Efficient solutions for the monitoring of the Internet

Chadi BARAKAT

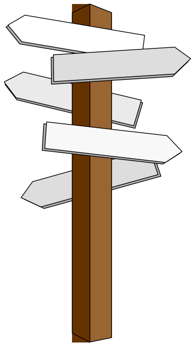
INRIA Sophia Antipolis, France
Planète research group

HDR defense – January 22, 2009

Email: Chadi.Barakat@sophia.inria.fr
WEB: http://www.inria.fr/planete/chadi

INRIA SOPHIA ANTIPOLIS

Outline



- Internet monitoring: Interests and Challenges
- Overview of the state of the art + selected contributions
  - Passive monitoring, sampling and traffic modeling
  - Active monitoring and network inference
- Zoom on some contributions
  - A framework for network wide sampling
  - TICP: A transport protocol for active probing and data collection
- Conclusions and future research

Internet monitoring: Interests

- From a network operator perspective
  - Real time monitoring of router and link load
  - Understanding the behavior of users, predicting SLAs
  - Routing optimization, provisioning
  - Detection and blocking of undesirable traffic
  - Topology and connectivity between other operators
- From a user perspective
  - Path characteristics for the optimization of applications
  - Network resource localization
  - Network troubleshooting

Network monitoring: Challenges

- Explosion of the Internet size
  - Around 1.5 billion users (23% of world population) (source internetworldstats)
  - Around 600 million hosts (source swivel)
  - Around 1 trillion web page (source google)
  - Around 30,000 advertised AS numbers (source potaroo)
- Hardware limitations
  - Fast links (~ 100 Mbps) vs. slow memory access (~ 10 ns)
- Completely decentralized architecture
  - No one knows how all this looks like and how it connects and behaves
  - We only know our neighbors and what do they tell us
  - Except routing information and the ICMP messages, operators and users don't exchange information on network performance

Network monitoring: Challenges

- Simplicity of the Internet service
  - Get an access and send whatever you want
  - To whomever you want
  - As much as your bandwidth allows

↗ The main reason behind the Internet success

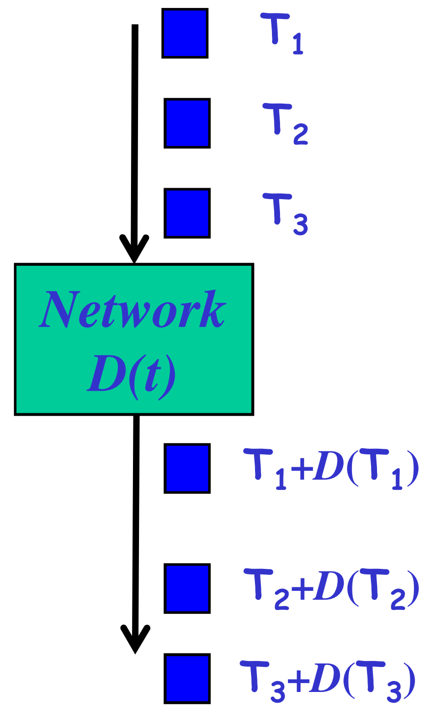
↘ Origin of many problems: attacks, congestion, traffic uncertainties

- Selfish policies adopted by operators
  - Very often announced routes are not the shortest ones
  - Some block control (ICMP) messages
- Security problems
  - Difficulties in placing measurement points
  - Difficulties in sharing measurement results

State-of-art: Two approaches (1)

- Passive measurements
  - Sniff traffic on one or multiple interfaces inside a network
  - Aggregate the traffic and send reports to a collector
  - Analyze the traffic and infer as much as possible
- Among the hot topics
  - Fast traffic collection and analysis
  - Traffic sampling
  - Bypass encryption and non standard port usage
  - Traffic modeling
  - Anomaly detection
  - Monitor placement

State-of-art: Two approaches (2)



- Active measurements
  - Send probe packets through the network
  - Packets got delayed differently
  - Infer what is in the box from the pattern of packets at the output
  - ICMP can be used to get feedback
- Among the hot topics
  - Path characterization (bandwidth, loss, delay, jitter)
  - Router and link characterization (a la traceroute)
  - Topology mapping and modeling
  - Network delay embedding (virtual coordinates)

Overview of main contributions: Three directions (1)

- Better modeling of the Internet using probabilities, stochastic analysis & machine learning
  - Poisson shot noise to model Internet traffic at the flow level
  - Packet size distribution and unsupervised machine learning for application identification
  - Kalman filter for tracking delay error in coordinate systems
  - Linear filters (wiener filter) for counting large populations (number of receivers, number of flows, number of entities)
- Real traces for validation

Overview of main contributions: Three directions (1)

- Better modeling of the Internet using probabilities, stochastic analysis & machine learning
  - Poisson shot noise to model Internet traffic at the flow level
  - Packet size distribution and unsupervised machine learning for application identification
  - Kalman filter for tracking delay error in coordinate systems
  - Linear filters (wiener filter) for counting large populations (number of receivers, number of flows, number of entities)
- Real traces for validation

INRIA SOPHIA ANTIPOLIS

Counting large populations

- Suppose a large size-varying population X_n to be tracked
  - Machines, flows, receivers, etc
- Exact counting not possible because of the overhead
- Probabilistic counting:
  - Members signal themselves with some low probability p
  - Count the received signals Y_n and infer the global
  - Simple inference: $X_n = Y_n / p$
  - Quadratic error proportional to $1 / p$ ☹
  - One can do better by accounting for the auto-correlation of X_n

INRIA SOPHIA ANTIPOLIS

Counting large populations

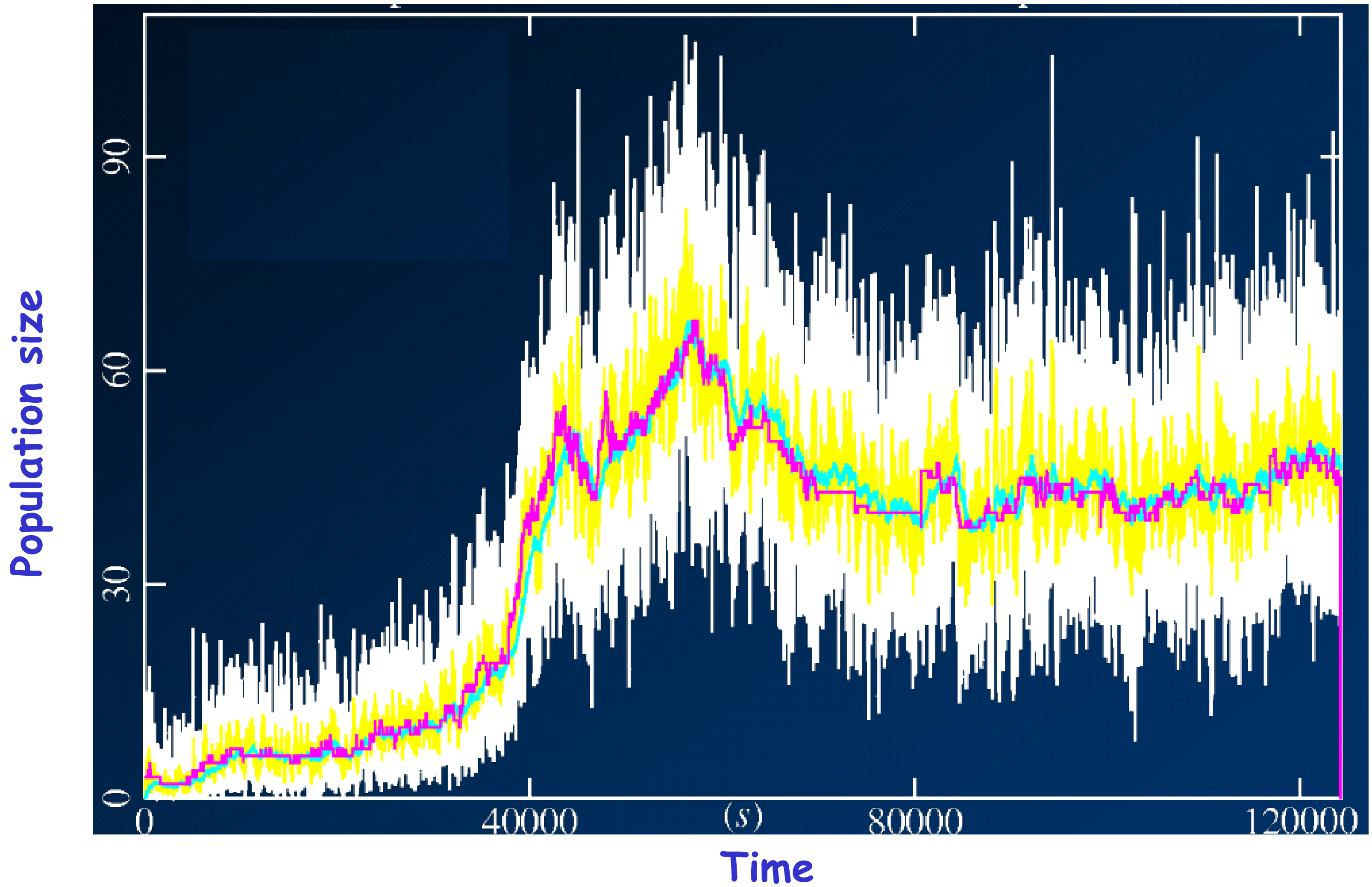
- Auto-correlation important when counting is done at faster than members' lifetime
- Actual measurement to add to previous estimation, e.g.,

$$X_n = A \,.\, X_{n-1} + B \,.\, Y_n / p$$

- How to set the weights ?
- Contribution:
  - Fit the problem in the context of Optimal Wiener filter
  - Optimal weights for Poisson arrivals
  - Optimal form over all linear filters for exponential lifetimes

