GBN 97

High Speed Web: An Application for Active Caching

Samrat Bhattacharjee   Kenneth L. Calvert   Ellen W. Zegura

Networking and Telecommunications Group
College of Computing
Georgia Institute of Technology
Atlanta, Georgia, USA

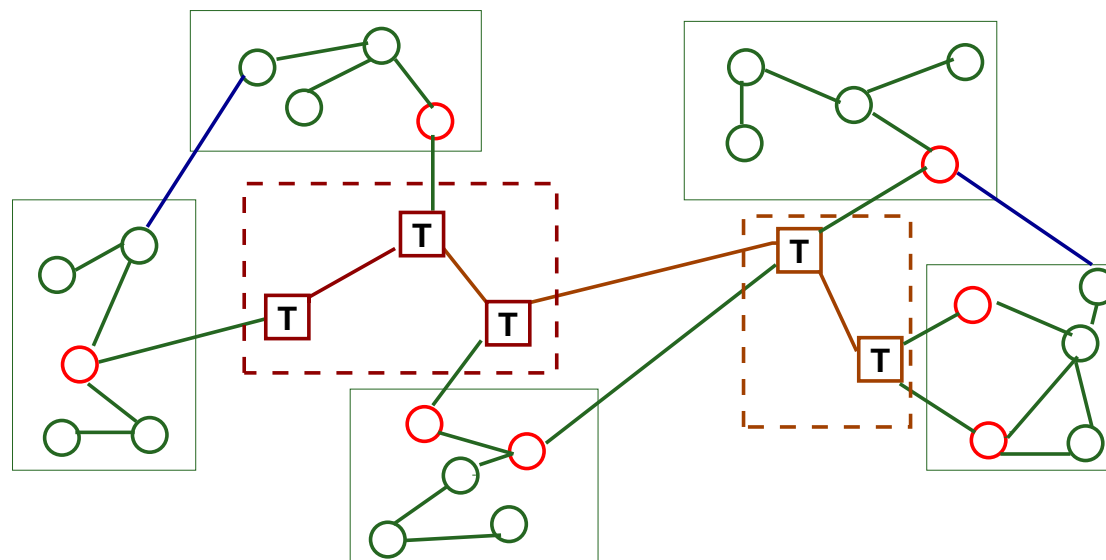# Caching and the Web

- Traditional web : short request-response transactions

- Ideal for some form of network caching

- Do gigabits help or *hurt ?*

- How to scale caches in a large and fast network?

backbone-only
caches

distributed caches

cache size
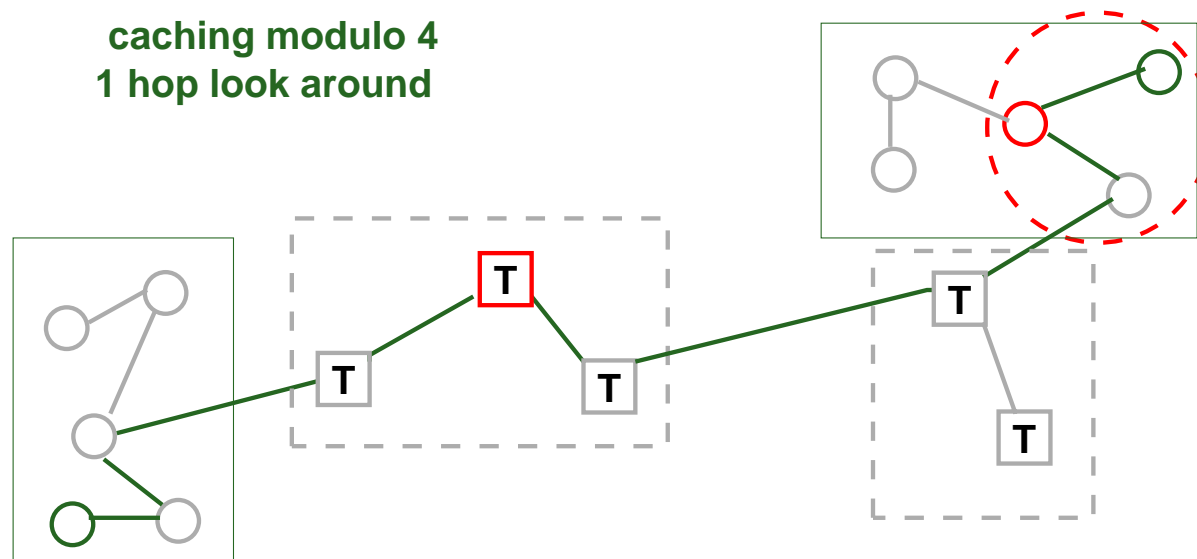
aggregate bandwidth
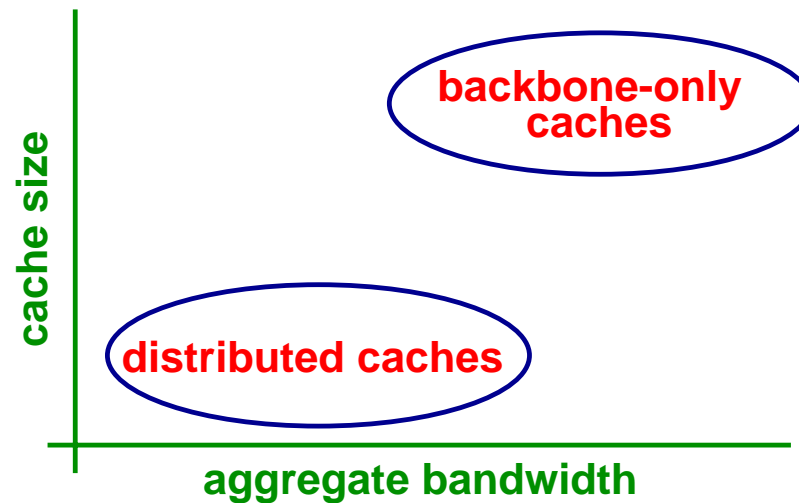
Approaches to Caching

