ReNets: Statically-Optimal Demand-Aware Networks

Chen Avin and Stefan Schmid

“We cannot direct the wind,
but we can adjust the sails.”

(Folklore)

Acknowledgements:

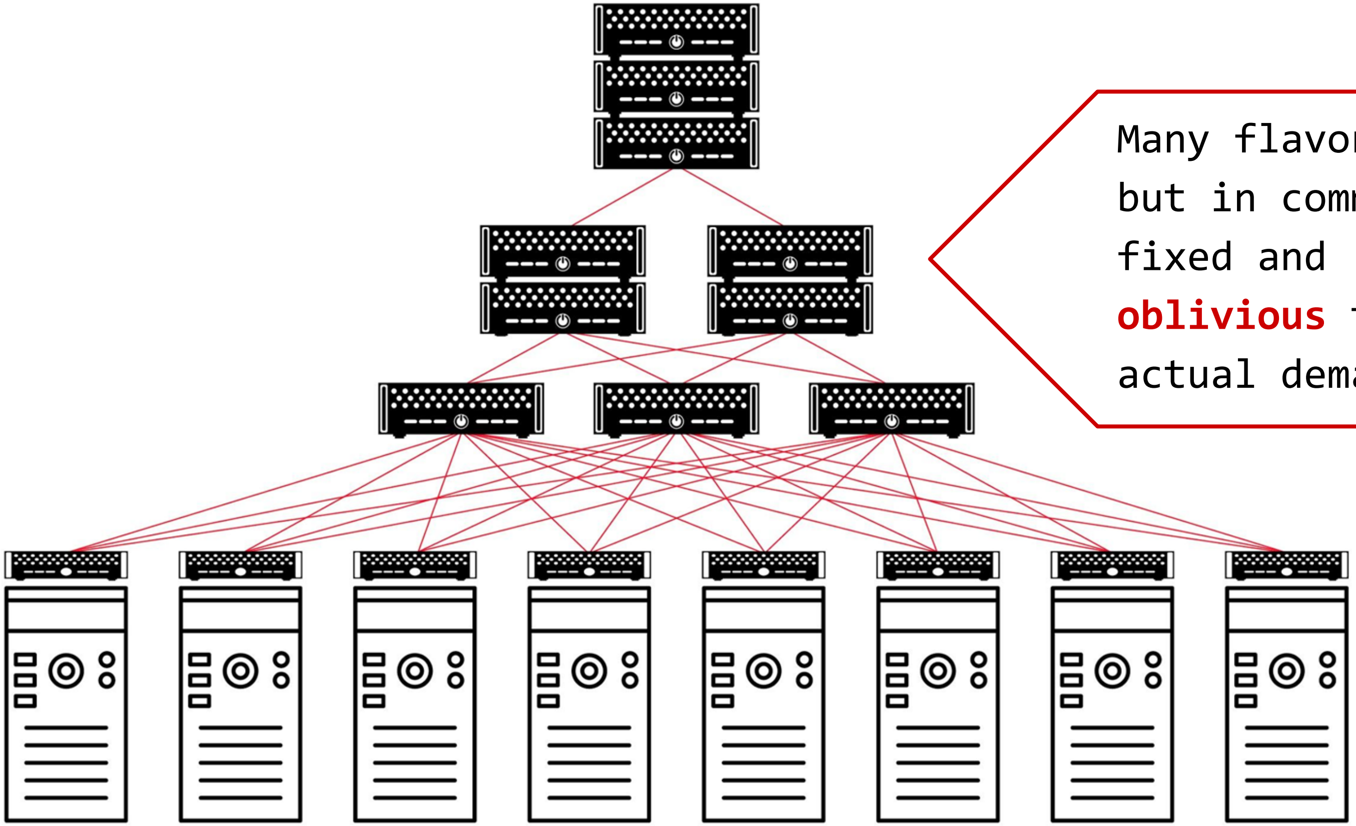
German-Israeli Foundation for Scientific Research and Development

European Research Council
Established by the European Commission

universität wien

Today’s Datacenters

Fixed and Demand-Oblivious Topology

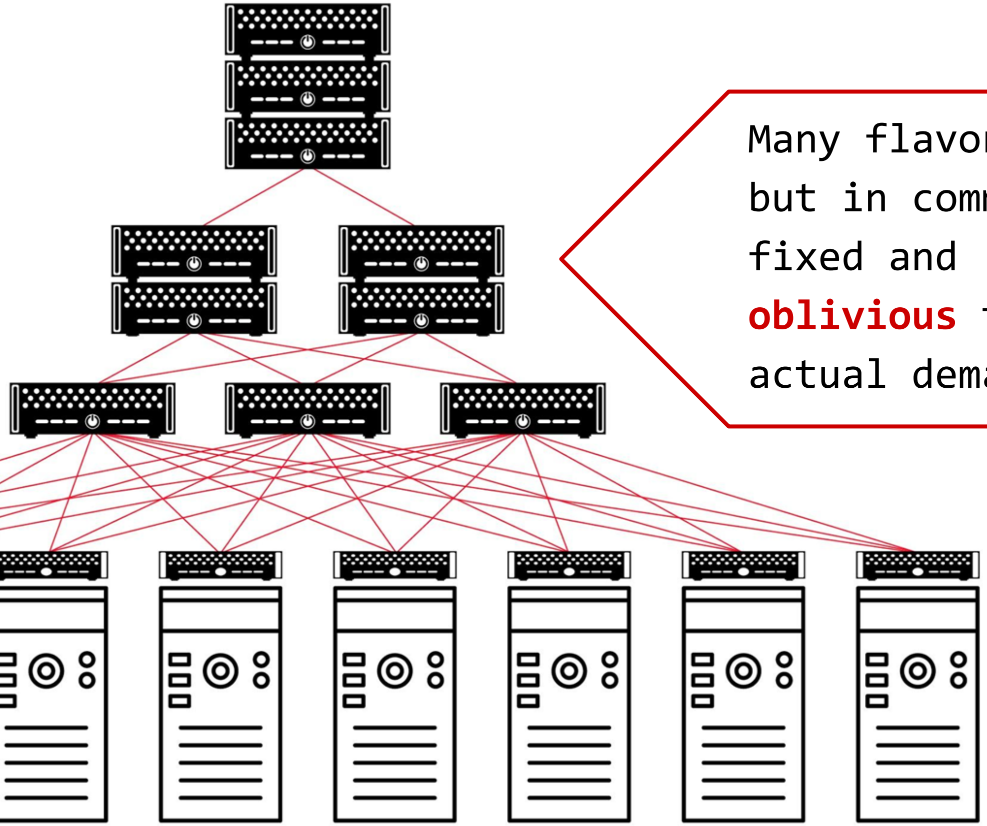


Many flavors, but in common: fixed and **oblivious** to actual demand.