Counting large populations

Number of receivers in a multicast session – p=0.01, S=1s

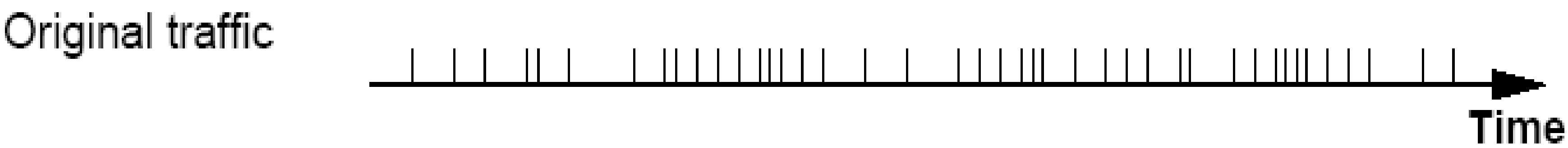


Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements

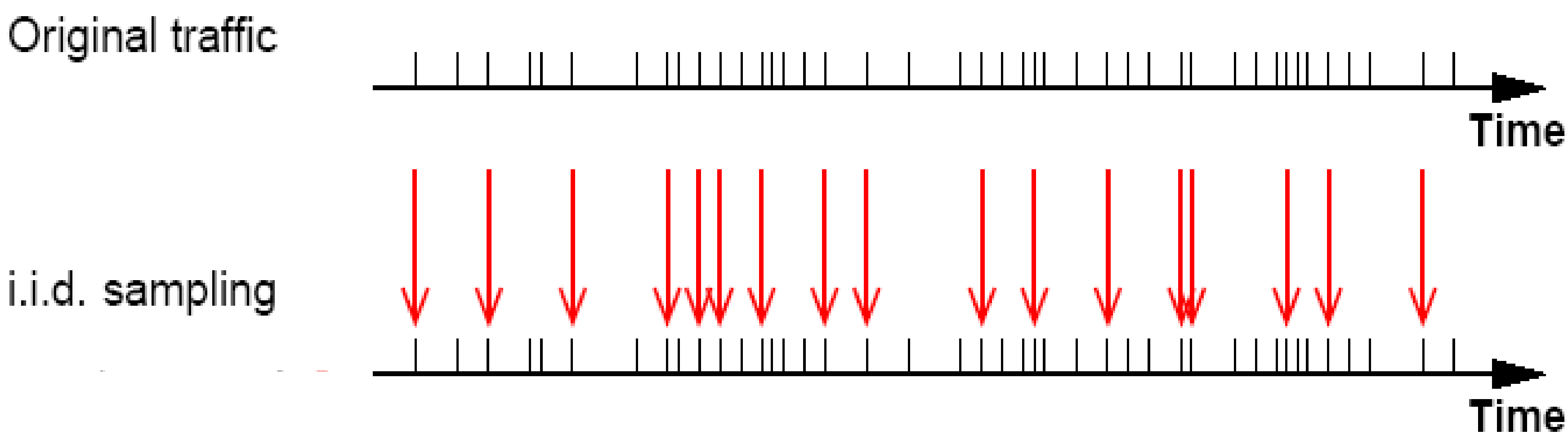
Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements



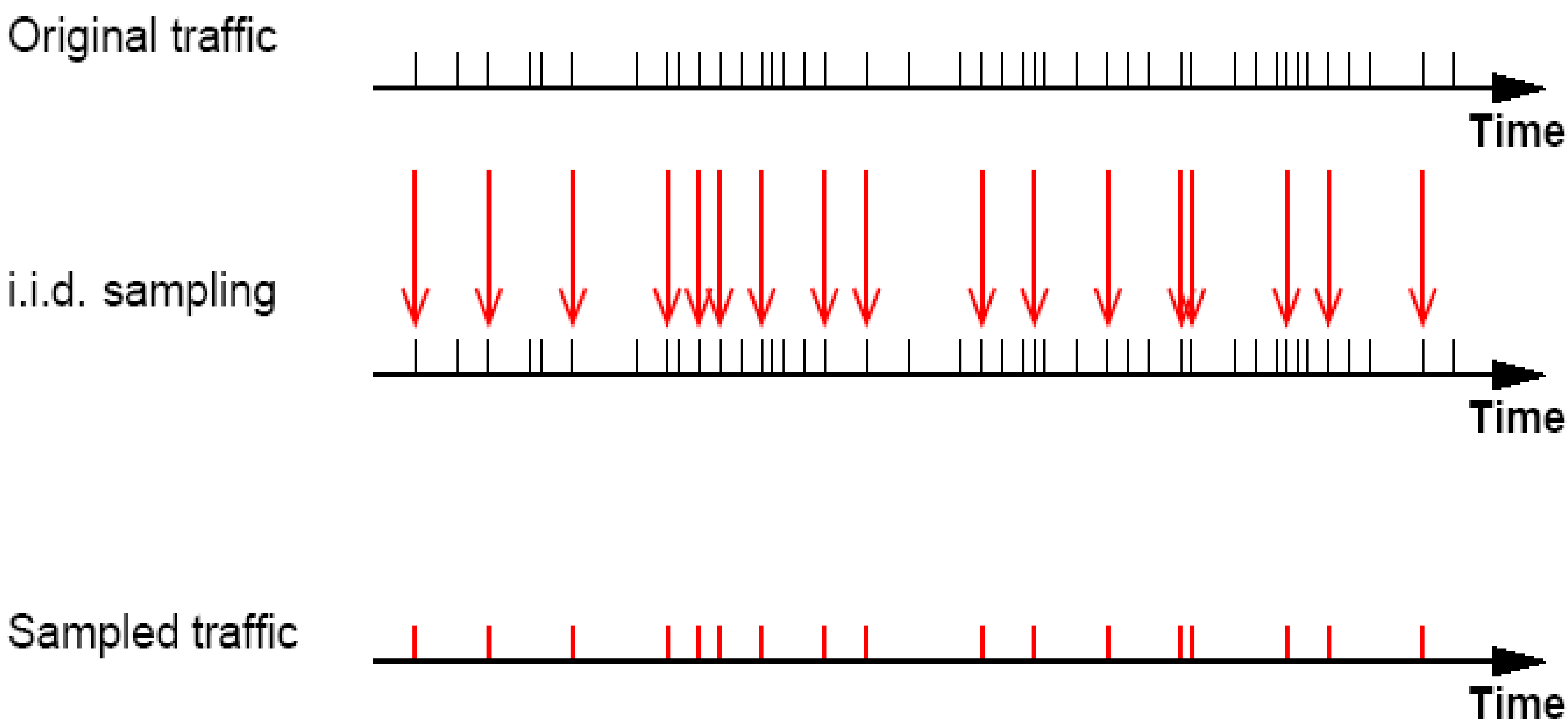
Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements



Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements



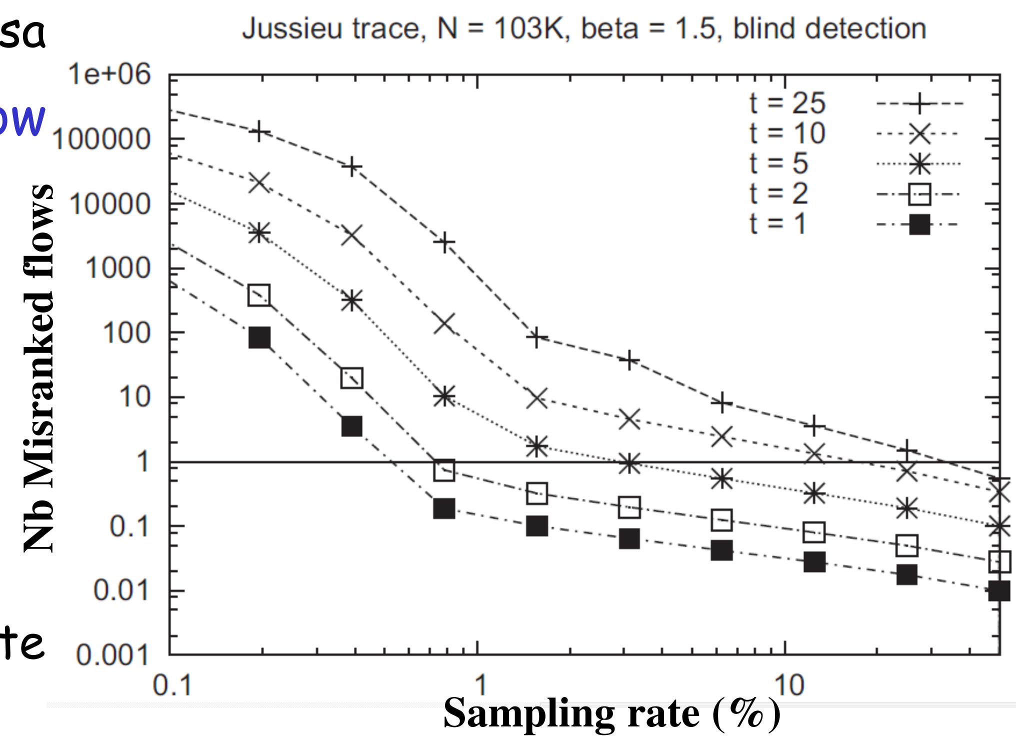
INRIA SOPHIA ANTIPOLIS

Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements
  - Serious impact on flow based statistics
    - Most flows are small and get disappear

Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements
  - Serious impact on flow based statistics
    - Most flows are small and get disa
  - We quantified the impact on flow size estimation and large flows detection and ranking
    - A sampling rate of order 10% for the detection of the few largest flows
    - Considered as a negative result given the practiced sampling rate



Overview of main contributions: Three directions (2)