- Caching by location
  - At *backbone* nodes
  - At stub domain borders

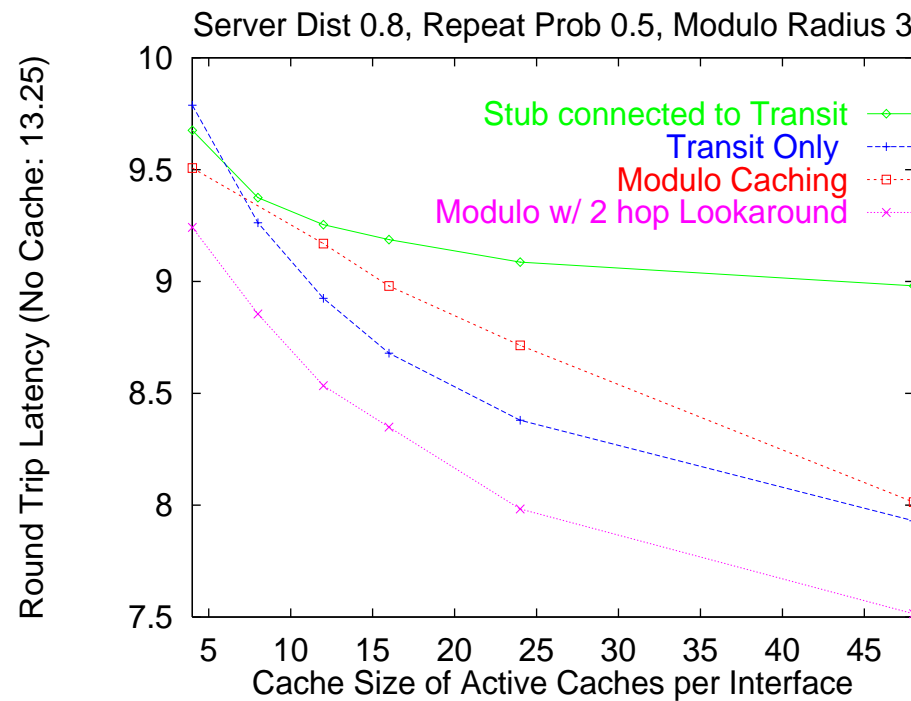- Effective use of widely distributed caches is difficult

