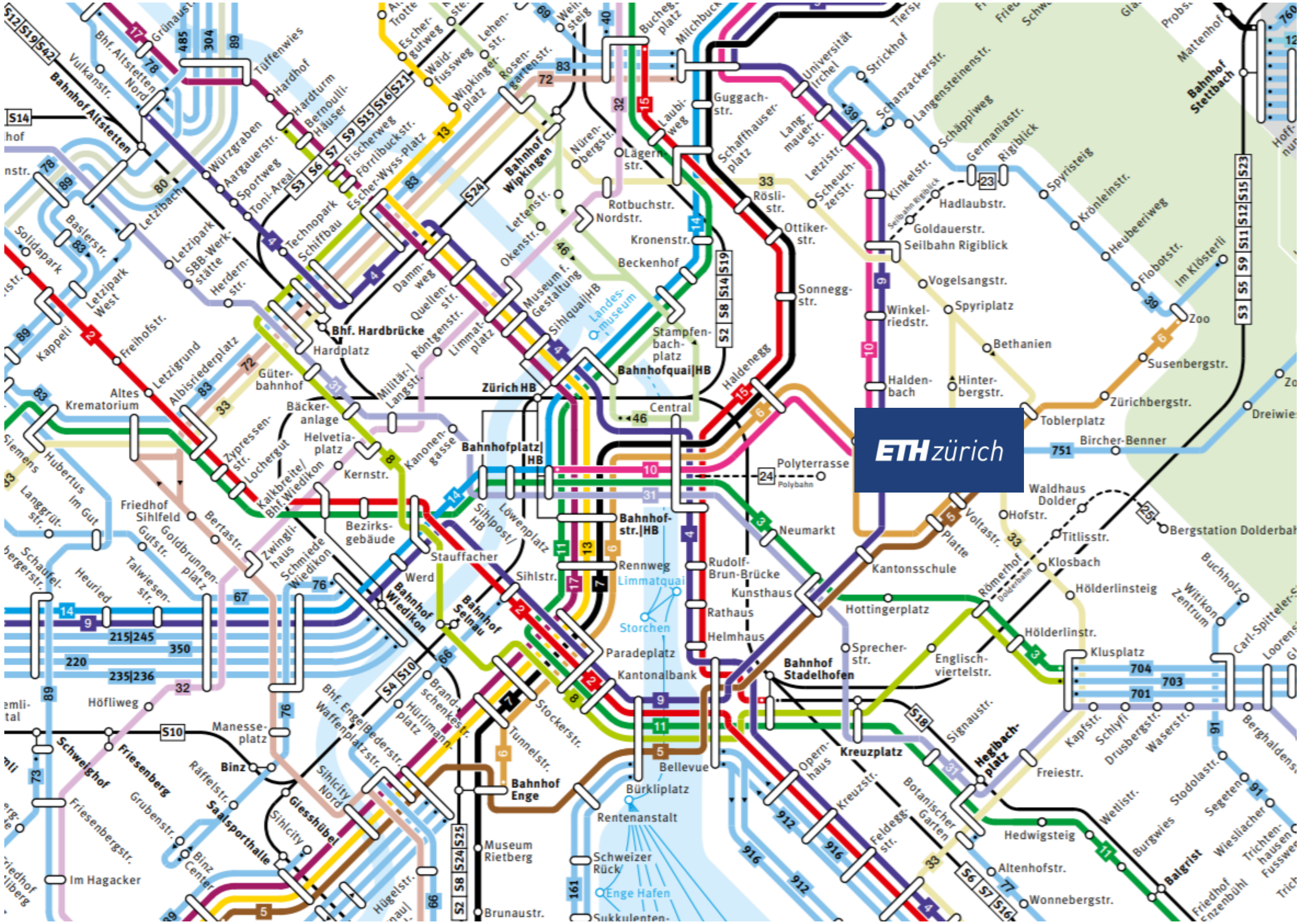
universität wien

Congestion and Stretch Aware Static Fast Rerouting [appeared @INFOCOM'19]

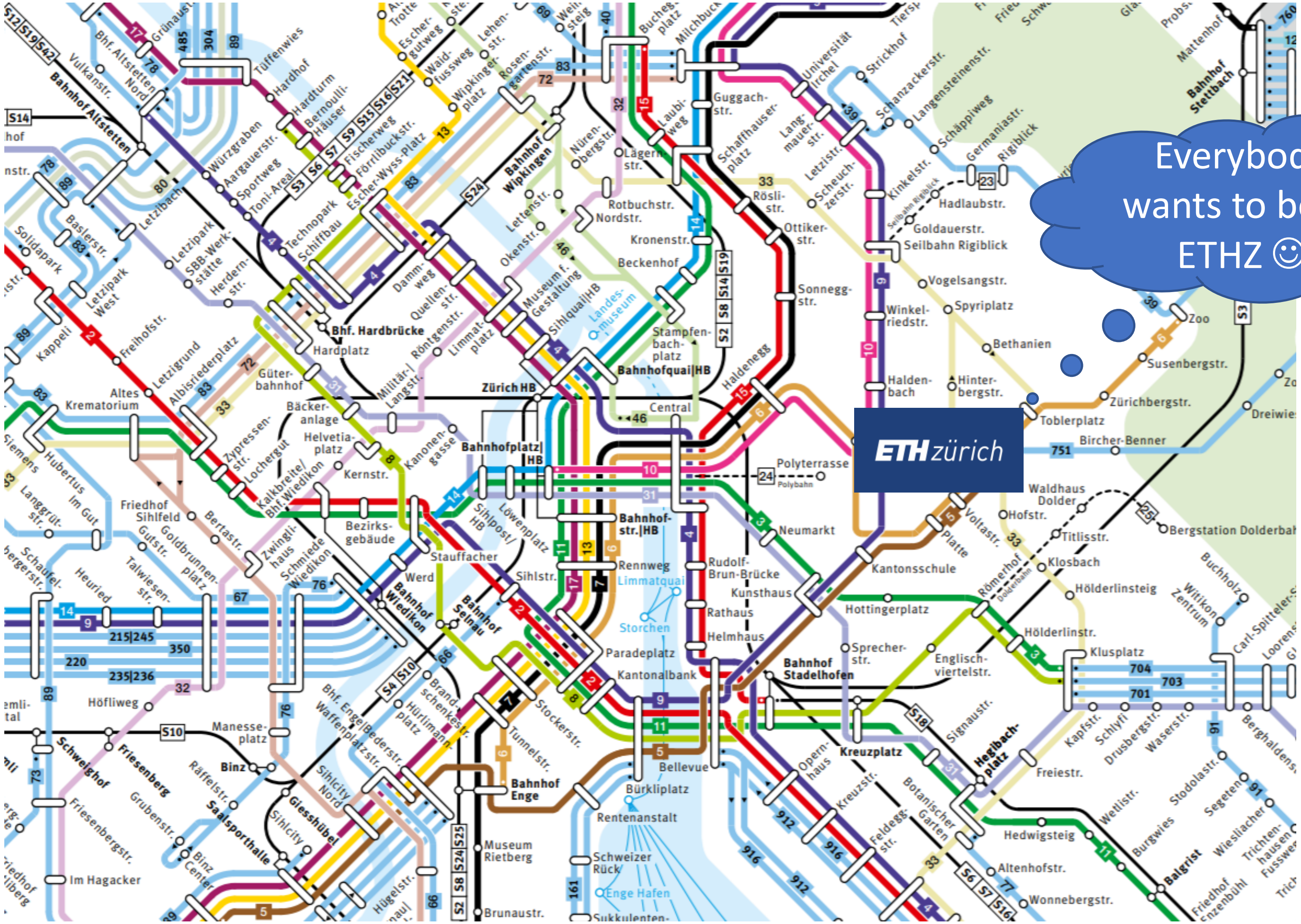
Klaus-Tycho Foerster, Yvonne-Anne Pignolet (DFINITY), Stefan Schmid, and Gilles Tredan (LAAS-CNRS)