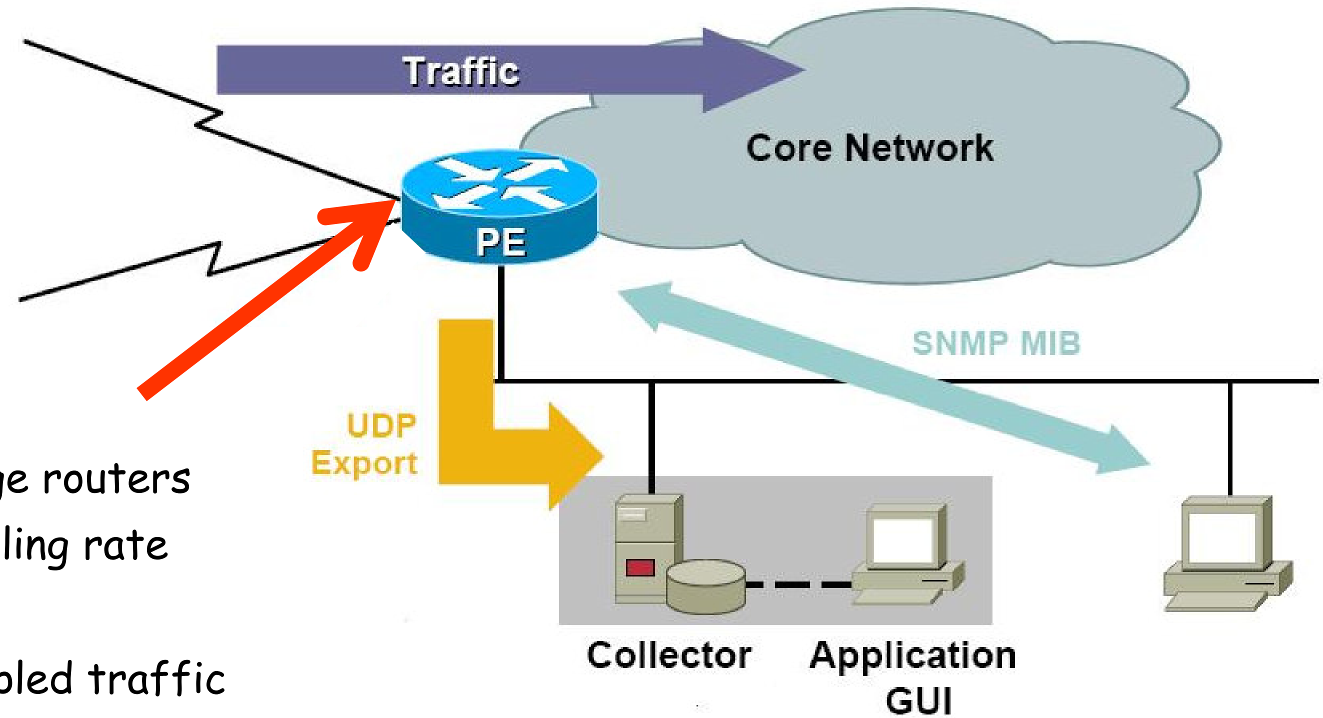
- Packet sampling as a solution to reduce the overhead of passive measurements
  - Serious impact on flow based statistics
    - Most flows are small and get disappear
  - We quantified the impact on flow size estimation and large flows detection and ranking
  - We introduced the notion of network-wide sampling
    - Allow sampling in all network routers (not only at the edge)
    - Run a global optimization problem to find the optimal sampling rate per router interface
    - Target: Maximize accuracy while minimizing overhead

Overview of main contributions: Three directions (2)

- Packet sampling as a solution to reduce the overhead of passive measurements
  - Serious impact on flow based statistics
    - Most flows are small and get disappear
  - We quantified the impact on flow size estimation and large flows detection and ranking
  - We introduced the notion of network-wide sampling
    - Allow sampling in all network routers (not only at the edge)
    - Run a global optimization problem to find the optimal sampling rate per router interface
    - Target: Maximize accuracy while minimizing overhead

INRIA SOPHIA ANTIPOLIS

Common configuration



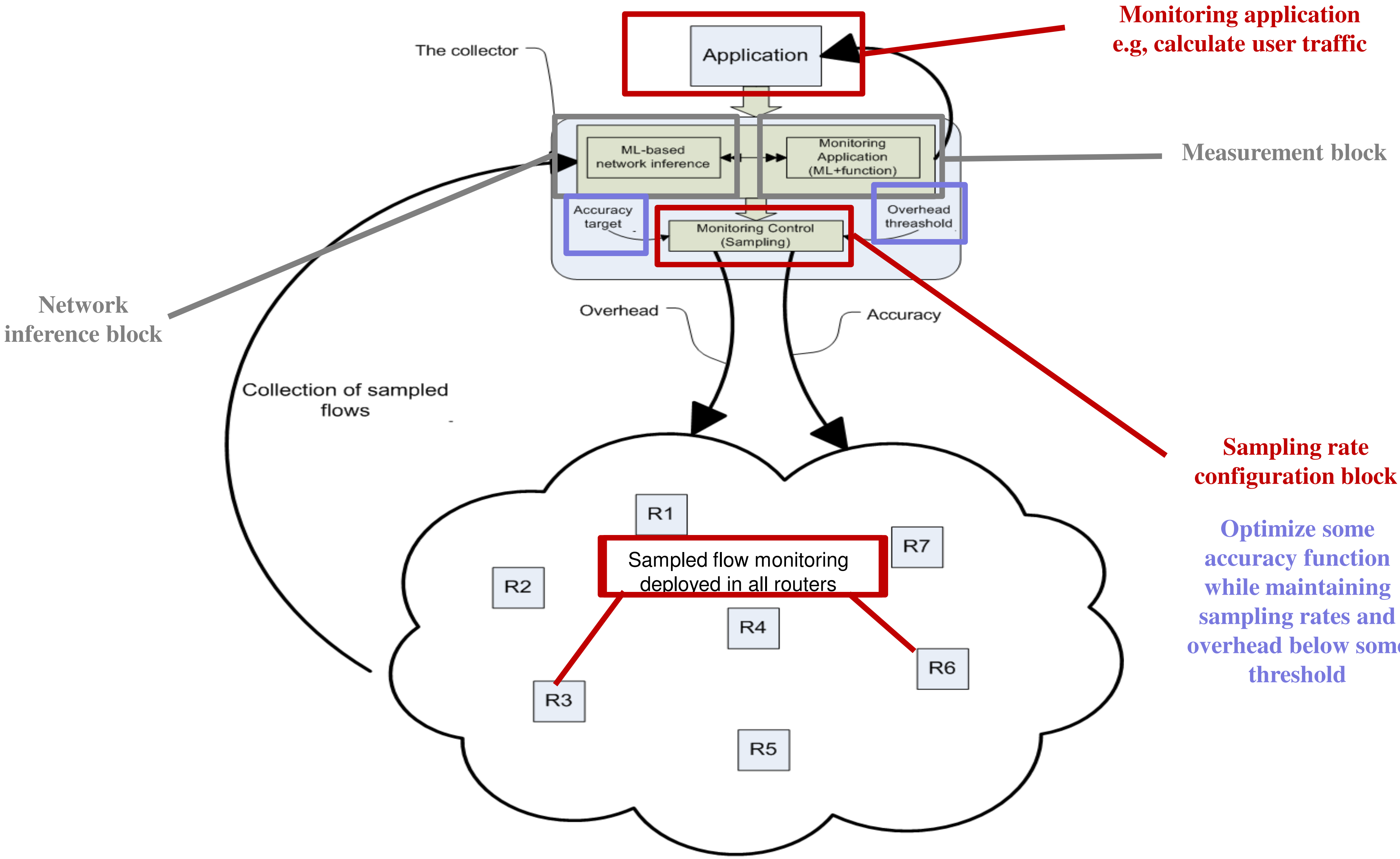
- Sample traffic at edge routers with some fixed sampling rate
  - Usually 1% or 0.1%
- Form flows from sampled traffic
- Invert flow sizes by a simple division by sampling rate
- Export
- A flow seen once. No optimization.

INRIA SOPHIA ANTIPOLIS

Common configuration: Discussion

- ❑ Simple, no duplicate flows
- ❑ Monitoring tasks have different requirements
  - Some don't require to sample all edge routers
    - Think about monitoring a point-to-point traffic
  - Other may require/tolerate different sampling rates
    - Lightly sample loaded routers
    - Heavily sample non loaded ones
- ❑ Limited choice: A flow is only seen once
  - More control by sampling and monitoring all network routers
  - In this case, the collector needs to merge flow measurements from different routers and invert

