Incrementally Improving Lookup Latency in Distributed Hash Table Systems

Hui Zhang[1], Ashish Goel[2], Ramesh Govindan[1]

[1]**University of Southern California**

[2]**Stanford University**

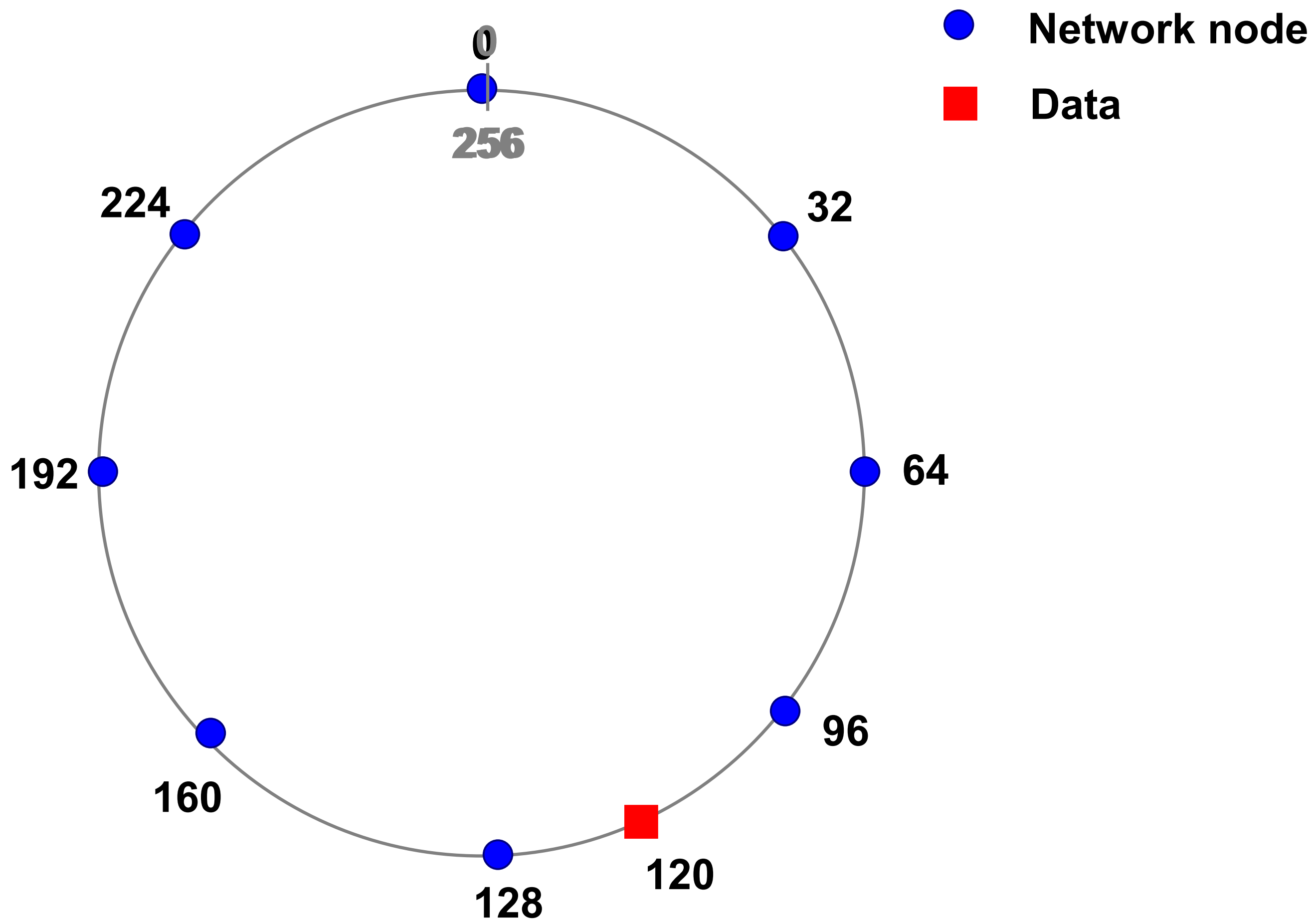
Outline

- Latency stretch problem in Distributed Hash Table (DHT) systems, with Chord as an example
- Two "latency stretch" theorems
- Lookup-Parasitic Random Sampling (LPRS)
- Simulation & Internet measurement results
- Conclusion & future work

