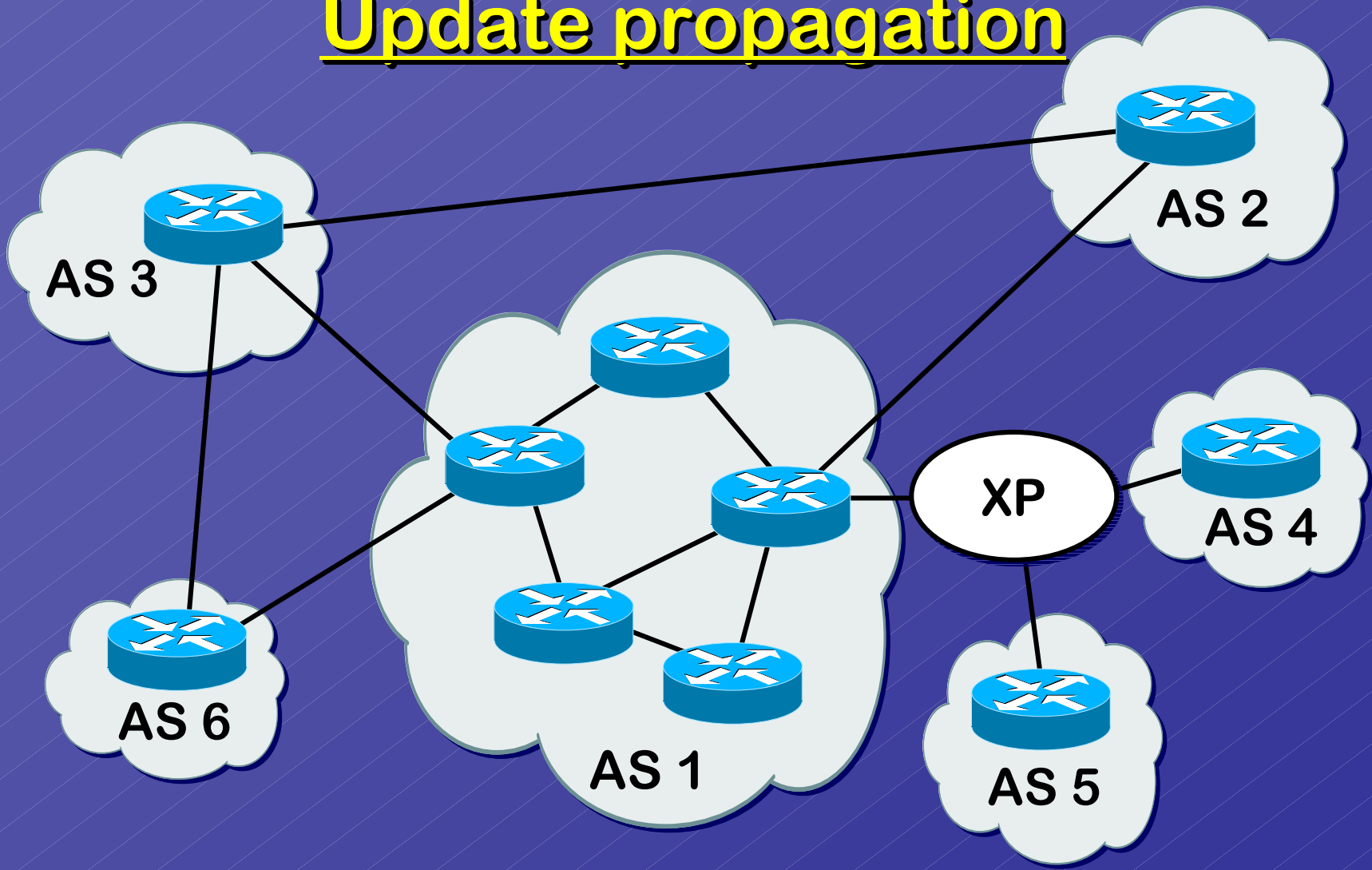
On BGP Convergence and Scalability

Olaf Maennel, Anja Feldmann
Technical University Munich, Germany

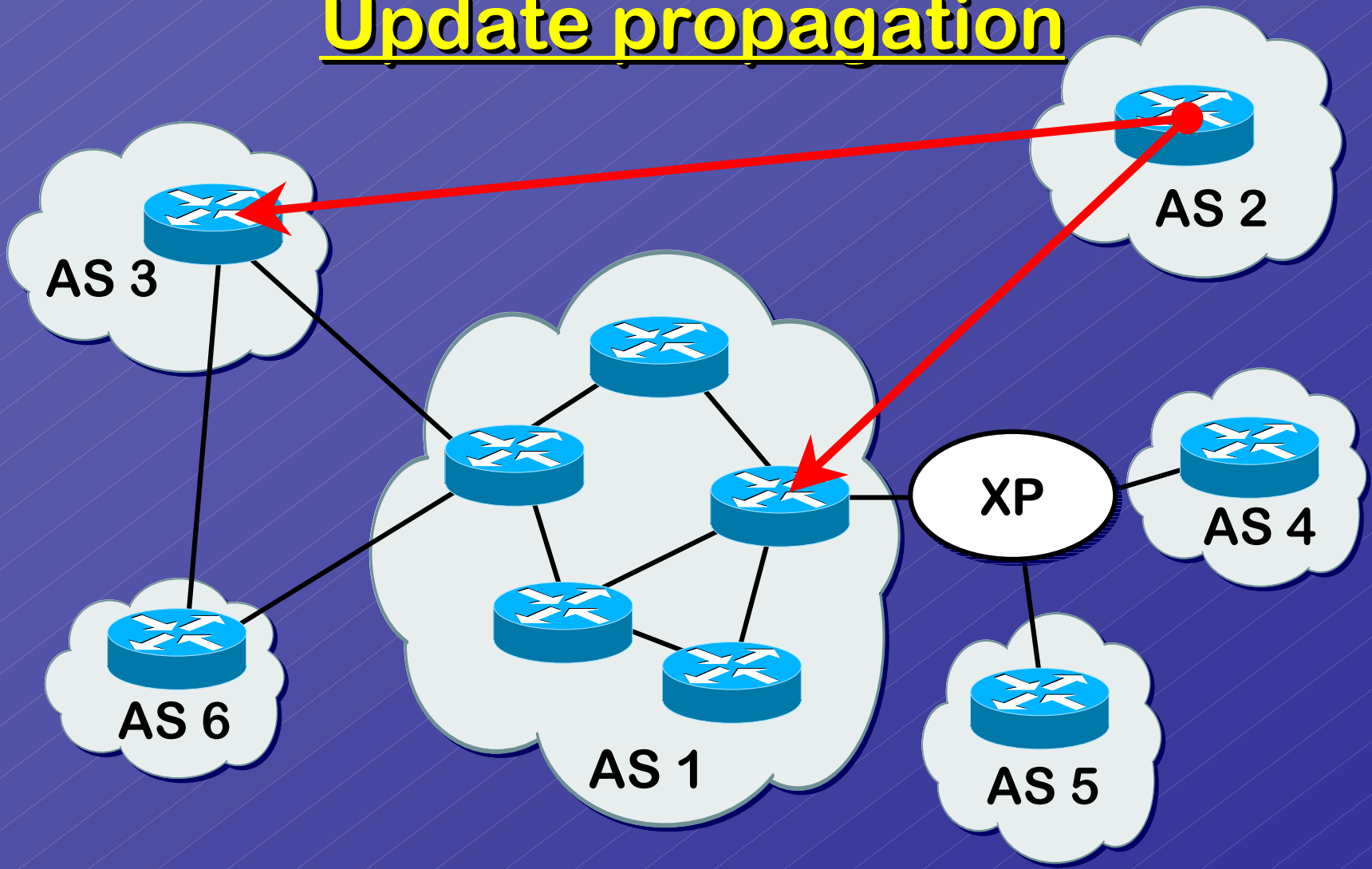
Outline

- Motivation
- Convergence
  - properties of updates bursts
- Scalability
  - a workload model for a BGP test-bed
- Summary

