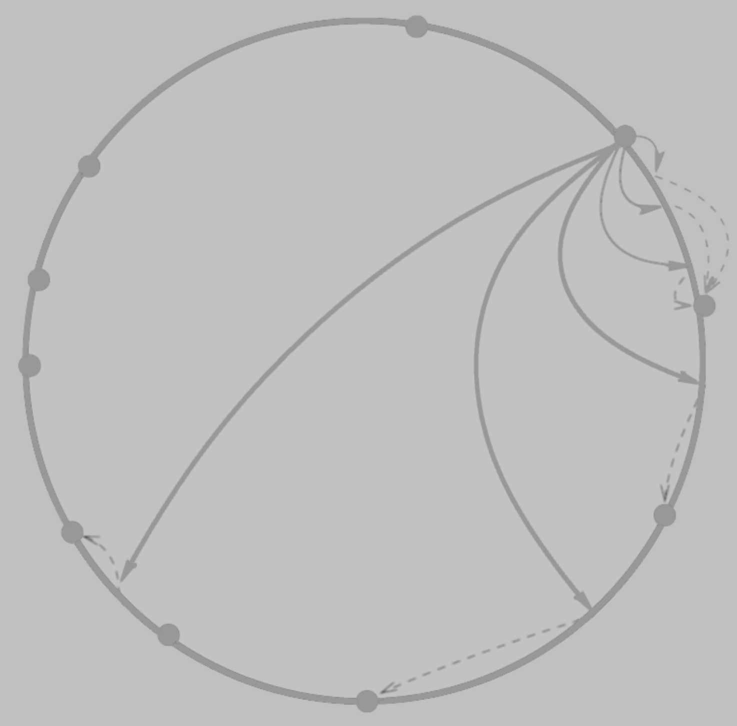
Examining The Tradeoffs Of Structured Overlays In A Dynamic Non-transitive Network

Steve Gerding Jeremy Stribling
{sgerding, strib}@csail.mit.edu



Motivation

- P2P overlays are a hot topic in networking research
- However, overlay performance research is still young
- Relatively unexplored areas:
    - Comparing several overlays in a fair setting, with a unified metric
    - Examining their behavior under real, pathological conditions
    - Determining how parameter tuning affects performance
- Important for system designers and wide area deployment

Our Goal

- Compare the performance of several structured P2P overlays under real world network conditions
- Explore the effects of parameter tuning for individual overlays

Accomplished by:

- Gathering and analyzing data about real world network conditions
- Using this data to compare the overlays in simulation
- Analyzing the simulation results and drawing conclusions

Presentation Overview

- Related work
- Real world dataset: **PlanetLab**
- Overlays in brief: **Chord**, **Tapestry**, **Kademlia**, **Kelips**
- Experimental methodology
- Results
- Discussion
- Future work

Related Work

- Gummadi et. al.: Effect of routing geometry on resilience, proximity
  The impact of DHT routing geometry on resilience and proximity, SIGCOMM 2003

- Rhea et. al.: App-level bmarks to encourage quality implementations
  Structured peer-to-peer overlays need application-driven benchmarks, IPTPS 2003

- Liben-Nowell et. al.: Chord stabilization traffic, with churn
  Analysis of the evolution of peer-to-peer systems, PODC 2002

- Xu: Routing state vs. network diameter: log(n) asymptotically optimal
  On the fundamental tradeoffs between routing table size and network diameter, Infocom 2003

- Countless structured and unstructured P2P overlays

The PlanetLab Dataset

- Topology data obtained from the PlanetLab federated testbed
- Extracted from PlanetLab All-Pairs-Pings data (http://pdos.lcs.mit.edu/~strib/pl_app)

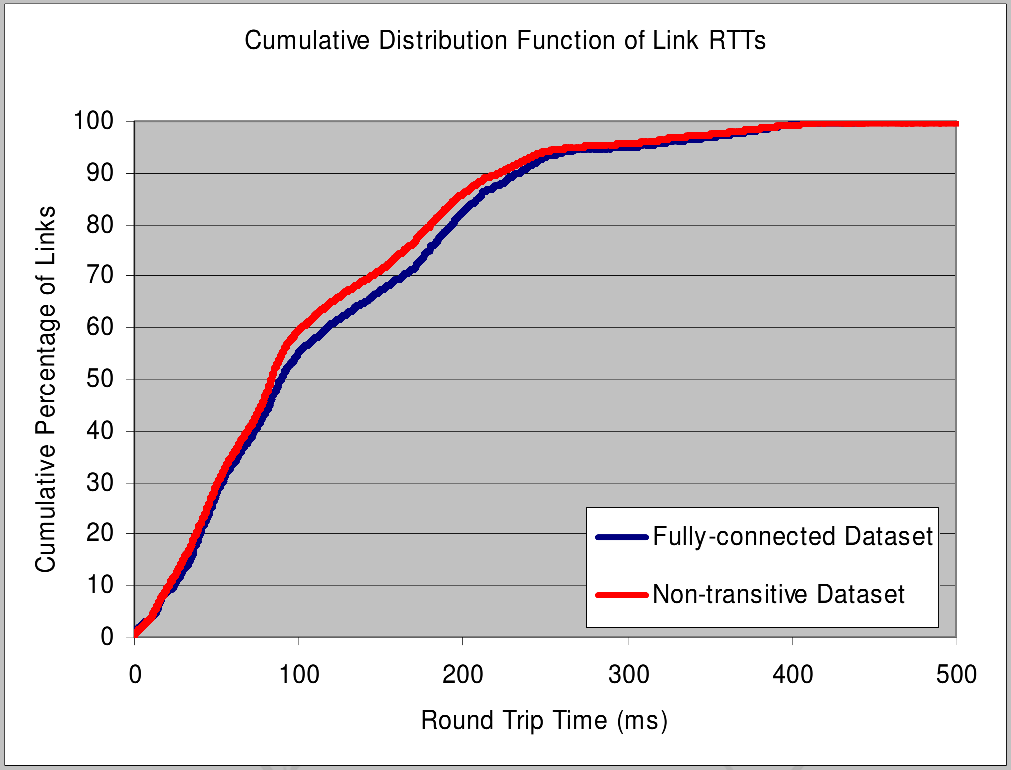
- Why is this interesting?
    - Global-scale testbed
    - Non-transitive links
    - Time-varying latency data
    - Real-world rates of churn (node failure and recovery)

