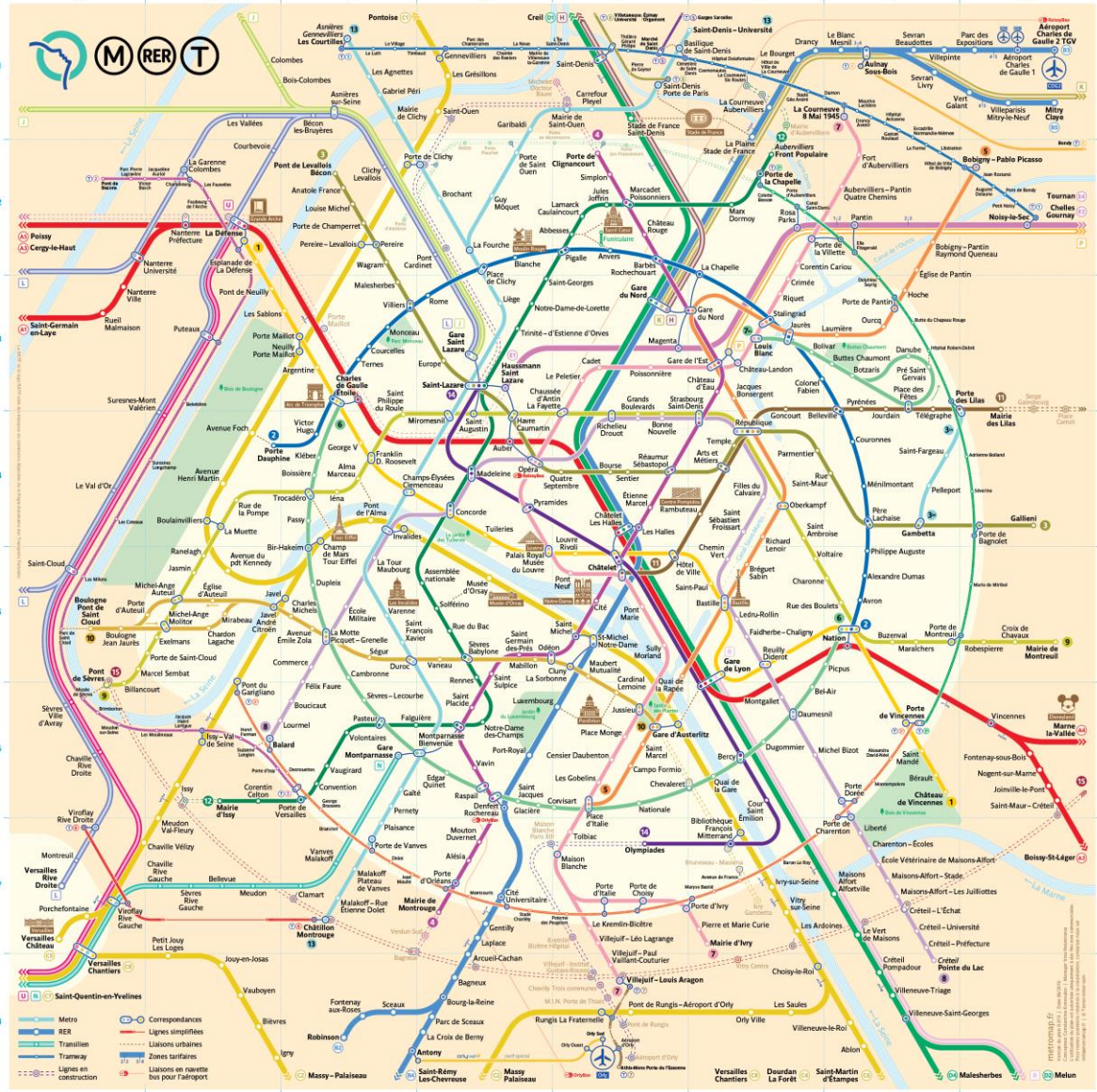
universität wien

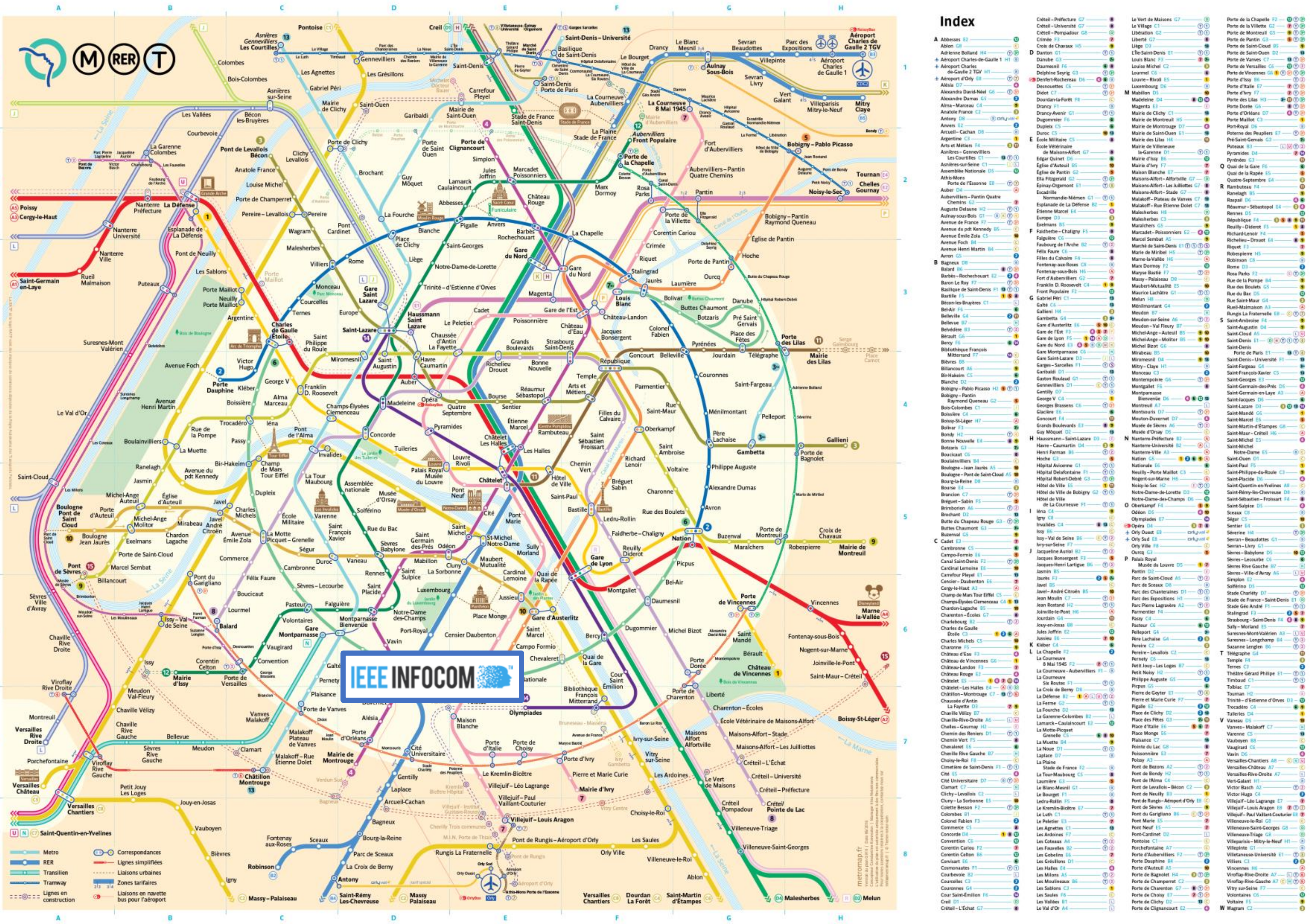
CASA: Congestion and Stretch Aware Static Fast Rerouting

