

# CoopNet: Cooperative Networking

Venkat Padmanabhan  
Microsoft Research

September 2002

Microsoft  
**Research**

CoopNet

# Collaborators

- MSR Researchers
  - Phil Chou
  - Helen Wang
- MSR Intern
  - Kay Sripanidkulchai (CMU)

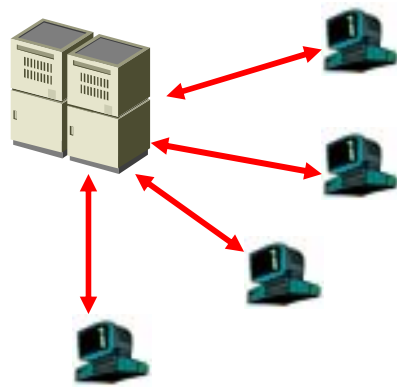
# Outline

- CoopNet
  - motivation and overview
  - web content distribution
  - streaming media content distribution
    - multiple description coding
    - multiple distribution trees
  - related work
  - summary and ongoing work
- Other networking projects at MSR

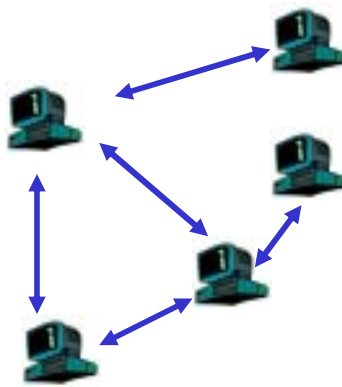
# Motivation

- A **flash crowd** can easily overwhelm a server
  - often due to news event of widespread interest...
  - ... but not always (e.g., Webcast of birthday party)
  - can affect relatively obscure sites (e.g., `election.dos.state.fl.us`, `firestone.com`, `nbaa.org`)
    - site becomes unreachable precisely when popular!
  - affects Web content as well as streaming content
  - infrastructure-based CDNs aren't for everyone
    - too expensive even for big sites (e.g., CNN)
    - uninteresting for CDN to support small sites
- **Goal:** solve the flash crowd problem without requiring new infrastructure!

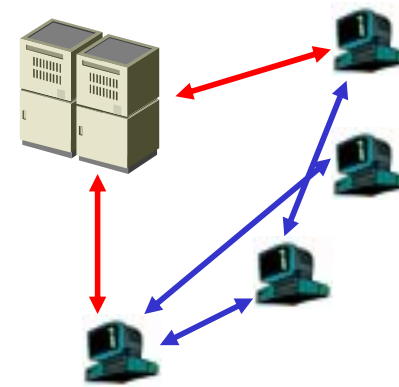
# Cooperative Networking



*Client-server*



*Pure peer-to-peer*



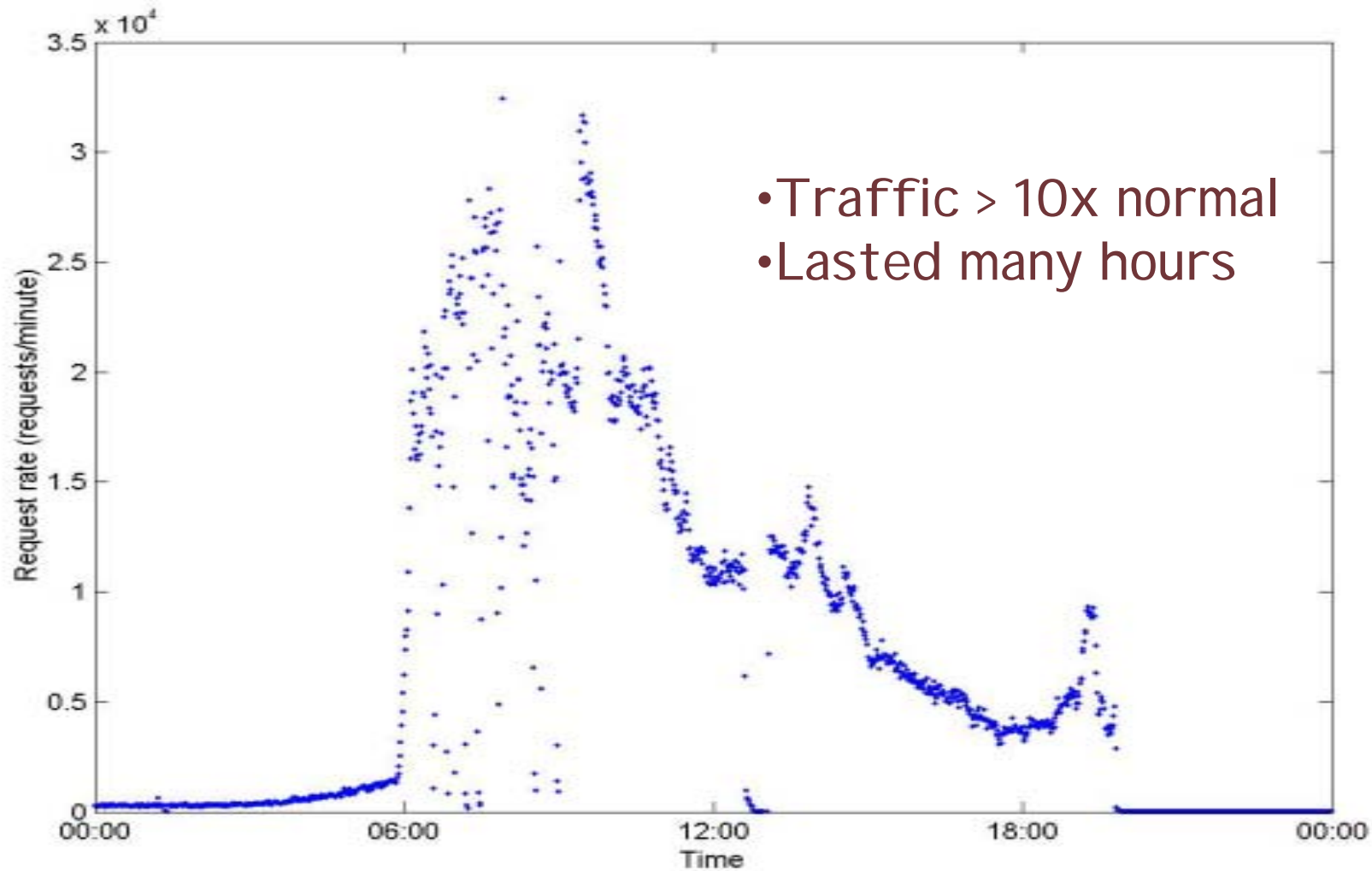
*CoopNet*

- CoopNet complements client-server system
  - client-server operation in normal times
  - P2P content distribution invoked on demand to alleviate server overload
  - clients participate only while interested in the content
  - server still plays a critical role

# CoopNet Tradeoffs

- Avoids dependence on expensive CDN infrastructure
  - but no performance “guarantees”
- P2P network size scales with load
- Availability of resourceful server simplifies many P2P tasks
  - but is the server a potential bottleneck?

# Flash Crowd Characteristics



# Where is the bottleneck?

- Disk?
  - no, most requests are for popular content
  - MSNBC: 90% of requests were for 141 files
- CPU?
  - perhaps for dynamic content
  - a single server node can pump out > 1 Gbps
- Network?
  - yes, most likely close to the server
  - 65% of servers have bottleneck bandwidths of less than 1.5 Mbps (Stefan Saroiu, U.W.)



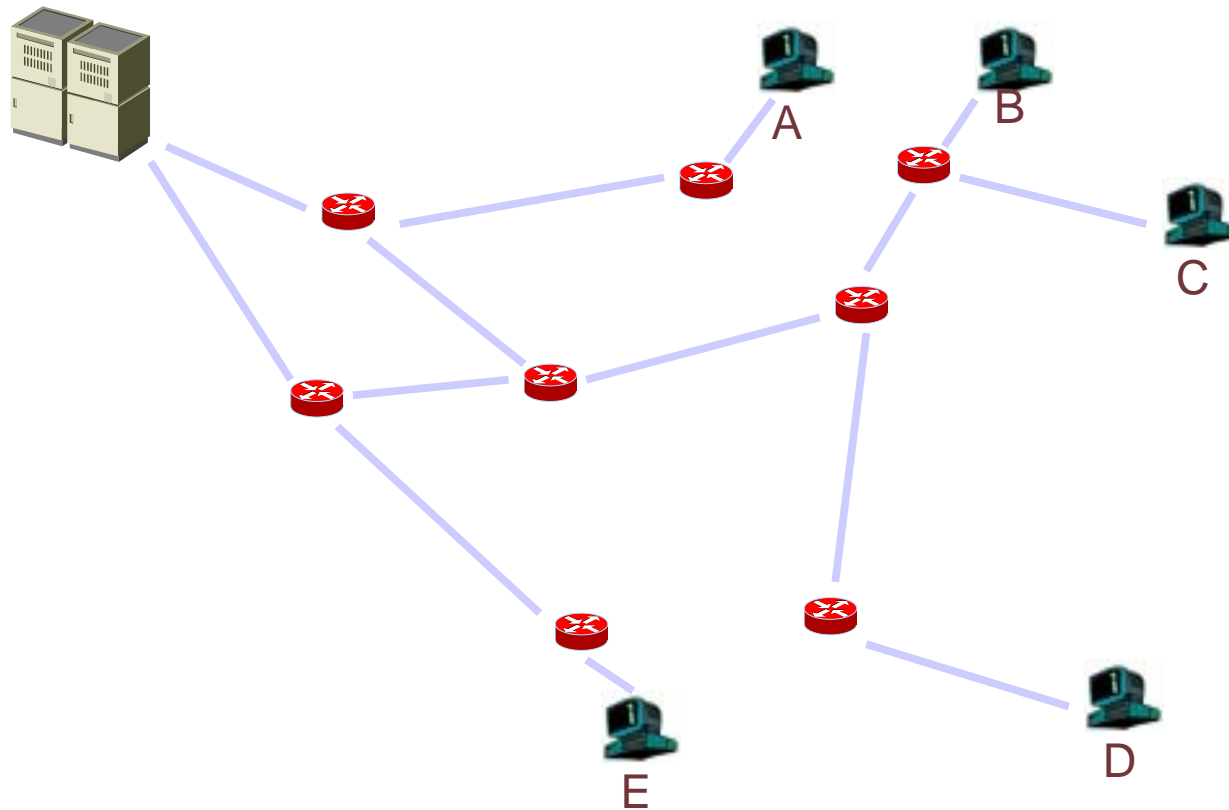
# Outline

- CoopNet
  - motivation and overview
  - web content distribution
  - streaming media content distribution
    - multiple description coding
    - multiple distribution trees
  - related work
  - summary and ongoing work
- Other networking projects at MSR

