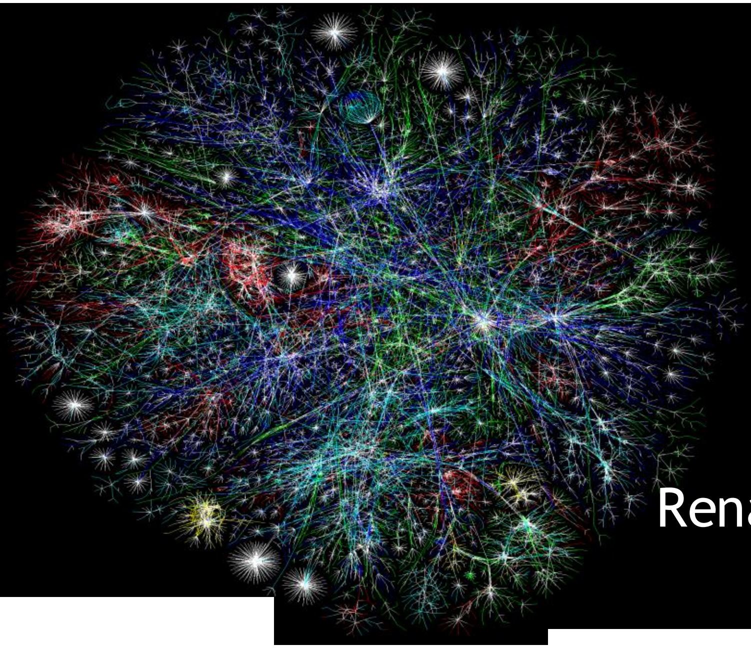


Predicting and Tracking Internet Path Changes

Ítalo Cunha
Renata Teixeira, Darryl Veitch, and Christophe Diot

UPMC SORBONNE UNIVERSITÉS
CNRS dépasser les frontières
THE UNIVERSITY OF MELBOURNE
technicolor

Problem statement

Goal: track large number of paths

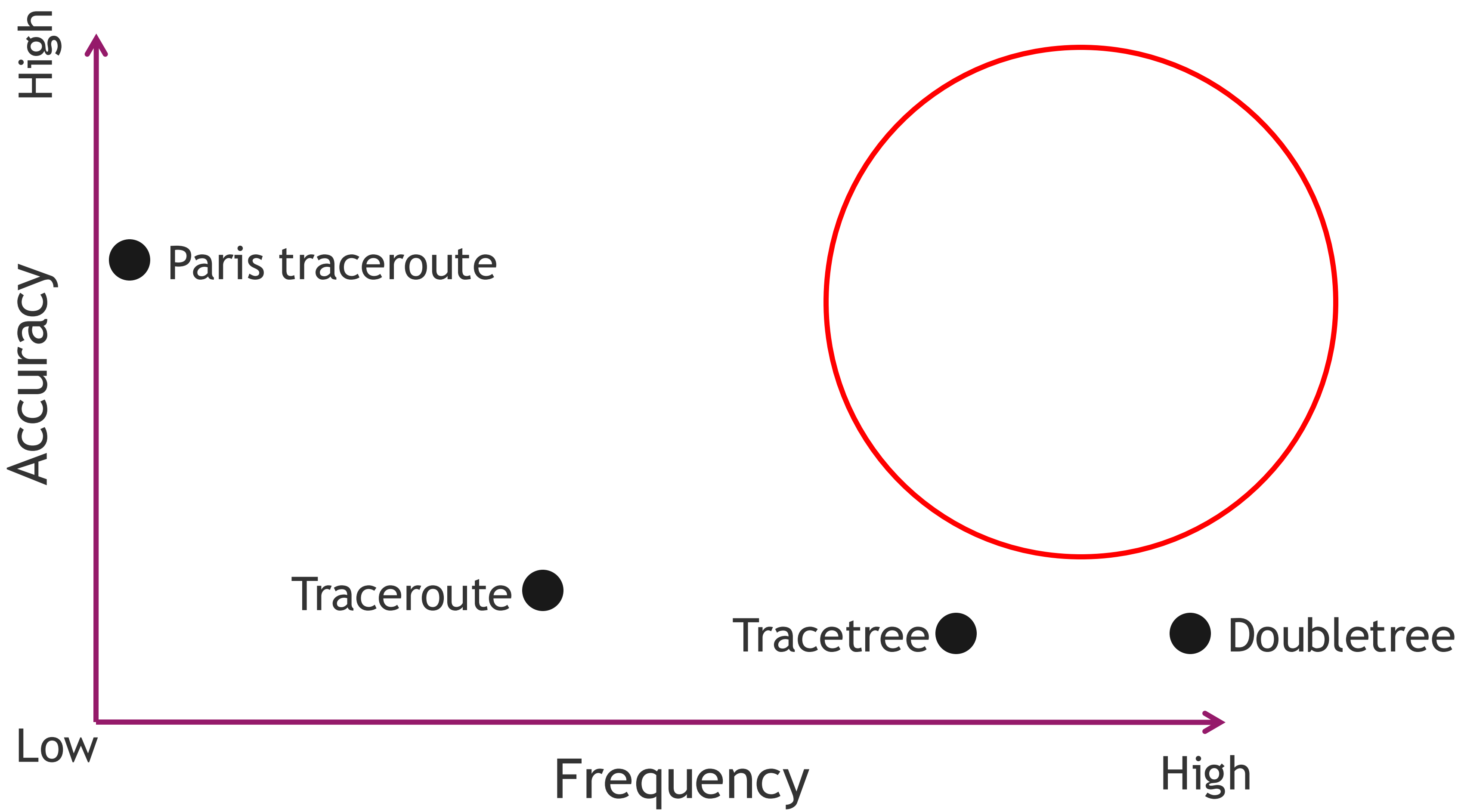
Current approach: traceroute-style measurements

Challenges

- Cannot measure frequently enough to detect all changes
  - Network and system limitations
- Accurate measurements require extra probes
  - Identify all paths under load balancing

technicolor

Frequent vs. accurate measurements



Approach

Observation: Internet paths are mostly stable

- Current techniques waste probes

Probe according to path stability

Separate tasks of change *detection* and change *remapping*

- Use lightweight probing to detect changes faster
- Remap with Paris traceroute to get accurate path measurements

Contributions

NN4: Predicting Internet path changes

- Distinguish between stable and unstable paths

DTrack: Tracking Internet path changes

- Lightweight probing process to detect changes
- Allocates more probes to unstable paths

Predicting path changes

Prediction goals

- Time until the next change
- Number of changes in a time interval
- Whether a path will change in a time interval

Identify path features that can help with prediction

- Features must be computable from traceroute measurements
  - Characteristics of the current path
  - Characteristics of the last path change
  - Behavior of the path in the recent past

technicolor

Feature selection

Use RuleFit to identify the relative importance of features

1. Fraction of time path was active in the past (prevalence)
2. Number of changes in the past
3. Number of previous occurrences of the current path instance
4. Path age

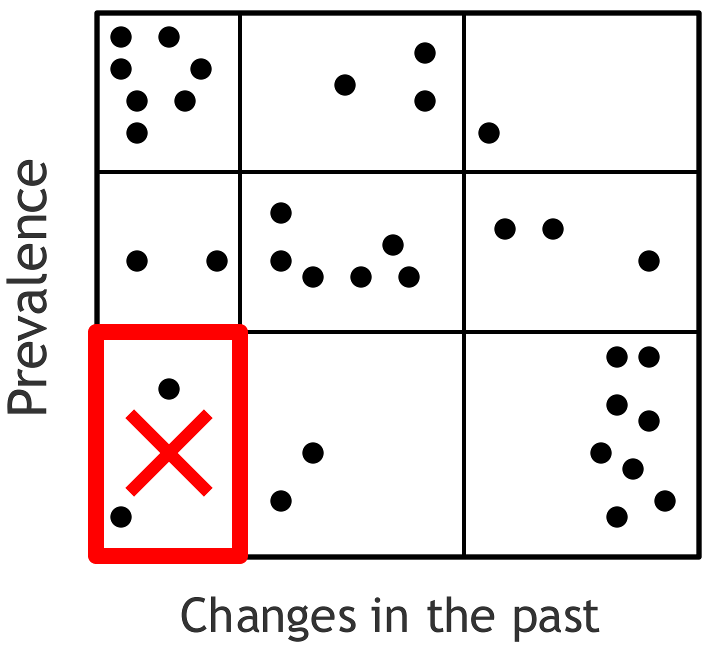
Four most important features carry all the predictive information

NN4 predictor

RuleFit is CPU-intensive and hard to integrate in other systems

NN4 is based on the nearest-neighbor scheme

- Compute neighbors by partitioning the path feature “state-space”
  - Boundaries computed from feature distributions
- Prediction computed as the average behavior of all neighbors



technicolor

FastMapping data

Frequent path measurements

- 5 times faster than Paris traceroute

Complete information about routers performing load balancing

- Required to differentiate load balancing from routing changes

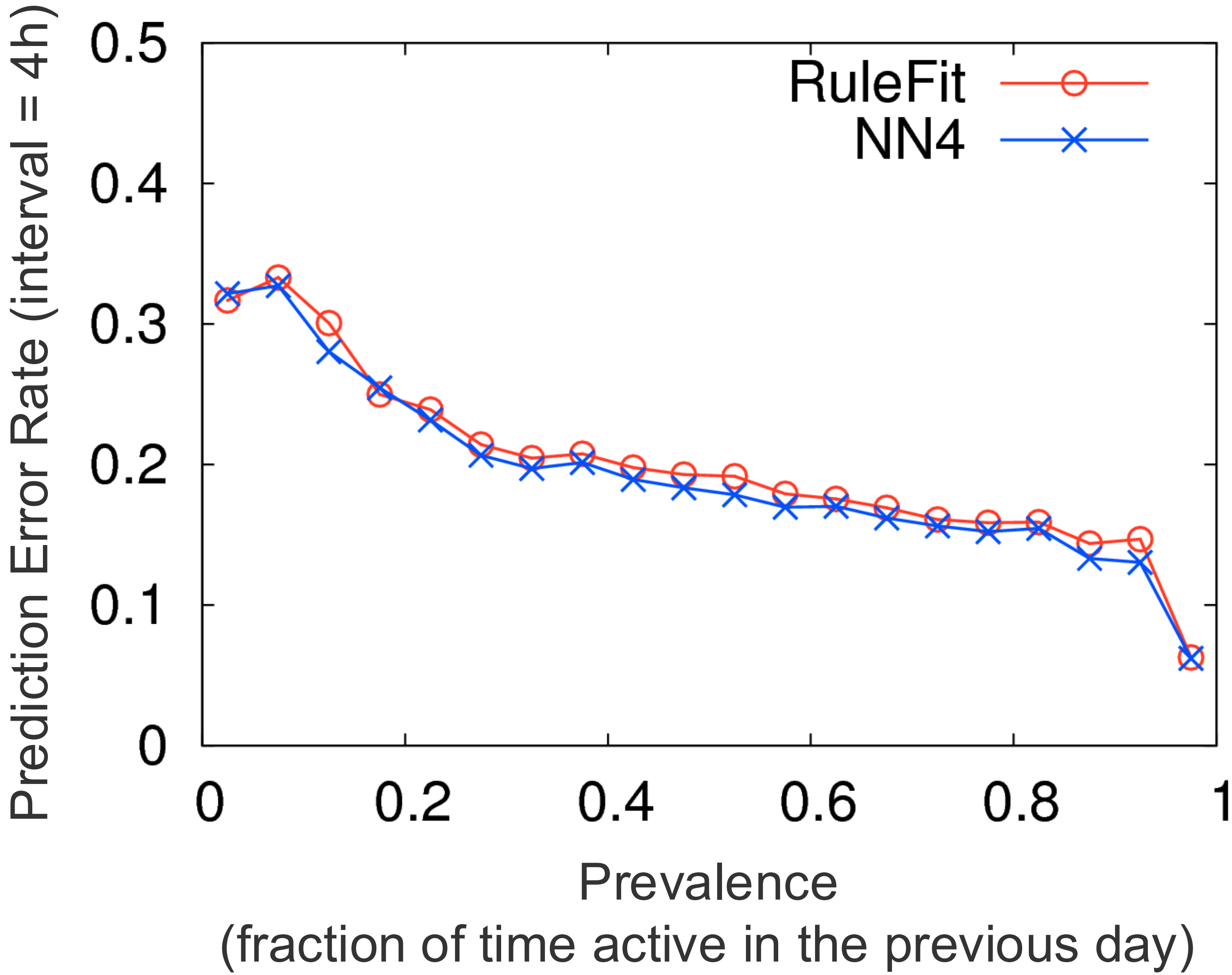
70 PlanetLab hosts probing 1000 destinations

