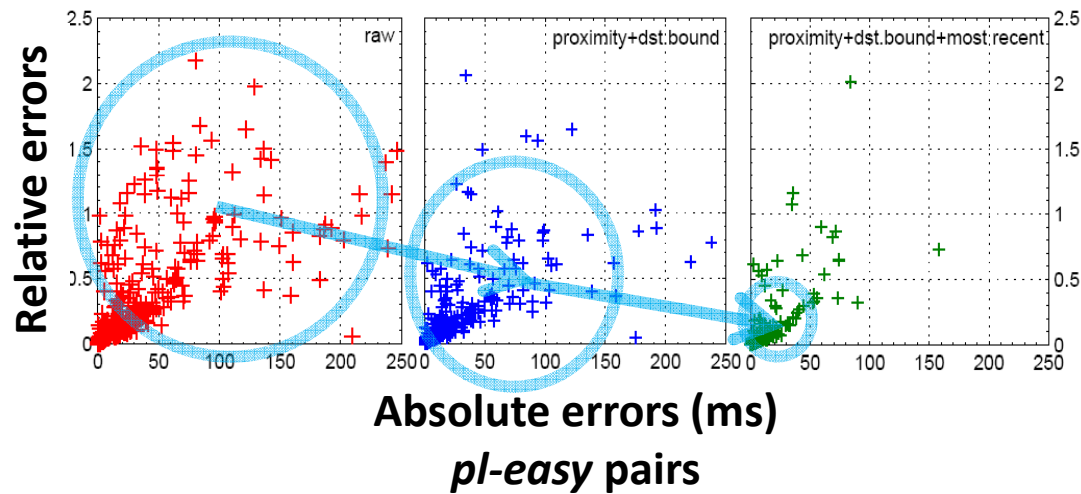
pl-easy pairs



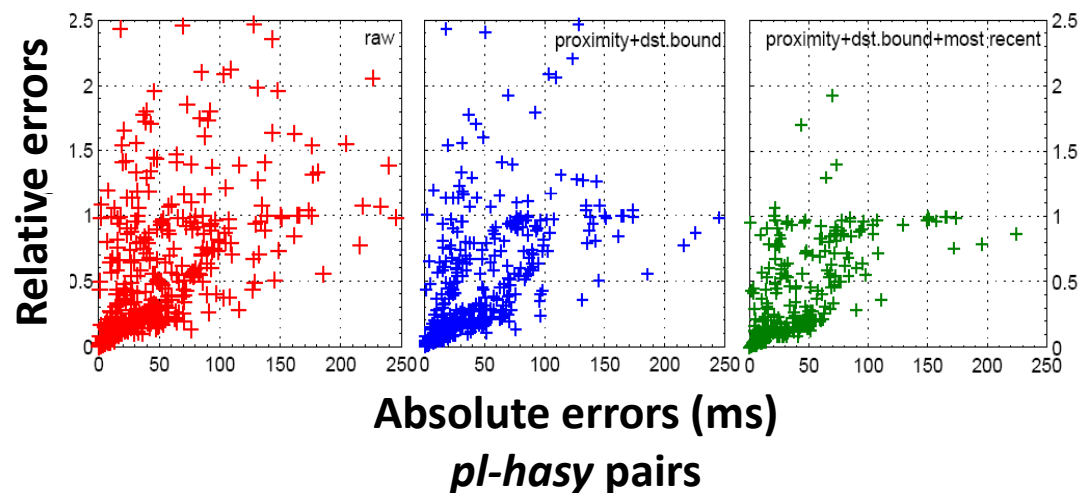
pl-hard pairs

Preference Rules – (3)

- Relative error vs. absolute error

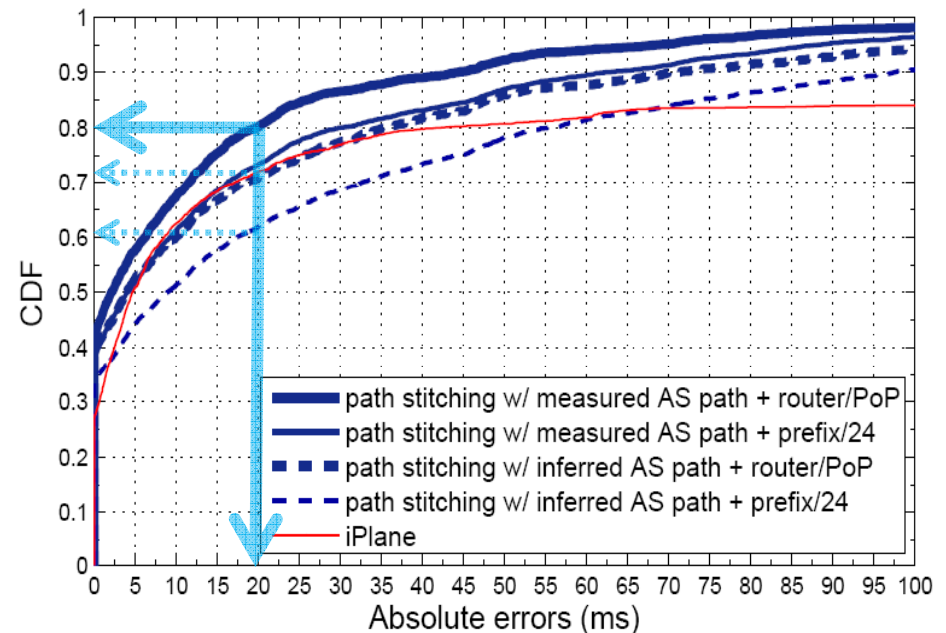
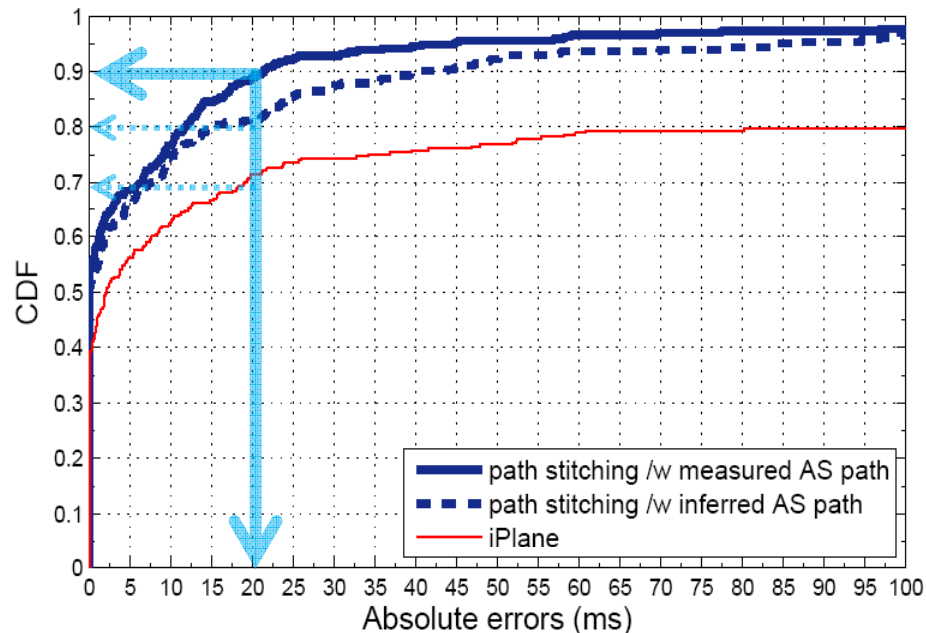