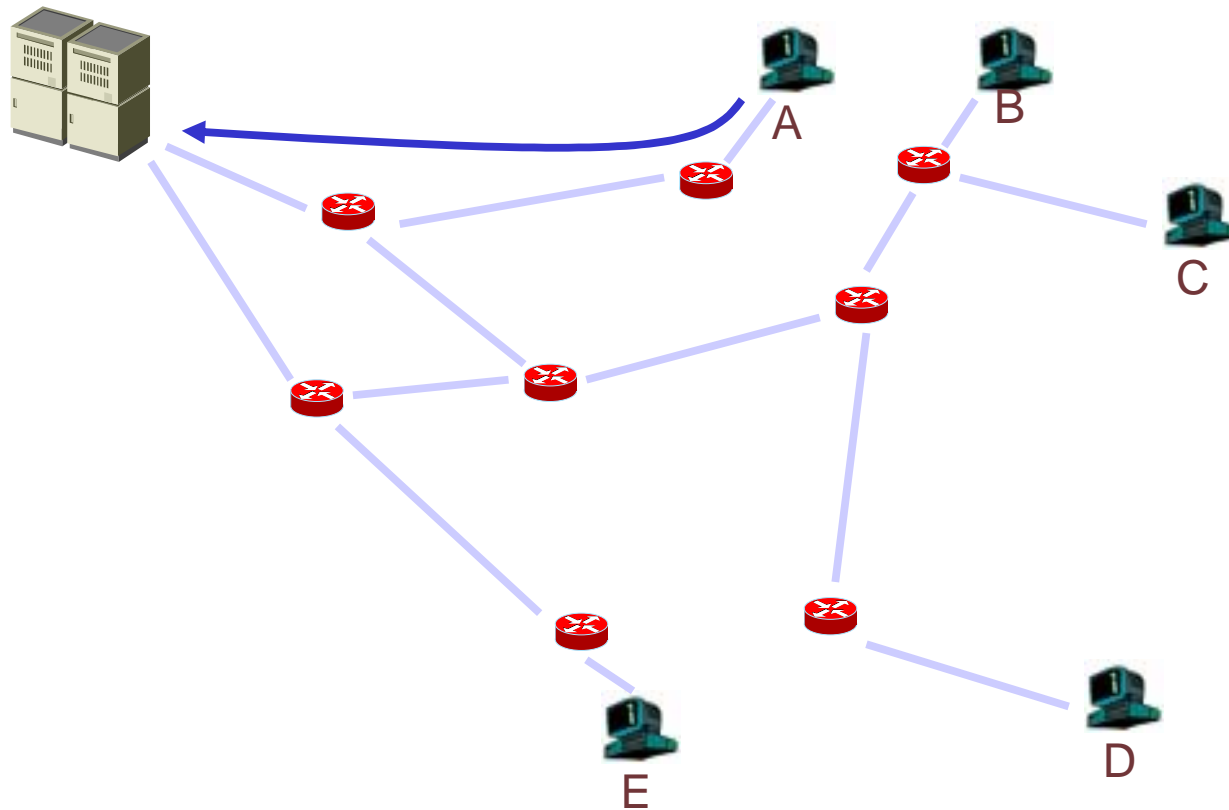
# CoopNet for Web Content

- Server maintains a cache of recent client IP addresses
- When overloaded, it redirects new clients to old ones that have the content
- Huge bandwidth savings (100X)
  - 200 B redirect instead of 20 KB page