5 weeks of data starting September 1st, 2010

Dataset covers 7942 ASes and 97% of the large ASes

technicolor

NN4 performance



technicolor

NN4: summary

NN4 is lightweight, easy to integrate, and as accurate as RuleFit

Prediction is not highly accurate

- It is possible to distinguish unstable from stable paths

DTrack

Goal: Given a probing budget, detect as many changes as possible

Allocates probing rates *per path* using NN4's predictions

Targets probes along each path

- Reduce redundant probes at shared links
- Spread probes over time

technicolor

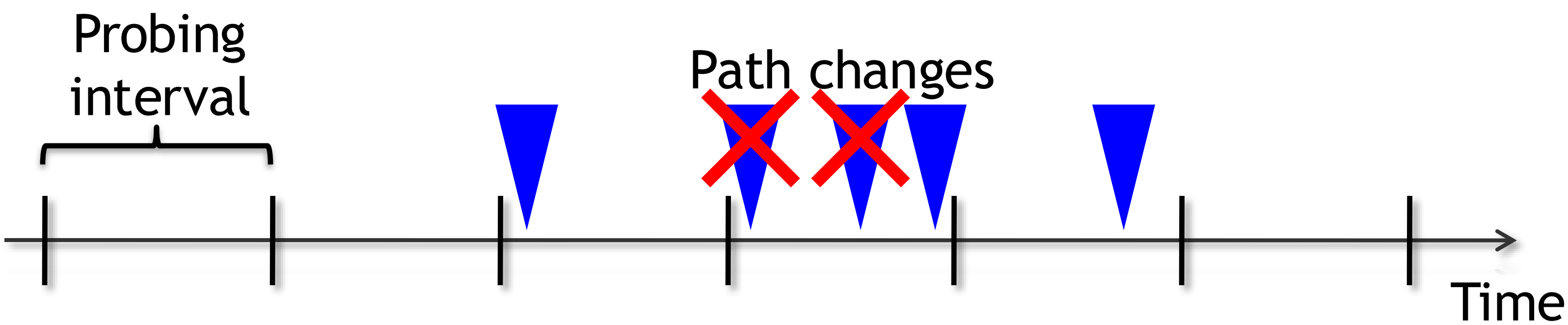
Probe rate allocation

Allocate rates that minimize total number of missed changes

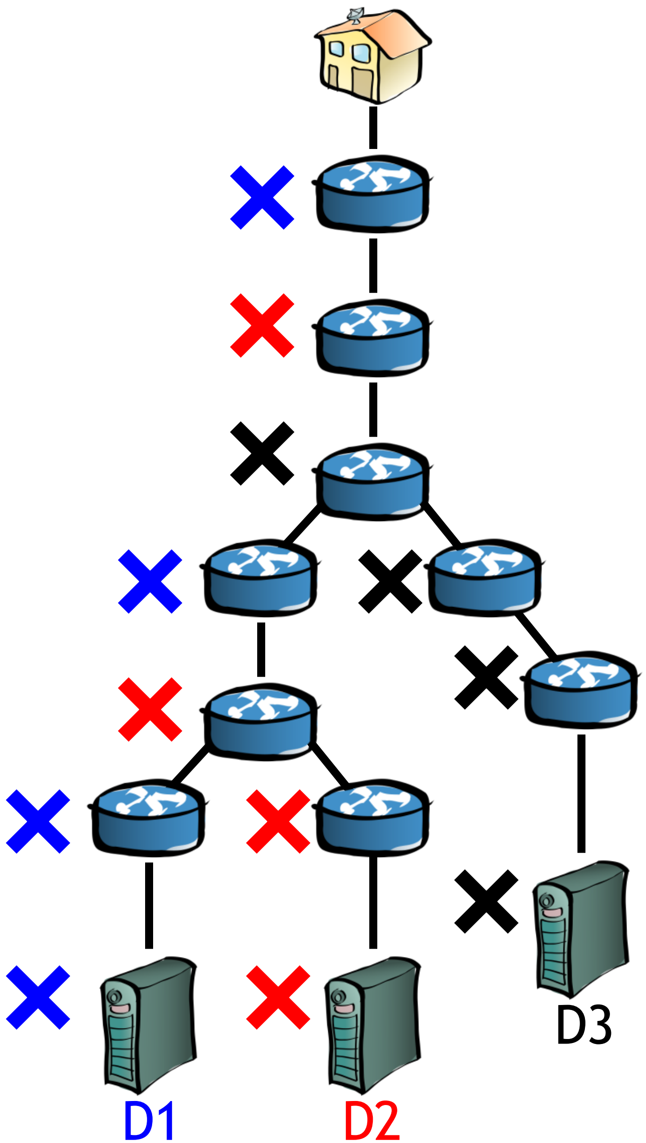
Model changes in each path as a Poisson process

- Estimate the rate of changes using NN4

Compute missed changes as function of probing rate



Probe targeting overview



Evaluation

Method

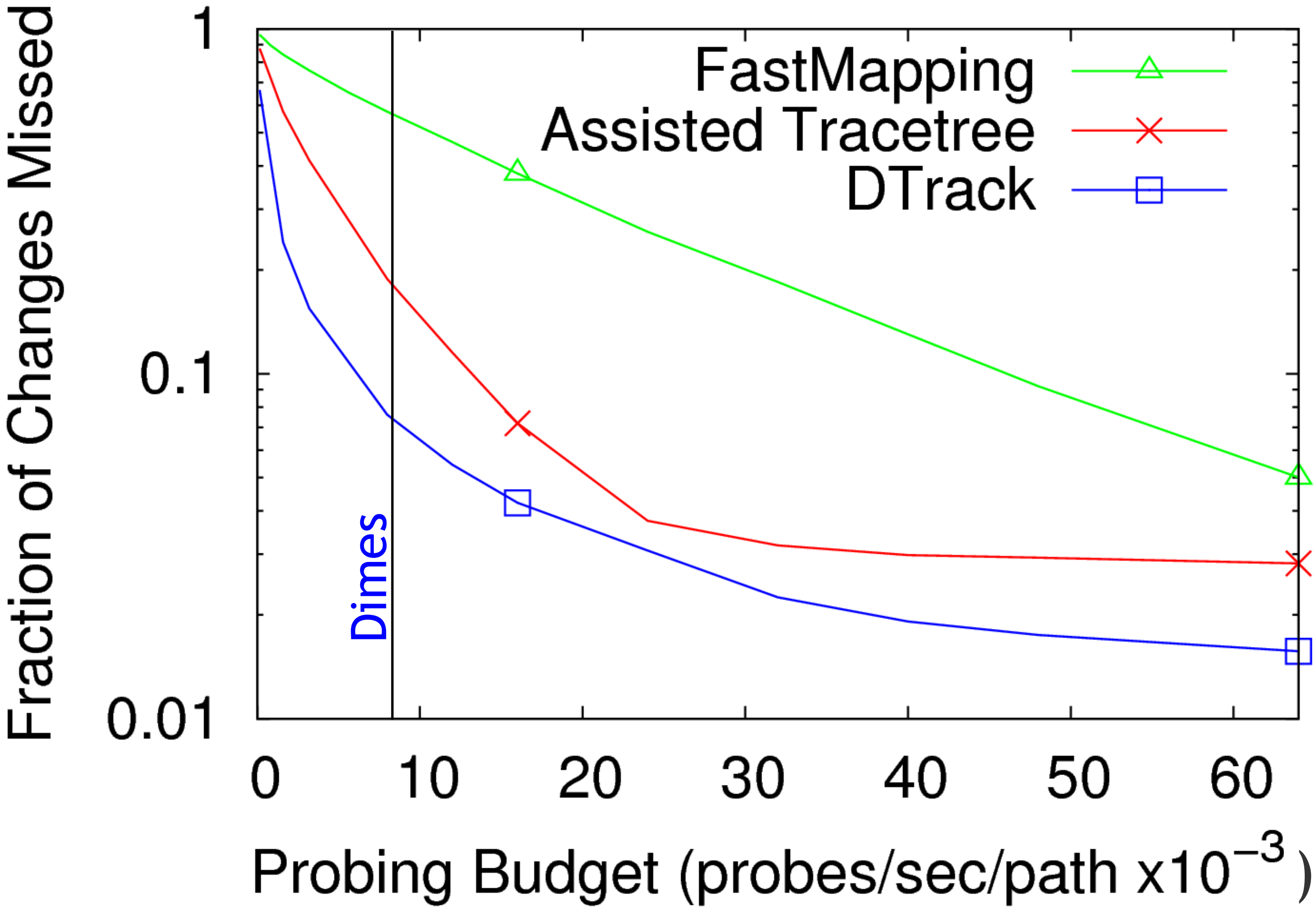
- Trace-driven simulations using the FastMapping dataset

Performance metrics

- Number of missed changes
- Change detection delay

Compare against FastMapping and Tracetree

Number of changes missed



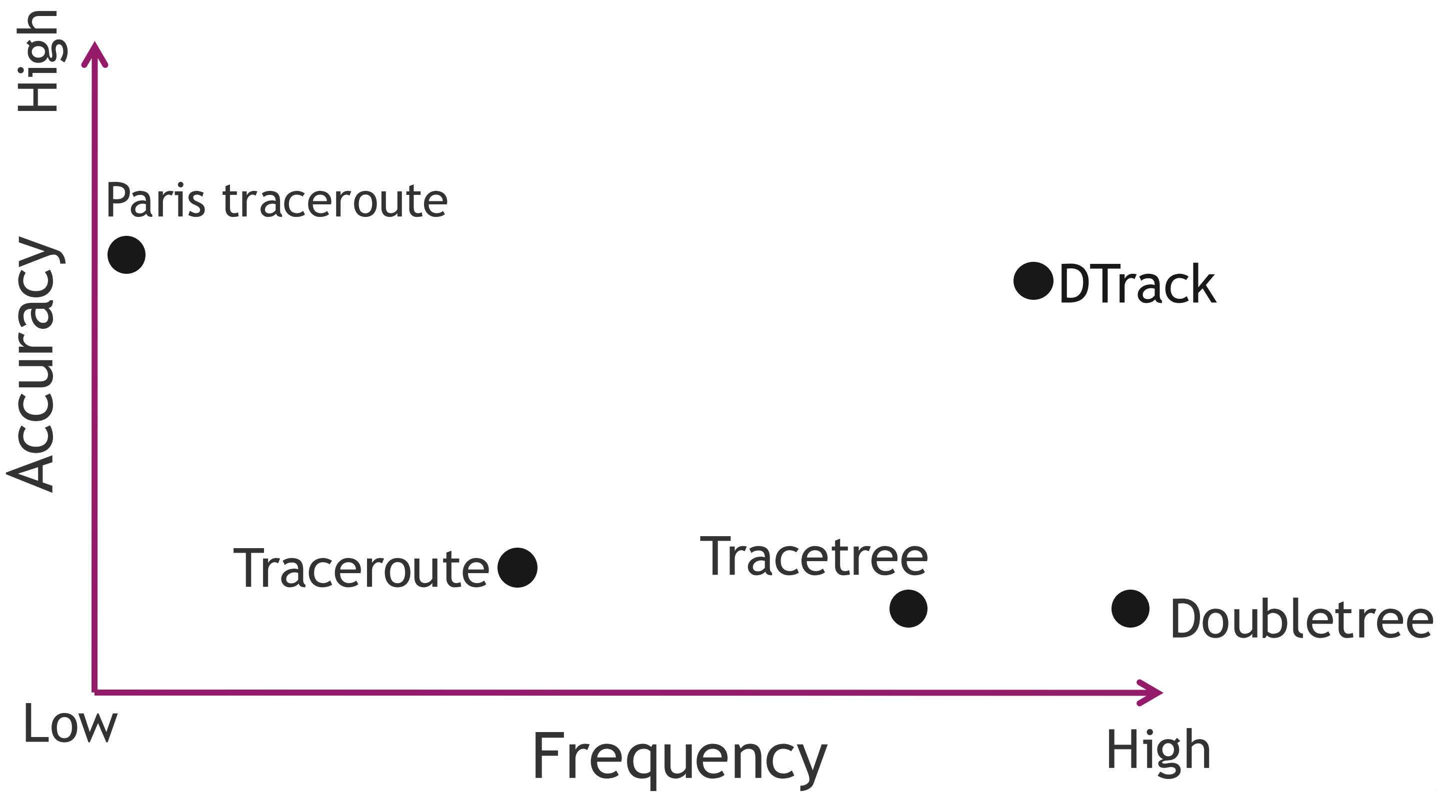
technicolor

Conclusion

NN4: A lightweight predictor of path changes

- Distinguishes stable and unstable paths

DTrack detects more changes than the current state-of-the-art



technicolor

Future work

Deploy DTrack on gateways

Improve NN4's prediction accuracy

- Use extra information like BGP updates

Extend DTrack

- Reduce remapping cost
- Coordinate probing across multiple monitors

Thank you!

Questions?