Update propagation

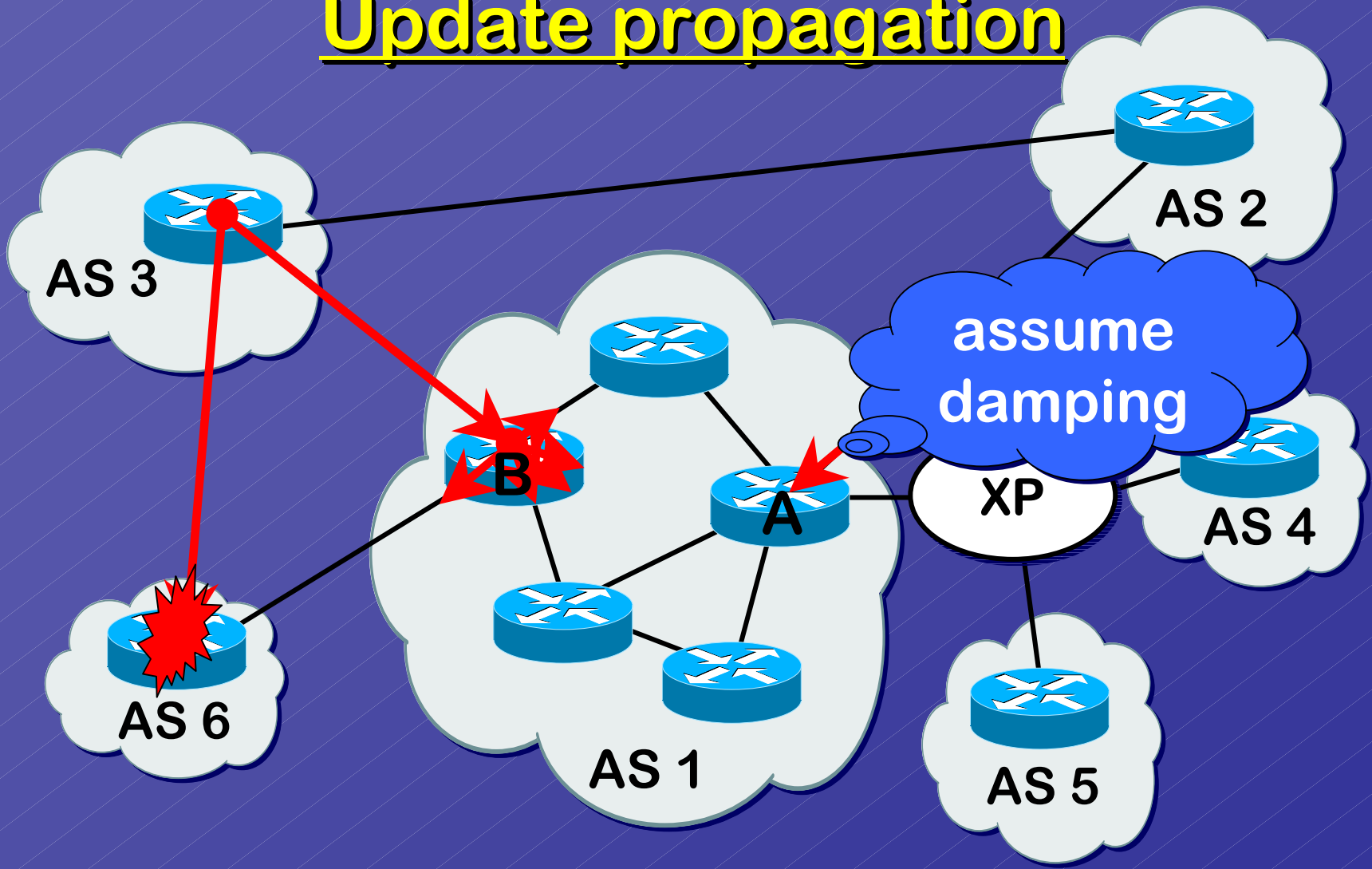


Update propagation



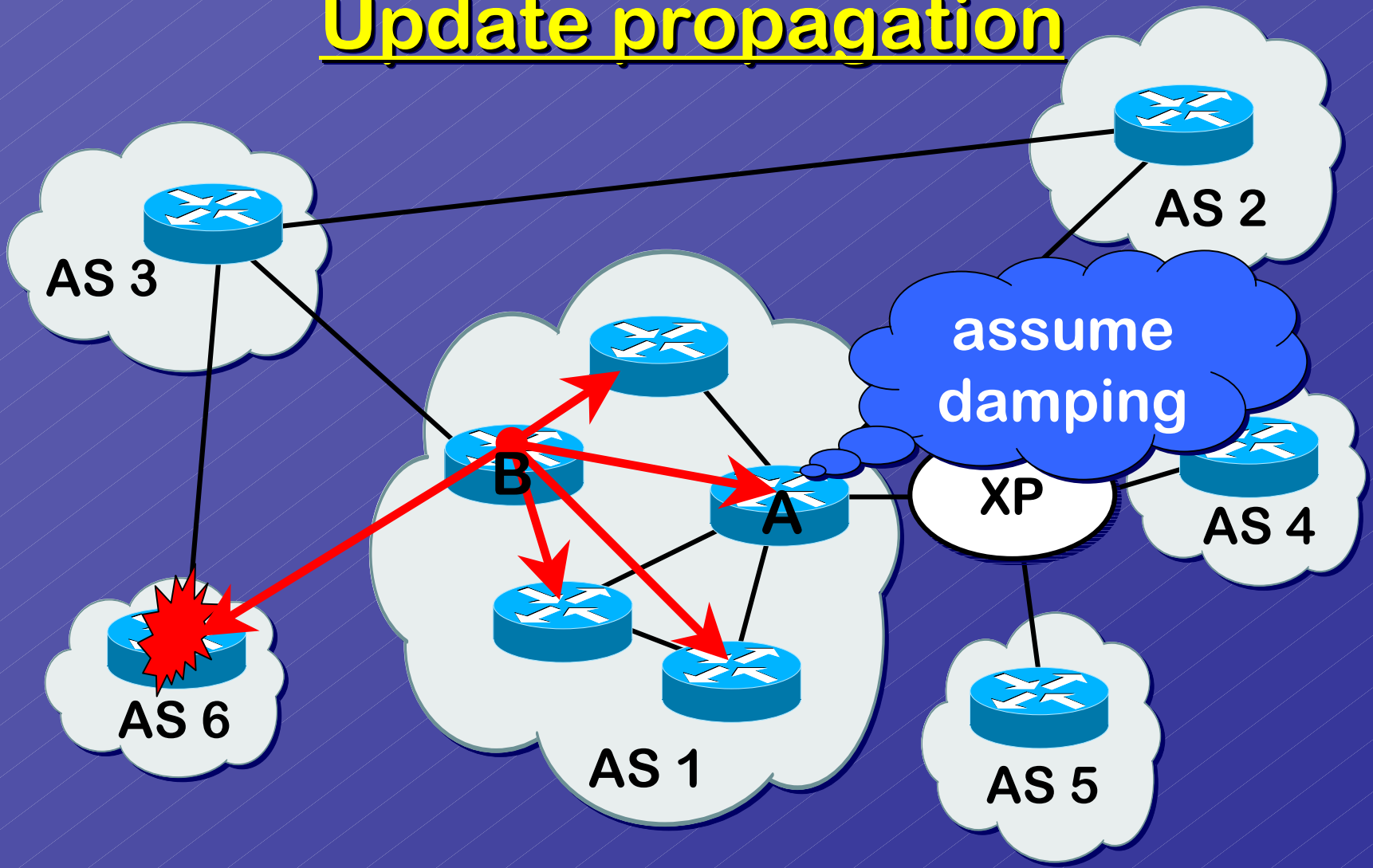
AS 2 issues an update

Update propagation



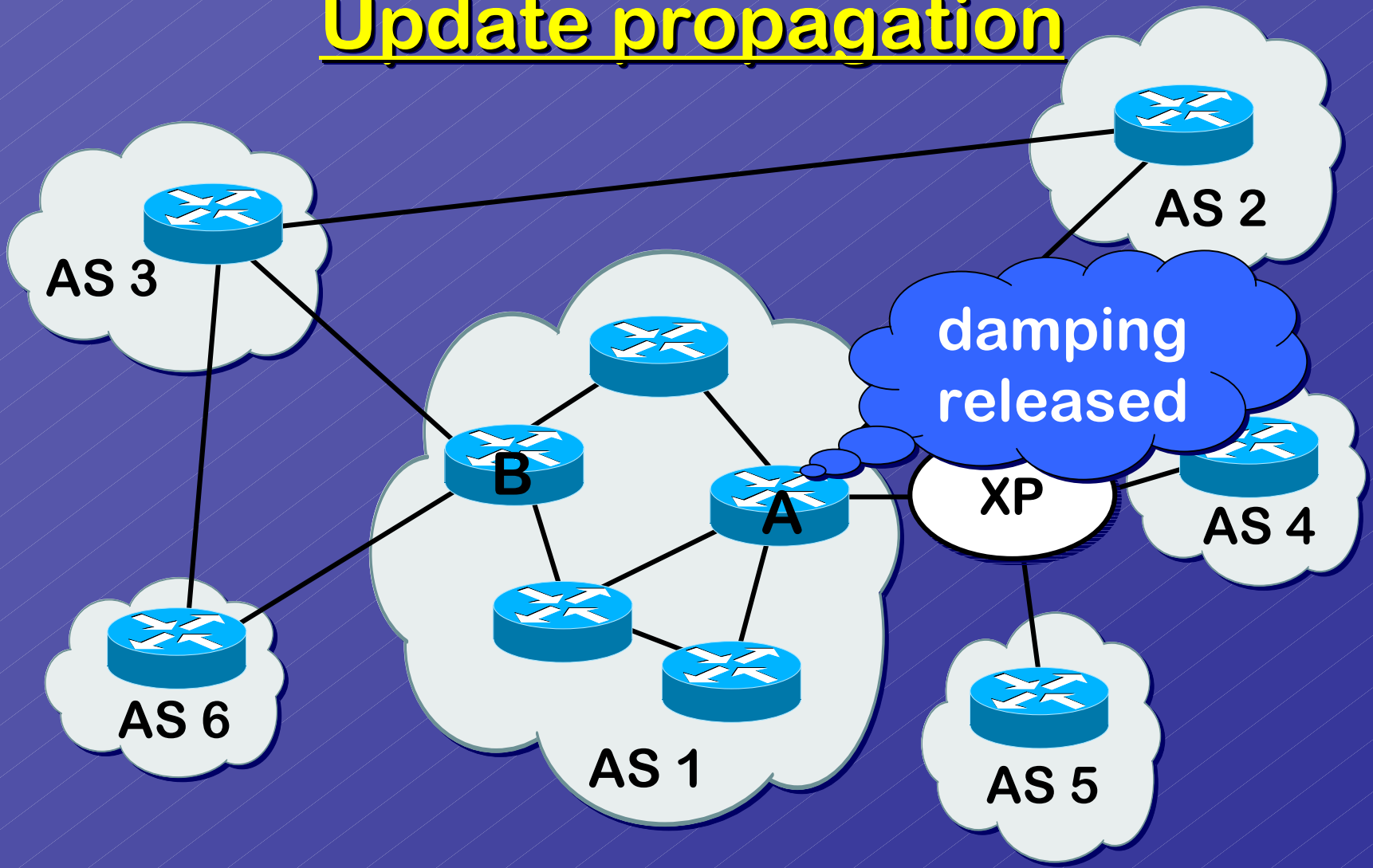
AS 6 may reach AS 2 via AS 3

Update propagation



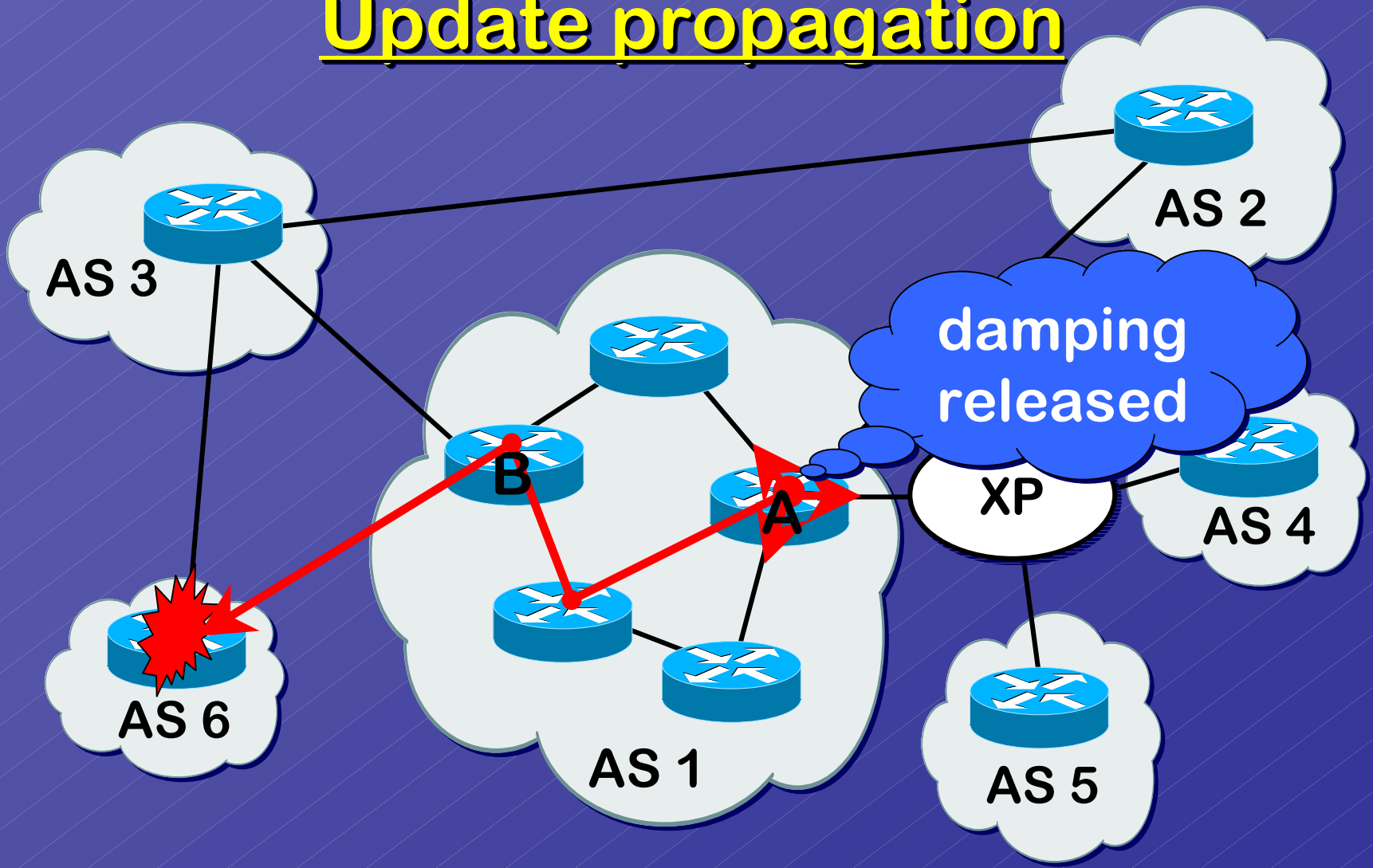
AS 6 may reach AS 2 via AS 3

Update propagation



AS 6 may reach AS 2 via AS 1

Update propagation

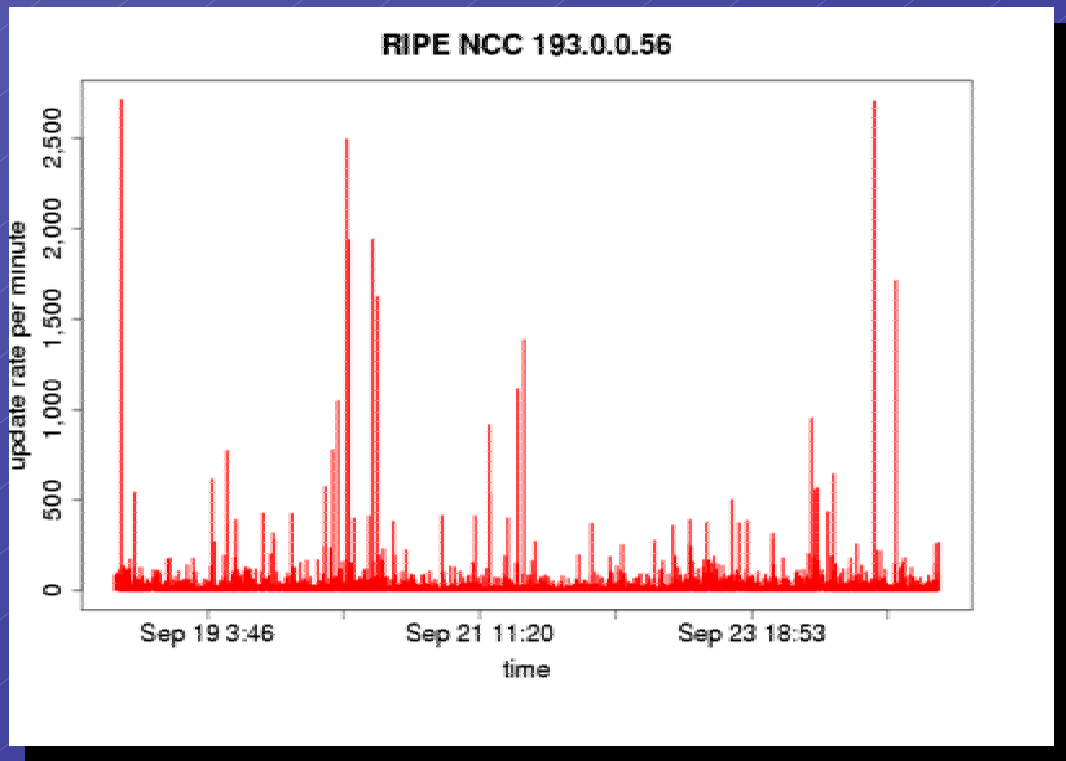


