RADWAN

Rate Adaptive Wide Area Networks

Rachee Singh / U. Massachusetts Amherst (now at MSR)

Manya Ghobadi / Microsoft Research (now at MIT)

Klaus-T. Foerster / University of Vienna

Mark Filer / Microsoft Research

Phillipa Gill / U. Massachusetts Amherst

The paper is available at https://dl.acm.org/doi/10.1145/3230543.3230570 . Slides based on those from Rachee Singh. Thanks a lot!

O(100) datacenters

Dedicated Wide Area Network

Costs O(100) million dollars per year

Achieving High Utilization with Software-Driven WAN

[SIGCOMM '13]

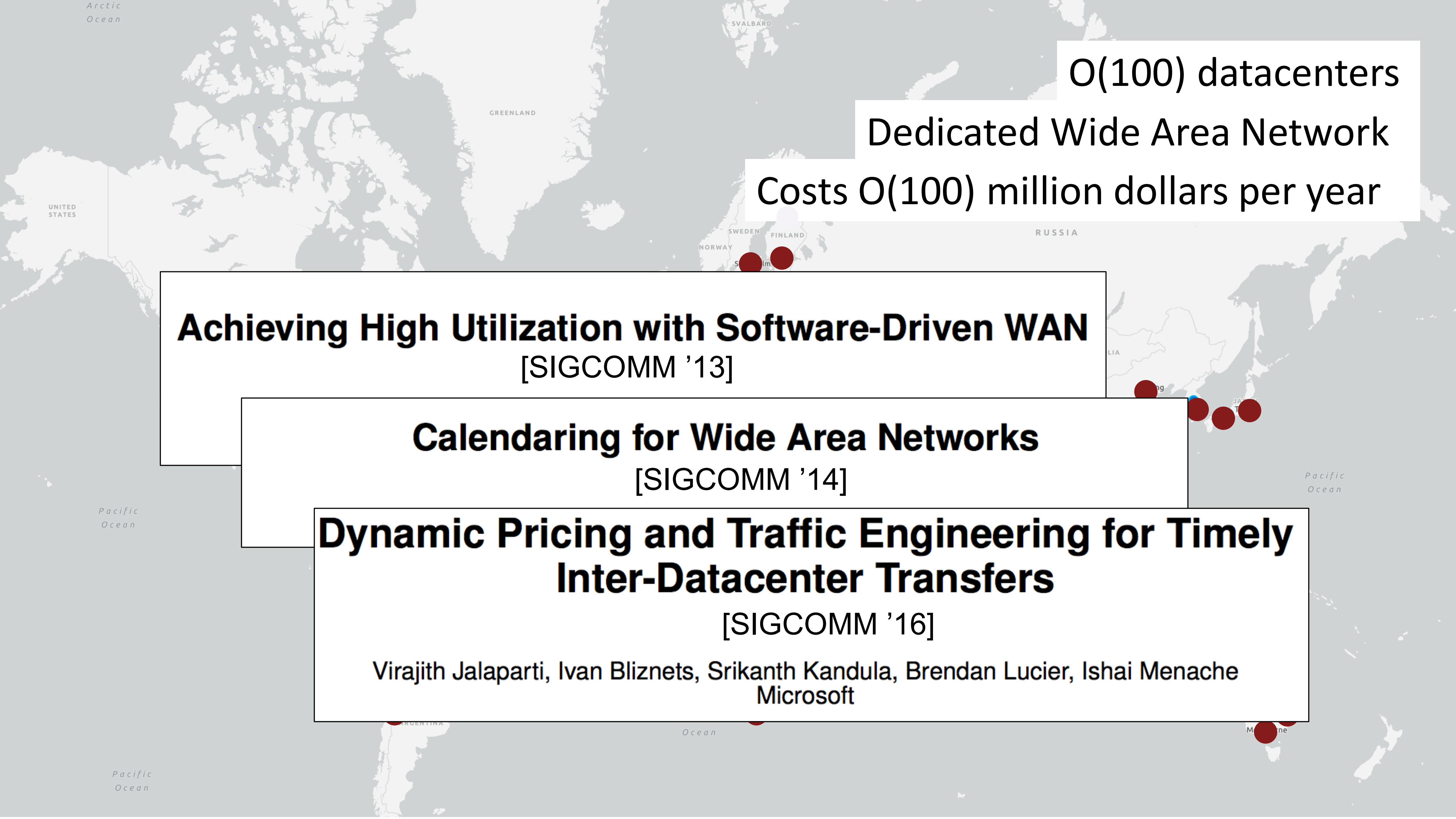
Calendaring for Wide Area Networks

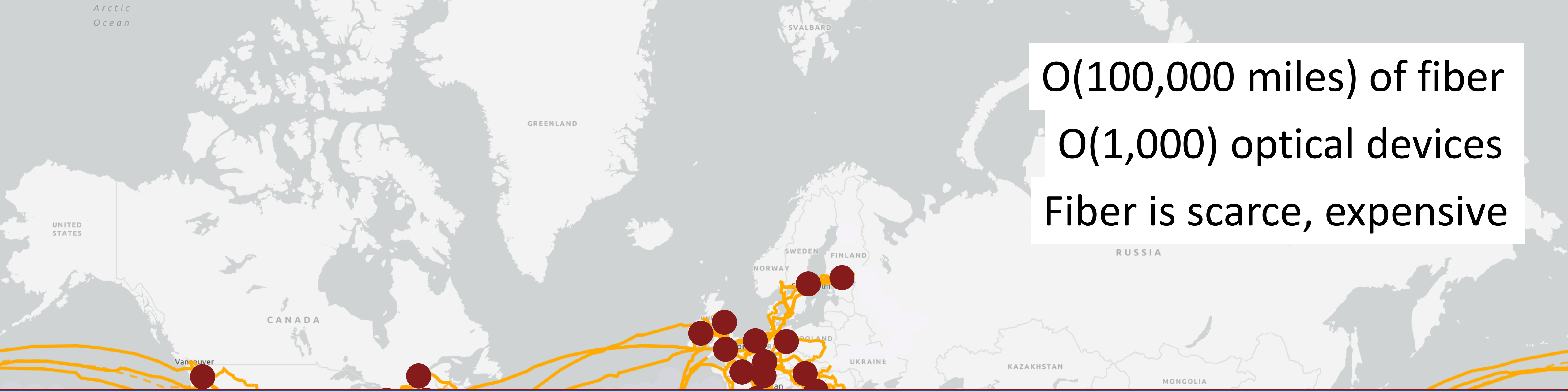
[SIGCOMM '14]

Dynamic Pricing and Traffic Engineering for Timely Inter-Datacenter Transfers

