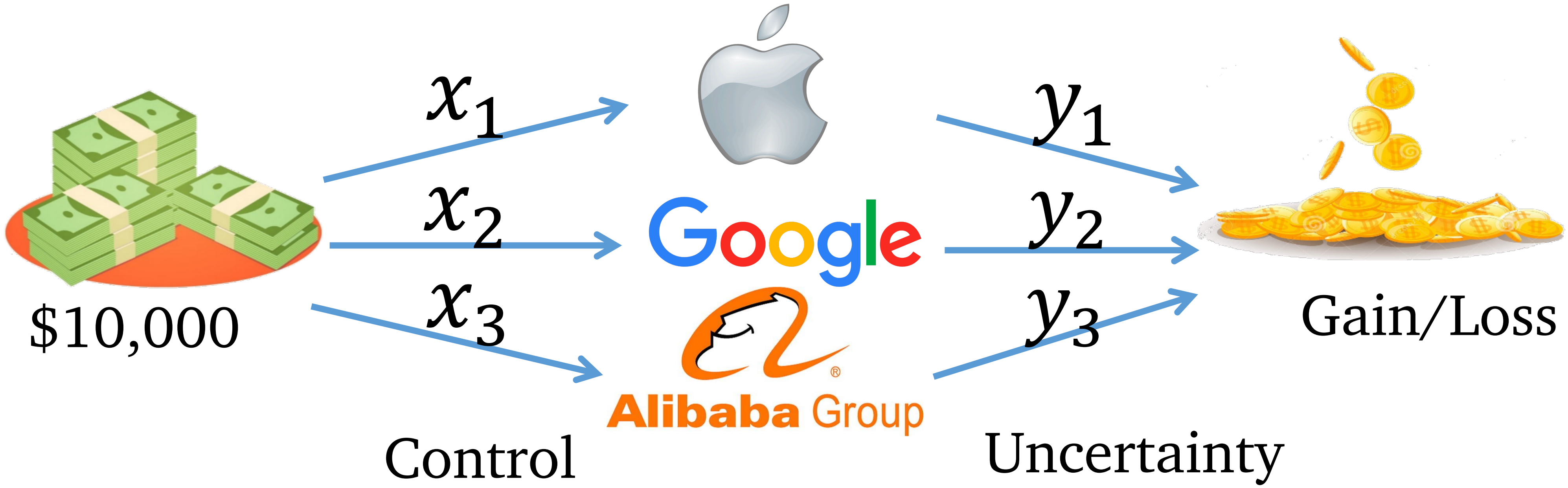
Striking the Right Utilization-Availability Balance in the WAN

Manya Ghobadi
MIT

Joint work with:
Jeremy Bogle, Nikhil Bhatia (MIT),
Ishai Menache, Nikolaj Bjorner (MSR),
Asaf Valadarsky, Michael Schapira (Hebrew U)

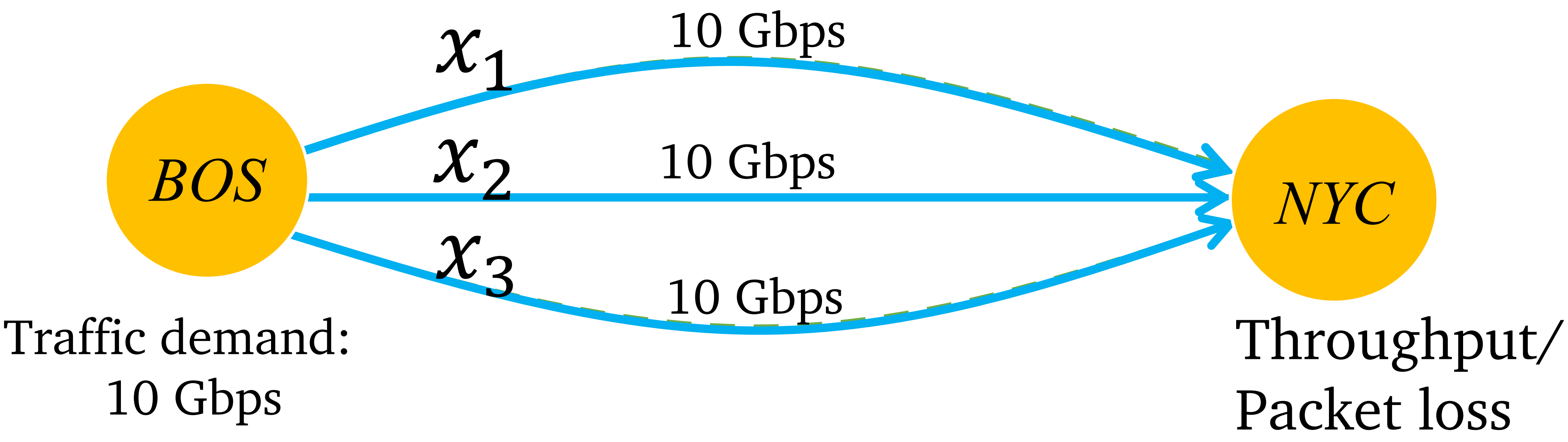
How to invest smartly in the stock market?



Loss will be ≤ $100 with probability 99%

Solution: financial risk theory

Traffic Engineering (TE) problem



- How to configure the allocation of traffic on network paths?
- Goal: efficiently utilizing the network to match the current traffic demand (periodic process)

Extensive research on TE in a broad variety of environments

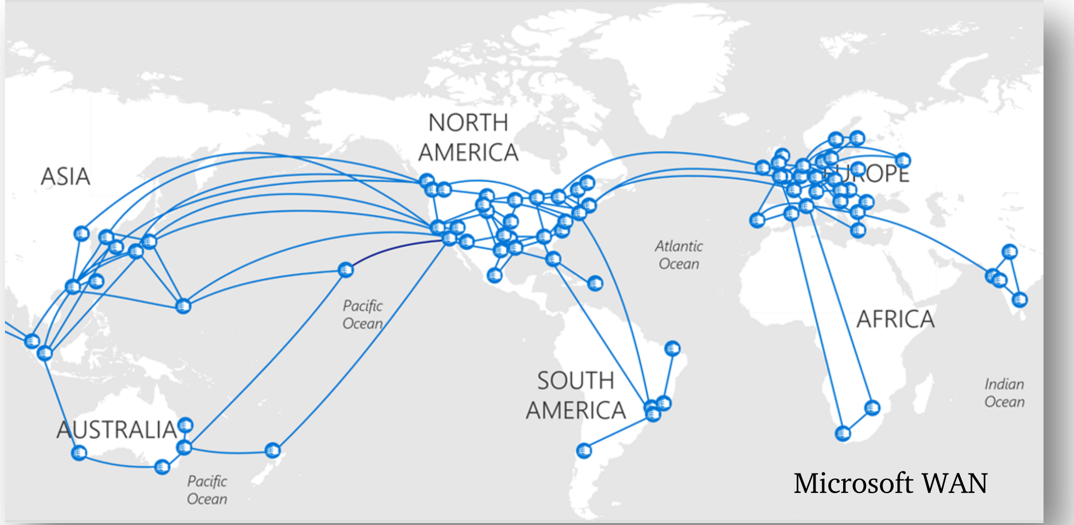
- Wide-area networks
  - Kumar et al. [NSDI'18]
  - Liu et al. [SIGCOMM'14]
  - Kumar et al. [SIGCOMM'15]
  - Jain et al. [SIGCOMM'13]
  - Hong et al. [SIGCOMM'13]
- ISP networks
  - Jiang et al.[SIGMETRICS'09]
  - Kandula et al. [SIGCOMM'05]
  - Fortz et al. [INFOCOM'2000]
- Data center networks
  - Alizadeh et al. [SIGCOMM'14]
  - Akyildiz et al. [Journal of Comp. Nets.'14]
  - Benson et al. [CoNEXT'11]