AS 6 may prefer to reach AS 2 via AS 1

Example data sets

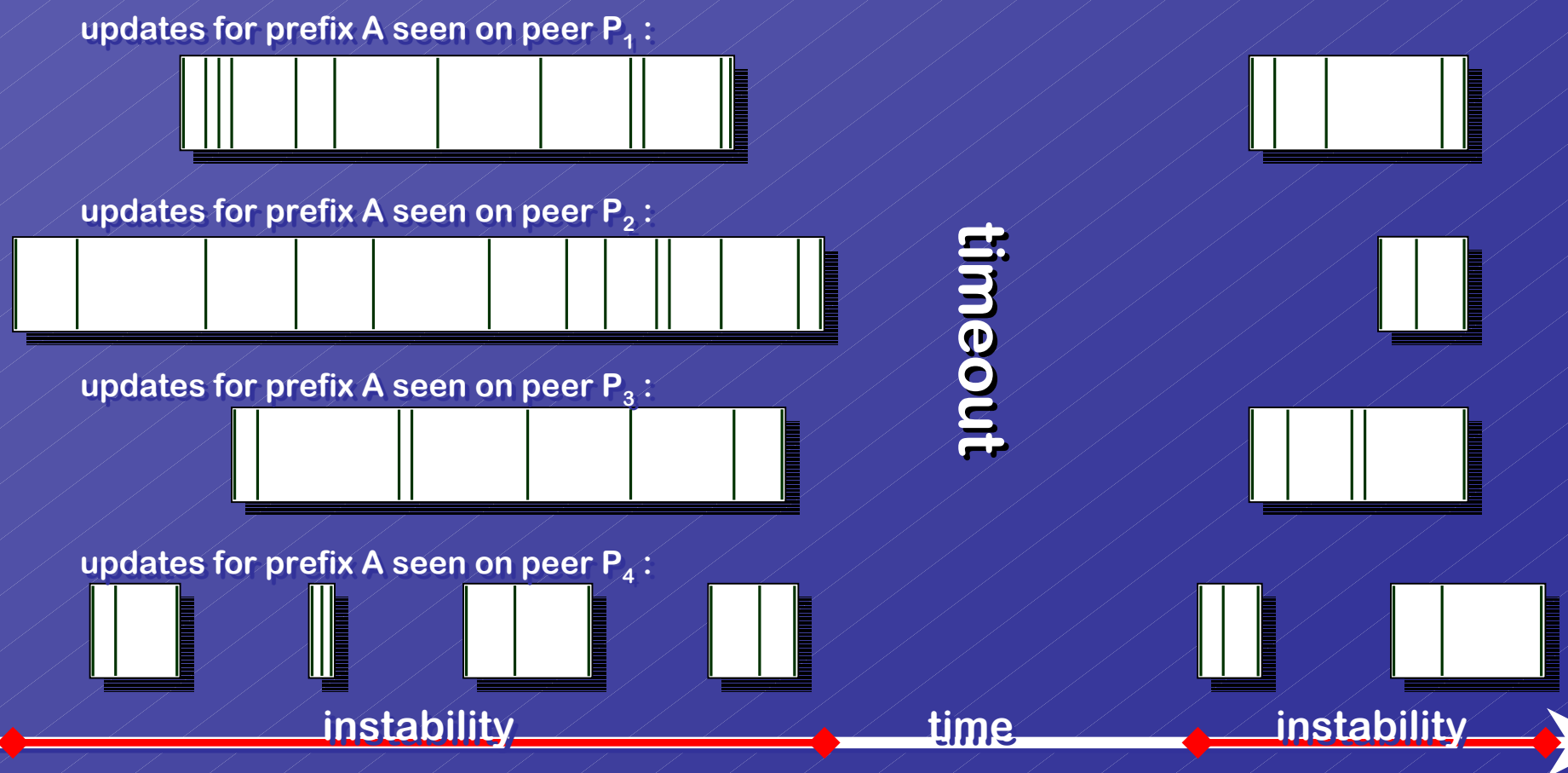
September 18-25, 2002:

- **RIPE RIS project**
  - 95 peers (all RRC*)
  - updates for 124,977 different prefixes
- **Routeviews**
  - 19 peers + 2 SaarGate
  - updates for 124,662 different prefixes



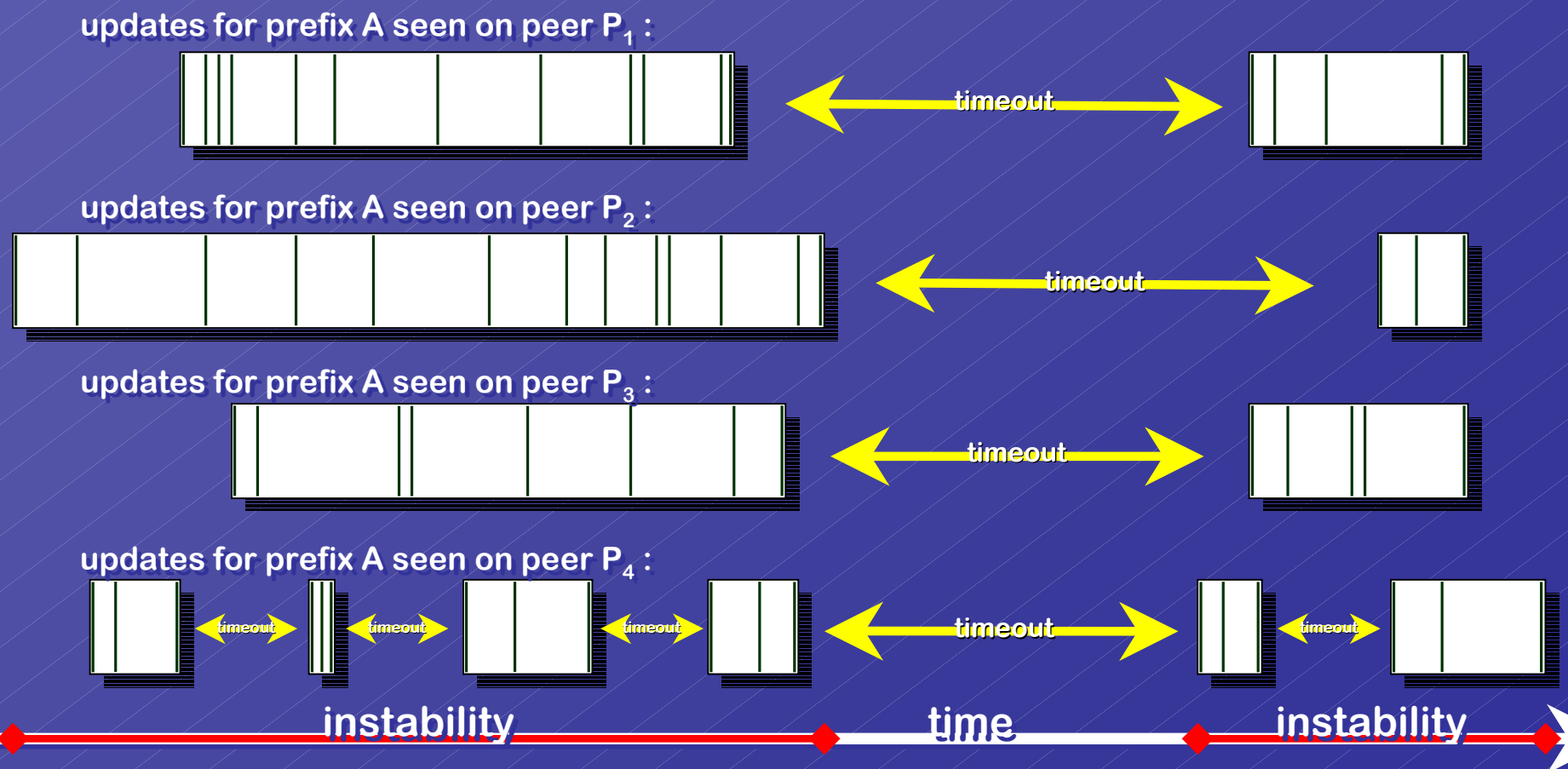
On BGP convergence

"Updates bursts" consists of several updates:
- same prefix / peer
- short time window