Klaus-Tycho Foerster, Yvonne-Anne Pignolet (DFINITY), Stefan Schmid, and Gilles Tredan (LAAS-CNRS)



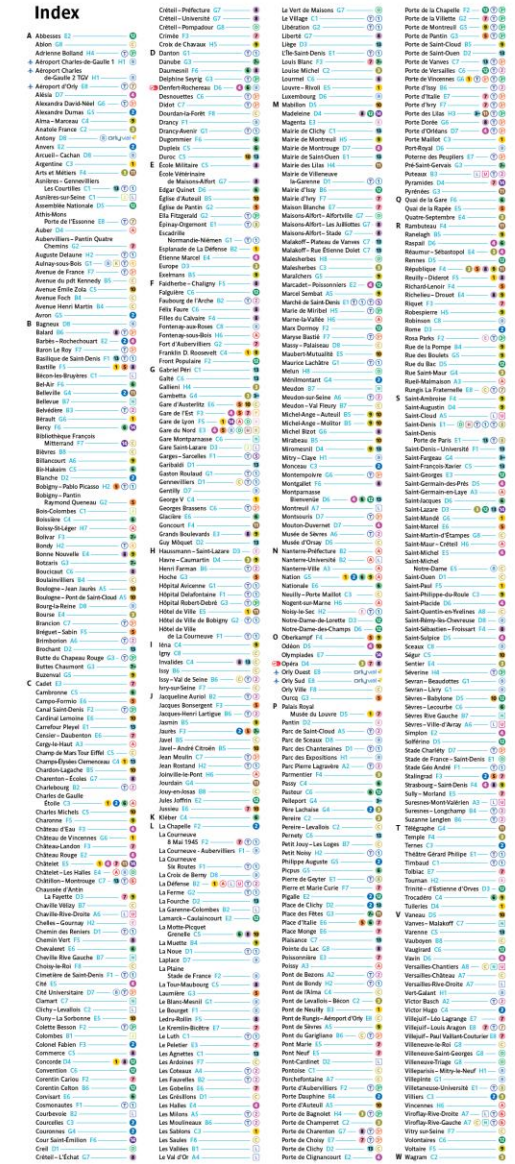
[Idea taken from Gilles Tredan]

http://metromap.fr

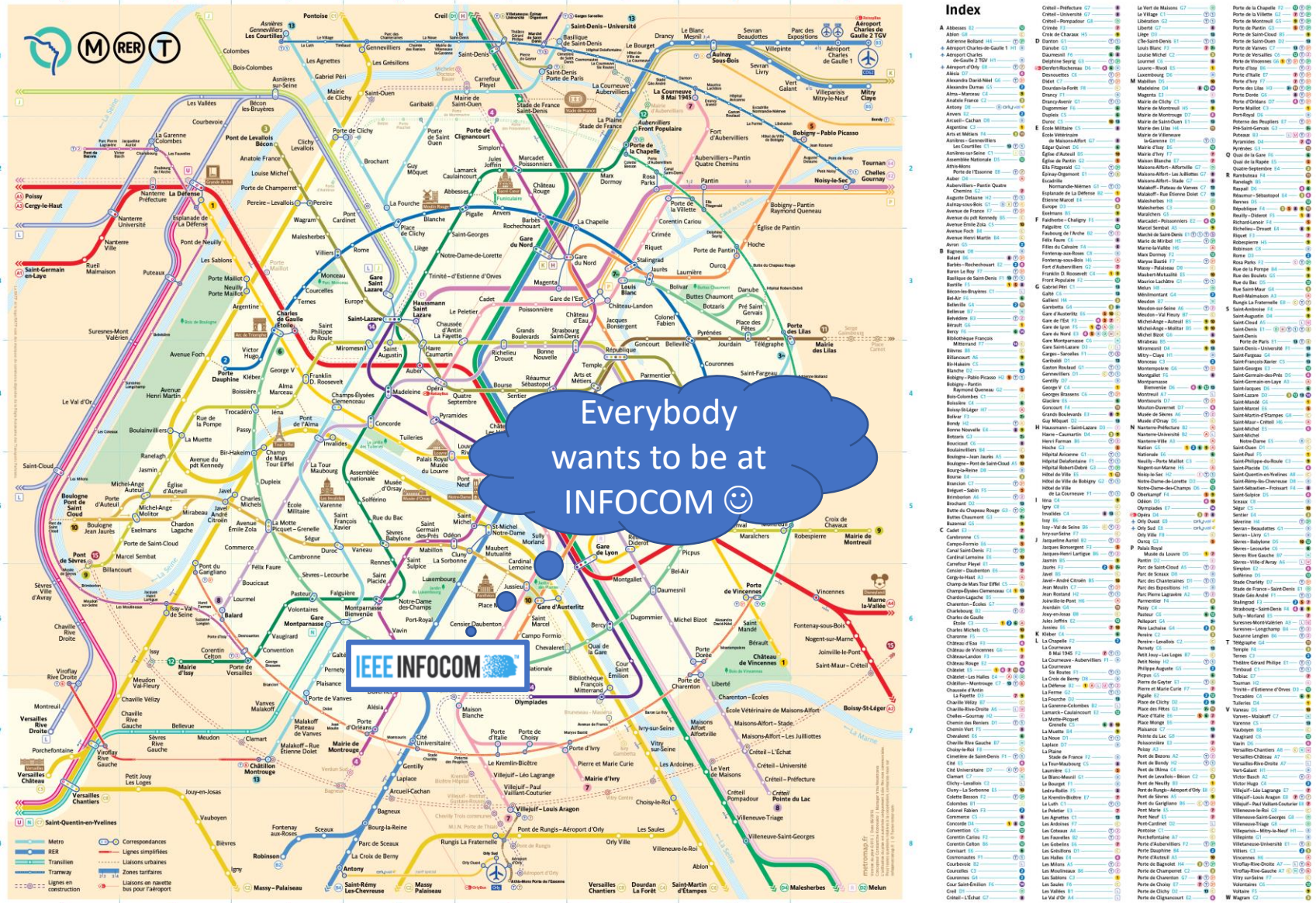


http://metromap.fr


Everybody
wants to be at
INFOCOM ☺
IEEE INFOCOM


Index

What if a link fails?



