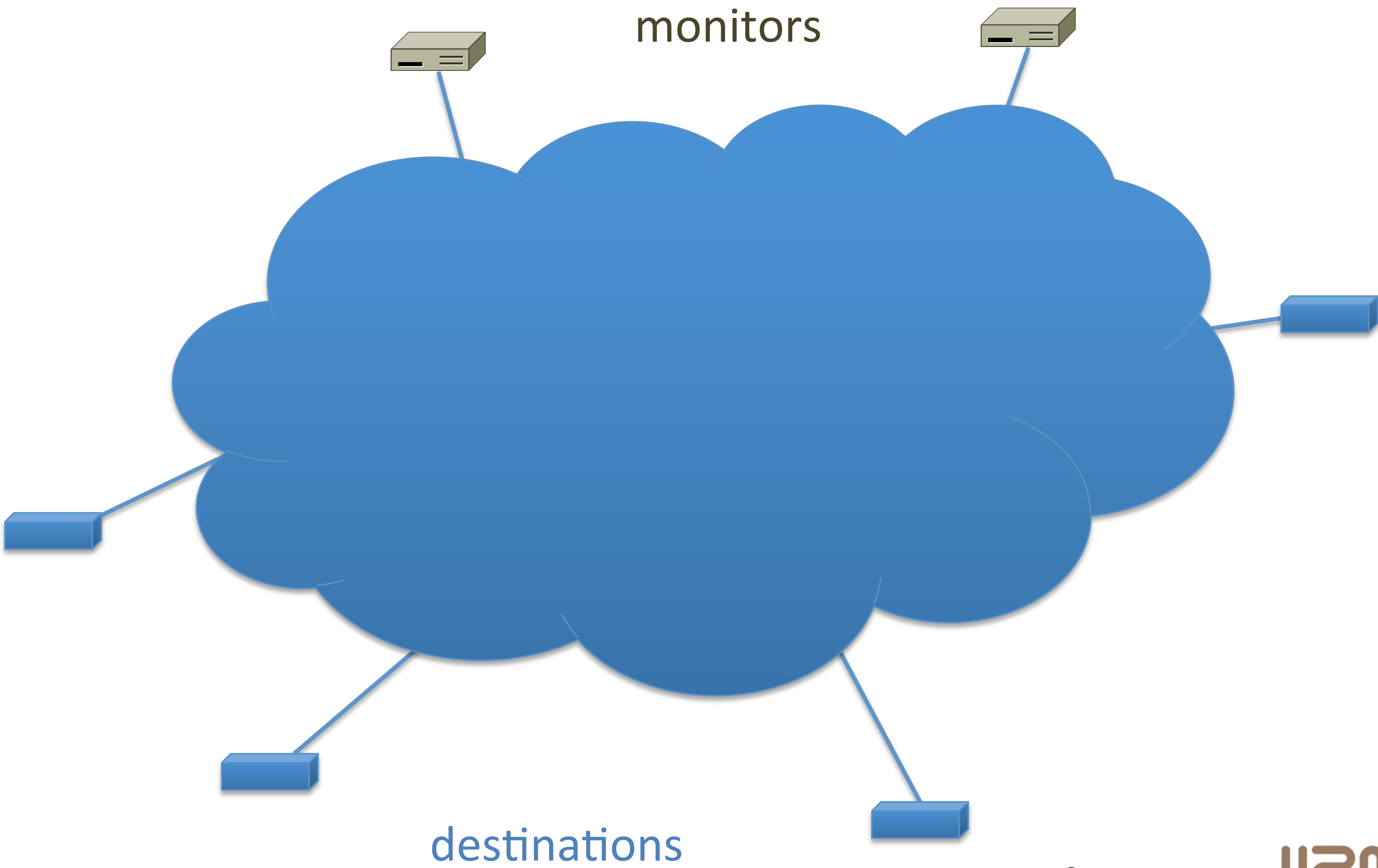
LINCS internal seminar, 26 January 2011

Towards improved measurement systems to capture internet dynamics

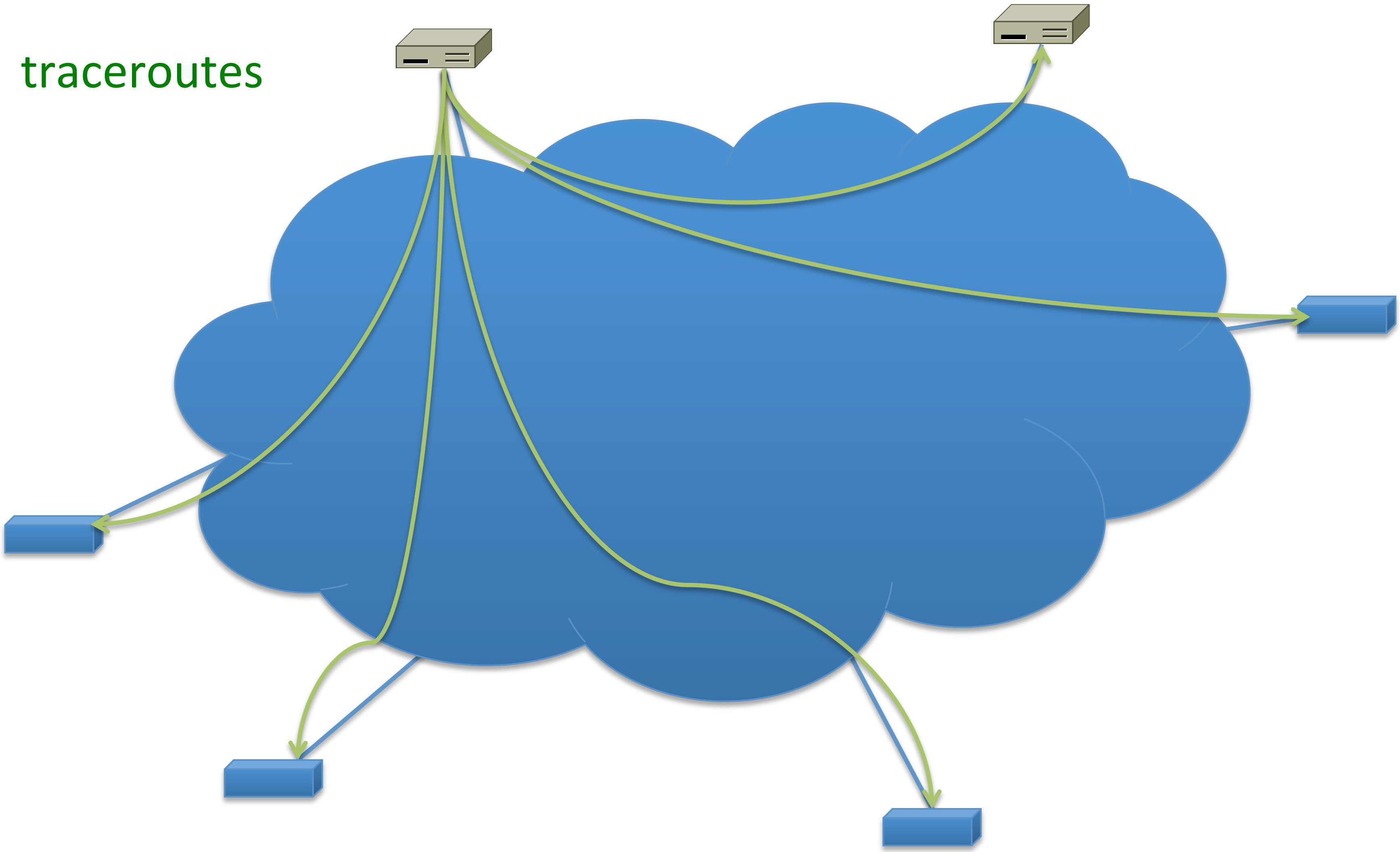
Timur Friedman

Distributed network measurement systems

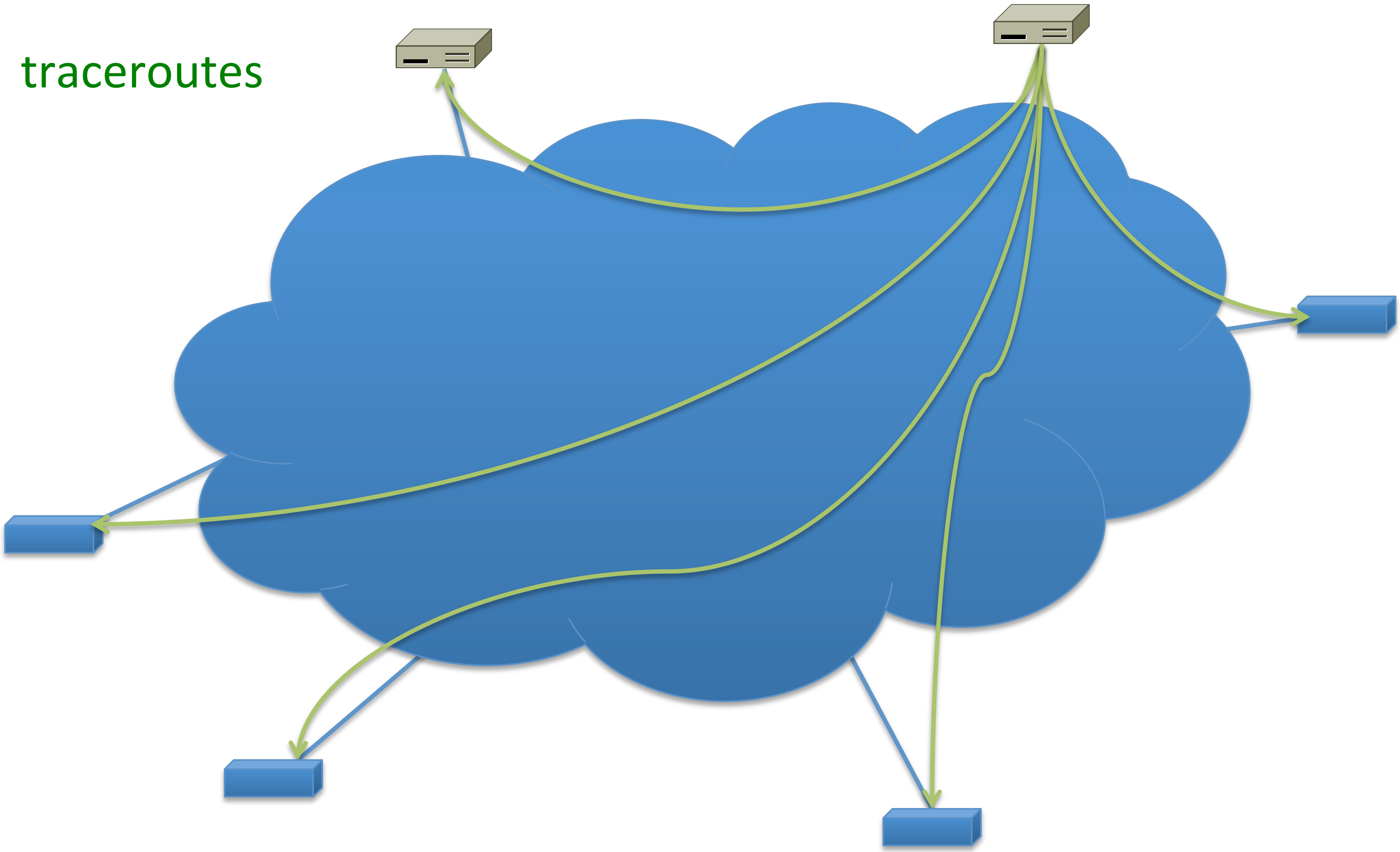


UPMC SORBONNE UNIVERSITÉS

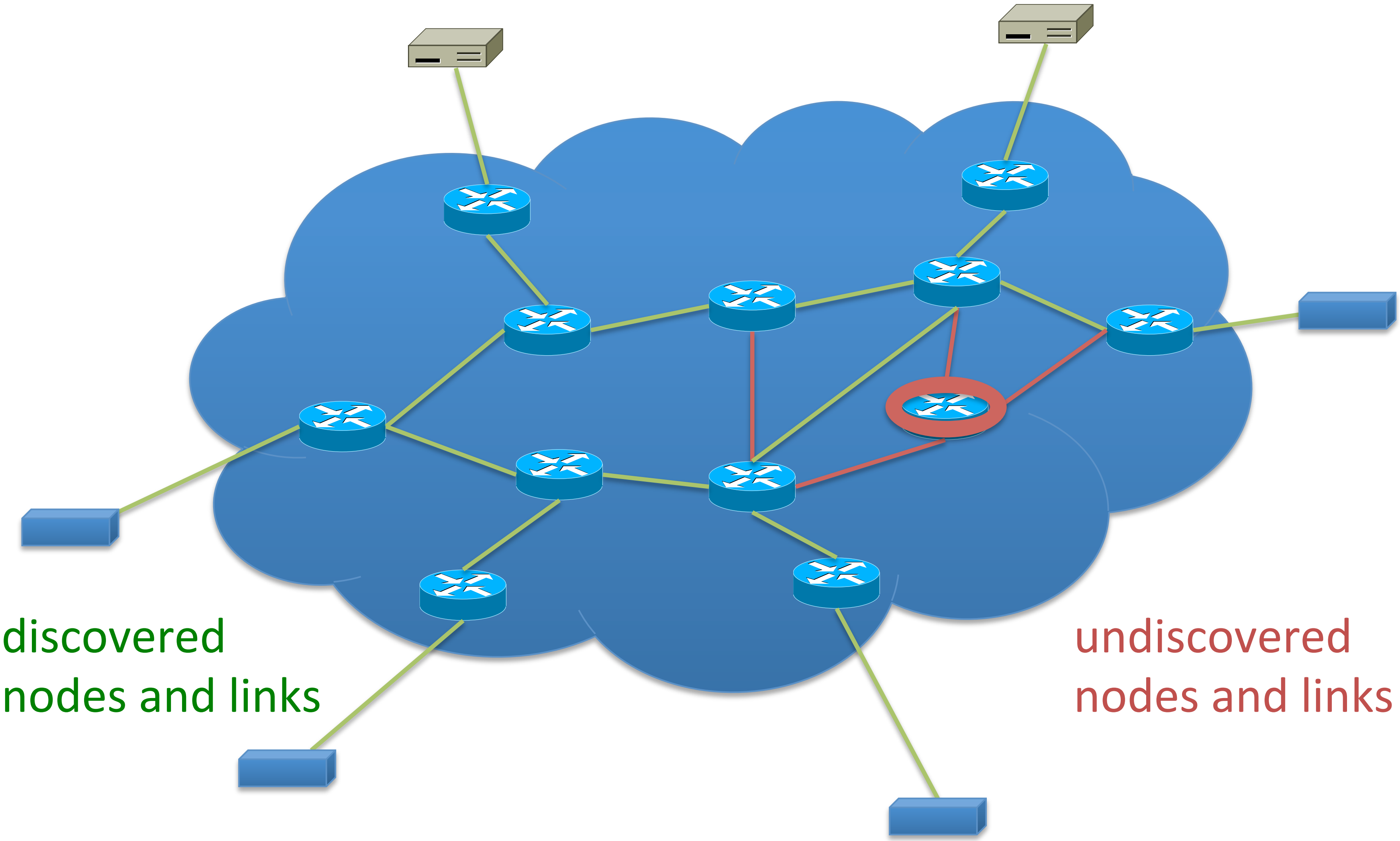
Distributed network measurement systems



Distributed network measurement systems



Distributed network measurement systems



discovered
nodes and links

undiscovered
nodes and links

UPMC SORBONNE UNIVERSITÉS

Principal measurement systems

Project	Institution	Number of monitors	Number of destinations	Measurement frequency
Archipelago (aka Ark, formerly Skitter)	CAIDA center at UC San Diego	45	9.1 million (all /24 prefixes)	2-3 days
EdgeScope (Ono BitTorrent plug-in)	Northwestern University AquaLab	~ 800,000	~ 800,000 (40,000 networks)	unknown
DIMES	Tel Aviv University	~ 1000	~ 10^5	7 days
iPlane	Washington University	~ 300	~150,000	1 day
TTM	RIPE	~ 200	~ 200	6 min

Why measure?

Obtain fundamental understanding

- What is the structure of the internet?
  - Graph properties: small world

Guide protocol design

- Multicast, content distribution, p2p, overlay protocols, etc. depend on assumptions about network structure
- Topology generators for simulators

Guide network planning

- Is the network robust against failure?
- How does it reflect demographics?

Challenges

Completeness

- Are we seeing all of the network?

Accuracy

- Are we getting an accurate picture of what we do see?

Efficiency

- Are we using our probing resources to maximum effect?

Challenge: Completeness

Lakhina et al. sampling bias work

INFOCOM 2003

- Measuring from too few vantage points can in principle introduce biases in the inferred graph properties.

Spring, Mahajan, and Wetherall Rocketfuel work

SIGCOMM 2002

- With enough vantage points and good techniques, we can get full and accurate maps of ISPs.

Shavitt & Weinsberg

INFOCOM 2009

- Broad distribution of vantage points can in actual fact yield good estimates of graph properties.

Challenge: Accuracy

Teixeira et al. path diversity work

IMC 2003

- Rocketfuel topologies suffer from many inaccuracies.

Augustin et al. Paris Traceroute work

IMC 2006

- Classic traceroute was producing inaccurate and incomplete maps. Largely corrected with Paris Traceroute.