The PlanetLab Dataset

Observed properties of the PlanetLab testbed:

- Size of datasets
  Fully-connected: 159
  Non-transitive: 248
- Non-transitivity
  9.9% of combinations are non-transitive
- Mean round trip time
  Fully-connected : 117.39 ms
  Non-transitive: 118.46 ms
- Churn rate
  MTTF: 321.1 hours
  MTTR: 2.7 hours

(Blind submission, SIGMETRICS 2004)



Overlays

Chord Tapestry Kademlia Kelips

Overlays

Chord Tapestry Kademlia Kelips

Properties of Chord (*Stoica et. al., SIGCOMM 2001*):

- Ring/Skiplist geometry
- Separates correctness (successors) and performance (finger table)
- **log(n)** state, **log(n)** hops

Parameters Explored:

# successors	4 – 32
Finger base	2 – 128
Finger stabilization	2 – 32 min
Succlist stabilization	1 – 32 min
Recursive routing	Yes / No

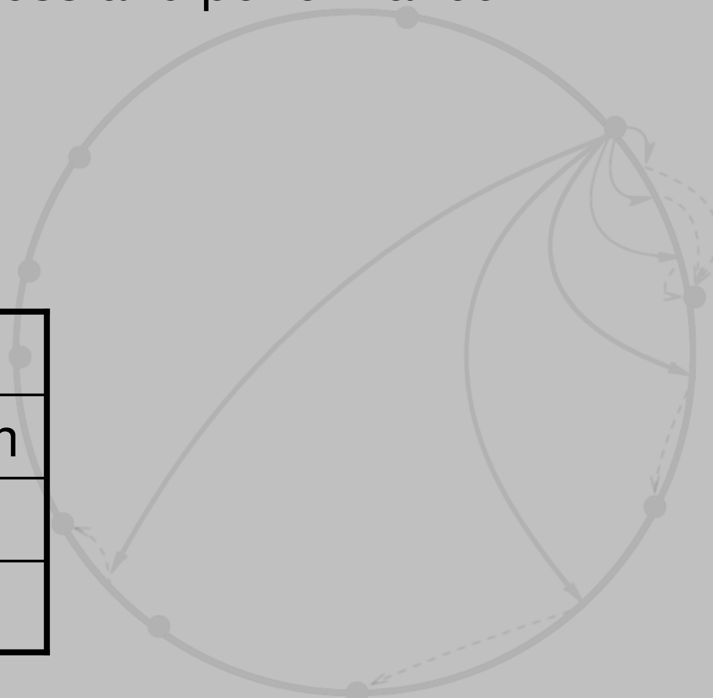
Overlays

Properties of Tapestry *(Zhao et. al., UC Berkeley TR 2001):*

- Tree-like geometry
- Rtg. table used for both correctness and performance
- Recursive routing
- **log(n)** state, **log(n)** hops

Parameters Explored:

ID Base	2 - 128
Stabilization	2 – 32 min
Backups per entry	1 – 4
Backups used in lookups	1 – 4



Overlays

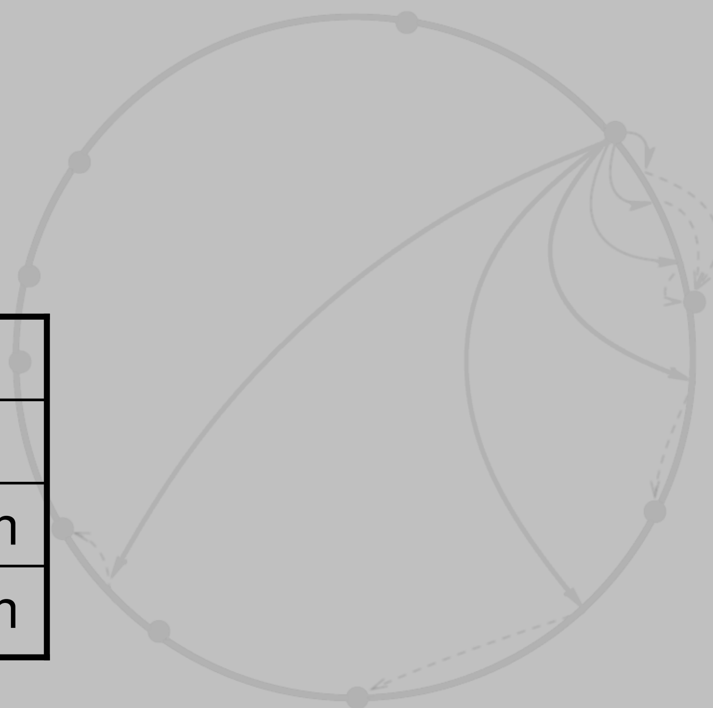
Chord   Tapestry   **Kademlia**   Kelips

Properties of Kademlia *(Maymounkov & Mazières, IPTPS 2002)*:

- **XOR** routing metric
- Lookups refresh routing state
- Iterative routing
- **log(n)** state, **log(n)** hops

Parameters Explored:

k (bucket size)	8 – 32
α (parallel lookups)	1 – 5
Stabilization timer	2 – 32 min
Refresh rate	2 – 32 min



Overlays

Properties of Kelips *(Gupta et. al., IPTPS 2003)*:

- Nodes hashed into $n^{1/2}$ groups
- Keep contacts in each other group
- Use p2p gossip state maintenance
- **$O(n^{1/2})$** state, **2** hops

(Some of the) Parameters Explored:

Gossip interval	.125 – 24 min
Contacts per group	2 – 8
New item gossip count	0 - 4
Routing entry timeout	5 – 40 min

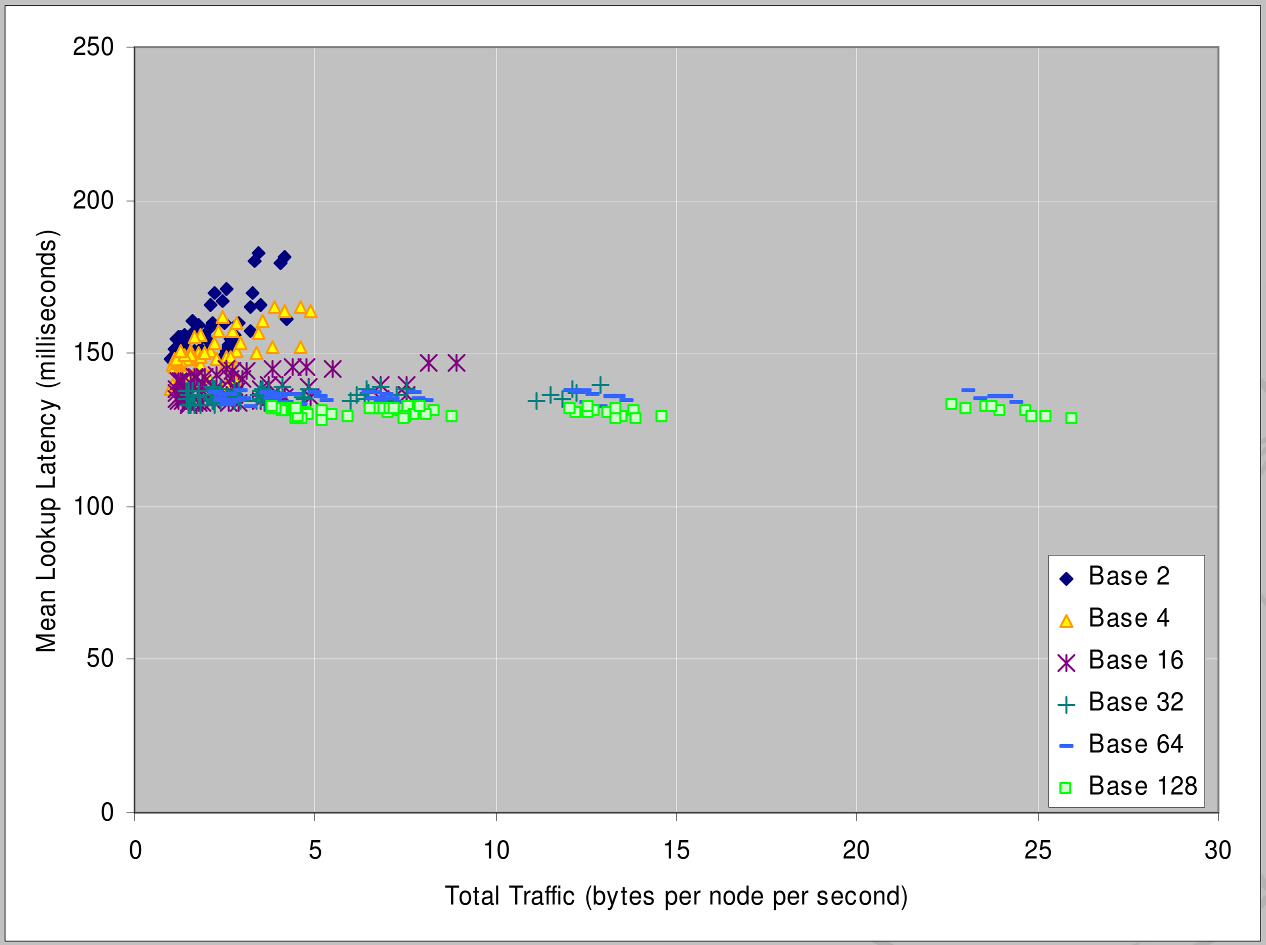
Experimental Methodology

- **p2psim**, a discrete event simulator *(http://pdos.lcs.mit.edu/p2psim)*
  - Simulates network delay