http://metromap.fr

What if a link fails? Take a detour ☺

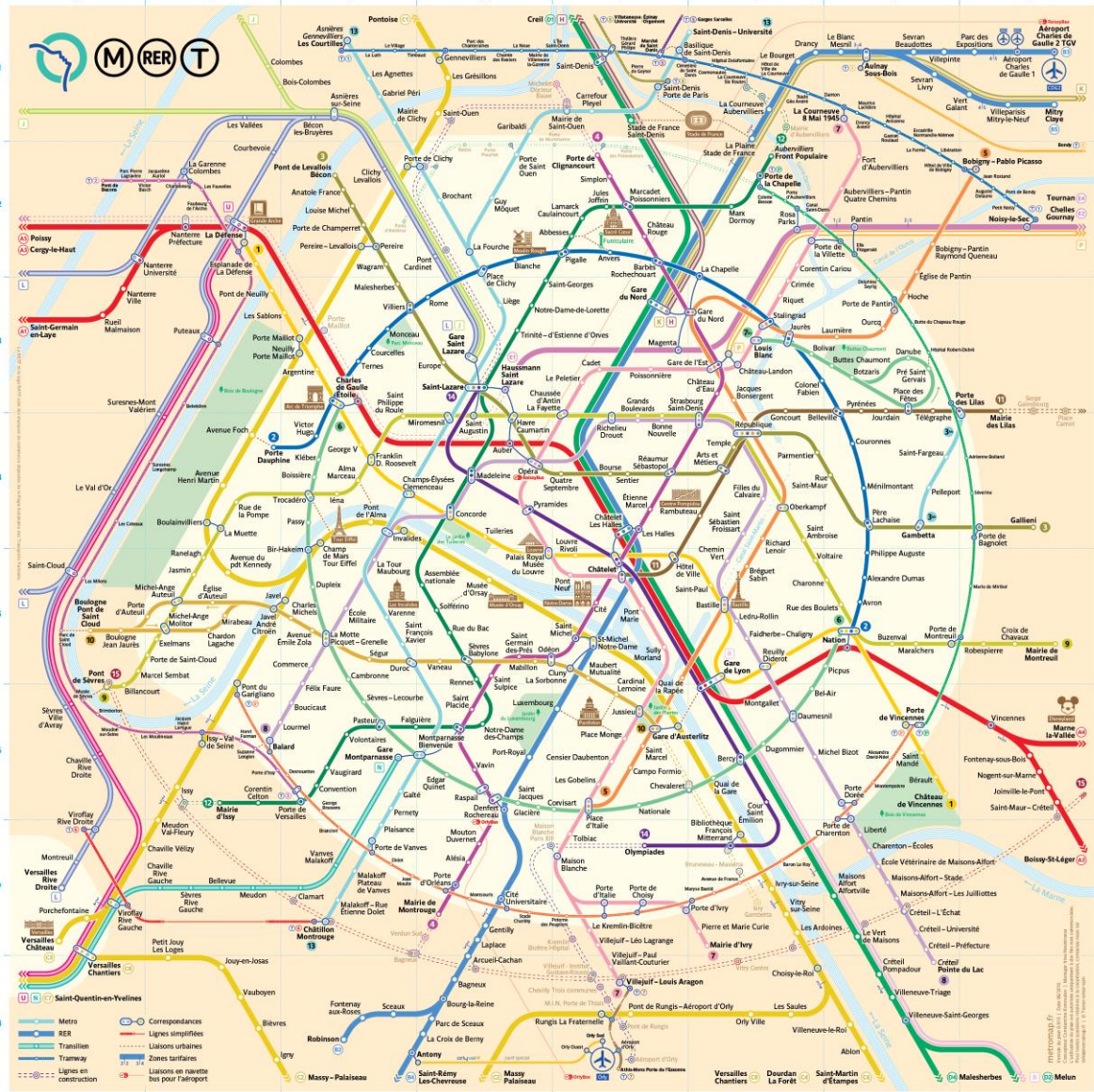
http://metromap.fr

Everybody takes the same detour? High load!

https://stephalvarez.wordpress.com/2011/03/06/bonjour-from-paris/

Distribute people over all detours? High path stretch!

https://www.elle.com/beauty/health-fitness/news/a35632/why-we-fall-asleep-on-trains/

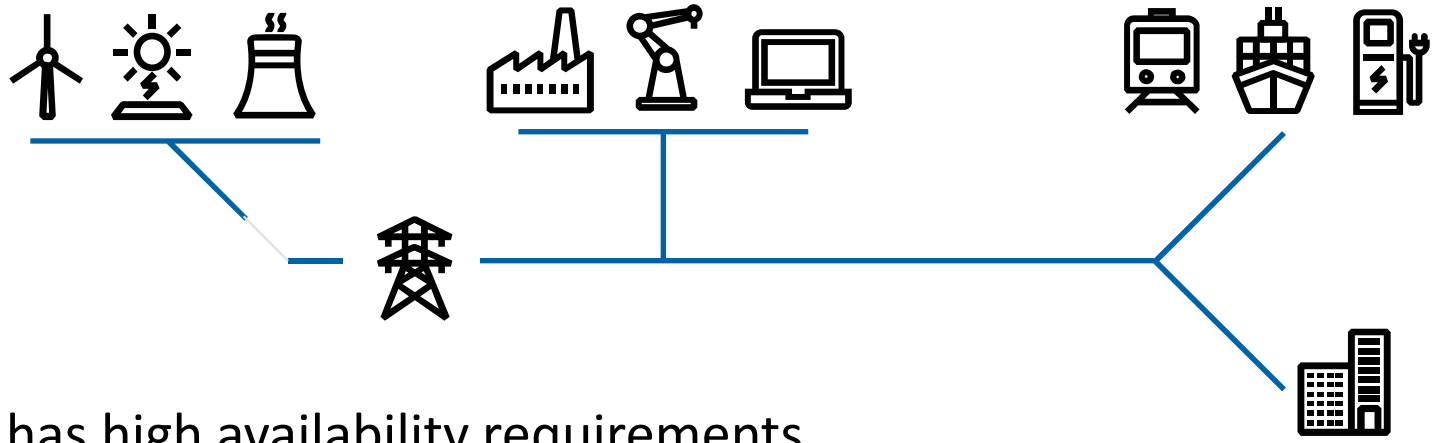


Index

universität wien

> "*The disparity in timescales between packet forwarding (which can be less than a microsecond) and control plane convergence (which can be as high as hundreds of milliseconds) means that failures often lead to unacceptably long outages*"
>
> Ensuring Connectivity via Data Plane Mechanisms: NSDI'13