[Idea taken from Gilles Tredan]

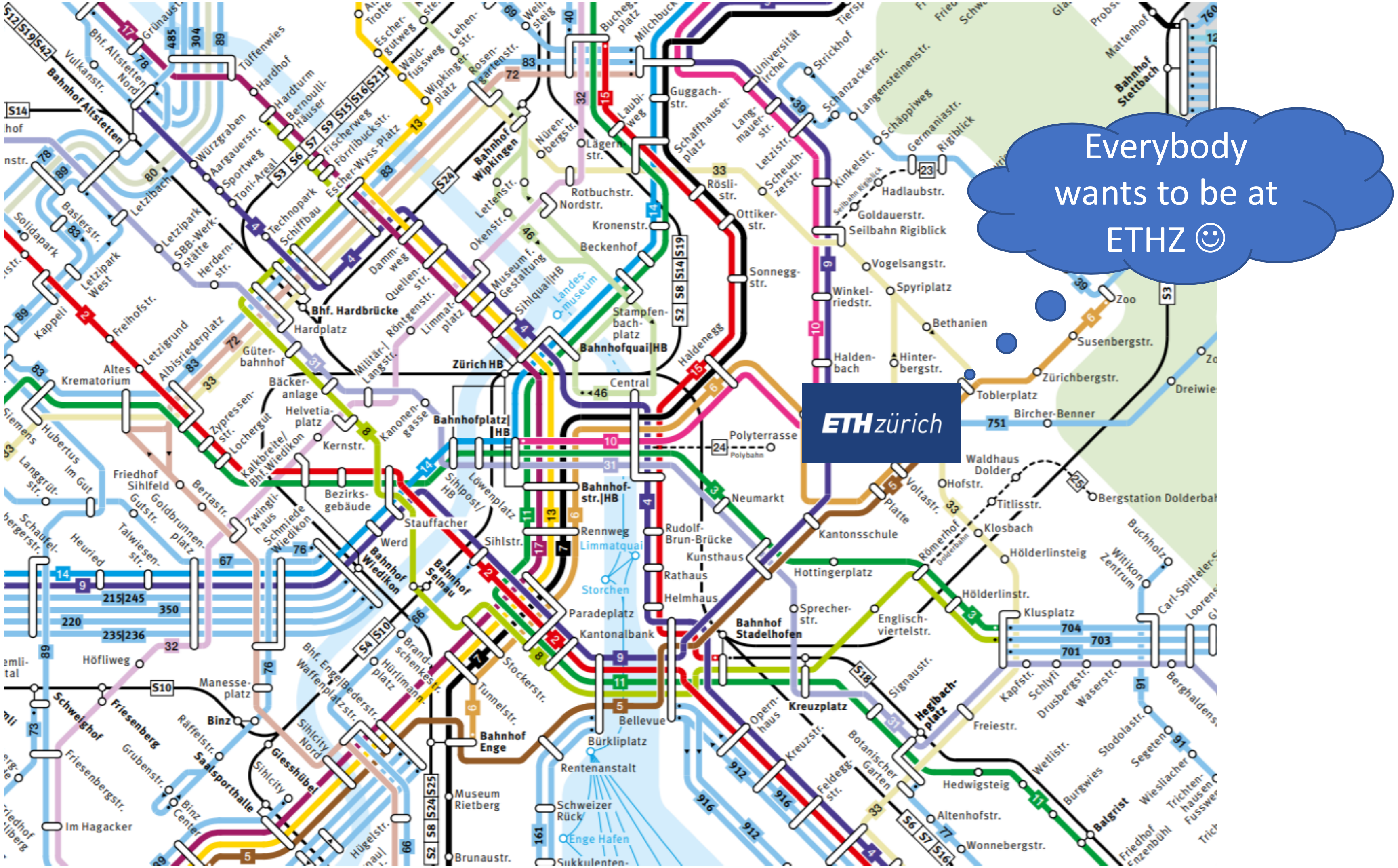


ETH zürich
Zürich HB
Bahnhofplatz/HB
Bahnhofquai/HB
Bahnhofstr./HB
Central
Polyterrasse
Polybahn
Neumarkt
Kantonsschule
Hottingerplatz
Kunsthaus
Rathaus
Helmhaus
Bahnhof Stadelhofen
Kreuzplatz
Bellevue
Bürkliplatz
Rentenanstalt
Paradeplatz
Kantonalbank
Rennweg
Rudolf-Brun-Brücke
Stauffacher
Sihlstr.
Werd
Bahnhof Selnau
Bahnhof Wiedikon
Bahnhof Enge
Bahnhof Wipkingen
Bahnhof Altstetten
Bahnhof Stettbach
Bhf. Hardbrücke
Hardplatz
Güterbahnhof
Bäckeranlage
Helvetiaplatz
Kernstr.
Bezirksgebäude
Löwenplatz
Limmatplatz
Museum für Gestaltung
Landesmuseum
Stampfenbachplatz
Beckenhof
Kronenstr.
Rotbuchstr.
Nordstr.
Guggachstr.
Schaffhauserplatz
Röslistr.
Ottikerstr.
Sonneggstr.
Haldenegg
Haldenbach
Winkelriedstr.
Seilbahn Rigiblick
Goldauerstr.
Hadlaubstr.
Vogelsangstr.
Spyriplatz
Bethanien
Hinterbergstr.
Toblerplatz
Zürichbergstr.
Susenbergstr.
Zoo
Bircher-Benner
Waldhaus Dolder
Hofstr.
Titlisstr.
Klosbach
Hölderlinsteig
Hölderlinstr.
Klusplatz
Bergstation Dolderbahn
Englischviertelstr.
Sprecherstr.
Hegibachplatz
Freiestr.
Hedwigsteig
Altenhofstr.
Wonnebergstr.
Balgrist
Museum Rietberg
Brunaustr.
Schweizer Rück
Enge Hafen
Storchen
Limmatquai
Manesseplatz
Binz
Höfliweg
Friedhof Sihlfeld
Altes Krematorium
Letzigrund
Albisriederplatz
Freihofstr.
Kappeli
Letzipark West
Letzipark
Letzibach
Hardhof
Tüffenwies
Technopark
Schiffbau
Escher-Wyss-Platz
Wipkingerplatz
Schweighof
Friesenberg
Saalsporthalle
Giesshübel
Im Hagacker
Zwinglihaus
Schmiede Wiedikon
Kalkbreite/Bhf. Wiedikon
Lochergut
Zypressenstr.
Stockerstr.
Tunnelstr.
Opernhaus
Kreuzstr.
Feldeggstr.
Botanischer Garten