Improvements in *absolute errors* reflect similar improvements in *relative errors*



Comparisons with iPlane

- CDF of absolute errors



We note that iPlane's performance observed in our results is comparable to the best cases reported in [Madhyastha *et al*, OSDI 2006]

With measured AS paths, errors $\leq 20\text{ms}$ for 90% of pl-easy and 80% of pl-hard pairs
With inferred AS paths and approximation methods, accuracy degrades

Conclusions

- “path stitching”
 - A new approach to improve the coverage of Internet-wide measurement infrastructures.
 - Fully *decouples* the data collection phases from the data analysis
 - *Enables the incremental integration of multiple data sets* in order to produce more accurate estimates
 - Achieves an accuracy similar or slightly better than previous solutions that require additional data collection

Part II. “10 years in the Internet”

What are the basic rules that govern the growth of the Internet? How has it grown? At what rate? When and why did the rate change?

Pairs of Points to Examine – (1)

- We have 11 skitter/ark monitors in two continents: America and Asia.

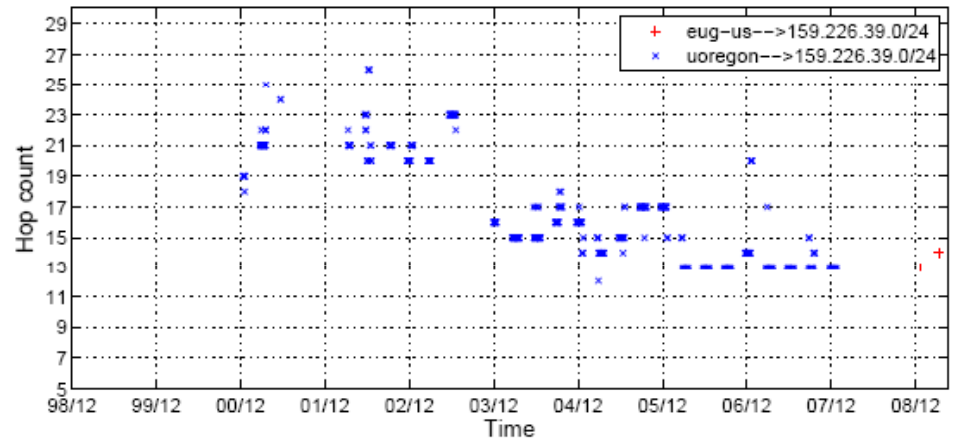
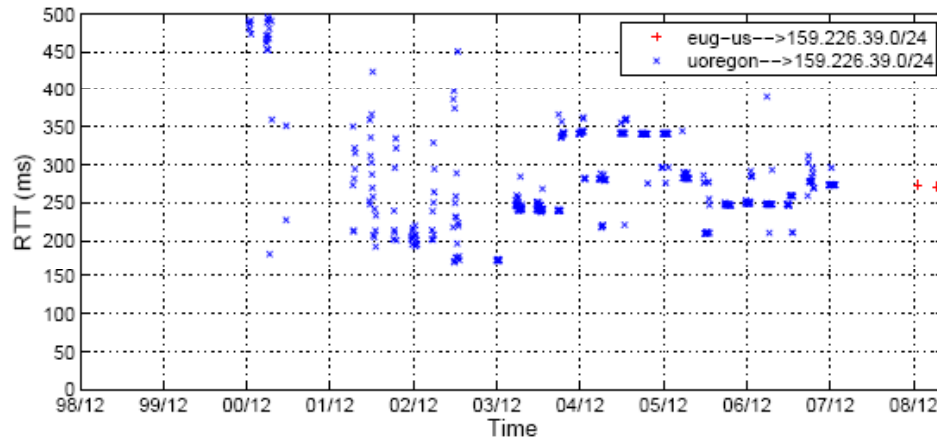
| Skitter/ Ark monitor | Organization | Location |
|-------------------------|---------------------------------------|---------------|
| riesling, priya, san-us | San Diego Supercomputer Center | San Diego, US |
| mwest, sjc-us | Verizon Business Network Services Inc | San Jose, US |
| uoregon, eug-us | University of Oregon | Eugene, US |
| yto, yto-ca | Canarie Inc | Ottawa, CA |
| apan-jp, nrt-jp | Asia Pacific Advanced Network - Japan | JP |

Table 1: Skitter/Ark monitors activated constantly for the past 10 years.