Motivation

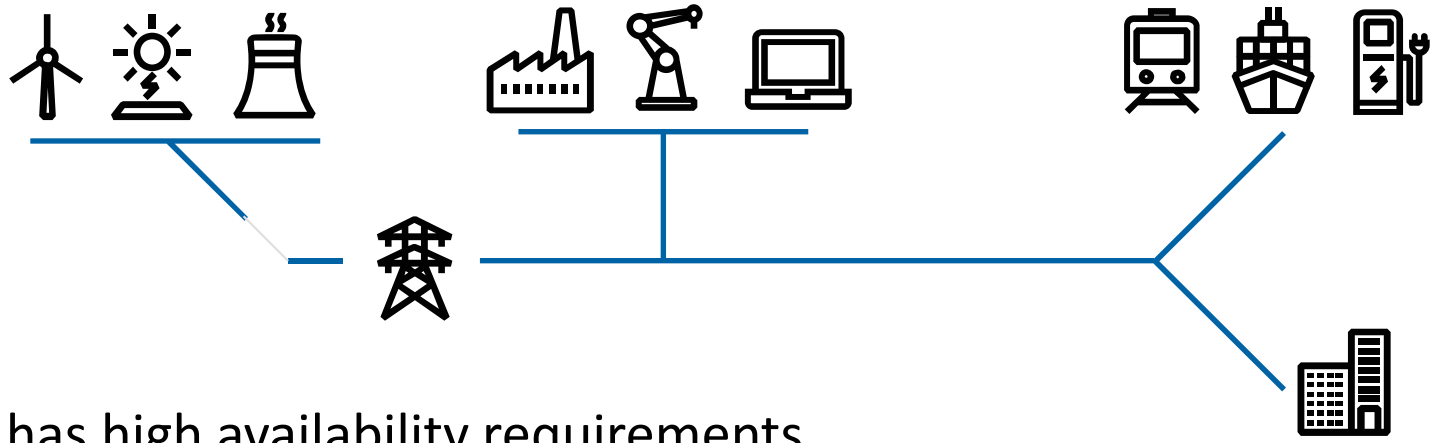


- Critical infrastructure has high availability requirements
- Industrial systems are more and more connected
- Hard real-time requirements

universität wien

> *"The disparity in timescales between packet forwarding (which can be less than a microsecond) and control plane convergence (which can be as high as hundreds of milliseconds) means that failures often lead to unacceptably long outages"*
>
> Ensuring Connectivity via Data Plane Mechanisms: NSDI'13

Motivation



- Critical infrastructure has high availability requirements
- Industrial systems are more and more connected
- Hard real-time requirements

⇒ How to provide dependability guarantee despite link failures in networks?
⇒ Possible without communication between nodes?
⇒ With low load? With low stretch?

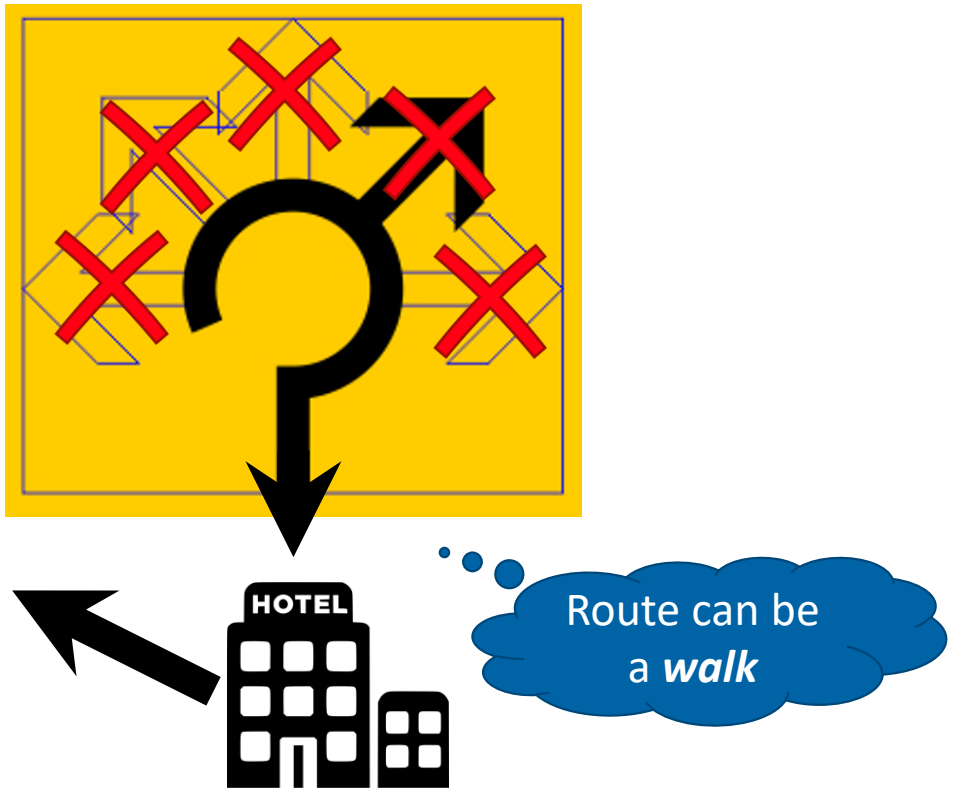
universität wien

Talk Structure

1. Model and Objectives
2. Background and Lower Bounds
3. Algorithms and Upper Bounds
4. Simulation Results
5. Conclusion and Outlook

Model I/II: Routing and Network

- Network is a strongly connected directed graph
- Forwarding may only match on:
  1. Source
  2. Destination
  3. Incident failures
  4. Incoming port
- No packet (header) changes allowed, no communication
- Static routing tables, deterministic behaviour
- Single destination routing, uniform flow sizes



universität wien

Model II/II: Quality *from a Worst-Case Perspective*

1. Resilience

- How many link failures can we survive and still guarantee delivery?
- Upper bound: (***r+1***)-link-connected graph: at most ***r***