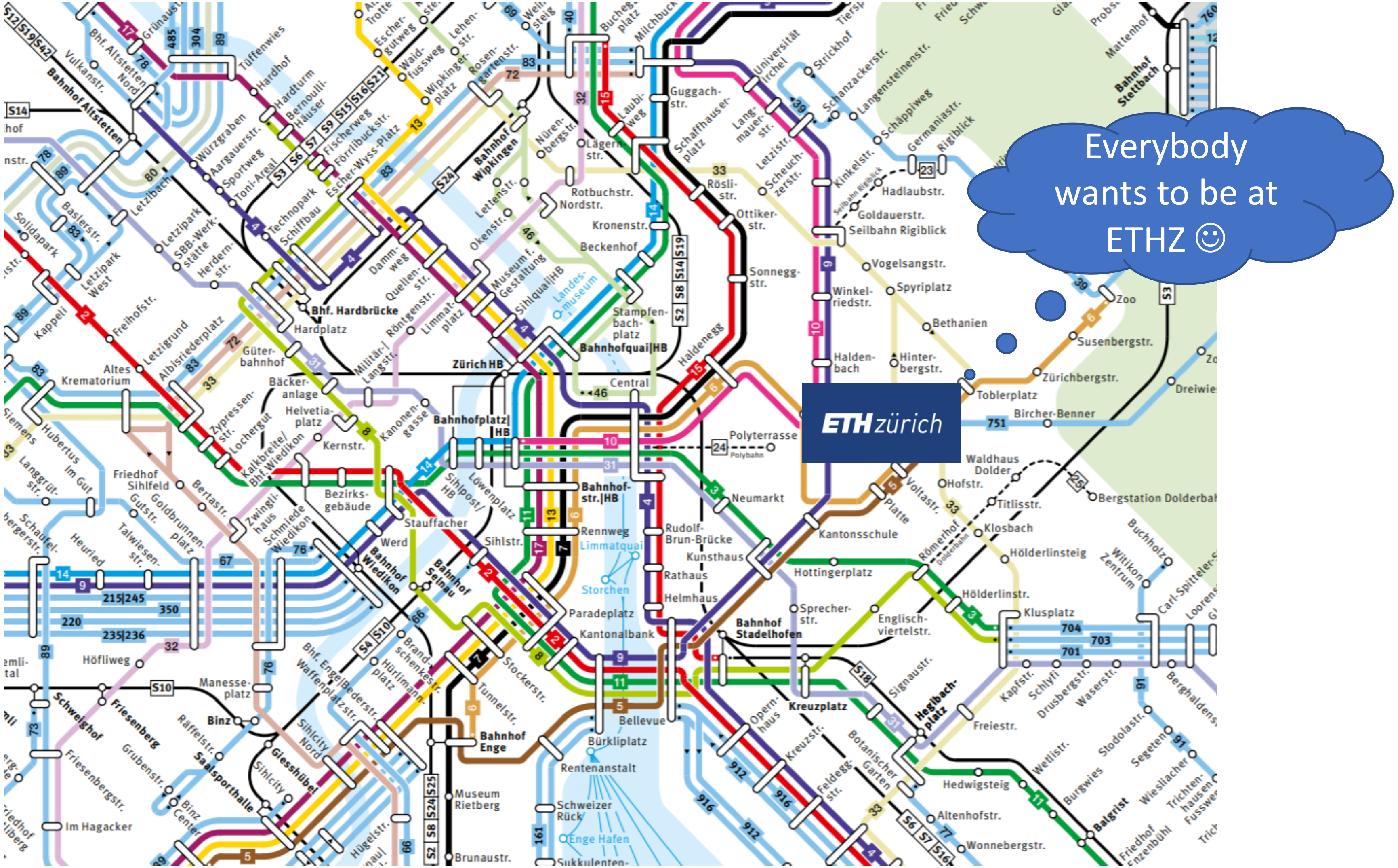
Everybody wants to be at ETHZ ☺



What if a link fails?



What if a link fails? Take a detour ☺



Everybody takes the same detour? High load!

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

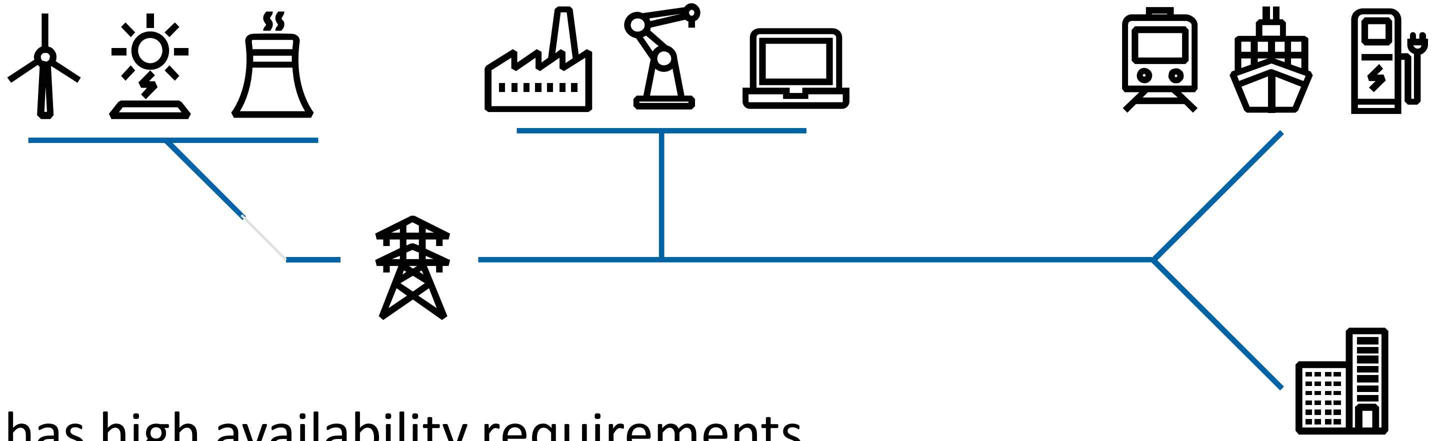
Distribute people over all detours? High path stretch!