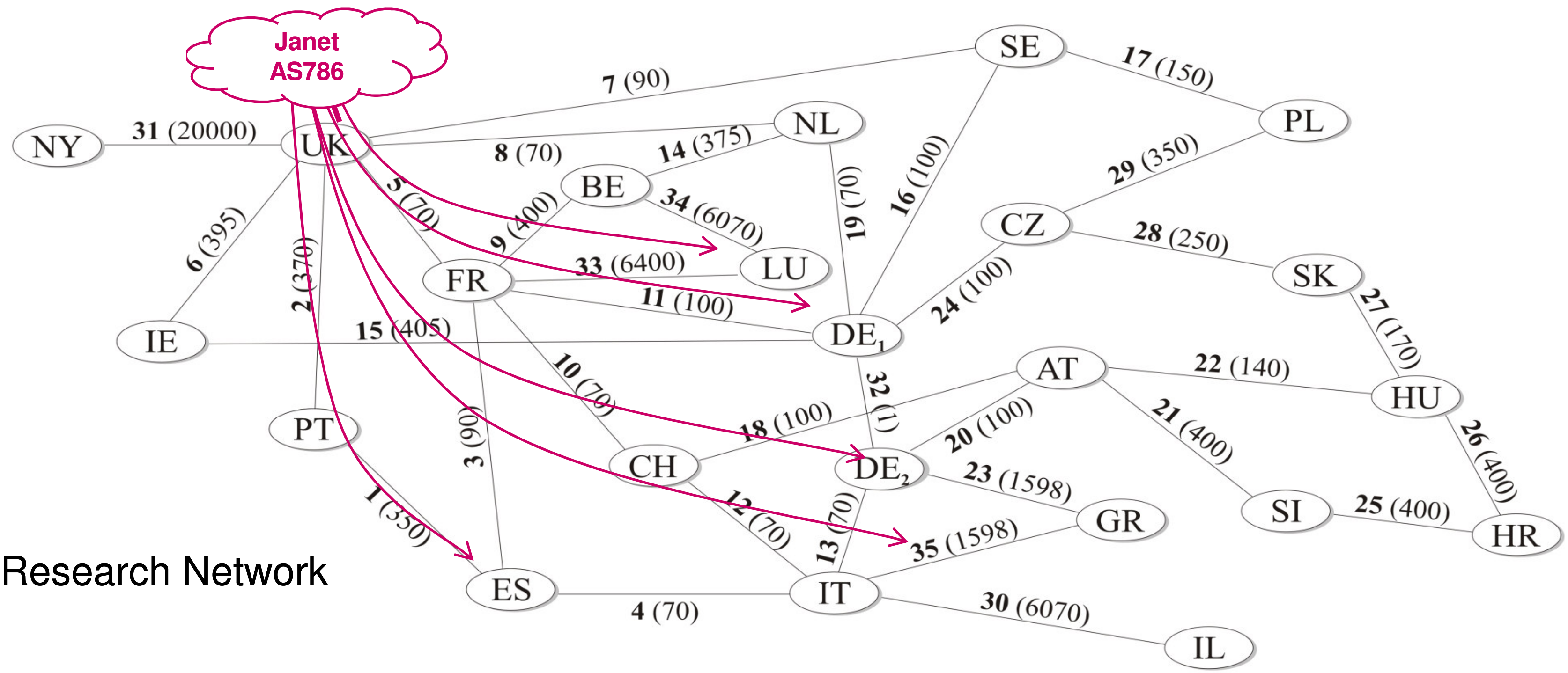
Our framework: network wide sampling



Optimize some accuracy function while maintaining sampling rates and overhead below some threshold

Case study: Traffic calculation

- Estimate amount of traffic flowing among a subset of origin-destination flows (common task for traffic engineering apps).
- Where and how to tune the sampling to estimate all UK sent traffic?



GEANT
European Research Network

An OD (Origin Destination) flow

INRIA SOPHIA ANTIPOLIS

Problem formulation

Choose vector of sampling rates *p* that maximizes

accuracy function for OD flow k

***"effective"* sampling rate for OD flow k**

$$\sum_{k \in \mathcal{F}} M\left(\rho_k\left(\boldsymbol{p}\right)\right),$$

sampling rate on link i

max sampling rate for link i

$$\begin{aligned} p_i &\geq 0 && \text{fo} \\ p_i &\leq \alpha_i && \text{for all } i \in \mathcal{L} \\ \sum_{i \in \mathcal{L}} p_i U_i &\leq \theta, \end{aligned}$$

packets traversing link i

system capacity constraint (in packets)

- Effective sampling rate approximated by sum of sampling rates
- All constraints are linear and define a convex solution space
  - Unique maximizer exists as long as **M(.)** is strictly concave
- Problem solved numerically
- Start from some default p vector and iterate until estimation converges

INRIA SOPHIA ANTIPOLIS

The accuracy function

- Measures the quality of sampling an OD flow
- Our example:
  - M = 1 - Mean Square Relative Error
  - MSRE = $E[((X/\rho - S) / S)^2]$

  where S is the actual estimation of the size of the OD flow

- Other functions could be possible to model other measurements tasks (left for future research):
  - accuracy of ranking/estimating the largest flows
  - accuracy of estimating the flow size distribution
  - accuracy of anomaly detection

Evaluation scenario

- Consider NetFlow data from GEANT
  - Collected using Juniper's Traffic Sampling
  - 1/1000 periodic sampling
  - We scale the measurement by 1000
- Get OD flow sizes and link loads every 5 minutes
- Solve the algorithm for the sampling rates that allow to estimate the sizes of the OD flows originated at UK
- Set θ to 100K packets. Don't limit the sampling rate.

Optimal sampling rates

For this example, one needs to sample 10 links at around 0.1% per link

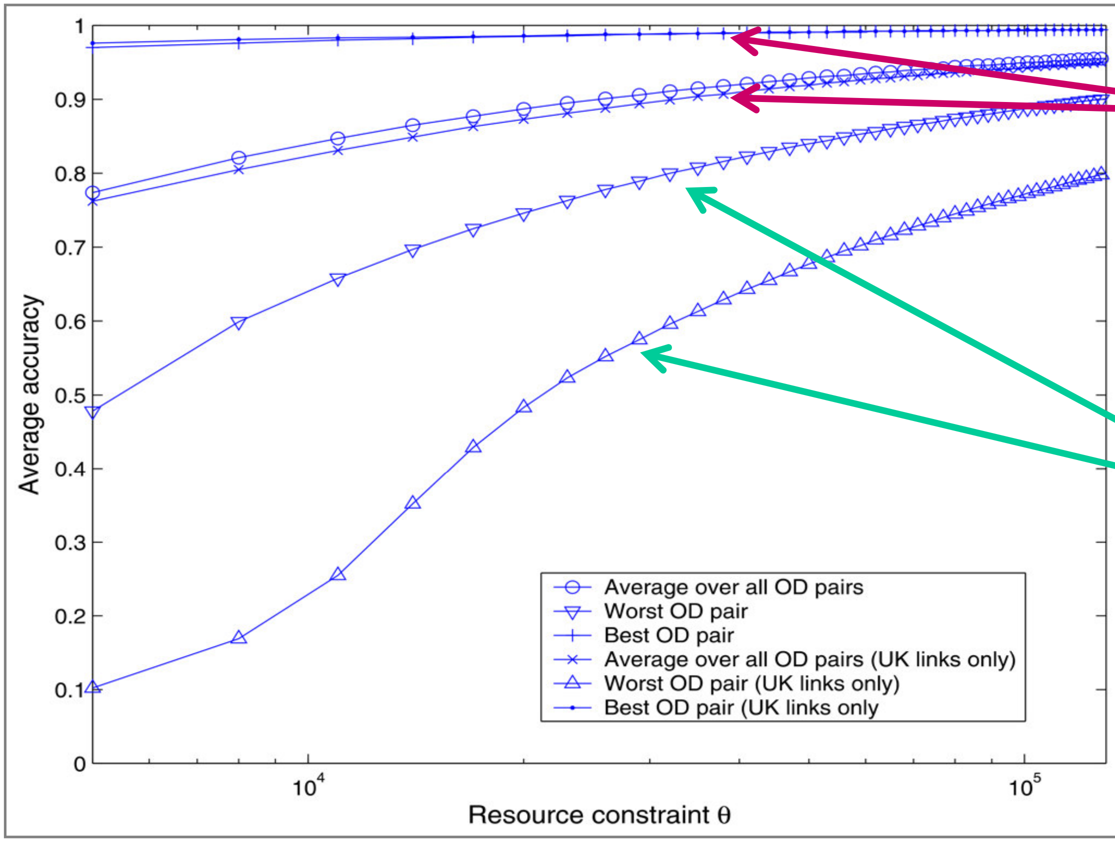
Sampled Link ID

OD flow

OD pair	pkt/s	p_5 UK-FR	p_7 UK-SE	p_8 UK-NL	p_9 UK-NY	p_{17} SE-PL	p_{30} UK-PT	p_{31} IT-IL	p_{33} FR-BE	p_2 FR-LU	p_{28} CZ-SK	Accuracy
JANET-NL	30123	-	-	0.0016	-	-	-	-	-	-	-	0.9999
JANET-NY	9387	-	-	-	0.0002	-	-	-	-	-	-	0.9982
JANET-DE	4300	-	-	0.0016	-	-	-	-	-	-	-	0.9995
JANET-SE	4080	-	0.0003	-	-	-	-	-	-	-	-	0.9973
JANET-CH	4033	0.0013	-	-	-	-	-	-	-	-	-	0.9994
JANET-FR	1723	0.0013	-	-	-	-	-	-	-	-	-	0.9985
JANET-PL	1400	-	0.0003	-	-	0.0003	-	-	-	-	-	0.9960
JANET-GR	1080	-	-	0.0016	-	-	-	-	-	-	-	0.9981
JANET-ES	1003	0.0013	-	-	-	-	-	-	-	-	-	0.9974
JANET-SI	913	-	-	0.0016	-	-	-	-	-	-	-	0.9977
JANET-IT	873	0.0013	-	-	-	-	-	-	-	-	-	0.9971
JANET-AT	790	0.0013	-	-	-	-	-	-	-	-	-	0.9968
JANET-CZ	590	-	-	0.0016	-	-	-	-	-	-	-	0.9965
JANET-BE	490	0.0013	-	-	-	-	-	-	0.0002	-	-	0.9955
JANET-PT	463	-	-	-	-	-	0.0011	-	-	-	-	0.9935
JANET-HU	377	-	-	0.0016	-	-	-	-	-	-	-	0.9945
JANET-HR	237	-	-	0.0016	-	-	-	-	-	-	-	0.9912
JANET-IL	87	0.0013	-	-	-	-	-	0.0018	-	-	-	0.9877
JANET-SK	43	-	-	0.0016	-	-	-	-	-	-	0.0092	0.9929
JANET-LU	20	0.0013	-	-	-	-	-	-	-	0.0090	-	0.9840
Link Loads (pkt/s)		63603	51833	57756	37286	23680	19950	15213	11173	6133	2600	
Contribution to θ		24.5%	5.1%	26.9%	2.1%	2.1%	6.8%	8.3%	0.7%	16.5%	7.1%	