2. Load

- Maximum additional link utilization due to rerouting

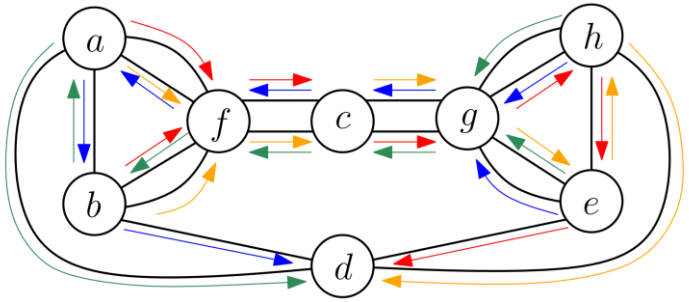
3. Stretch

- Maximum additional hops due to rerouting

Background: Static Fast Rerouting for Multiple Failures

Resiliency on **General Graphs**

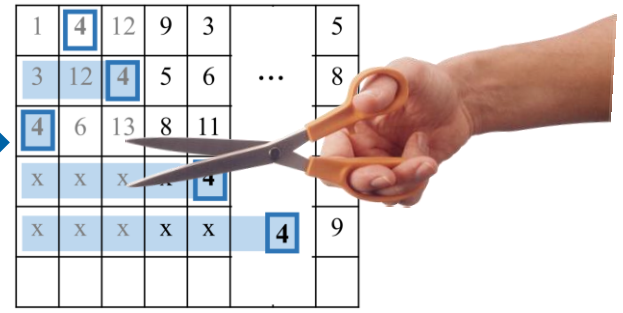
- Elhourani *et al.* [ToN'16] / Chiesa *et al.* [INFOCOM'16 etc]:
  - Employ directed link-disjoint arborescences
    - *i.e.* disjoint spanning routing trees
    - after failure: change tree (*e.g.* in circular fashion)
    - incoming port defines current tree



From Chiesa *et al.* 2016

Resiliency & Load on **Complete Graphs**

- Borokhovich & Schmid [OPODIS'13]
  - Bounds and handcrafted schemes
- Pignolet *et al.* [DSN'17]
  - Connection to Balanced Incomplete Block Designs (BIBDs)
    - General scheme how to distribute well after failures



From Pignolet *et al.* 2017

Resiliency & **Load** on **General Graphs**
this paper

With improved BIBDs!

universität wien

The Price of Locality (for *every* Scheme and Graph)

Stretch under ***r*** failures:

- Adversary can force to visit ***r+1*** neighbors of destination

Fail ***r*** links incident to the destination

Load under ***r*** failures:

- Adversary can force additional load of $\sqrt{r}$

Previously only weaker bound known, without incoming port

Let's try to meet this bound for many flows

universität wien

CASA: Rerouting on Arborescences

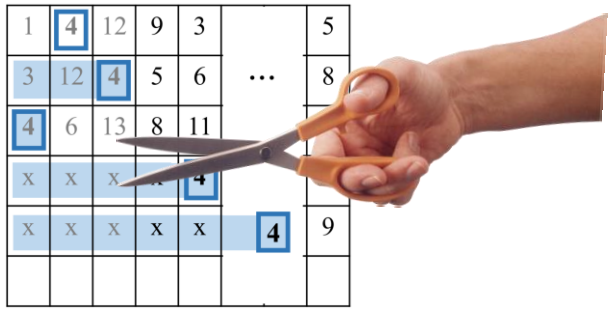
- Takes arborescences as input *e.g.* generated by Chiesa *et al.*
  - Influences the stretch, we get good bounds for *e.g.* so-called *independent spanning trees*

Algorithm

1: *Determine current arborescence T from in-port*

2: If next hop in T alive, use it, else

3: Pick next arborescence T' from ***BIBD-Matrix***



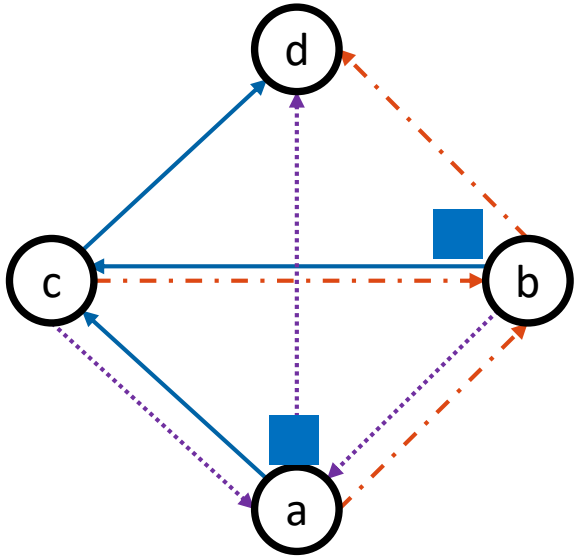
until the next hop is alive

different flows use different *T'*

We re-structure BIBD-matrix to be good for many flows

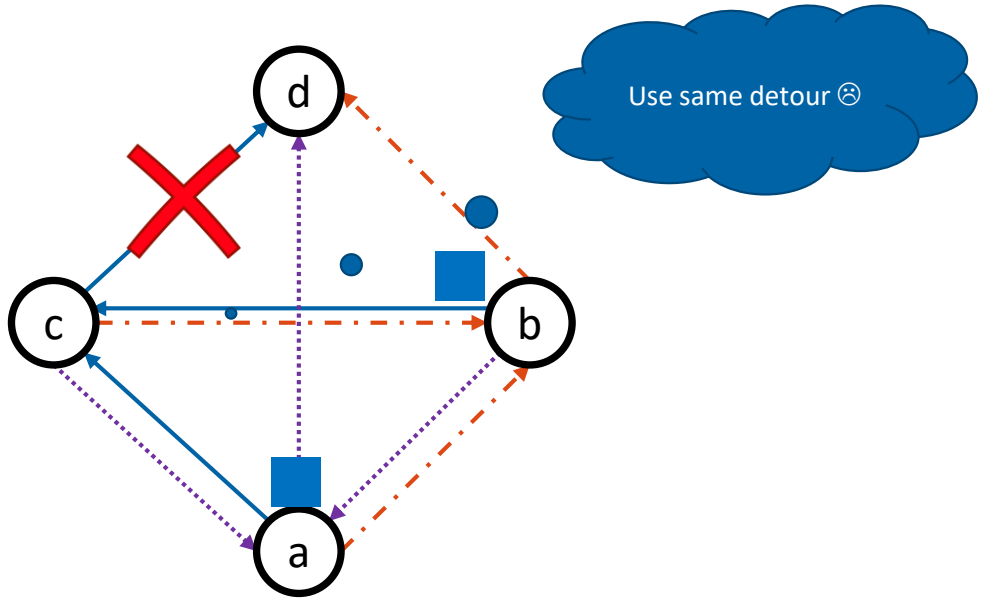
universität wien

CASA: Example *without* BIBD



universität wien

CASA: Example *without* BIBD

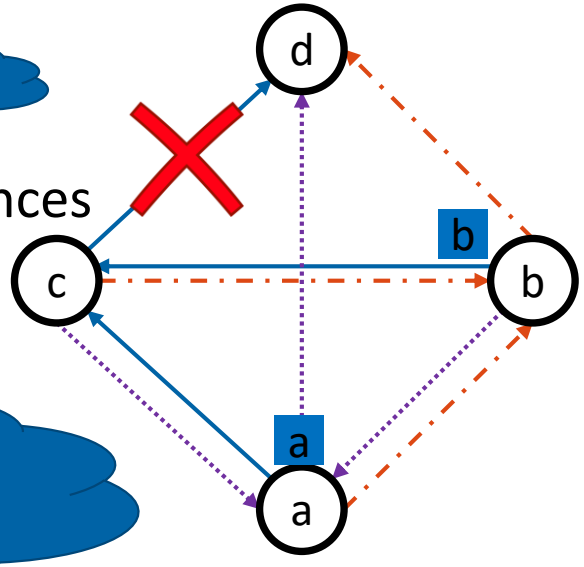


CASA: Example *with* BIBD

How much extra load?