Today's Datacenters

Fixed and Demand-Oblivious Topology

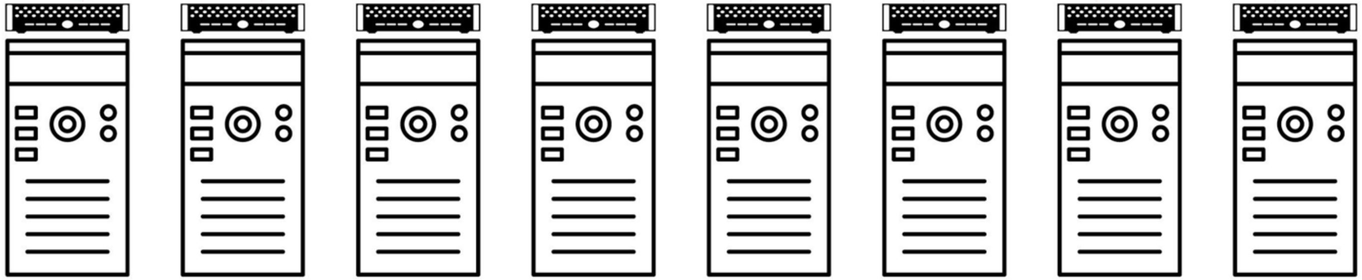
Highway which ignores actual traffic: frustrating!



Many flavors, but in common: fixed and **oblivious** to actual demand.