Katz-Basset et al. Reverse Traceroute work

NSDI 2010

- Use the IP Record Route option to obtain accurate paths.

Challenge: Efficiency

Govindan & Tangmunarunkit Mercator work

INFOCOM 2000

- When tracing from a single vantage point, trace backwards from the destinations and stop when you encounter familiar nodes.

Donnet et al. Doubletree work

SIGMETRICS 2005

- When tracing from multiple vantage points, trace both backwards and forwards from a medium distance. Communicate between monitors to know when to stop forward tracing.

A new challenge

Capture network dynamics

- Observe network changes over time
  - Long timescale: understand the evolution of the internet
  - Short timescale: detect routing changes, system maintenance, failures, attacks, etc.
- Current timescales:
  - days (Ark, DIMES, iPlane)
  - minutes (TTM) – problem: TTM is just a small mesh

Can we capture graphs at Ark-scale every few minutes?

Our aim: 1000x speedup.

UPMC SORBONNE UNIVERSITÉS

‘La vérité’

La photographie c’est la vérité. Et le cinéma c’est vingt-quatre fois la vérité par seconde.

- Bruno Forestier, dans *Le Soldat* de Jean-Luc Goddard (1960)

The scope of the challenge

Size of the graph

largest measured graphs today:

15 million nodes, 60 million links

- IPv4 allows for up to 4.3 billion addresses
  (9.1 million /24 prefixes)
- IPv6 allows for up to 3.4×10^{38} addresses

Frequency

speedup in two phases:

- 50x speedup from 2 days to 1 hour
- 20x speedup from 1 hour to 3 minutes

Bandwidth and storage back-of-the-envelope calculations

- one traceroute query: 40 byte probe, 36 byte reply: 76 bytes total
- 120 million probes = 9.12 GB
  - 9.12 GB / 3 minutes = 405 Mbps
- for reference, 3 years of Ark data consumes 3.1 TB

The scope of the challenge

Back-of-the-envelope calculations

Bandwidth

- one traceroute query:
  - 40 byte probe
  - 36 byte reply
  - 76 bytes total
- 120 million probes = 9.12 GB
  - 9.12 GB / 3 minutes = 405 Mbps
- ➢ 405 Mbps for a distributed system is entirely possible

Storage

- 3 years of Ark data consumes 3.1 TB
- ➢ 1000 TB/year would be daunting
  but significant compression should be possible

UPMC SORBONNE UNIVERSITÉS

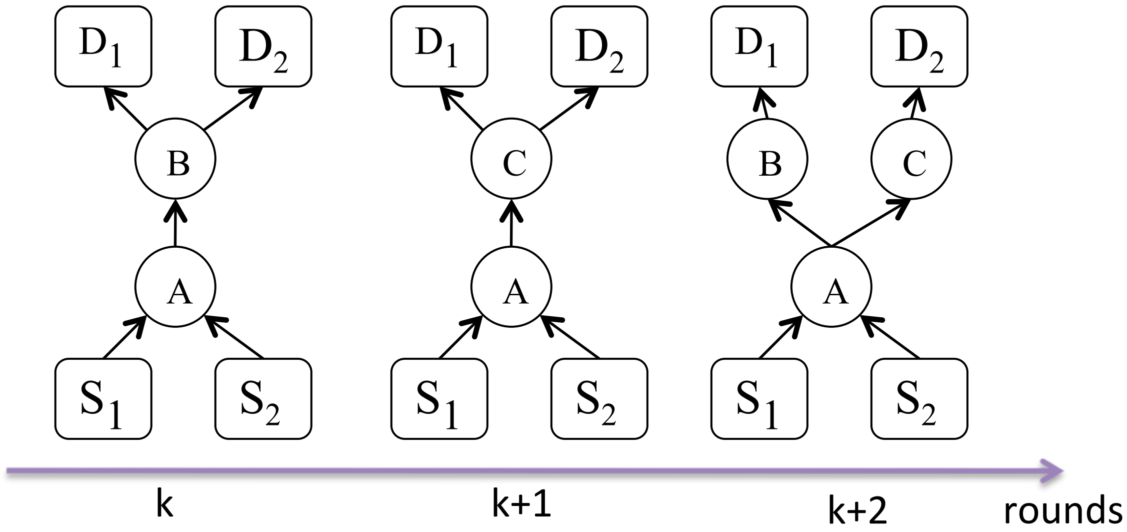
Preliminary work

Research recently begun by Thomas Bourgeau

Initial study:

- How much information is being lost by probing slowly?
- What simple algorithms would allow us to speed up capture?

