

# Designing a Predictable Internet Backbone with Valiant Load-balancing

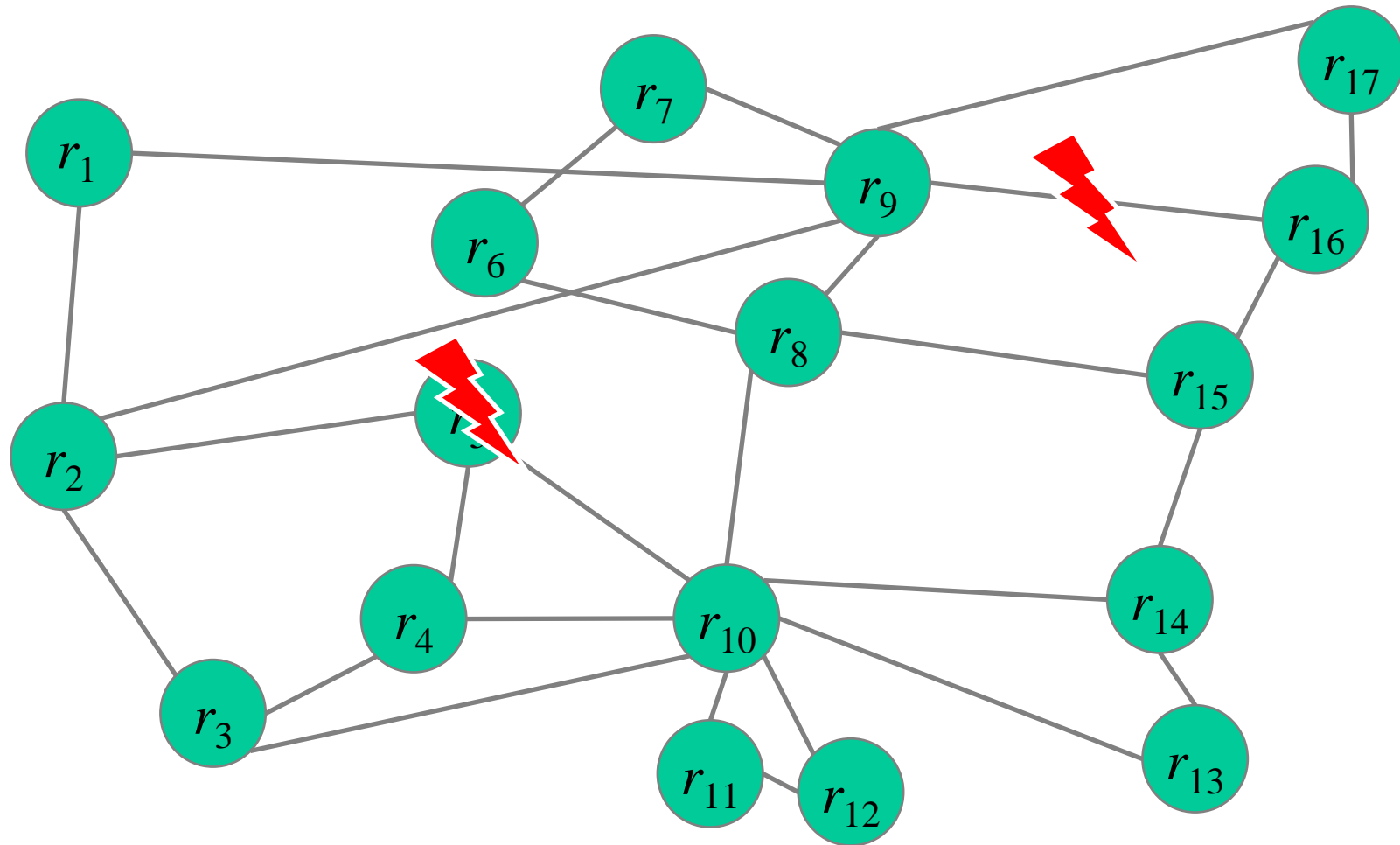
Rui Zhang-Shen, Nick McKeown  
{rzhang, nickm}@stanford.edu