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


Bahnhof Altstetten
Vulkanstr.
Bhf. Altstetten Nord
Grünaustr.
Tüffenwies
Hardhof
Hardturm
Bernoulli-Häuser
Würzgraben
Aargauerstr.
Sportweg
Toni-Areal
Fischerweg
Förrlibuckstr.
Escher-Wyss-Platz
Technopark
Schiffbau
Letzibach
Letzipark
SBB-Werkstätte
Herdernstr.
Letzipark West
Solidpark
Kappeli
Freihofstr.
Letzigrund
Albisriederplatz
Altes Krematorium
Siemens
Hubertus
Langgrütstr.
Im Gut
Friedhof Sihlfeld
Goldbrunnenplatz
Gutstr.
Talwiesenstr.
Heuried
Schaufelbergerstr.
Bertastr.
Zypressenstr.
Lochergut
Kalkbreite/Bhf. Wiedikon
Zwinglihaus
Schmiede Wiedikon
Bahnhof Wiedikon
Bhf. Hardbrücke
Hardplatz
Güterbahnhof
Bäckeranlage
Helvetiaplatz
Kernstr.
Bezirksgebäude
Stauffacher
Werd
Bahnhof Selnau
Sihlstr.
Sihlpost/HB
Löwenplatz
Militär-/Langstr.
Röntgenstr.
Limmatplatz
Quellenstr.
Dammweg
Kanonengasse
Museum f. Gestaltung
Sihlquai/HB
Landesmuseum
Zürich HB
Bahnhofquai/HB
Bahnhofplatz/HB
Bahnhofstr./HB
Rennweg
Limmatquai
Storchen
Paradeplatz
Kantonalbank
Stockerstr.
Tunnelstr.
Bahnhof Enge
Museum Rietberg
Brunaustr.
Bhf. Enge/Bederstr.
Waffenplatzstr.
Hürlimannplatz
Brandschenkestr.
Sihlcity Nord
Sihlcity
Giesshübel
Saalsporthalle
Binz
Binz Center
Räffelstr.
Manesseplatz
Höfliweg
Schweighof
Friesenberg
Grubenstr.
Friesenbergstr.
Im Hagacker
Hügelstr.
Escher-Gutweg
Waidfussweg
Wipkingerplatz
Lehenstr.
Rosengartenstr.
Bahnhof Wipkingen
Nürenbergstr.
Letten
Okenstr.
Rotbuchstr.
Nordstr.
Lägernstr.
Kronenstr.
Beckenhof
Stampfenbachplatz
Guggachstr.
Laubiweg
Milchbuck
Bucheggplatz
Schaffhauserplatz
Röslistr.
Ottikerstr.
Sonneggstr.
Haldenegg
Central
Polyterrasse
Polybahn
ETH/Universitätsspital
Haldenbach
Winkelriedstr.
Neumarkt
Rudolf-Brun-Brücke
Kunsthaus
Rathaus
Helmhaus
Bahnhof Stadelhofen
Bellevue
Bürkliplatz
Rentenanstalt
Schweizer Rück
Enge Hafen
Opernhaus
Kreuzstr.
Feldeggstr.
Kreuzplatz
Botanischer Garten
Kantonsschule
Hottingerplatz
Sprecherstr.
Englischviertelstr.
Universität Irchel
Strickhof
Langmauerstr.
Letzistr.
Scheuchzerstr.
Kinkelstr.
Schanzackerstr.
Langensteinenstr.
Schäppiweg
Germaniastr.
Rigiblick
Hadlaubstr.
Goldauerstr.
Seilbahn Rigiblick
Vogelsangstr.
Spyriplatz
Bethanien
Hinterbergstr.
Kirche Fluntern
Toblerplatz
Zürichbergstr.
Susenbergstr.
Zoo
Spyristeig
Krönleinstr.
Heubeeriweg
Flobotstr.
Im Klösterli
Bahnhof Stettbach
Mattenhof
Dreiwiesen
Bircher-Benner
Waldhaus Dolder
Hofstr.
Titlisstr.
Klosbach
Bergstation Dolderbahn
Voltastr.
Platte
Römerhof
Dolderbahn
Hölderlinsteig
Hölderlinstr.
Klusplatz
Witikon Zentrum
Buchholz
Carl-Spitteler-Str.
Kapfstr.
Schlyfi
Drusbergstr.
Waserstr.
Freiestr.
Hegibachplatz
Signaustr.
Hedwigsteig
Altenhofstr.
Wonnebergstr.
Wetlistr.
Burgwies
Balgrist
Stodolastr.
Segeten
Wieslacher
Trichtenhausen Fussweg
Friedhof Enzenbühl
Berghaldenstr.

> *"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

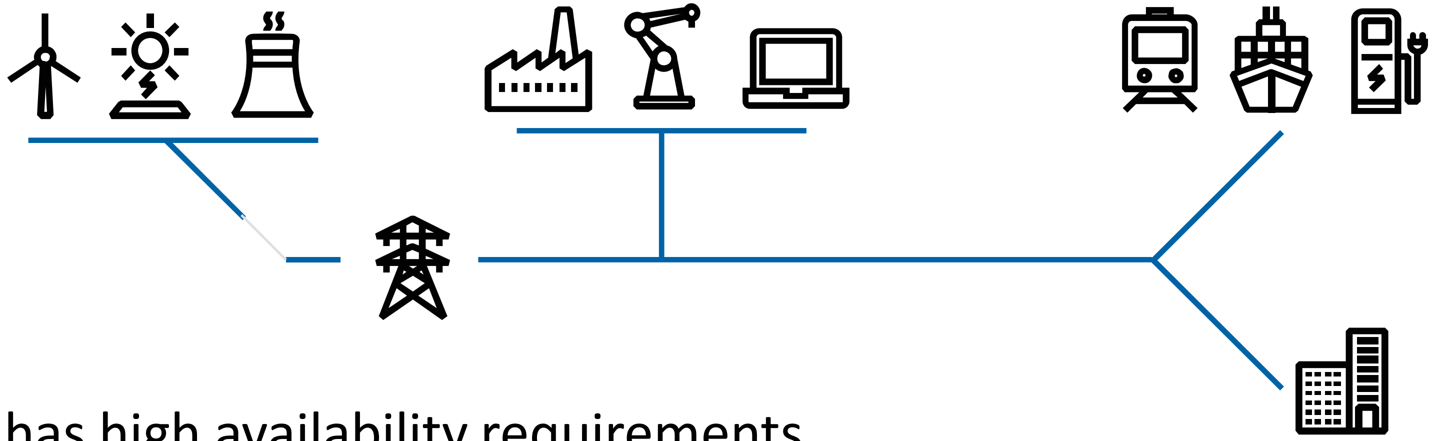


- Critical infrastructure has high availability requirements
- Industrial systems are more and more connected
- Hard real-time requirements

"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?