INRIA SOPHIA ANTIPOLIS

Comparing to common configuration



Almost same performance over all OD flows

Small OD flows are better captured by our method

- Why does our method work better?
  - It looks across the entire network to find where small OD flows manifest themselves without hiding behind large flows

INRIA SOPHIA ANTIPOLIS

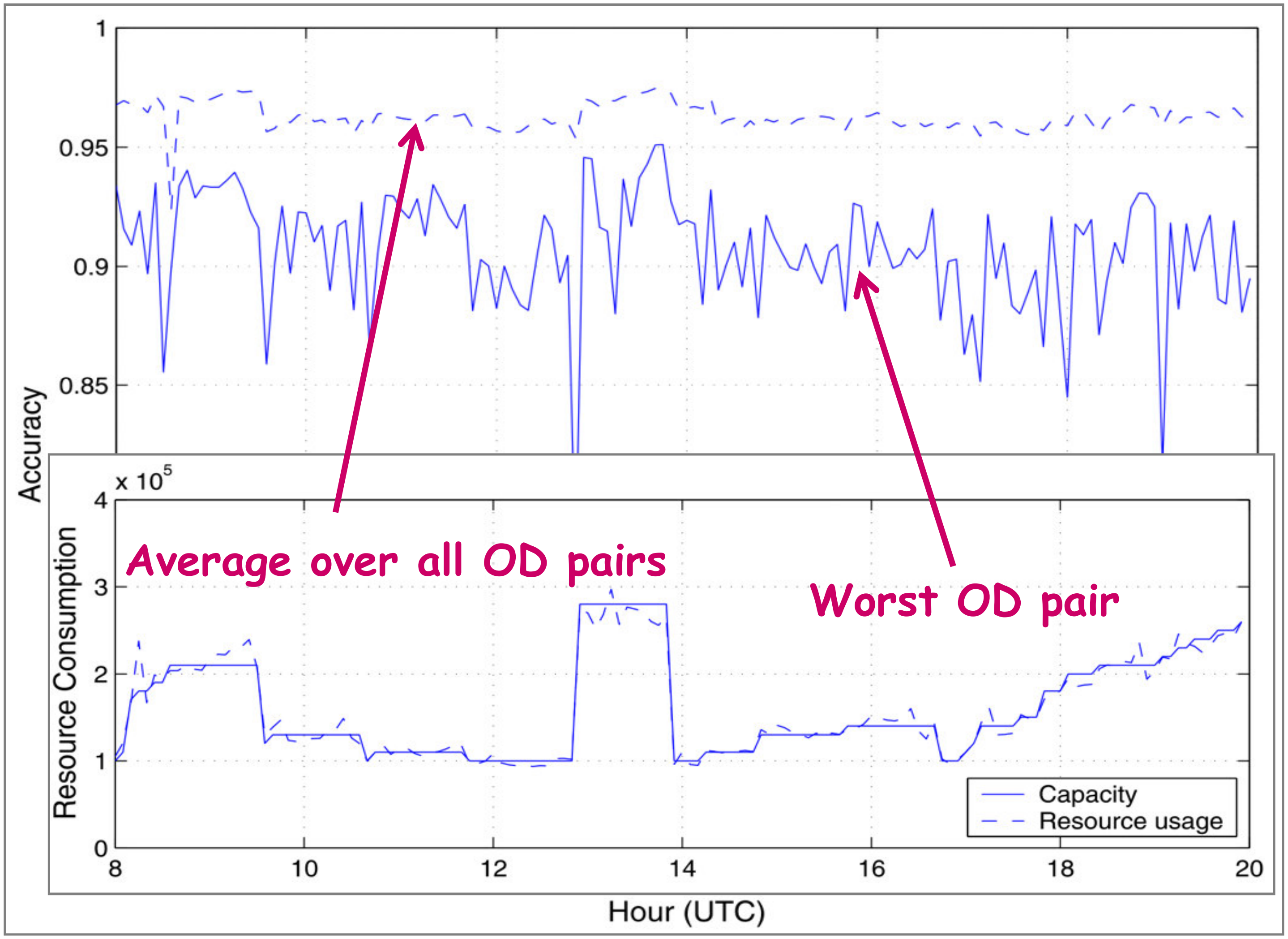
Dynamic version of our algorithm

- Compute new sampling rates when
  - estimated accuracy drops below target
  - collected traffic exceeds capacity
- If the estimated accuracy is still below target, increase capacity constraint by some factor say 10%
- Decrease capacity constraint if estimated accuracy is above target for more than some time (say one hour)

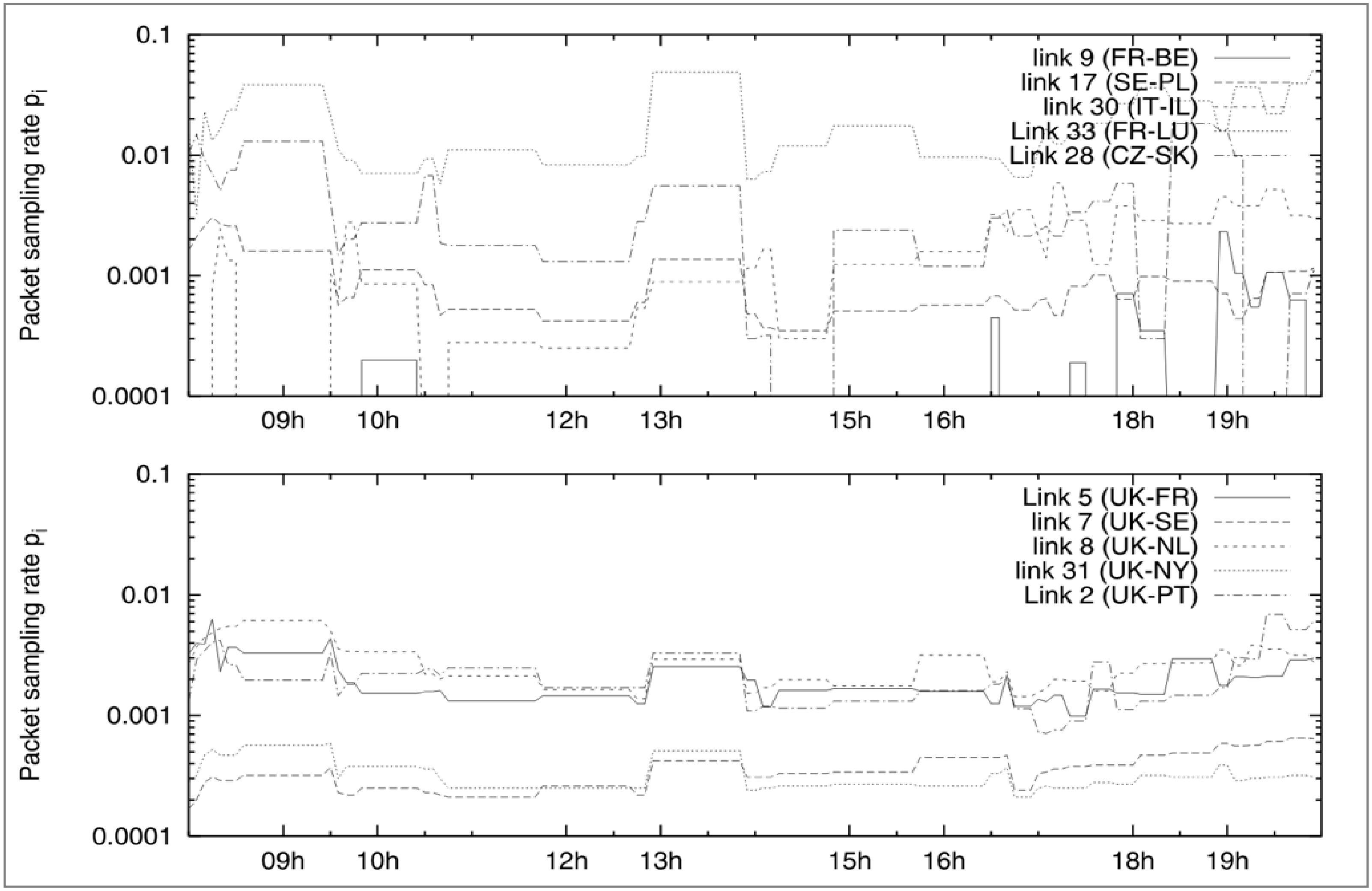
Implementation of dynamic version

Target accuracy 85%

Resource consumption



Implementation of dynamic version



INRIA SOPHIA ANTIPOLIS

Overview of main contributions: Three directions (3)

- Network monitoring by active probing
  - Embedding network delays and securing coordinate calculation
  - Correlation and compressibility of network path characteristics
    - Over the same path, among different paths
  - Congestion and error control for active probing
    - TICP: TCP-friendly Information Collection Protocol
    - Initially designed for data collection in large networks
    - Regulate the rate of probes and ensures reliability
    - A component absent in existing measurement infrastructures (Periodic probing, Poisson probing, round-robin, etc)

Overview of main contributions: Three directions (3)

- Network monitoring by active probing
  - Embedding network delays and securing coordinate calculation
  - Correlation and compressibility of network path characteristics
    - Over the same path, among different paths
  - Congestion and error control for active probes
    - TICP: TCP-friendly Information Collection Protocol
    - Initially designed for data collection in large networks
    - Regulate the rate of probes and ensures reliability
    - A component absent in existing measurement infrastructures (Periodic probing, Poisson probing, round-robin, etc)