[SIGCOMM '16]

Virajith Jalaparti, Ivan Bliznets, Srikanth Kandula, Brendan Lucier, Ishai Menache
Microsoft





O(100,000 miles) of fiber

O(1,000) optical devices

Fiber is scarce, expensive

Identify inefficiencies in the optical backbone to gain capacity, availability at reduced cost.

Gain 134 Tbps of capacity and prevent 25% link failures in large North American WAN.

universität wien

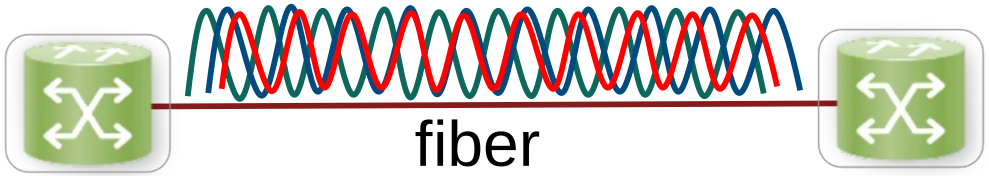
Outline

- How inefficient are optical backbones?
- Dynamic capacity links in WANs
- Challenges in dynamically adapting link capacities
- Rate Adaptive WANs

Optical Backbone Networks

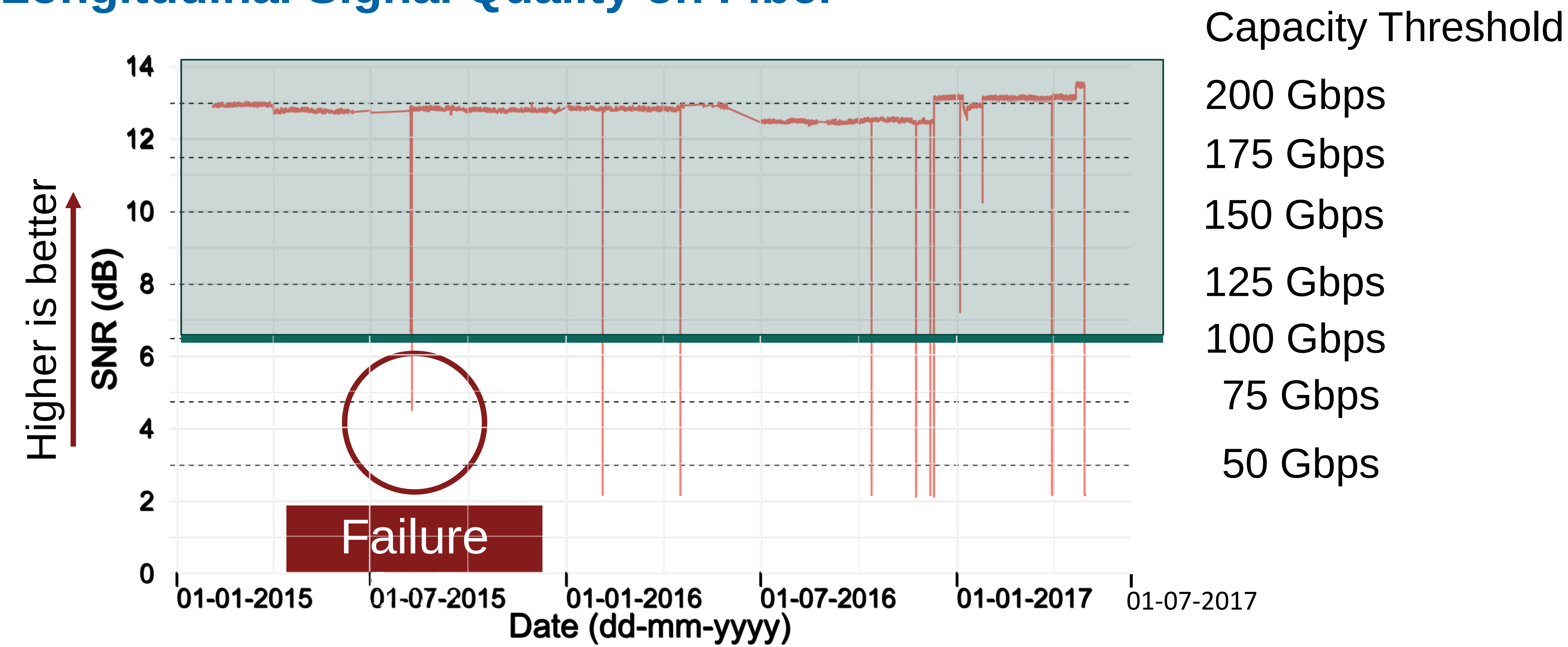
Optical cross-connects (OXCs)