What we study



- Consecutive rounds of measurements
  $k = 1, 2, 3, \ldots$
- Assemble a graph for each round
  $G_k = (V_k, E_k)$
- Discovered for each round
  nodes V_k and links E_k

Measurement setup

TopHat measurement system on PlanetLab

230 sources

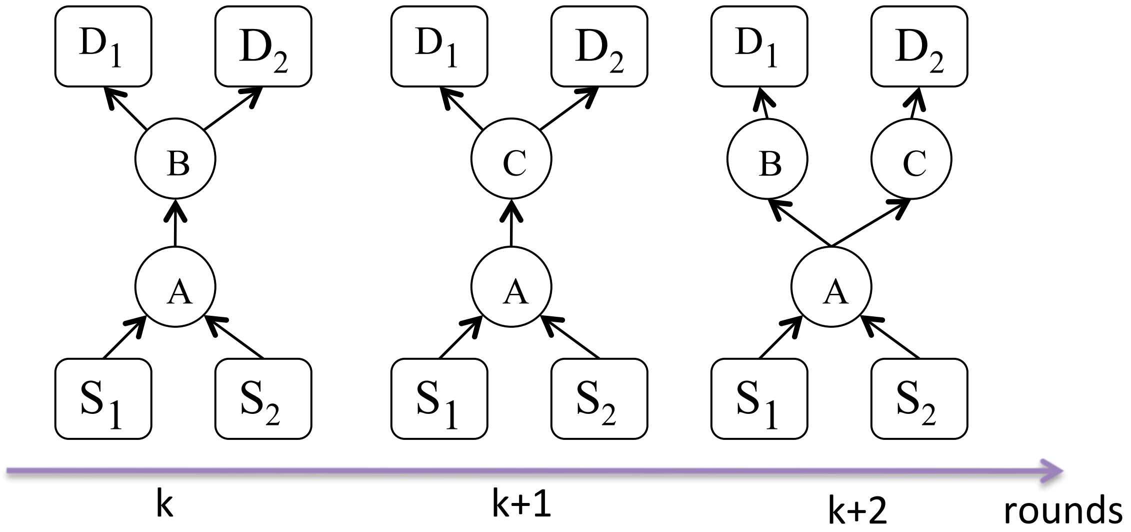
800 destinations

one measurement round per hour

- F_1 (fine grained) time resolution

from 25 May to 25 July 2010

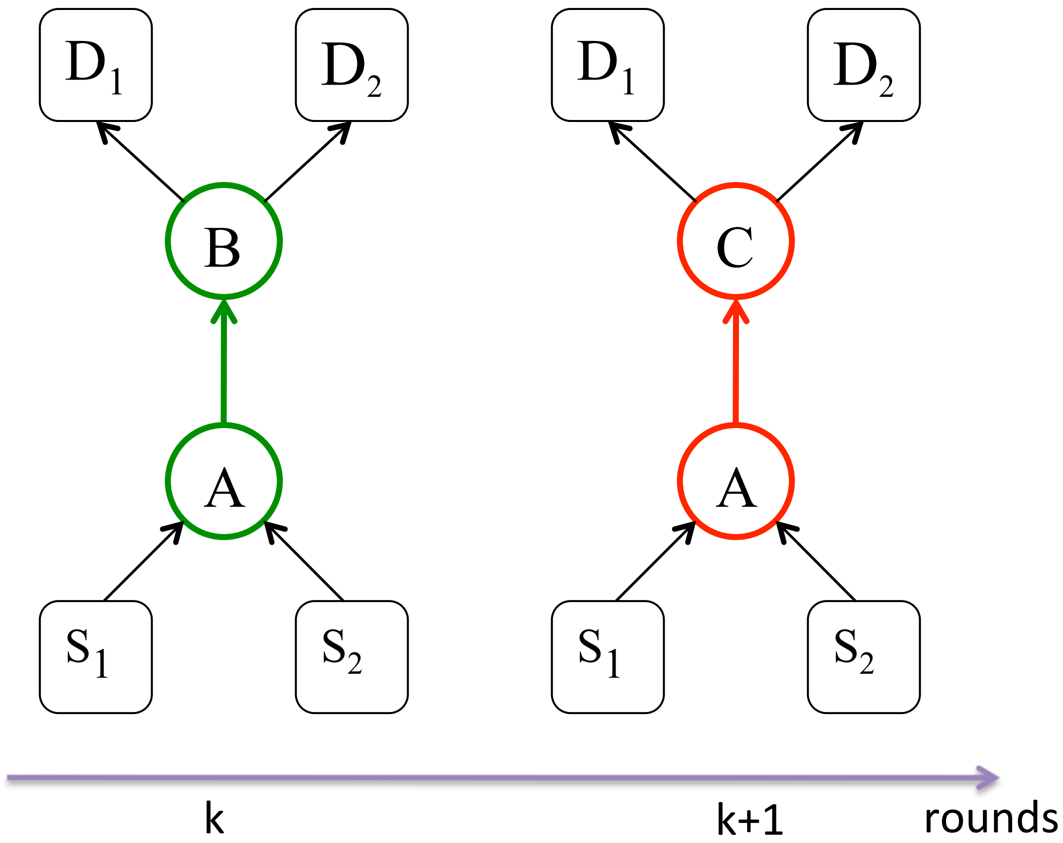
What we study



- Measurement experiments:
  - Each experiment rounds (k) is a full-mesh traceroute probing between a fixed sources-destinations set at a measurement frequency F_n.
- Dynamism observation:
  - Compare consecutive measured traceroute graphs $G_k(V_k,E_k)$

Appearances and disappearances

Event: appearance or disappearance of a node or link between consecutive measurement rounds.



Appearances:

$$A_k = G_{k+1} \setminus G_k$$

Disappearances:

$$D_k = G_k \setminus G_{k+1}$$

Events:

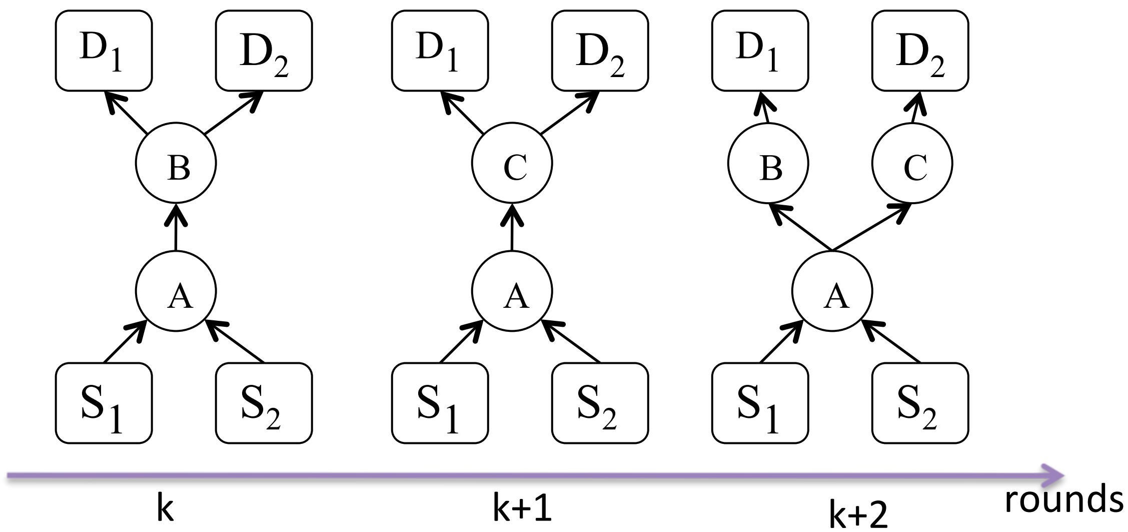
$$T_k = A_k \cup D_k$$

UPMC SORBONNE UNIVERSITÉS

Static states

	k	k+1	k+2
C	**0**	**1**	**1**
B	1	0	1
A	1	1	1

rounds



State variable:

– Presence: $\delta = 1$

– Absence: $\delta = 0$

Static state:

A series of consecutive presences or absences of a node or link

UPMC SORBONNE UNIVERSITÉS

Simulating lower probing frequencies

<table>
<tr><td>F_1</td><td>0</td><td>1</td><td>1</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr>
<tr><td>F_3</td><td colspan="3">1</td><td colspan="3">1</td><td colspan="3">0</td><td colspan="3">1</td></tr>
<tr><td>F_6</td><td colspan="6">1</td><td colspan="6">0</td></tr>
</table>