DTrack vs. FastMapping PlanetLab deployment

Run DTrack and FastMapping simultaneously

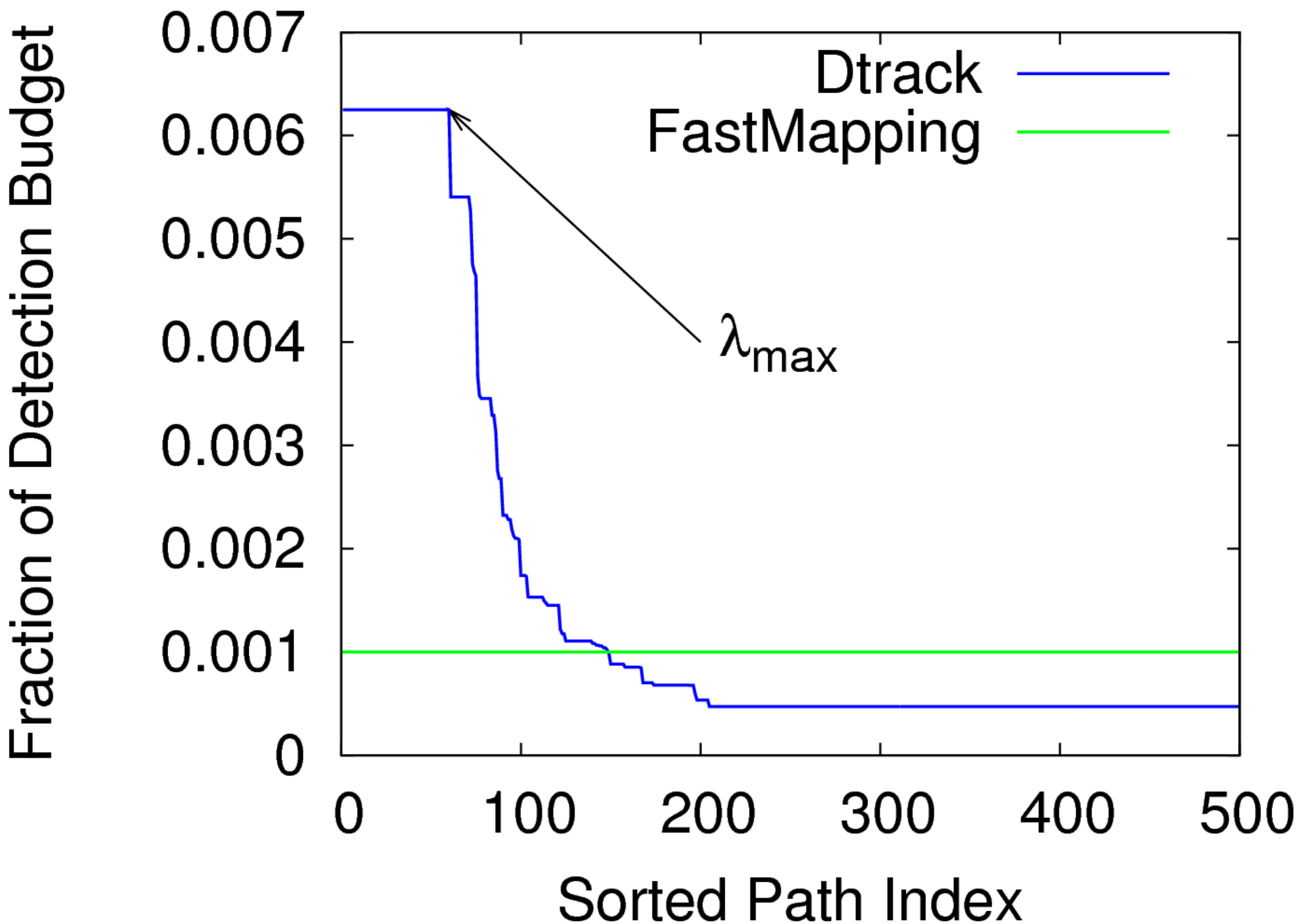
with the same sampling budget

No ground truth

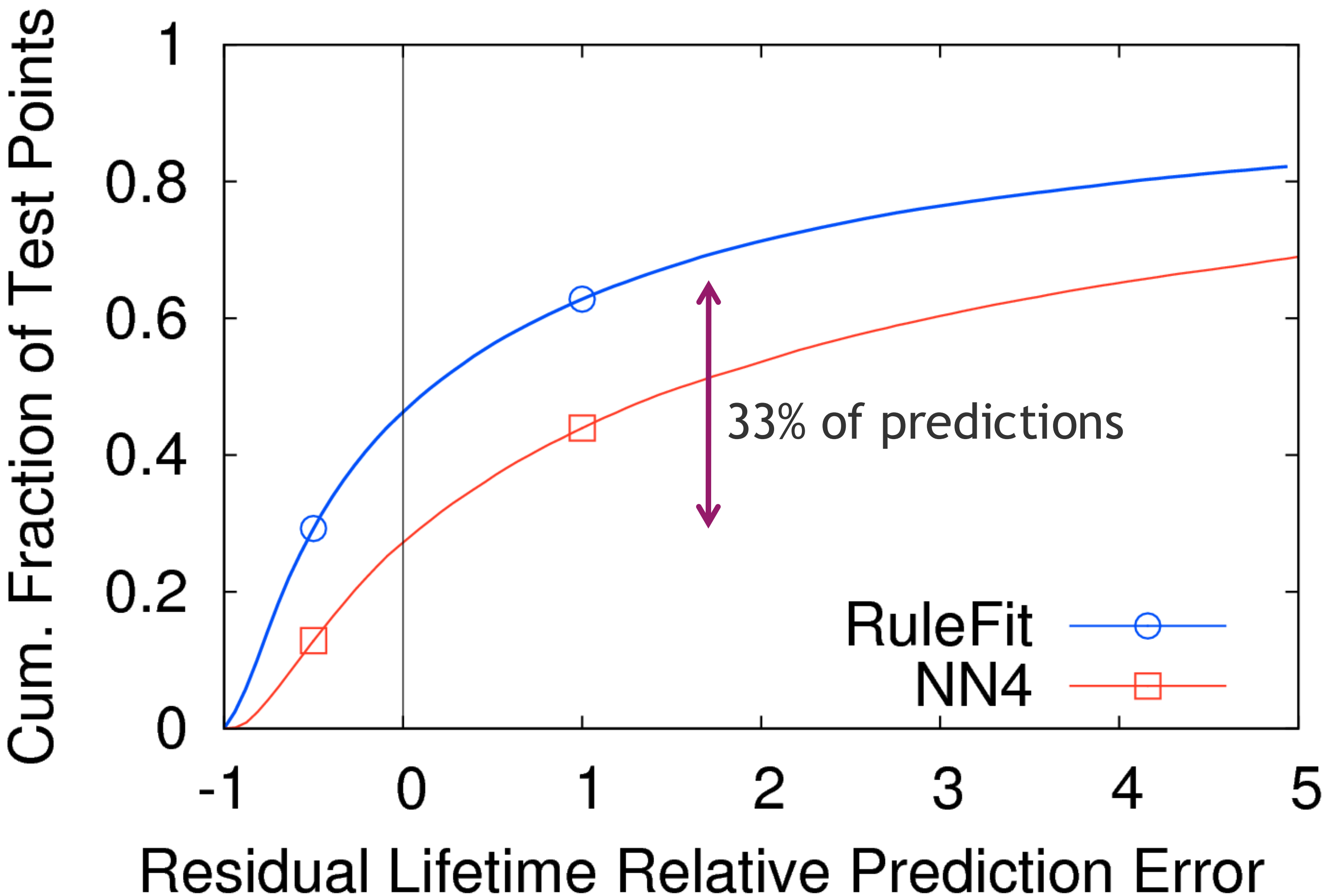
DTrack detected almost 5 times more changes than FastMapping

- Up from 2.2 times in the trace-driven simulations

Probing rate allocation



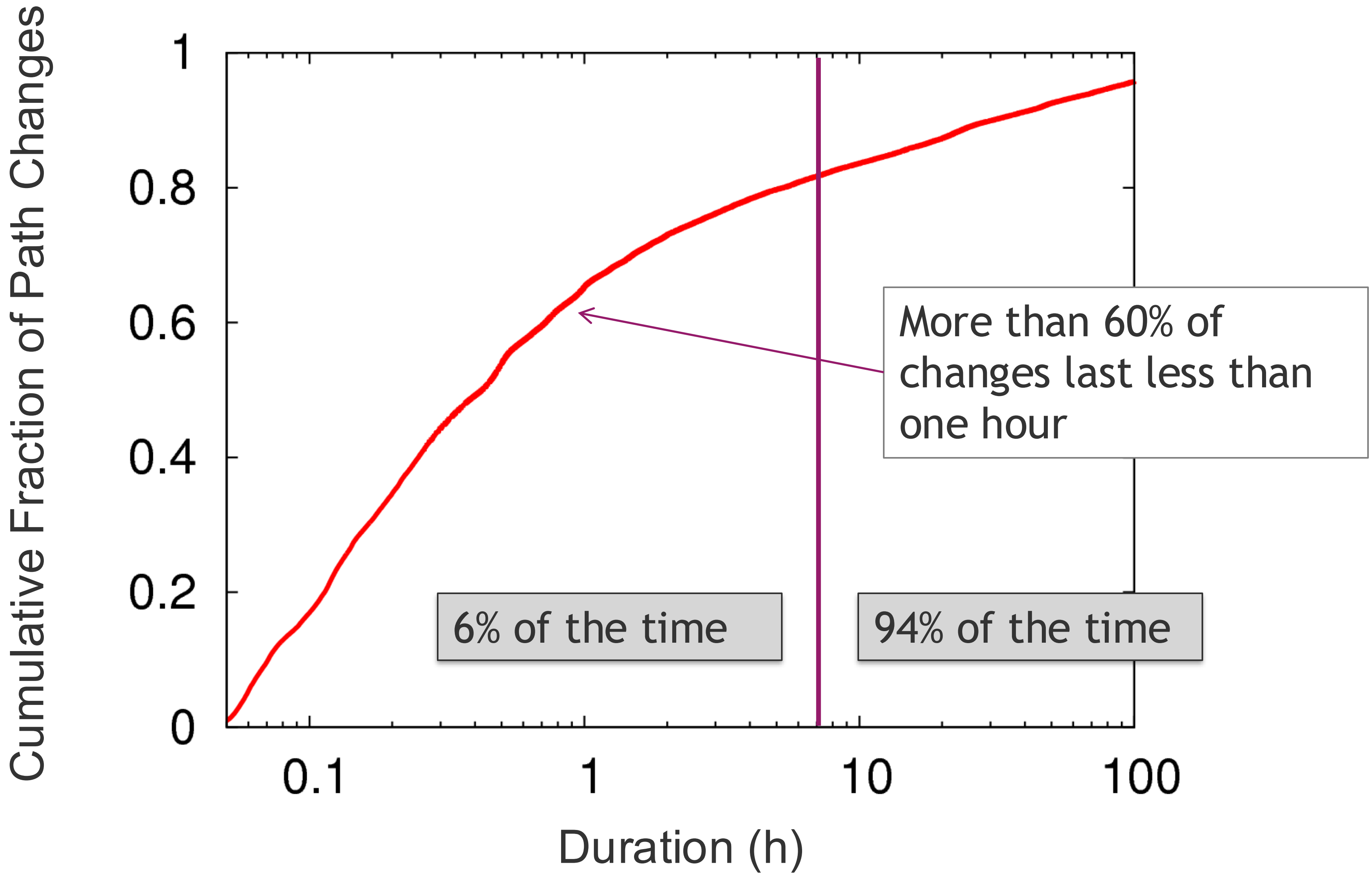
Prediction accuracy of residual lifetime