DHT systems

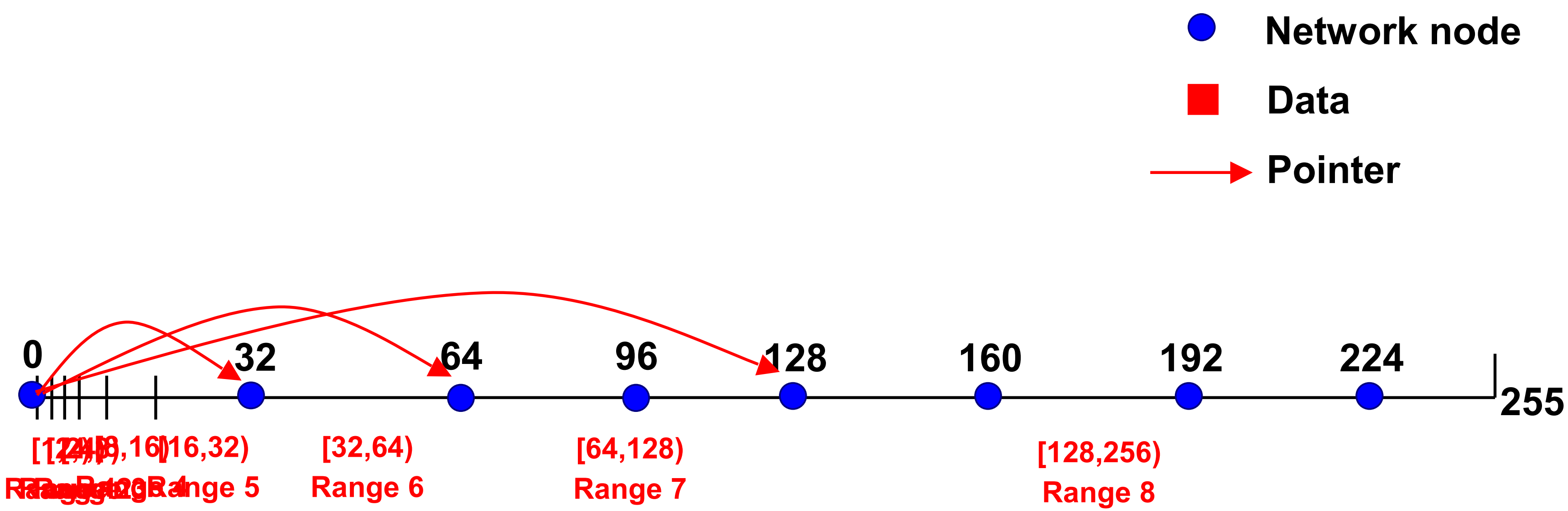
- A new class of peer-to-peer routing infrastructures
  - CAN, Chord, Pastry, Tapestry, etc.
- Support a hash table-like functionality on Internet-like scale
  - a global key space: each data item is a key in the space, and each node is responsible for a portion of the key space.
  - given a key, map it onto a node.
- Our research results apply to *frugal* DHT systems.
  - The search space for the key decreases by a constant factor after each lookup hop.
  - Examples: Chord, Pastry, Tapestry.