On BGP convergence

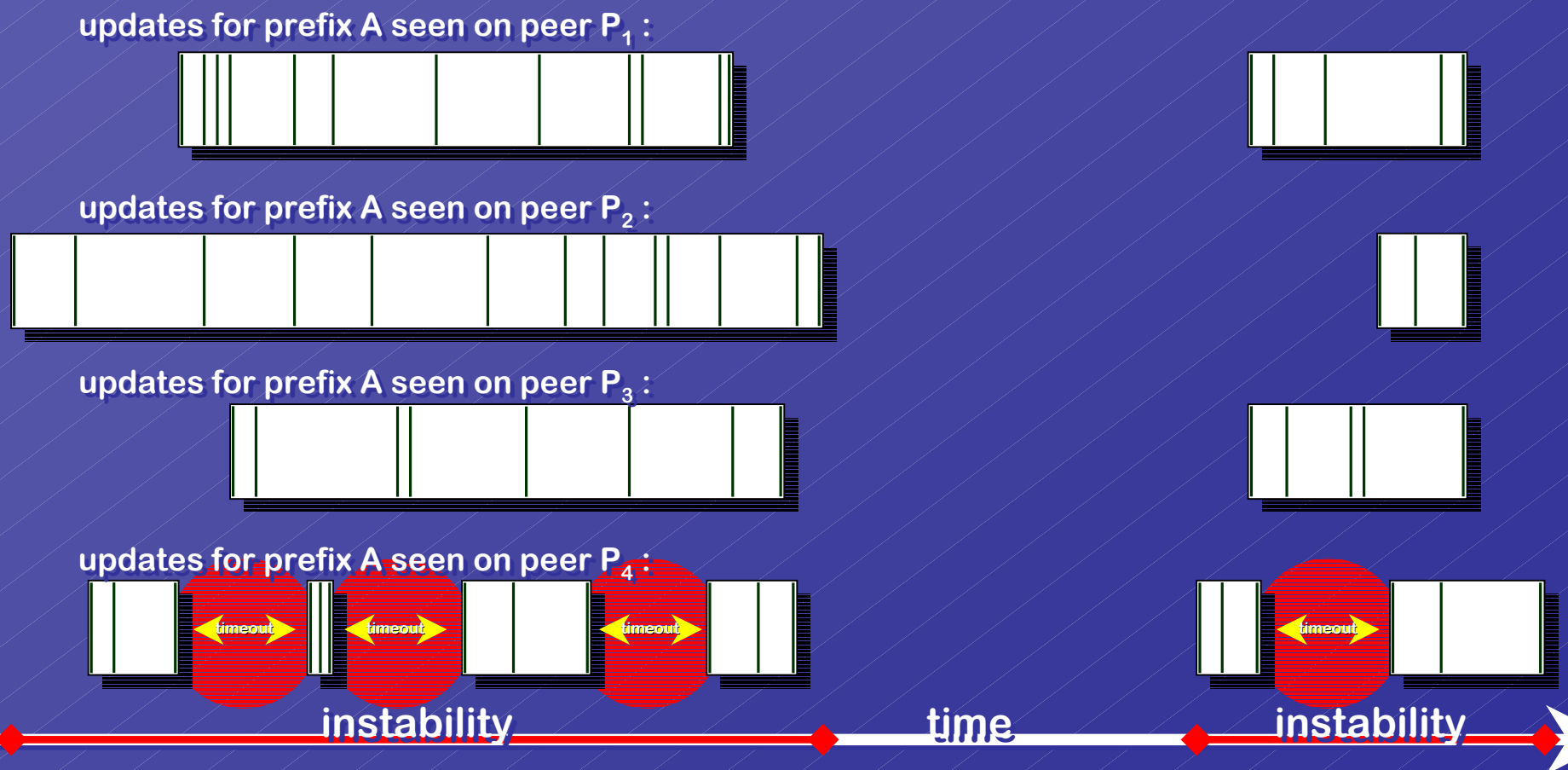
Prefixes are stable for at least *"timeout"* seconds.



On BGP convergence

It's only a heuristic!

- timeout too small: can't capture all effects (e.g., damping)
- timeout too large: combine several instabilities in one burst

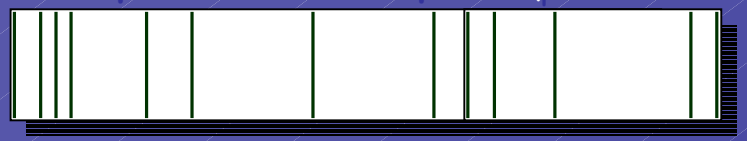


On BGP convergence

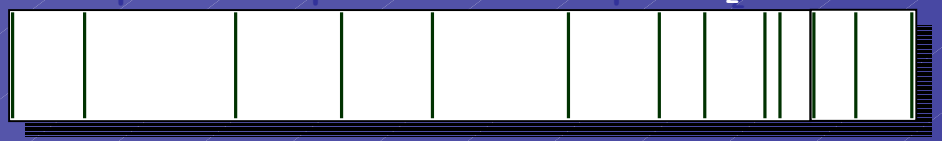
It's only a heuristic!

- timeout too small: can't capture all effects (e.g., damping)
- timeout too large: combine several instabilities in one burst

updates for prefix A seen on peer P_1 :

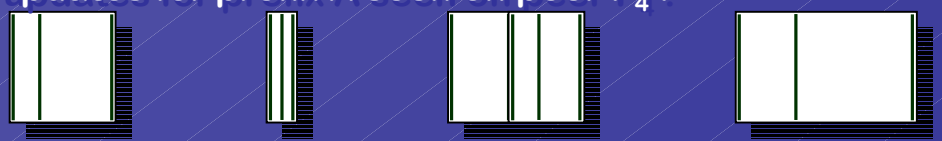


updates for prefix A seen on peer P_2 :



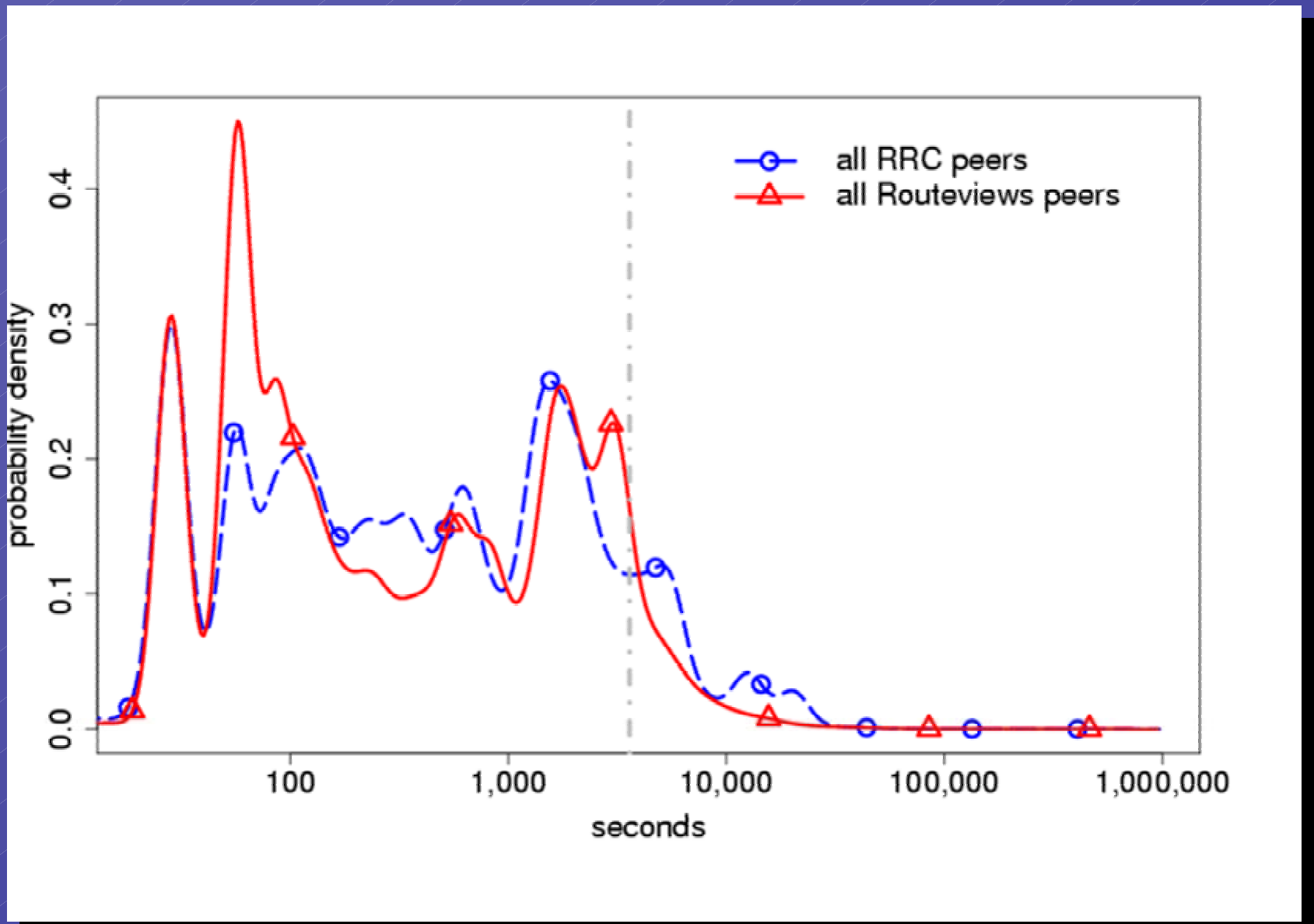
updates for prefix A seen on peer P_3 :

updates for prefix A seen on peer P_4 :