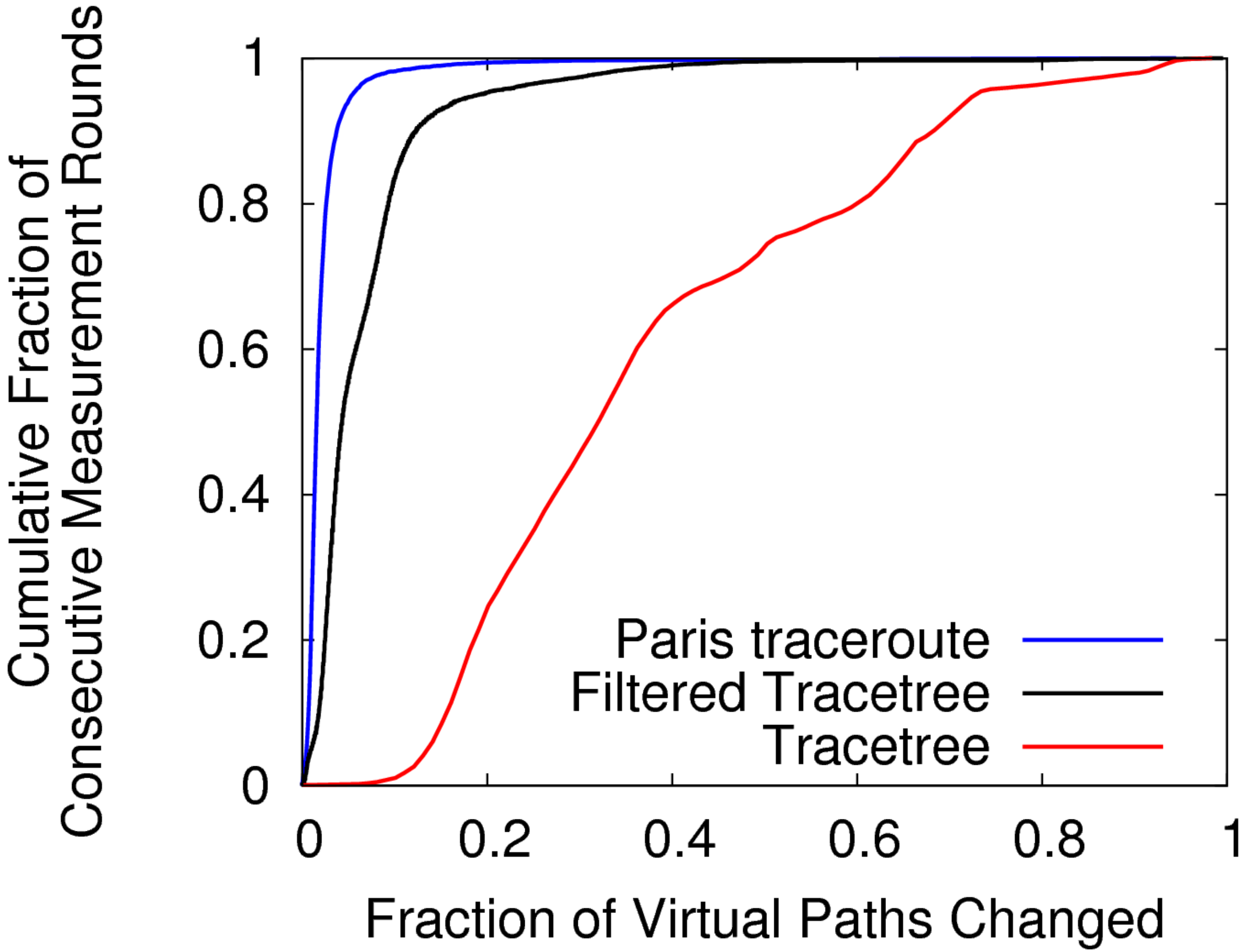
technicolor

Characterization / FastMapping

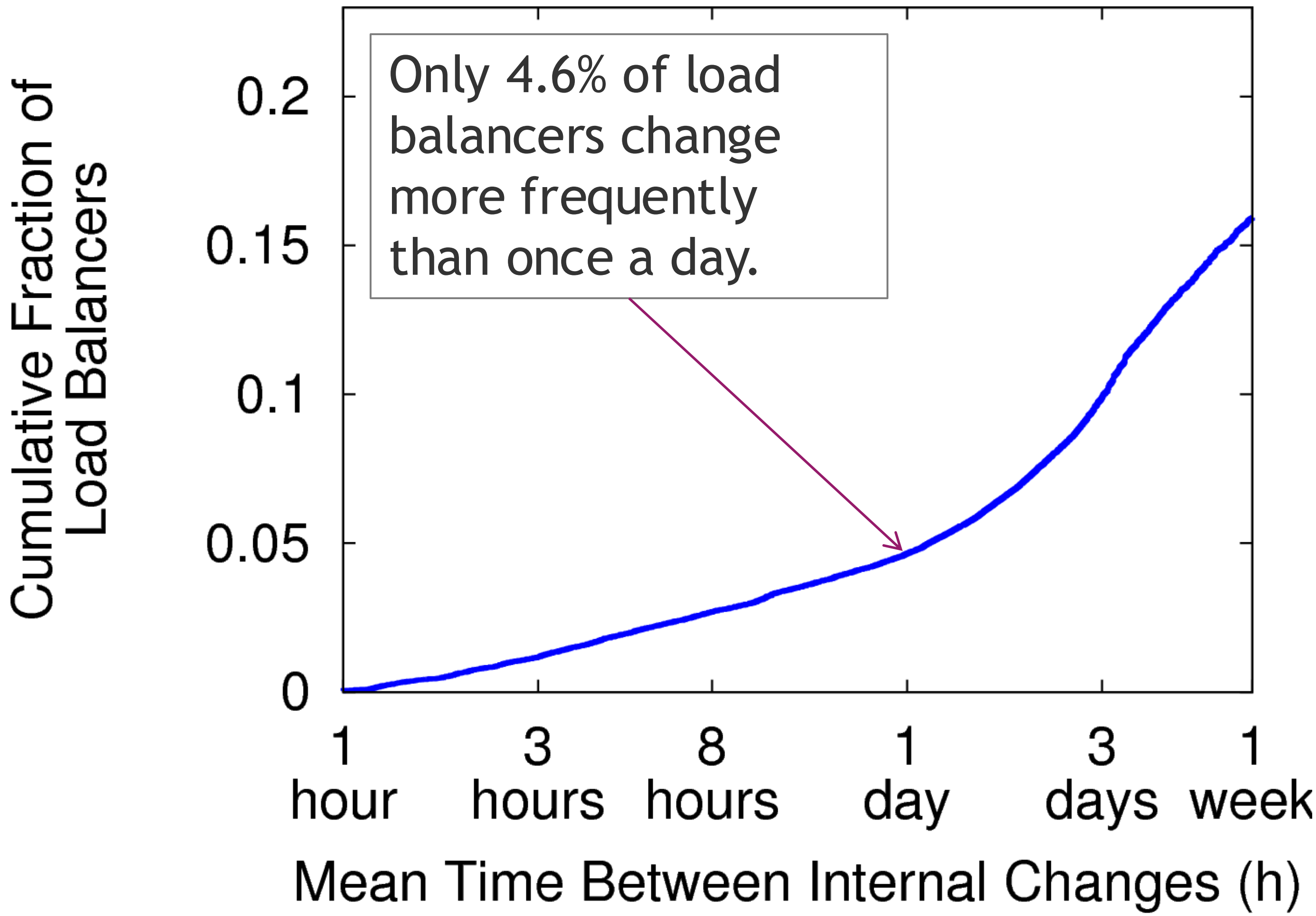
Path change duration



Filtering load balancers from Tracetree



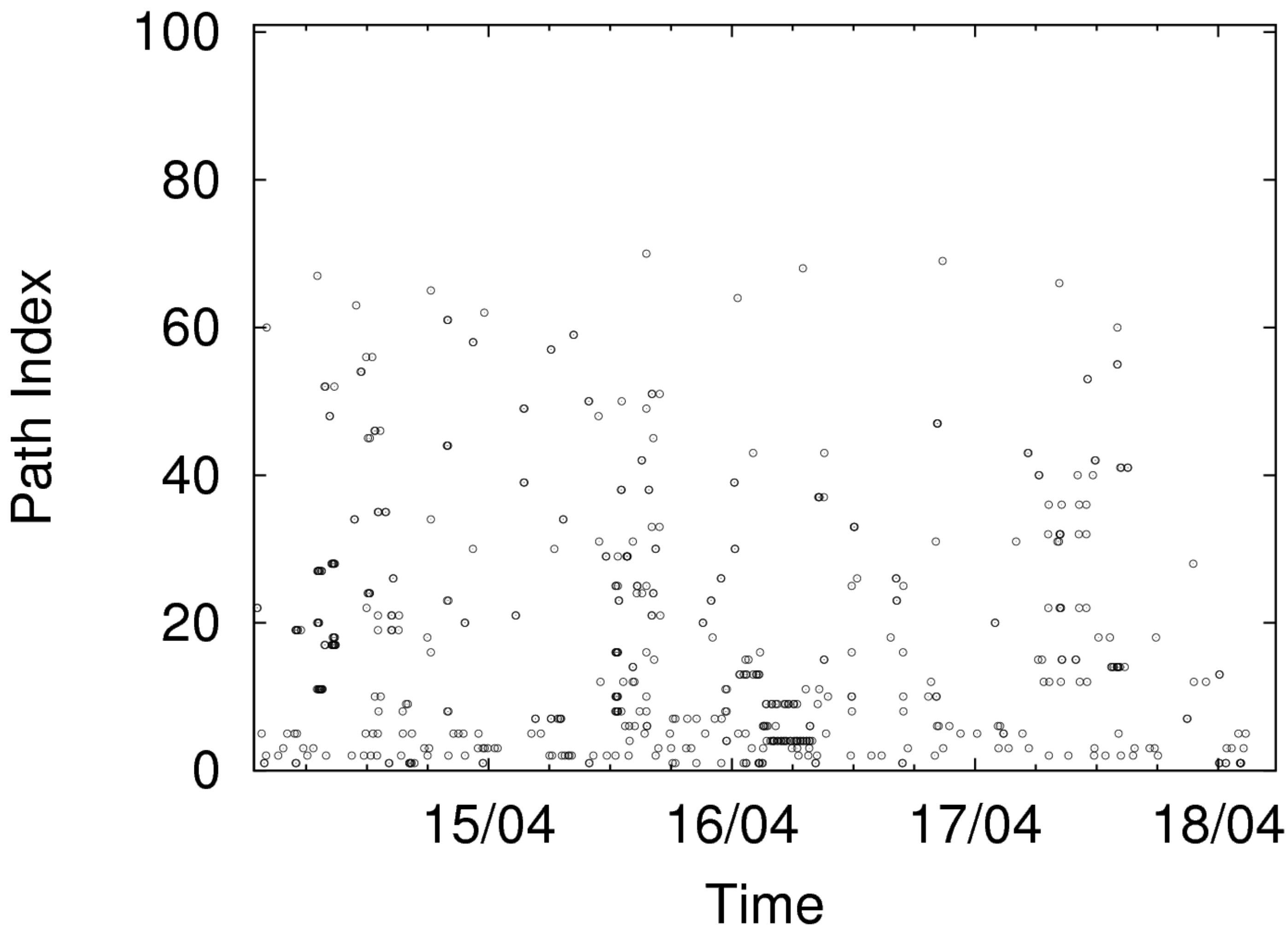
Load balancer stability



technicolor

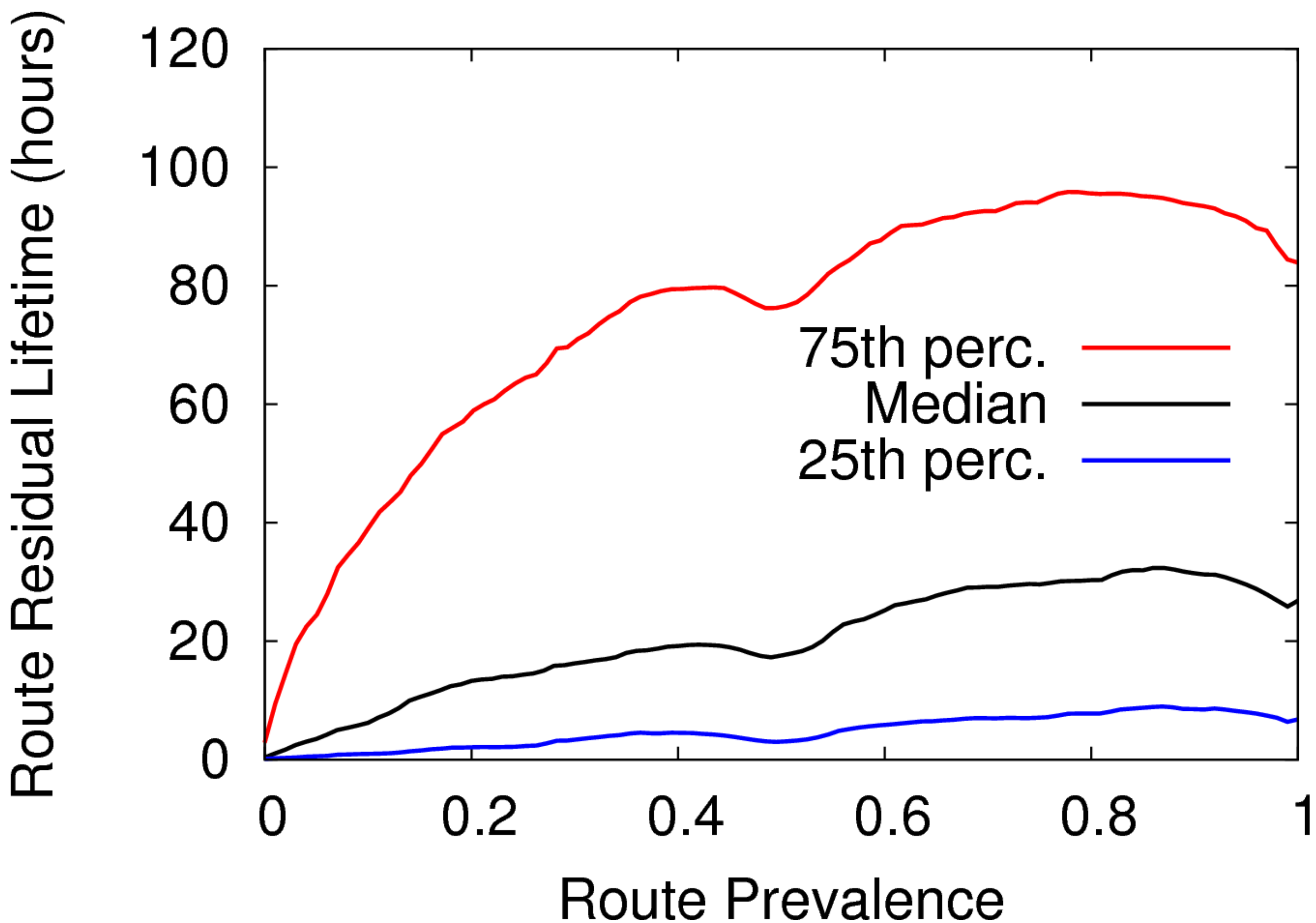
Overview of route dynamics

Paths are stable most of the time,
but go through short-lived instability periods



Prediction / NN4

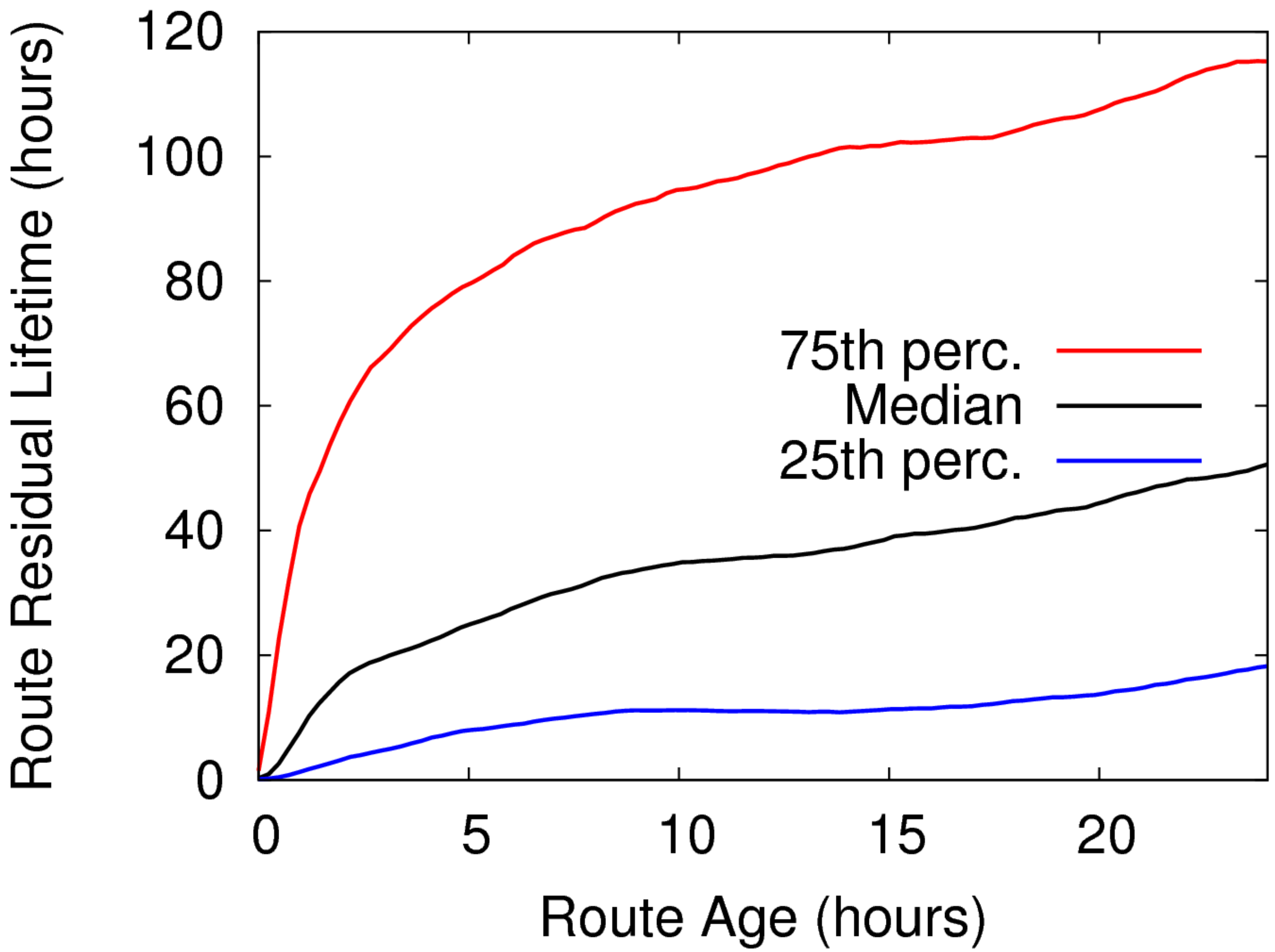
A first look into path change prediction