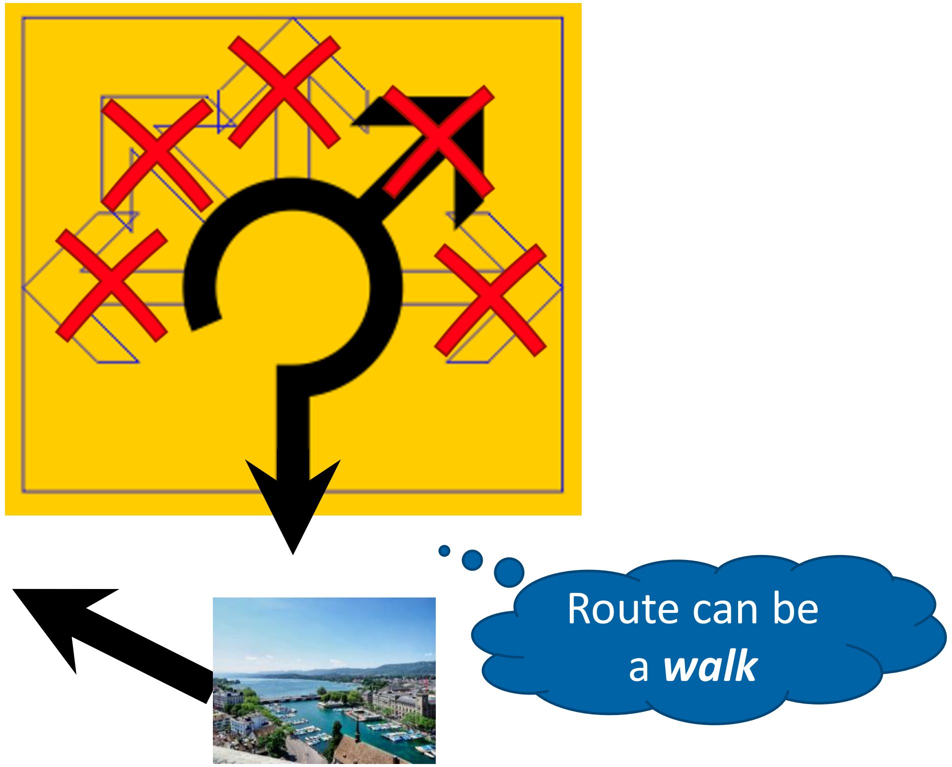
[Content taken from Yvonne-Anne Pignolet]

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



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

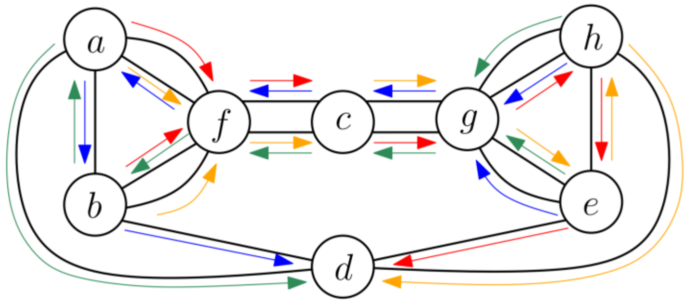
Background: Static Fast Rerouting for Multiple Failures

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

Resiliency & Load on **Complete Graphs**

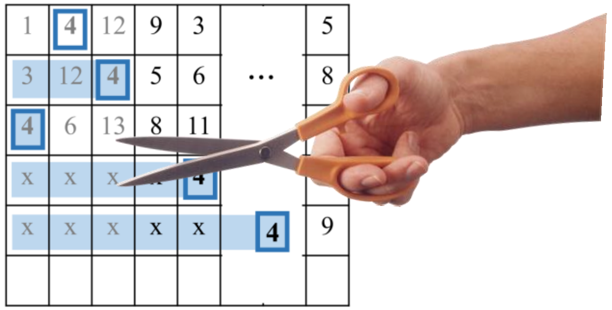
- Borokhovich & Schmid [OPODIS'13]
  - Bounds and handcrafted schemes
- Pignolet *et al.* [DSN'17]
  - Connection to Balanced Incomplete Block Designs (BIBDs)
    - General scheme how to distribute well after failures



From Chiesa *et al.* 2016

Resiliency & **Load** on **General Graphs**

this paper



From Pignolet *et al.* 2017

With improved BIBDs!

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

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

Stretch under ***r*** failures:

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

CASA: Rerouting on Arborescences

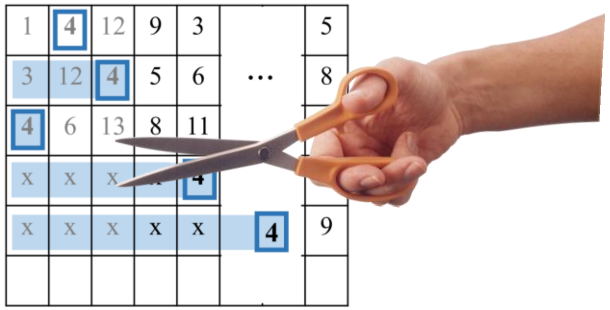
- Takes arborescences as input *e.g*. generated by Chiesa *et al.*
  - Influences the stretch, we get good bounds for *e.g.* so-called *independent spanning trees*