TE problem in Wide-Area Networks

TE problem in Wide-Area Networks

Challenging:

- Billion dollar infrastructure
- High efficiency and availability



Microsoft WAN

Solution:

- Model the network as a graph
- Solve a Linear Program

$$\min \Phi = \sum_{a \in A} \Phi_a$$

Objective

subject to

$$\sum_{x:(x,y)\in A} f^{(s,t)}_{(x,y)} - \sum_{z:(y,z)\in A} f^{(s,t)}_{(y,z)} = \begin{cases} -D(s,t) & \text{if } y=s, \\ D(s,t) & \text{if } y=t, \\ 0 & \text{otherwise}, \end{cases}$$

$$y, s, t \in N, \qquad (1)$$

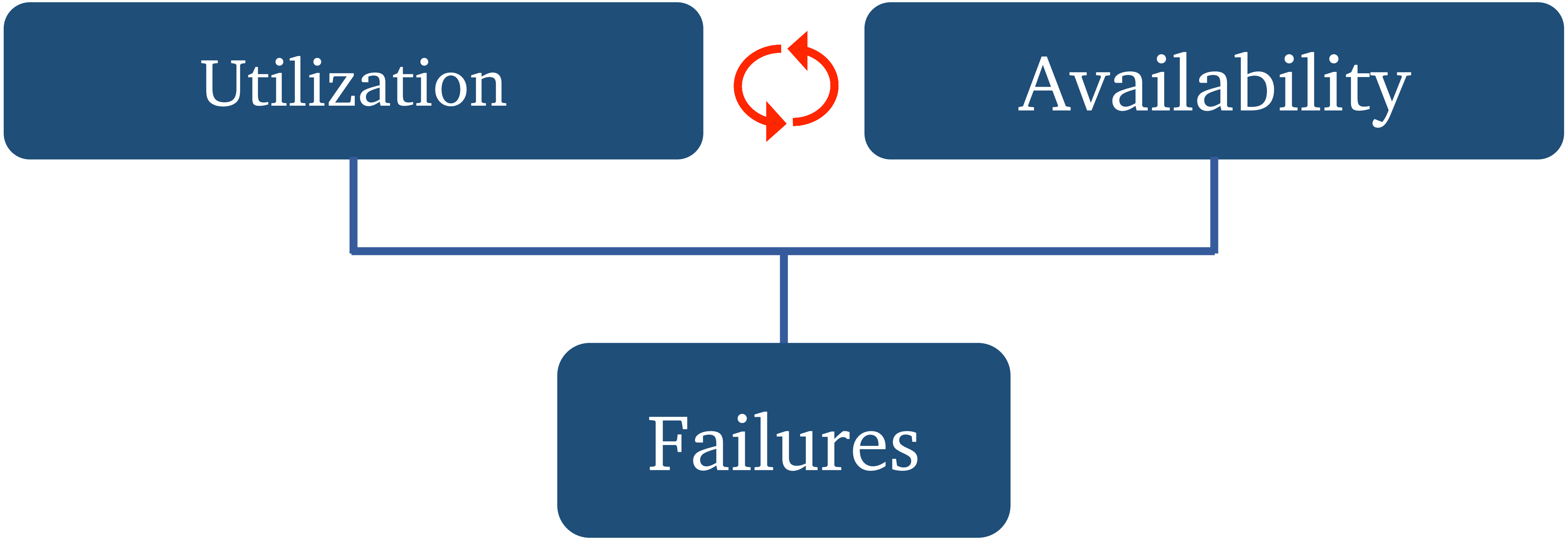
$$\ell(a) = \sum_{(s,t)\in N\times N} f^{(s,t)}_a \qquad a \in A, \qquad (2)$$

$$\Phi_a \geq \ell(a) \qquad a \in A, \qquad (3)$$

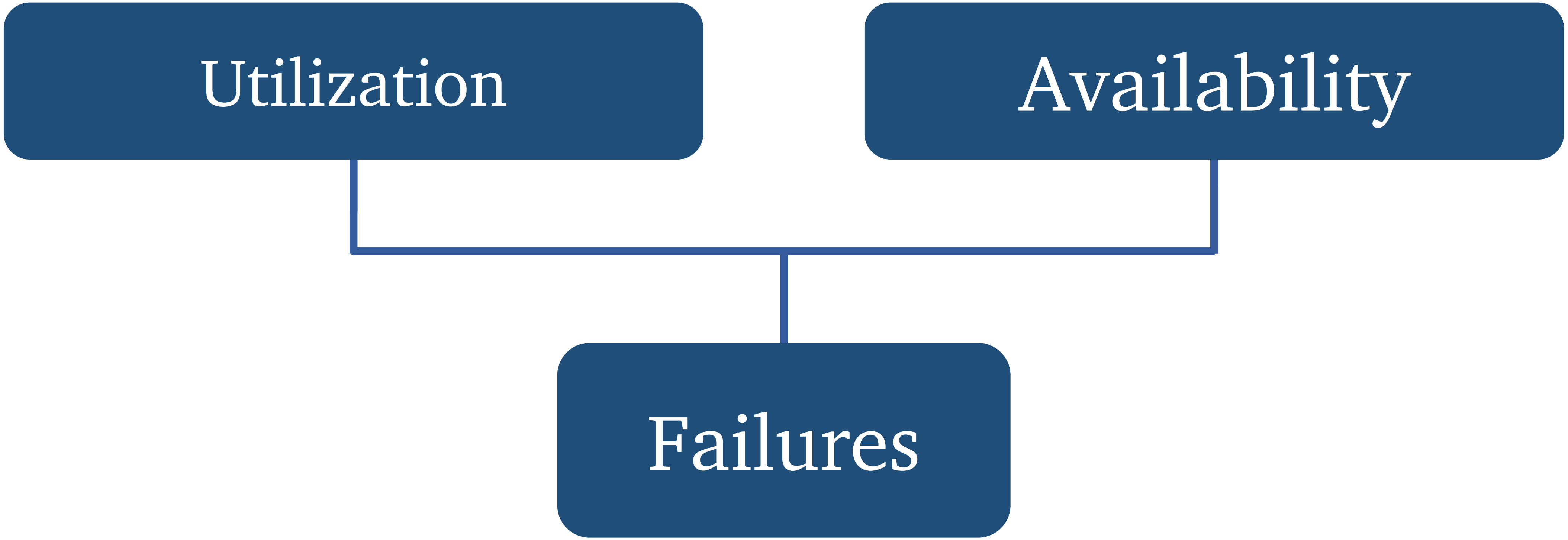
Constraints

[B. Fortz, Internet Traffic Engineering by Optimizing OSPF Weights, INFOCOM'2000]