- OXC: switches optical signals



- Signal-to-noise ratio (SNR) measures signal quality
- At OXC, measure signal quality
  - 8,000 wavelengths
  - Every 15 minutes
  - February 2015 to June 2017

Longitudinal Signal Quality on Fiber



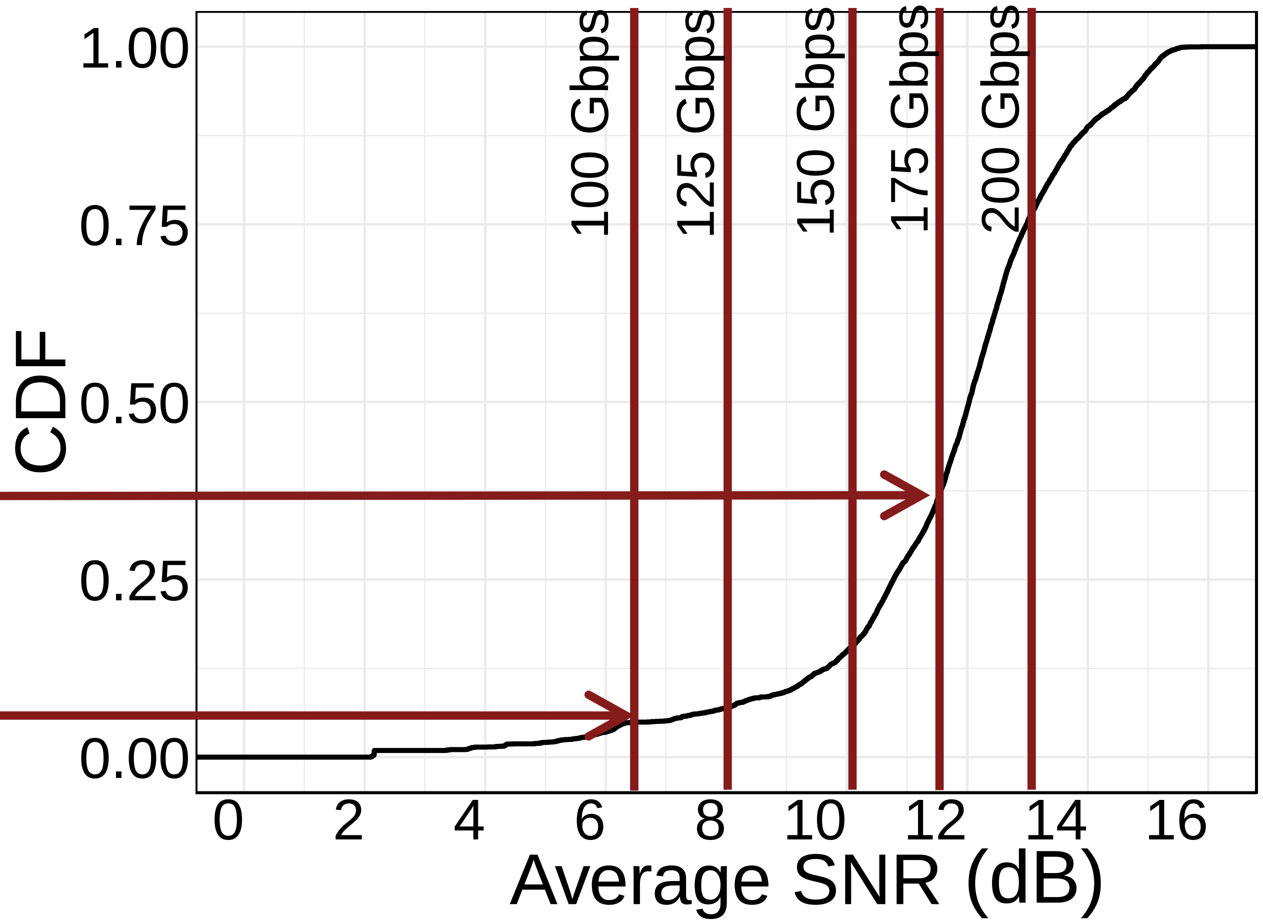
Opportunity for capacity gain

For 8,000 wavelengths in WAN:

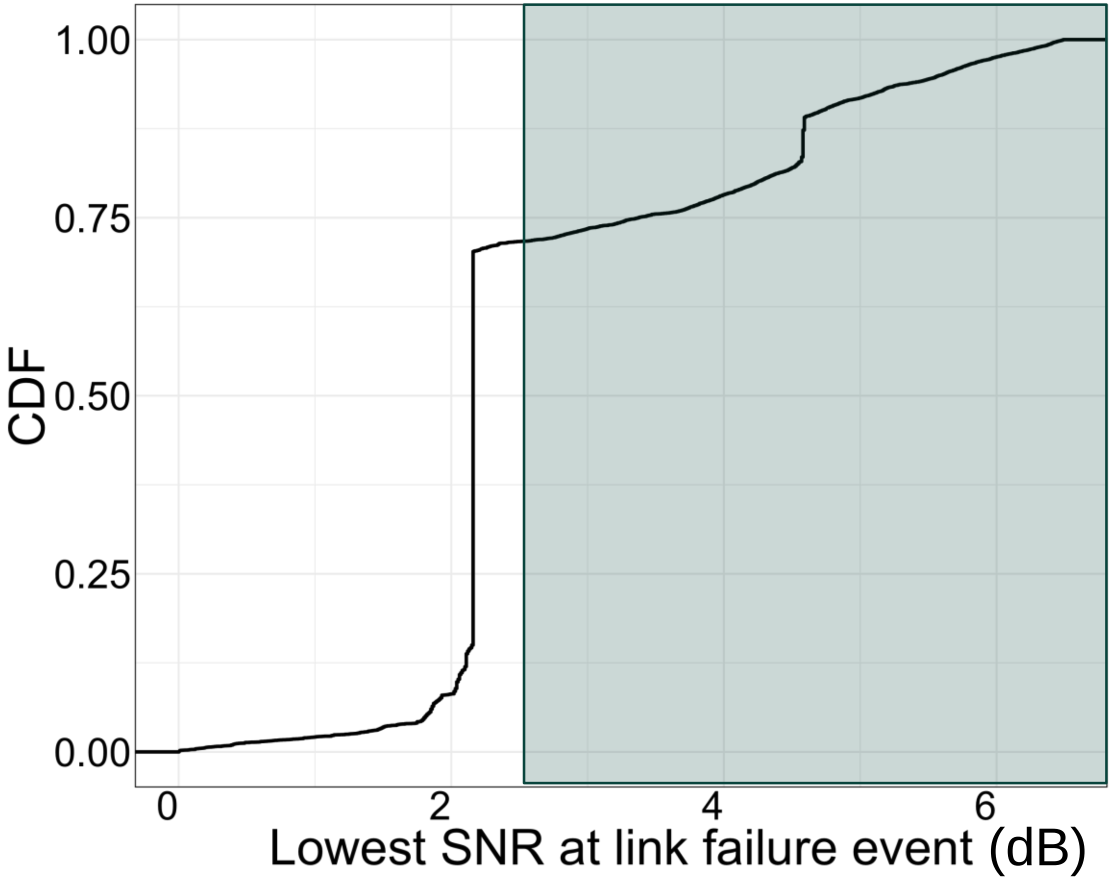
- Analyze average SNR
- Compare with thresholds for link capacity

64% of optical wavelengths can operate at 175 Gbps

95% of optical wavelengths can operate at higher than 100



Opportunity for availability gain



- Distribution of link failure SNR
  - Across WAN links
  - For 2.5 years

25% of failures have SNR > 2.5dB

These failures can be prevented by reducing link capacity to 50 Gbps

Our proposal

- Dynamically adapt link capacities in response to changes in SNR.

Gain 134 Tbps capacity	Prevent 25% link failures
By increasing link capacity when high SNR	By reducing link capacity when low SNR

Outline

- How inefficient are optical backbones?
- Dynamic capacity links in WANs
- Challenges in dynamically adapting link capacities
- Rate Adaptive WANs

Challenges in dynamically adapting link capacities

- Requires hardware support for capacity reconfiguration

- Requires re-thinking IP layer traffic engineering

Hardware support for capacity reconfiguration

Small scale lab experiments show:

- Commodity hardware takes over **1 minute of link downtime** to change capacity
- Able to reduce to 35ms with evaluation board

universität
wien

Question

How should traffic engineering incorporate dynamic capacity links?