Caching in an Active Network

- Cache radius — modulo caching

- Trade cache memory for location information
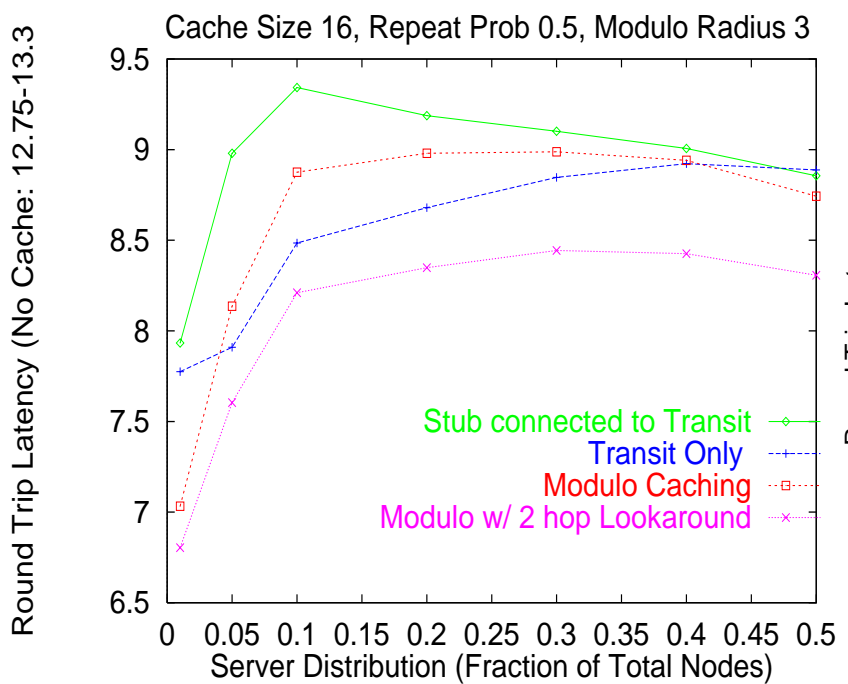
- Co-ordination between active caches — look-around

**caching modulo 4
1 hop look around**

AN-Sim Environment, Results

- Total cache size constant across methods
- Total cache size increases along x-axis
- Each transit-only cache is *25 times* larger than each active cache

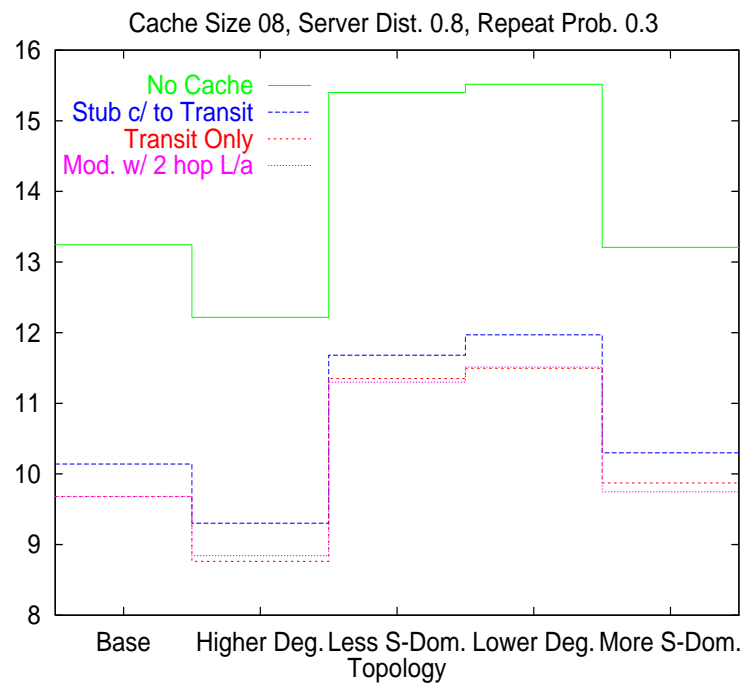Server Dist 0.8, Repeat Prob 0.5, Modulo Radius 3

Legend:
- Stub connected to Transit
- Transit Only
- Modulo Caching
- Modulo w/ 2 hop Lookaround

Y-axis: Round Trip Latency (No Cache: 13.25)

X-axis: Cache Size of Active Caches per Interface

More Results

- Effect of server distribution and topology

Variation in Server Distribution

Variation in Topology

How do Gigabit Networks Change the Web?

Q.1 What is the Web?

HTTP or Other Protocols?

A.0 *Reduce congestion?*

The WEB will expand to fill all available bandwidth

A.1 Not much. Round trip times remain unchanged.

How does the Web Change Gigabit Networks?

A.0 May depend on the type of gigabit network

- Connection-less network (e.g. IP)

- Connection-oriented network (e.g. ATM)

A.1 Makes the network do more connection-setup/teardown processing than other (bulk) applications

- Reduces them to *Megabit* networks