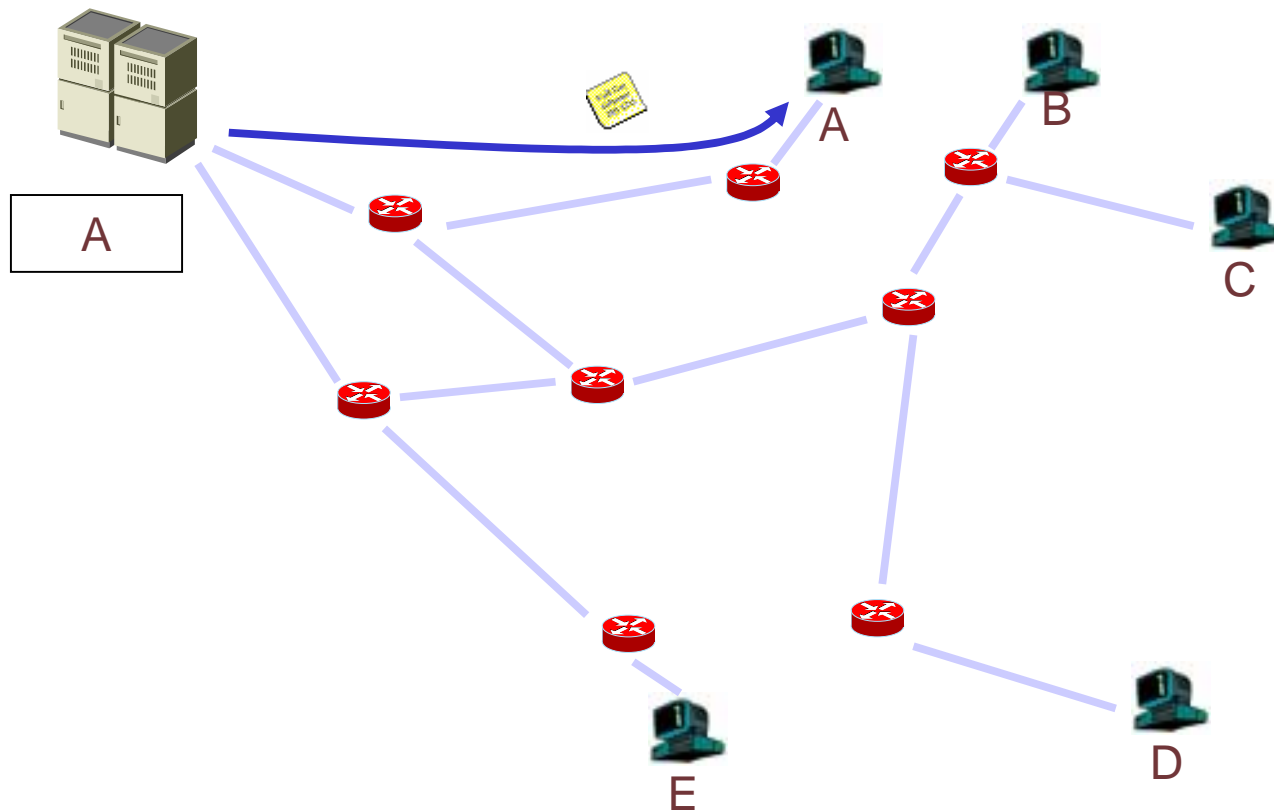
# Operation of CoopNet



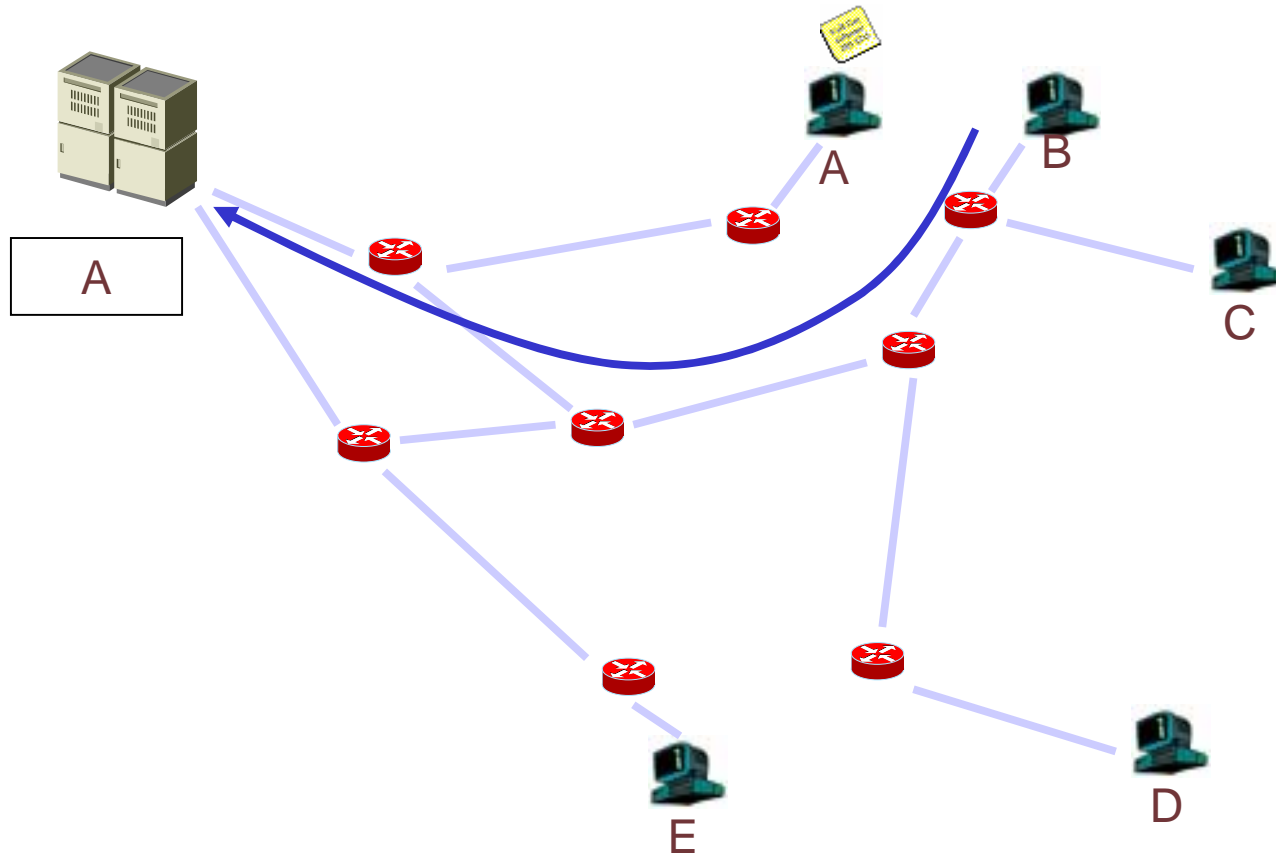
# Operation of CoopNet



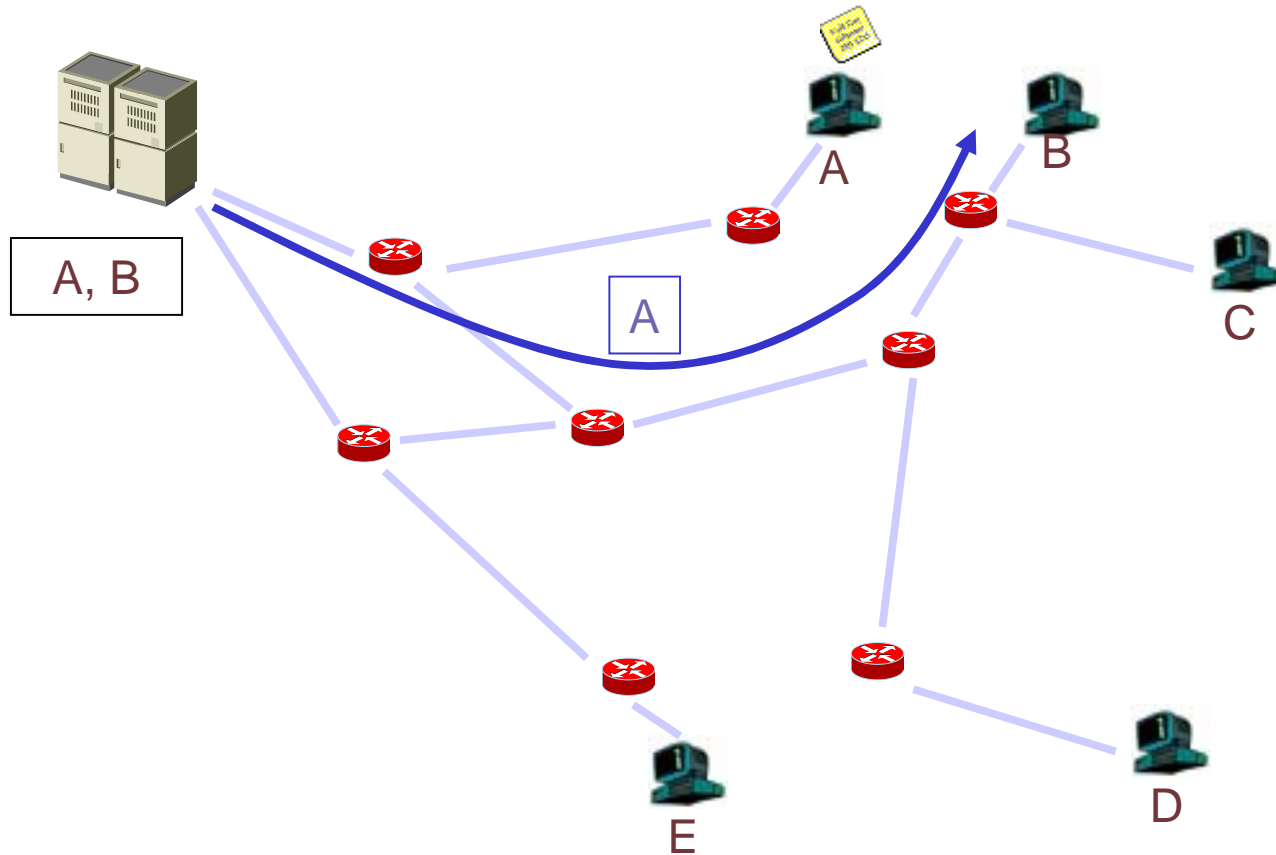
# Operation of CoopNet



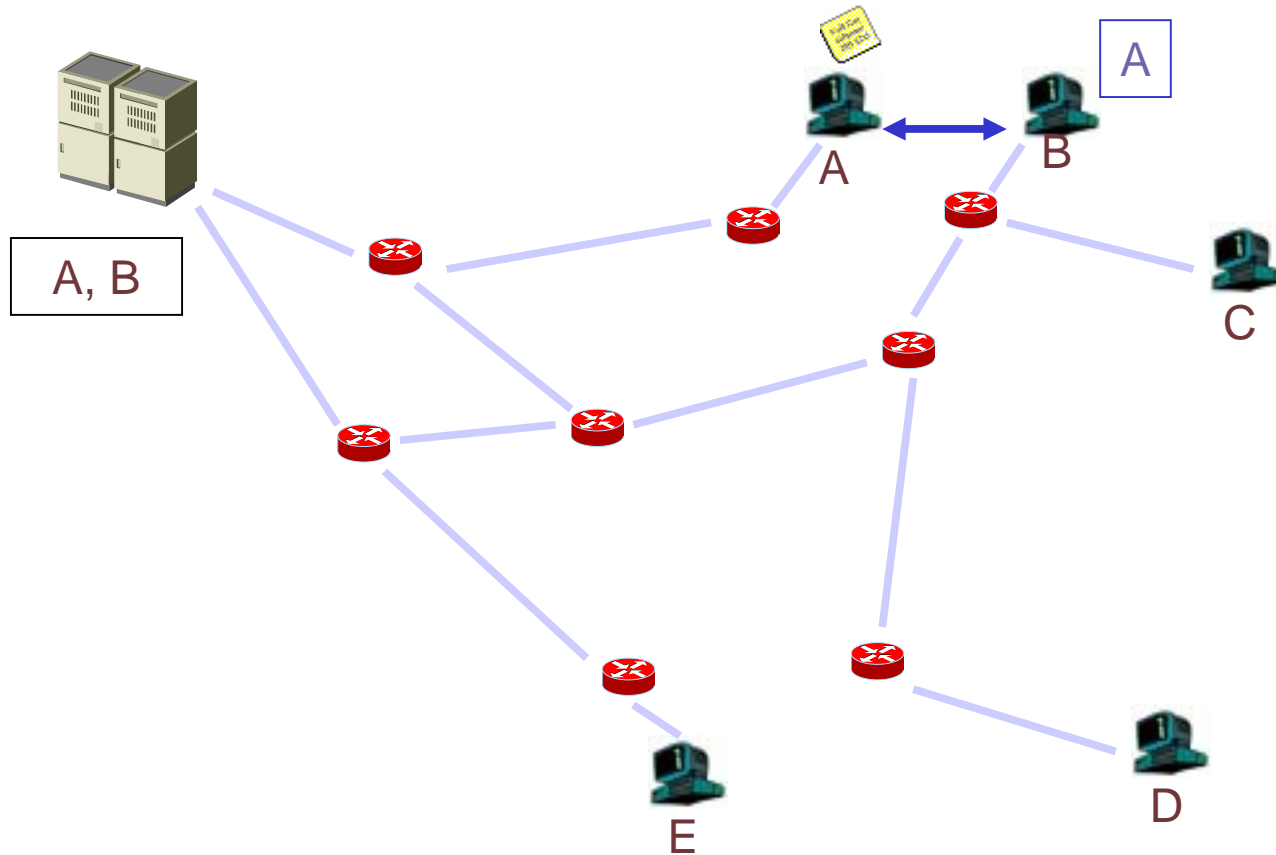
# Operation of CoopNet



# Operation of CoopNet

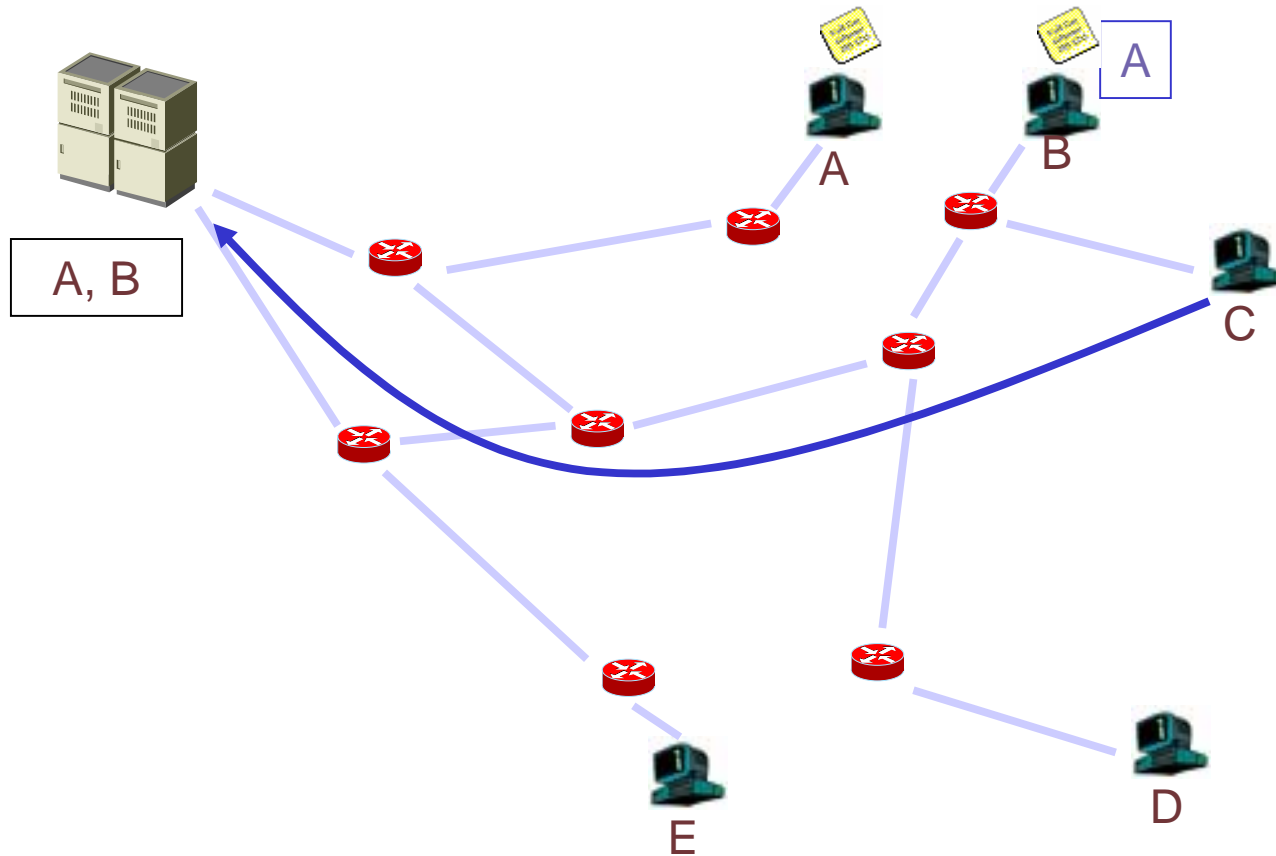


# Operation of CoopNet

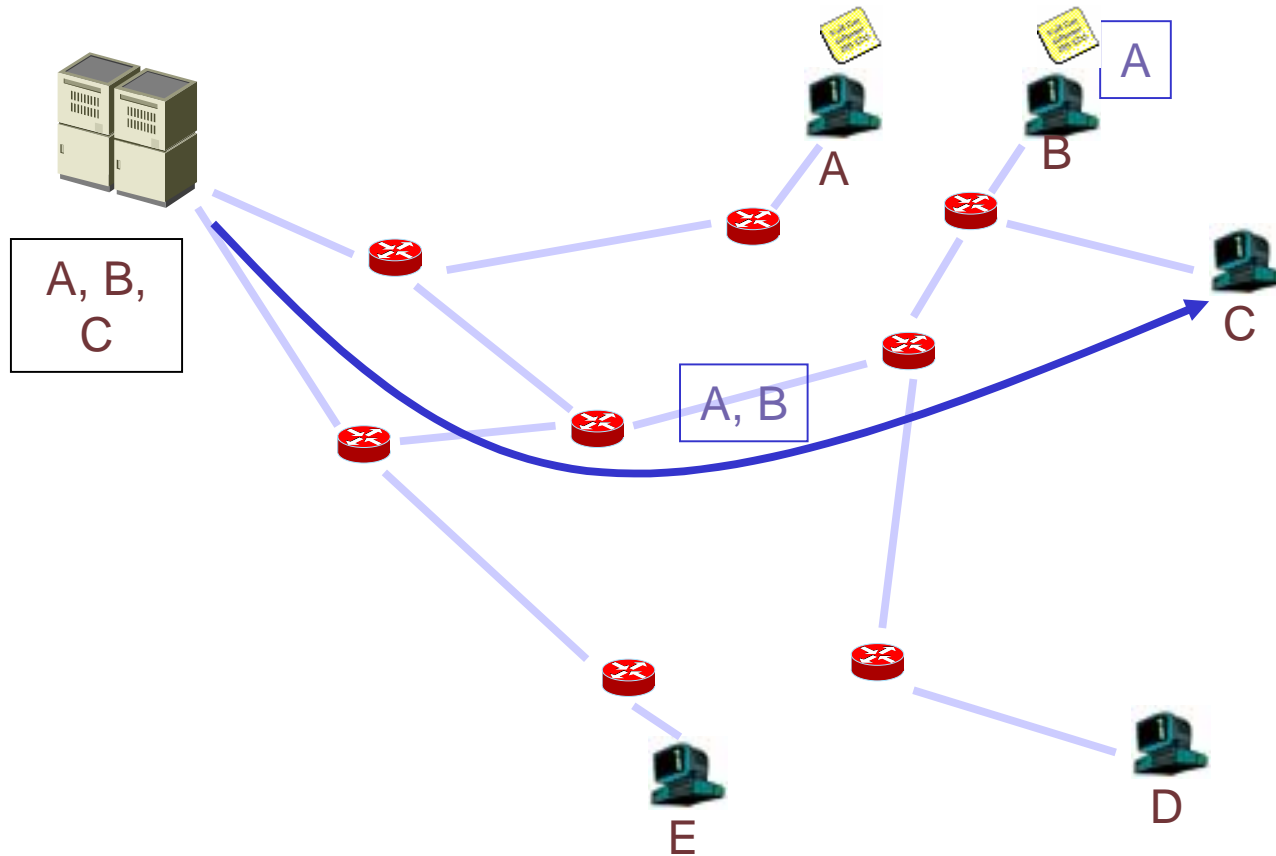




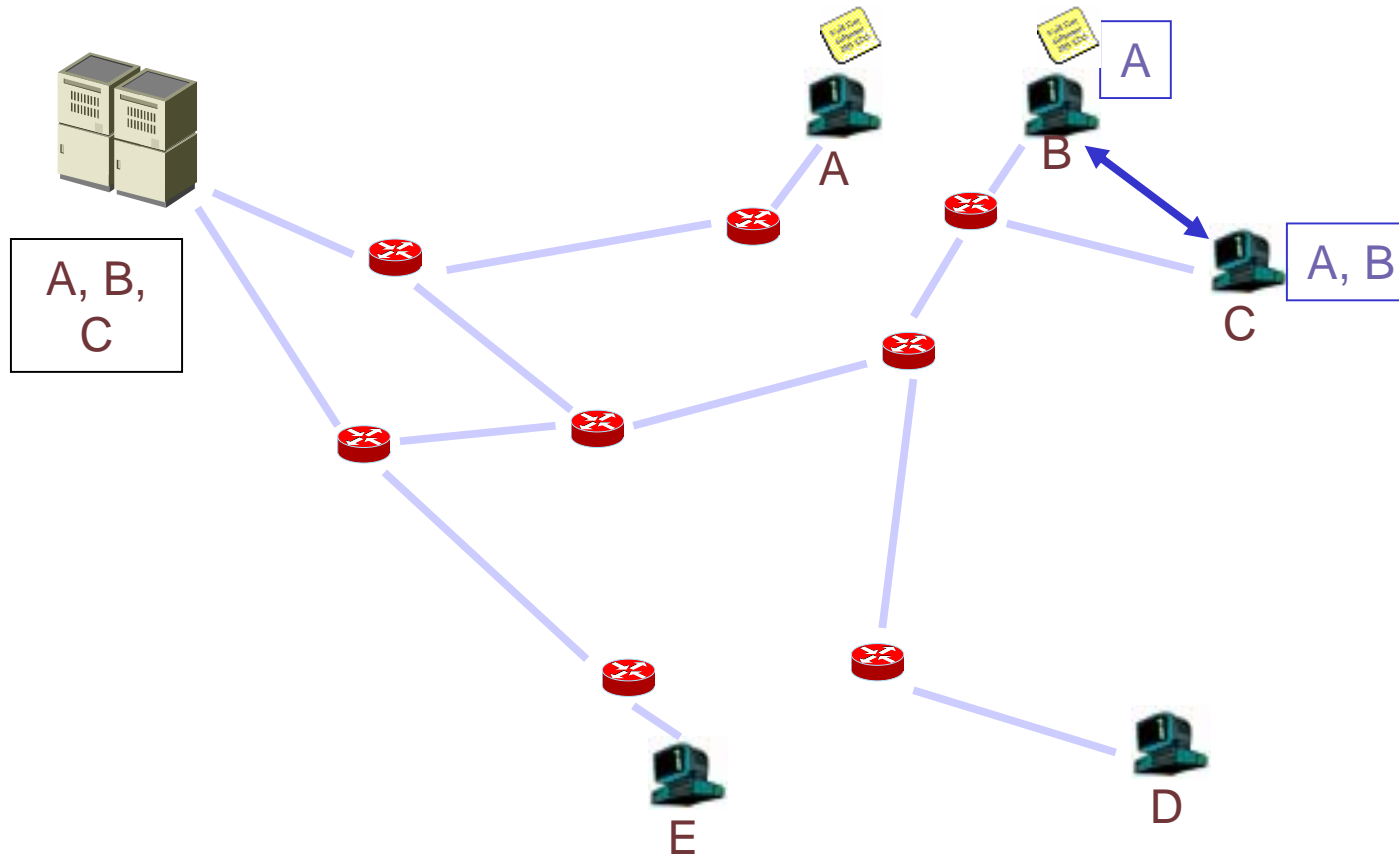
# Operation of CoopNet



# Operation of CoopNet



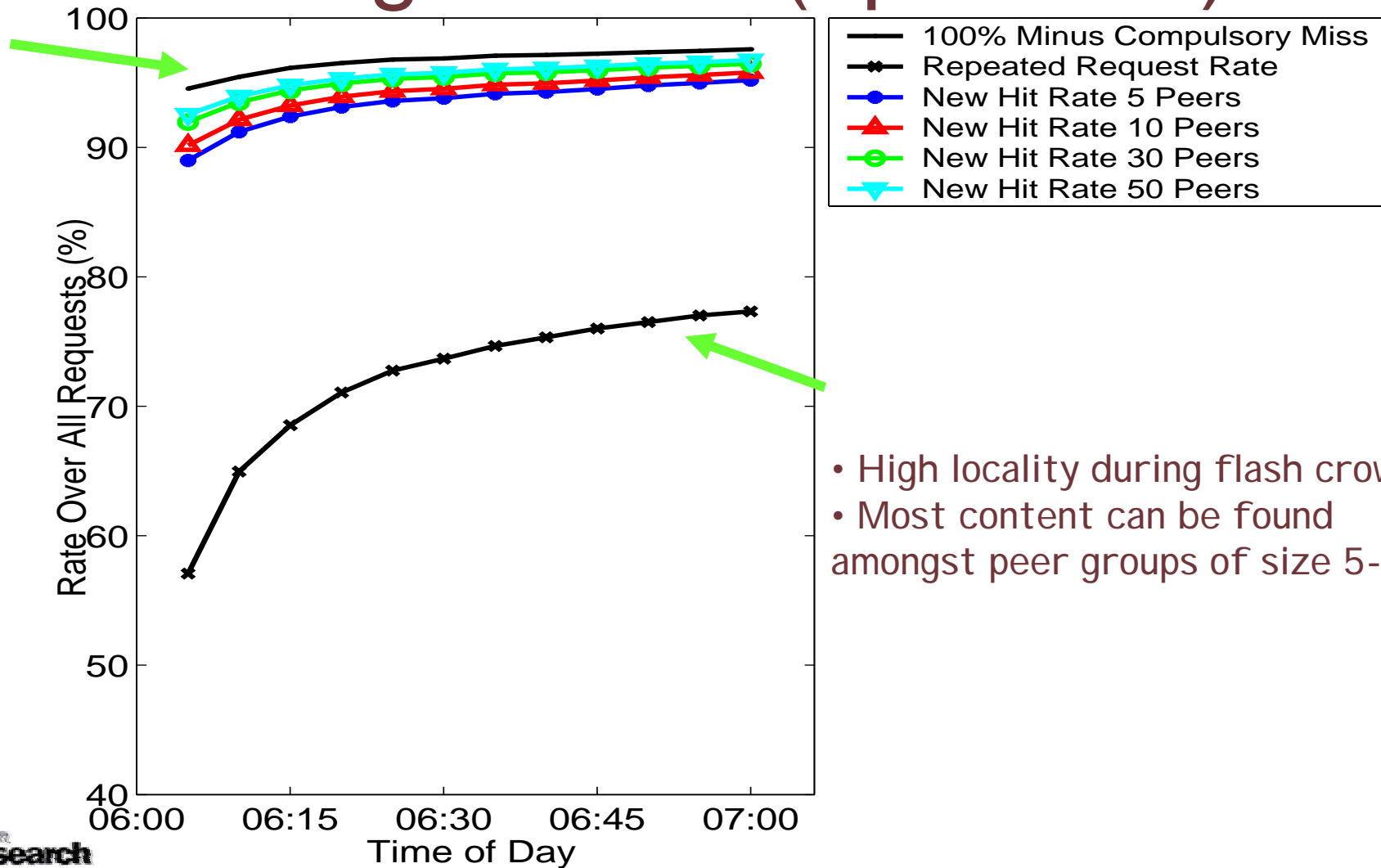
# Operation of CoopNet



# Issues

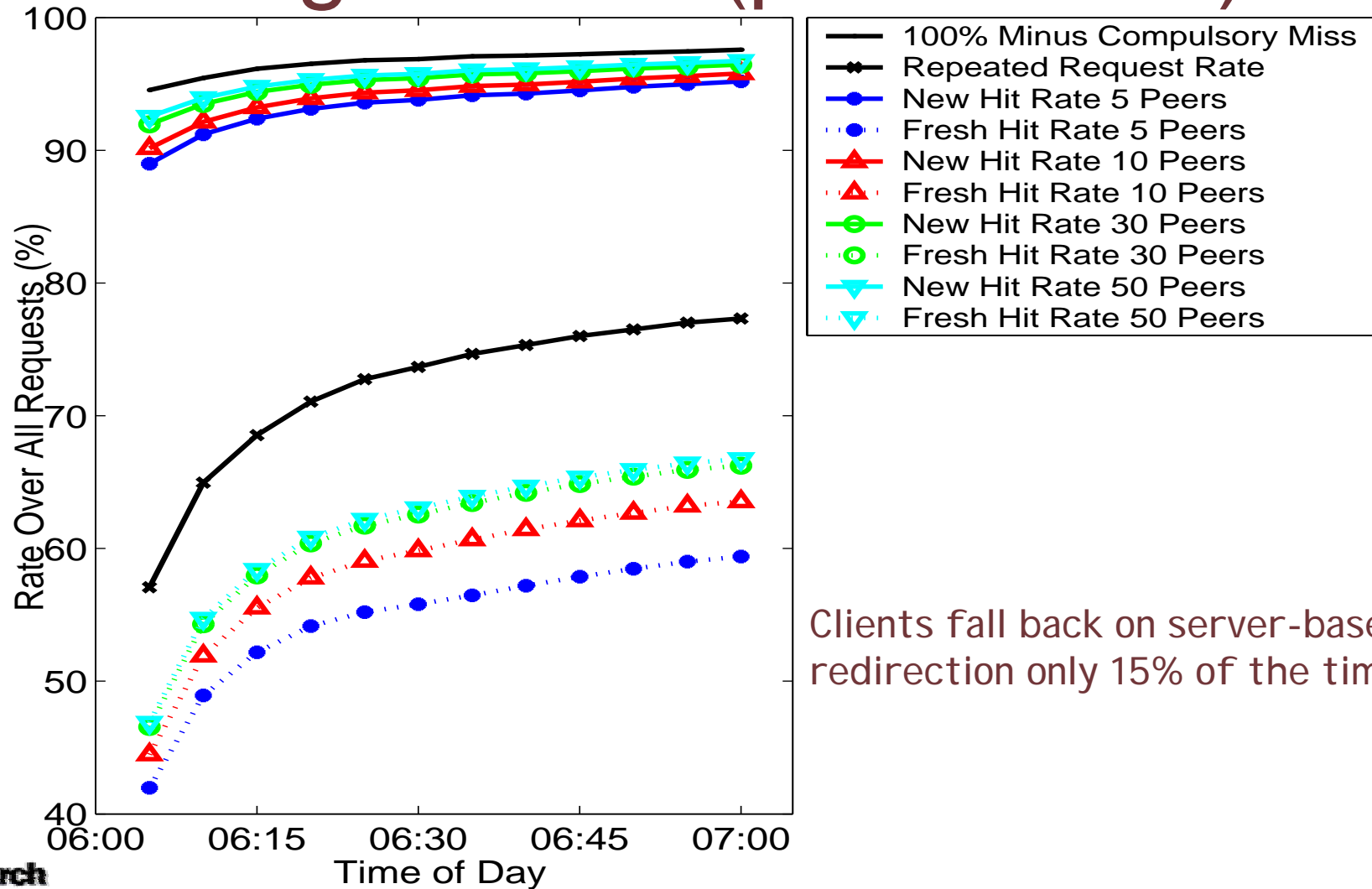
- Peer selection
  - network proximity: BGP prefix, delay-based coordinates
  - matching peer bandwidth
- Server bottleneck
  - large # of CoopNet peers ♥ large volume of redirects
  - small # of CoopNet peers ♥ server remains overloaded
  - CoopNet still beneficial, but initial redirect can take long
  - solution: initial search in **peer group**
    - high locality ♥ small group size suffices
    - greatly simplifies distributed search
    - fall back to server-based redirect upon miss
- Privacy

# Finding content (optimistic)



- High locality during flash crowd
- Most content can be found amongst peer groups of size 5-30

# Finding content (pessimistic)



Clients fall back on server-based redirection only 15% of the time

# Alternative approaches

- Proxy caching
  - deployment barriers
  - not effective when clients are scattered across the Internet
- Commercial CDNs (e.g., Akamai)
  - not cost-effective for small sites
- P2P system of servers (e.g., Backslash)
  - feasible in practice?