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***

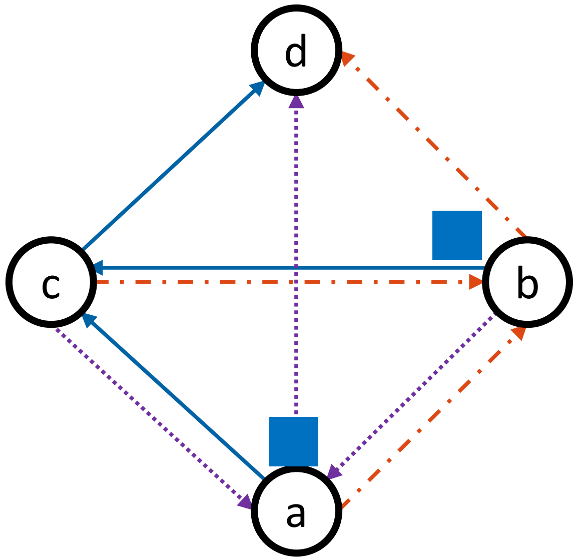


until the next hop is alive

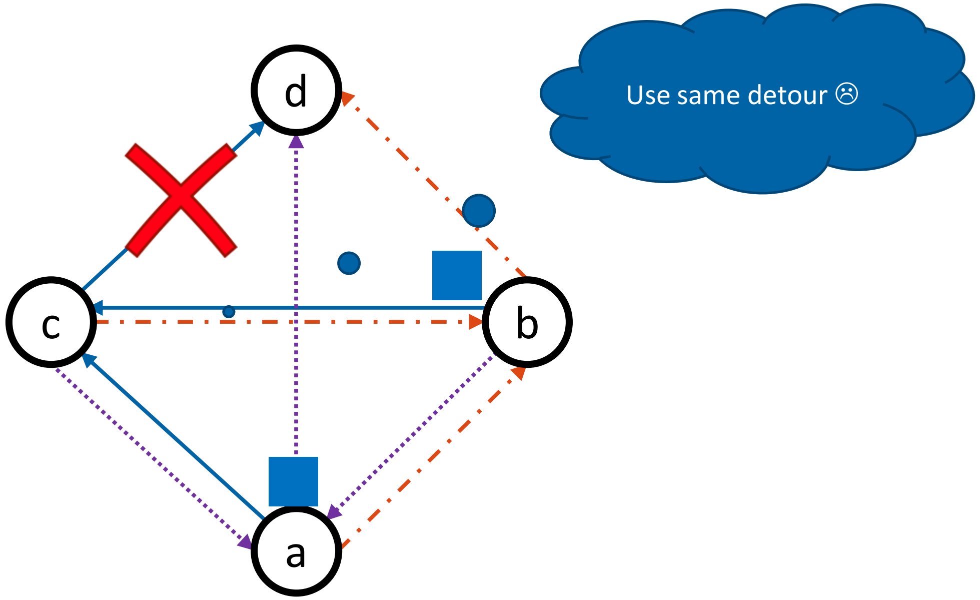
different flows use different *T'*

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

CASA: Example *without* BIBD



CASA: Example *without* BIBD



CASA: Example *with* BIBD

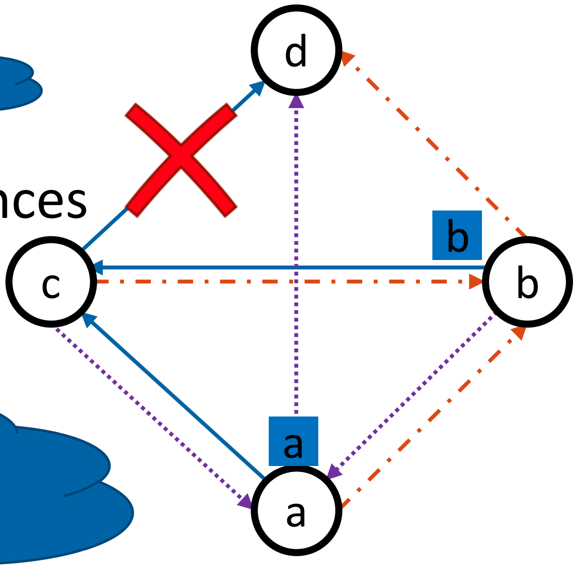
How much extra load?

- Up to $O(\sqrt{r})$

Lower bound: $\sqrt{r}$

- For more flows than #arborescences

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



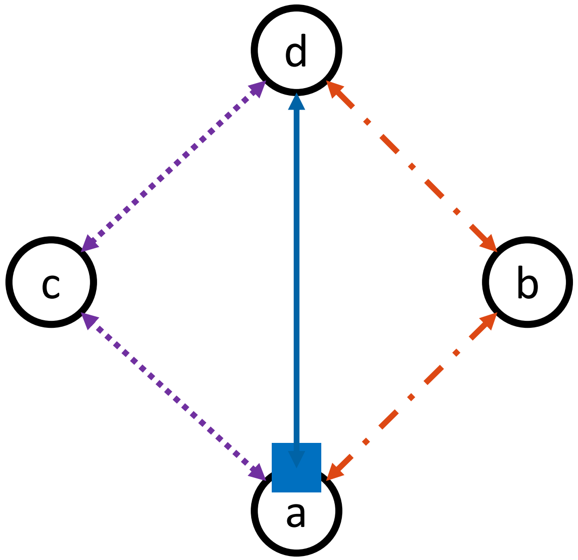
Beyond CASA

- ***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

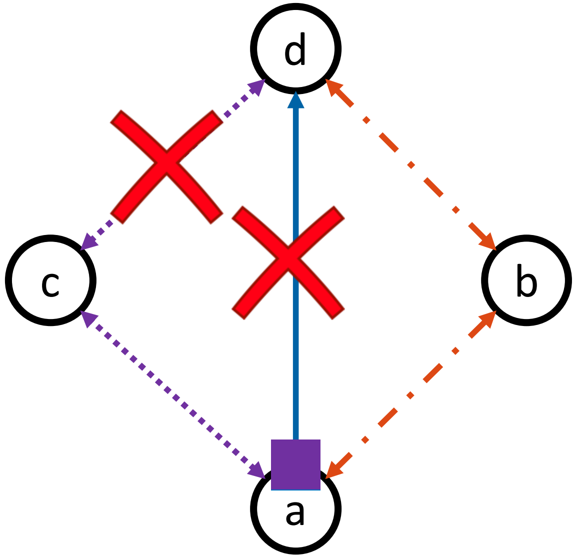
SquareOne



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

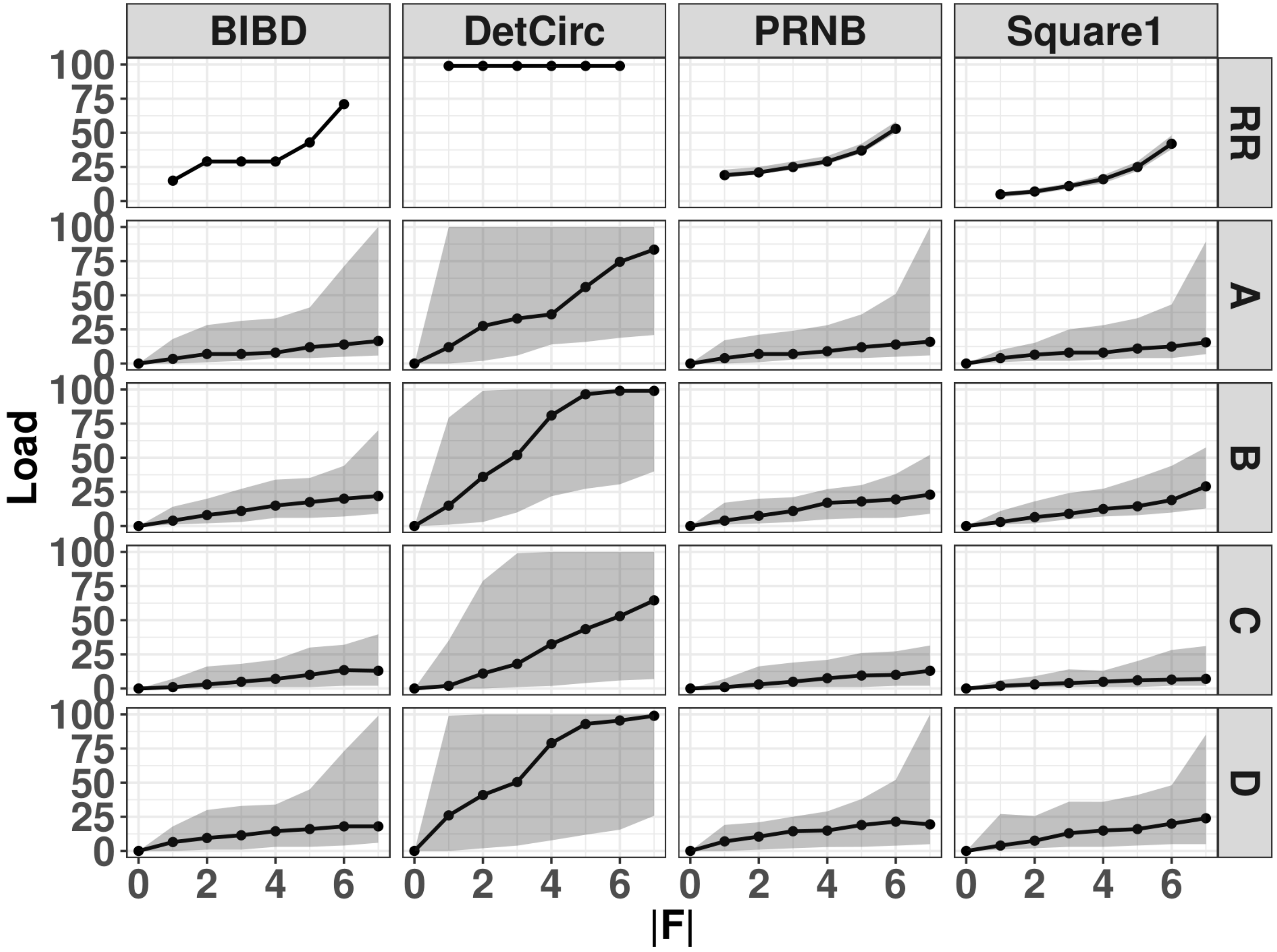
- 8-connected 8-regular random graphs (**RR**, 100 routers each)
- well-connected cores of real-world ASes (*Rocketfuel*) (204-387 routers, 1667-4736 links)

- 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?