- Up to $O(\sqrt{r})$ · · · Lower bound: $\sqrt{r}$
- For more flows than #arborescences

$$\sqrt{\#failures} < \frac{(\#arborescences)^{\frac{3}{2}}}{\#flows}$$



universität wien

Beyond CASA

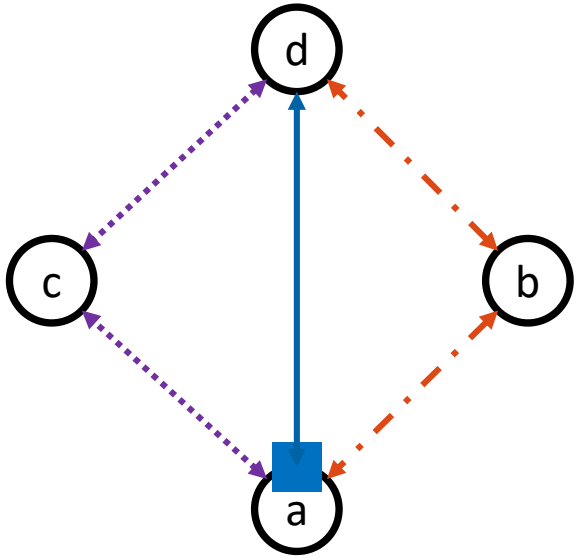
- ***r+1*** arborescences give ***r***-resiliency under directed link failures
  - But unclear how to obtain ***r***-resiliency under bi-directed link failures

- Motivation for a simplified heuristic: ***SquareOne***
  - Pick ***r+1*** bi-directed link-disjoint source-destination paths
    - Under failure: bounce back to the source, pick next path

https://Netflix.com

universität wien

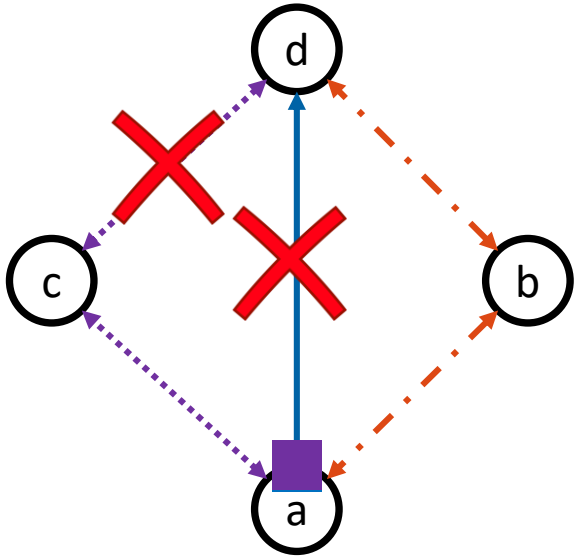
SquareOne



universität wien

https://Netflix.com

SquareOne



How good in practice?

No theoretical guarantees beyond resiliency

Easy to compute via *e.g.* max-flow formulations. Order path priority *e.g.* by length

universität wien

Selected Evaluations

Setting from prior work

- 8-connected 8-regular random graphs (**RR**, 100 routers each)
- well-connected cores of real-world ASes (*Rocketfuel*) (204-387 routers, 1667-4736 links)

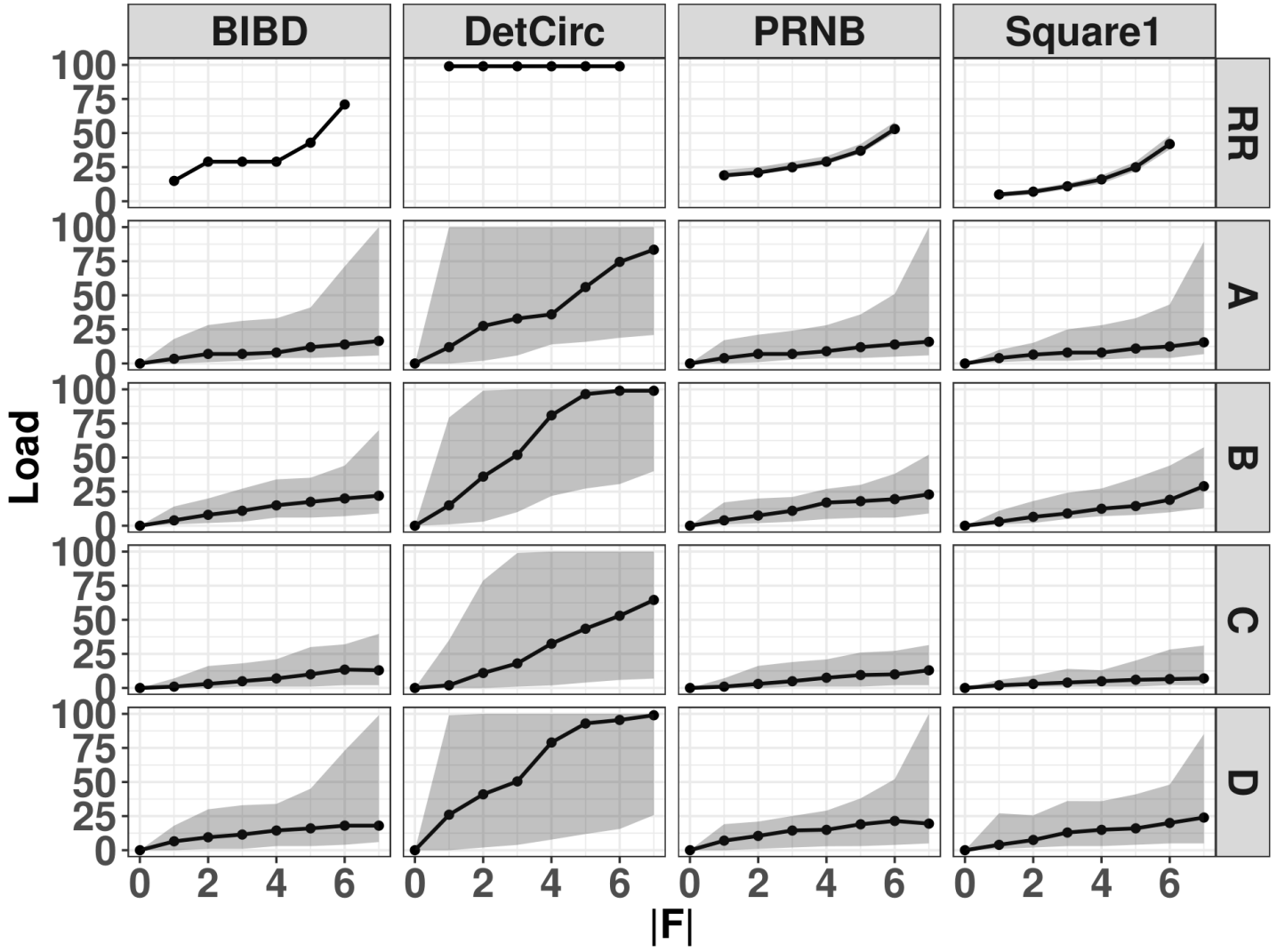
- Three arborescence methods (using the *same* arborescences)
  - *CASA **(BIBD)***
  - Deterministic Circular (***DetCirc***) from Chiesa *et al.*
  - Random (***PRNB***) from Chiesa *et al.*
- Also: **SquareOne**