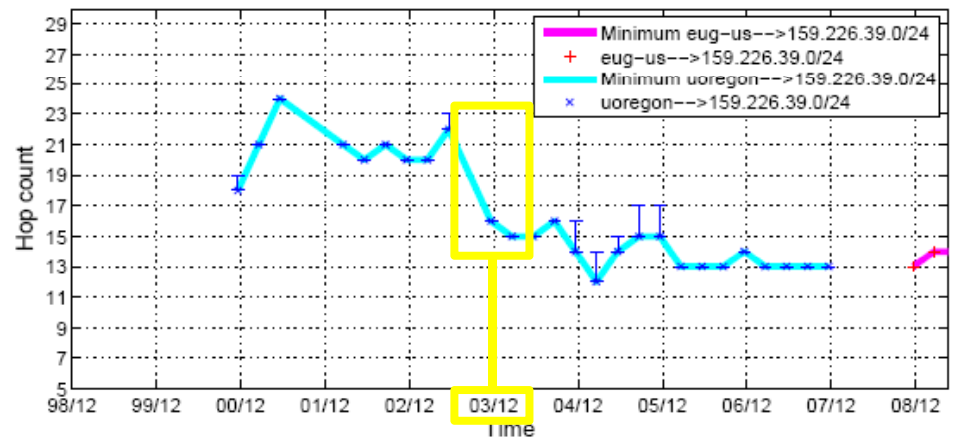
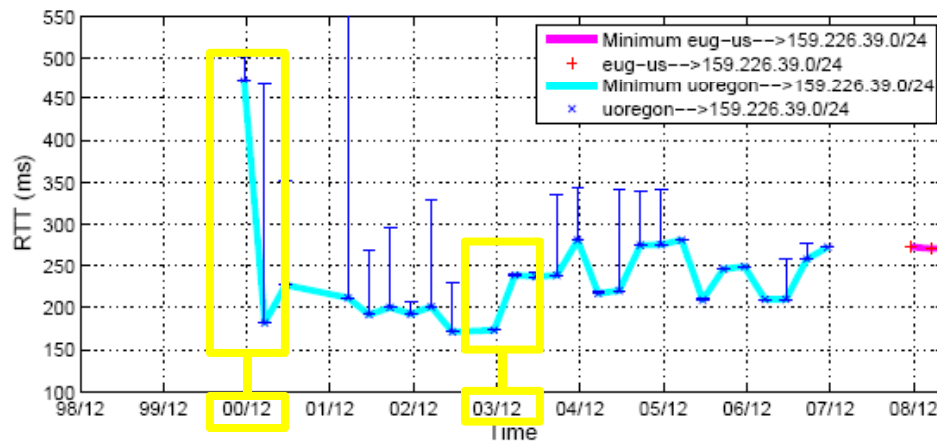
Pairs of Points to Examine – (2)

- We choose 977 destination prefixes (/24 or /23) from 14 different ASes in *five* continents: Europe, Africa, Australia, Asia, and America.
- Those destination prefixes are used for the past 10 years.
- Then, we have 10 ($5C2$) inter-continental pairs to examine.
 - **7 pairs** are directly covered by Skitter/Ark dataset
 - **Other 3 pairs** (Europe to {Australia, Africa}, and Australia to Africa) need to be estimated by **path stitching**

From US to China



(a) RTT and Hop count: Scatter plots



(b) RTT and Hop count: Errorbar plots

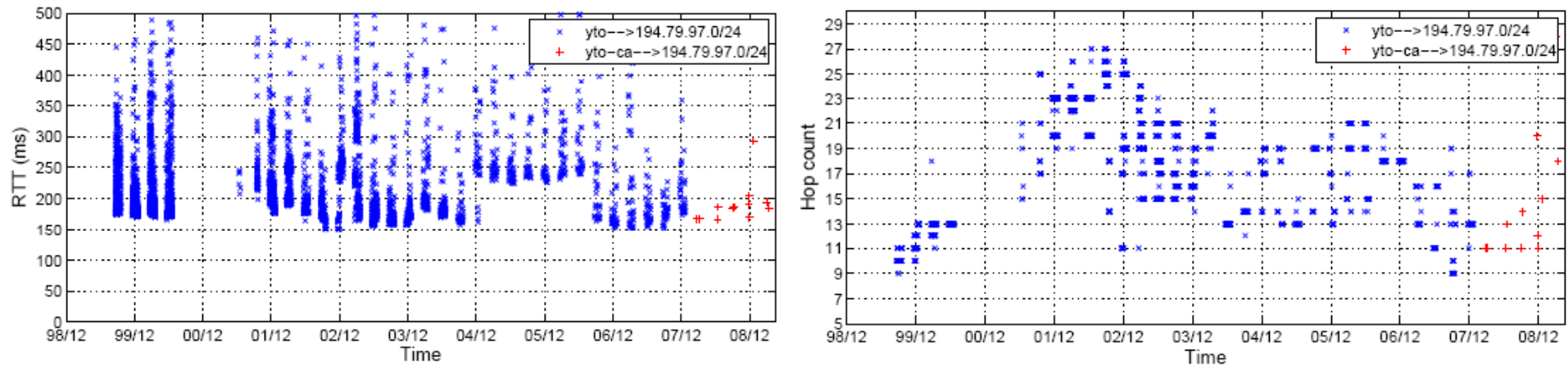
China-US Cable Network (2000)

- Major upgrades in the Internet

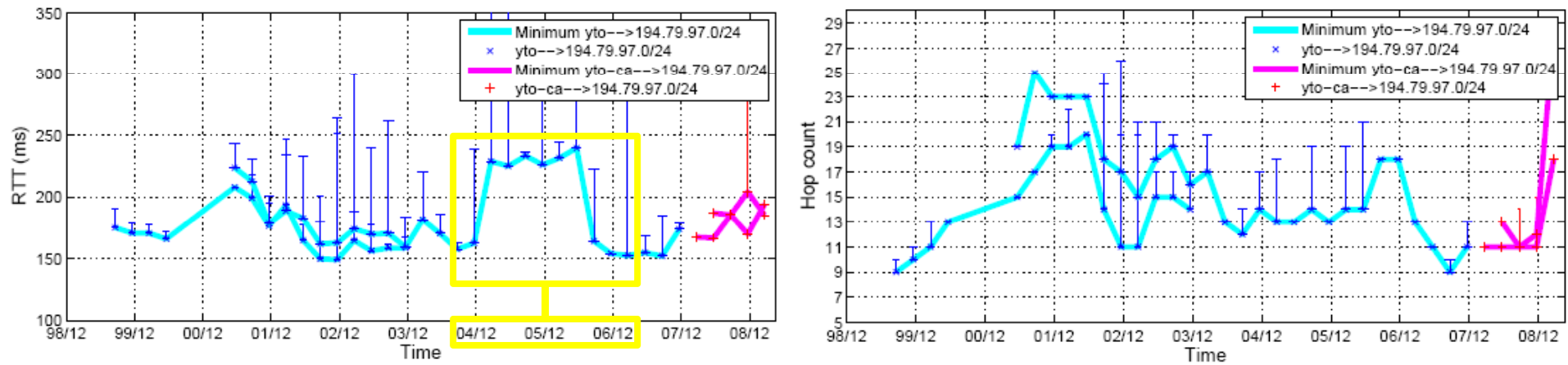


Figure 7: China-US Cable Network. 2000. The first un-dersea fiber optic cable network between US and China. [should be redrawn for the publication]