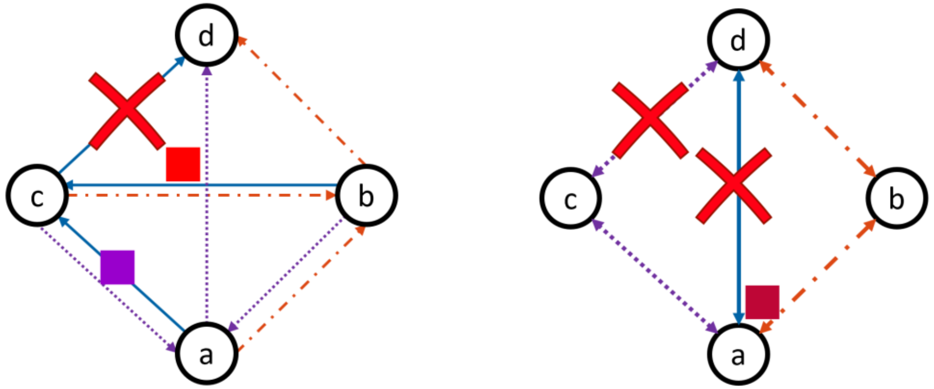
Deterministic Worst-Case Failures



universität wien

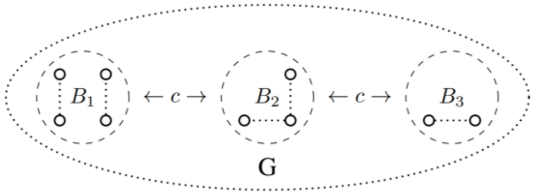
Conclusion

- 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 slide: Further related problems we work on



Some More Related Problems

- Improving arborescence decompositions
  - #1: Build small stretch arborescences in parallel
    - Current approach: build sequentially in greedy fashion
    - Benefit: Resilient to more failures under nice distributions



  - #2: Account for e.g. Shared Risk Link Groups (SRLGs)
    - Leverage post-processing according to objective function
    - Ideally: A SRLG is contained in a single arborescence

Appears at #1: DSN 2019, #2: SRDS 2019

- Allowing packet header modification (MPLS, SR)
  - #1: More powerful, but harder to verify correctness?
    - MPLS w. multiple link failures: verification in polynomial time!

	P-Rex	NetKAT	HSA	VeriFlow	Anteater
Protocol Support	SR/MPLS	OF	Agn.	OF	Agn.
Approach	Autom.	Alg.	Geom.	Tries	SAT
Complexity	Polynom.	PSPACE	Polynom.	NP	NP
Static	✓	✓	✓	χ	✓
Reachability	✓	✓	✓	✓	✓
Loop Queries	✓	✓	✓	✓	✓
What-if	✓	N/A	✓	N/A	χ
Unlim. Header	✓	N/A	χ	χ	N/A
Performance	✓	✓ [1]	✓	✓	✓
Waypointing	✓	✓	✓	✓	χ
Language	Py., C	OCaml	Py., C	Py.	C++, Ruby

  - #2: Leverage Segment Routing (in Linux kernel for IPv6)
    - Allows maximal link protection e.g. in Hypercubes

Appears at #1: CoNEXT 2018, #2: OPODIS 2018

Papers

- *Improved Fast Rerouting Using Postprocessing*
  Klaus-T. Foerster, Andrzej Kamisinski, Yvonne-Anne Pignolet, Stefan Schmid, and Gilles Tredan. *SRDS 2019*
- *Bonsai: Efficient Fast Failover Routing Using Small Arborescences*
  Klaus-T. Foerster, Andrzej Kamisinski, Yvonne-Anne Pignolet, Stefan Schmid, and Gilles Tredan. *DSN 2019*
- *CASA: Congestion and Stretch Aware Static Fast Rerouting*
  Klaus-T. Foerster, Yvonne-Anne Pignolet, Stefan Schmid, and Gilles Tredan. *INFOCOM 2019*
- *P-Rex: Fast Verification of MPLS Networks with Multiple Link Failures*
  Jesper S. Jensen, Troels B. Krogh, Jonas S. Madsen, S. Schmid, Jiri Srba, and Marc T. Thorgersen. *CoNEXT 2018*
- *Local Fast Segment Rerouting on Hypercubes*
  Klaus-T. Foerster, Mahmoud Parham, Stefan Schmid, and Tao Wen. *OPODIS 2018*

universität
wien

Congestion and Stretch Aware Static Fast Rerouting [appeared @INFOCOM'19]

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

Papers Referenced

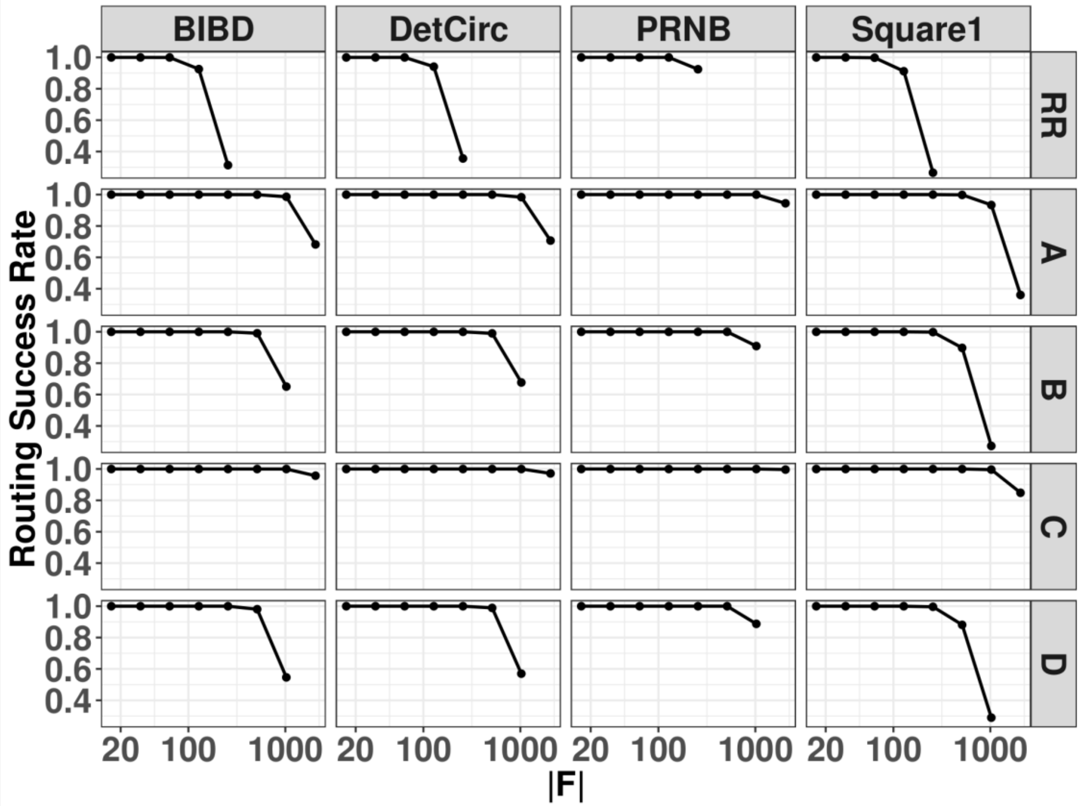
- *How (Not) to Shoot in Your Foot with SDN Local Fast Failover: A Load-Connectivity Tradeoff*
  Michael Borokhovich and Stefan Schmid. *OPODIS 2013*
- *Load-Optimal Local Fast Rerouting for Dependable Networks*
  Yvonne-Anne Pignolet, Stefan Schmid, and Gilles Tredan. *DSN 2013*
- *IP Fast Rerouting 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*