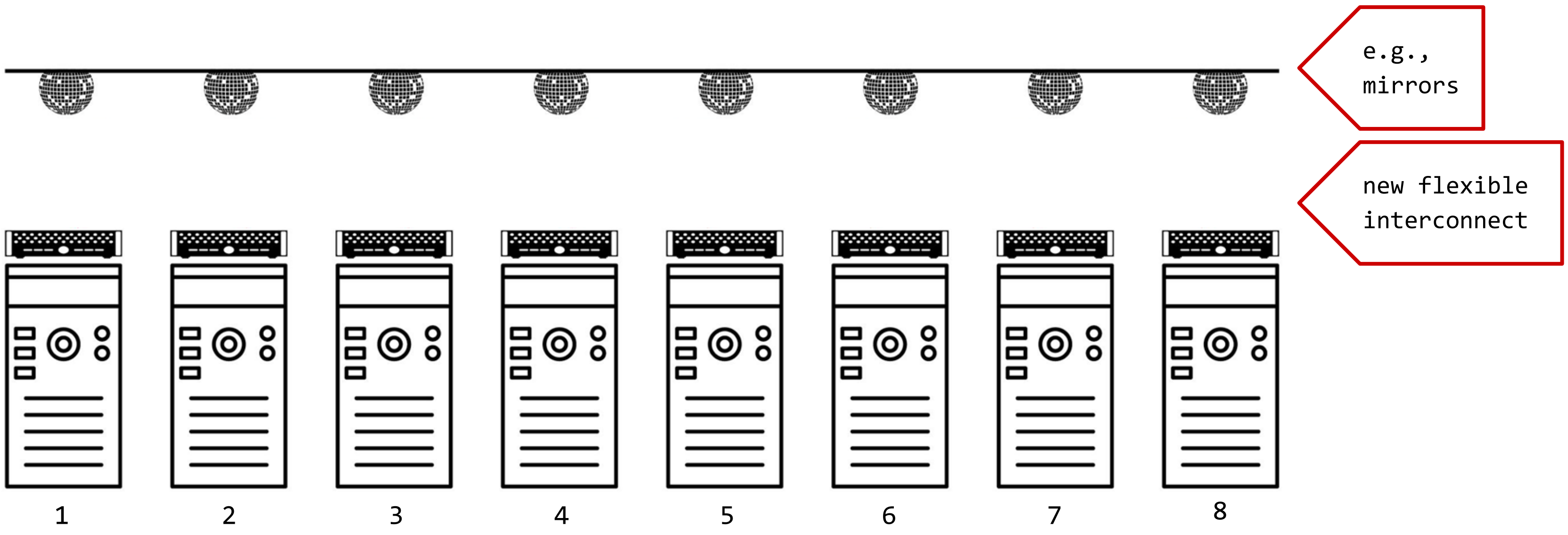
Our Vision

Flexible and Demand-Aware Topologies



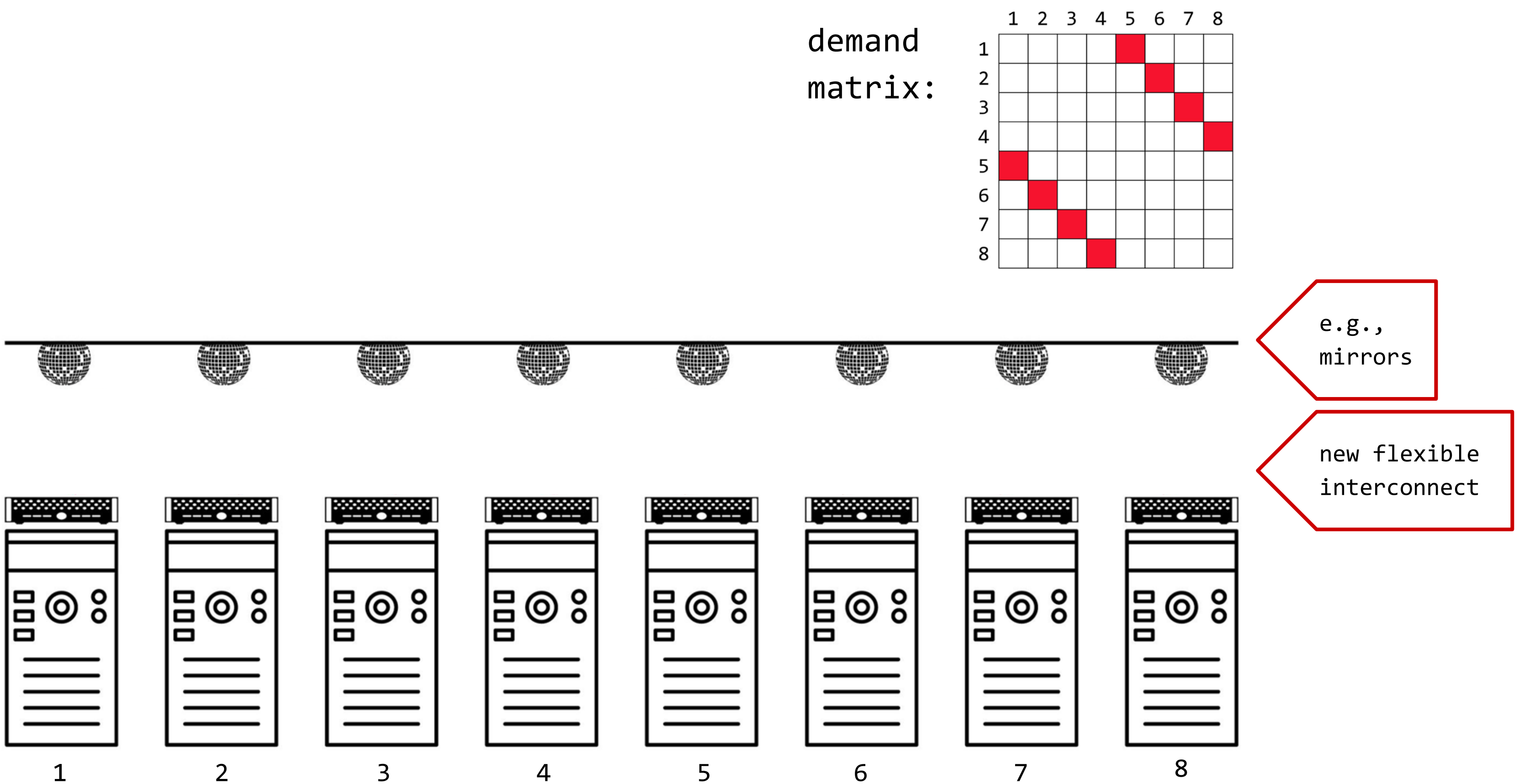
Our Vision

Flexible and Demand-Aware Topologies



Our Vision

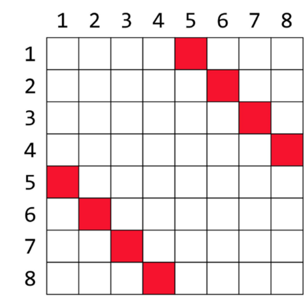
Flexible and Demand-Aware Topologies



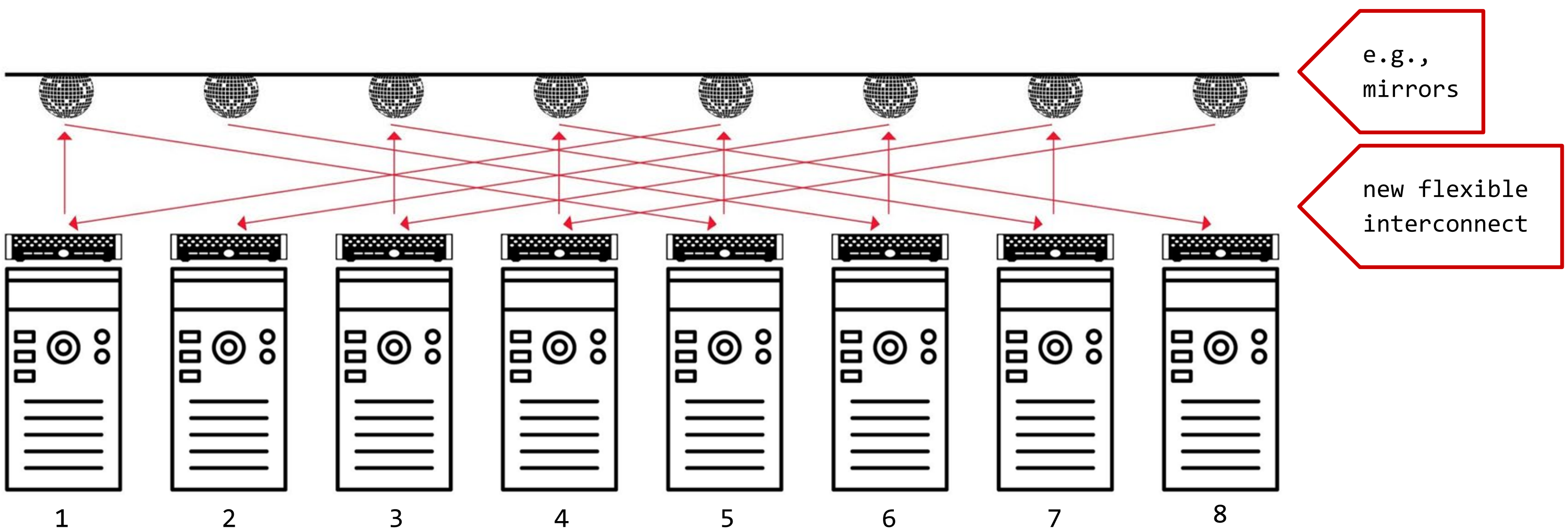
Our Vision

Flexible and Demand-Aware Topologies

demand matrix:

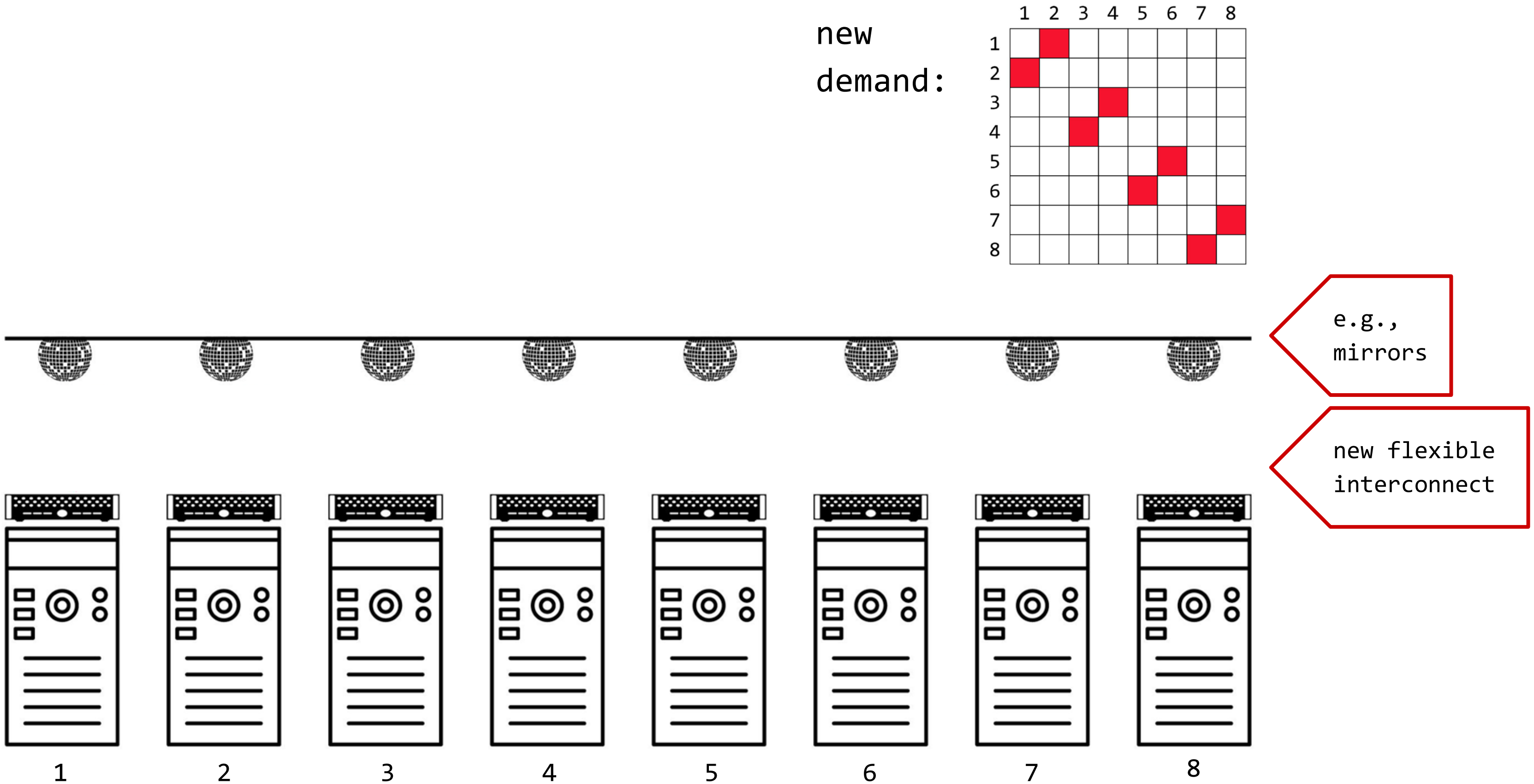


Matches demand



Our Vision

Flexible and Demand-Aware Topologies

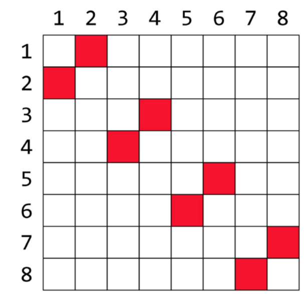


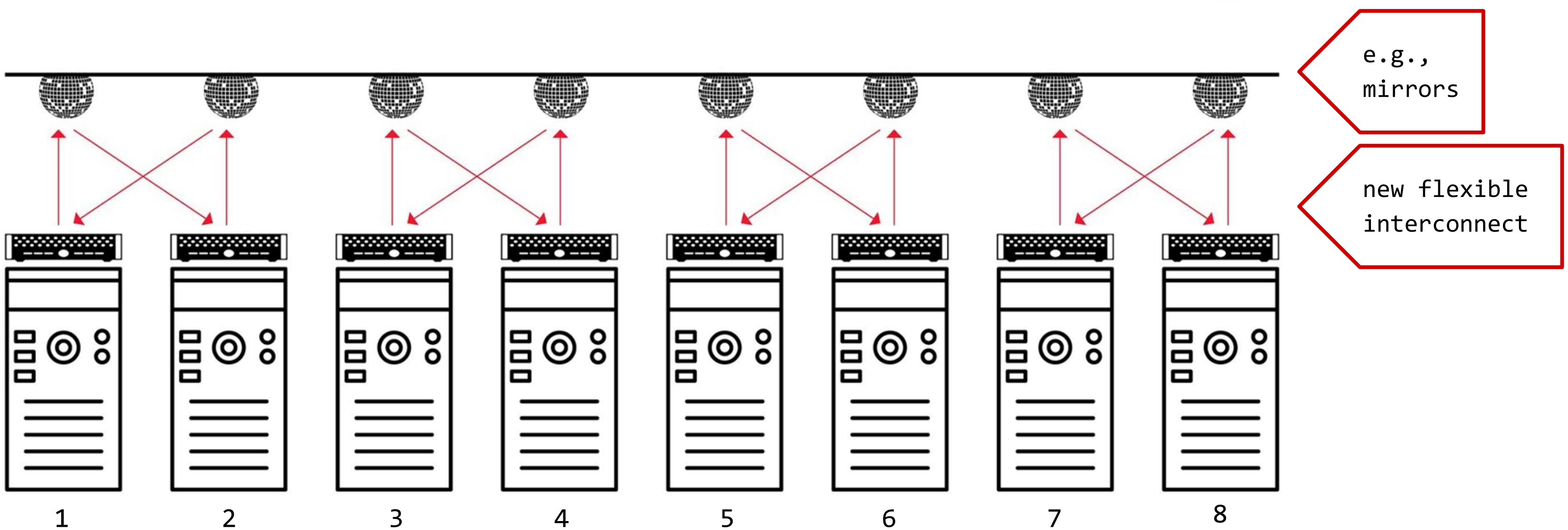
Our Vision

Flexible and Demand-Aware Topologies

Matches demand

new demand:



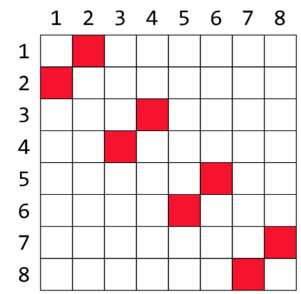


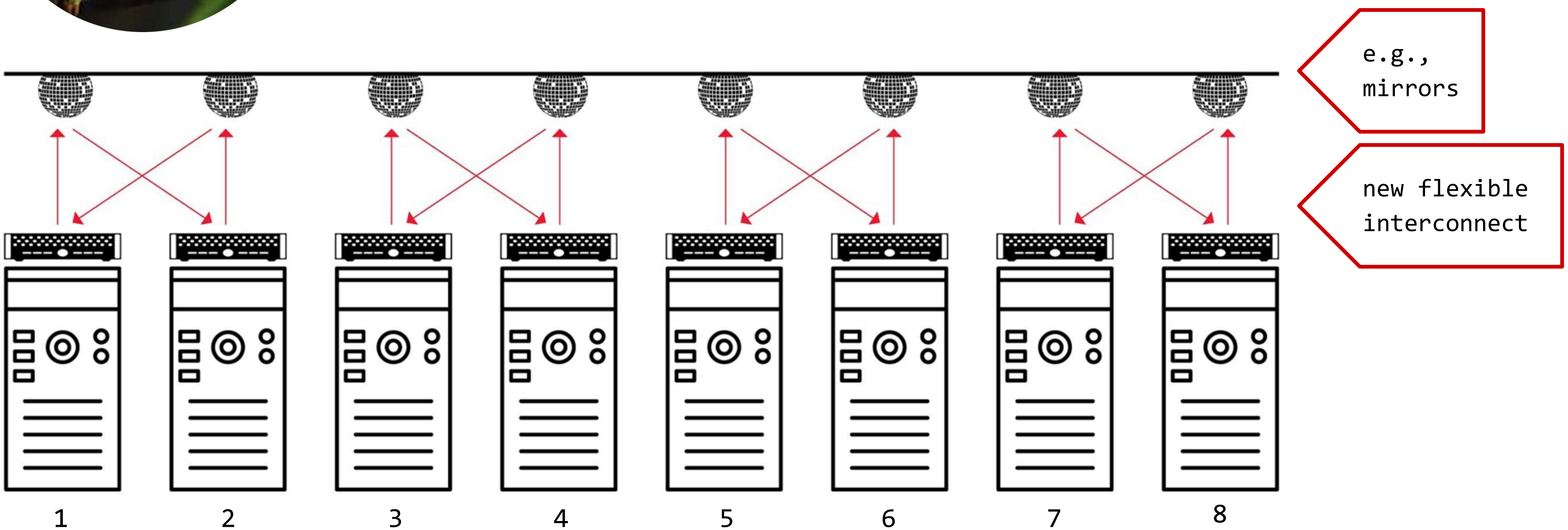
Our Vision

Flexible and Demand-Aware Topologies

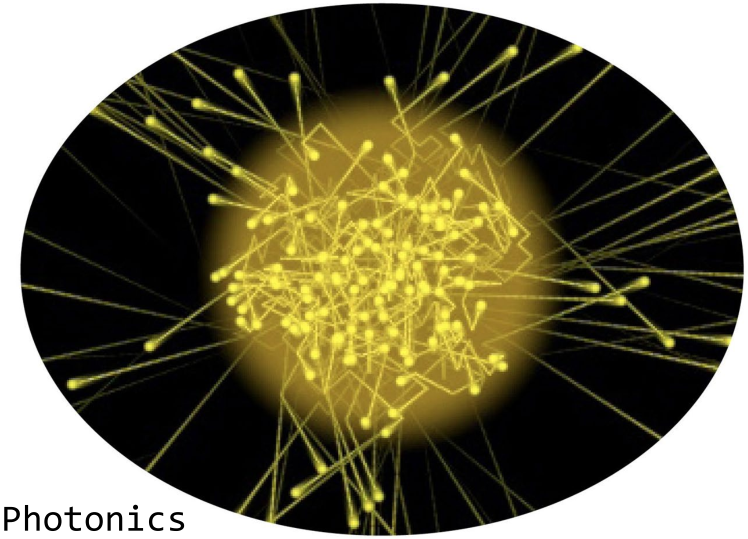
Self-Adjusting Networks

new demand:





Sounds Crazy?
Emerging Enabling Technology.