F_1: one round per hour “fine-grained” timescale

F_3: one simulated round every 3 hours

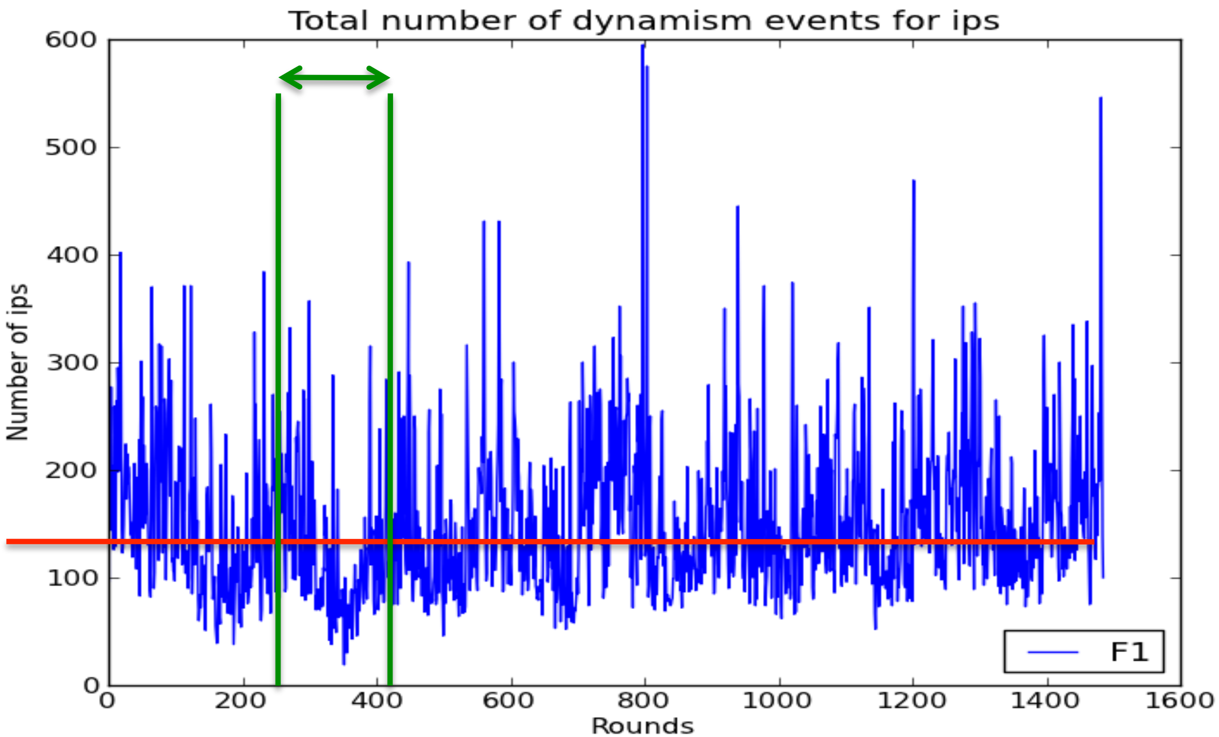
F_6: one simulated round every 6 hours

…

F_{48}: one simulated round every two days

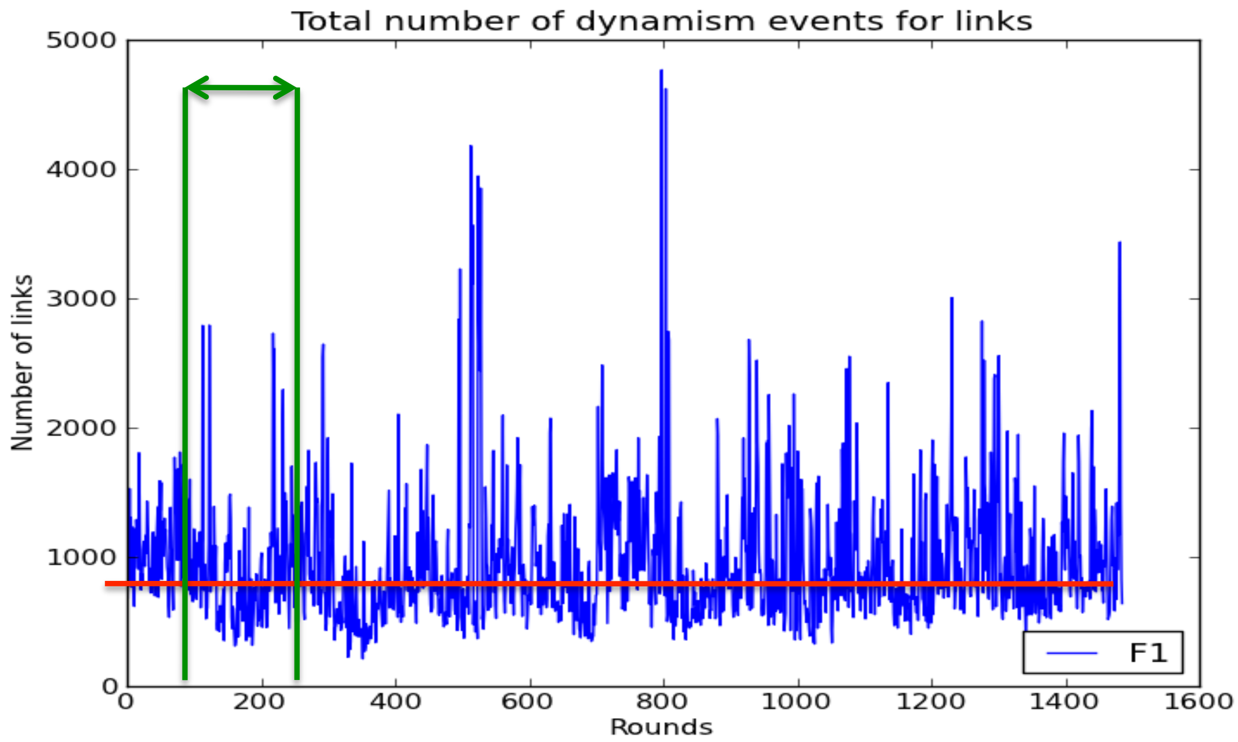
Simulation based upon choosing at random an observation from the prior timescale

Number of node events seen at F_1



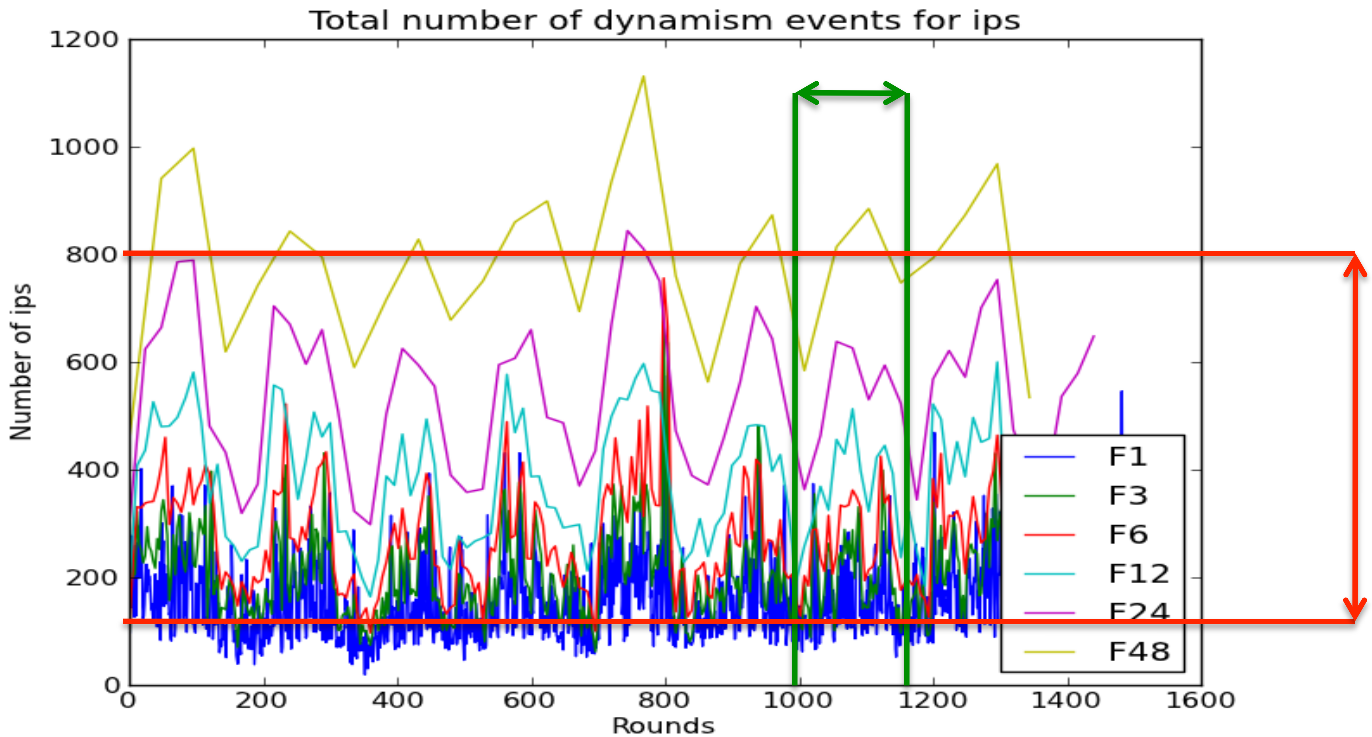
- 14,322 nodes seen in an average round
- 110 node events in an average round (0.8% of all IPs)
- periodic behavior: ~160 rounds (7 days)