Competing goals: high utilization and availability



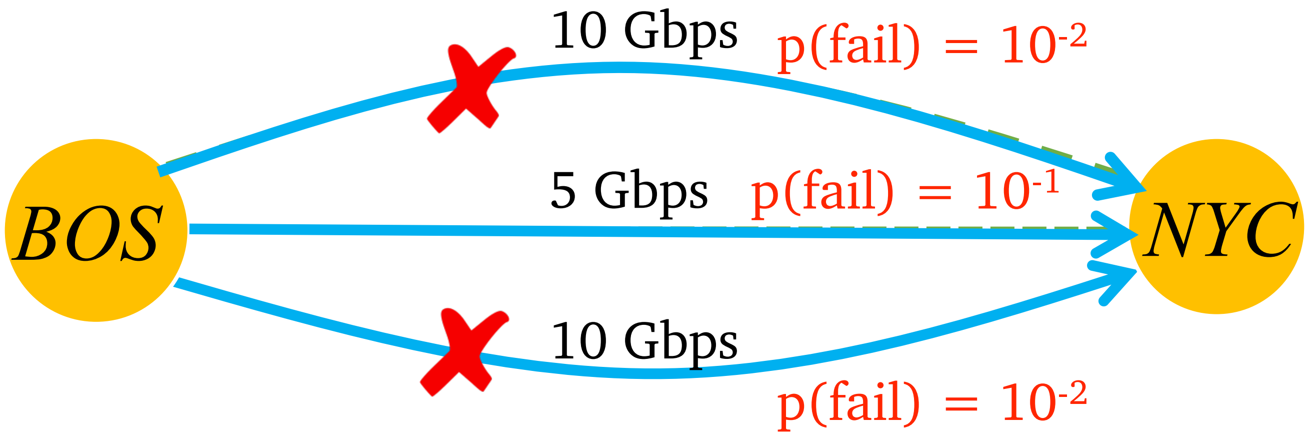
Competing goals: high utilization and availability



Traffic engineering under failures

- Today: optimize for the worst conceivable (potentially unlikely) failure scenarios
- Problem: under-utilizing the network

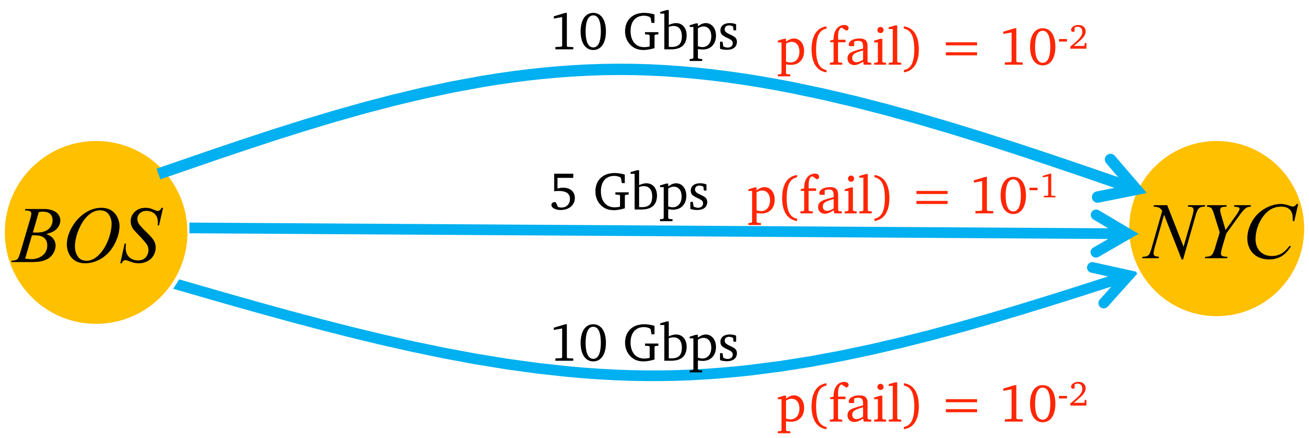
Robust against *k* simultaneous link failures



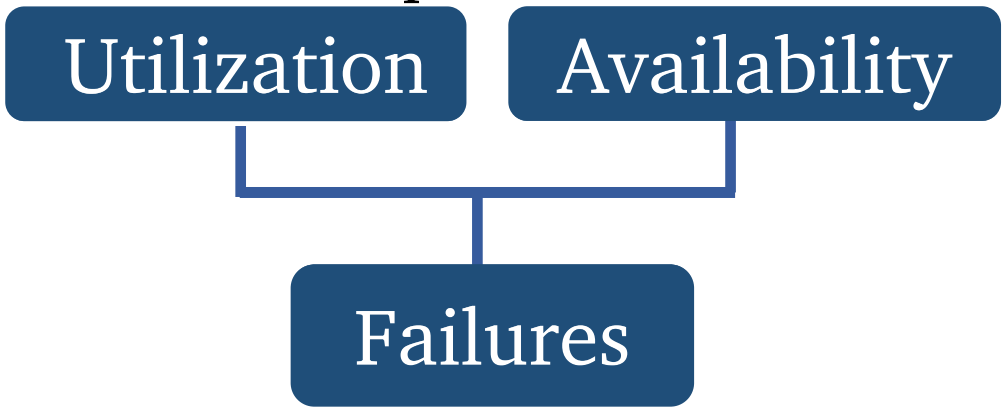
Admissible traffic: 5 Gbps 99.999% of the time