Photonics

H2020:
“Photonics one of only five key enabling technologies for future prosperity.”

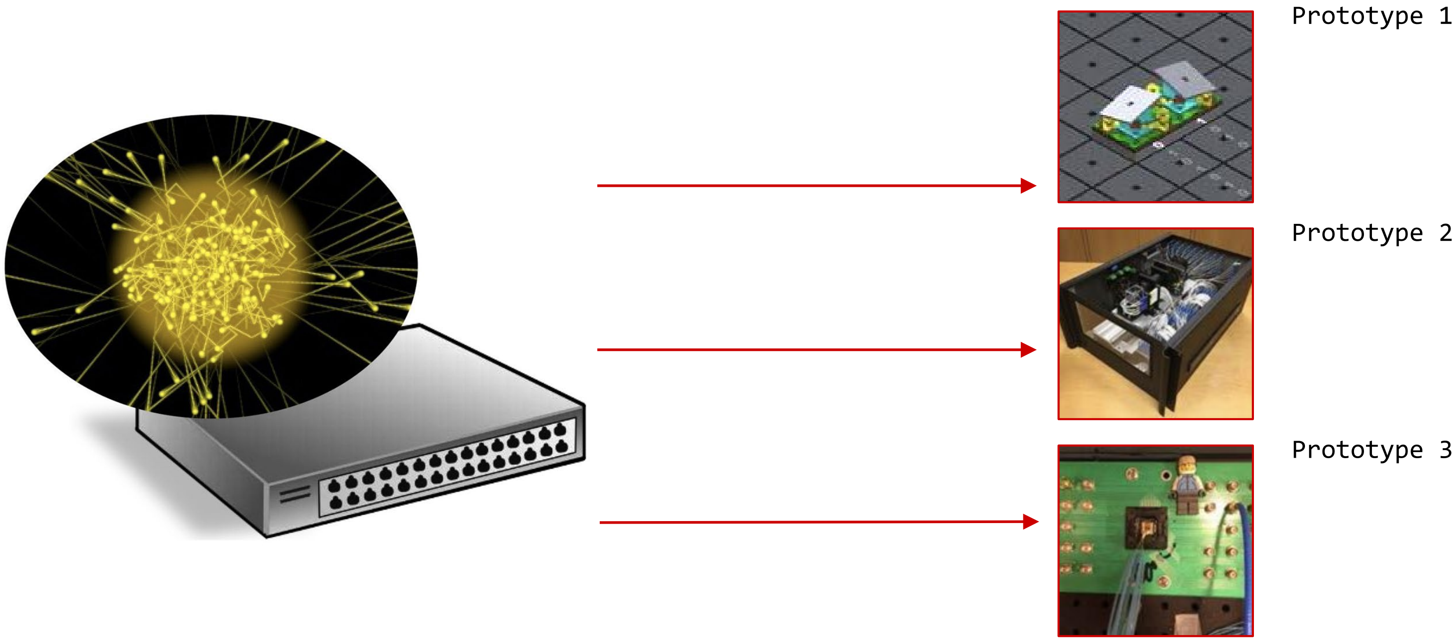
US National Research Council:
“Photons are the new Electrons.”

Enabler

Novel Reconfigurable Optical Switches

⇢ **Spectrum** of prototypes

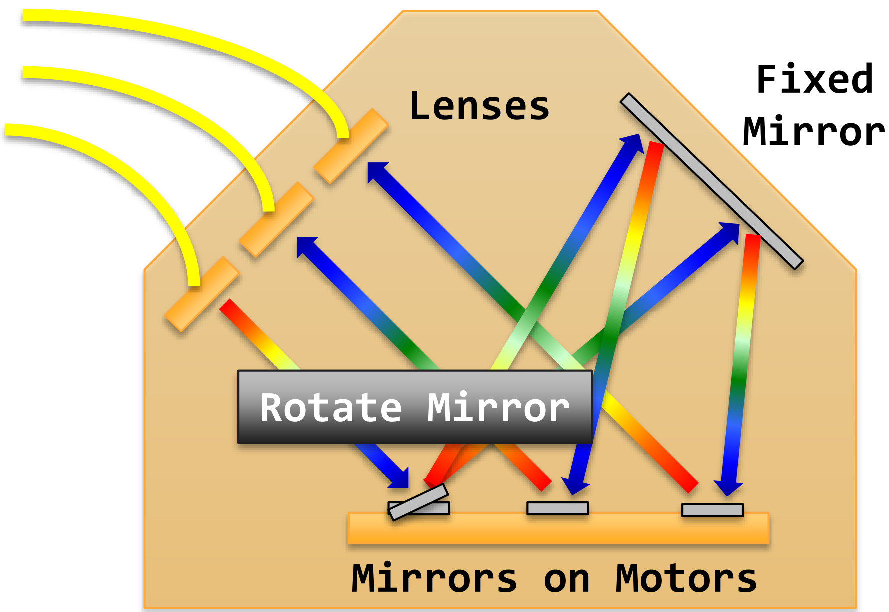
- ⇀ Different sizes, different reconfiguration times
- ⇀ From our recent ACM SIGCOMM **OptSys'19** workshop



Example

Optical Circuit Switch

⇢ Optical Circuit Switch rapid adaption of physical layer
  → based on rotating mirrors