Number of link events seen at F_1



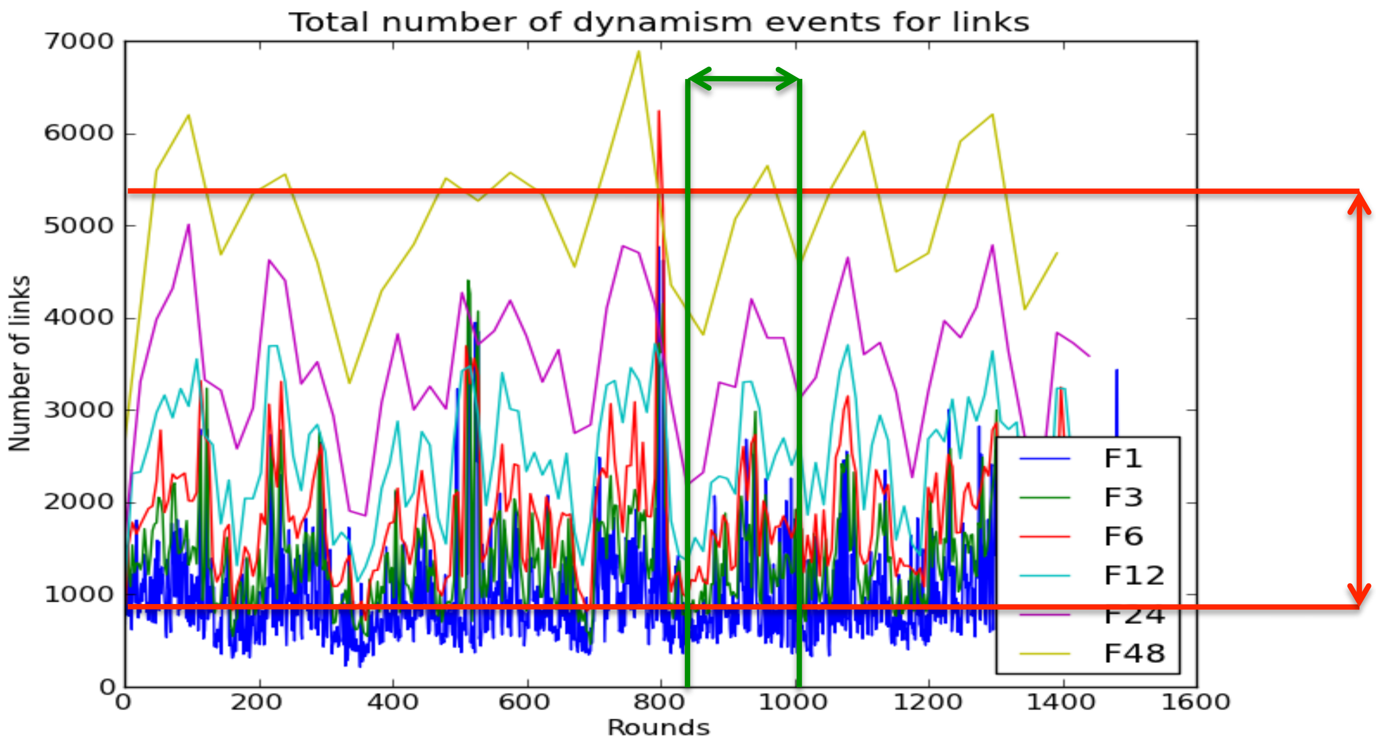
- 40,850 links seen in an average round
- 900 link events in an average round (2.2% of all links)
- Same periodic behavior as for nodes

Number of node events at different scales



- As expected, longer rounds mean more changes per round.
- From ~0.8% of nodes at F_1 to ~ 5.5 % at F_{48}
- Periodic behavior remains.

Number of link events at different scales



- Similar observation as for nodes: more changes per round.
- From ~ 2.2% at F_1 to ~ 12.2 % at F_{48}
- Periodic behavior remains.

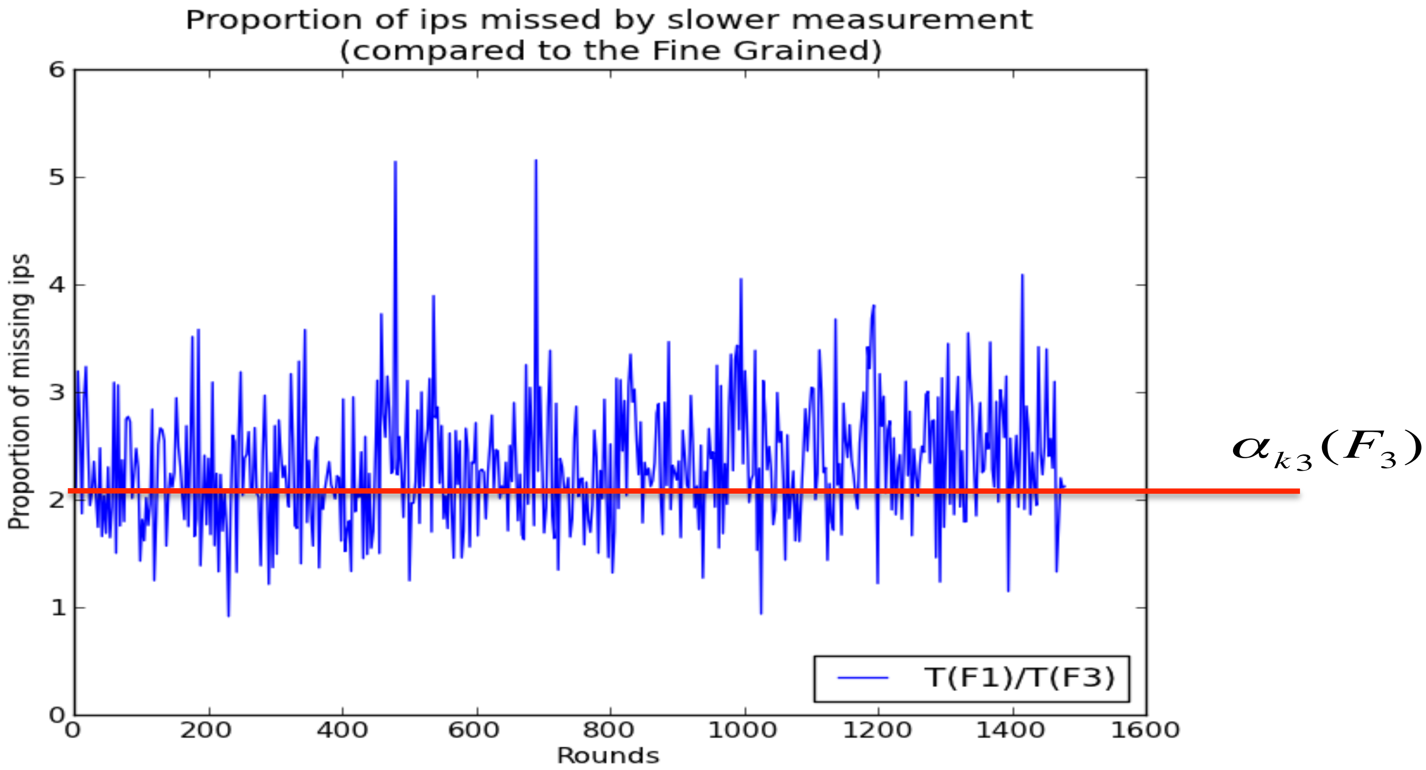
UPMC SORBONNE UNIVERSITÉS

Events missed

- Longer rounds capture more events *per round*, but what do they miss over comparable timescales?
- Proportion of events missed:

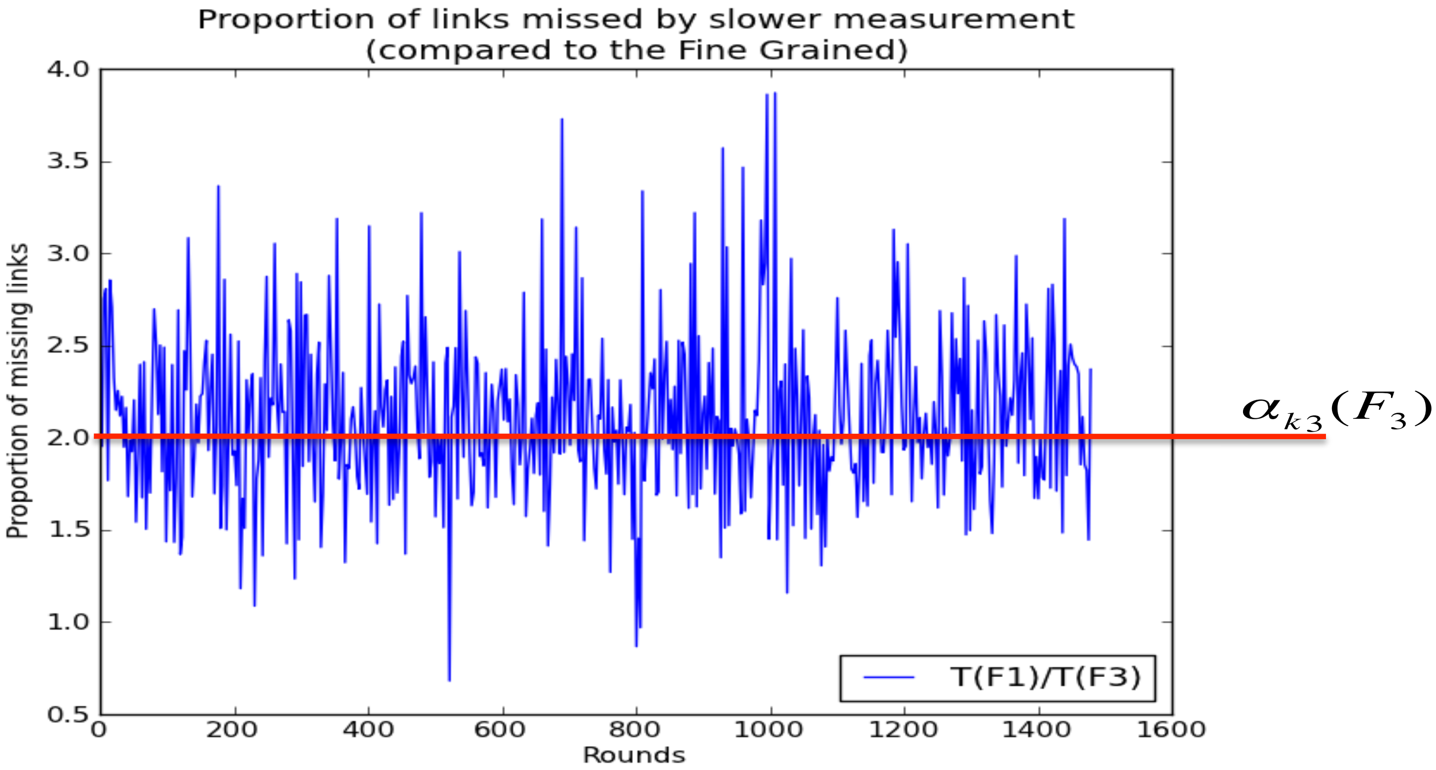
$$\alpha_{k_n}(F_n) = \frac{\left|T_{k_1}(F_1)\right|}{\left|T_{k_n}(F_n)\right|}$$

Proportion of node events missed at F_3



- Half of the node events observed at F_1 are missed when probing 3 times slower.