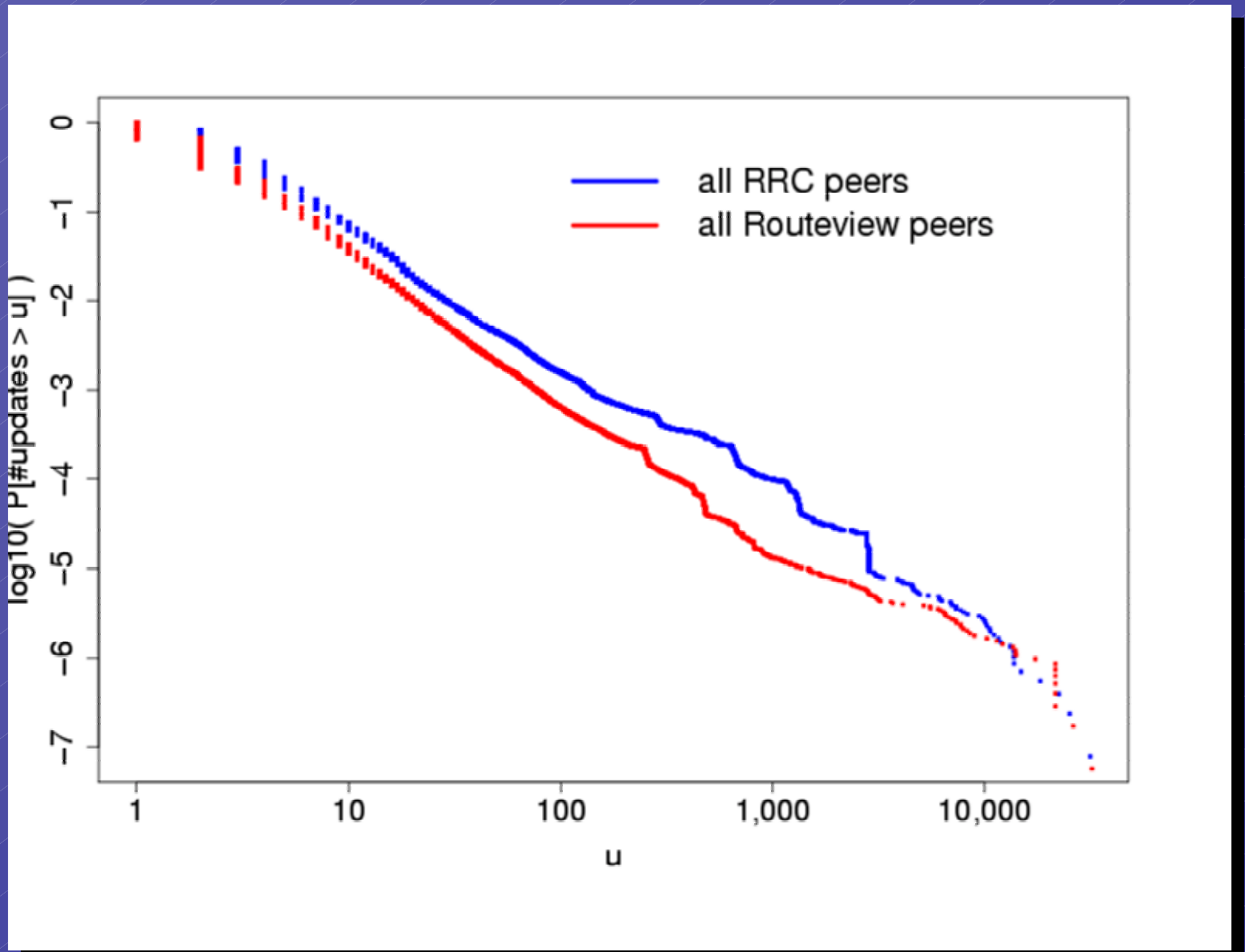


failure repair time

Burst duration

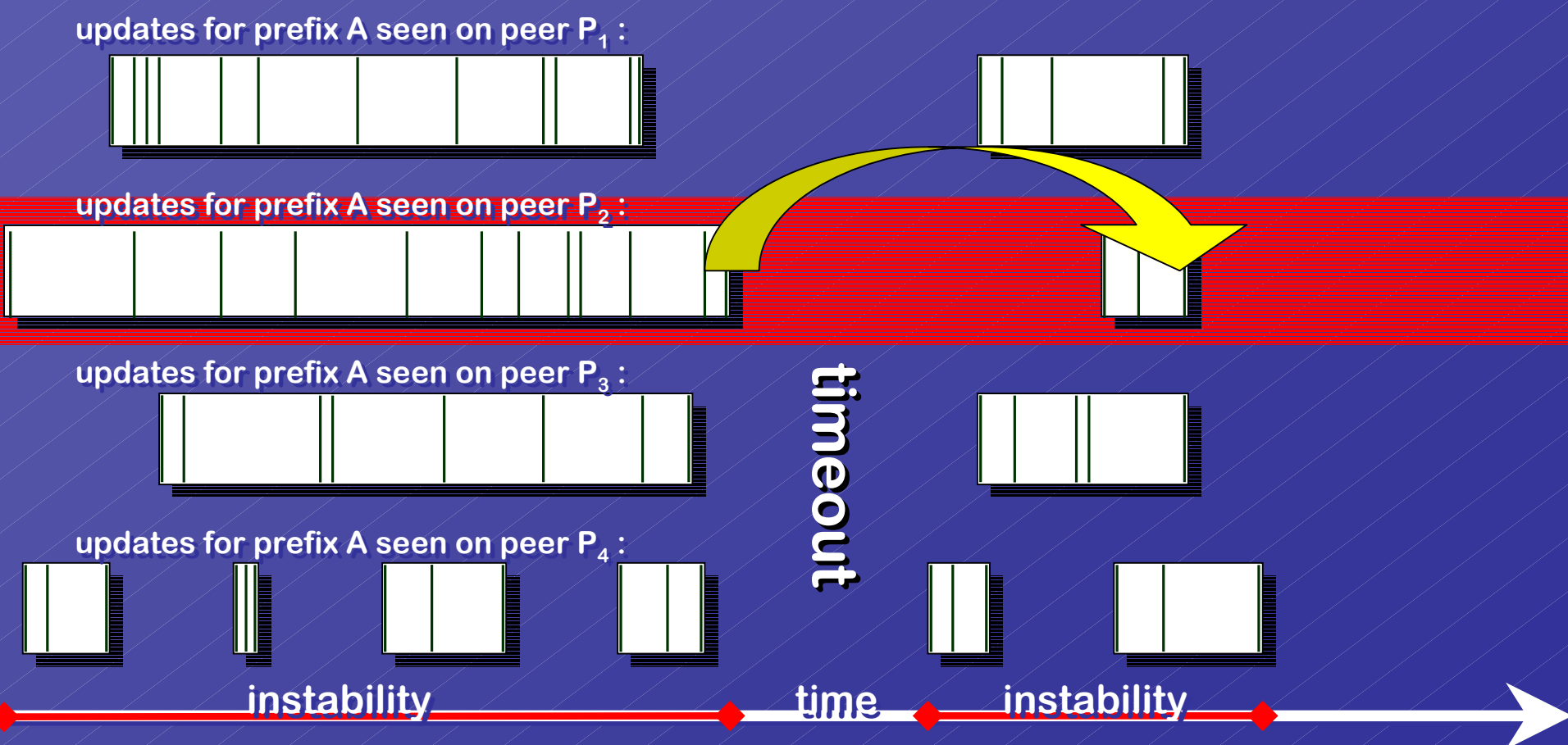


Updates in burst

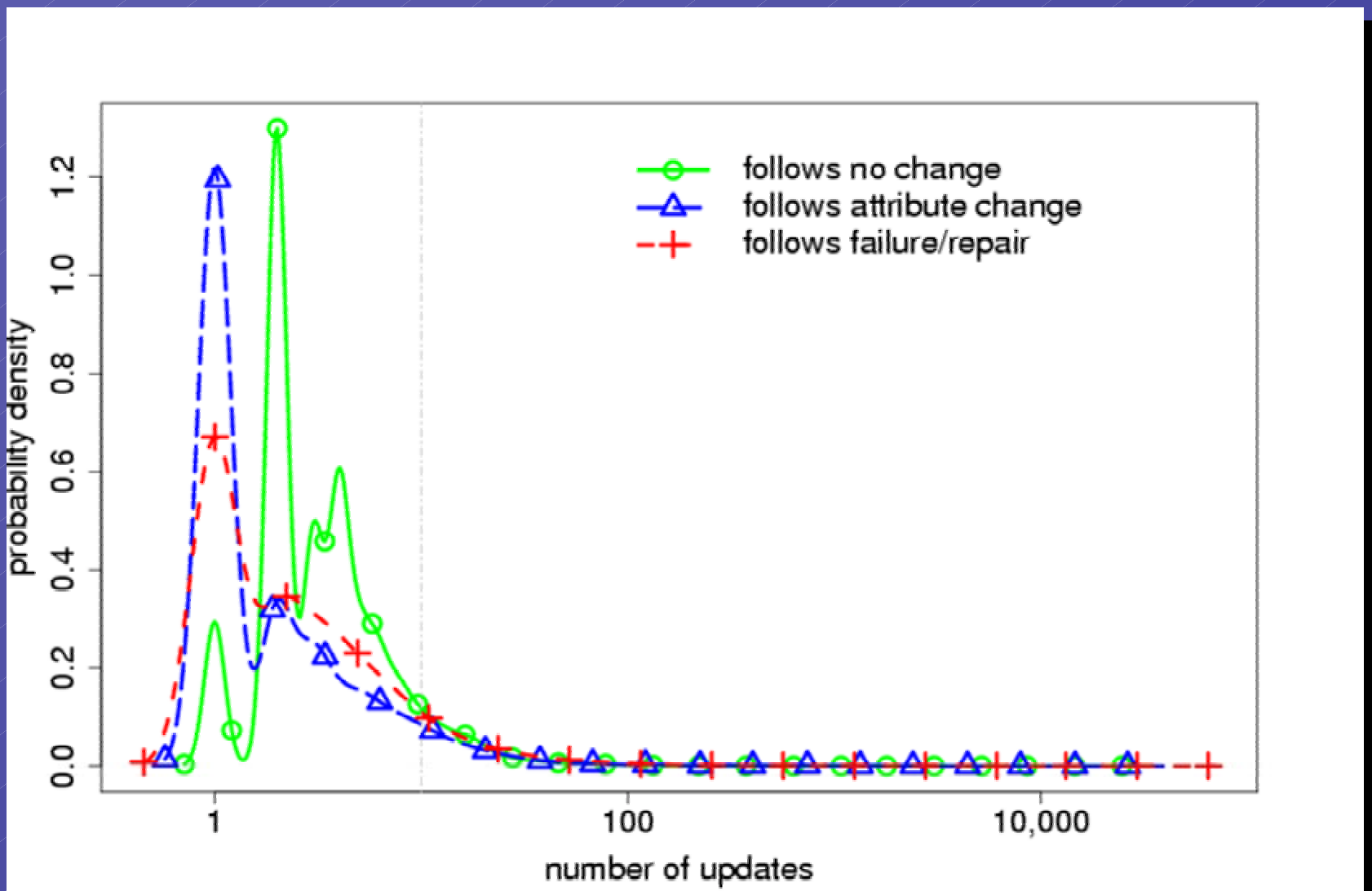


On BGP convergence

Last / stable updates are important (= result of "best path" selection process)
Question: compare what has changed!