Capacity changes cause links to be **unavailable for carrying traffic.**

Capacity changes lead to **network churn** and can be **disruptive**.

Outline

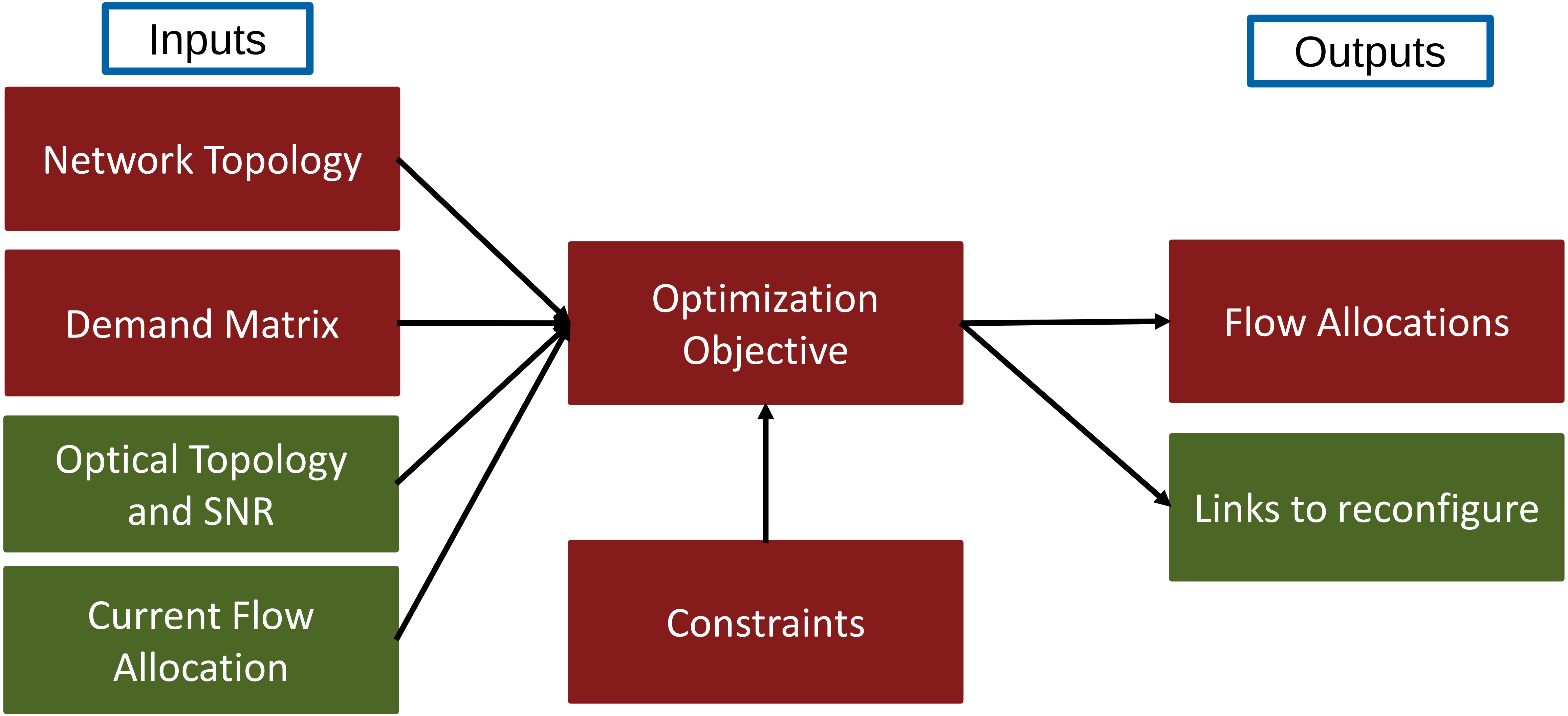
- How inefficient are optical backbones?
- Dynamic capacity links in WANs
- Challenges in dynamically adapting link capacities
- Rate Adaptive WANs

Solution

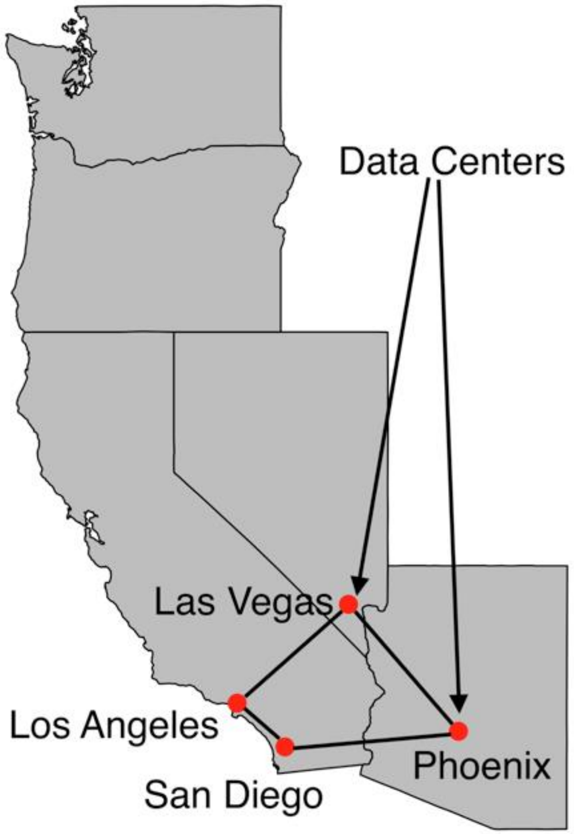
We design a Rate Adaptive Wide Area Network (RADWAN) traffic engineering controller.