# Outline

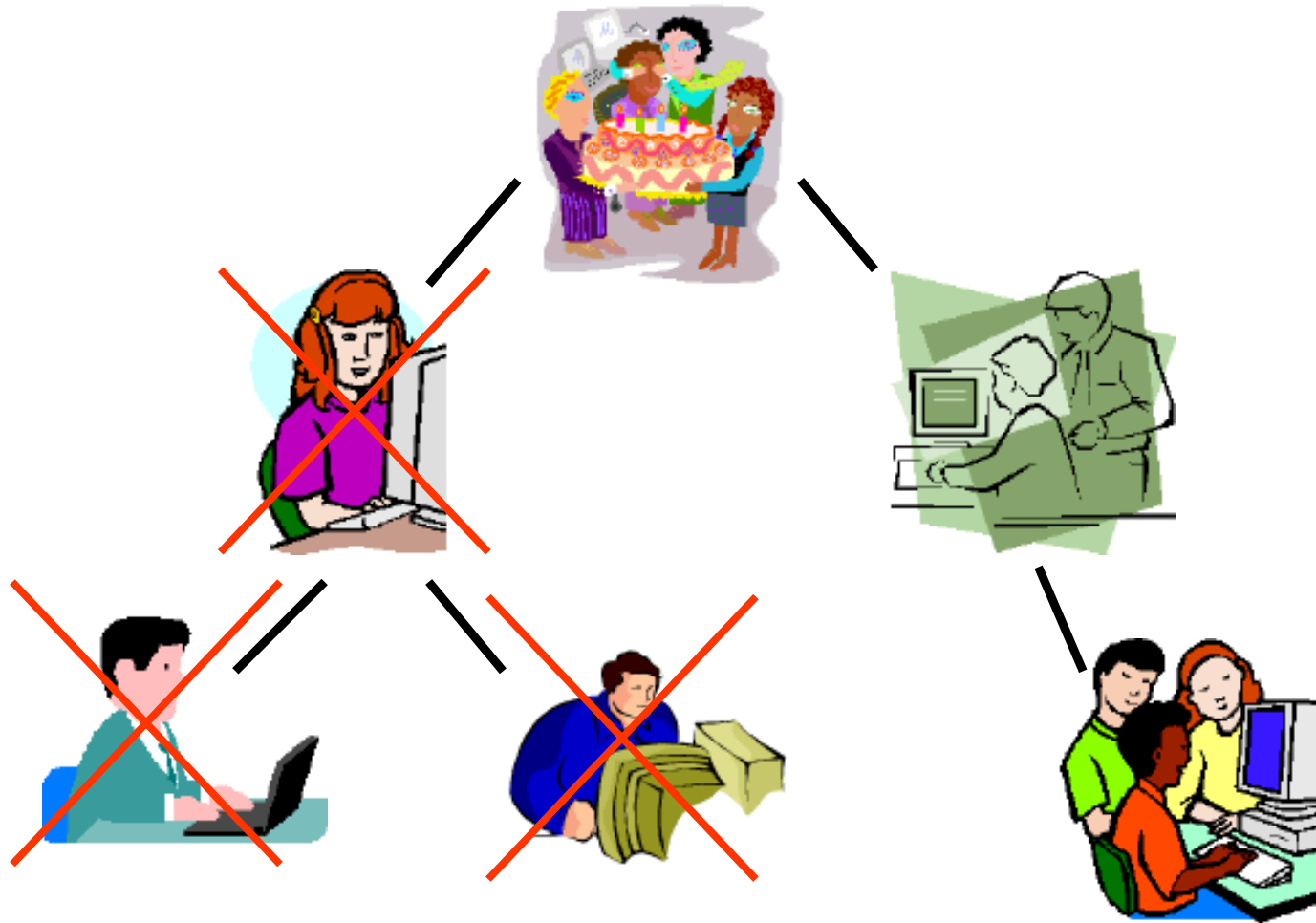
- CoopNet
  - motivation and overview
  - web content distribution
  - streaming media content distribution
    - multiple description coding
    - multiple distribution trees
  - related work
  - summary and ongoing work
- Other networking projects at MSR



# CoopNet for Live Streaming

- More likely that server will be overwhelmed
- Key issue: robustness
  - peers are not dedicated servers ♥ potential disruption due to:
    - node departures and failures
    - higher priority traffic
  - traditional application-level multicast (ALM) falls short

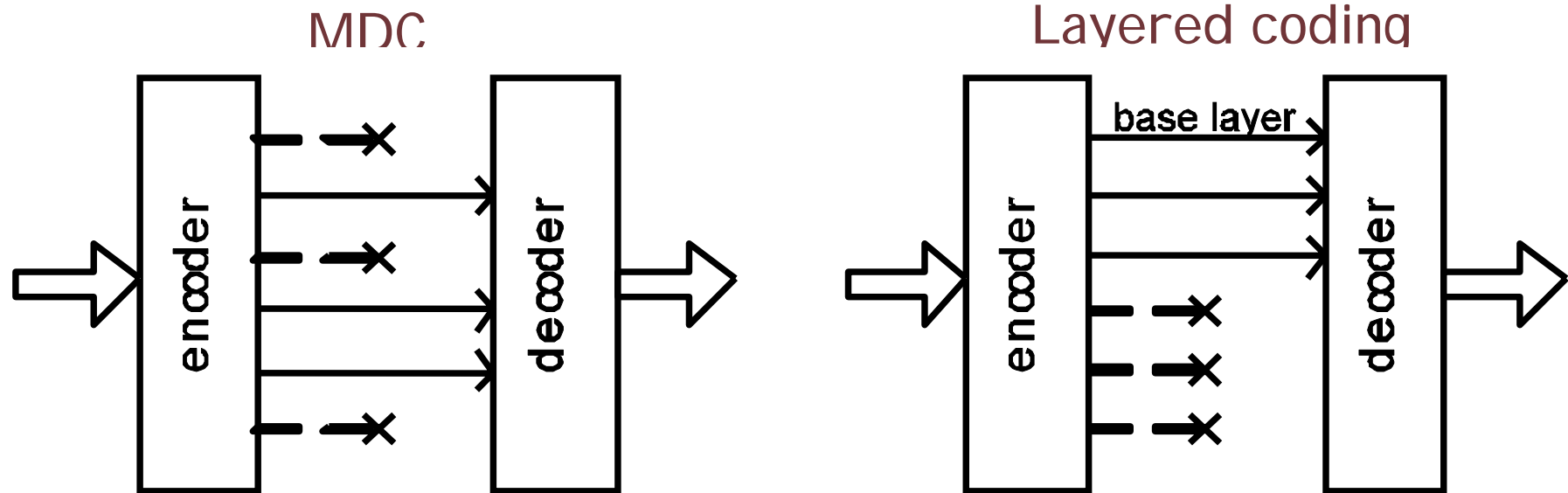
# Traditional Application-level Multicast



# CoopNet Approach to Robustness

- Add redundancy in data...
  - multiple description coding (MDC)
- ...and in network paths
  - multiple, diverse distribution trees

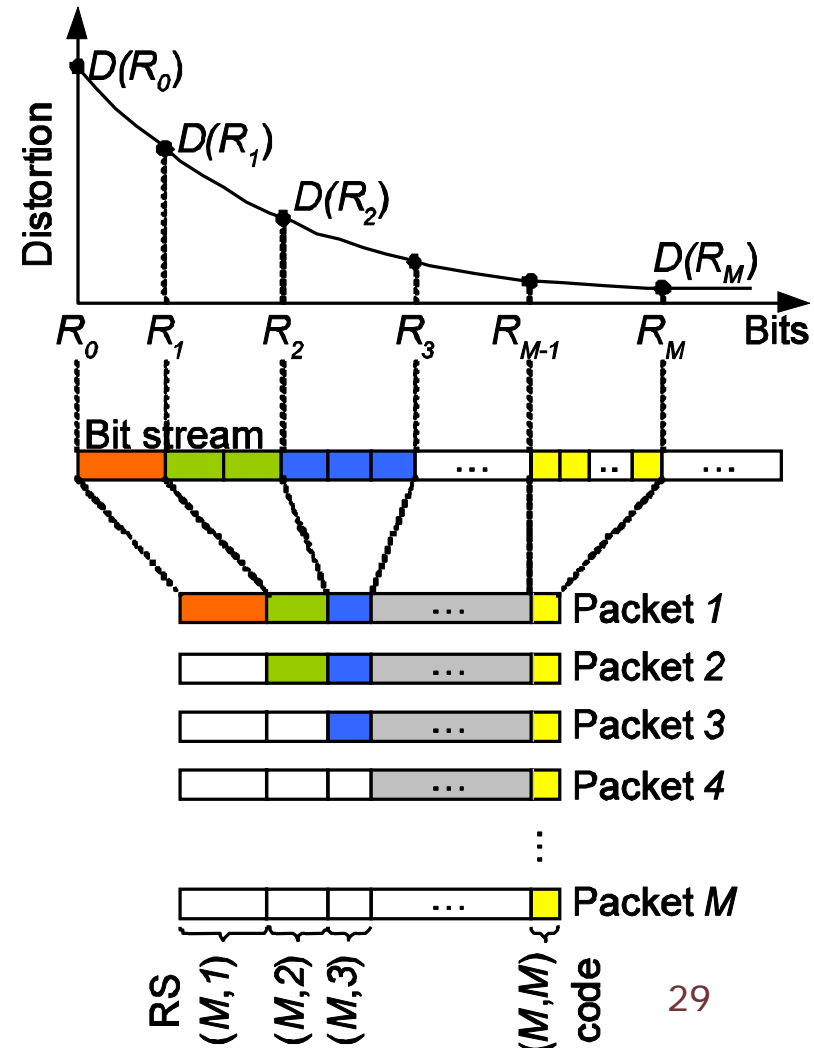
# Multiple Description Coding



- Unlike layered coding, there isn't an ordering of the descriptions
- Every subset of descriptions must be decodable
- Modest penalty relative to layered coding

# Multiple Description Coding

- Simple MDC:
  - every  $M^{\text{th}}$  frame forms a description
- More sophisticated MDC combines:
  - layered coding
  - Reed-Solomon coding
  - priority encoded transmission
  - optimized bit allocation