From US to Egypt

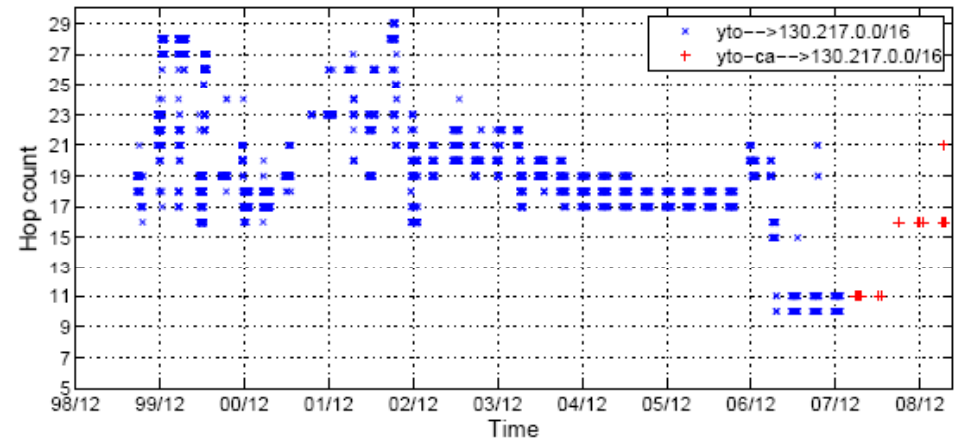
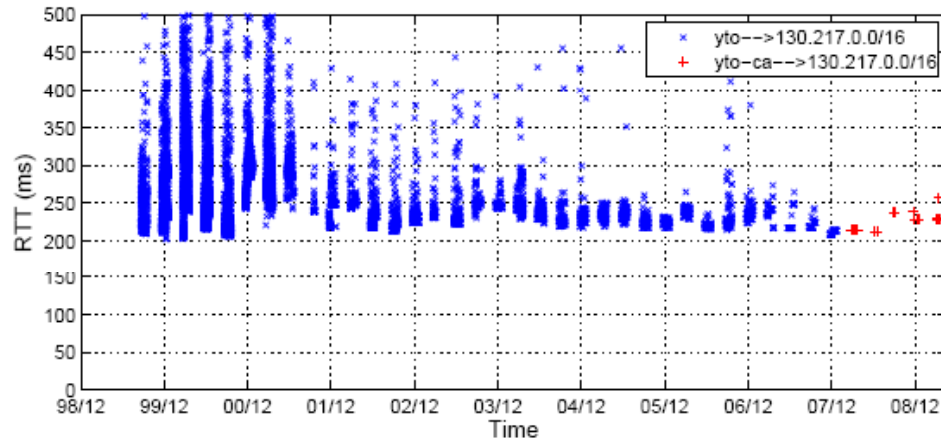


(a) RTT and Hop count: Scatter plots

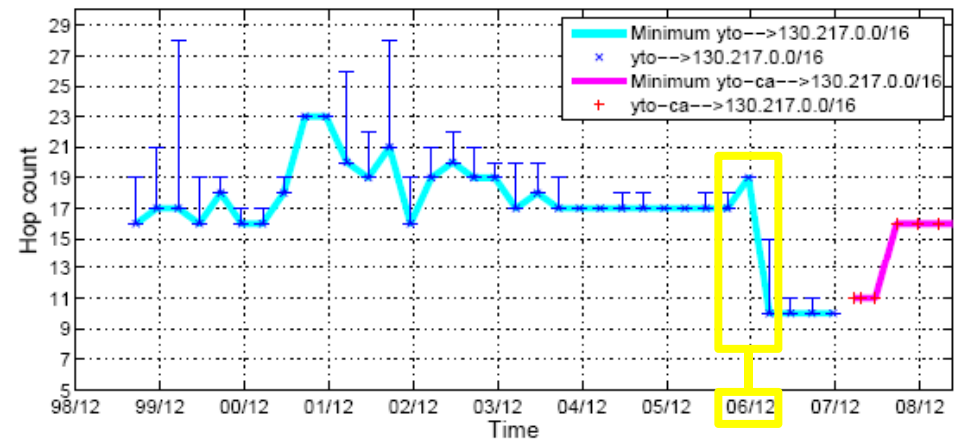
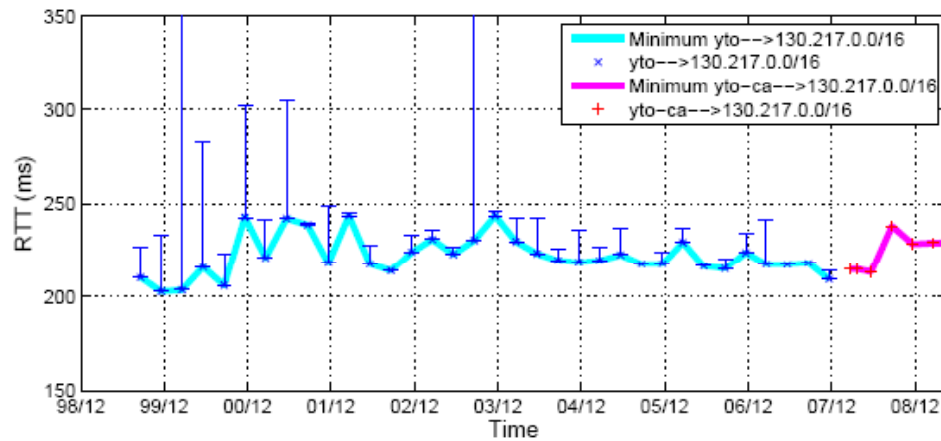


(b) RTT and Hop count: Errorbar plots

From US to New Zealand



(a) RTT and Hop count: Scatter plots



(b) RTT and Hop count: Errorbar plots

Trend Detection – (1)

- What are the basic rules that govern the growth of the Internet? How has it grown? At what rate? When and why did the rate change?
- Has the Internet grown shorter in delay?
 - (1) Are we reaching the speed limit of light yet?
 - (2) Did short delay play a role in proliferation of p2p applications?
Is the Internet ready for the next big killer application?