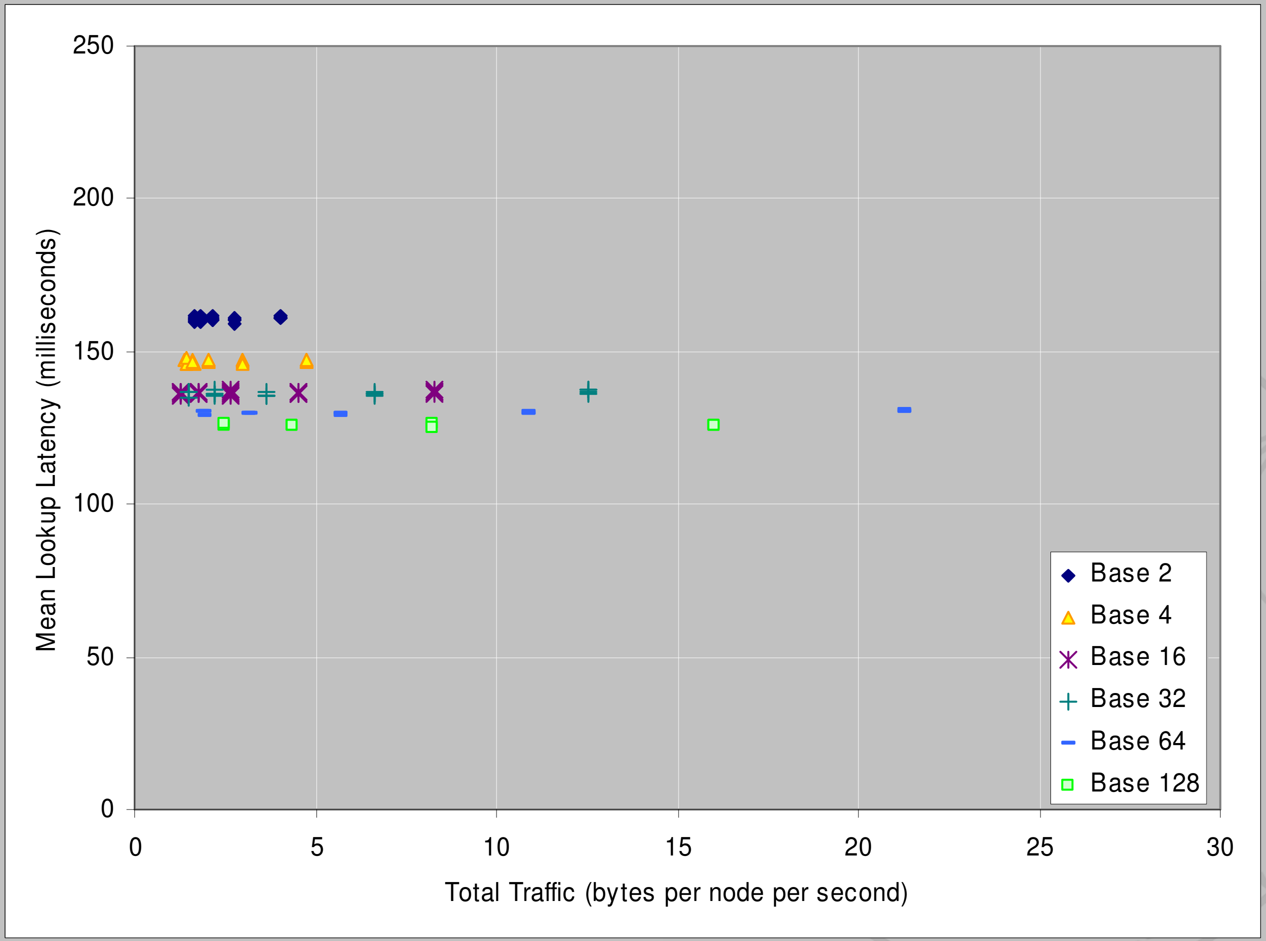
- Nodes generate lookups for random keys every 116 seconds
  - As observed by Saroiu et. al. for Kazaa traffic
    An analysis of content delivery systems, OSDI 2002

- Observed tradeoff between bandwidth and latency
  - Background maintenance traffic
  - Timeouts incurred during lookups

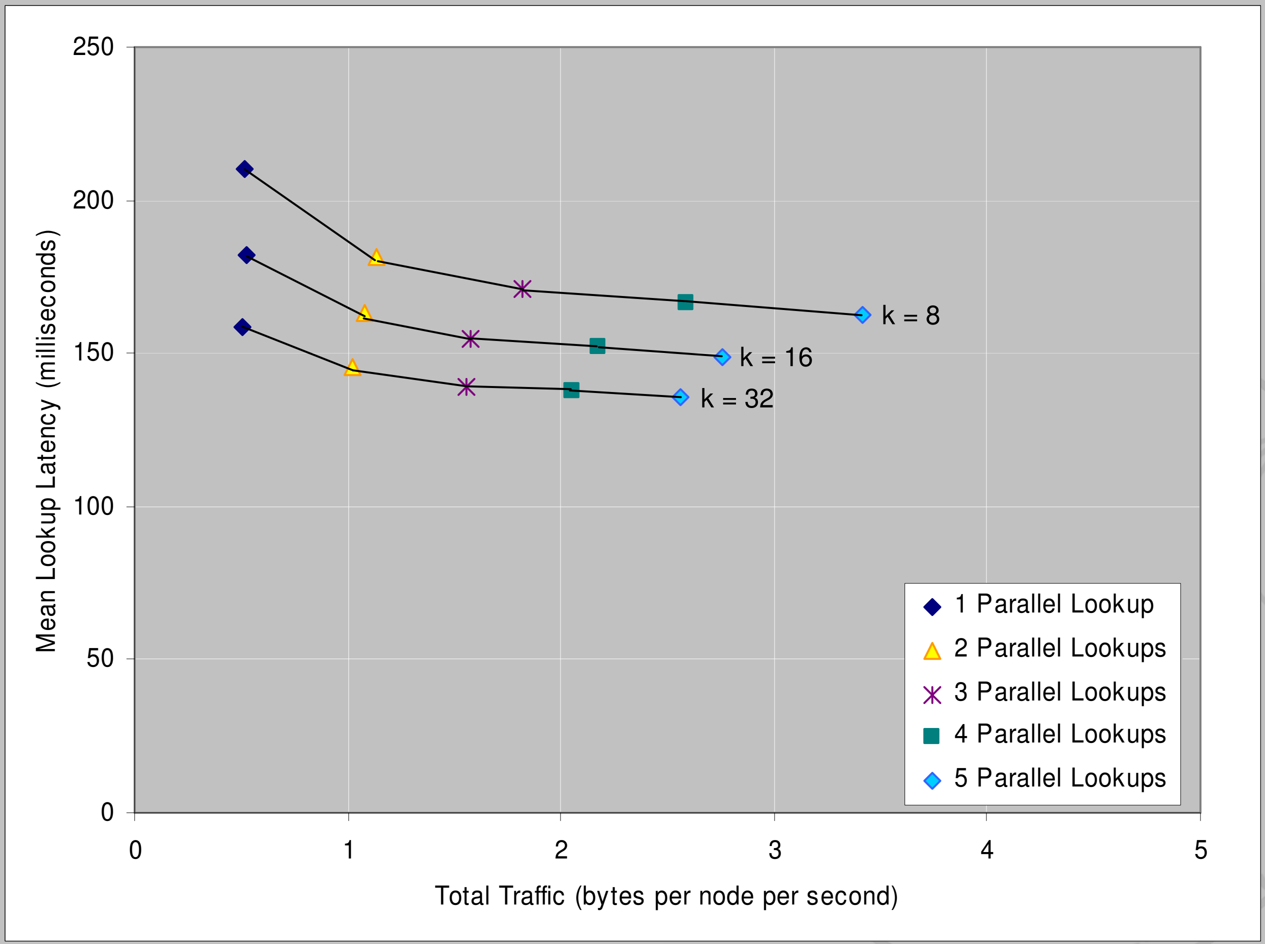
Baseline Results – Chord (Recursive)