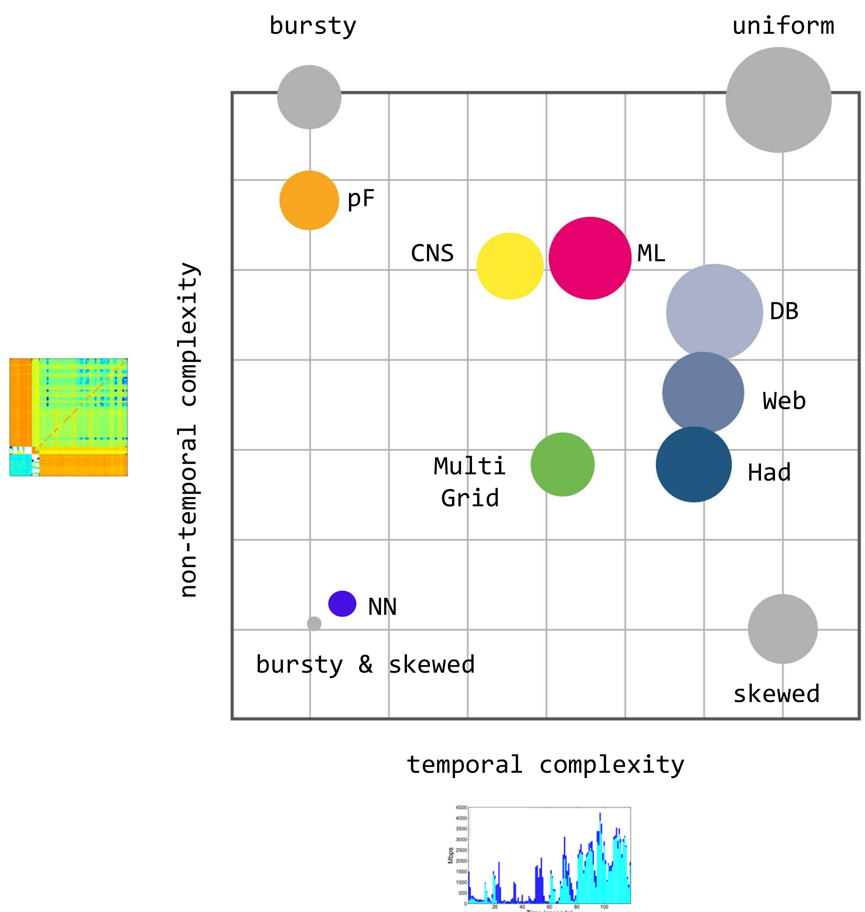


Optical Circuit Switch

By Nathan Farrington, SIGCOMM 2010

Empirical Motivation

Temporal and Spatial Structure



No structure

Avin et al. SIGMETRICS 2020

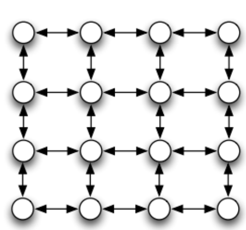
Source: Avin et al., On the Complexity of Traffic Traces and Implications, SIGMETRICS 2020

The Potential

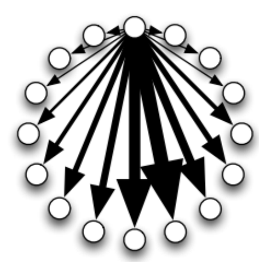
Example: Expected Route Length

- **Expected path length**: number of hops times demand
- Consider design of **constant degree** topologies (e.g., 4)
- Note: diameter is at least **logarithmic**

Demand 1: Low Degree



Demand 2: Skewed



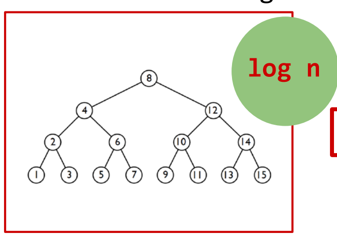
Expected route length in demand-aware network is **constant** in these cases (while diameter is Ω(log n)).

Connection to Entropy

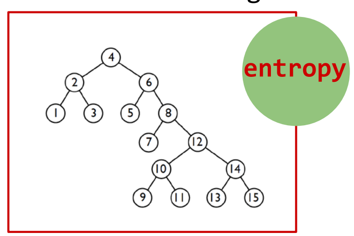
- Achievable expected route length is proportional to **conditional entropy** of the demand matrix
- Similar to coding and data structures:

Avin et al. DISC 2017

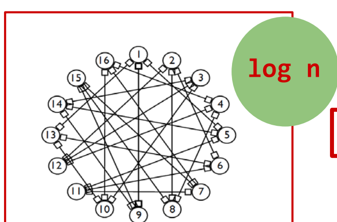
Traditional BST / worst-case coding



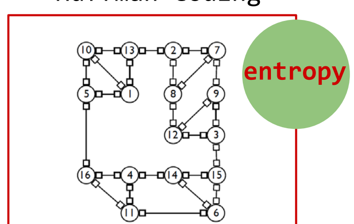
Demand-aware BST / Huffman coding



Traditional networks / worst-case traffic



Demand-aware BST / Huffman coding



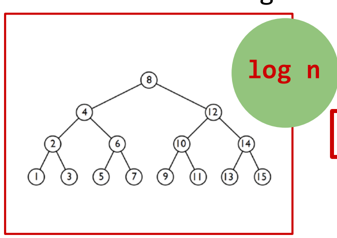
Connection to Entropy

- Achievable expected route length is proportional to **conditional entropy** of the demand matrix

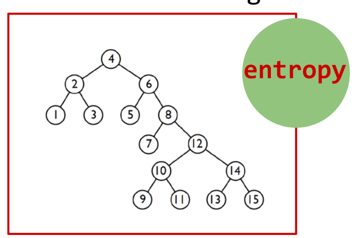
- Similar to coding and data structures:

Avin et al. DISC 2017

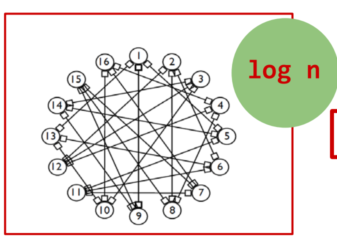
Traditional BST / worst-case coding



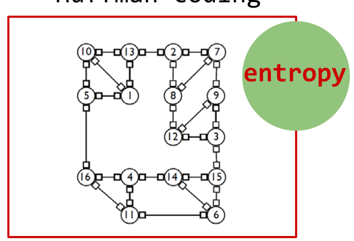
Demand-aware BST / Huffman coding



Traditional networks / worst-case traffic



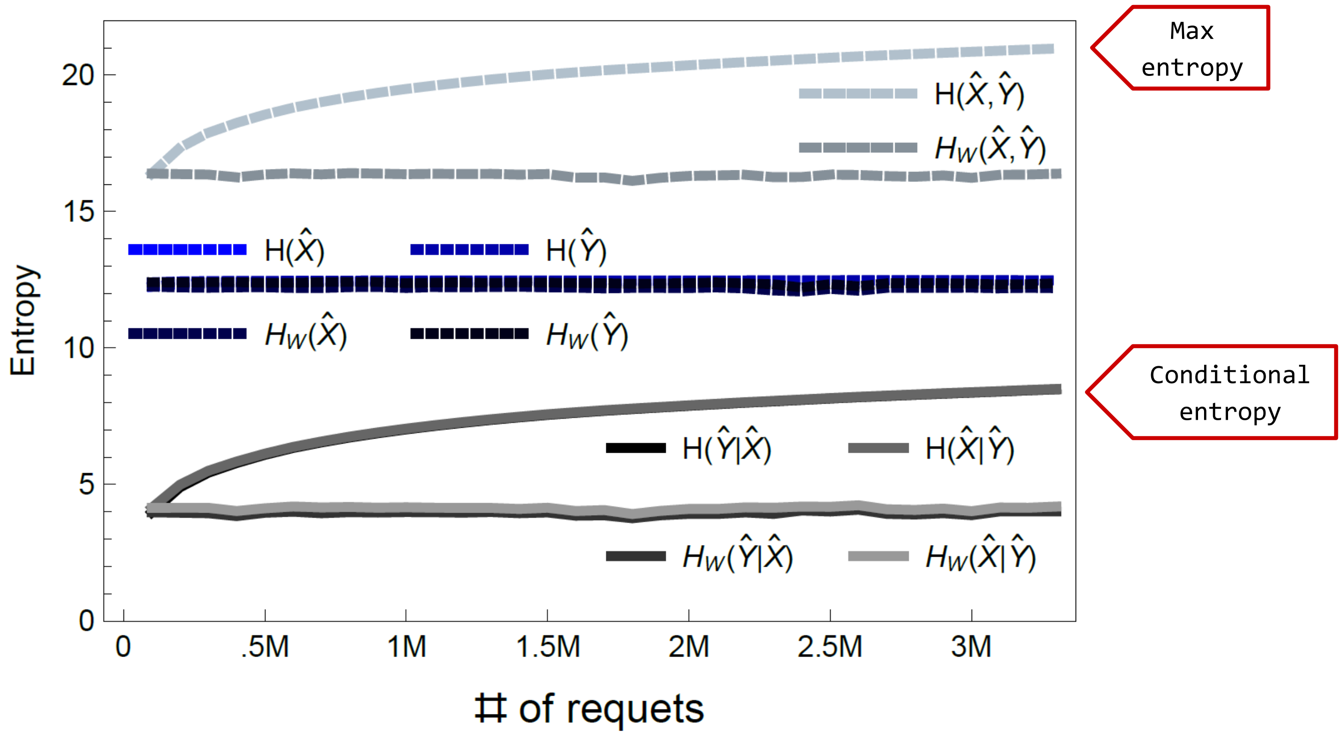
Demand-aware BST / Huffman coding



But how to achieve short routes if the demand is not known ahead of time and we have to account for reconfiguration costs?

Example

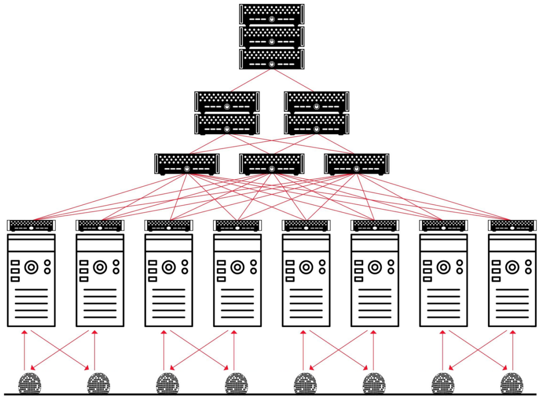
Facebook's Datacenter Traces



Our Contribution:

ReNet, A Statically Optimal Demand-Aware Network

- Model: **hybrid architecture**
  - Fixed network of diameter log n plus reconfigurable network (**constant** number of direct links)
  - **Segregated** routing
  - **Online** sequence of requests: σ = (σ1, σ2, σ3, ...)
  - Global controller



fixed

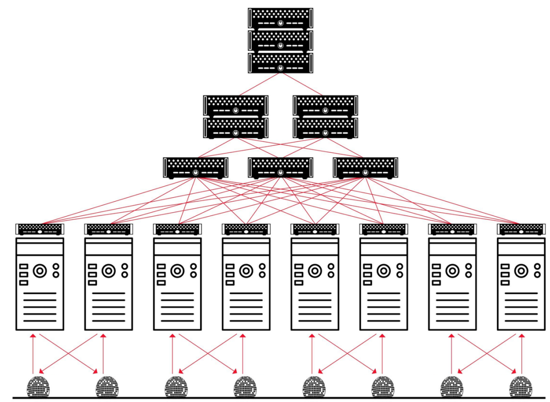
reconfigurable

- **Objective:** Minimize route length plus reconfigurations
  - More specifically: be **statically optimal**
  - Compared to a fixed algorithm which knows σ ahead of time

Our Contribution:

ReNet, A Statically Optimal Demand-Aware Network

- Model: **hybrid architecture**
  - Fixed network of diameter log n plus reconfigurable network (**constant** number of direct links)
  - **Segregated** routing
  - **Online** sequence of requests: σ = (σ1, σ2, σ3, ...)
  - Global controller



fixed

reconfigurable

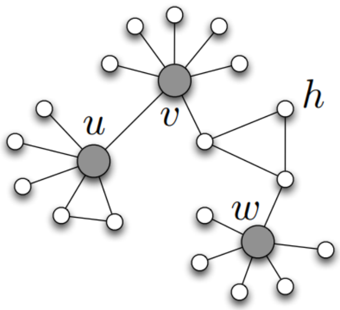
- **Objective:** Minimize route length plus reconfigurations
  - More specifically: be **statically optimal**
  - Compared to a fixed algorithm which knows σ ahead of time

- Compact routing (constant tables)
- Local routing (greedy)
- Arbitrary addressing

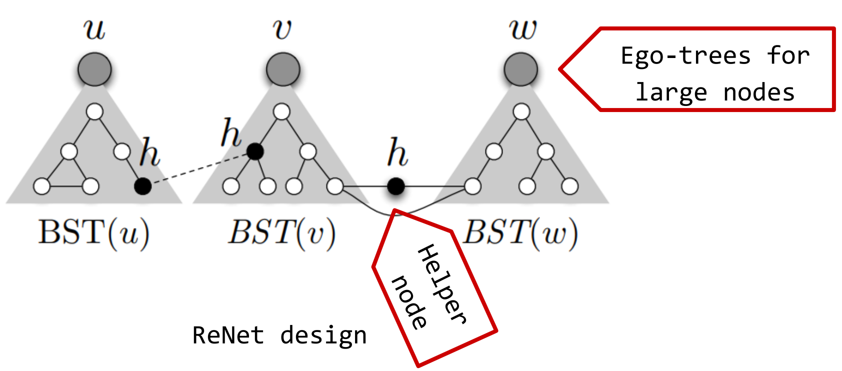
The ReNet Algorithm (1)

Algorithmic building blocks:

1. **Working Set** (WS)
   → Nodes keep track of recent communication partners in σ.
2. Small/large nodes and **Ego-Tree**
   → Nodes with small WS connect to WS directly, nodes with large WS via a self-adjusting binary search tree (e.g., a **splay tree**)
3. **Helper nodes** to reduce the degree
   → Large nodes may appear in many ego-trees, so get help of small nodes



Demand graph

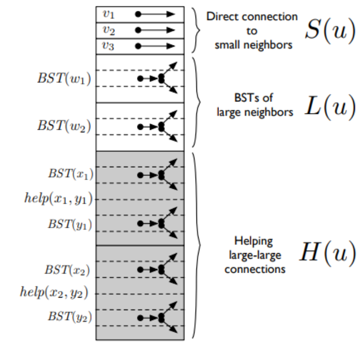


ReNet design

The ReNet Algorithm (2)

Continued:

4. **Self adjustments**
   - Keep track of WS; when too large: **flush-when-full**
5. Centralized coordination
   - Fairly **decentralized**: coordinator only needs to keep track of which nodes are large and which small
   - Nodes inform coordinator when adding node to working set
   - Coordinator then assigns helper node on demand



Analytical Results (1)

Theorem 1:

For any sparse communication sequence of a certain length, ReNets are statically optimal while ensuring a bounded degree.