Traffic engineering under failures

Robust against *k* simultaneous link failures

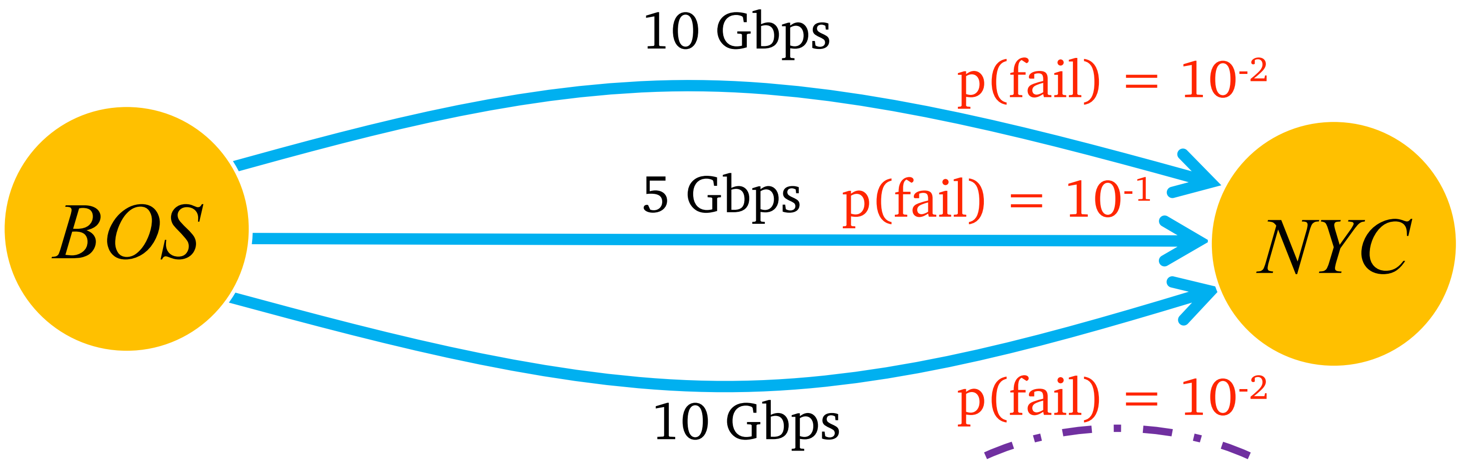


Admissible traffic: 5 Gbps   99.999% of the time



Our approach to traffic engineering under failures

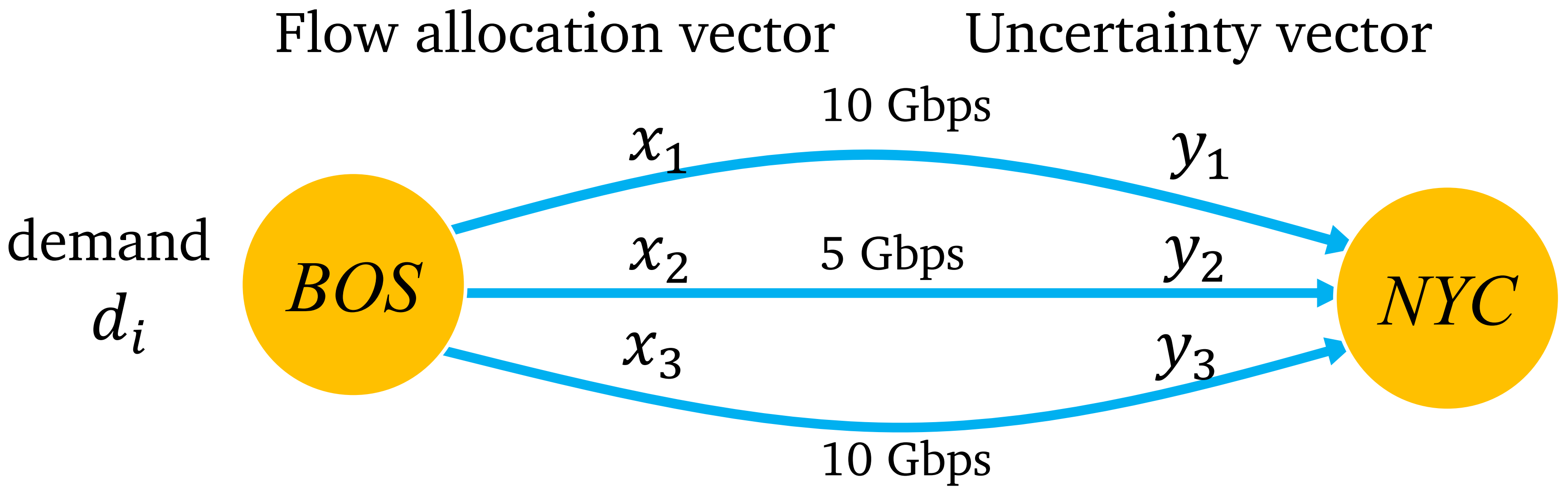
- Use the failure probabilities to reason about the likelihood of failure scenarios
- Provide a mathematical probabilistic guarantee for availability



Admissible traffic	Availability
5 Gbps	99.999%
10 Gbps	99.99%
15 Gbps	99.8%
20 Gbps	98%

Our approach to traffic engineering under failures

- Use the failure probabilities to reason about the likelihood of failure scenarios
- Provide a mathematical probabilistic guarantee for availability

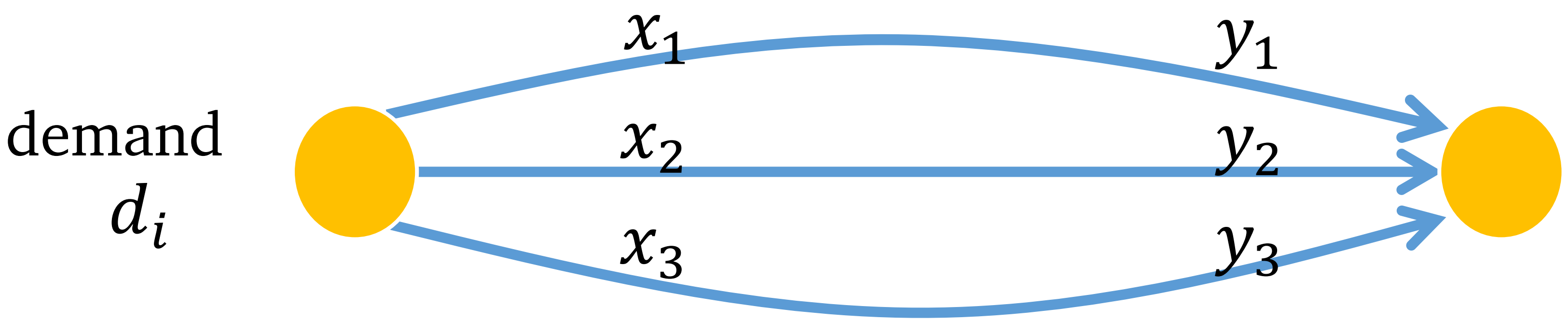


For all flows, 90% of the demand is satisfied 99.9% of the time
For all flows, loss will be ≤ 10% of the demand 99.9% of the time