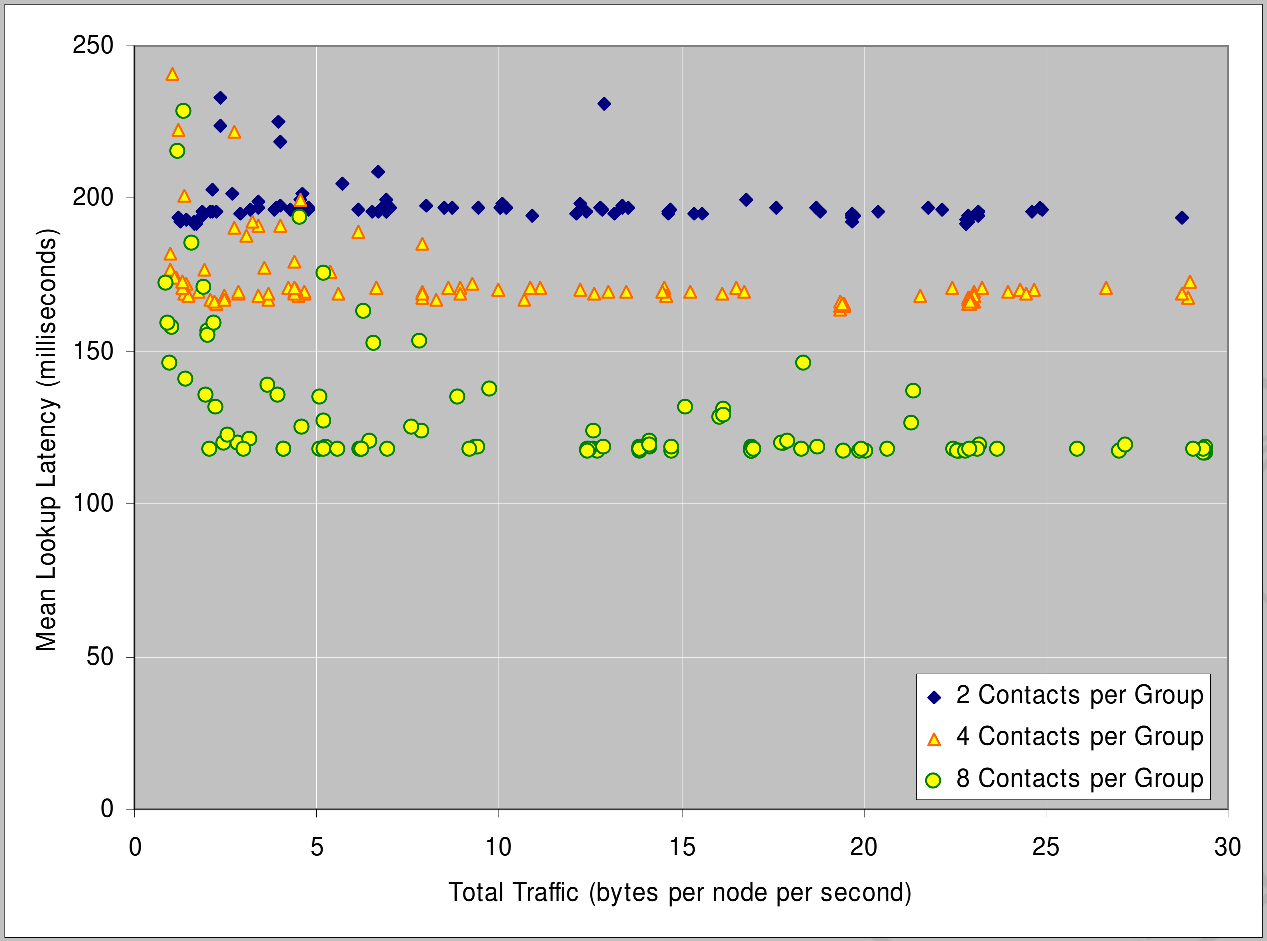
Baseline Results - Tapestry



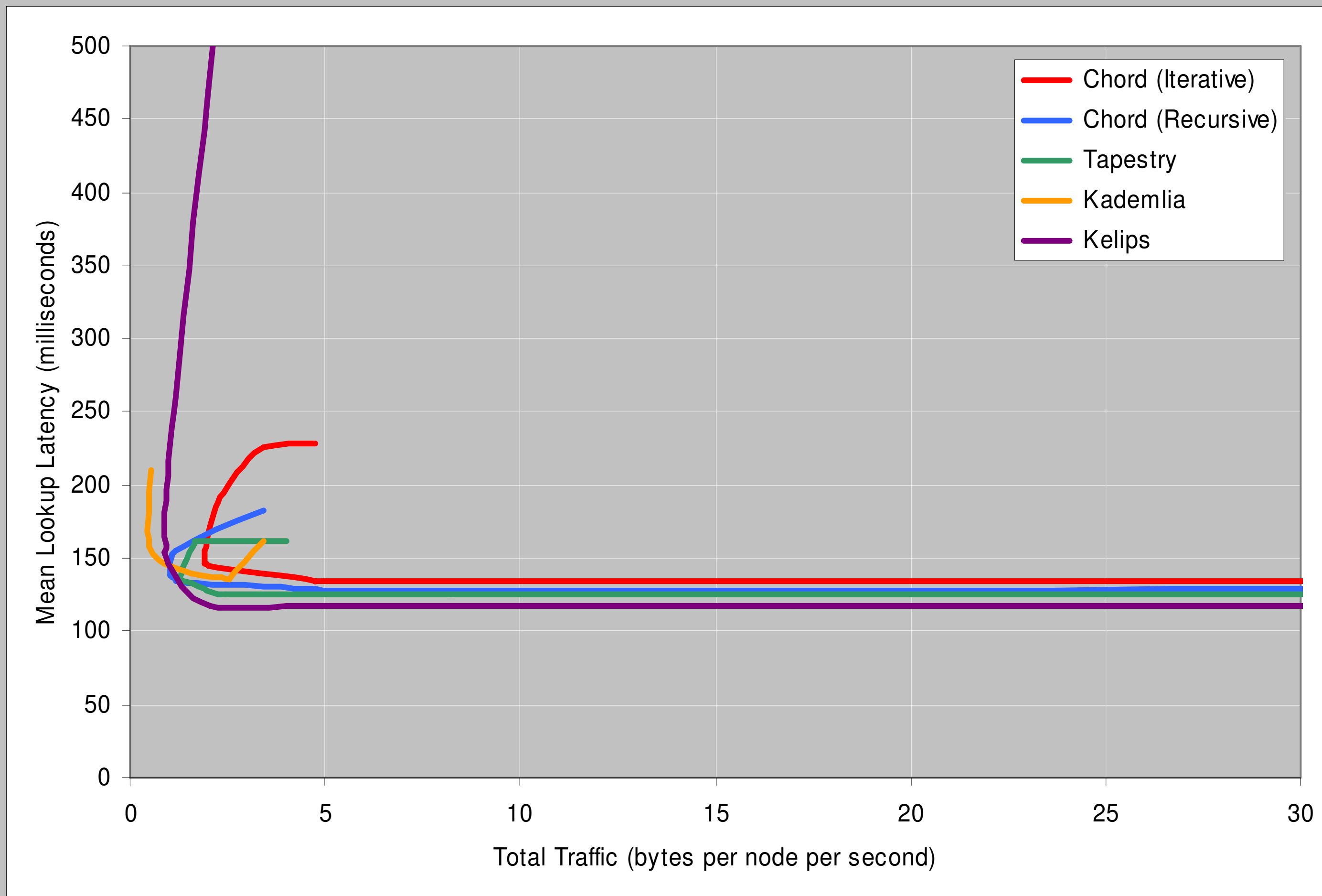
Baseline Results - Kademlia



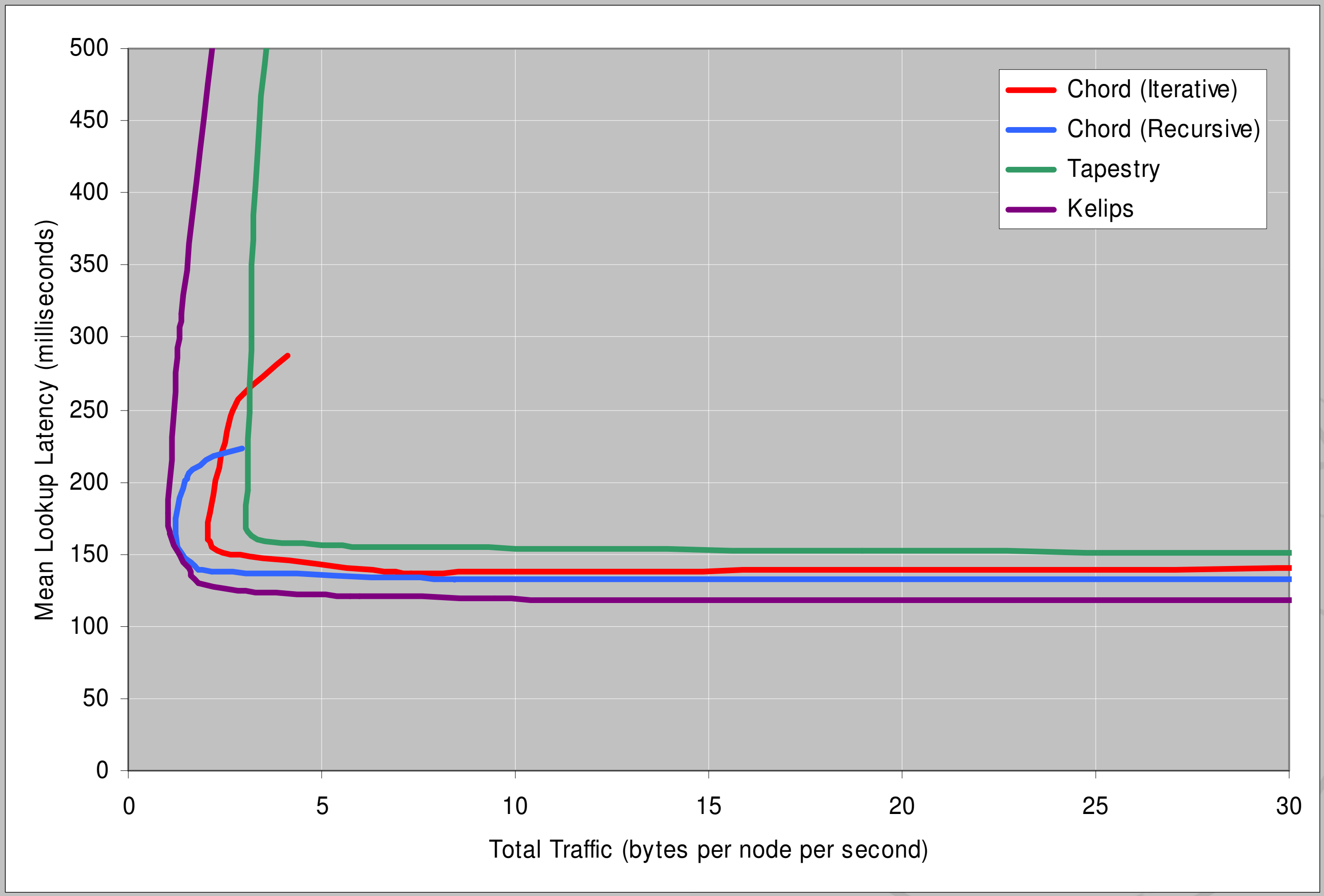
Baseline Results - Kelips



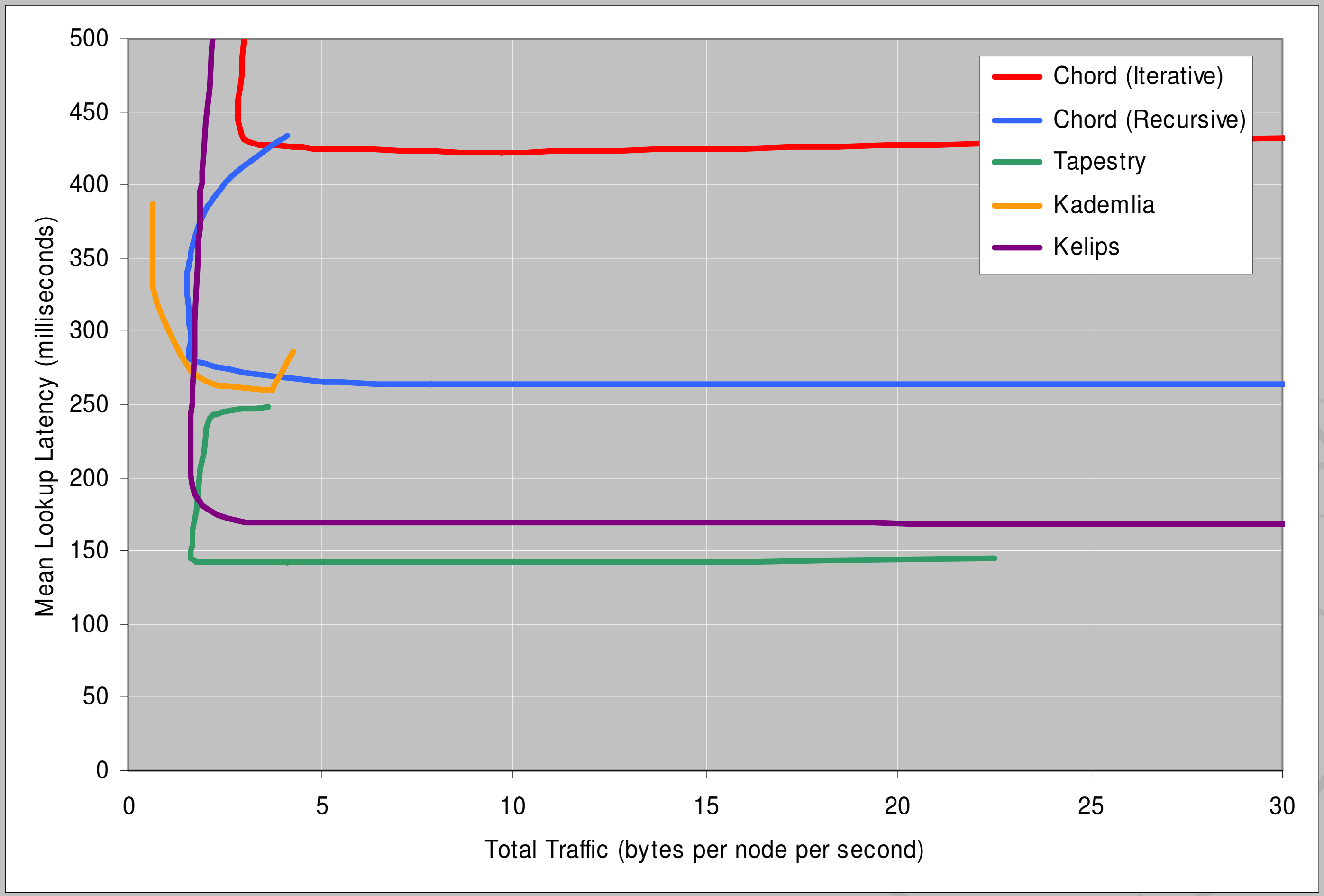
Baseline Results - All



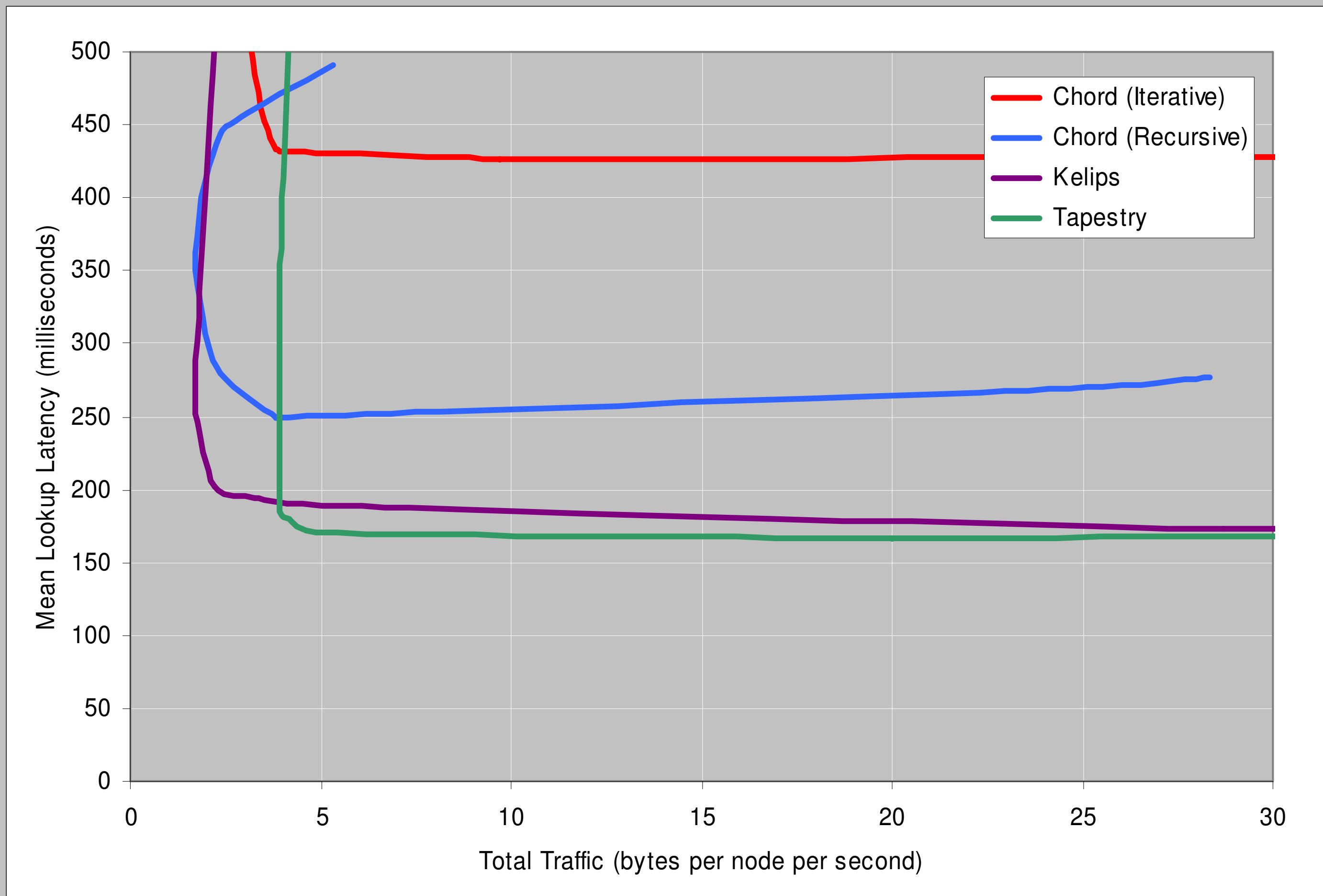
Churn Results - All



Non-transitive Results - All



Non-transitive + Churn Results - All



Discussion

- Performance of a particular protocol can vary widely
  - Careful tuning of parameters greatly improves performance
- Low rate of churn on PlanetLab has little effect on most protocols
  - Optimal configuration:
    - Large number of neighbors (base)
    - Low maintenance traffic (stabilization)
- Non-transitivity has a greater effect
  - Recursive routing a big win
  - Strictness of Chord hinders its performance

Future Work

- By next Friday
  - Analysis of overlays in the presence of variable-latency links
  - Data for Kademlia in churn scenario

- Future research topics
  - More overlays (Koorde, one-hop, etc.)
  - Effects of link failures
  - Effects of asymmetric links
  - Scaling simulation up to thousands of nodes
  - Adaptive, self-tuning parameters

Summary

- Our goal: Explore the effects of real world conditions and parameter tuning on the performance of structured overlays
- Real world data was collected from the PlanetLab testbed
- Illustrated tradeoffs within and between four overlay protocols
- Non-transitivity has a large effect on performance
- Recommendations for system designers:
  - Choose an appropriate overlay for target environment
  - Carefully tune parameters for that overlay

Why Non-transitivity Breaks Chord