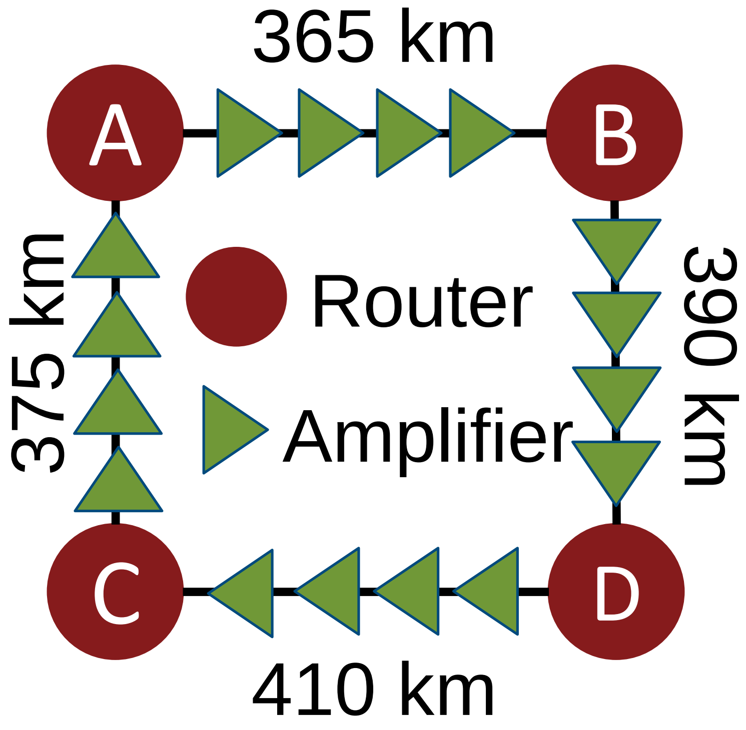
SNR-aware	Rate Adaptive	Minimally disruptive
Knows possible capacity gain of each link	Adapts link rates to meet demands and improve availability	Reconfigure capacity while minimizing network churn