(frac. of time active in the previous day)

technicolor

Old routes have higher residual lifetimes



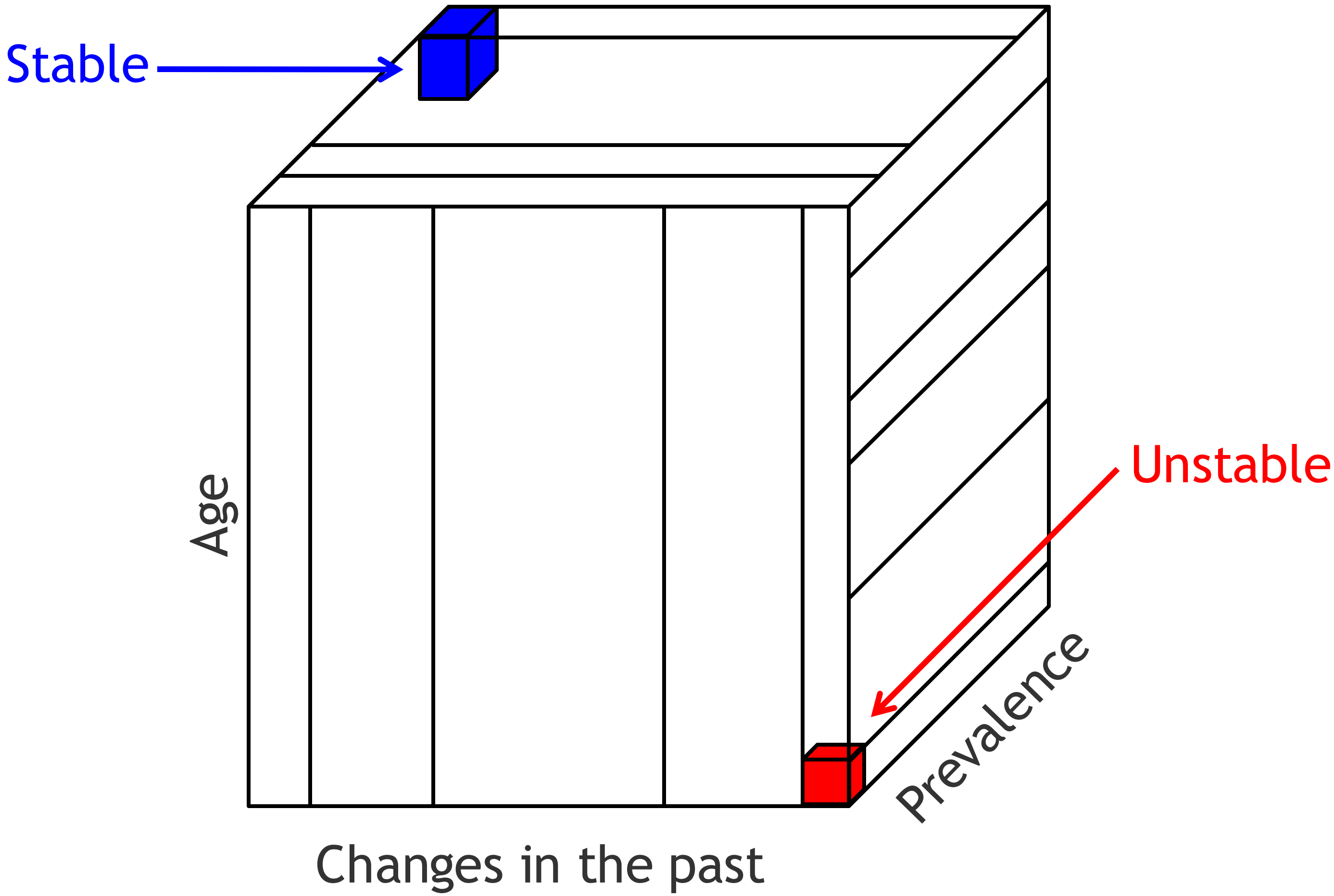
technicolor

NN4 predictor

Based on the nearest-neighbor scheme

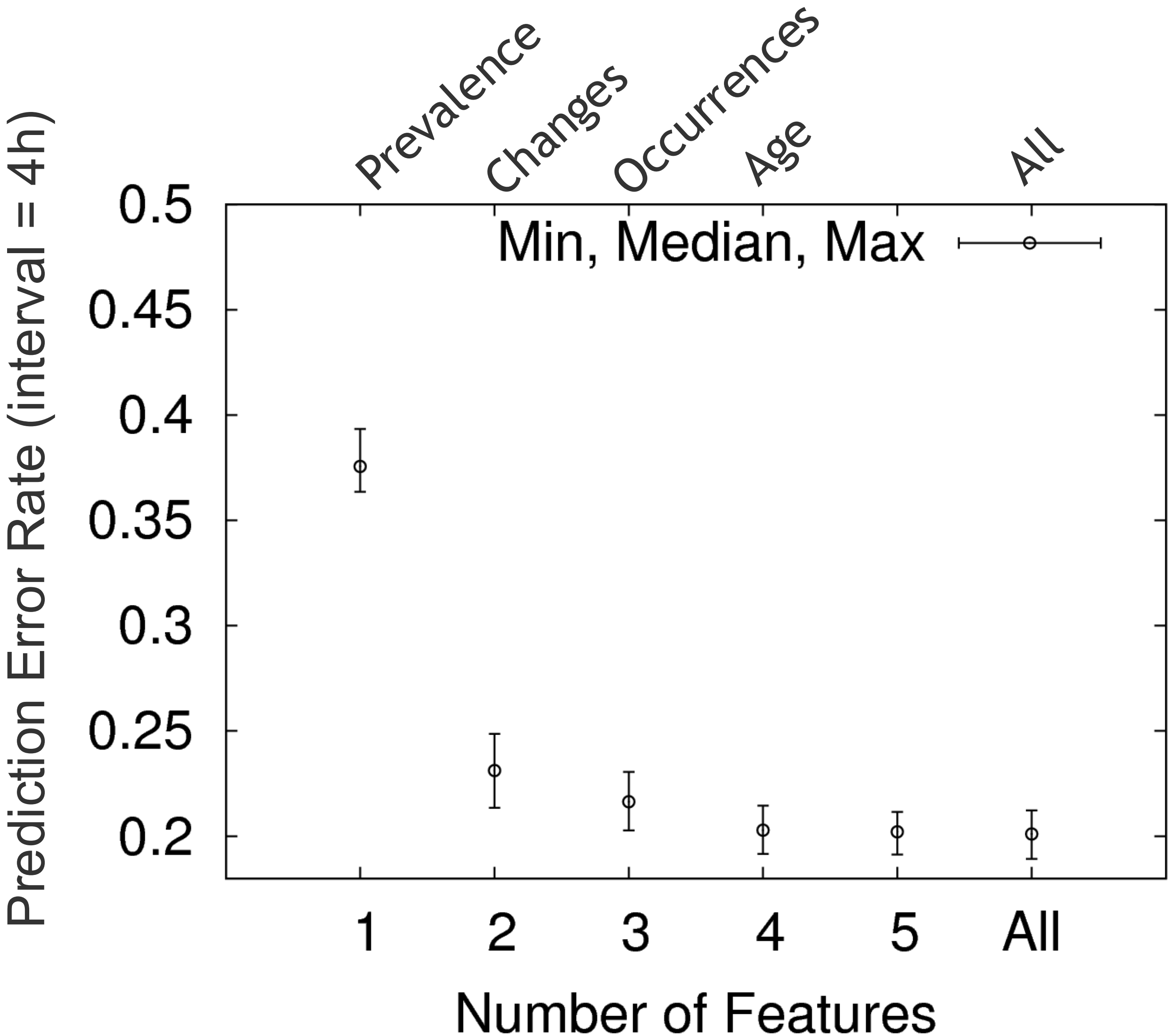
Compute neighbors by partitioning the path feature “state-space”