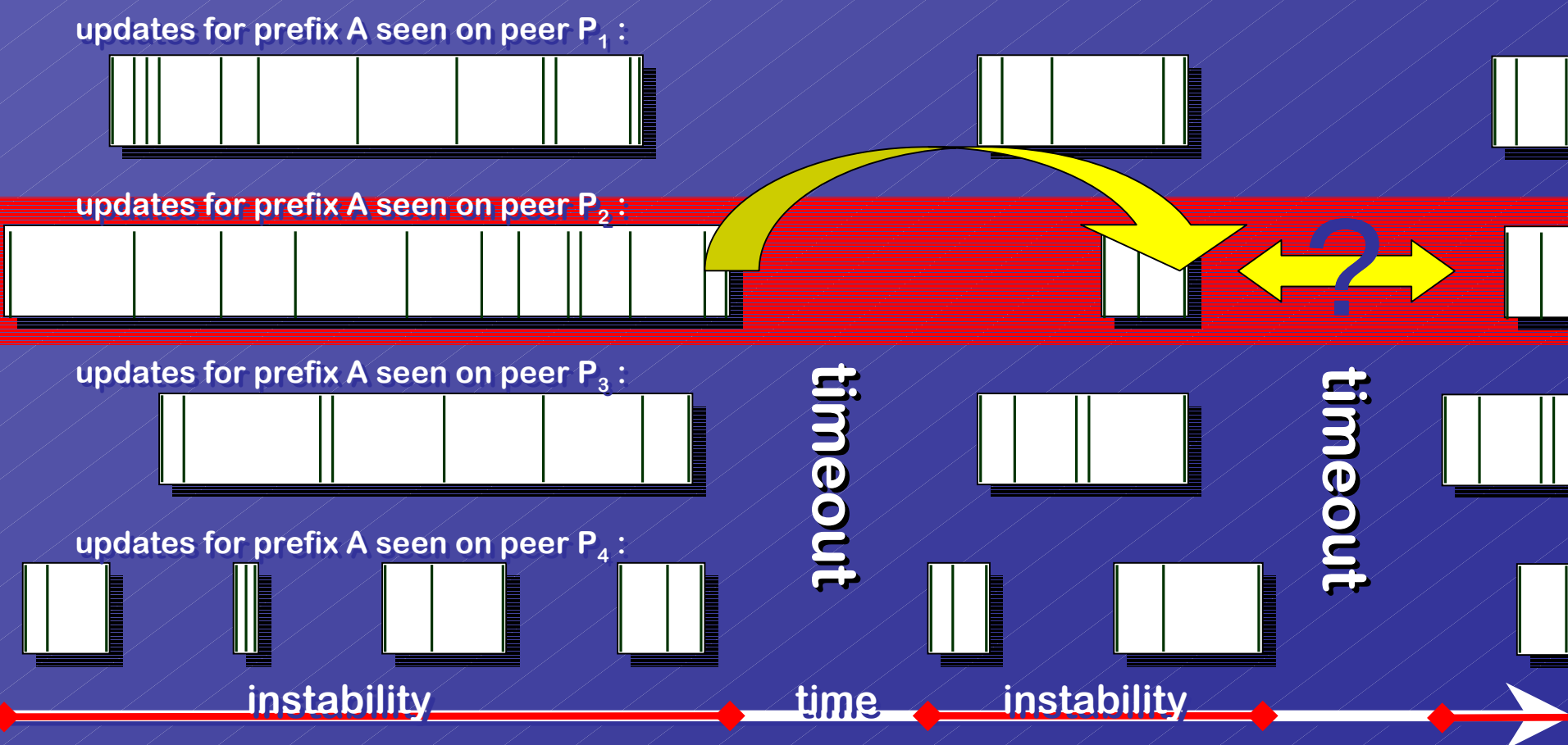
Updates in burst



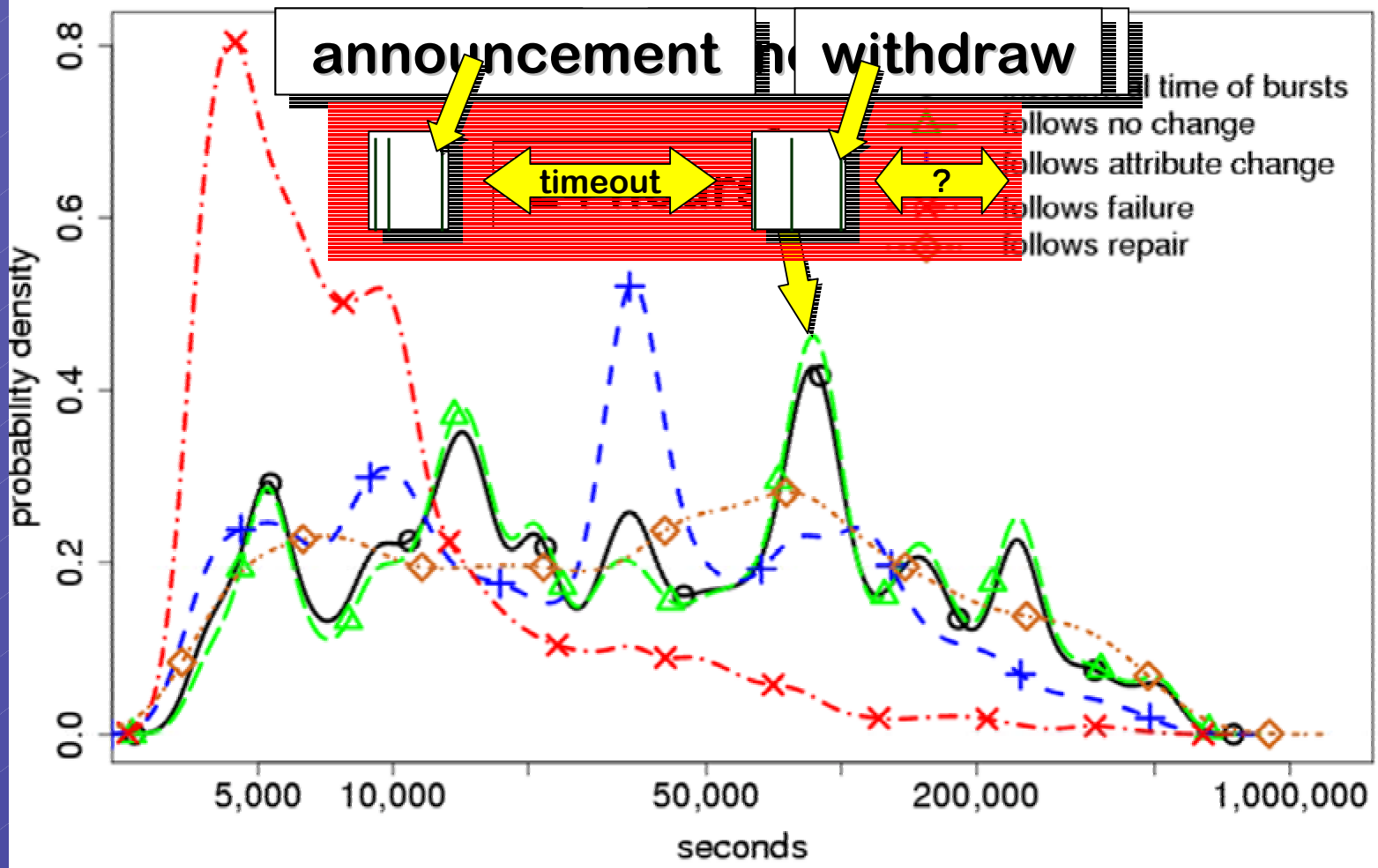
Datasets: all RRC peers; Time: 09/18-09/25/2002

On BGP convergence

Last / stable updates are important (= result of "best path" selection process)
Question: when does the next burst starts?