*IWQoS, June 22, 2005*



# Current Backbone Design



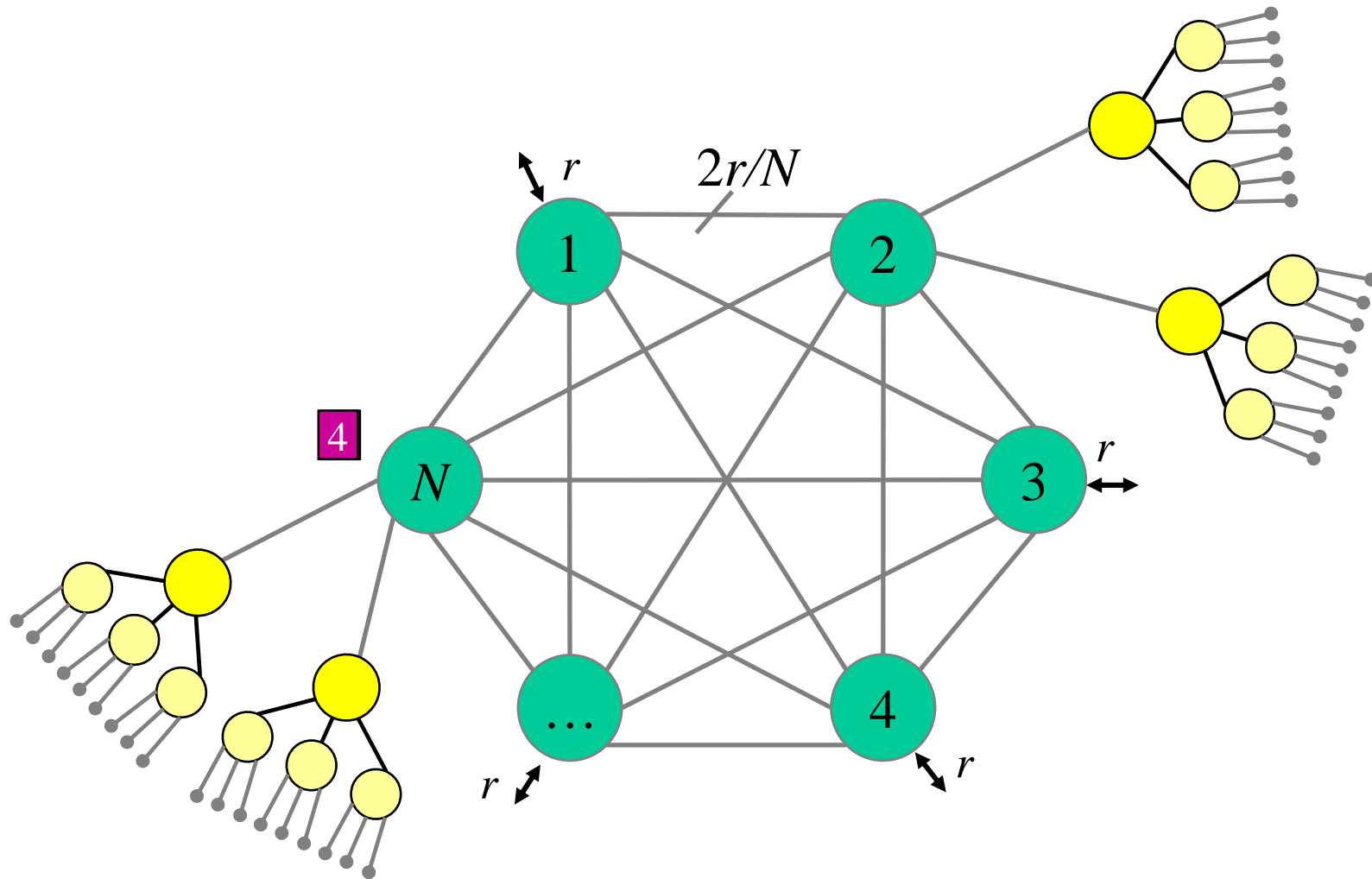
# What makes network design hard?

- Hard to measure current traffic matrix
- Futile to estimate the future traffic matrix
- Need to provision for failures

# Approach

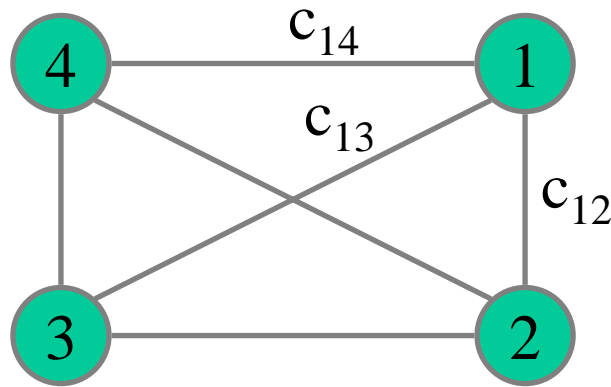
- Assume we know/estimate traffic entering and leaving each Regional Network
  - Requires only local knowledge of users and market estimates
- Use Valiant Load Balancing (VLB) over whole network
  - Enables support of all traffic matrices
  - Efficient