- Partition boundaries computed automatically

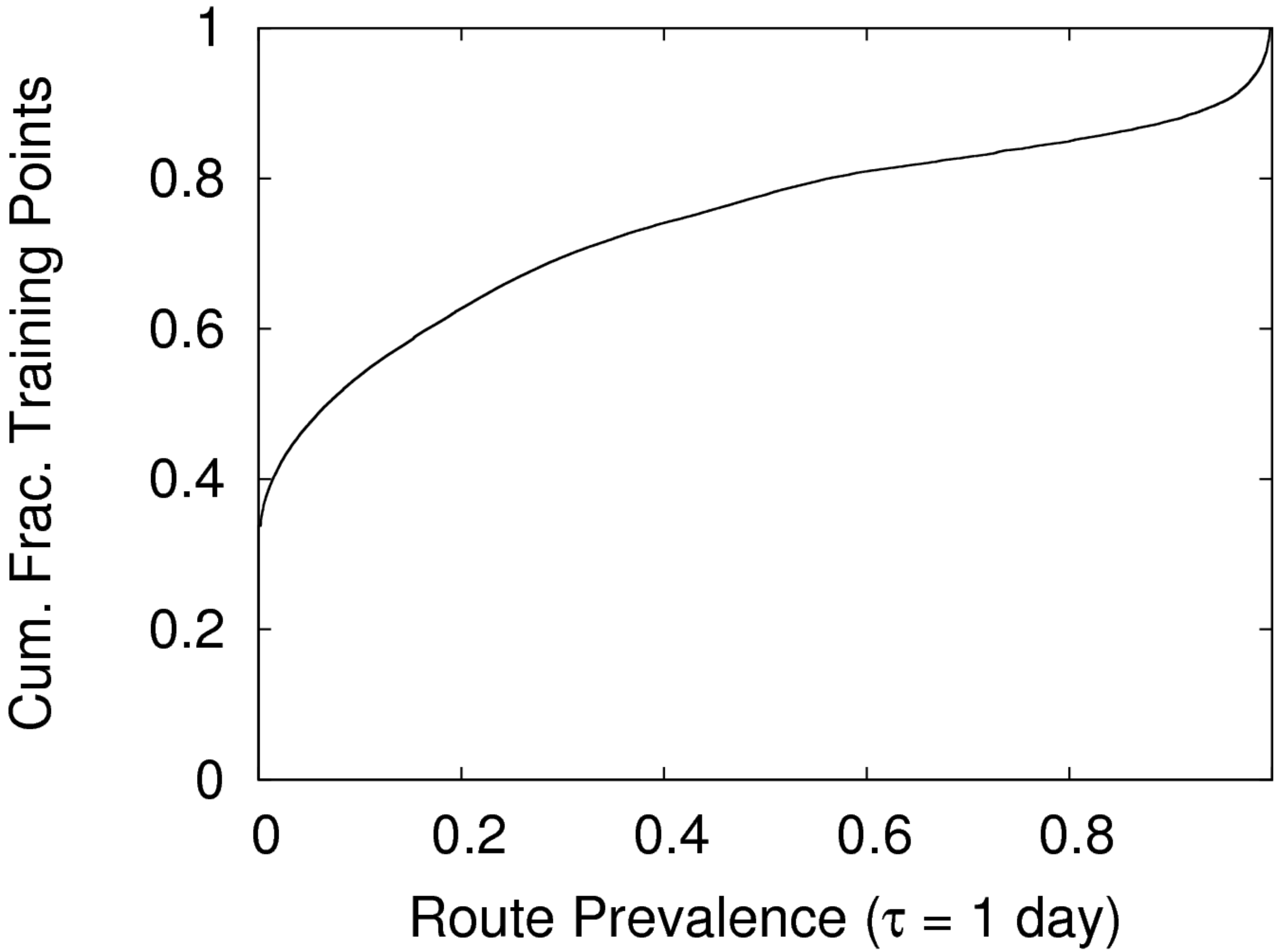


technicolor

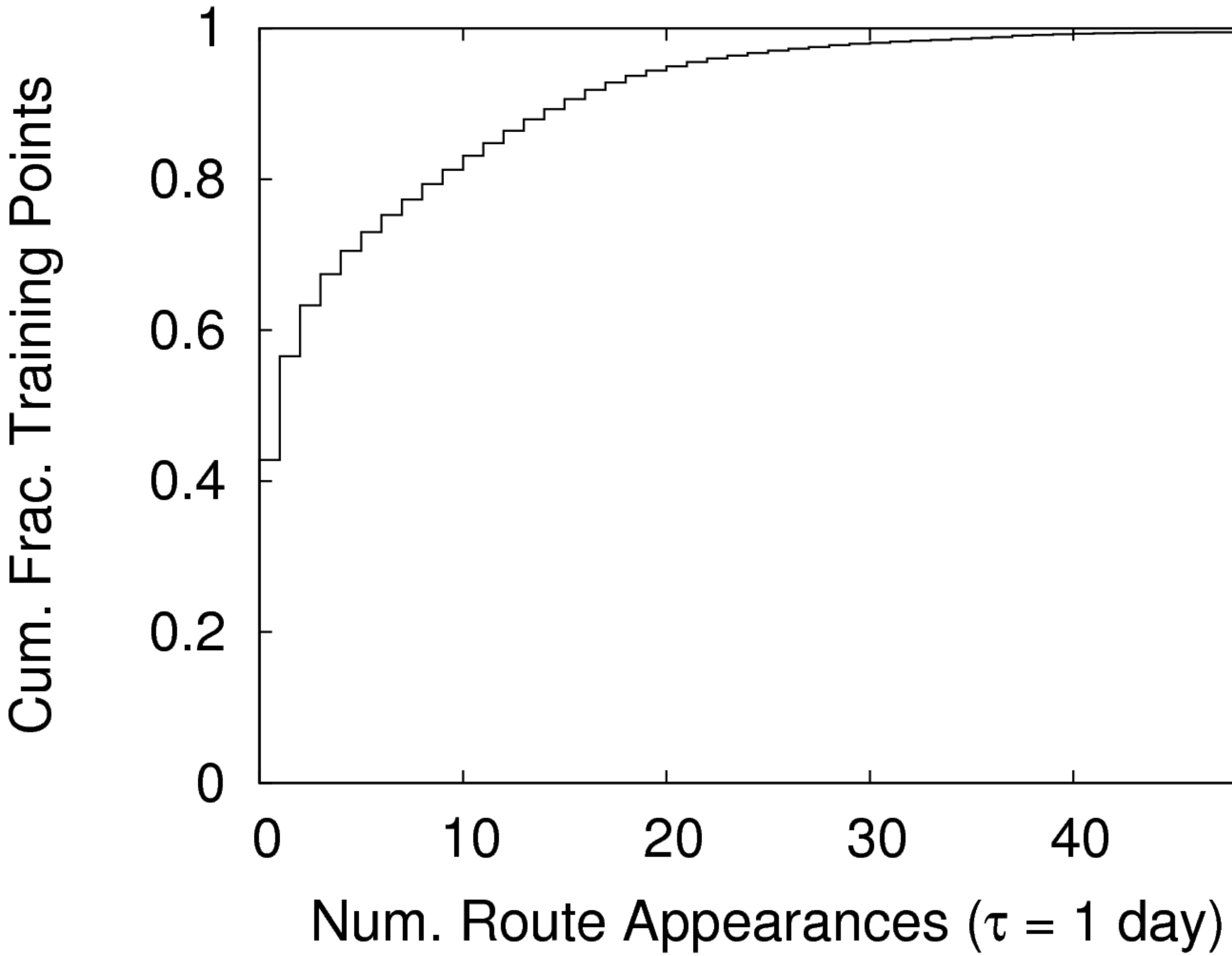
Impact of the number of features



Distribution of route prevalence



Distribution of number of occurrences of a route

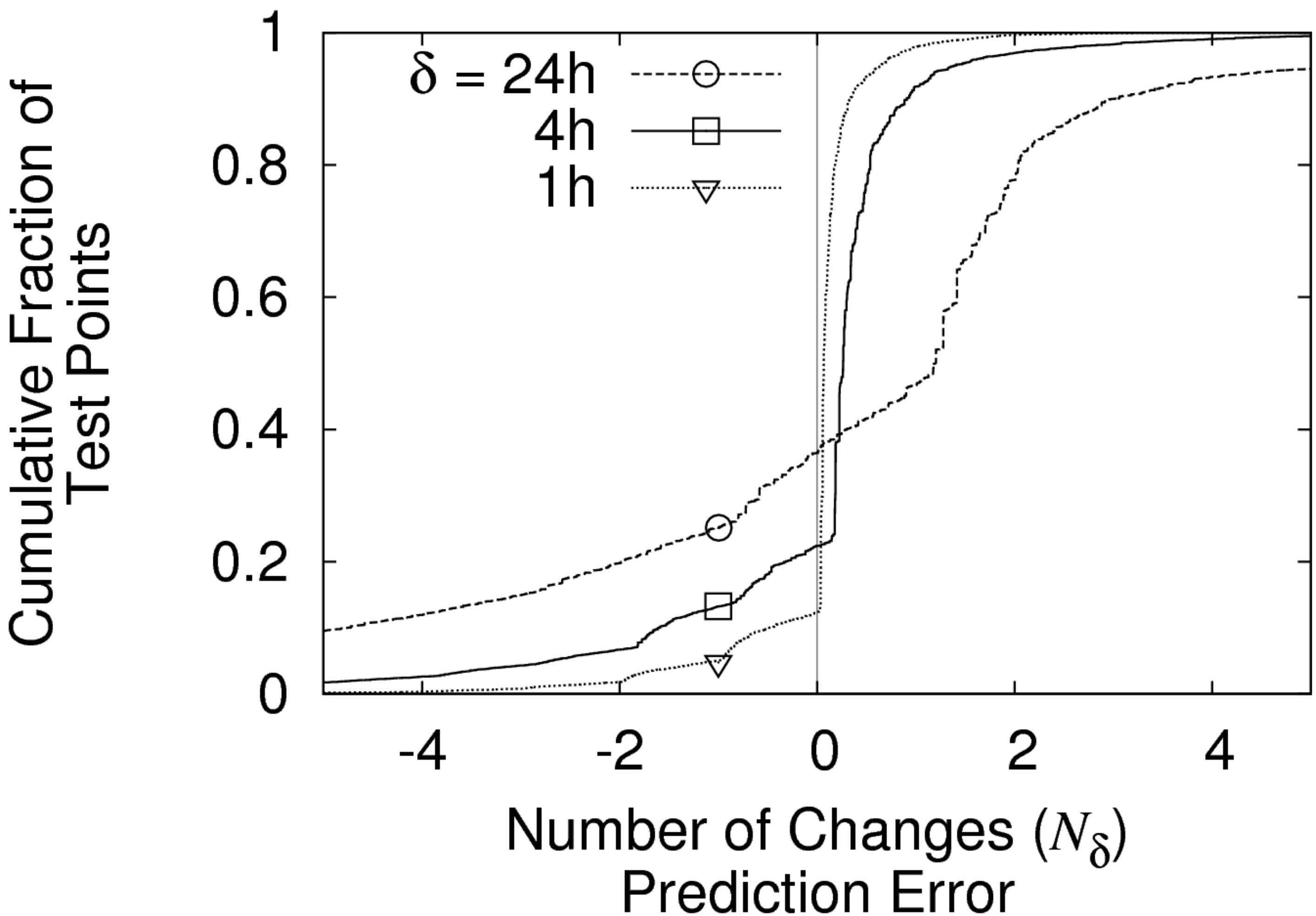


technicolor

Feature importance

Path feature	Importance
Prevalence (fraction of time active in the previous day)	1.0
Number of virtual path changes (1 day)	0.624
Number of previous occurrences of the current route (1 day)	0.216
Age	0.116
Times since most recent occurrences of the current route	≤ 0.072
Edit distance (last change)	0.015
Duration of the previous route	0.014
Standard deviation of route durations (1 day)	0.014
Length difference (last change)	0.012
All other features	≤ 0.010

Number of changes is hard to predict



technicolor

High accuracy in many experimental conditions

No need for big training sets

- 50,000 path changes are enough (we have 2,000,000+ changes in 5 weeks)
- 10 monitors almost as good as 70
- Infrequent measurements are still useful
  - Probing every 2 hours has 3% higher change error rate than probing every 5 minutes

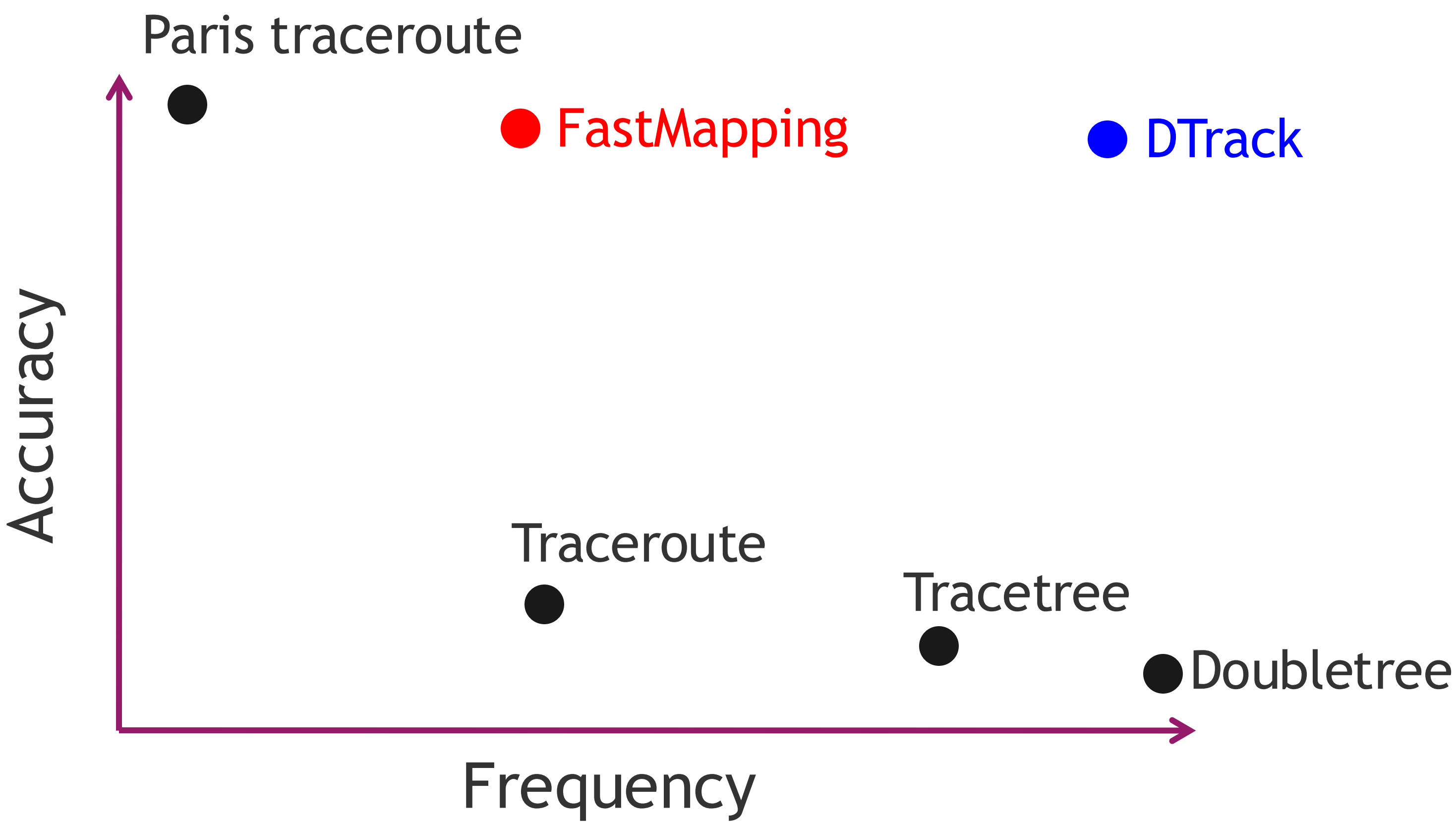
Few causes for path changes

- Network management and engineering (ISIS link weights)
- Hardware failures (power outage, link failure)
- Natural /external effects (fiber cut)

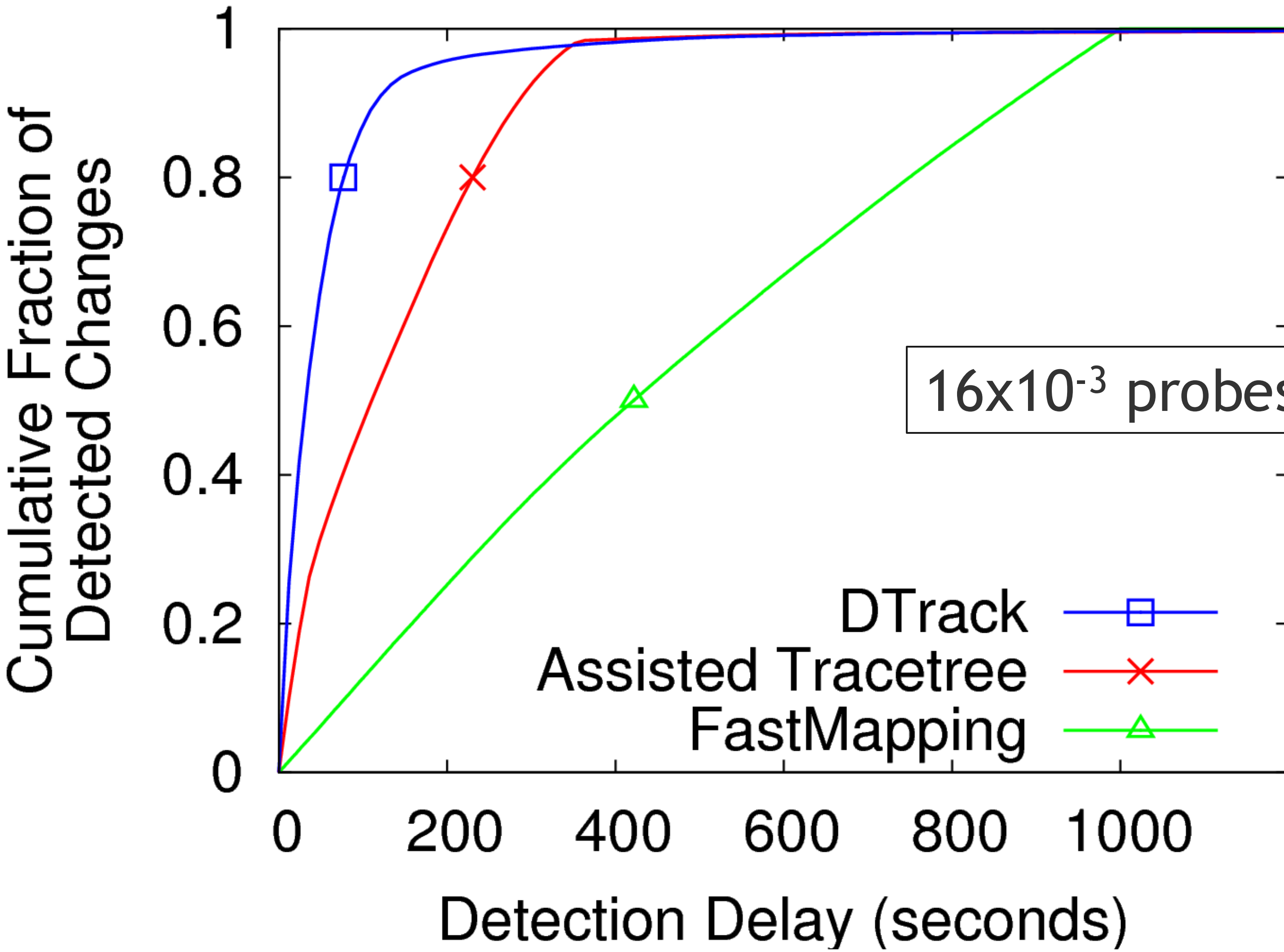
Training sets need only capture enough change diversity