Interarrival time of burst

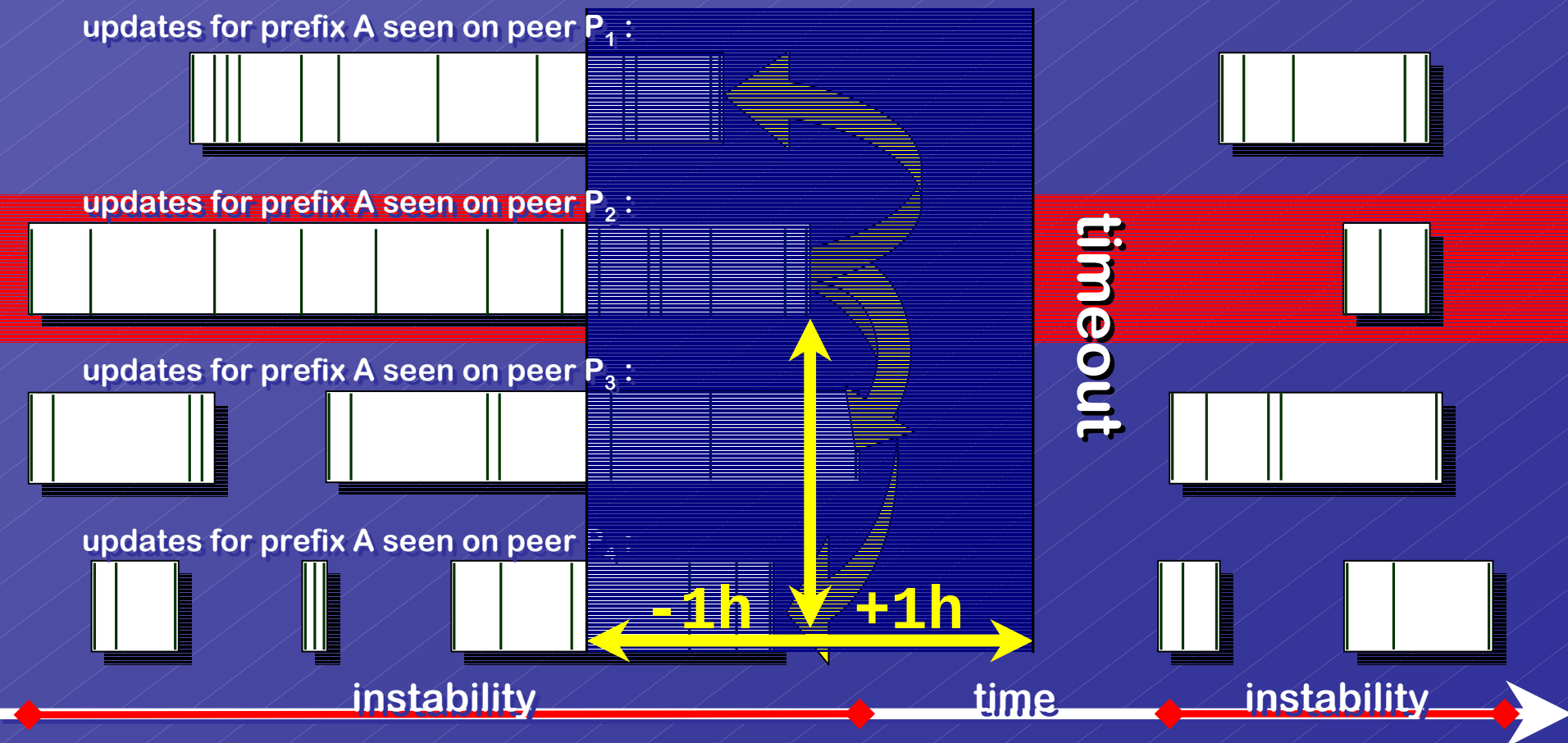


Datasets: Routeviews, SaarGate; Time: 09/18-09/25/2002

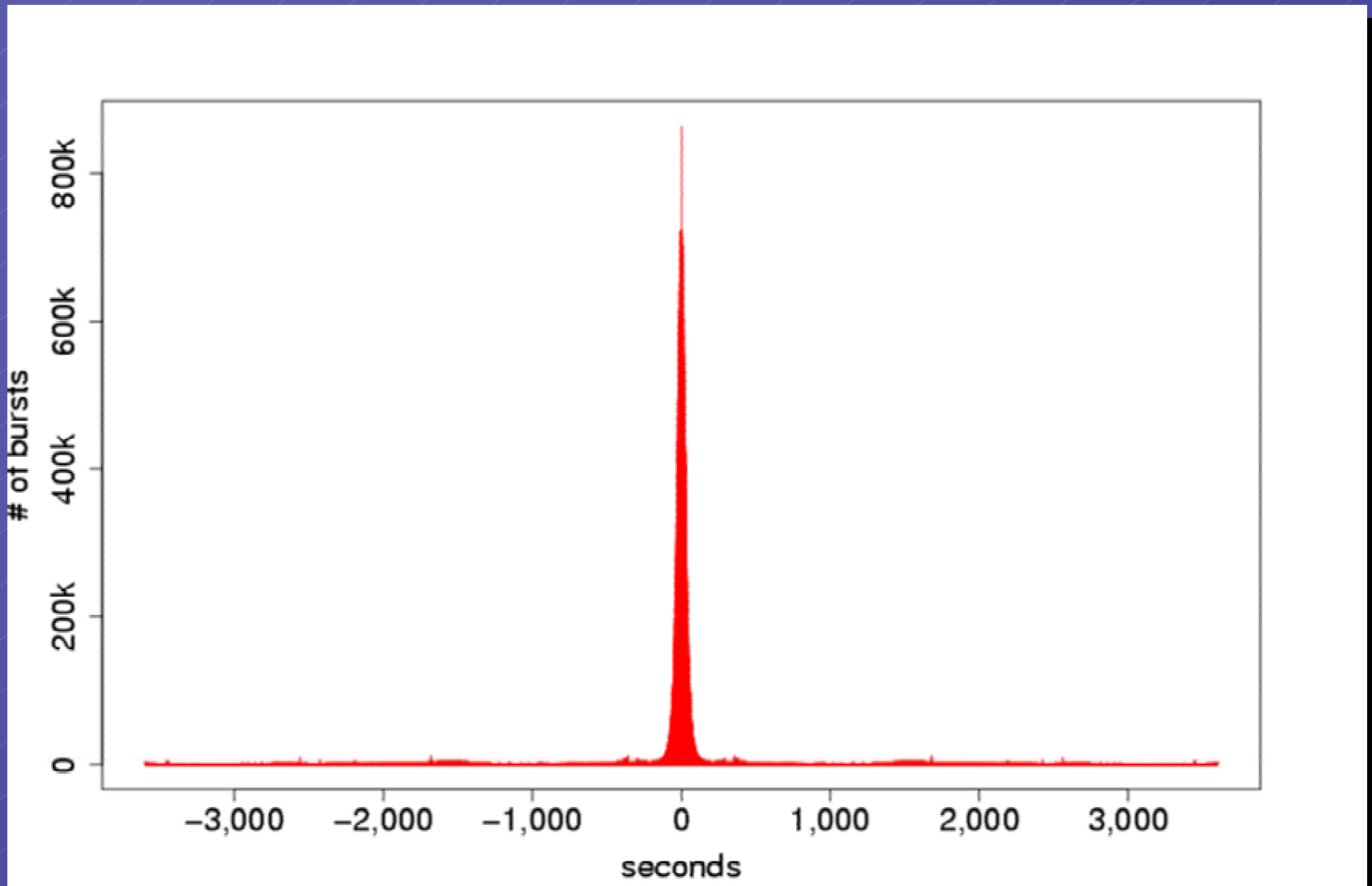
On BGP convergence

Interarrival properties of updates for the same prefix on different peers.

- pick one peer
- analyze interarrival process based on the selected peer.

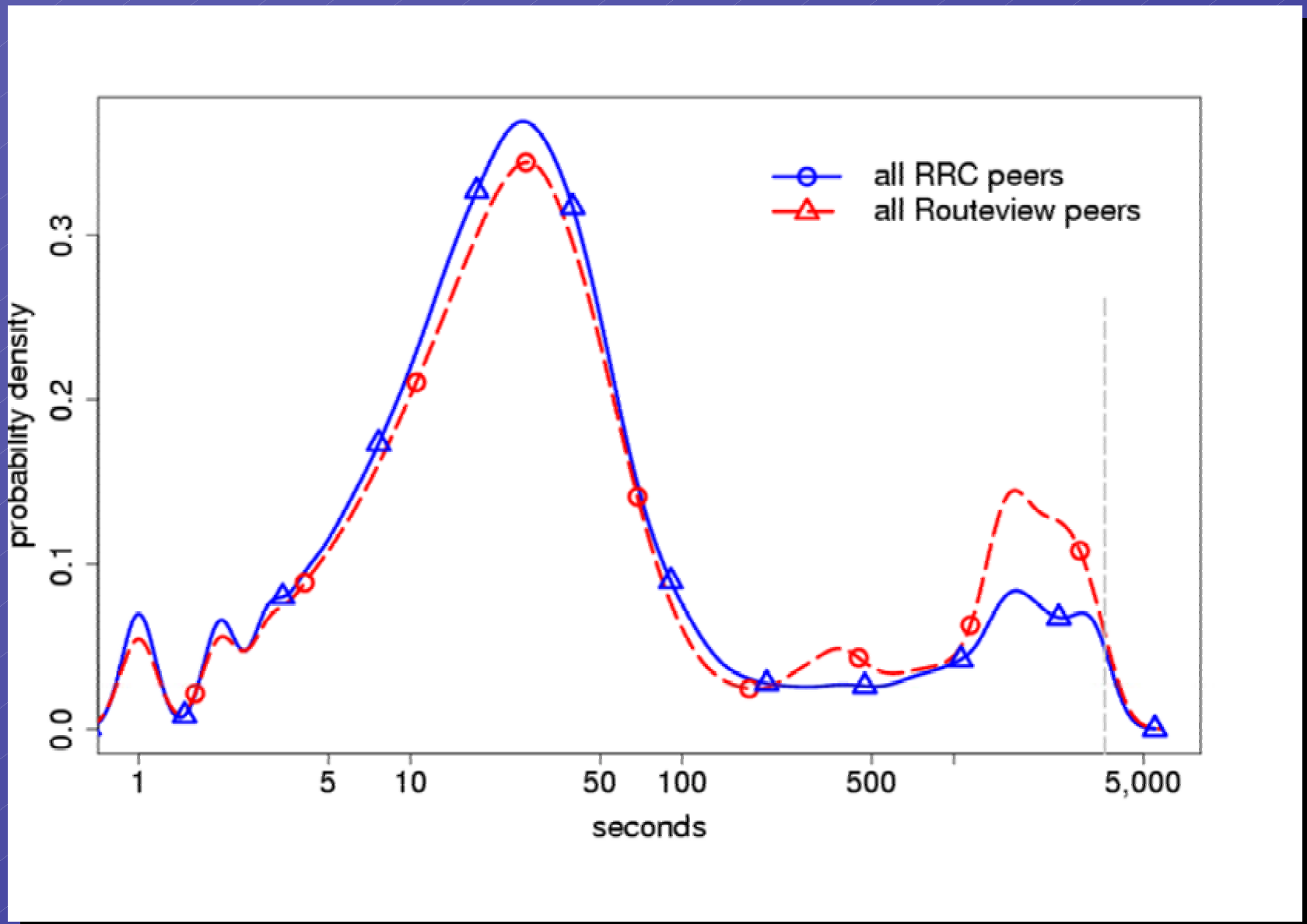


Interarrival time on peers

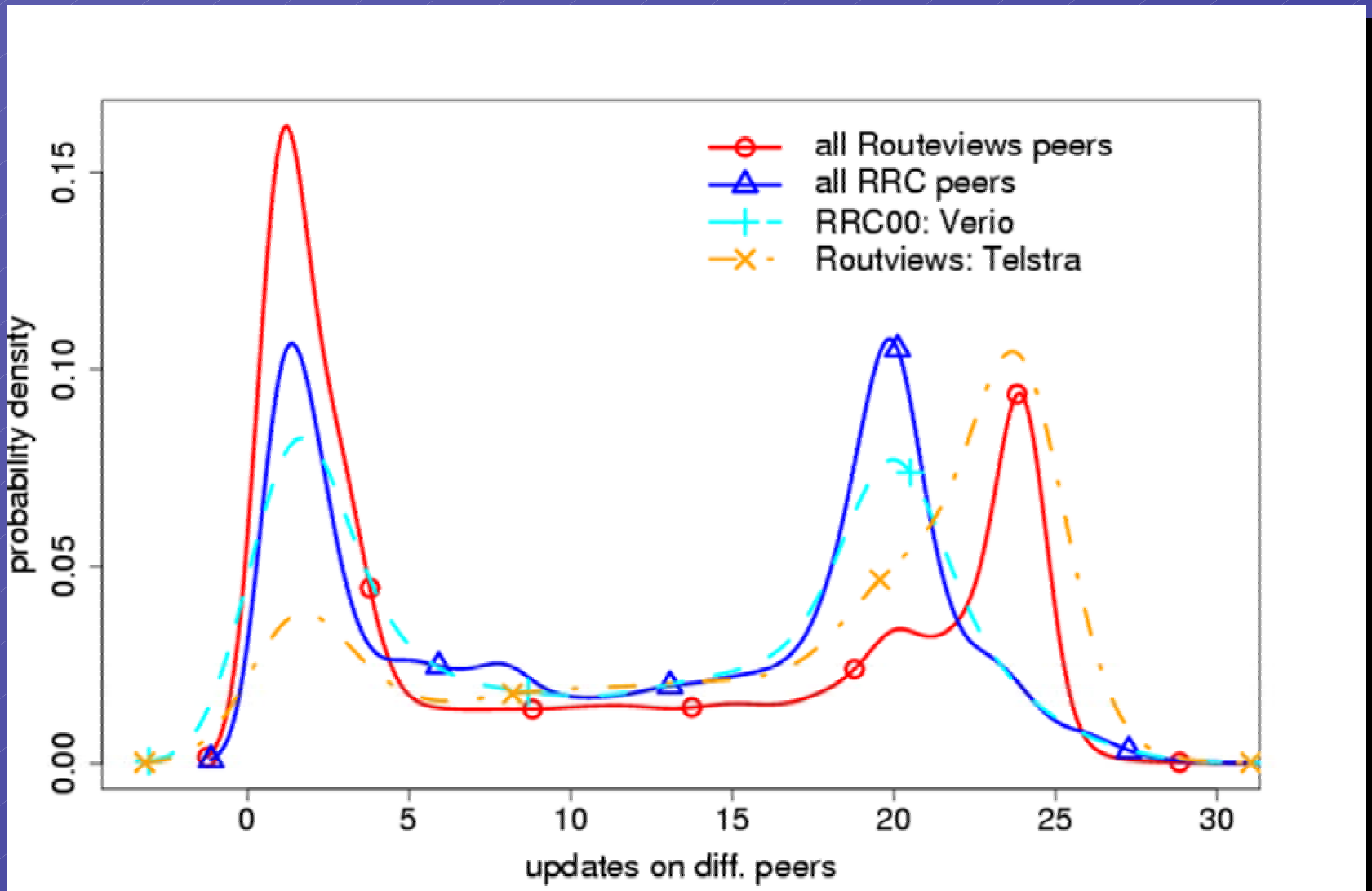


Datasets: Routeviews, SaarGate; Time: 09/18-09/25/2002

Interarrival time on peers



Updates on different peers



Testing scalability of BGP

Internet evolution

today ·····> future

How to judge:

- Router or BGP performance?
- Evaluate BGP's scalability?