# Multiple Distribution Trees

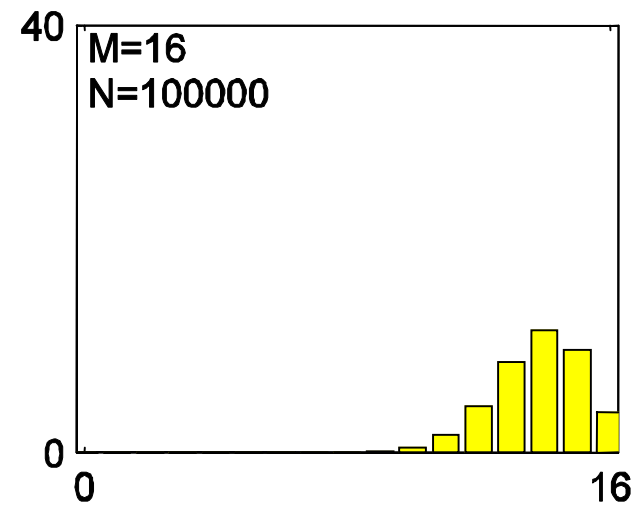
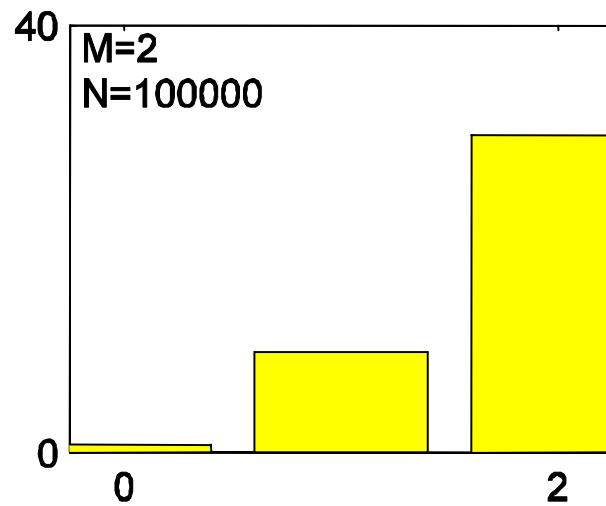
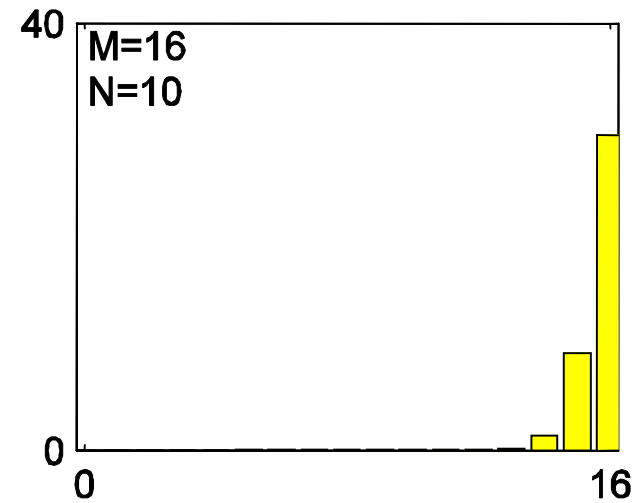
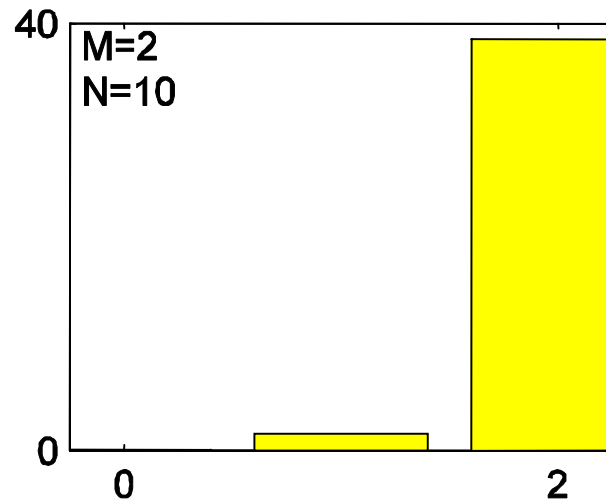


Tree diversity provides robustness to node failures

# MDC Analysis

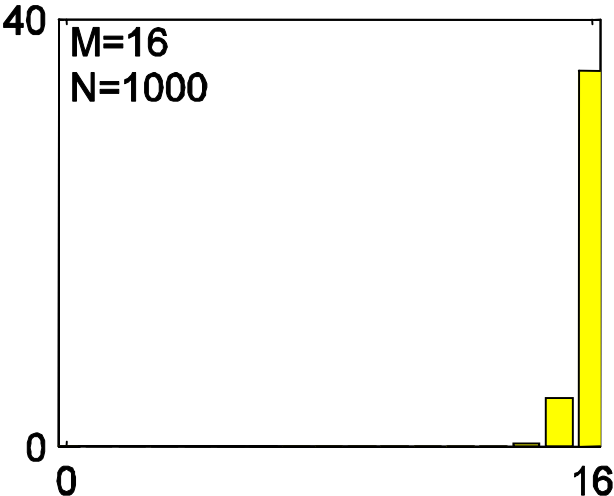
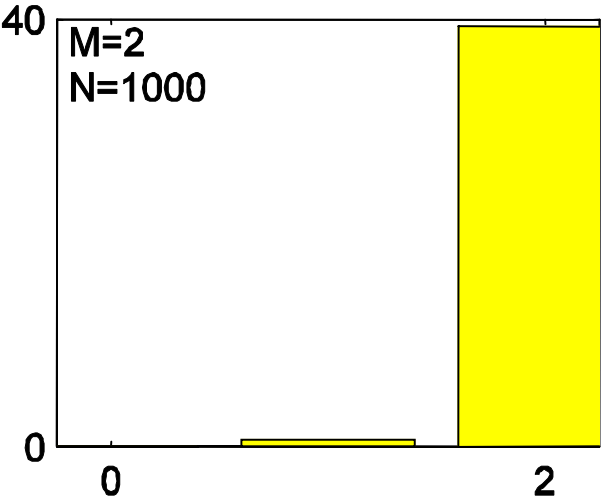
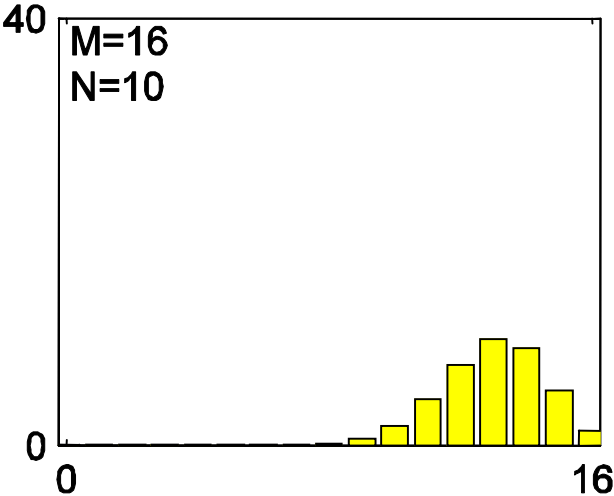
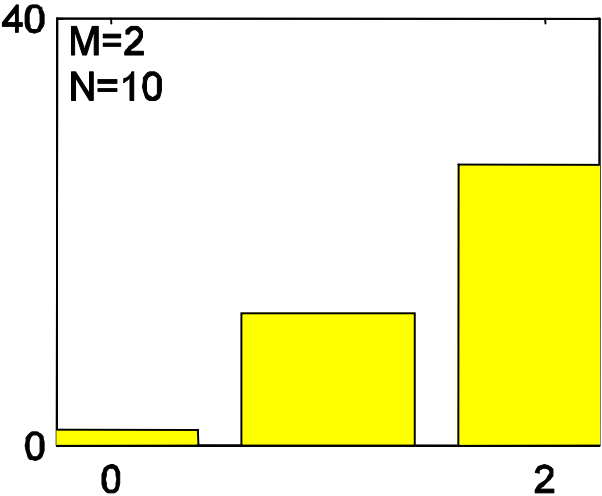
- Key parameters:
  - number of nodes (N)
  - number of descriptions (M)
  - out-degree of each node
  - repair time
  - node departure rate
- Two scenarios of interest
  - large N, high churn ♥ multiple node failures in repair interval
  - small N, stable ♥ occasional, single node failures

# Quality During Multiple Failures





# Quality During Single Failure



# Tree Management

- Goals:
  - short and wide trees
  - efficiency
  - diversity
  - quick join and leave processing
  - scalability
- CoopNet approach: centralized protocol anchored at the server
  - single point of failure...
  - ...but server is source of data anyway

# Basic Tree Management Protocol

- Nodes inform server of their arrival and departure
- Server tracks node capacity and tells new nodes where to join
  - high up in the tree but randomized
  - fan out of server is typically much larger
- Each node monitors its packet loss rate and takes action when loss rate becomes too high
- Simple, scales to 1000+ joins/leaves per sec.

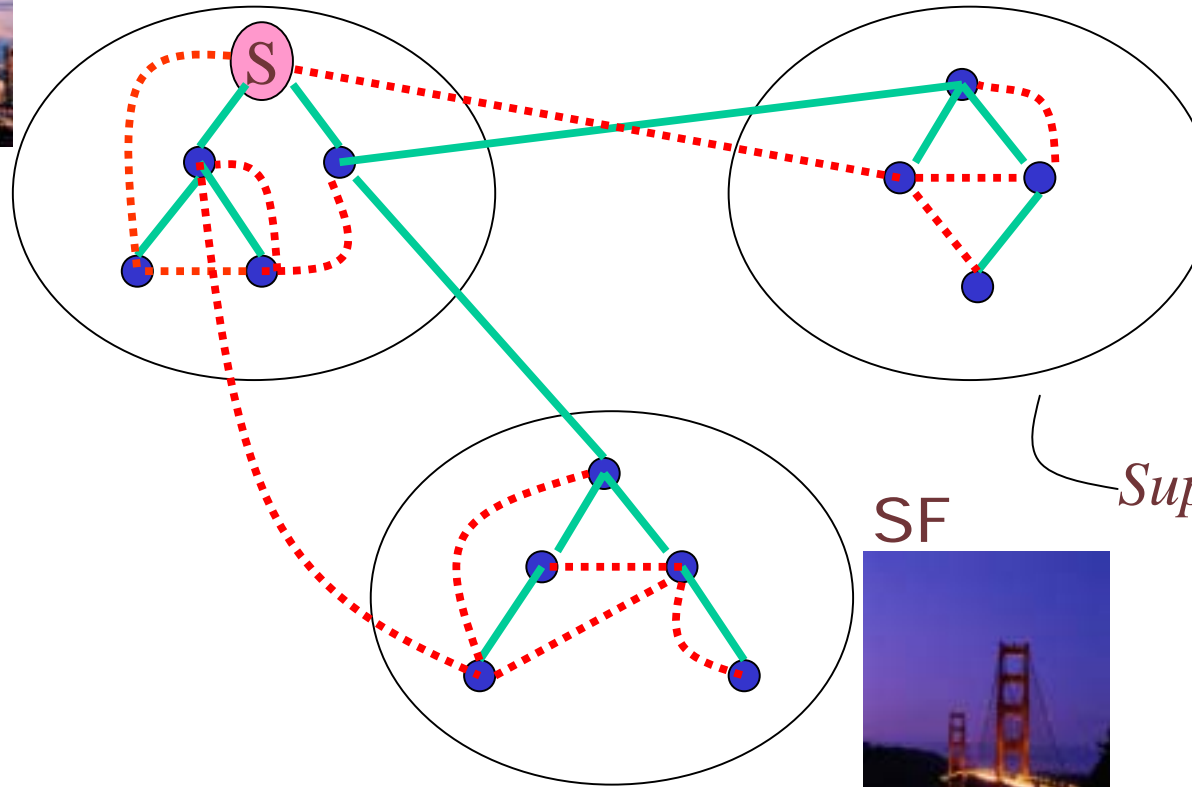
# Optimizations

- Achieving efficiency and diversity
  - cluster nodes into **super nodes** using delay-based coordinates (akin to GeoPing)
  - logical topology matches physical topology at the macroscopic level
- Migrate “stable” nodes to higher levels in the tree