Congestion and error control

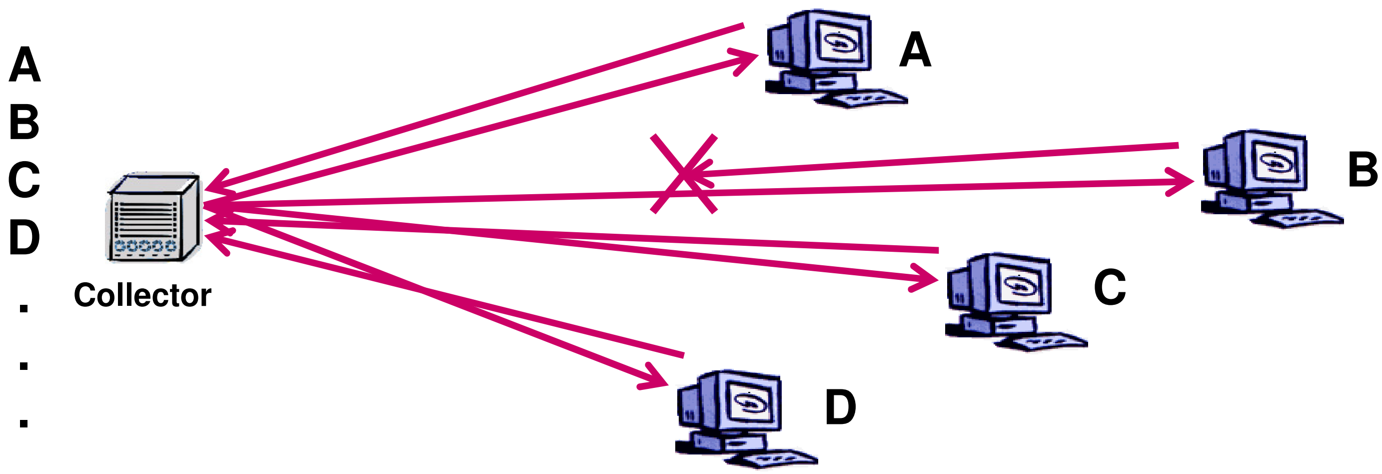
Challenges

- End-to-end, scalability and reliability
  - Retransmit probes when lost
  - No help from inside the network
- Congestion control in the forward and the reverse directions
  - High throughput and low loss rate
- Probes and reported information of different sizes, but generally small (many TCP connections would not work)
  - IP address of a router
  - Availability and statistics per a machine
  - Experienced Quality of Service
- Can be seen as many-to-one TCP session

Congestion and error control

Requirements

- ❑ Probed entities known by collector
  - E.g. PlanetLab IP adresses, List of machines to traceroute, etc.
- ❑ Probes directly answered
  - *Any delay is interpreted as network delay*



- ❑ How to regulate the rate of these probes ?

Protocol in brief: Congestion control

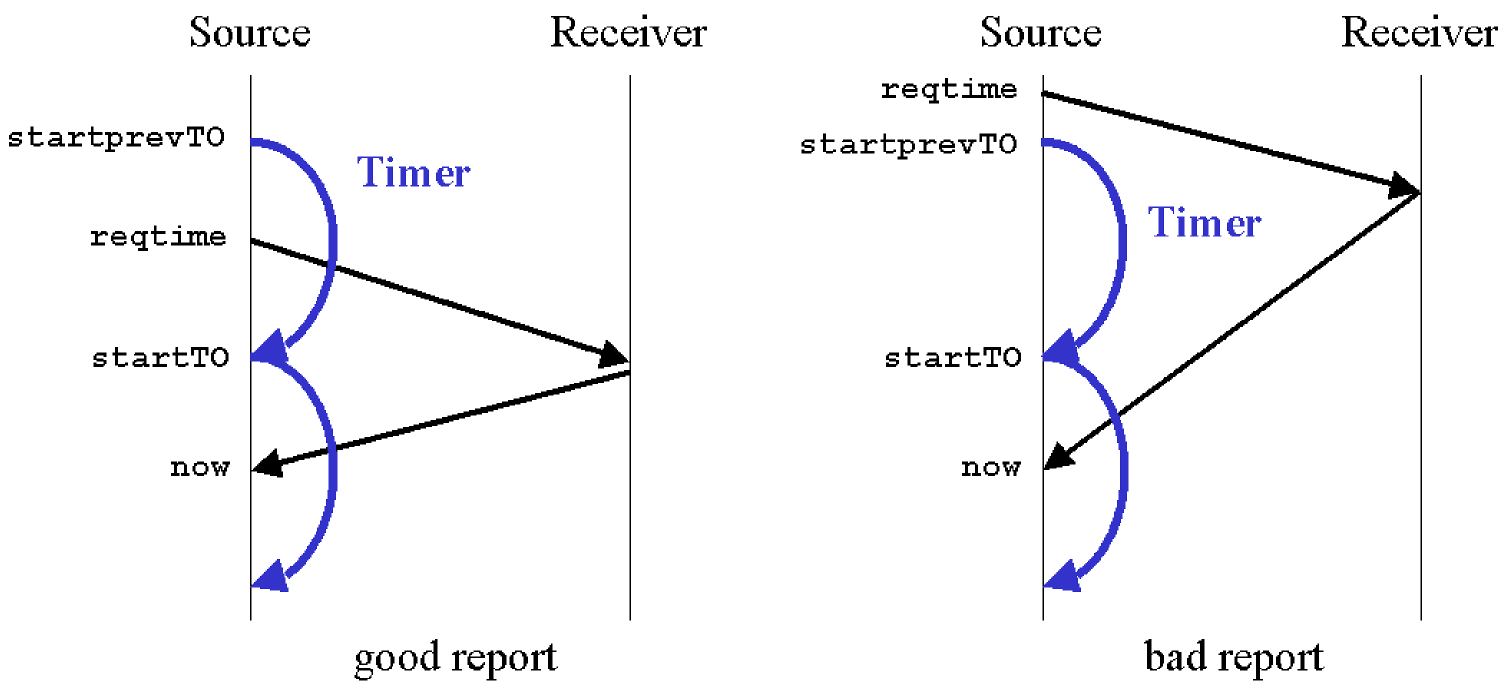
- A window-based flow control
  - **cwnd:** maximum number of machines the collector can probe before receiving any information
- The collector increases **cwnd** and monitors at the same time the loss rate of probes (during a time window in the past)
  - The protocol has two modes: slow start and congestion avoidance
- Congestion of the network is inferred when the loss rate of probes exceeds some threshold
- Upon congestion, divide **cwnd** by 2, and restart its increase

Protocol in brief: Error control

- The protocol is reliable
  - It ensures that all probes came back
- To reduce the duration of the session
  - In the first round, the protocol probes all machines
    - Order to be defined later
  - In the second round, the protocol probes machines whose reports were lost in the first round
  - In the third round, the protocol probes machines whose reports were lost in the first two rounds
  - Continues in rounds until all reports are received

Measuring the loss ratio

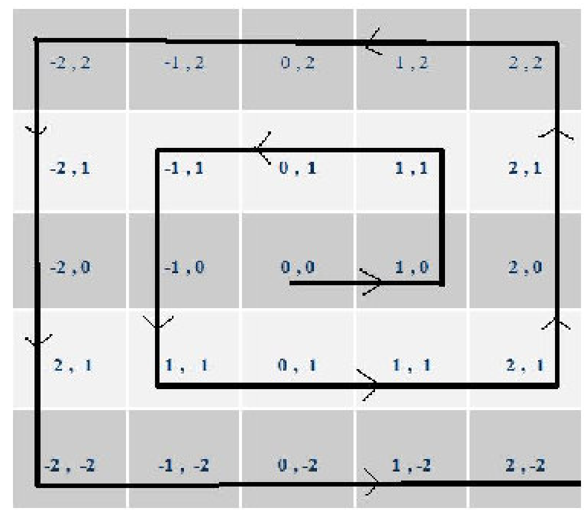
- The collector disposes of a timer, denoted TO, over which the loss rate is measured
  - Probes sent during one cycle of the timer have to arrive the next cycle, otherwise they are supposed lost
- Time
  - Se
    - –
    - –
  - R
  - Th
  - Th



)