⇢ Sparse: subsequences of only involve a linear number of nodes

⇢ Required to ensure availability of helper nodes (DISC 2017)

Analytical Results (2)

Theorem 2:

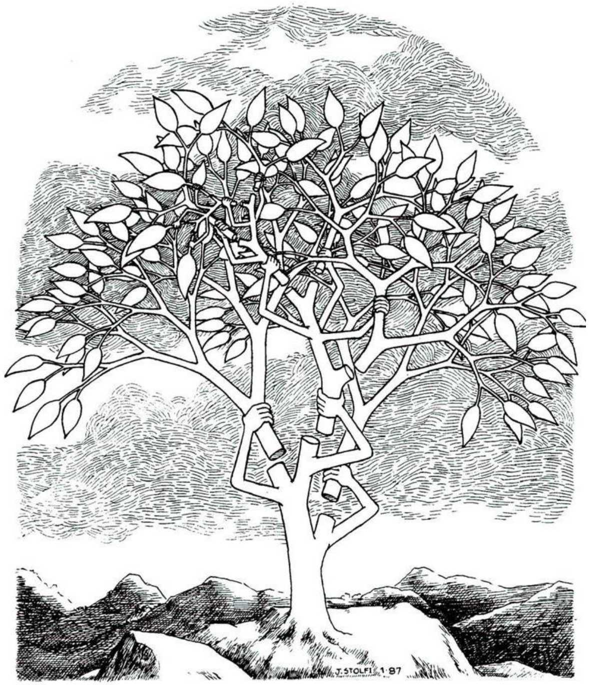
Under certain communication patterns, the amortized cost of ReNet can be significantly lower than the static optimum, i.e., $\Omega(\log n)$.

- Example: consider sequence of $\sigma = (\sigma^{(1)}, \sigma^{(2)}, \sigma^{(3)}, ...)$ where each $\sigma^{(i)}$ is of length n log n, sparse and corresponds to different **2-dimensional grid**.
- In this example, the cost of ReNet is **constant** for each $\sigma^{(i)}$.
- Overall, the union of the grids form a uniform pattern, so the cost of the static algorithm is **log n** (for constant degree).

Conclusion

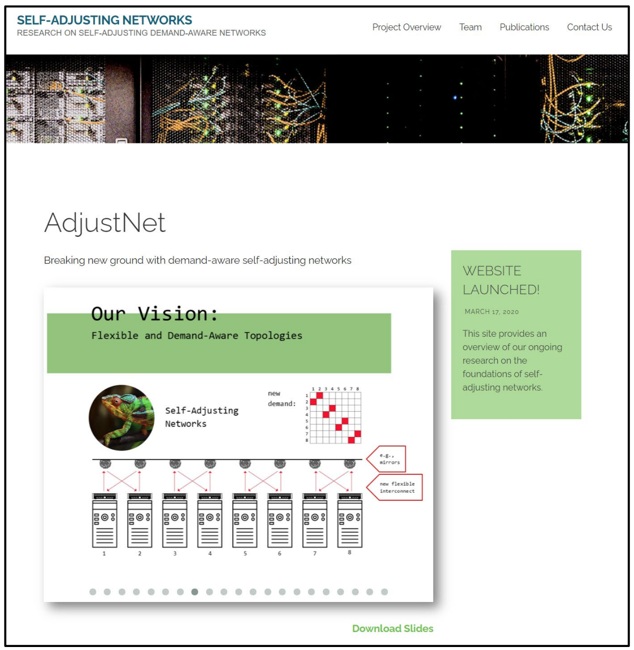
- ReNet: statically optimal and
  - compact routing
  - local routing
  - arbitrary addressing
- Avenues for future work
  - dense communication
  - dynamic optimality

Thank you!

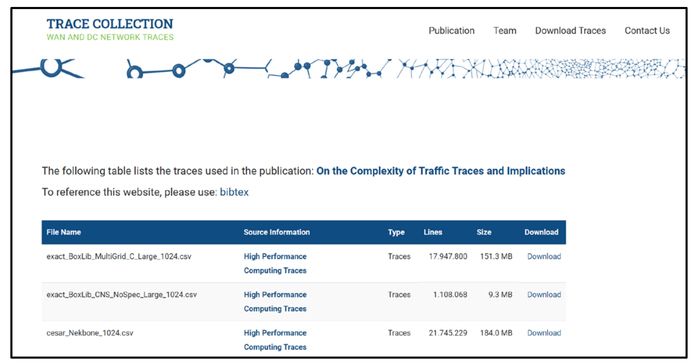


A Self-Adjusting Search Tree
by Jorge Stolfi (1987)

Websites



http://self-adjusting.net/
Project website



https://trace-collection.net/
Trace collection website

Selected References

On the Complexity of Traffic Traces and Implications
Chen Avin, Manya Ghobadi, Chen Griner, and Stefan Schmid.
ACM SIGMETRICS, Boston, Massachusetts, USA, June 2020.

Survey of Reconfigurable Data Center Networks: Enablers, Algorithms, Complexity
Klaus-Tycho Foerster and Stefan Schmid.
SIGACT News, June 2019.

Toward Demand-Aware Networking: A Theory for Self-Adjusting Networks (Editorial)
Chen Avin and Stefan Schmid.
ACM SIGCOMM Computer Communication Review (**CCR**), October 2018.

Measuring the Complexity of Network Traffic Traces
Chen Griner, Chen Avin, Manya Ghobadi, and Stefan Schmid.
arXiv, 2019.

Demand-Aware Network Design with Minimal Congestion and Route Lengths
Chen Avin, Kaushik Mondal, and Stefan Schmid.
38th IEEE Conference on Computer Communications (**INFOCOM**), Paris, France, April 2019.

Distributed Self-Adjusting Tree Networks
Bruna Peres, Otavio Augusto de Oliveira Souza, Olga Goussevskaia, Chen Avin, and Stefan Schmid.
38th IEEE Conference on Computer Communications (**INFOCOM**), Paris, France, April 2019.

Efficient Non-Segregated Routing for Reconfigurable Demand-Aware Networks
Thomas Fenz, Klaus-Tycho Foerster, Stefan Schmid, and Anaïs Villedieu.
IFIP Networking, Warsaw, Poland, May 2019.

DaRTree: Deadline-Aware Multicast Transfers in Reconfigurable Wide-Area Networks
Long Luo, Klaus-Tycho Foerster, Stefan Schmid, and Hongfang Yu.
IEEE/ACM International Symposium on Quality of Service (**IWQoS**), Phoenix, Arizona, USA, June 2019.

Demand-Aware Network Designs of Bounded Degree
Chen Avin, Kaushik Mondal, and Stefan Schmid.
31st International Symposium on Distributed Computing (**DISC**), Vienna, Austria, October 2017.

SplayNet: Towards Locally Self-Adjusting Networks
Stefan Schmid, Chen Avin, Christian Scheideler, Michael Borokhovich, Bernhard Haeupler, and Zvi Lotker.
IEEE/ACM Transactions on Networking (**TON**), Volume 24, Issue 3, 2016. Early version: IEEE **IPDPS** 2013.

Characterizing the Algorithmic Complexity of Reconfigurable Data Center Architectures
Klaus-Tycho Foerster, Monia Ghobadi, and Stefan Schmid.
ACM/IEEE Symposium on Architectures for Networking and Communications Systems (**ANCS**), Ithaca, New York, USA, July 2018.