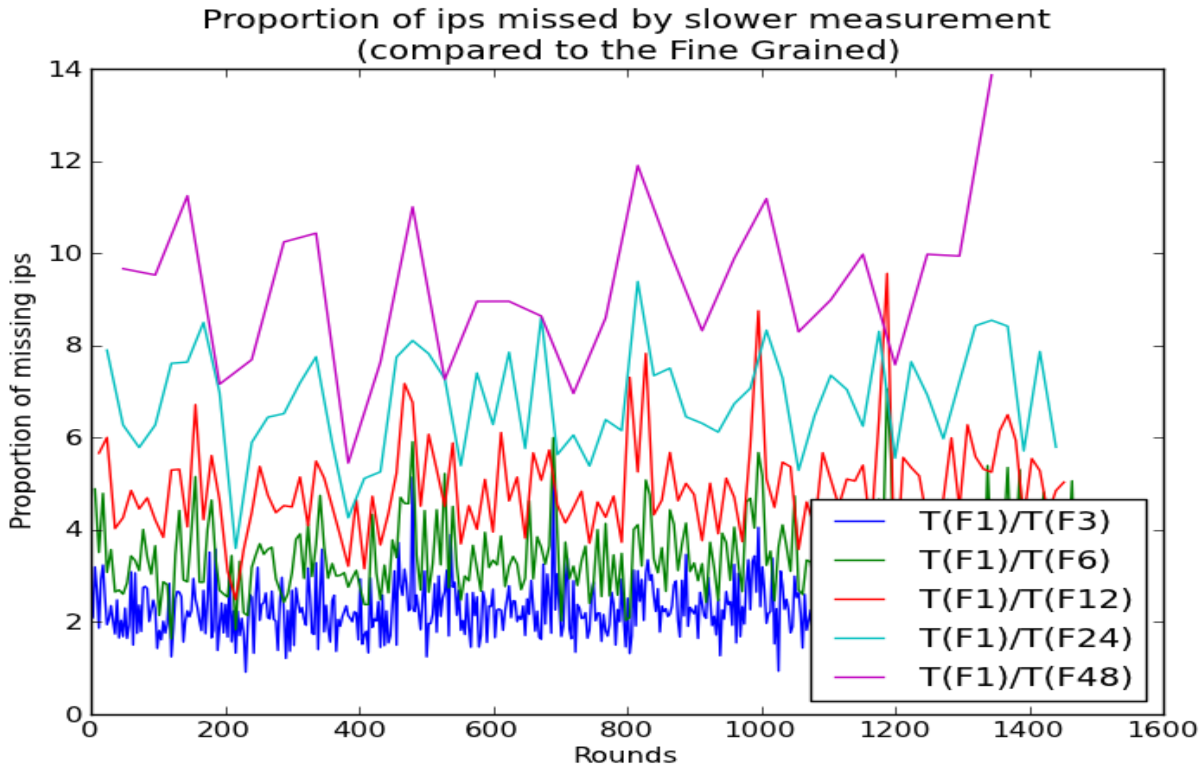
Proportion of link events missed at F_3



- Half of the link events observed at F_1 are missed when probing 3 times slower.

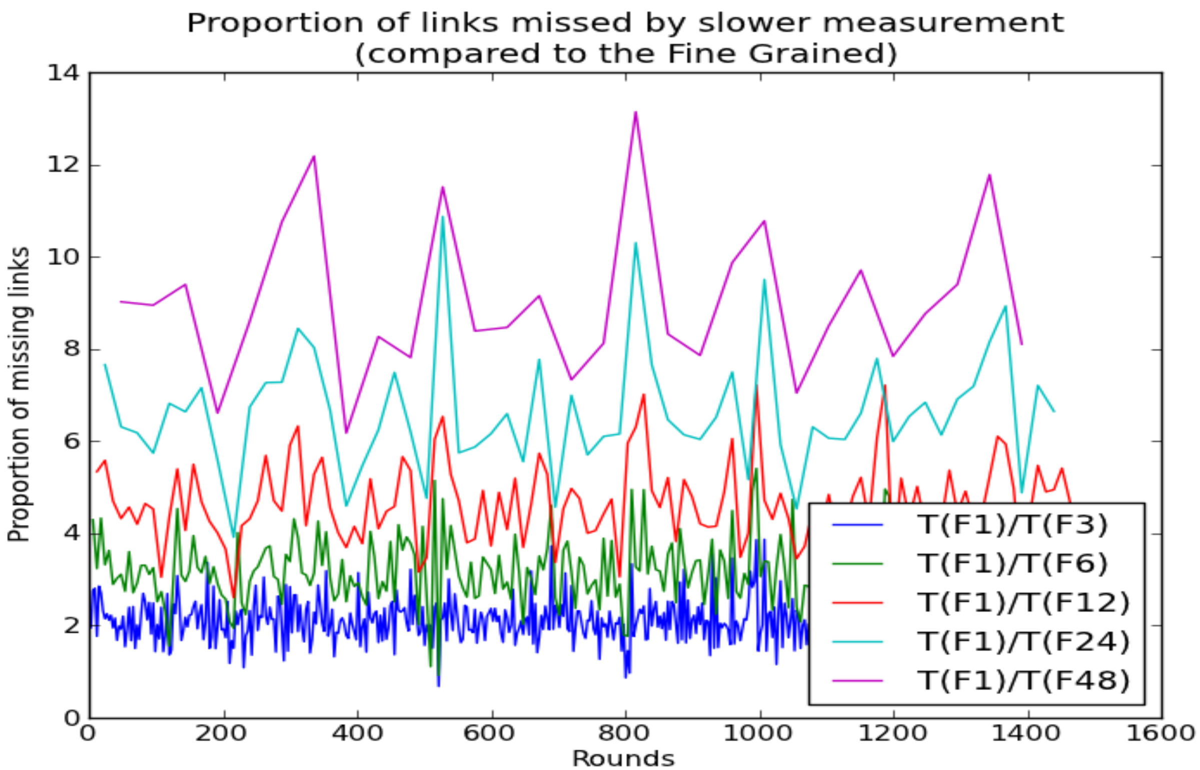
UPMC SORBONNE UNIVERSITÉS

Proportions missed at different scales: nodes



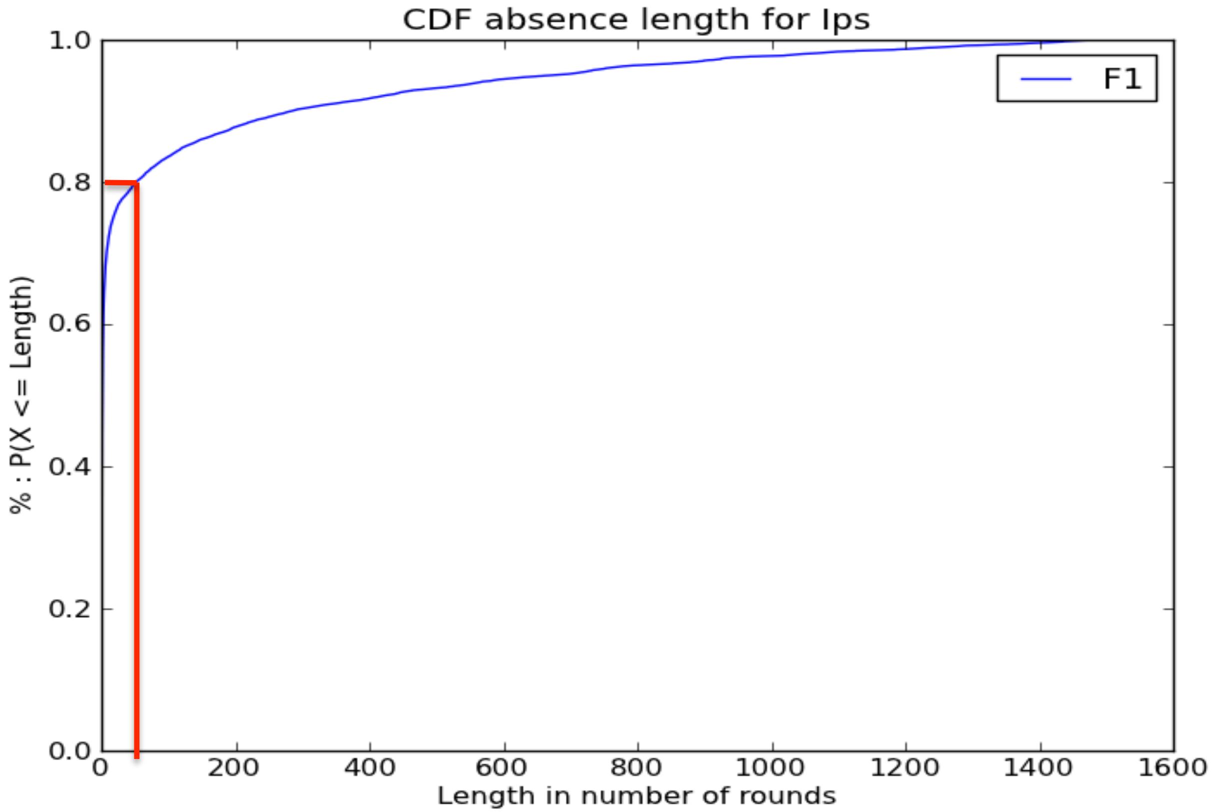
- As we probe more slowly, we miss more and more node events.
  - Probing every 2 days (as does Ark) may reveal 9 times fewer node events compared to probing every hour.