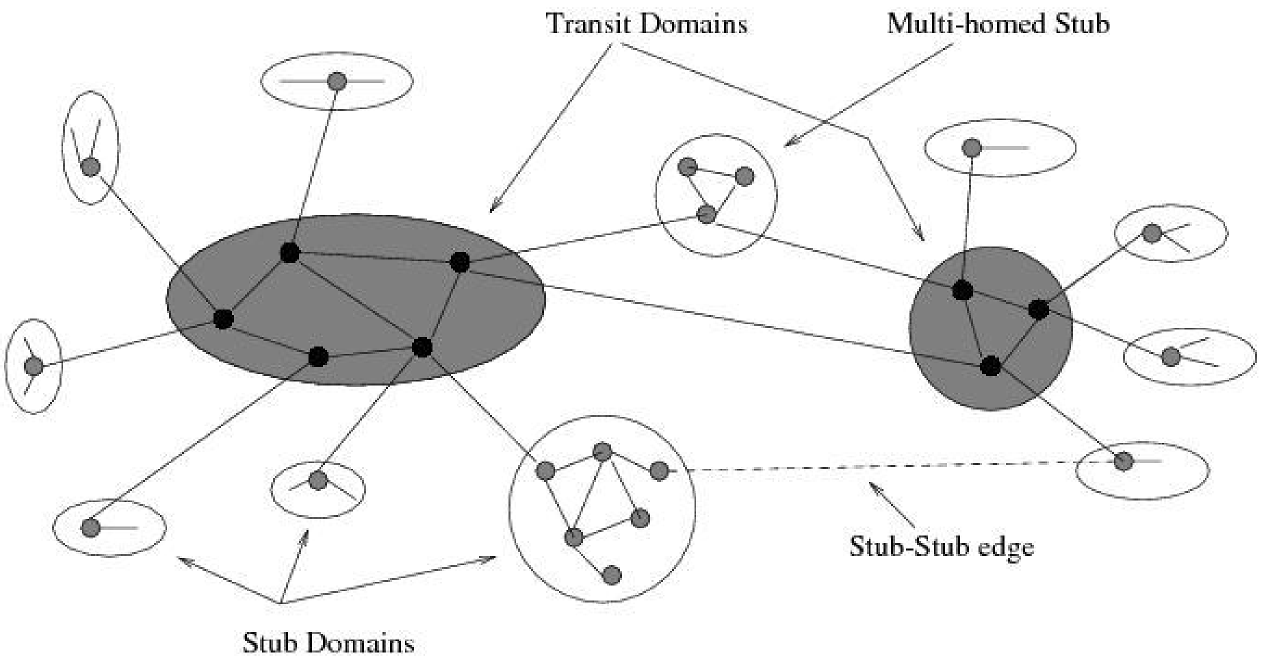
Ordering of probed machines

- Serious problem
- Random, topology independent
  - Inefficient.
  - Hard to handle multiple bottlenecks at once
  - RTT hard to predict (bad setting of the timer)
- Topology dependent
  - Cluster sources and rank clusters from closest to the collector to the farthest
  - Use this ordering to probe sources
  - Sources inside a cluster probed randomly
  - We use Internet coordinates for clustering



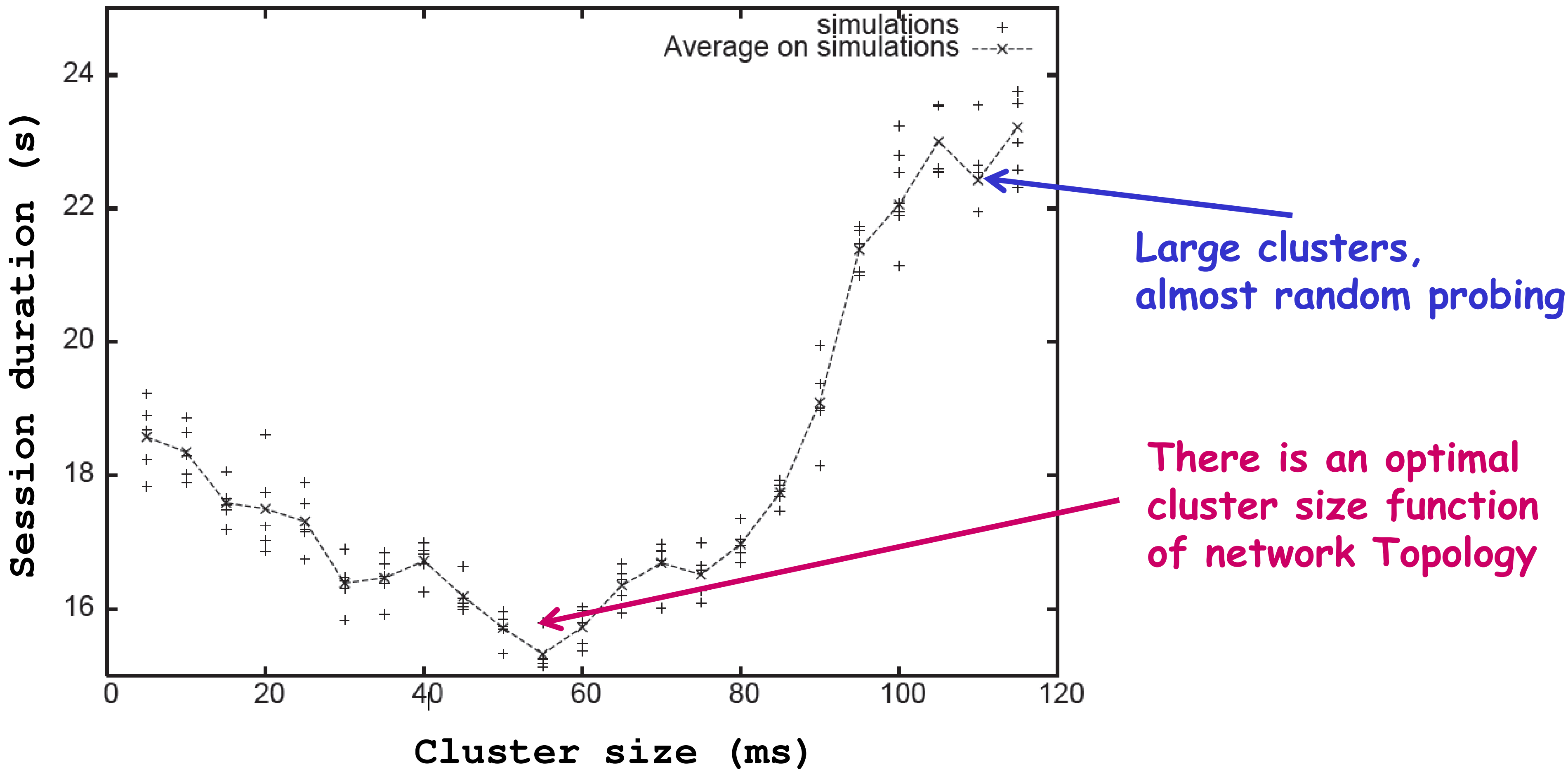
Performance of TICP

- Included in the ns-2 simulator and implemented over PlanetLab
- For ns-2, almost 2000 nodes in a Transit-Stub topology
- 500 probed machines generating a packet each



INRIA SOPHIA ANTIPOLIS

Performance of TICP: Cluster size



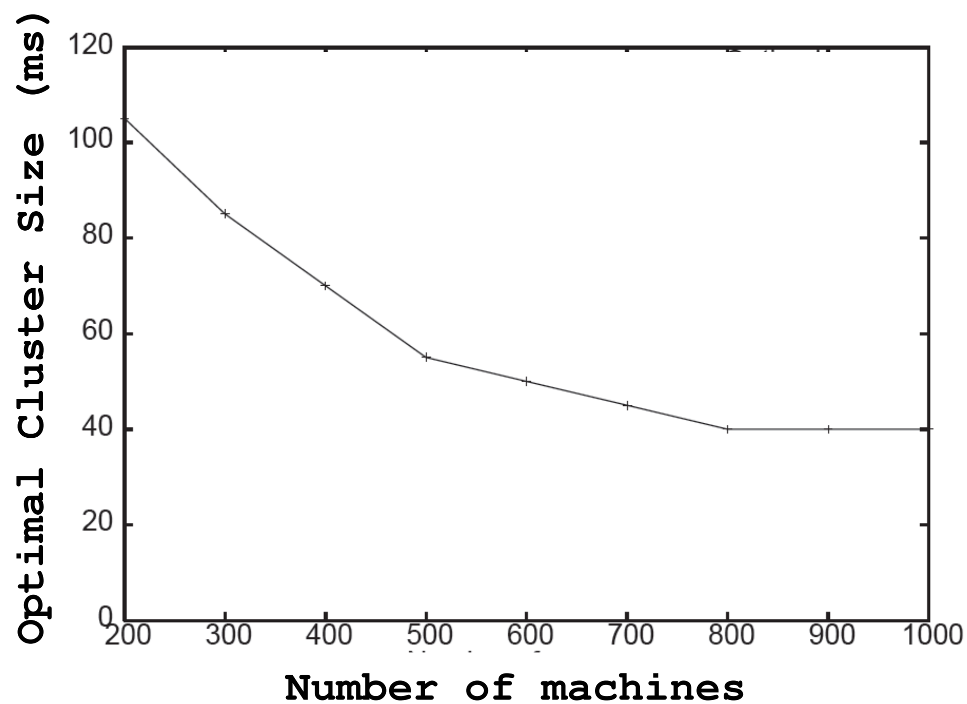
INRIA SOPHIA ANTIPOLIS

Performance of TICP: Cluster size

- Important parameter of the protocol to set.
- Our observation: As the number of sources increases, it converges to some constant value function of the underlying topology and the distribution of bottlenecks.