Main idea

The loss will be ≤ \$100 with probability 99%

Find *x* that minimizes the loss with probability β



The loss will be ≤ 10% of the demand with probability 99%

Find *x* that minimizes the loss with probability β

Key technique: scenario-based formulation

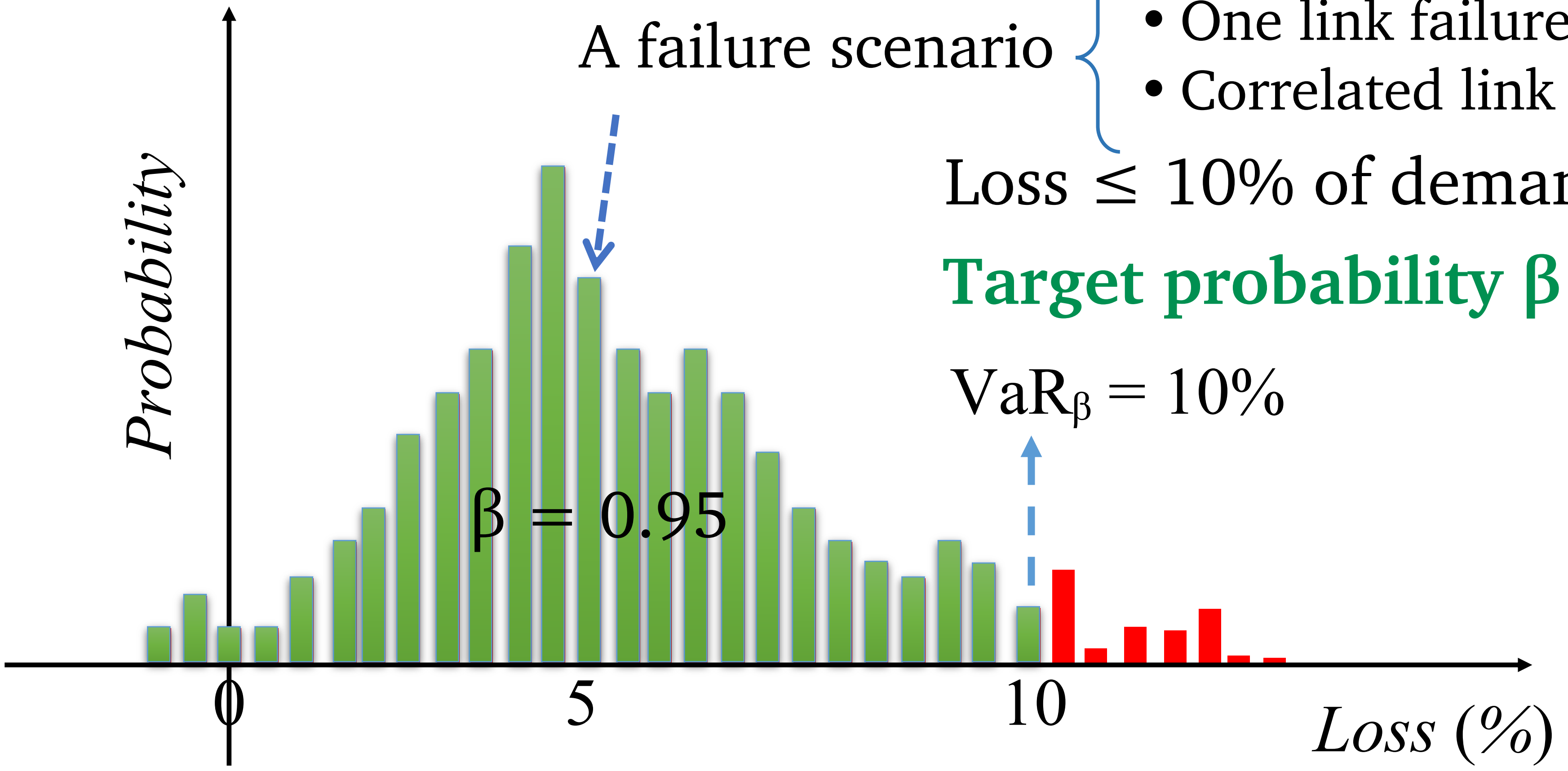
A failure scenario

- One link failure
- Correlated link failures

Loss ≤ 10% of demand w probability 0.95

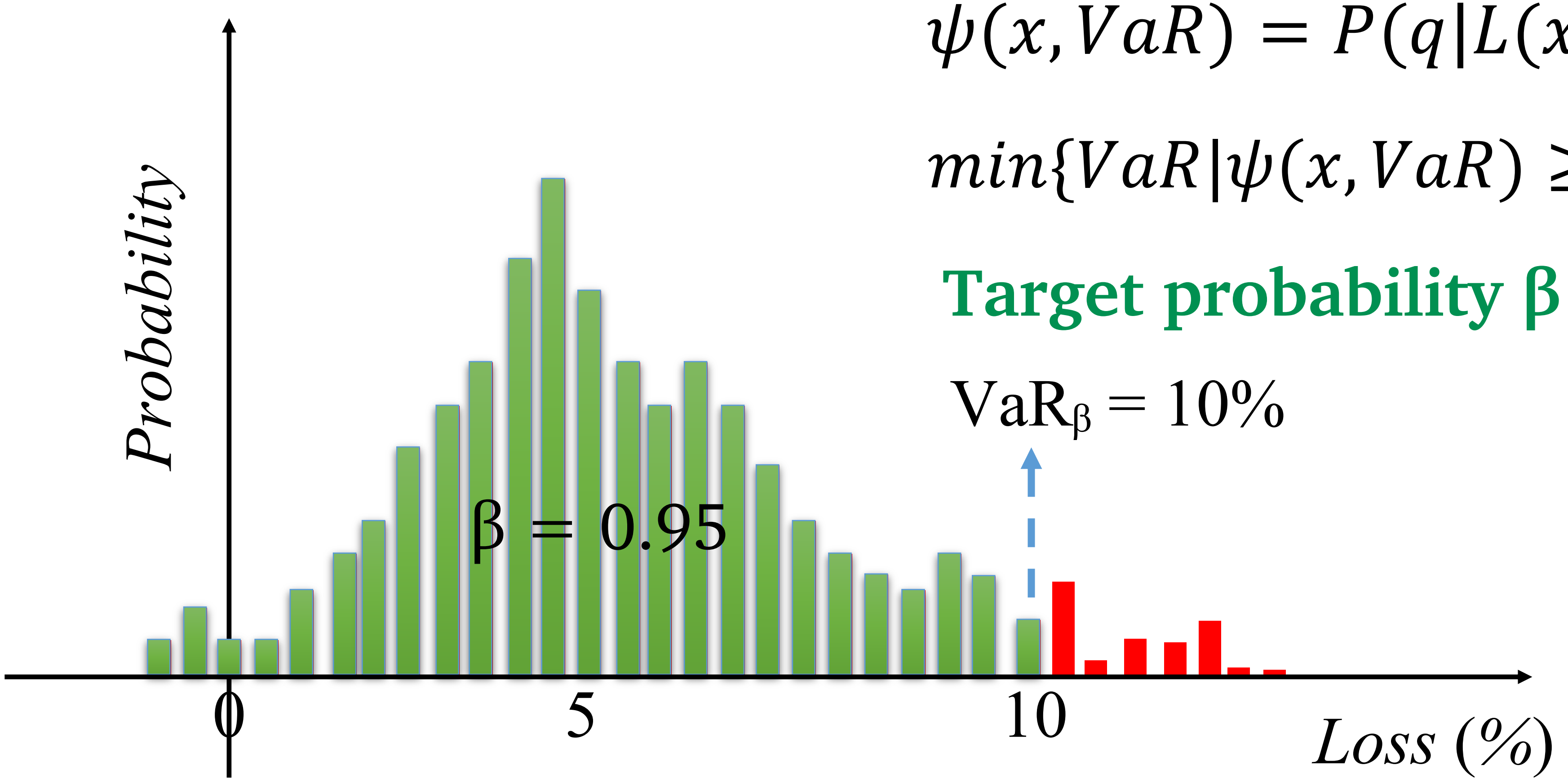
Target probability $\beta = 0.95$

$\mathrm{VaR}_\beta = 10\%$



- Unsatisfied demand
- Packet loss

Key technique: scenario-based formulation



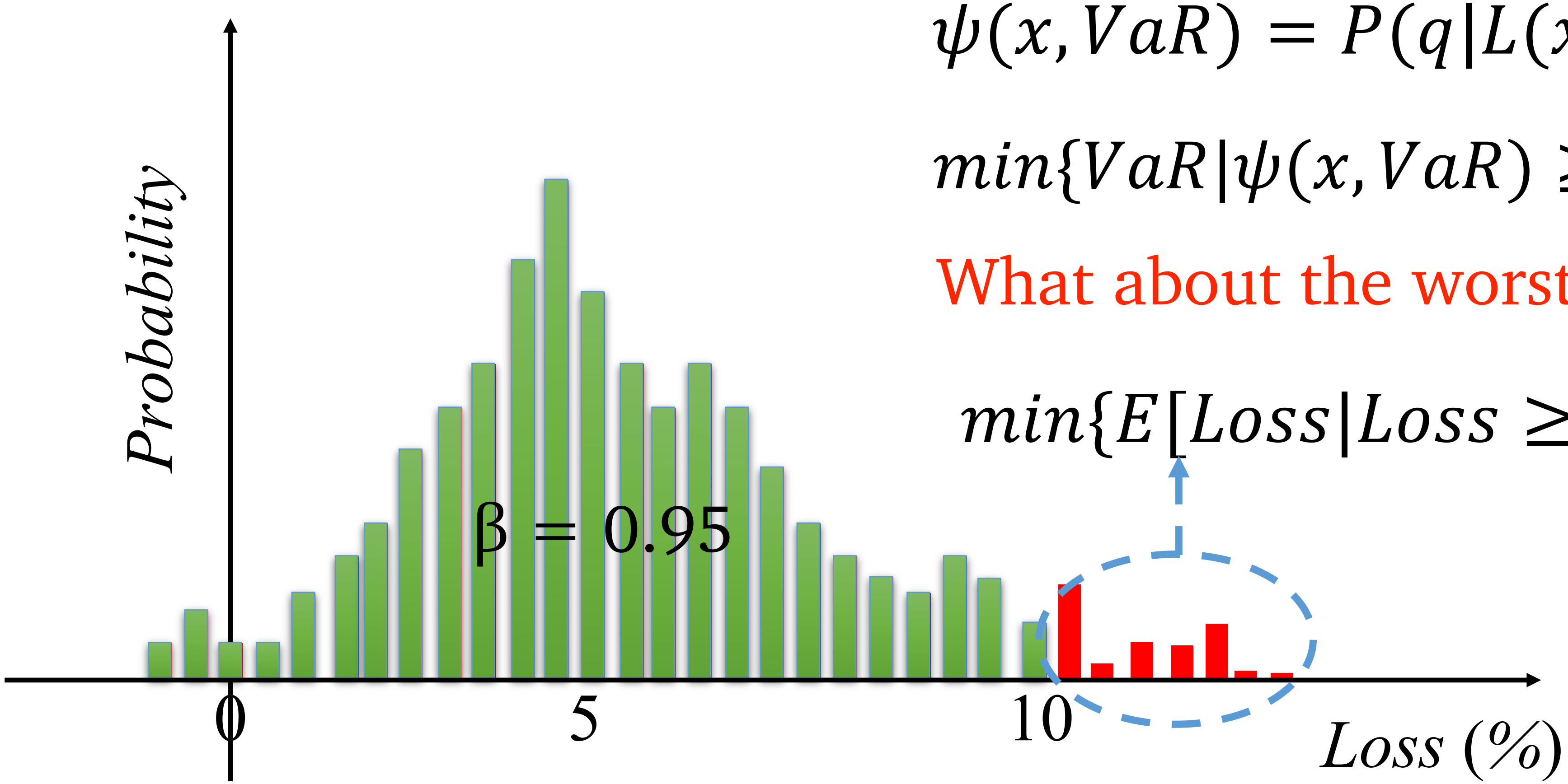
$$\psi(x, VaR) = P(q | L(x, y(q)) \leq VaR)$$

$$min\{VaR | \psi(x, VaR) \geq \beta\}$$

Target probability β = 0.95

$VaR_{\beta} = 10\%$