# Achieving Efficiency and Diversity



SEA



NY



SF

*Supernode*



# Performance Evaluation

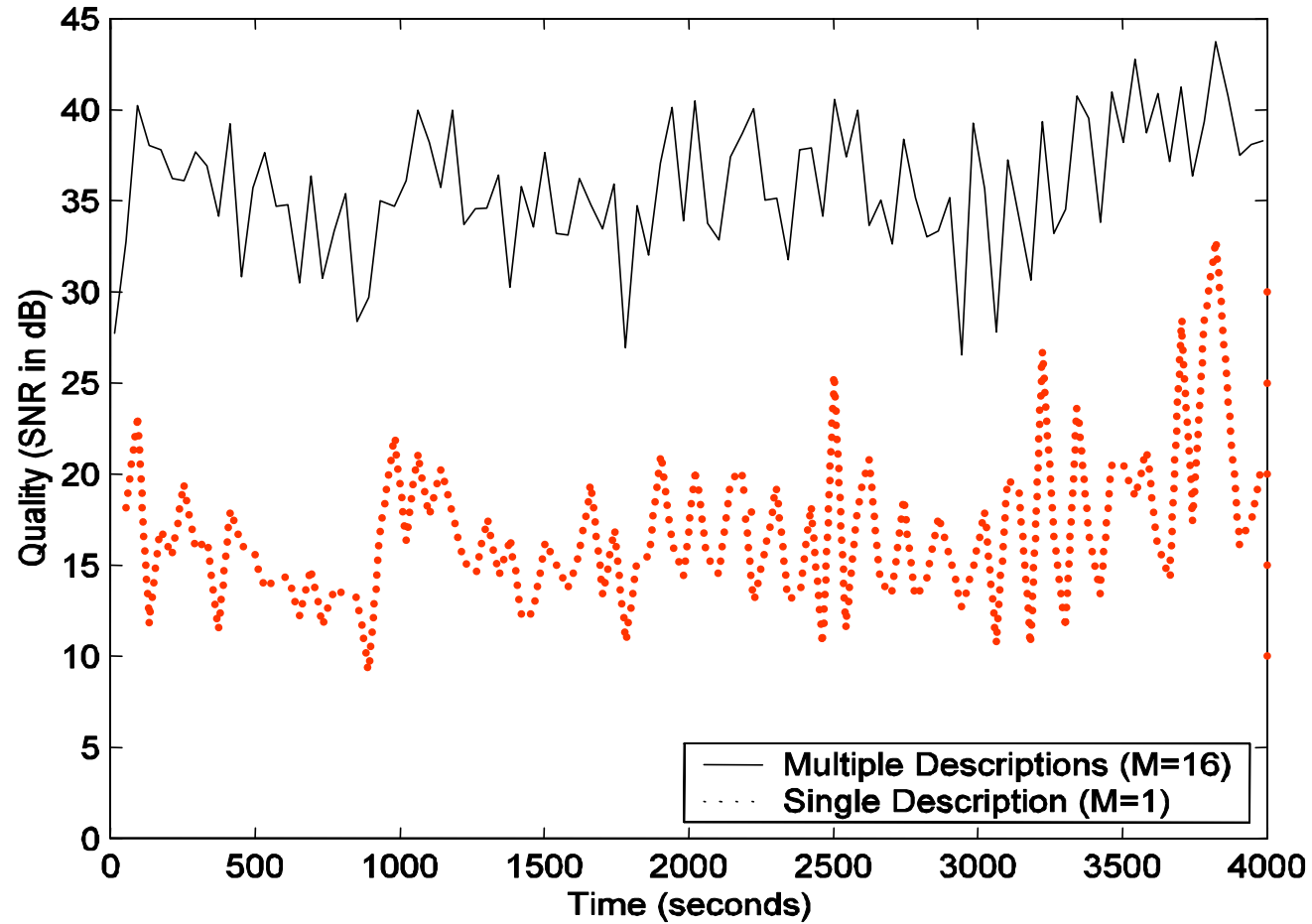
- MSNBC access logs from Sep 11, 2001
- Live streaming
  - ~18,000 simultaneous clients
  - ~180 joins/leaves per second on average; peak rate of ~1000 per second
  - ~70% of clients tuned in for less than a minute
- On-demand streaming
  - 300,000 requests in a 2-hour period

# Live Streaming

- Key questions
  - how beneficial is MDC?
  - does well is diversity preserved as trees evolve?
  - how does repair time impact performance?