DTrack

Overview of path measurement techniques



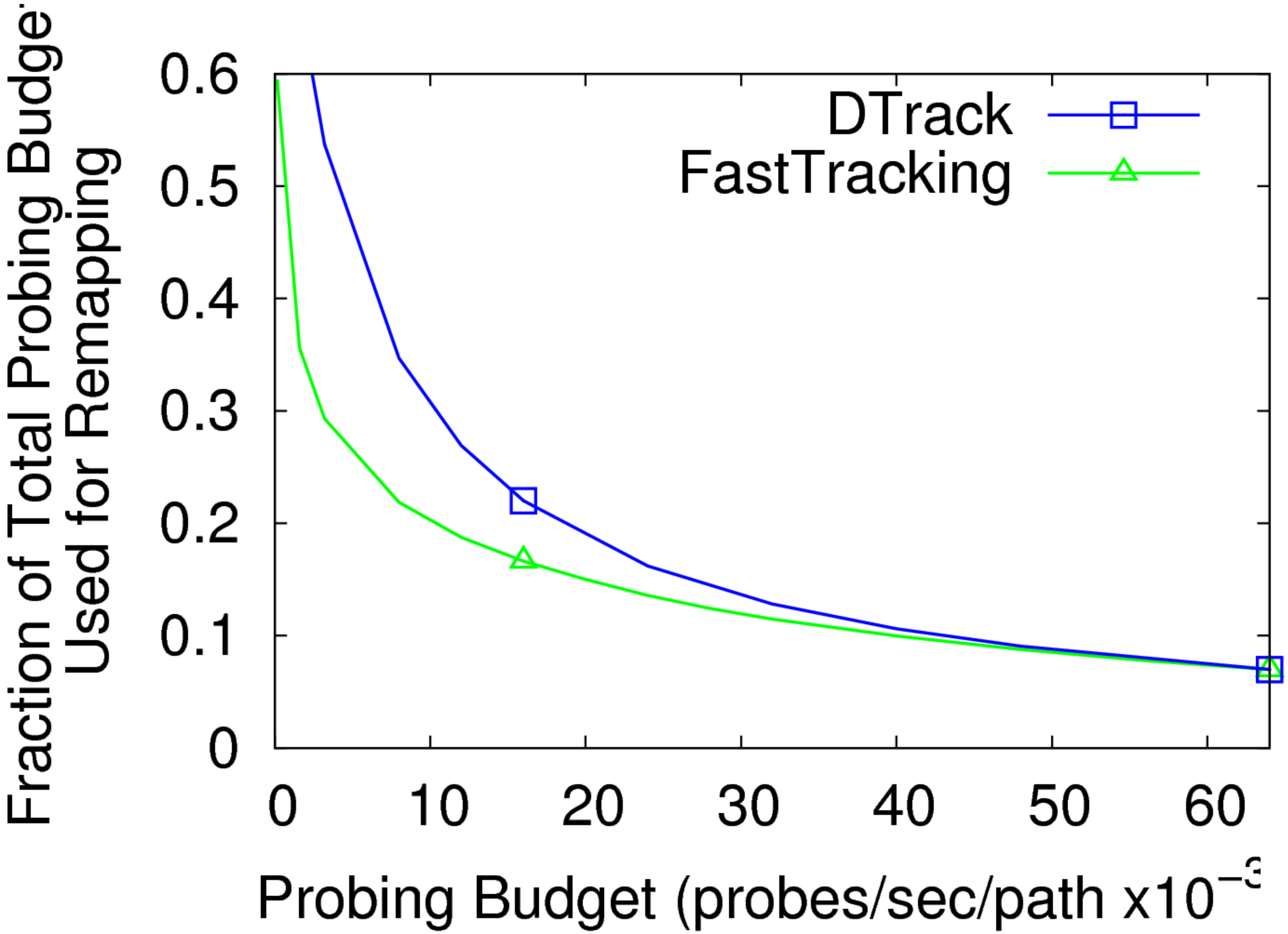
Change detection delay



16x10^{-3} probes/sec/path

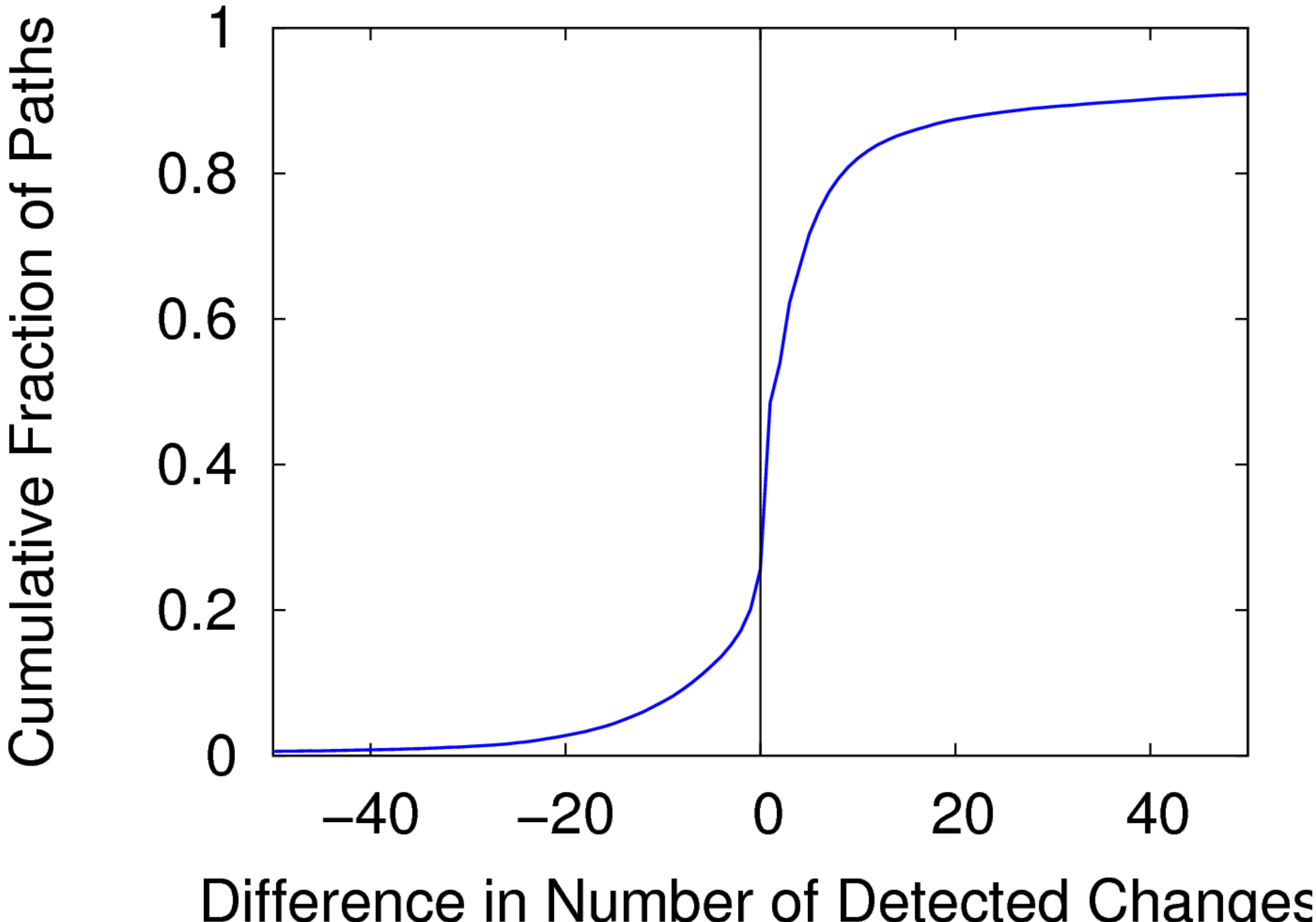
technicolor

Increase in remapping cost is small



technicolor

DTrack deployment confirms simulations



(DTrack – FastMapping)

Credits

Internet rendering on the first slide by The Opte Project.

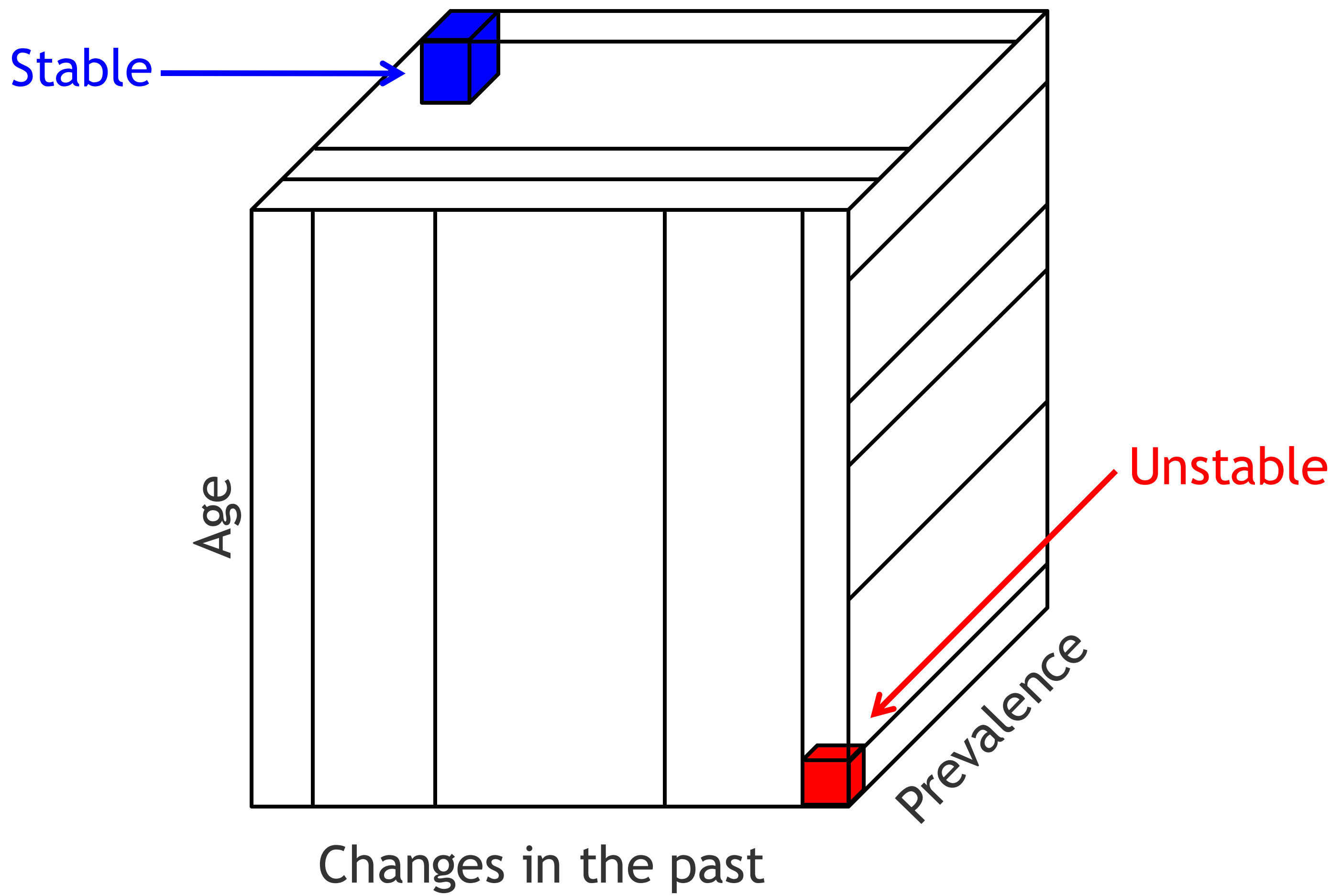
Server, router, and house cliparts from OpenClipArt.org.

NN4 predictor

Based on the nearest-neighbor scheme

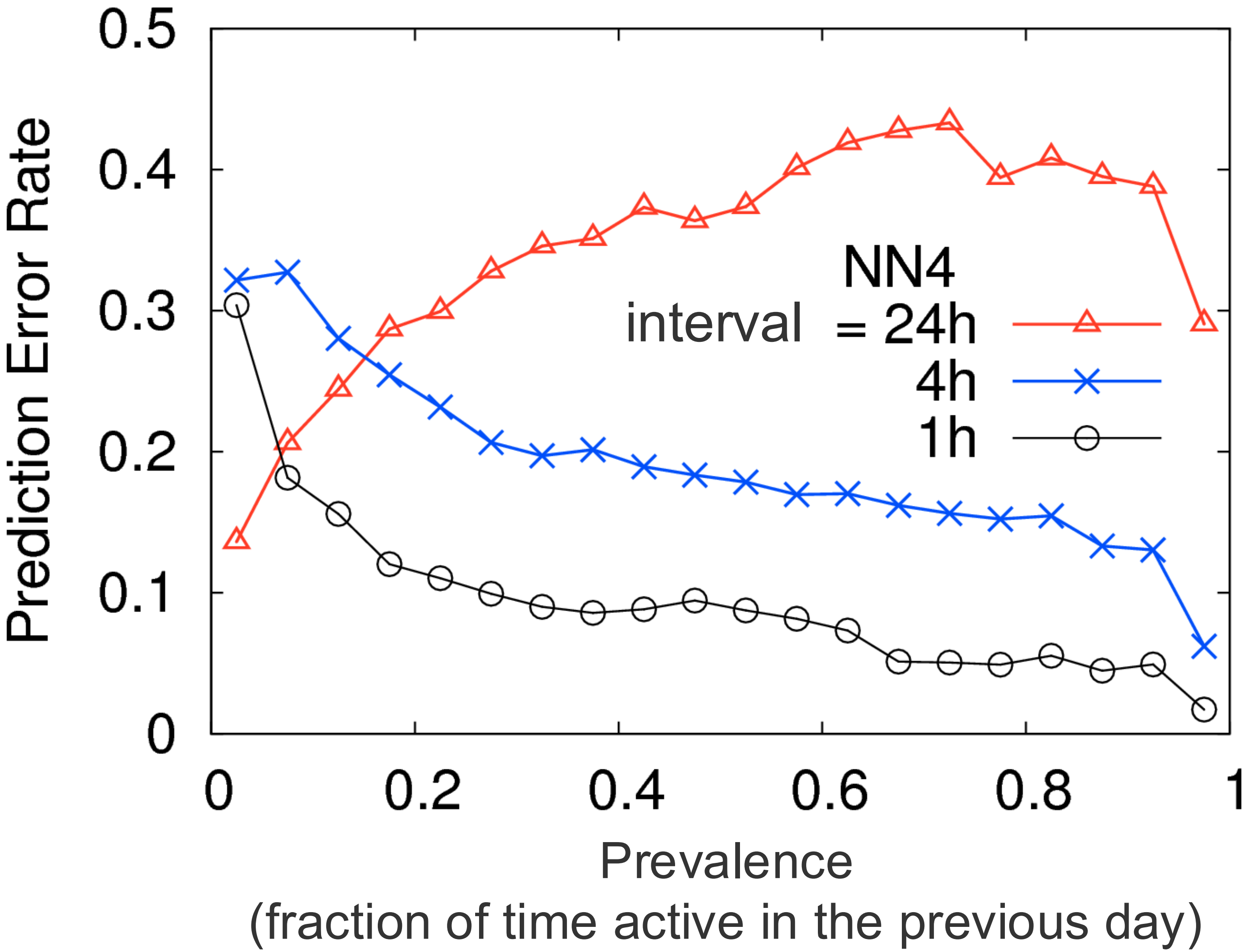
Compute neighbors by partitioning the path feature “state-space”