# Valiant Load-Balancing



# Non-homogeneous Networks

- $N$  nodes of capacities  $r_1 \geq r_2 \geq \dots \geq r_N$
- Total access capacity:  $R = \sum_i r_i$
- Link capacity  $c_{ij}$ ,  $l_i = \sum_{j \neq i} c_{ij}$ ,  $L = \sum_i l_i$
- Fanout  $f_i = l_i/r_i$ , network fanout  $f = \max_i f_i$



# Which to Minimize?

- Total link capacity  $L$
- Network fanout  $f$ 
  - Small nodes have small link capacities
- They are related:

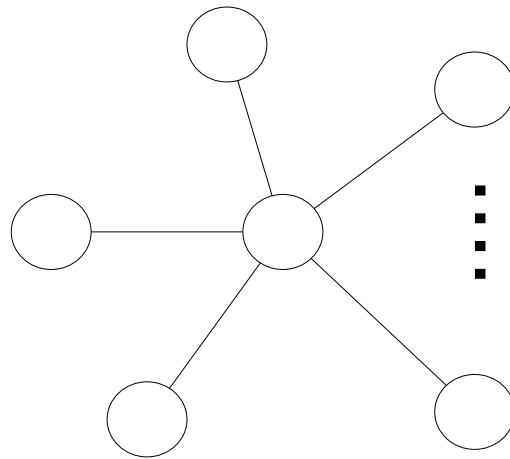
$$f = \max_i f_i = \max_i \frac{l_i}{r_i} \geq \frac{\sum_i l_i}{\sum_i r_i} = \frac{L}{R}$$

# Minimum Total Capacity

Theorem 1:

The minimum total link capacity required to serve all traffic matrices is  $L=2(R - \max_i r_i)$ , and it is achievable.

# Min. Total Capacity - Sufficiency



Theorem 1:  $\min L = 2 \sum_{i=2}^N r_i$

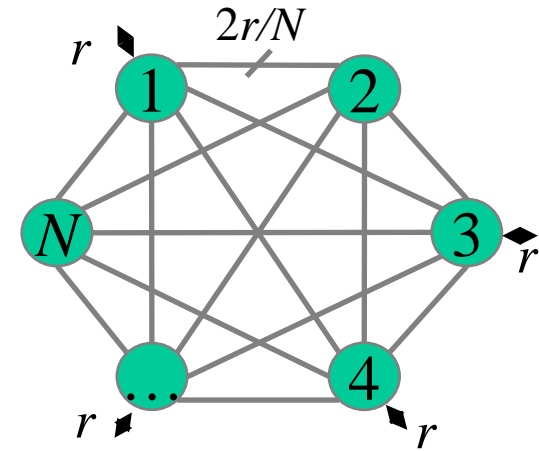
# Lower Bound on $f$

The minimum  $L$  gives a lower bound on  $f$ :

$$f \geq \frac{L}{R} \geq 2\left(1 - \frac{r_1}{R}\right)$$

# Homogeneous Full-mesh

- Each node has  $N-1$  links
- Each link has capacity  $2r/N$
- Node fanout is  $f = 2(1-1/N)$
  
- Min total link capacity is  $2r(N-1)$
- Network fanout lower bound is  $2(1-1/N)$



⇒ Homogeneous Full-mesh is optimal

# Minimizing $f$

- Assume *oblivious spreading*:  $p_i$  of every flow is spread to node  $i$ ,  $\sum_i p_i = 1$ .
- Theorem 2: Under oblivious spreading,

$$\min f = 1 + \frac{1}{\sum_j \frac{r_j}{R-2r_j}} \quad p_n = \frac{\frac{r_n}{R-2r_n}}{\sum_j \frac{r_j}{R-2r_j}}$$

$$c_{ij} = r_i p_j + r_j p_i$$

# VLB Is Close to Lower Bound

Theorem 3: The oblivious optimal capacity is at most 1.2 times of the capacity lower bound

$$\frac{R \min f}{\min L} \leq \frac{\sqrt{2} + 1}{2} \simeq 1.207$$

# Summary

- Minimize  $L$ :  $\min L = 2(\sum_i r_i - \max_i r_i)$

- Minimize  $f$  (under oblivious routing):

$$\min f = 1 + \frac{1}{\sum_j \frac{r_j}{R-2r_j}} \quad p_n = \frac{\frac{r_n}{R-2r_n}}{\sum_j \frac{r_j}{R-2r_j}}$$

- VLB is efficient:

$$R \min f \leq 1.2 \min L$$

# Conclusion

- VLB can guarantee service to *all* traffic matrices
- VLB is efficient
- VLB makes network design easier
- Fast failure recovery
- Simpler routers

# Open Issues

- Worst case propagation delay doubled
  - Low variance in delay
  - There are “express paths”
  - Adaptive algorithm with less LB under low load
- Understand physical network constraints
  - Optimal mapping of logical full mesh
  - Provision for failures
- Create a competitive environment