Rocketfuel ASes

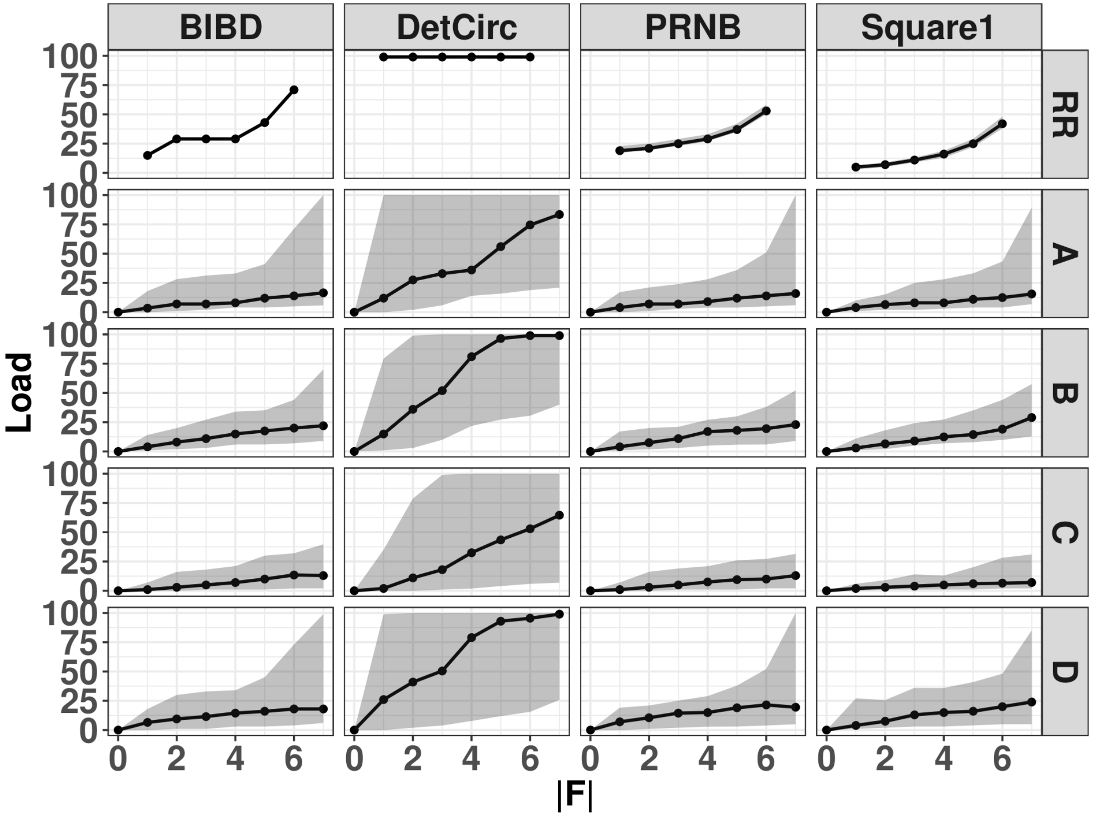
AS	1239 A	2914 B	3356 C	7018 D
Number of nodes	389	225	377	204
Number of links	3621	1696	4736	1667
Eccentricity	6	6	6	6
Avg shortest path length	3.06	2.48	3.14	3.17

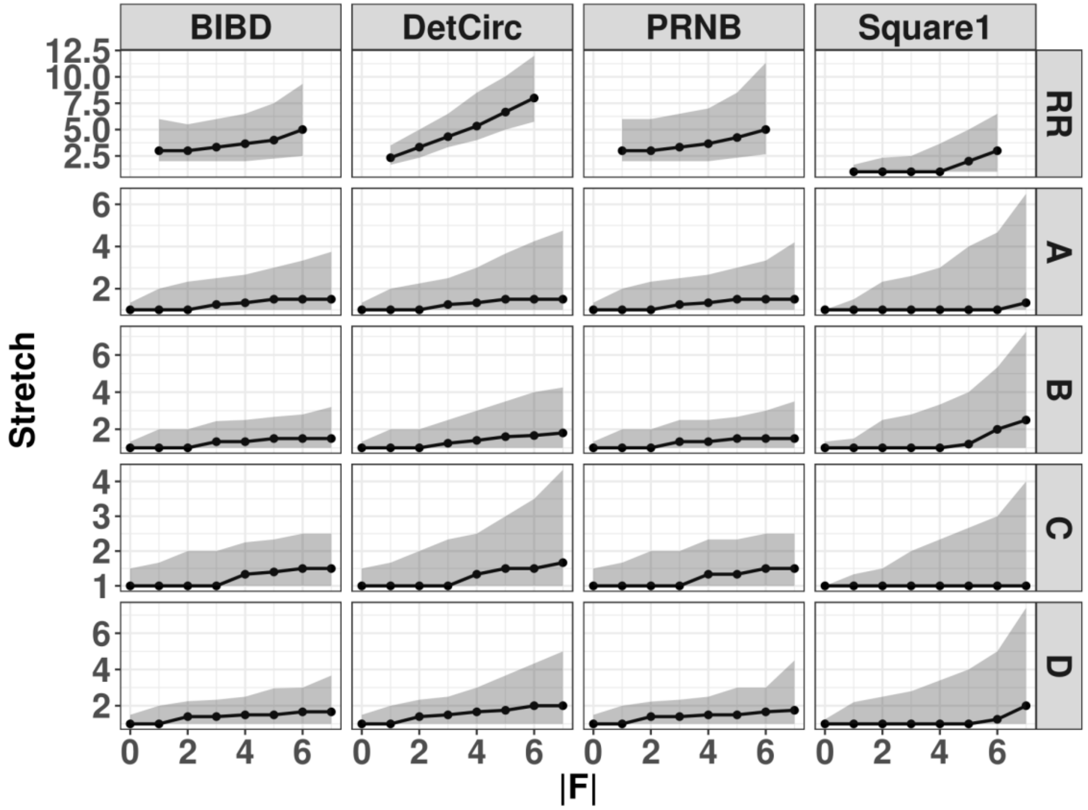
TABLE I: Properties of 8-connected cores of various ASes

Evaluation: Resiliency



Evaluation: Deterministic Worst-Case Failures





Evaluation: Random Failures