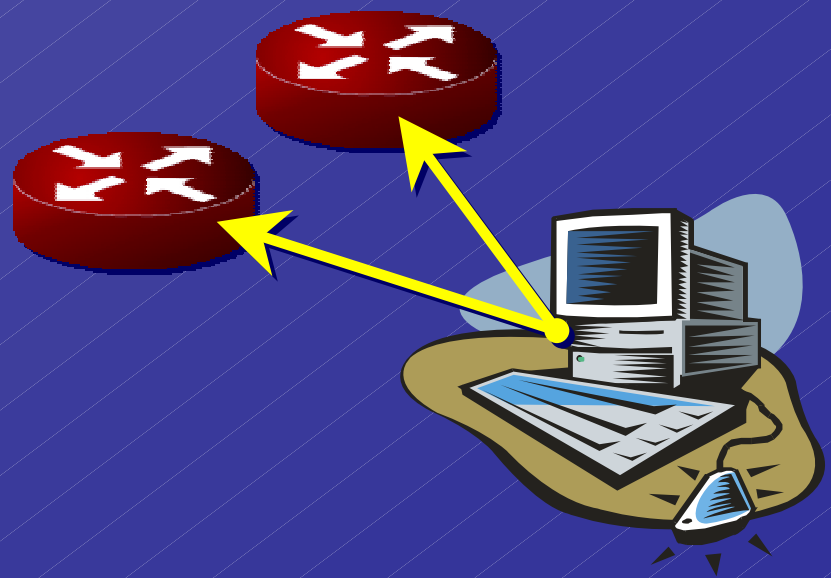
In test lab or real network?

⇓

NEED INTERNET VARIABILITY
IN A TEST LAB

Internet in a lab

- Topology
  - Tool: “Dummynet” (Pisa)
- User behavior
  - Tool: “Surge” (BU) / “Surge++” (work-in-progress)
- Routing
  - Tool: “RTG” (this talk)

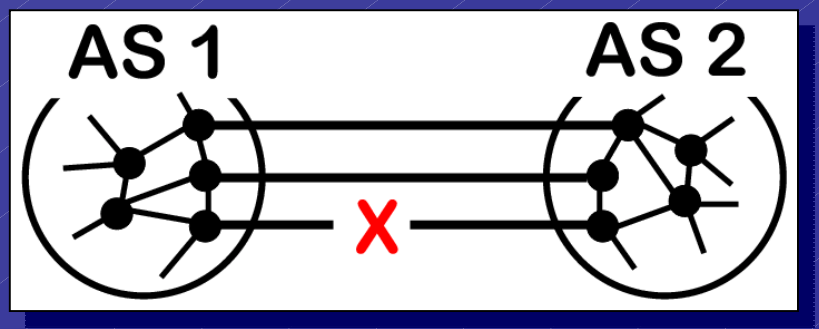


BGP workload ingredients

- Cause of routing instability
  - Instability creator
- Effect of routing instability
  - Instability bursts
- Baseline for prefix structure/hierarchy
  - Prefix forest
- AS topology and peering policies
  - AS path properties
- Correlations within instability
  - Attribute changes

Cause of routing instability

- Instability creators:
  - BGP session establishment/teardown/reset
  - Session parameter change
  - Link failure/repair
  - Addition/deletion of prefixes
  - Prefix policy changes
- Instability creator
  - Two peering ASes
  - Session AS
  - Prefix



Effect of routing instability

- Instability events:
  - Generated by the instability creator
  - Propagates through the network
  - Hard to capture
  - Generates related updates:
    update sequence observed in data
- Update bursts
  - Resulting set of updates
  - Observable in data
- Instability creator
  - Prefix: responsible for single update burst
  - AS: set of update bursts

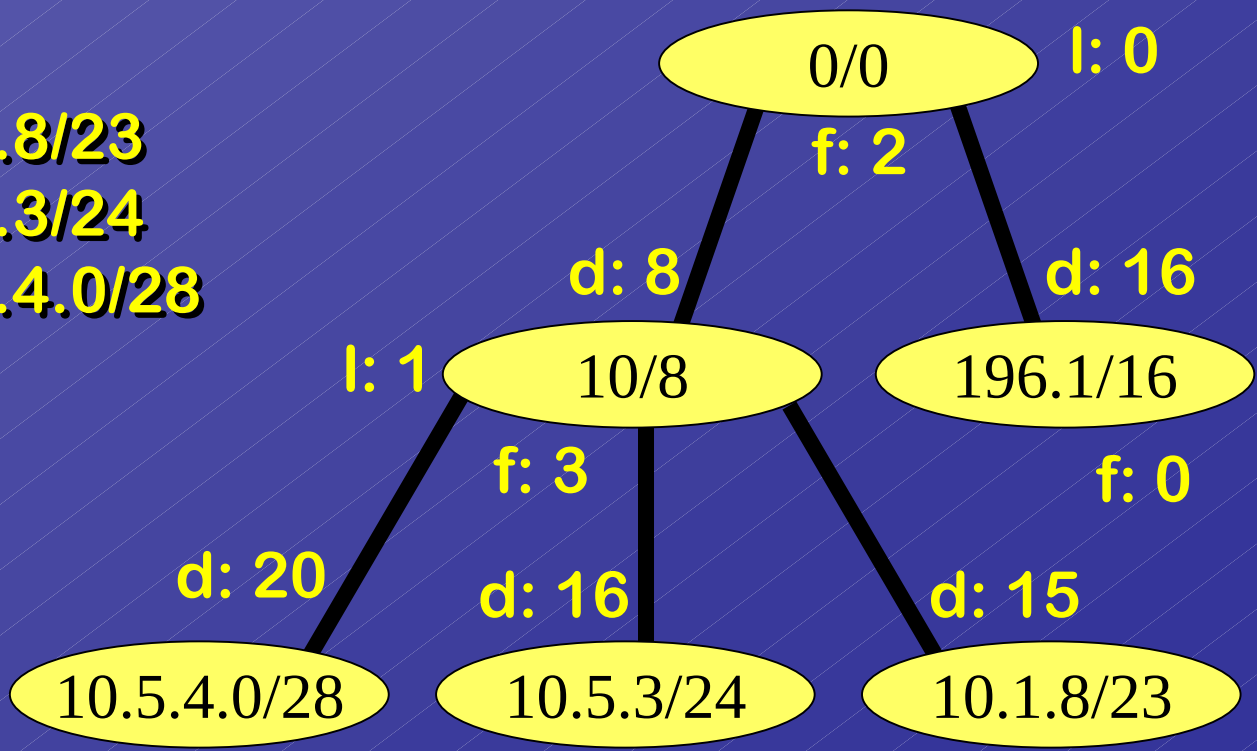
Baseline: prefix structure/hierarchy

- **Prefix forest**
  - Node == address space
  - Link == subset relationship

Prefix:

0/0	10.1.8/23
10/8	10.5.3/24
196.1/16	10.5.4.0/28

Distance: d
Depth: l
Fanout: f



AS topology/peering policies

- Correlation between instabilities: AS-path
- Peer view:
  - Directed acyclic graph (DAG)
    - Distance: BGP hops to remote AS
    - Connectivity: # of ASes at distance x
- Per AS information:
  - # of routes: originating/transiting
- Per AS path information:
  - Location of replication

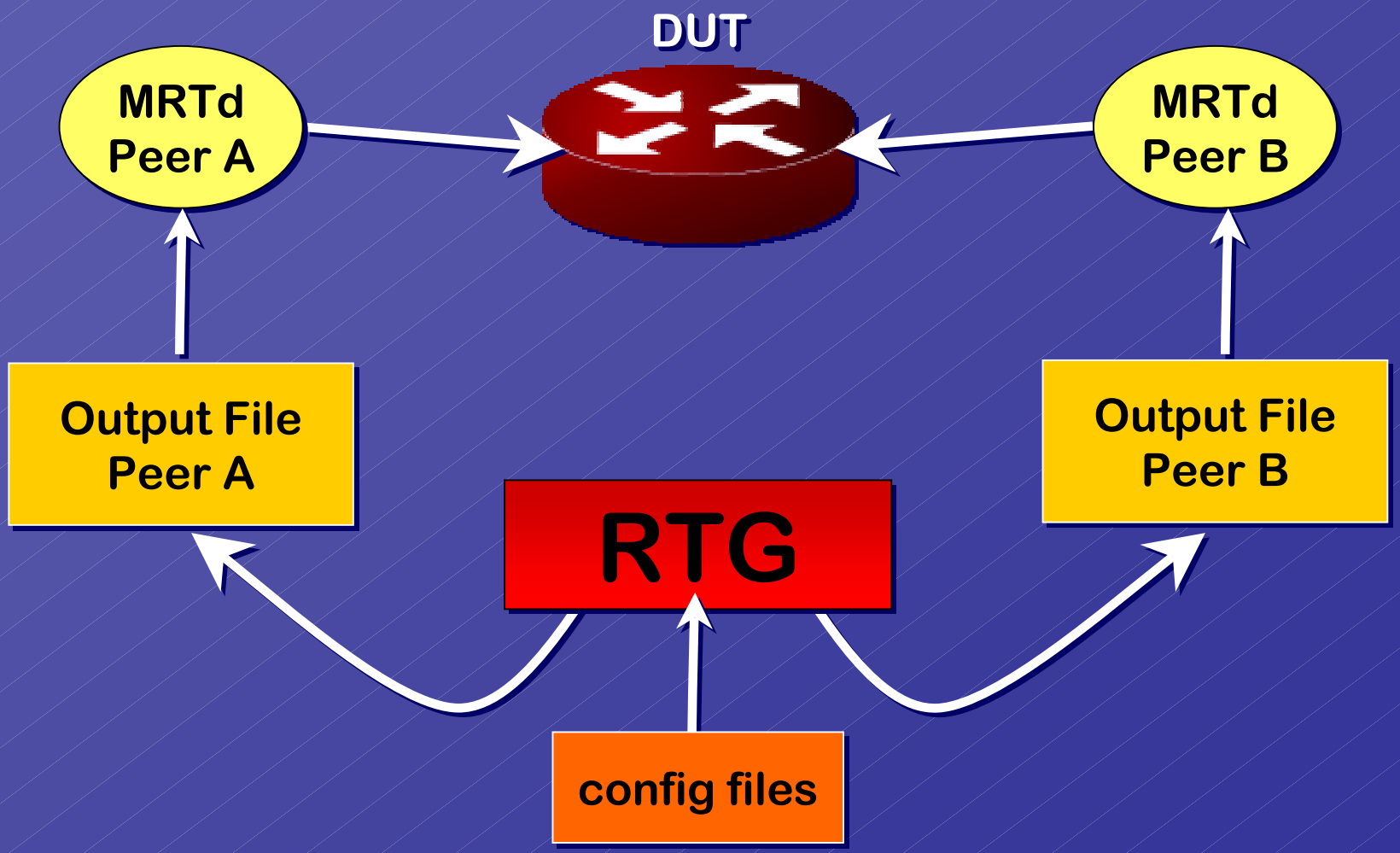
Correlations within instability

- Attribute change
  - New prefix?
  - Old prefix
  - Same attribute set
    - vs. previous update
    - vs. n‘s previous update
- Attributes
  - Fixed: Originator
  - Policy dependent: communities, MEDs
  - Convergence process: AS path, community

RTG: workload realization

- Idea
  - Generate updates off-line (stored in file): RTG
  - Feed them to system under test: e.g., MRTd
- RTG
  - Build routing table
  - Generate BGP attributes
  - Create BGP updates
- Parameters
  - Configuration files (automatic, semi-manual, manual)

BGP test bed setup



Benefits of BGP in a test-lab

- Test settings of BGP parameters
- Test interactions IGP vs. BGP
- Test BGP‘s scalability
- Test BGP protocol extensions
- Experiment with possible future workloads

Summary

- BGP workload model
  - identify structure in BGP traffic
  - characterize the structure using actual measurements
  - exploit the structure for a workload model
  - propose a tool, RTG, to realize the workload model
- One more component for an Internet Lab
- Towards a better testing methodology

Conclusion

If you are interested, please visit our website:

http://www.olafm.de/

Questions? Comments?!

Thanks !