Chord – key space



A Chord network with 8 nodes and 8-bit key space

Chord – routing table setup

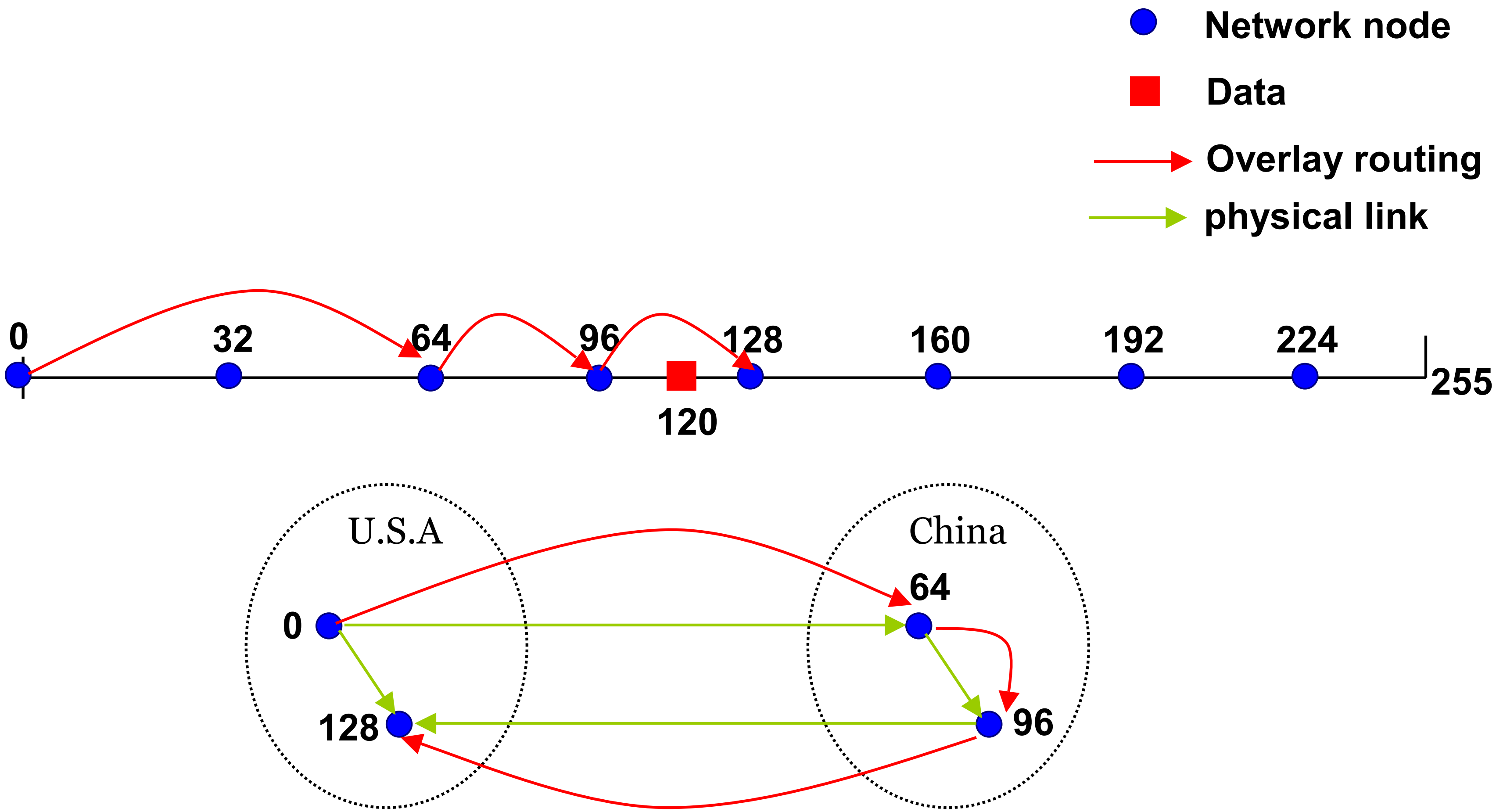


In node i's routing table:

One entry is created to point to to the first node in its jth ranges $[i+2^{j-1}, i+2^{j})$, $1 \leq j \leq m$.

A Chord network with N(=8) nodes and m(=8)-bit key space

Latency stretch in Chord



A Chord network with N(=8) nodes and m(=8)-bit key space

Latency stretch [Ratnasamy et al. 2001]

- $= \dfrac{\text{latency for each lookup on the overlay topology}}{\text{average latency on the underlying topology}}$

- In Chord, $\theta(\log N)$ hops per lookup in average
  - ❑ $\theta(\log N)$ stretch in original Chord.
  - ❑ Could Chord do better, e.g., O(1) stretch, without much change?

Our contributions

- Theory
  - ❑ Latency expansion characteristic of the underlying network topology decides latency optimization in frugal DHT systems.
    - Exponential latency expansion: bad news.
    - Power-law latency expansion: good news.
- System
  - ❑ Lookup-Parasitic Random Sample (LPRS), an incremental latency optimization technique.
    - Achieve O(1) stretch under power-law latency topologies.
- Internet measurement.
  - ❑ The Internet router-level topology resembles power-law latency expansion.