Proportions missed at different scales: links



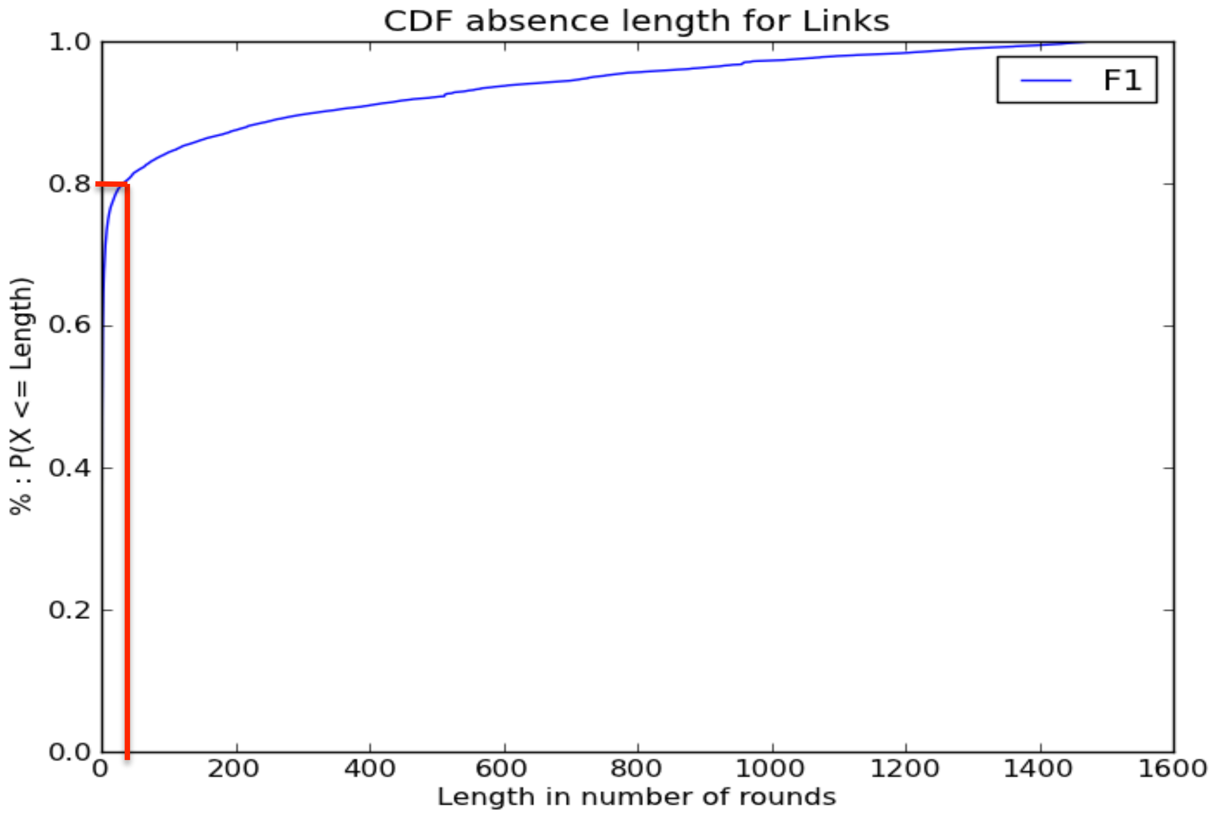
- As we probe more slowly, we miss more and more link events.
  - Probing every 2 days (as does Ark) may reveal 9 times fewer link events compared to probing every hour.

Static states for nodes at F_1



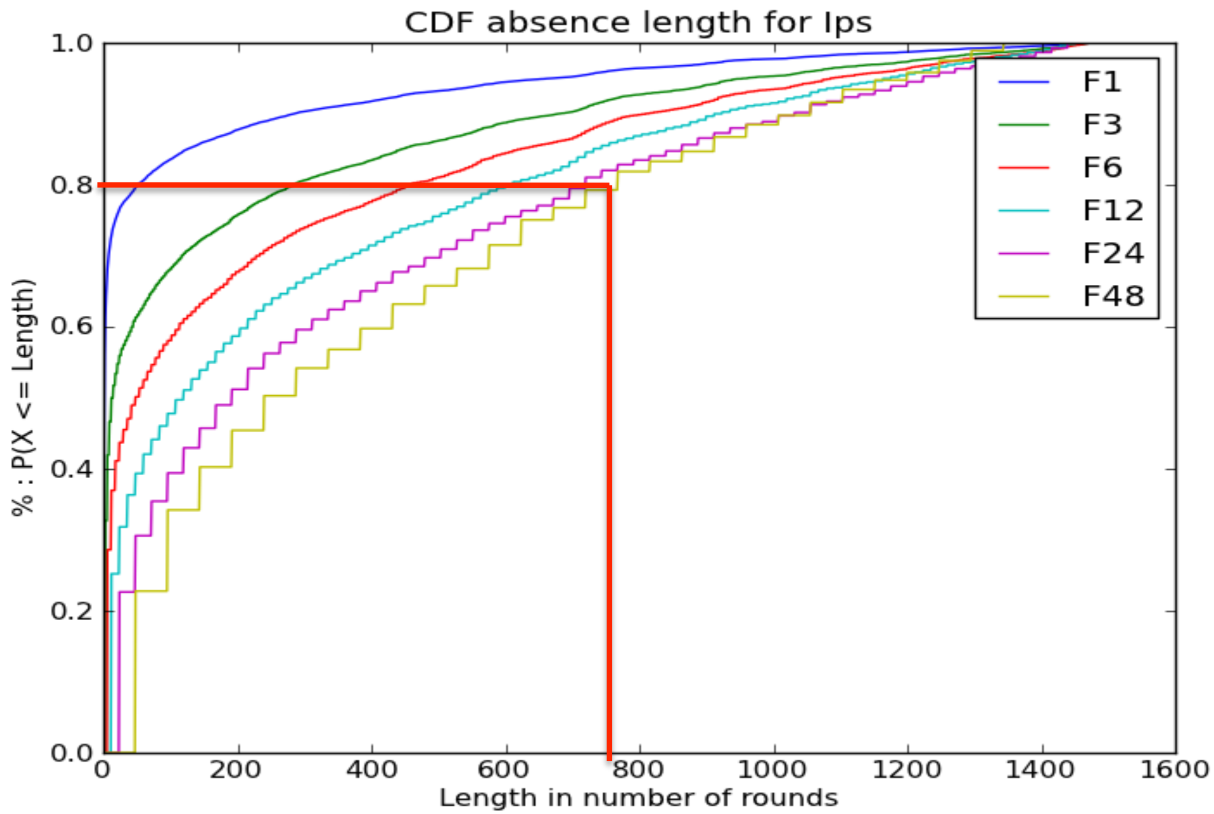
- When an IP is absent, it is typically (80% of the time) absent for a day (20 rounds) or less in our fine-grained measurement.

Static states for links at F_1



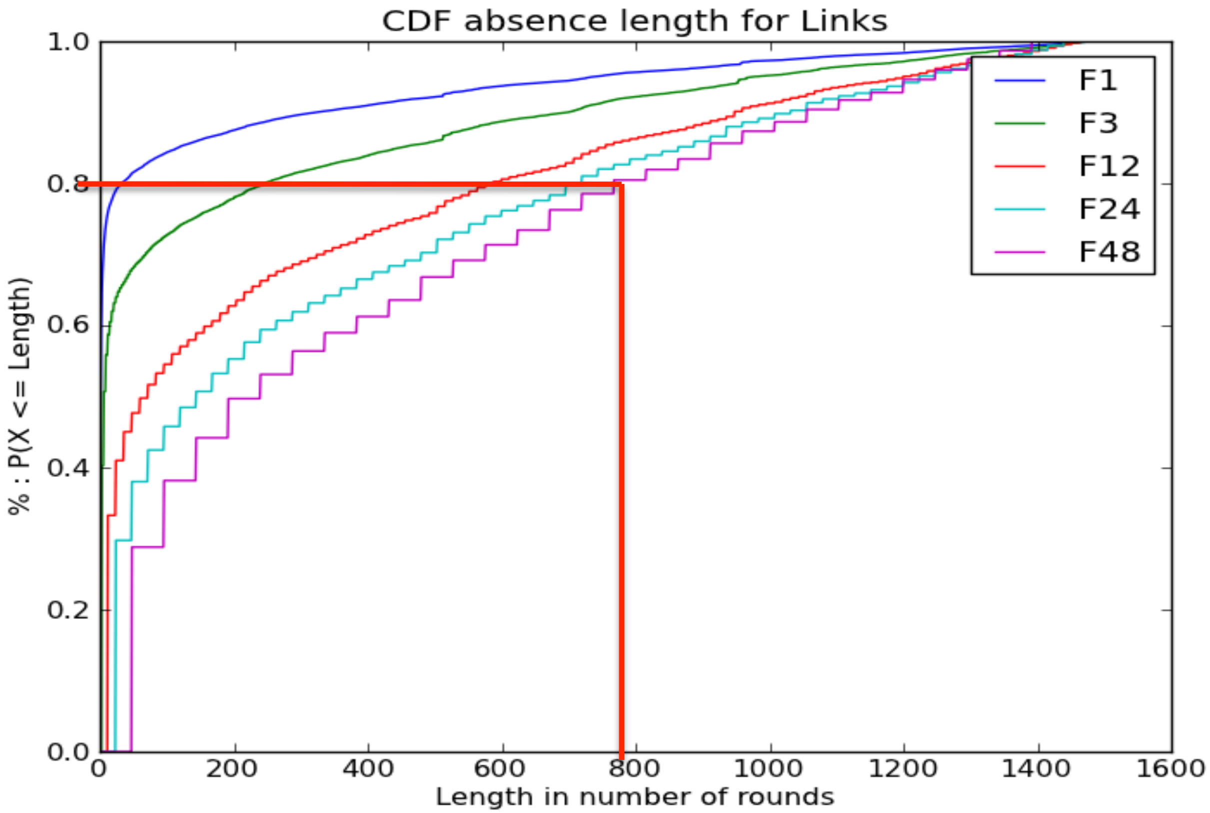
- When a link is absent, it is typically (80% of the time) absent for a day (20 rounds) or less in our fine-grained measurement.

Static states at different scales: nodes



- As we probe more slowly, typical absences for IPs get longer (80% are 30 days or less at F_{48})

Static states at different scales: links



- As we probe more slowly, typical absences for links get longer (80% are 30 days or less at F_{48}).

Summary of results

- Longer intervals between measurements mean more changes from one graph to the next.
  - But: by probing more slowly, we're missing a high proportion of events.