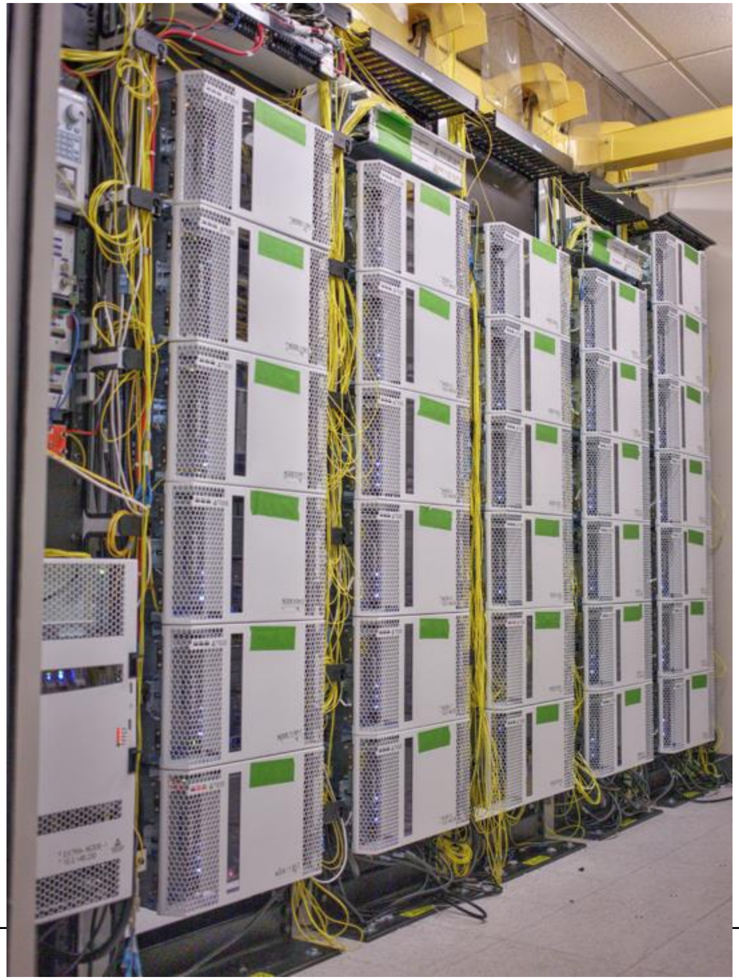
RADWAN Traffic Engineering Formulation



Proof of concept: RADWAN







Throughput Gains with RADWAN

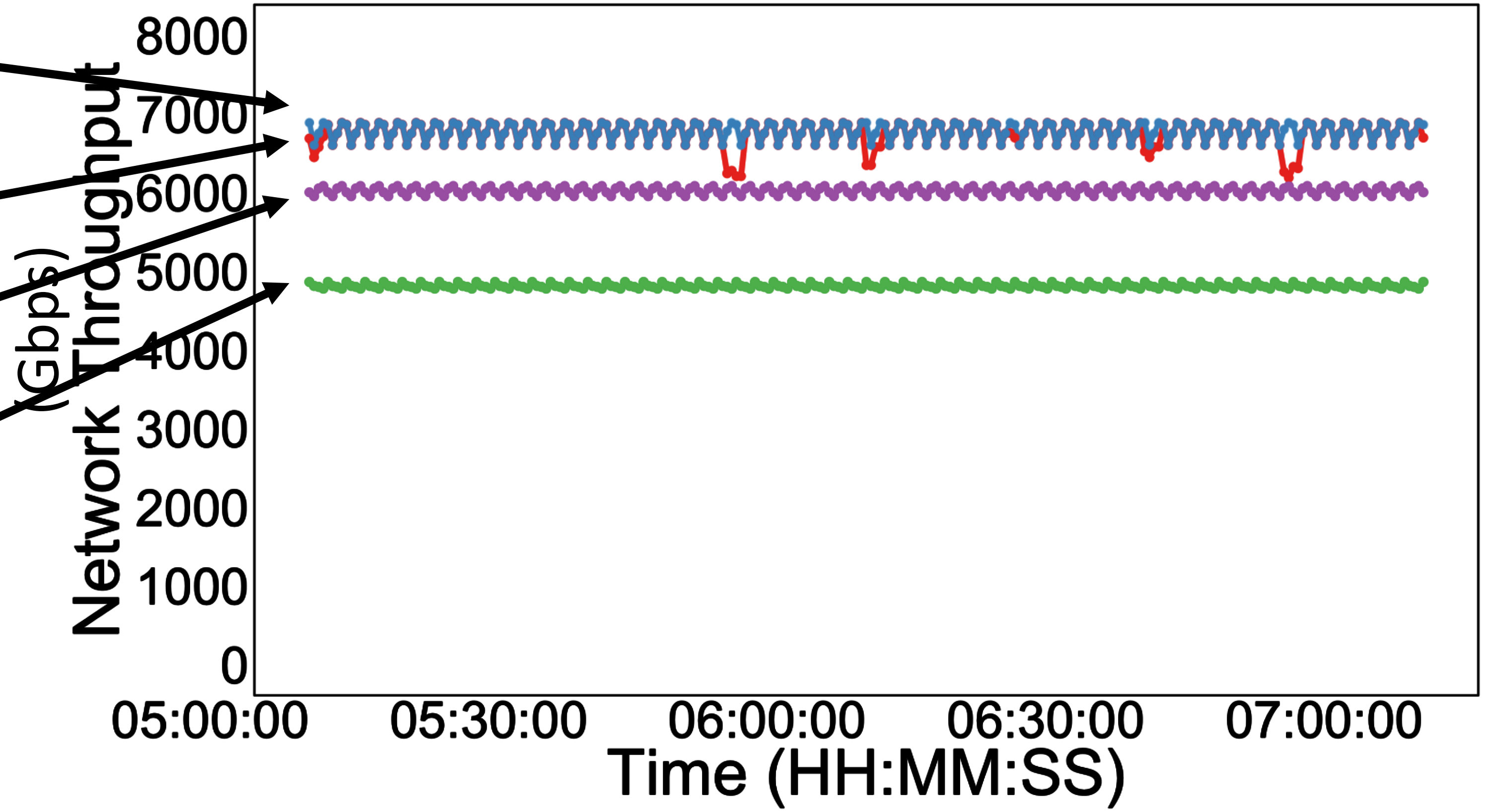
RADWAN-hitless

RADWAN

SWAN-150

SWAN [SIGCOMM '13]

RADWAN has 40%
Higher network throughput
compared to SWAN



Conclusion

- Physical layer today is configured statically
- We show that this leaves money on the table, in terms of
  - Network performance capacity
  - Link availability
  - Equipment cost ($/Gbps)

- **RADWAN** introduces programmability in Layer 1
  - **Improves network throughput by 40%**
  - **Reduces link downtime by a factor of 18**
  - **Reduces equipment cost ($/Gbps) by 32%**

RADWAN

Rate Adaptive Wide Area Networks

Rachee Singh / U. Massachusetts Amherst

Manya Ghobadi / Microsoft Research

Klaus-T. Foerster / University of Vienna

Mark Filer / Microsoft Research

Phillipa Gill / U. Massachusetts Amherst

Interested in an Overview on Algorithmic Problems in Reconfigurable Networks?