Trend Detection – (2)

- Instead of specific sites, can we review overall Internet performance historically?
- Has the Internet grown smaller in the number of AS and IP hops?
- Has the Internet grown larger in terms of the Internet diameter? (e.i., average latency, average path length)
- Has the growth been even? Or skewed? Regional vs. international?

Pairs of Points to Examine – (3)

- Internet sampling ideas

Thank You!

- Any question?
- We would appreciate any feedbacks on our ongoing work:
dklee@an.kaist.ac.kr

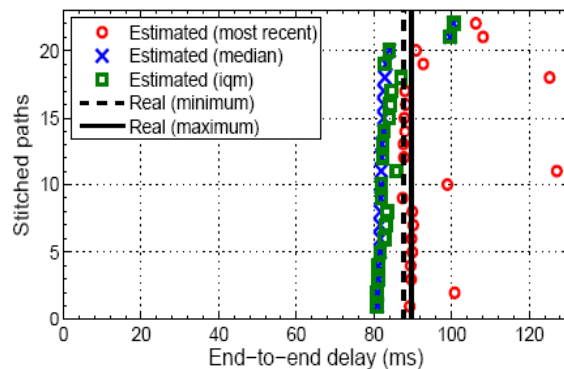
Appendix I. Backup Slides

*“To get to the essence of things,
one has to work long and hard”*

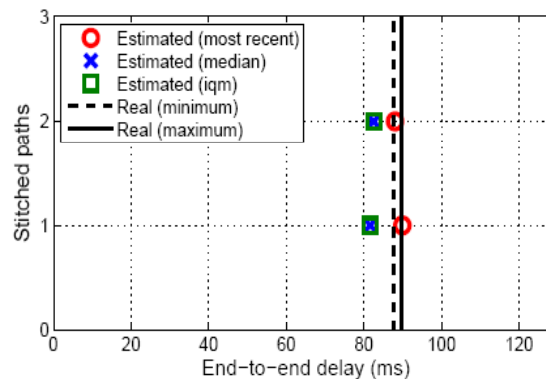
-- Vincent van Gogh

Finding Clues to Preference Rules

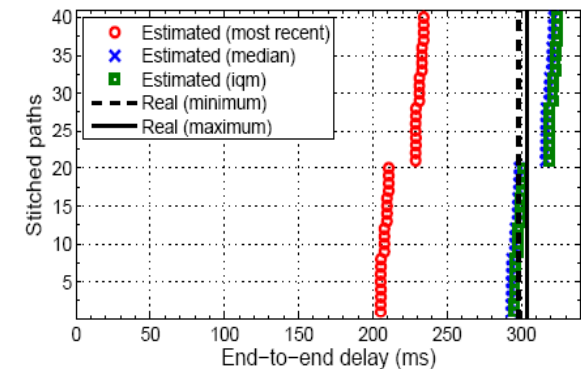
- Two examples to demonstrate differences between stitched paths



(a) planetlab1.csail.mit.edu → planet2.scs.stanford.edu



(b) filtering planetlab1.csail.mit.edu → planet2.scs.stanford.edu



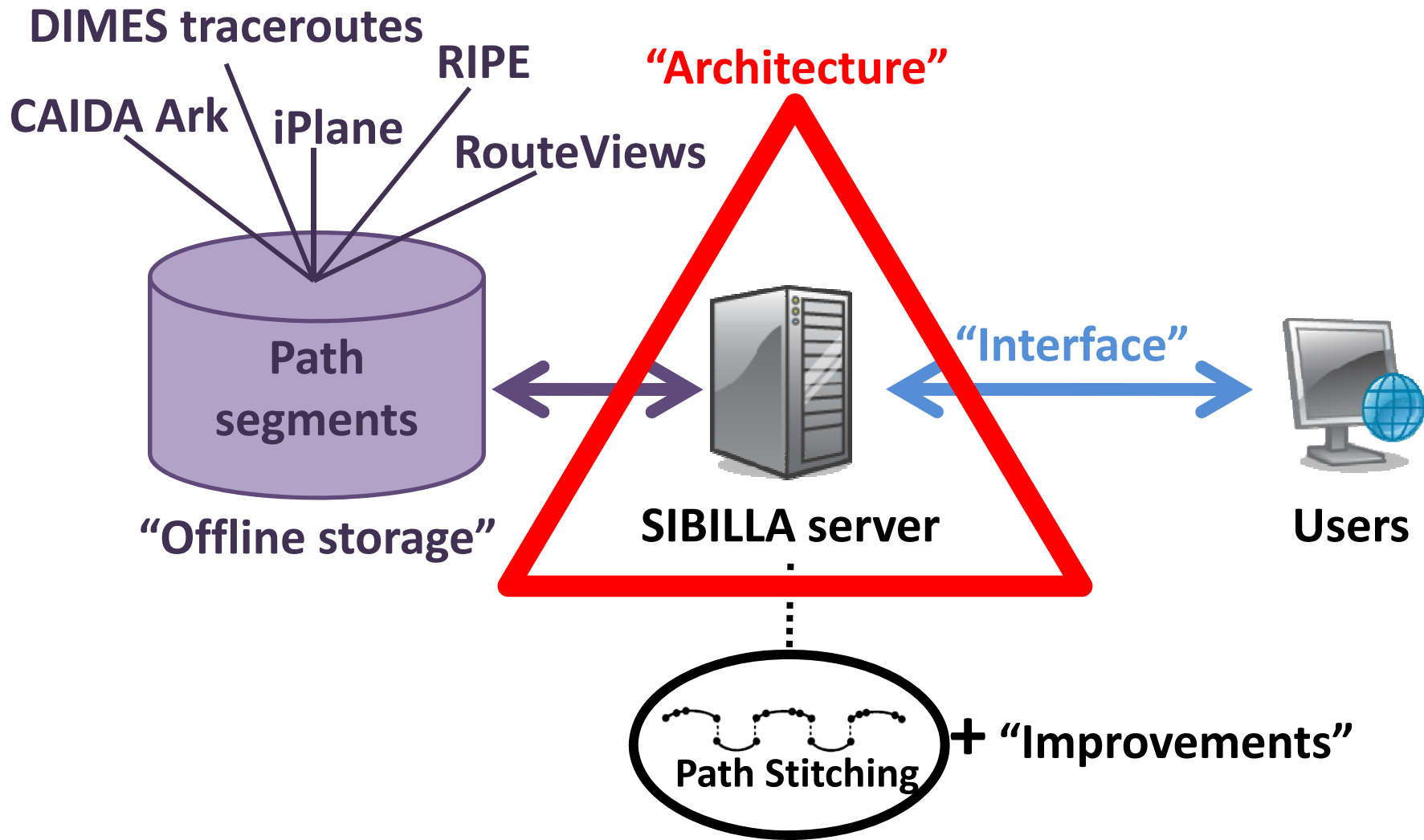
(c) planetlab2.xeno.cl.cam.ac.uk → pl1-higashi.ics.es.osaka-u.ac.jp