We must probe more frequently if we want to capture this detail

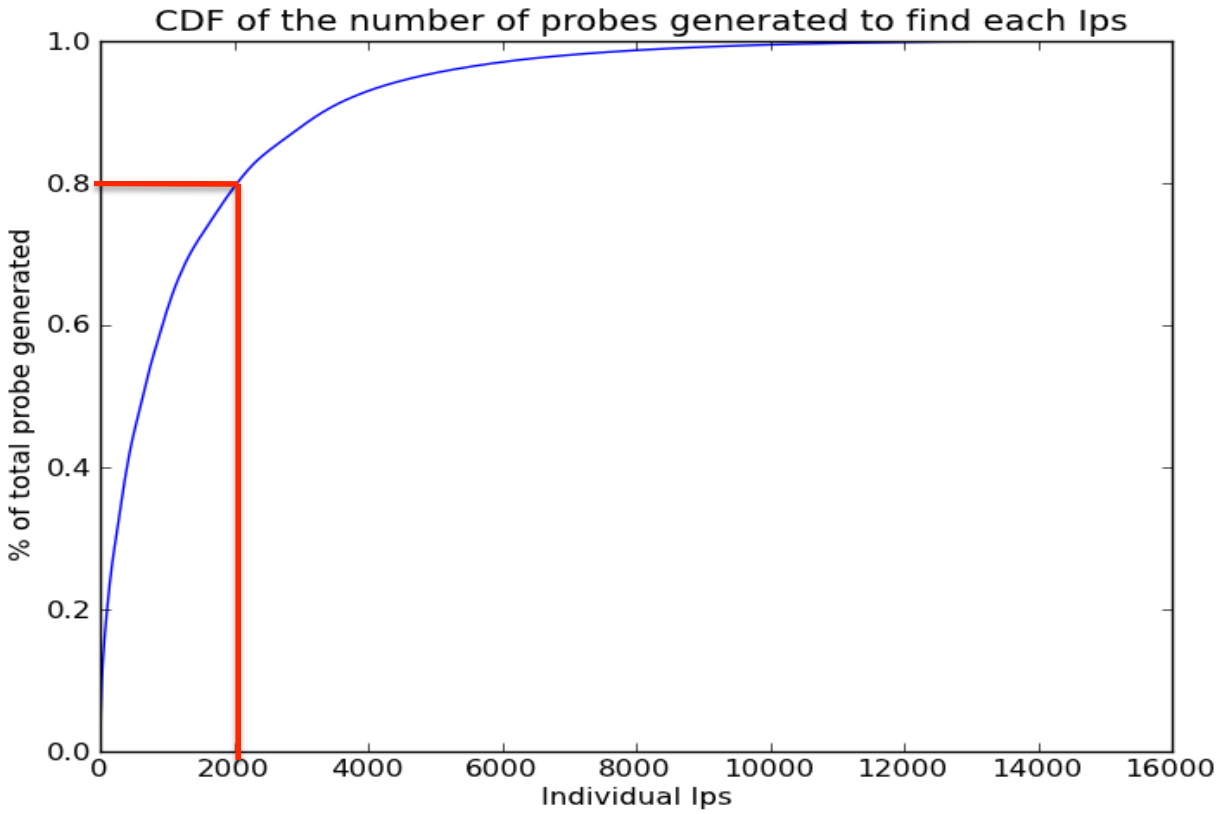
Probing more efficiently

- As we know from Donnet et al.'s Doubletree work, there is considerable probing redundancy
  - We examine redundancy in our data

    node redundancy = number of packets sent to discover a node

    link redundancy = ½ number of packets sent to discover a link
  - Similar results would mean that we have room for greater efficiency without data loss.

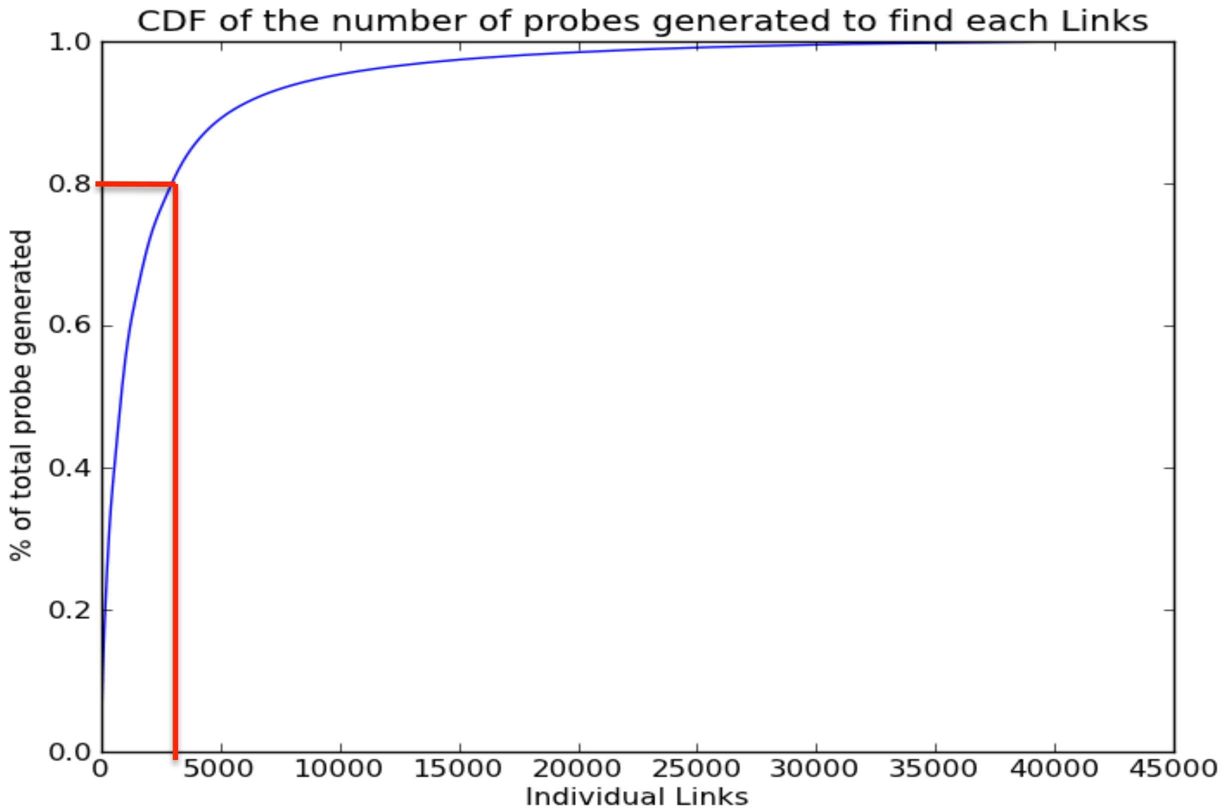
Redundancy of measurement probes: nodes



- IPs sorted in decreasing order of redundant discovery
- ~80% of measurement probes only discover ~14% of all IPs in each round

UPMC SORBONNE UNIVERSITÉS

Redundancy of measurement probes: links



- links sorted in decreasing order of redundant discovery
- ~80% of measurement probes only discover ~9% of all links

Next steps

- Develop more efficient probing algorithms:
  - Increase the frequency to catch short timescale events while lowering the load generated on the network.
  - Reduce measurement redundancy while capturing most events.

- Scale up the system:
  - Increase the size of the topology measured (millions of nodes) and increase the frequency (every few minutes).

This work has been supported by:

OneLab

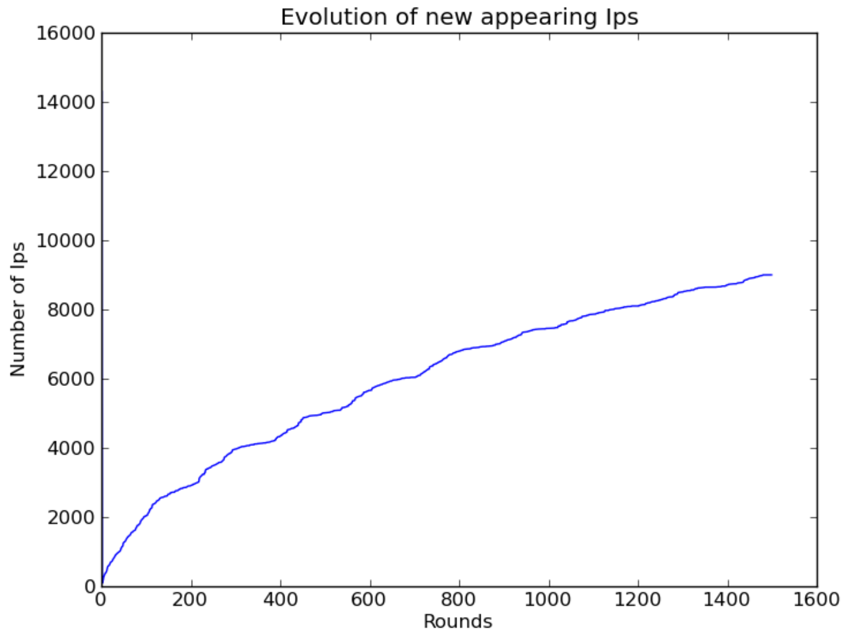
FUTURE INTERNET TESTBEDS

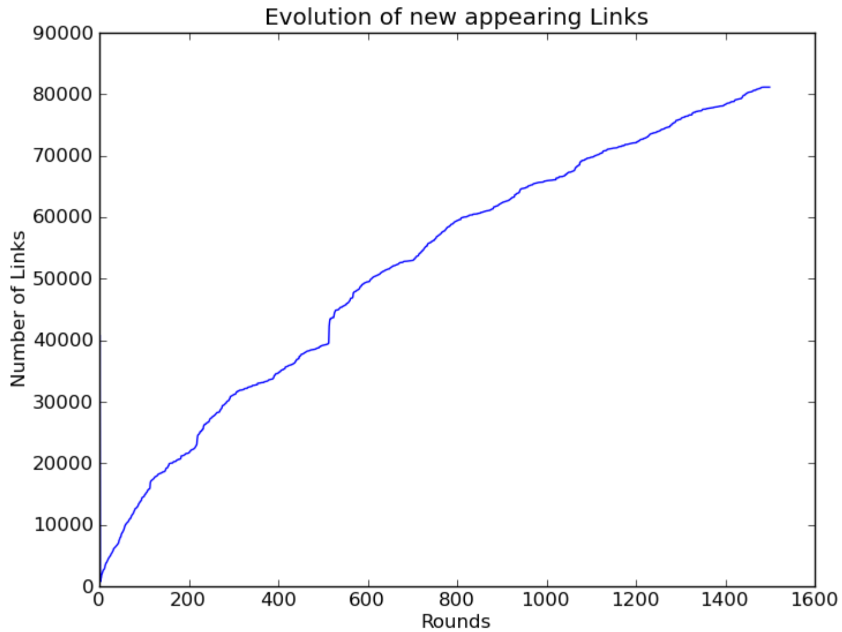
http://onelab.eu/

UPMC

SORBONNE UNIVERSITÉS

Dynamism events observation





- Over the whole experiment, we observe the birth of new IPs/links (confirming Latapy et al.'s observations).