Optimizations for Locality-Aware Structured Peer-to-Peer Overlays

Jeremy Stribling
strib@mit.edu

Collaborators:
Kris Hildrum
John D. Kubiatowicz

The First IRIS Student Workshop
August 10, 2003
Object Location in Tapestry
Is This Always Optimal?
Discussion

• **Why is this a problem?**
  – Latency, efficiency, availability

• **Metric**: Relative Delay Penalty (RDP)
  – Distance through Tapestry vs. IP distance

• **Solution**: trade storage for low local area RDP
  – Will work in DOLRs with a pointer indirection layer
Optimization 1: Backups

- **Redundancy**: Store up to $c$ nodes in each entry
  - $c-1$ nodes are backups
- **A simple optimization**: publish to $b$ backups
  - Limit to first $h$ hops of publish path
- **Result**
  - Nodes near the object more likely to encounter pointers
  - Cost: $b^*h$ additional pointers per object
Optimization 1: Backups

Experiments run in simulation on a GT-ITM transit stub topology

90th percentile RDP

Client to object round-trip (ping) time (10 ms buckets)

Unoptimized
Opt(1 back, 1 hop) [+1]
Opt(2 back, 2 hops) [+4]
Opt(3 back, 3 hops) [+9]
RDP = 1
Optimization 2: Nearest Neighbors

• **Observation:** In Opt. 1, choice for backups is limited
  – But lots of nodes at each level, many may be nearby

• **Optimization:** publish to $n$ nearest neighbors
  – Limit to first $h$ hops of the publish path

• **Result**
  – If $n$ is large, essentially local area flooding
  – Analytical cost: $n^*h$ additional pointers per object
Optimization 2: Nearest Neighbors

Experiments run in simulation on a GT-ITM transit-stub topology

90th percentile RDP

Client to object round-trip (ping) time (10 ms buckets)

Unoptimized
Opt(1 near, 1 hop) [+1]
Opt(1 near, 3 hops) [+3]
Opt(5 near, 2 hops) [+10]
Opt(8 near, 3 hops) [+24]
RDP = 1
Optimization 3: Local Surrogate
Optimization 3: Local Surrogate

- **Solution**: Check local node before leaving
  - When publishing, place a pointer on *local surrogate*
  - Occurs naturally on Coral, LAND, SkipNet
  - In practice, storage cost is very low

- **Issue**: What determines a wide area hop?
  - One metric: if next hop is more than $t$ times longer than last hop, consider it wide area
Optimization 3: Local Surrogate

Experiments run in simulation on a GT-ITM transit-stub topology
Future Work

- Automatically adjust $t$ when using local surrogate
- Take measurements on actual networks
- Test optimizations with real workloads
- Evaluate the maintenance cost

Questions?