Prefixes with MOAS conflicts

| IP Prefix | MOASes in the same country | % |
|------------------|----------------------------|-------|
| 205.189.33.0/24 | 6327 6509 | 26.18 |
| 134.75.20.0/24 | 1237 17579 | 25.41 |
| 207.231.241.0/24 | 293 14221 101 | 3.26 |
| IP Prefix | MOASes caused by IXPs | % |
| 80.81.192.0/23 | 12956 8365 | 10.90 |
| 198.32.176.0/24 | 701 2914 65517 4355 6461 | 7.72 |
| 198.32.160.0/24 | 6461 22691 12989 | 3.58 |
| 206.223.115.0/24 | 293 2914 1273 | 3.02 |
| 195.69.144.0/23 | 286 12956 1200 30132 31283 | 2.78 |
| 206.223.119.0/25 | 2914 293 | 2.33 |
| IP Prefix | Other MOASes | % |
| 69.28.128.0/18 | 22822 21318 | 4.85 |

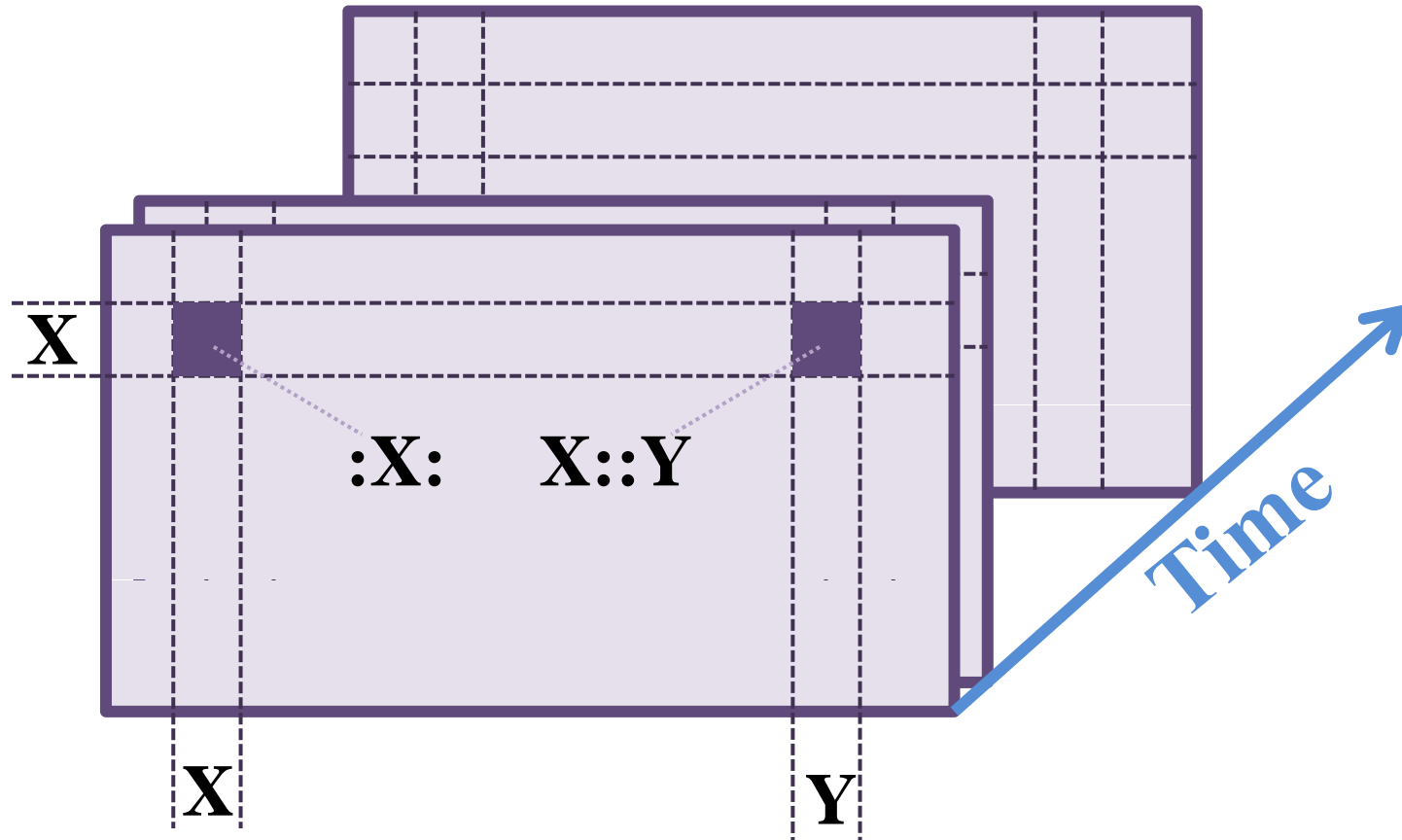
Table 4: Prefixes with MOAS conflicts. The percentage refers to the portion of the total number of traceroutes that exhibit a MOAS conflict.