Thanks to Marco Chiesa and Ilya Nikolaevskiy for their support

Issues in practice:
Real randomness on routers?
Packet reordering?

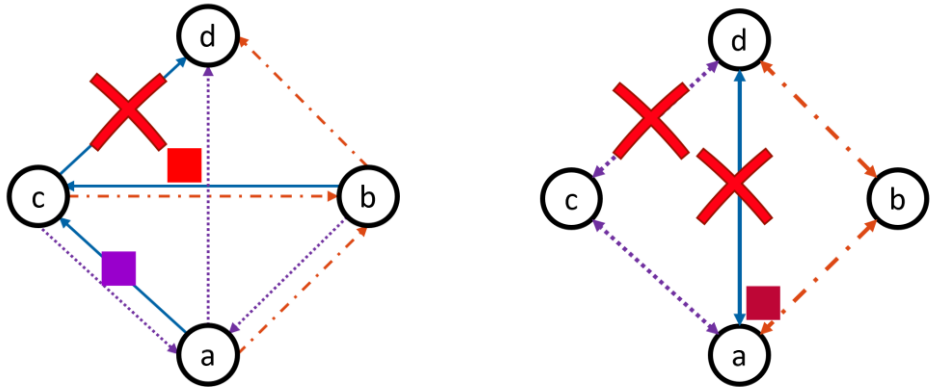
universität wien

Deterministic Worst-Case Failures



Conclusion

- We present **efficient static fast failover schemes** on general graphs
  - **CASA**: Combines **arborescences** and improved **block-designs** (BIBDs)
    - With theoretical guarantees
  - **SquareOne**: Well performing resilient heuristic
    - Based on edge-disjoint paths
- Next step: investigate specific topologies for better bounds



universität wien

Thanks a lot ☺! Slides are available at foerster.me/infocom19.pdf

CASA: Congestion and Stretch Aware Static Fast Rerouting

Klaus-Tycho Foerster, Yvonne-Anne Pignolet (DFINITY), Stefan Schmid, and Gilles Tredan (LAAS-CNRS)

universität wien

Papers Referenced (more in the paper)

- *How (Not) to Shoot in Your Foot with SDN Local Fast Failover: A Load-Connectivity Tradeoff*
  Michael Borokhovich and Stefan Schmid. OPODIS 2013
- *Load-Optimal Local Fast Rerouting for Dependable Networks*
  Yvonne-Anne Pignolet, Stefan Schmid, and Gilles Tredan. DSN 2013
- IP Fast Rerouting for Multi-Link Failures.
  Theodore Elhourani, Abishek Gopalan, Srinivasan Ramasubramanian.
  IEEE/ACM Trans. Netw. 24(5): 3014-3025 (2016)
- The Quest for Resilient (Static) Forwarding Tables.
  Marco Chiesa and Ilya Nikolaevskiy *et al.* INFOCOM 2016.

universität wien

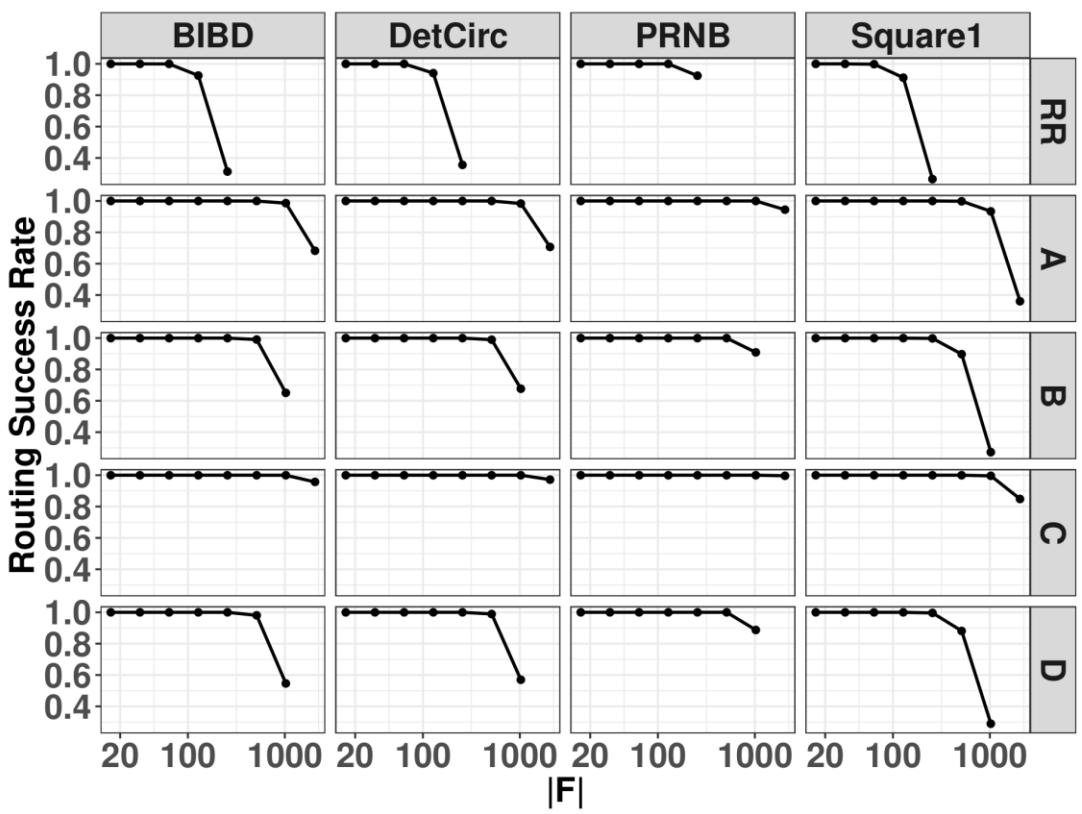
Rocketfuel ASes

AS	1239 A	2914 B	3356 C	7018 D
Number of nodes	389	225	377	204
Number of links	3621	1696	4736	1667
Eccentricity	6	6	6	6
Avg shortest path length	3.06	2.48	3.14	3.17