- Partition boundaries computed automatically



technicolor

NN4 performance over different intervals

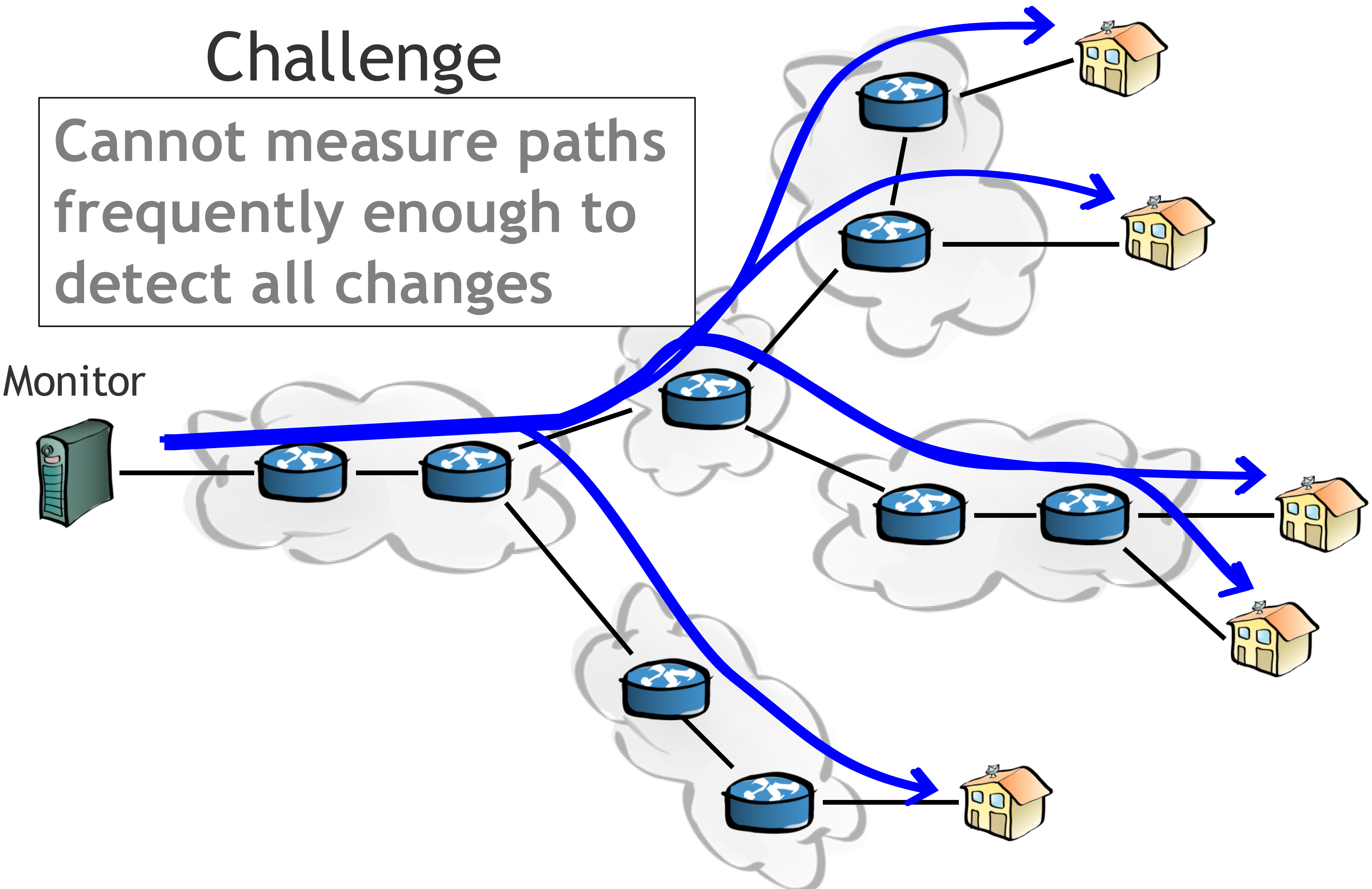


technicolor

Internet path measurements

Challenge

Cannot measure paths frequently enough to detect all changes



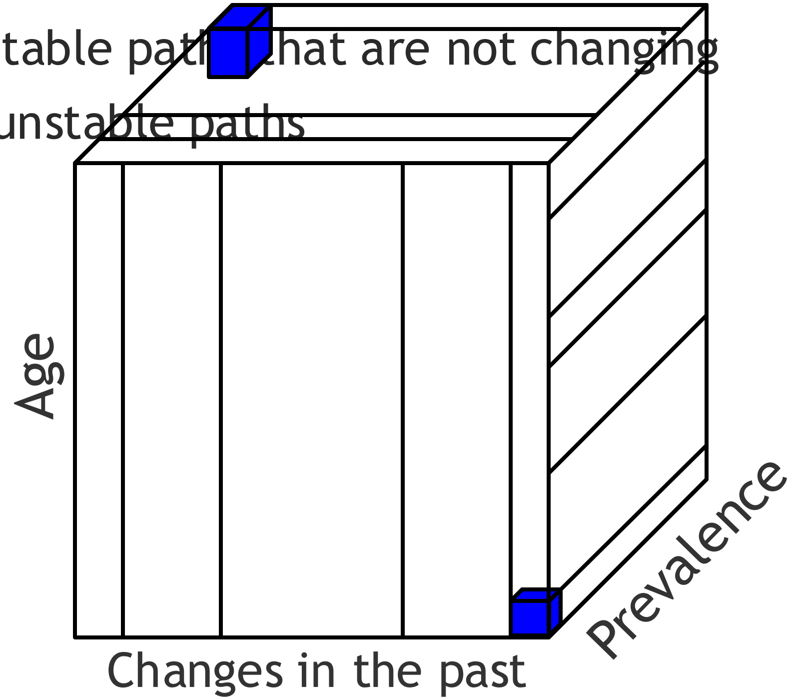
Limitations of current techniques

Internet paths are mostly stable

No need to remeasure paths that have not changed

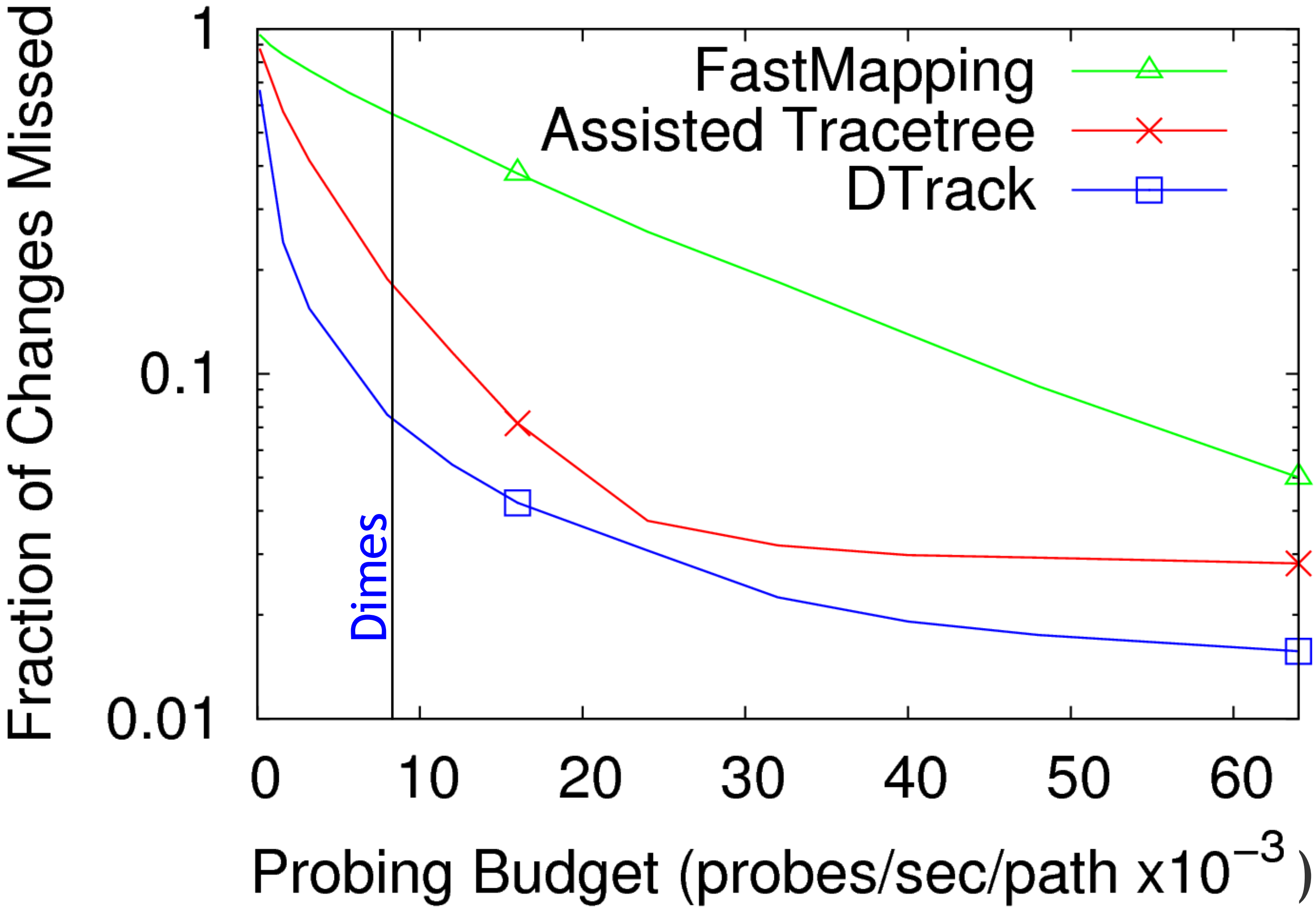
Current approach is to measure paths periodically

- Wastes probes on stable paths that are not changing
- Misses changes on unstable paths



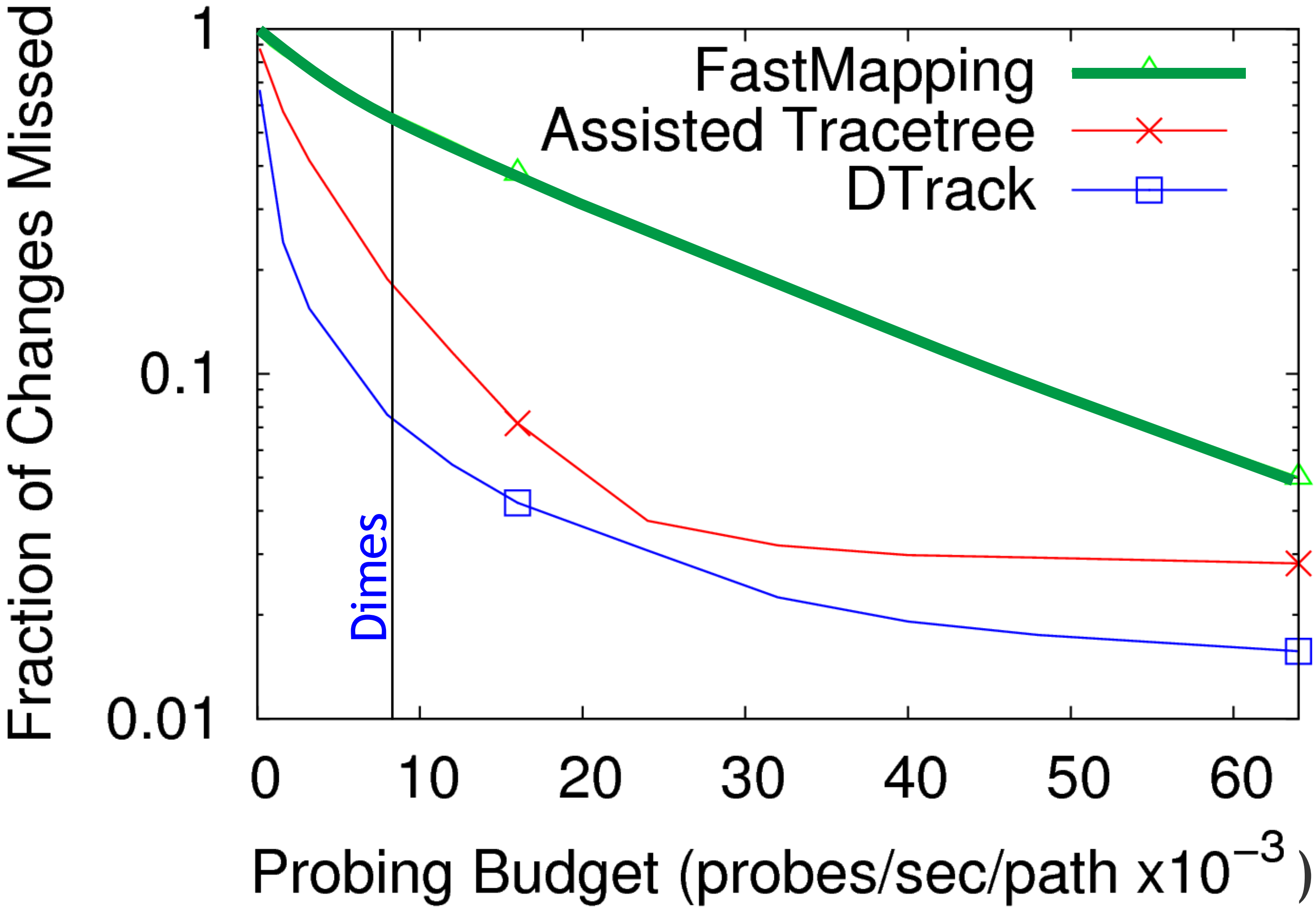
technicolor

Number of changes missed



technicolor

Number of changes missed



technicolor