Beyond the "Path Stitching" algorithm



Storage Model: BigTable

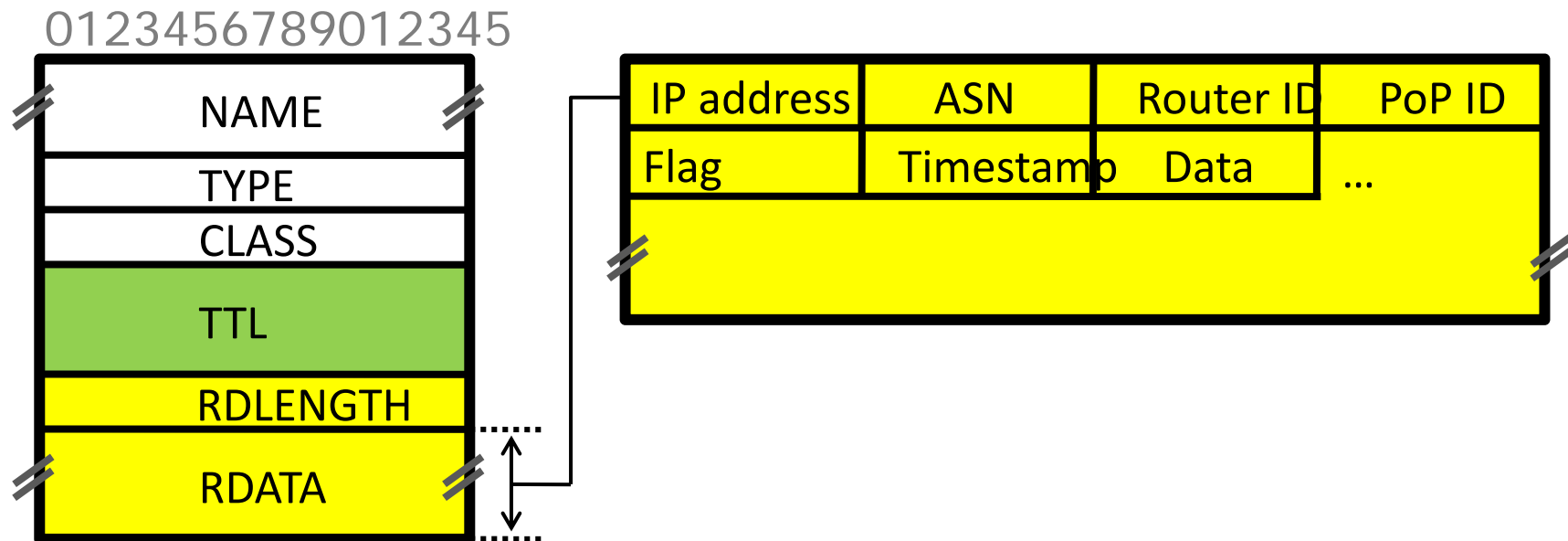
- Storage for massive amount of path segments
(row: ASN, col: ASN, time: int64) → path segments



Query Interface - (2)

- **Responses:**

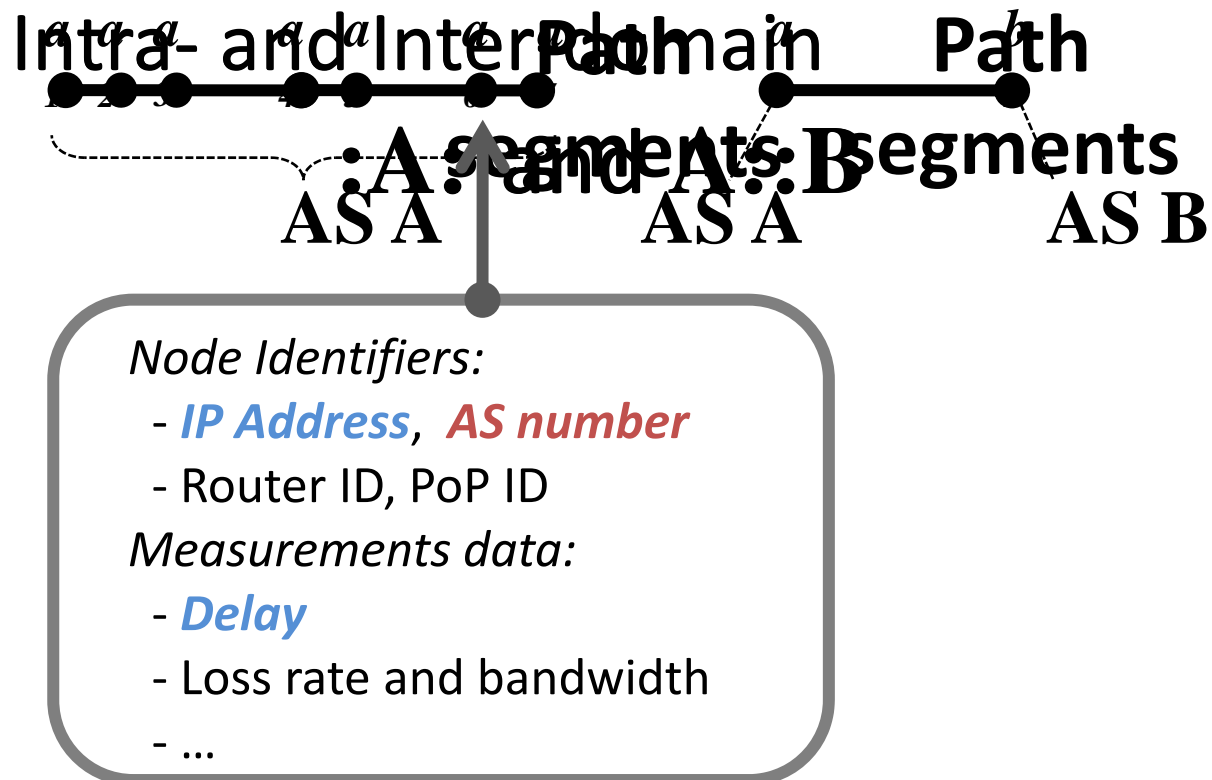
- Exploit Resource records (RR, record type: A or AAAA)