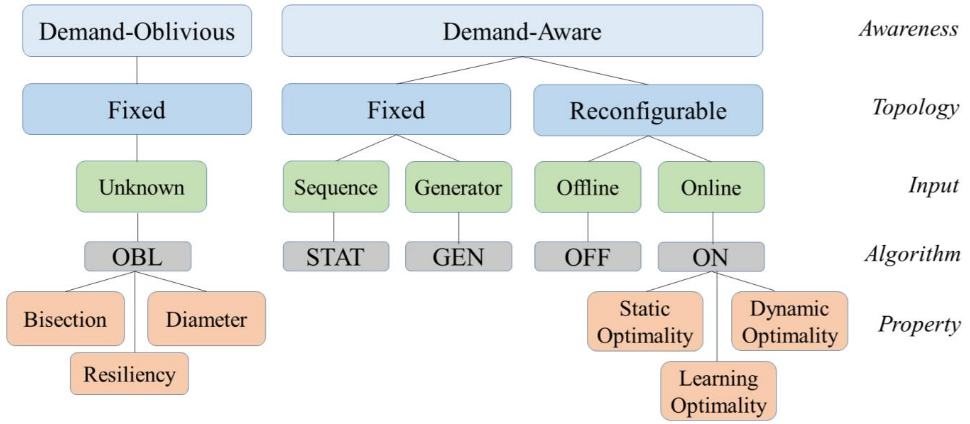


TABLE 1 Selected timeline of reconfigurable data centers

2009	– *Flyways* [51]: Steerable antennas (narrow beamwidth at 60 GHz [78]) to serve hotspots
2010	– *Helios* [33]/*c-Through* [98, 99]: Hybrid switch architecture, maximum matching (Edmond's algorithm [30]), single-hop reconfigurable connections ($O(10)ms$ reconfiguration time).
	– *Proteus* [21, 89]: k reconfigurable connections per ToR, multi-hop path stitching, multi-hop reconfigurable connections (weighted b-matching [69], edge-exchanges for connectivity [72], wavelength assignment via edge-coloring [67] on multigraphs)
2011	– Extension of *Flyways* [51] to better handle practical concerns such as stability and interference for 60GHz links, along with greedy heuristics for dynamic link placement [45]
2012	– *Mirror Mirror on the ceiling* [106]: 3D-beamforming (60 Ghz wireless), signals bounce off the ceiling
2013	– *Mordia* [31, 32, 77]: Traffic matrix scheduling, matrix decomposition (Birkhoff-von-Neumann (BvN) [18, 97]), fiber ring structure with wavelengths ($O(10)\mu s$ reconfiguration time)
	– *SplayNets* [6, 76, 82]: Fine-grained and online reconfigurations in the spirit of self-adjusting datastructures (all links are reconfigurable), aiming to strike a balance between short route lengths and reconfiguration costs
2014	– *REACToR* [56]: Buffer burst of packets at end-hosts until circuit provisioned, employs [77]
	– *Firefly* [14] Combination of Free Space Optics and Galvo/switchable mirrors (small fan-out)
2015	– *Solstice* [57]: Greedy perfect matching based hybrid scheduling heuristic that outperforms BvN [77]
	– Designs for optical switches with a reconfiguration latency of $O(10)ns$ [3]
2016	– *ProjecToR* [39]: Distributed Free Space Optics with digital micromirrors (high fan-out) [38] (Stable Matching [26]), goal of (starvation-free) low latency
	– *Eclipse* [95, 96]: $(1-1/e^{(1-\varepsilon)})$-approximation for throughput in traffic matrix scheduling (single-hop reconfigurable connections, hybrid switch architecture), outperforms heuristics in [57]
2017	– *DAN* [7, 8, 11, 12]: Demand-aware networks based on reconfigurable links only and optimized for a demand snapshot, to minimized average route length and/or minimize load
	– *MegaSwitch* [23]: Non-blocking circuits over multiple fiber rings (stacking rings in [77] doesn't suffice)
	– *Rotornet* [63]: Oblivious cyclical reconfiguration w. selector switches [64] (Valiant load balancing [94])
	– *Tale of Two Topologies* [105]: Convert locally between Clos [24] topology and random graphs [87, 88]
2018	– *DeepConf* [81]/*xWeaver* [102]: Machine learning approaches for topology reconfiguration
2019	– Complexity classifications for weighted average path lengths in reconfigurable topologies [34, 35, 36]
	– *ReNet* [13] and *Push-Down-Trees* [9] providing statically and dynamically optimal reconfigurations
	– *DisSplayNets* [75]: fully decentralized *SplayNets*
	– *Opera* [60]: Maintaining expander-based topologies under (oblivious) reconfiguration

Toward Demand-Aware Networking: A Theory for Self-Adjusting Networks
C. Avin, S. Schmid
ACM SIGCOMM Computer Communication Review, October 2018

Survey of Reconfigurable Data Center Networks: Enablers, Algorithms, Complexity
K.-T. Foerster, S. Schmid
ACM SIGACT News, June 2019