# MDC versus SDC

Random Trees

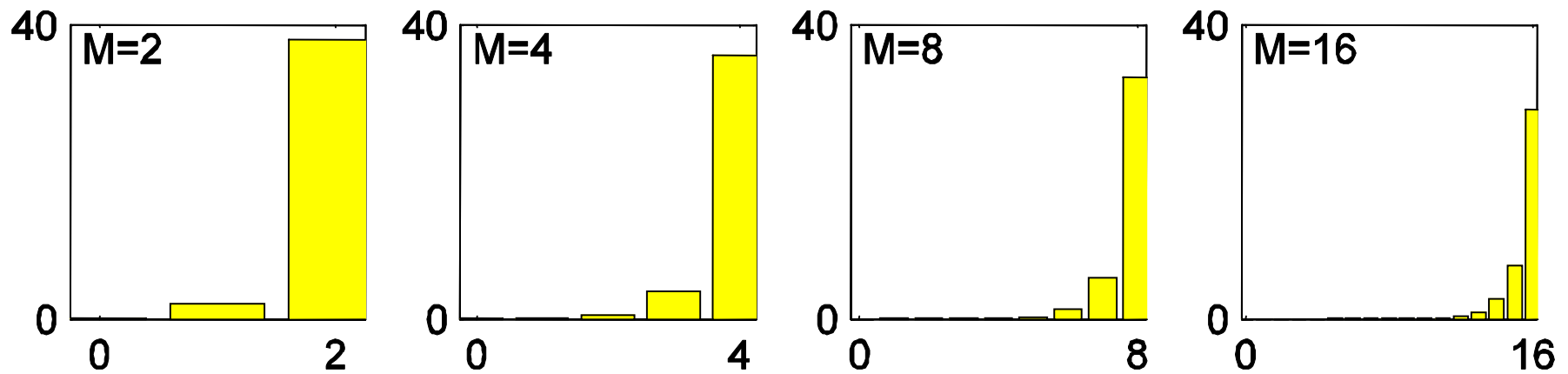


Based on MSNBC traces from Sep 11

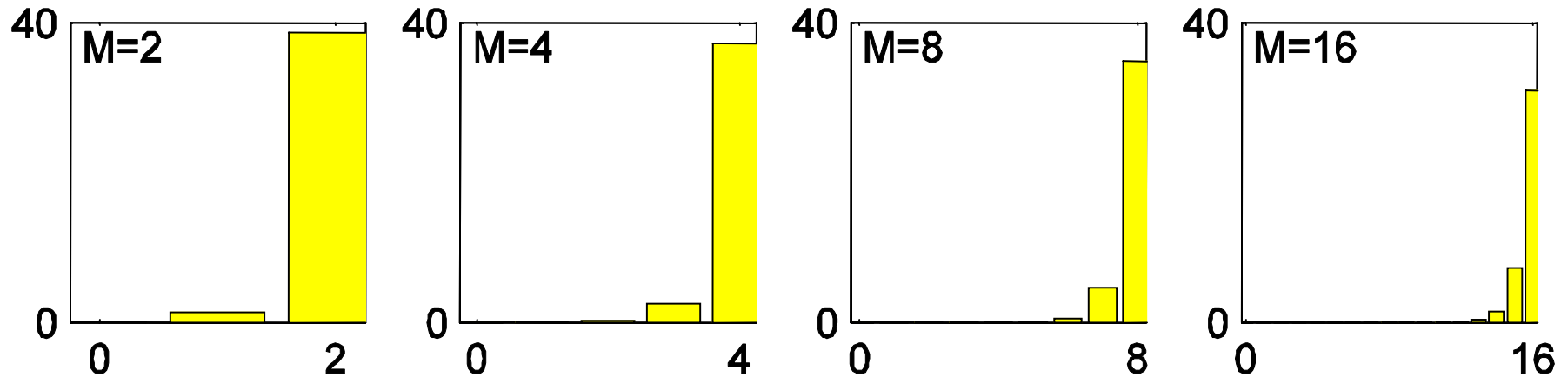


# Random Trees vs. Evolved Trees

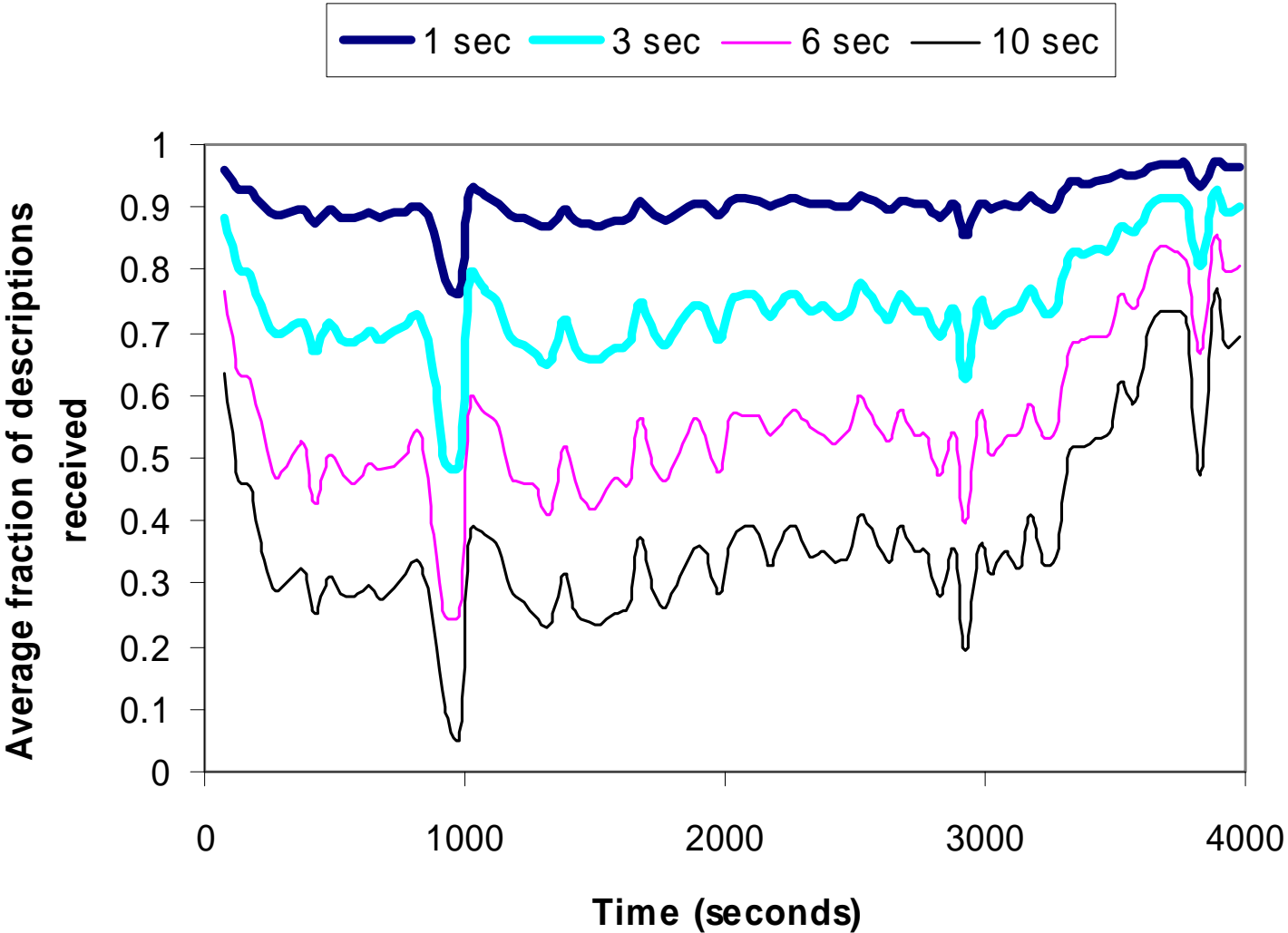
## Random Trees



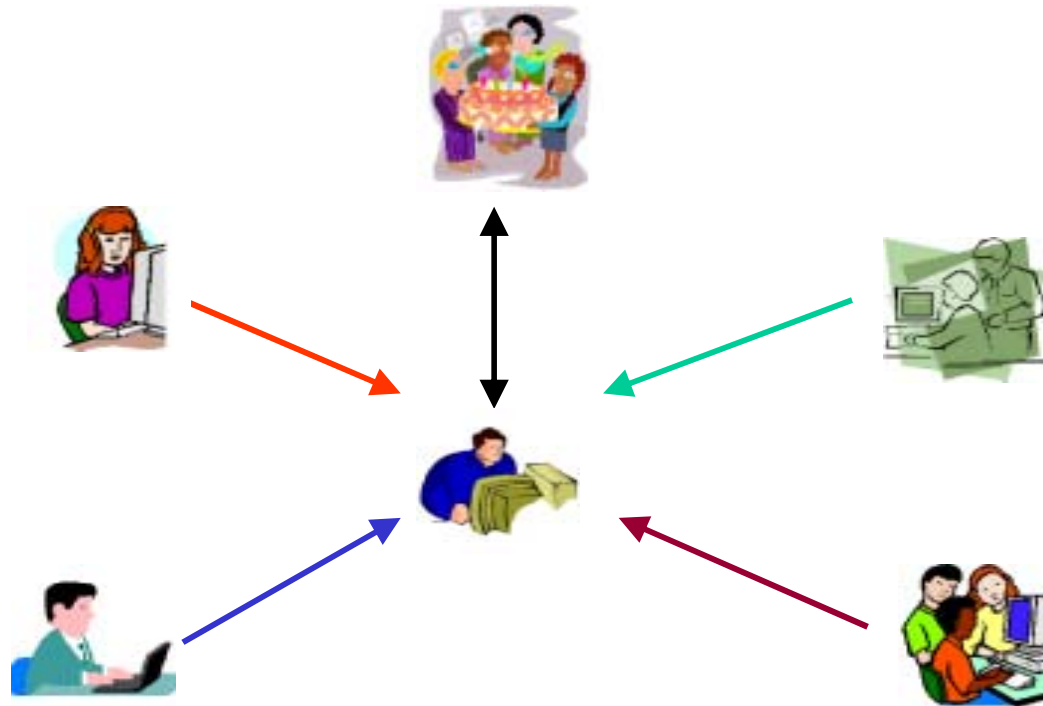
## Evolved Trees



# Impact of Repair Time



# CoopNet for On-demand Streaming

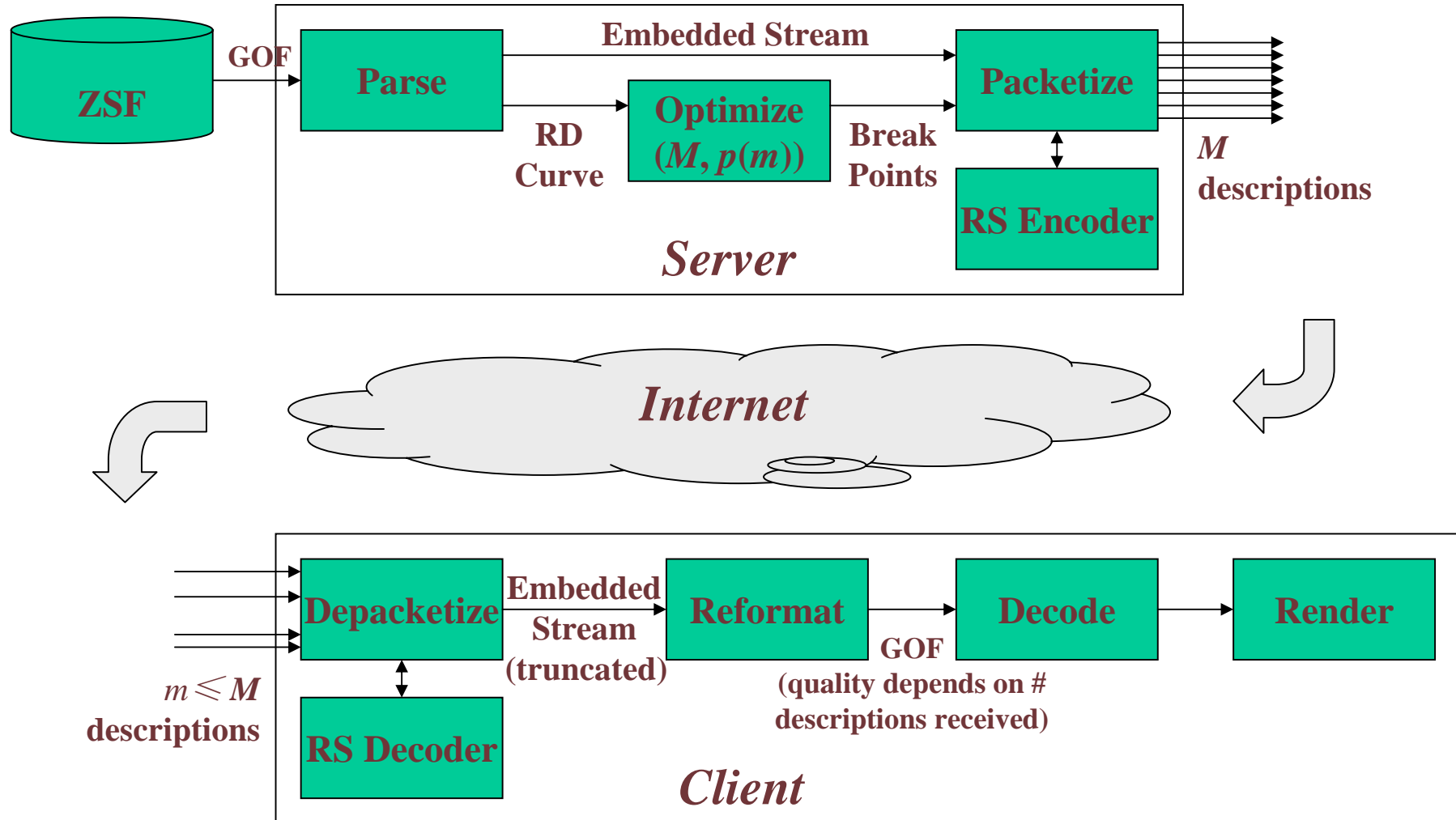


- Distributed streaming of multiple descriptions
- Improves robustness and load distribution

# On-demand Streaming

- Key results:
  - server bandwidth requirement drops from 20 Mbps to 300 Kbps
  - peer bandwidth requirement:
    - average over all peers is 45 Kbps
    - average over active peers is 465 Kbps
  - storage requirement at a peer is less than 100 MB
  - probability of finding peer in the same BGP prefix cluster is under 20%

# CoopNet Transport Architecture



# Outline

- CoopNet
  - motivation and overview
  - web content distribution
  - streaming media content distribution
    - multiple description coding
    - multiple distribution trees
  - related work
  - summary and ongoing work
- Other networking projects at MSR

# Related Work

- Infrastructure-based CDNs
  - Akamai, Digital Island
- P2P CDNs
  - Pseudo-serving, PROOFS, Backslash
  - SpreadIt, Allcast, vTrails
- Application-level multicast
  - ALMI, Narada, Scattercast
  - Bayeux, Scribe
- Multi-path content delivery
  - Byers et al. 1999, Nguyen & Zakhor 2002, Apostolopoulos et al. 2002

# Summary

- Client-server applications can benefit from selective use of peer-to-peer communications
- Availability of server simplifies system design
- Web content
  - high degree of locality
  - server-based redirection plus small peer group
- Streaming content
  - robustness to dynamic membership is the key challenge
  - MDC with multiple, diverse distribution trees improves robustness in peer-to-peer media streaming
  - centralized tree management is efficient and can scale



# Ongoing Work

- Prototype implementation
- Dealing with client heterogeneity for live streaming
  - combine MDC with layering

- More info:

[research.microsoft.com/~padmanab/projects/CoopNet](http://research.microsoft.com/~padmanab/projects/CoopNet)

- Papers at IPTPS '02 and NOSSDAV '02

# Outline

- CoopNet
  - motivation and overview
  - web content distribution
  - streaming media content distribution
    - multiple description coding
    - multiple distribution trees
  - related work
  - summary and ongoing work
- Other networking projects at MSR

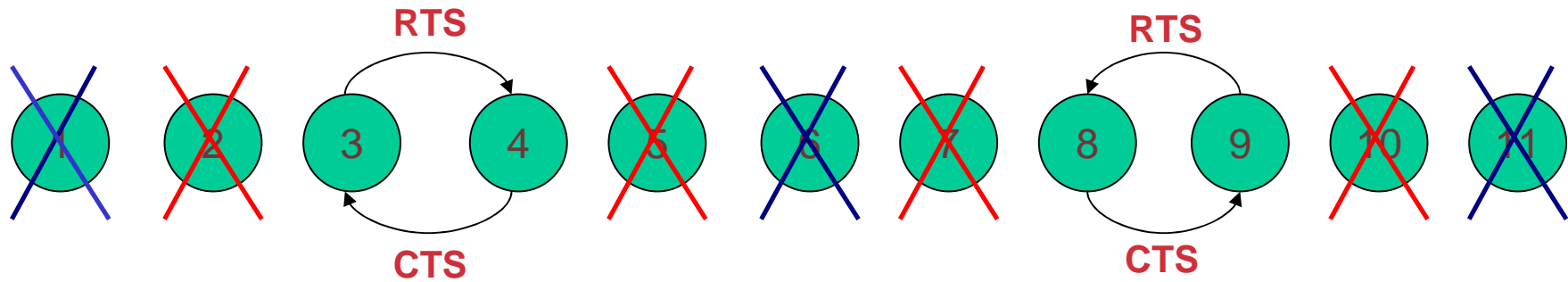
# Networking Research at MSR

- Internet measurement and performance
  - Passive Network Tomography
  - IP2Geo: Internet Geography
  - PeerMetric: broadband network performance
- Peer-to-Peer networking
  - Herald: scalable event notification system
  - CoopNet: P2P content distribution
- Wireless networking
  - UCoM: energy-efficient networking
  - Mesh Networks: multi-hop wireless access network

# PeerMetric

- Goal: characterize broadband network performance
  - DSL, cable modem, satellite, etc.
- P2P as well as client-server performance
- Deployment on ~25 distributed nodes underway
  - none in Atlanta — volunteers welcome!
- Joint work with Karthik Lakshminarayanan (MSR intern from Berkeley)

# Mesh Networks: Capacity is the Key Challenge



4 nodes are active, 2 packets in flight  
(example courtesy of Victor Bahl)