TeaVaR: Traffic Engineering Applying Value-at-Risk



$$\psi(x, VaR) = P(q \mid L(x, y(q)) \leq VaR)$$

$$min\{VaR \mid \psi(x, VaR) \geq \beta\}$$

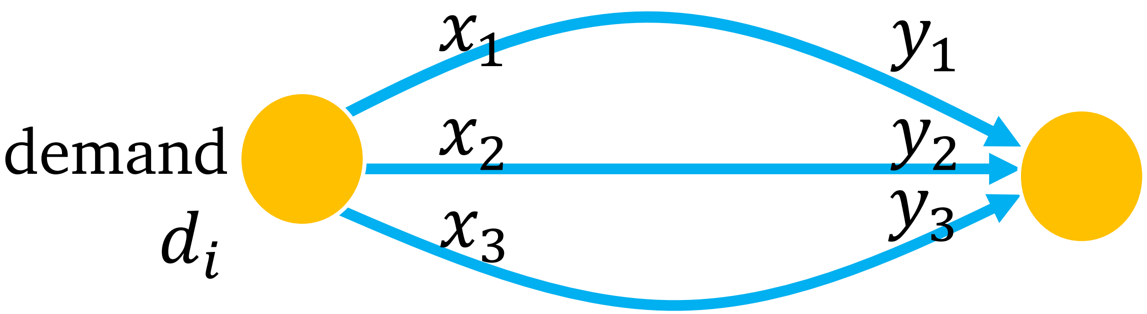
What about the worst 5% of scenarios?

$$min\{E[Loss \mid Loss \geq VaR]\}$$

Challenges unique to networking

- Achieving fairness across network users.
- Enabling computational tractability as the network scales.
- Capturing fast rerouting of traffic in data plane.
- Accounting for correlated failures.

Achieving fairness



Objective: Find x that minimizes the loss with probability β

R_i: routes for flow i
x_r: flow allocation on route $r \in R_i$
y_r: binary variable indicating if route r is up

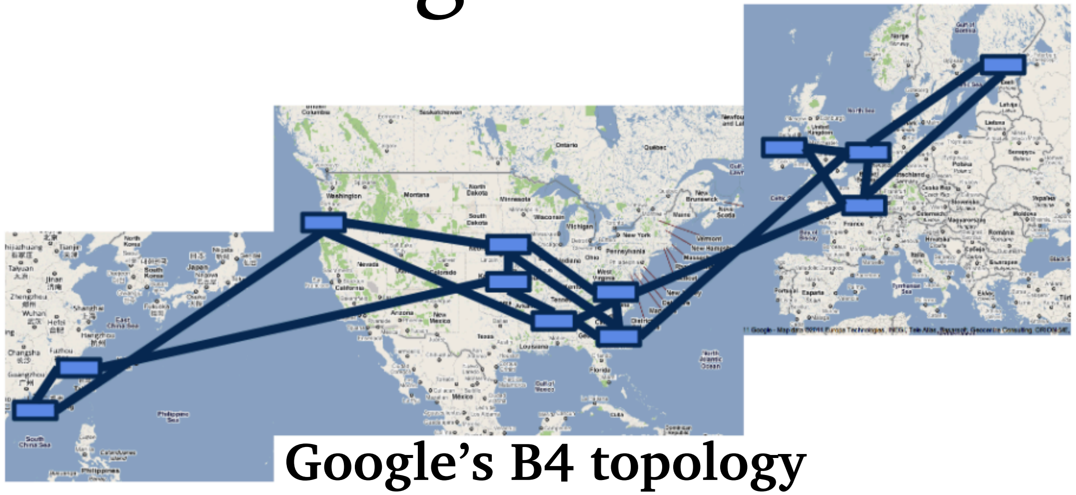
Satisfied demand for flow i: $\sum_{r \in R_i} x_r y_r$

Starvation-aware loss function:

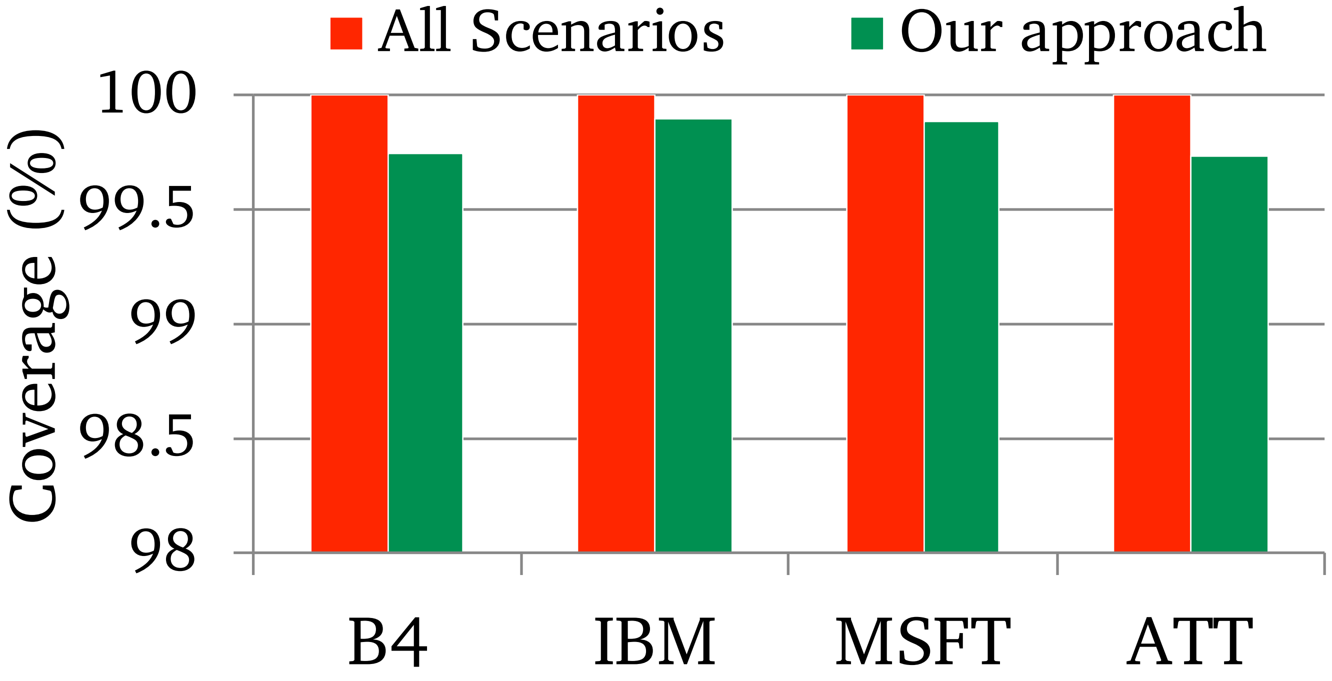
- Worst case normalized unmet demand

$$L(x,y) = max_i\left[1 - \frac{\sum_{r \in R_i} x_r y_r}{d_i}\right]^+$$

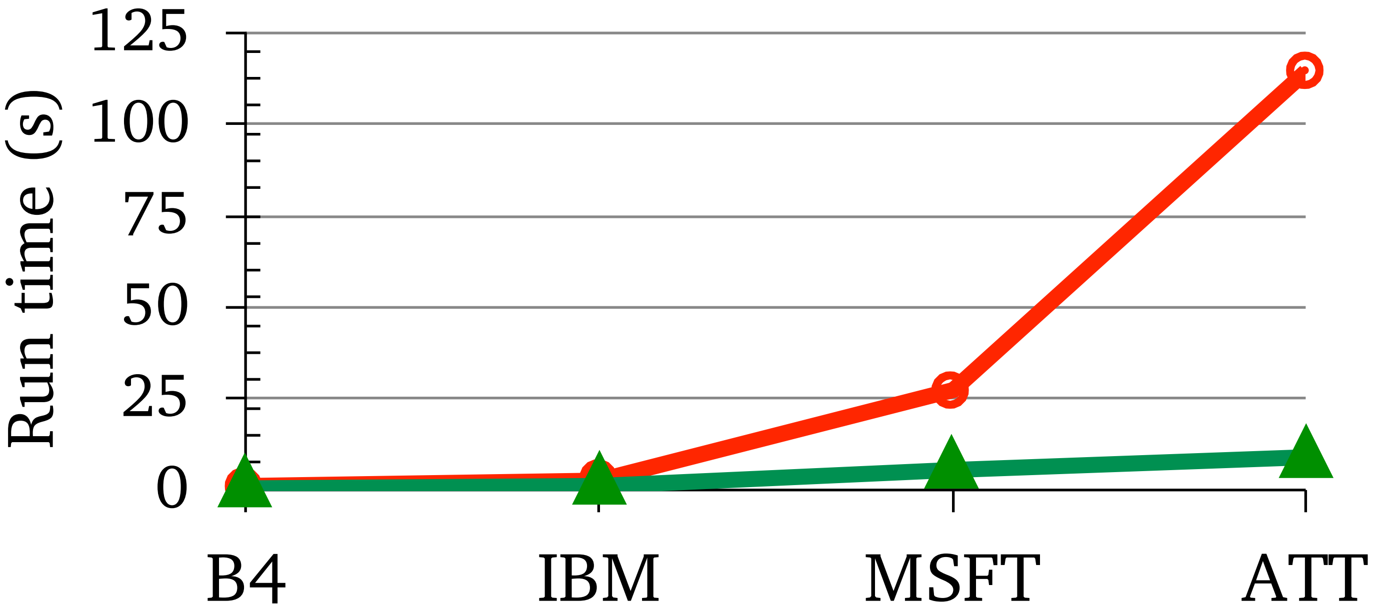
Handling scale



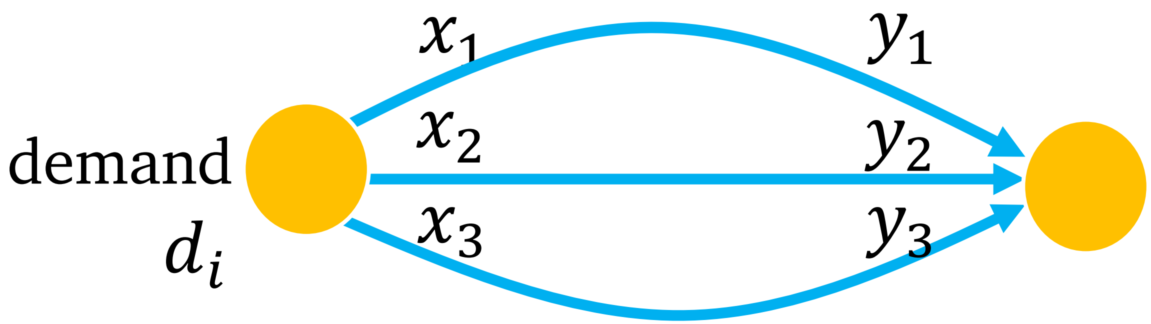
Google's B4 topology

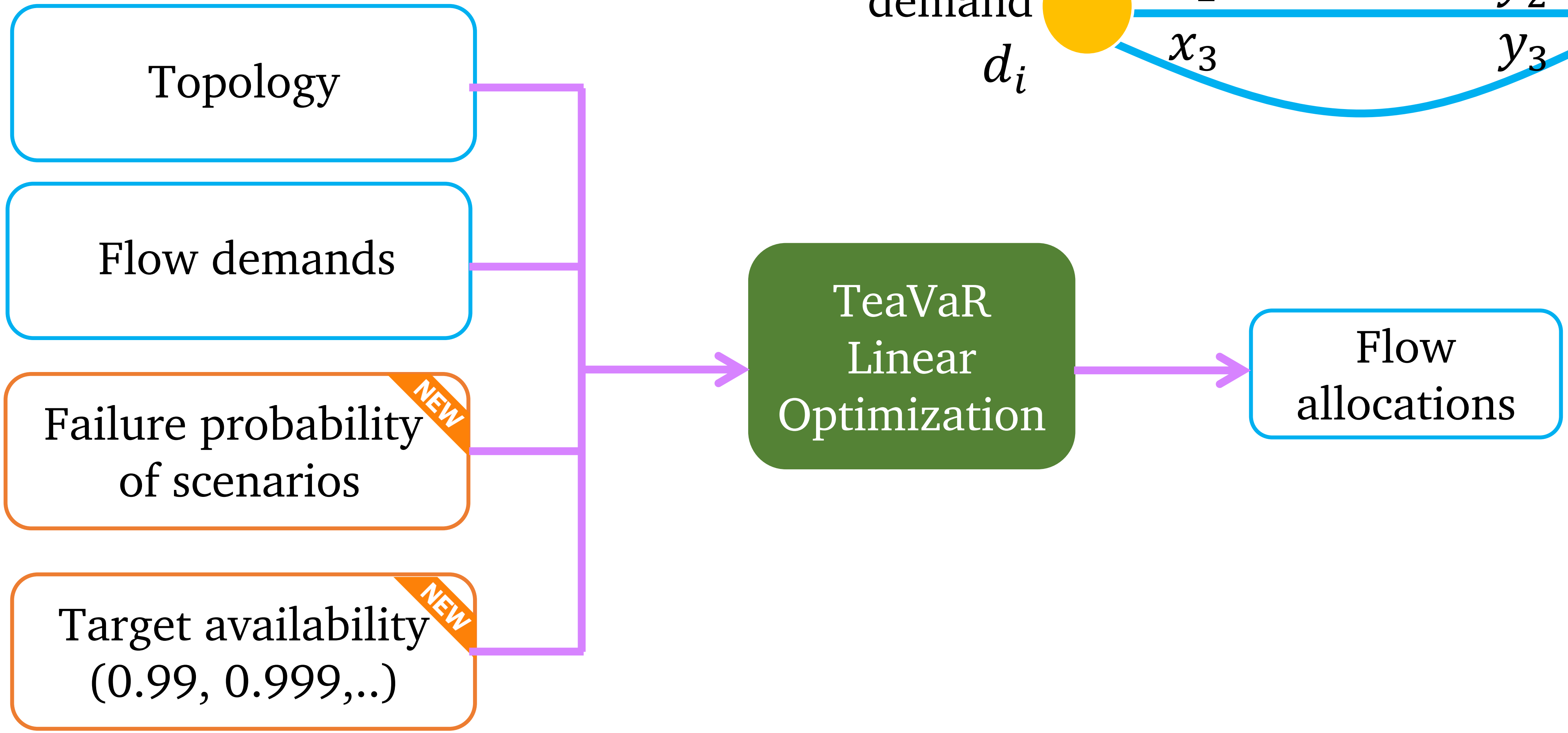


Topology	# Edges	# Scenarios
B4	38	O(1E11)
IBM	48	O(1E14)
MSFT	100	O(1E30)
ATT	112	O(1E33)



System architecture





TEAVAR Demo

Stocks

Topology	B4
Demand	1
Paths	ED
Beta (%)	0.9
Cutoff (%)	0.0001

Submit

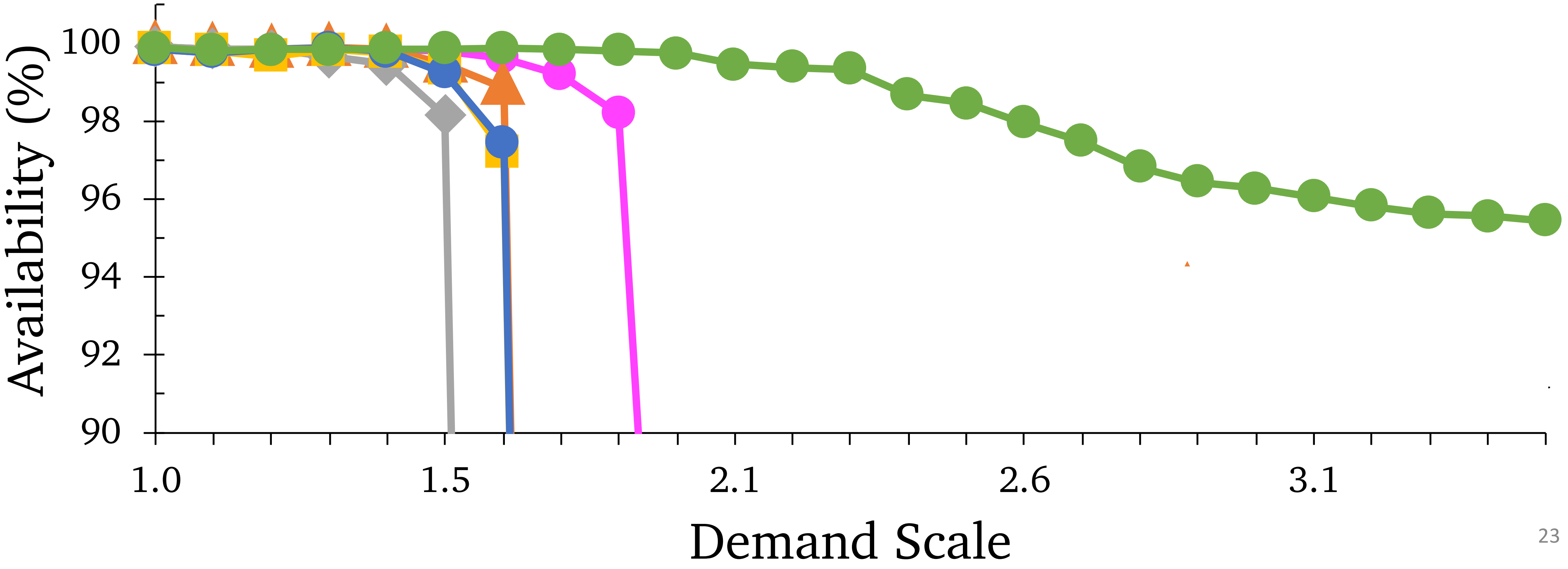
Waiting for 128.30.92.156...

Evaluations

- Topologies: B4, IBM, ATT, and MSFT
- Traffic matrix:
  - Four months of MSFT traffic matrix (one sample/hour), for the rest of topologies, used 24 TMs from YATES [SOSR’18]
- Tunnel selection:
  - Our optimization framework is orthogonal to tunnel selection
  - Oblivious paths, link disjoint paths, and k-shortest paths
- Baselines:
  - SMORE [NSDI’18]
  - FFC [SIGCOMM’14]
  - B4 [SIGCOMM’13]
  - ECMP

Availability vs. demand scale

- Availability is measured as the probability mass of scenarios in which demand is fully satisfied ("all-or-nothing" requirement)
  - If a TE scheme's bandwidth allocation is unable to fully satisfy demand in 0.1% of scenarios, it has an availability of 99.9%



Robustness to probability estimates

Noise in probability estimations	% error in throughput
1%	1.43%
5%	2.95%
10%	3.07%
15%	3.95%
20%	6.73%

Summary

- TeaVaR uses financial risk theory for solving Traffic Engineering under failures.

- TeaVaR's approach is applicable to networking resource allocation problems such as capacity planning.

- Code and demo available at: http://teavar.csail.mit.edu/