TABLE I: Properties of 8-connected cores of various ASes

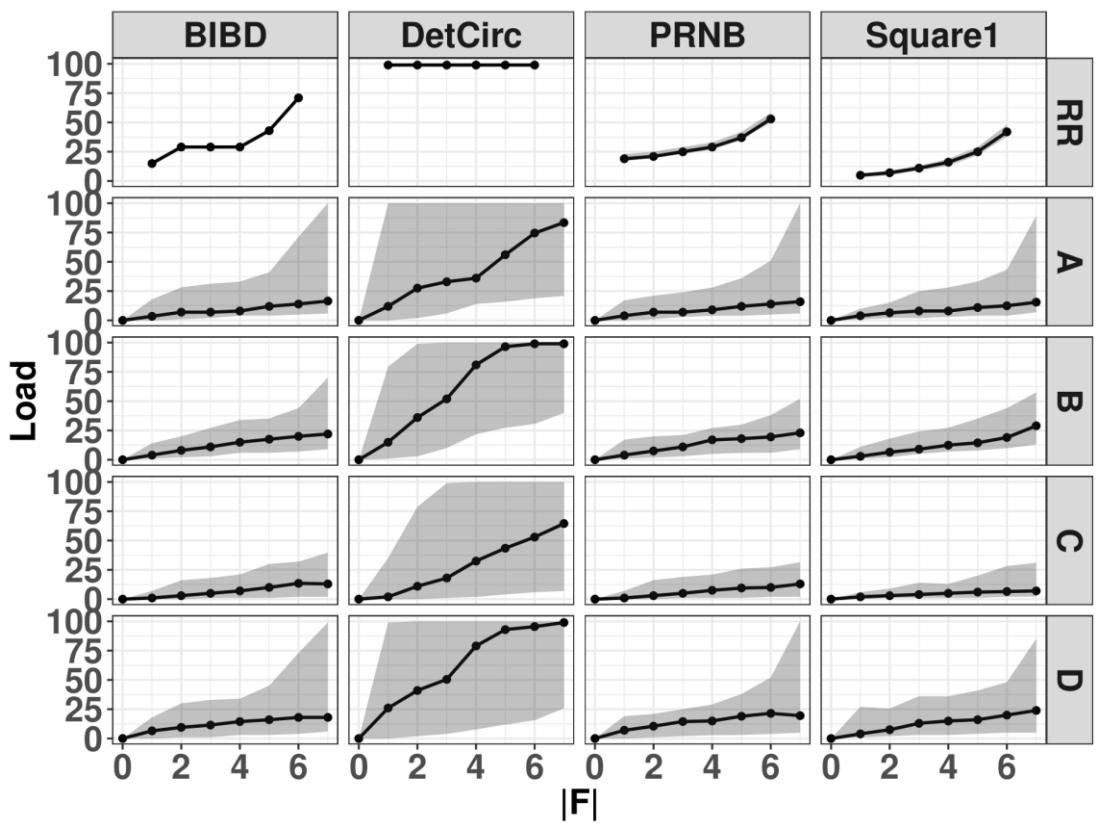
universität wien

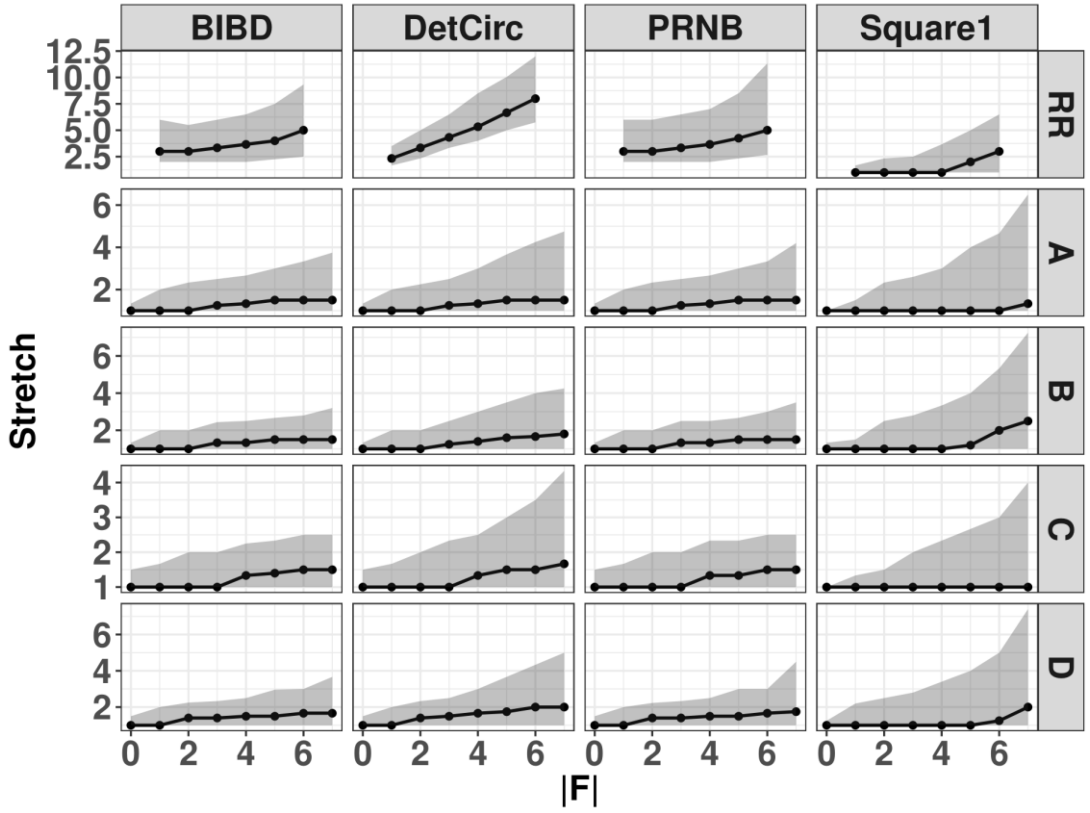
Evaluation: Resiliency



universität wien

Evaluation: Deterministic Worst-Case Failures





universität wien

Evaluation: Random Failures