- We may define special record types: PATH or DELAY
- *EDNS* for messages larger than 512 bytes. (RFC 2671)

Path Segments: Unit of Data Storage

- Tradeoffs: information loss vs. size vs. efficiency



Big Table Clones

- ***HBase***

- As a part of Apache Software Foundation's *Hadoop* project
- Implemented in Java language

- ***Neptune*** by NHN

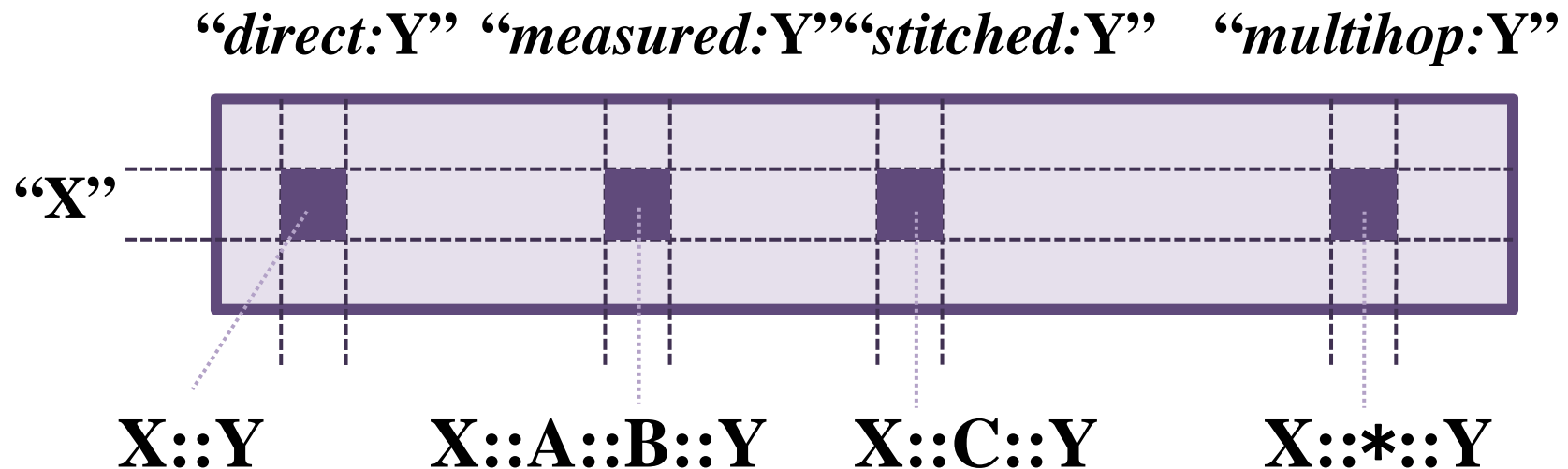
- ***HyperTable*** by Doug Judd

- Implemented in C++, Open source project
- Built on Hadoop file system

- <http://www.hypertable.org>

Offline Storage – Open Issues

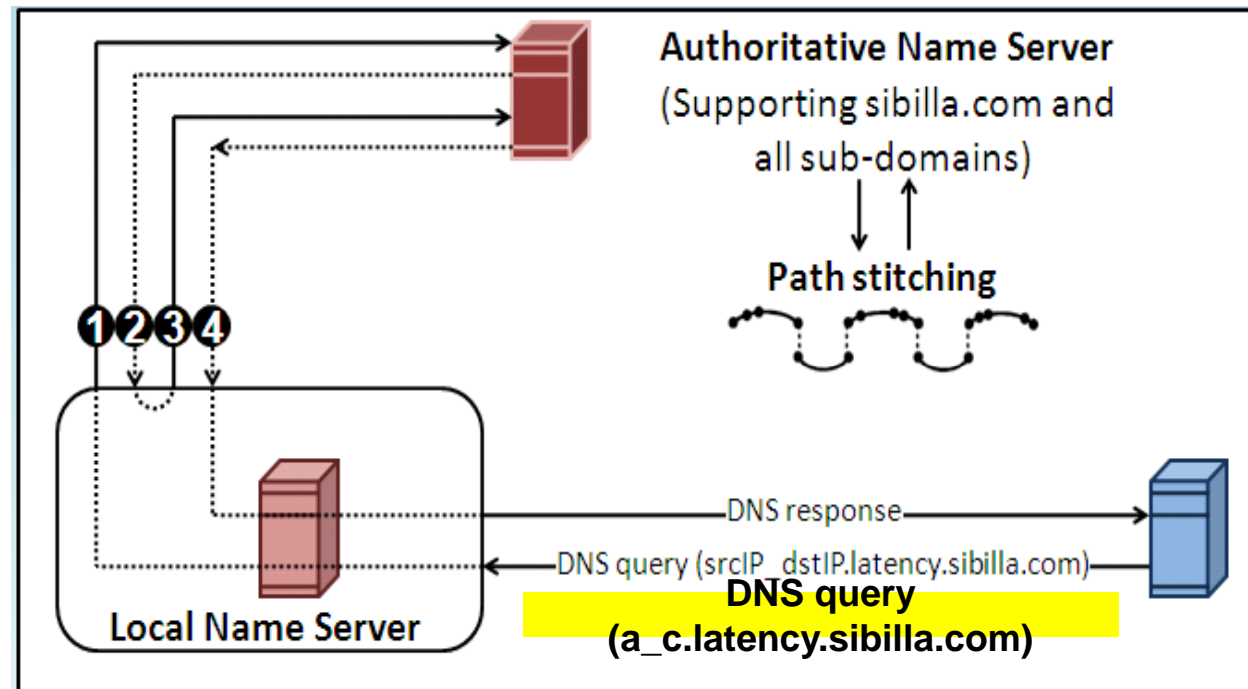
- Column families?



- Implementation (1 month)
- Performance evaluation

For the 100 % Response Rate

- When a client queries from *itself* to somewhere.



- When a client queries between two arbitrary hosts.
 - Additional data source is the solution.

Architecture

- To be *Peer-to-peer* or not to be?
- Advantages of P2P
 - Low budget requirement
 - Availability
 - Anonymity
- P2P is not appropriate for applications that need
 - lower latency
 - more than just distributed hash tables