Latency expansion

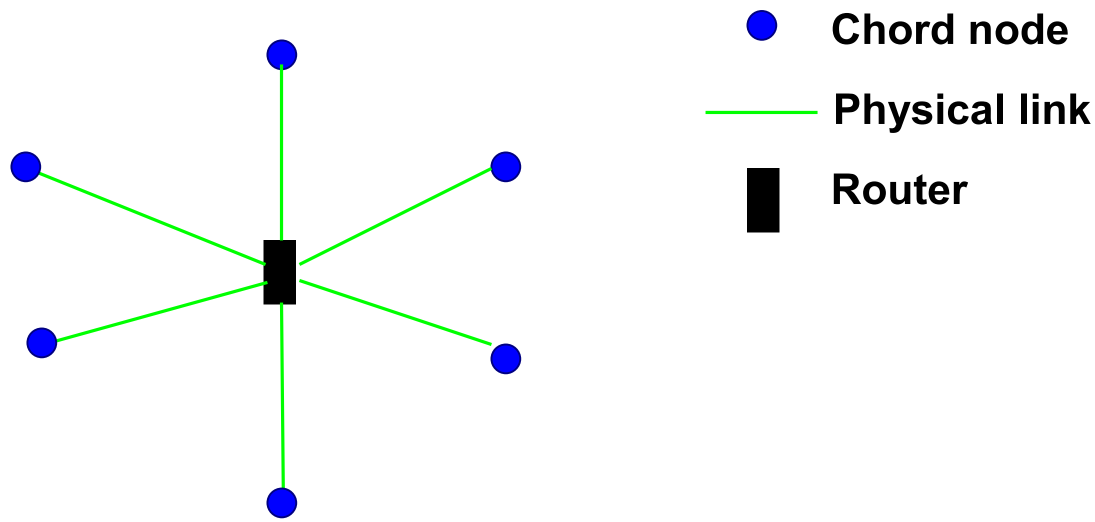
- Let $N_u(x)$ denote the number of nodes in the network G that are within latency x of node u.

  - power-law latency expansion: $N_u(x)$ grows (i.e. ``expands'') proportionally to x^d, for all nodes u.
    - Examples: ring (d=1), mesh (d=2).
  - exponential latency expansion: $N_u(x)$ grows proportionally to α^x for some constant $\alpha > 1$.
    - Examples: random graphs.

“Latency-stretch” theorem - I

- “Bad news” Theorem

  If the underlying topology G is drawn from a family of graphs with exponential latency expansion, then the expected latency of Chord is $\Omega(L \bullet \log N)$, where L is the expected latency between pairs of nodes in G.



The worse-case scenario: equal-distance

"Latency-stretch" theorem - II

- "Good news" Theorem

  If

  (1) the underlying topology G is drawn from a family of graphs with d-power-law latency expansion, and

  (2) for each node u in the Chord network, it samples *$(\log N)^d$* nodes in each range with uniform randomness and keeps the pointer to the nearest node for future routing,

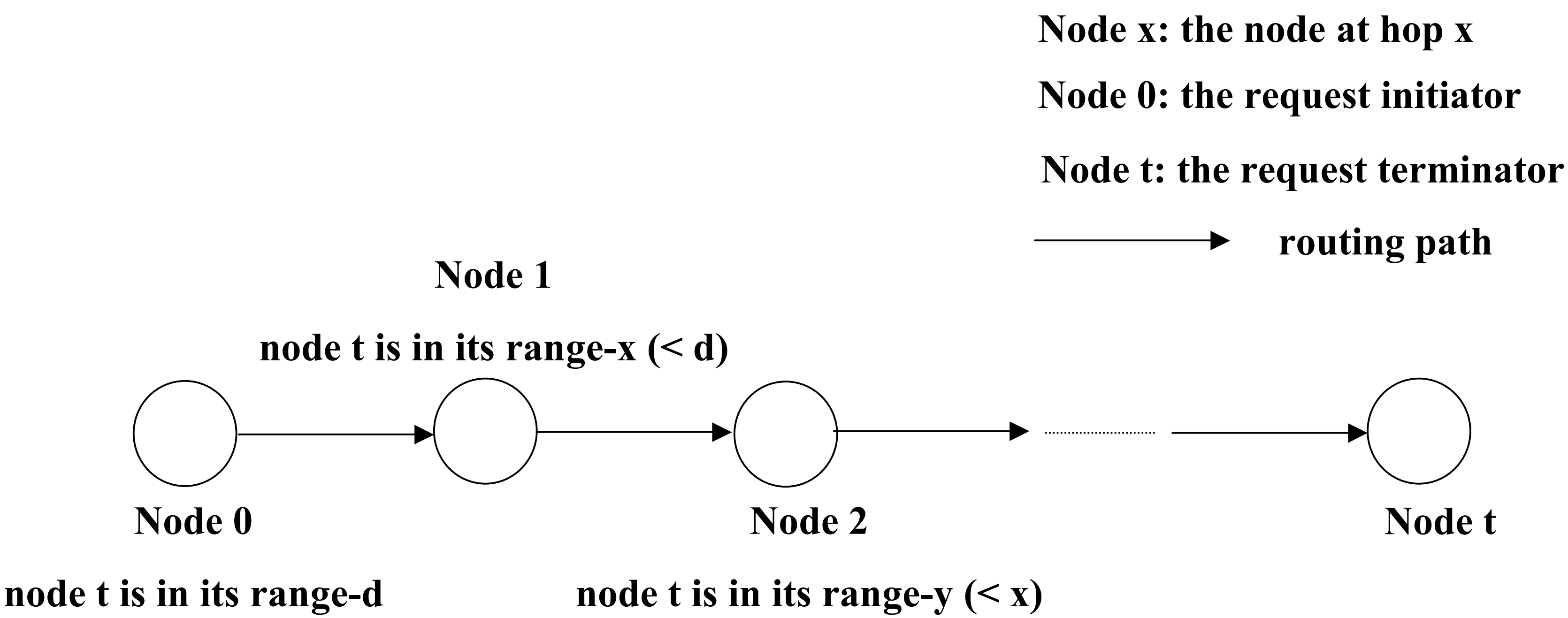
  then the expected latency of a request is *$O(L)$*, where L is the expected latency between pairs of nodes in G.