For example, over our ns-2 topology,

40 ms is a good choice …



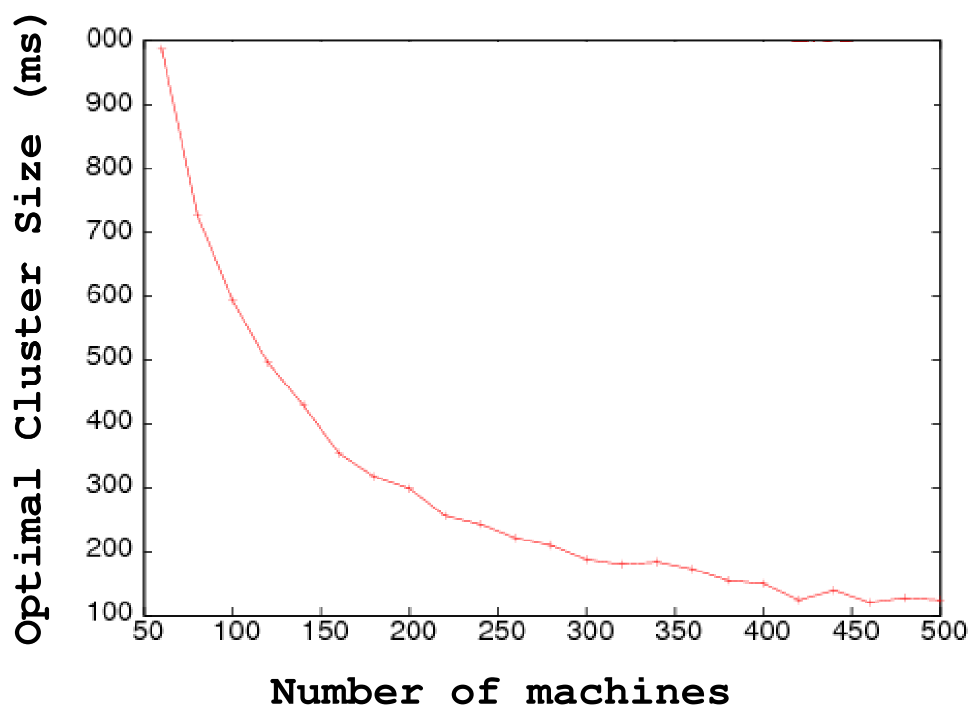
INRIA SOPHIA ANTIPOLIS

Performance of TICP: Cluster size

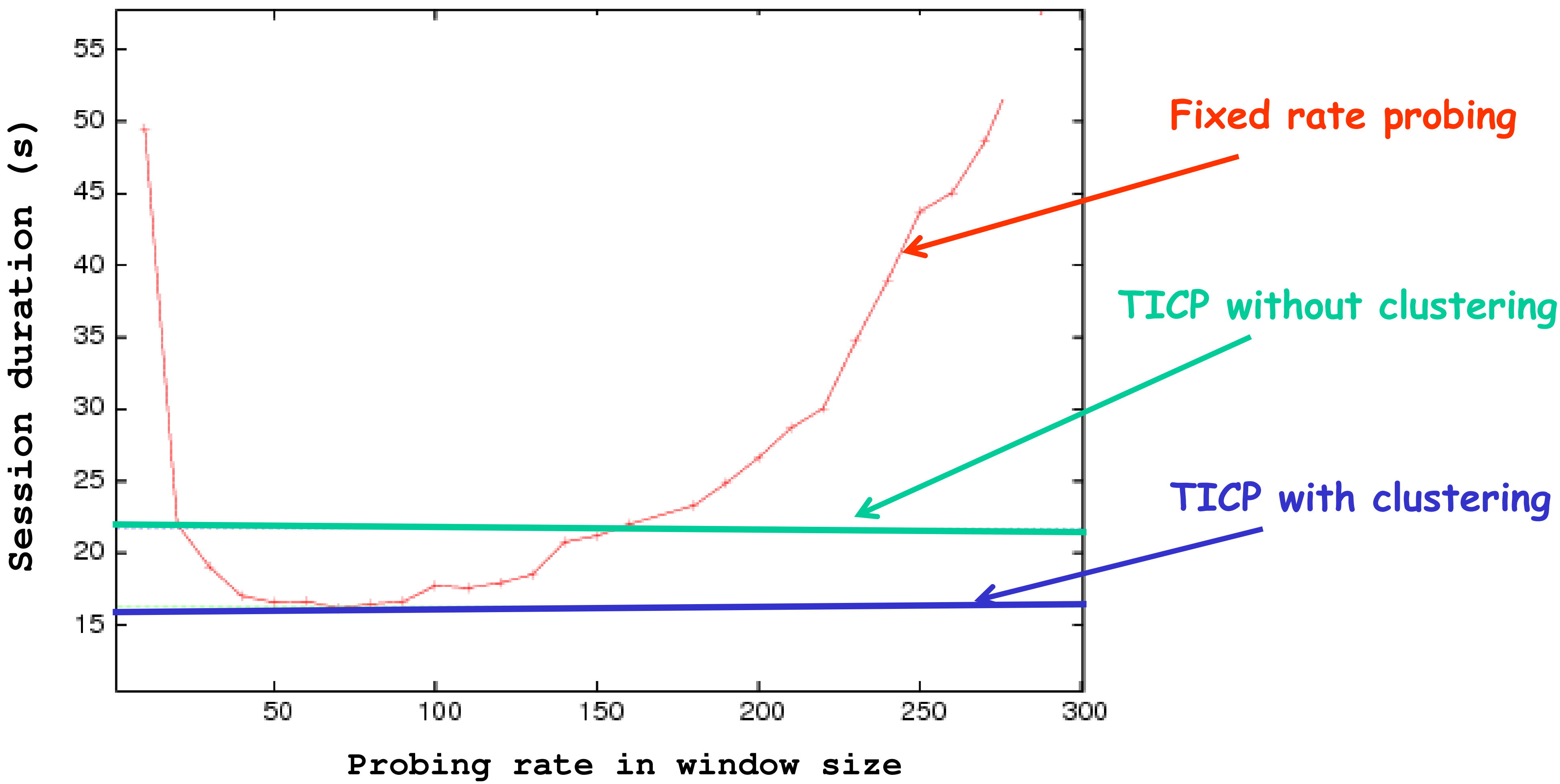
- Important parameter of the protocol to set.
- Our observation: As the number of sources increases, it converges to some constant value function of the underlying topology and the distribution of bottlenecks.

For example, over PlanetLab,

100 ms is a good choice …

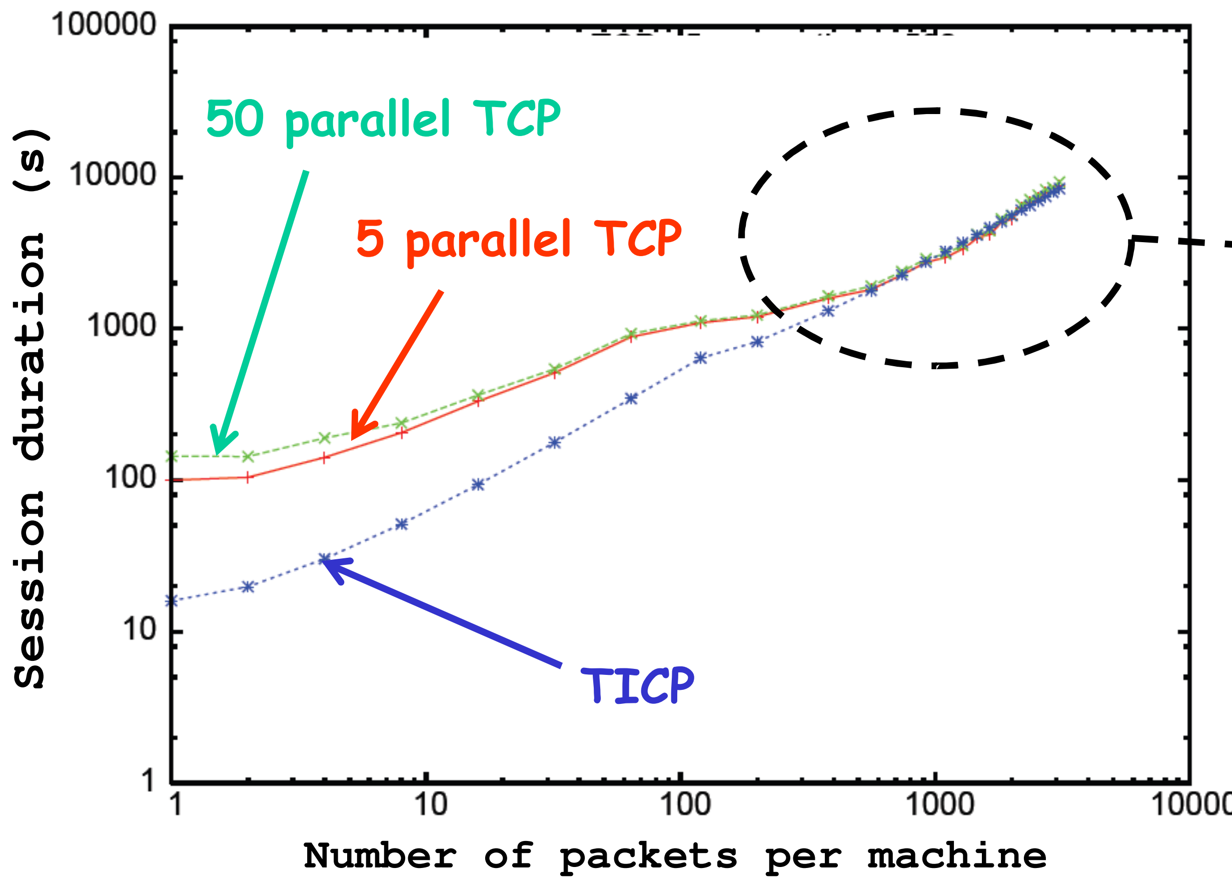


TICP vs. constant probing rate



Compared to parallel TCP

- What if parallel TCP connections were used ?
- TICP behaves better due to its multiplexing capability



TICP stops overperforming parallel TCP for large data/per machine

References on TICP

Visit http://www.inria.fr/planete/chadi/ticp

Karim Sbai and Chadi Barakat, “Experiences on enhancing data collection in large networks”, to appear in Computer Networks.

Chadi Barakat, Mohammad Malli, Naomichi Nonaka, "TICP: Transport Information Collection Protocol", Annals of Telecommunications, vol. 61, no. 1-2, pp. 167-192, January-February 2006.

Conclusions - Perspectives

- A set of solutions for efficient network monitoring
  - Network wide traffic sampling
  - Congestion control for network probing
- Domain will keep evolving
  - Deal with new applications (social networks, P2P) and architectures
  - Measurements at the service of applications and users (localization, topology-aware adaptation, diagnosis)
- Our future research will focus on leveraging correlation (spatial and temporal) to achieve better monitoring
  - Correlating sampled flow measurements made by routers
    - The ECODE FP7 project, 2008 - 2011
  - Correlating end-to-end measurements for network diagnosis
    - The CMON project with Thomson and the Grenouille.com, 2009-2012