Two remaining questions

- How does each node efficiently achieve $(\log N)^d$ samples from each range?

- Do real networks have power-law latency expansion characteristic?

Uniform sampling in terms of ranges

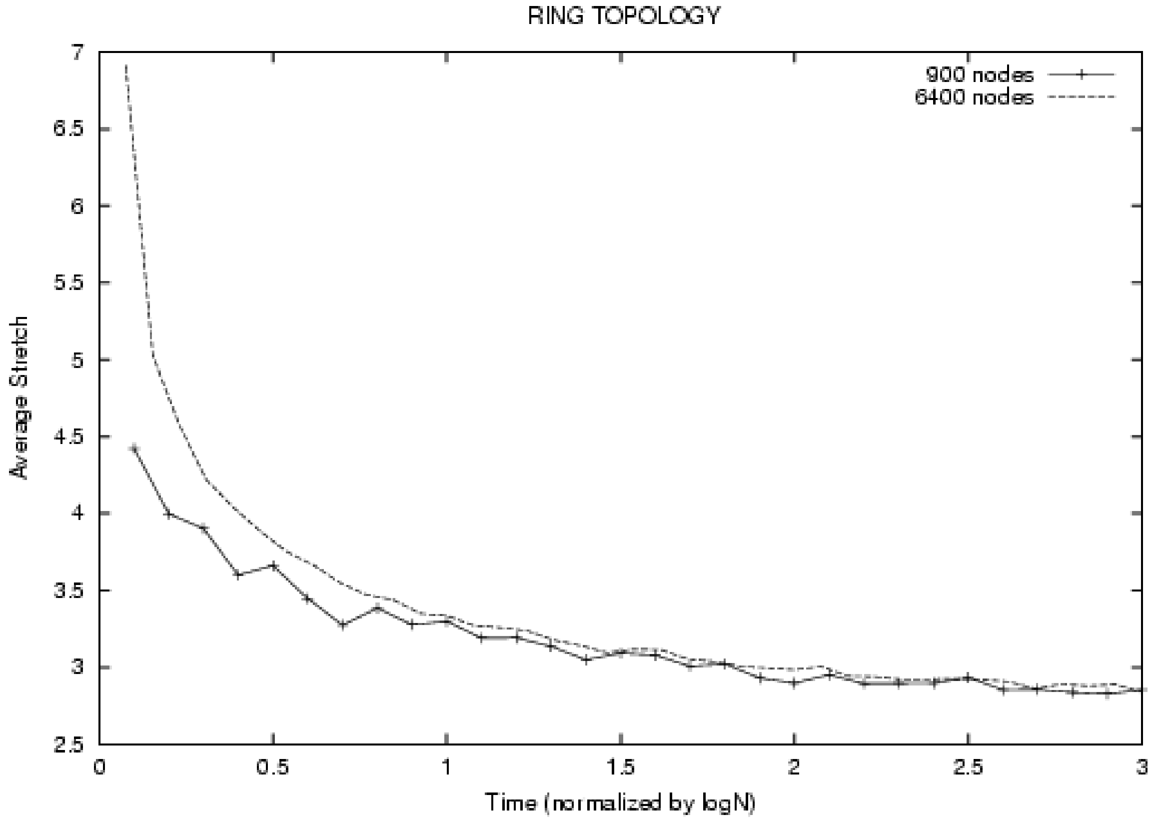


For a routing request with $\theta(log\ N)$ hops, final node t will be a random node in $\theta(log\ N)$ different ranges

Lookup-Parasitic Random Sampling

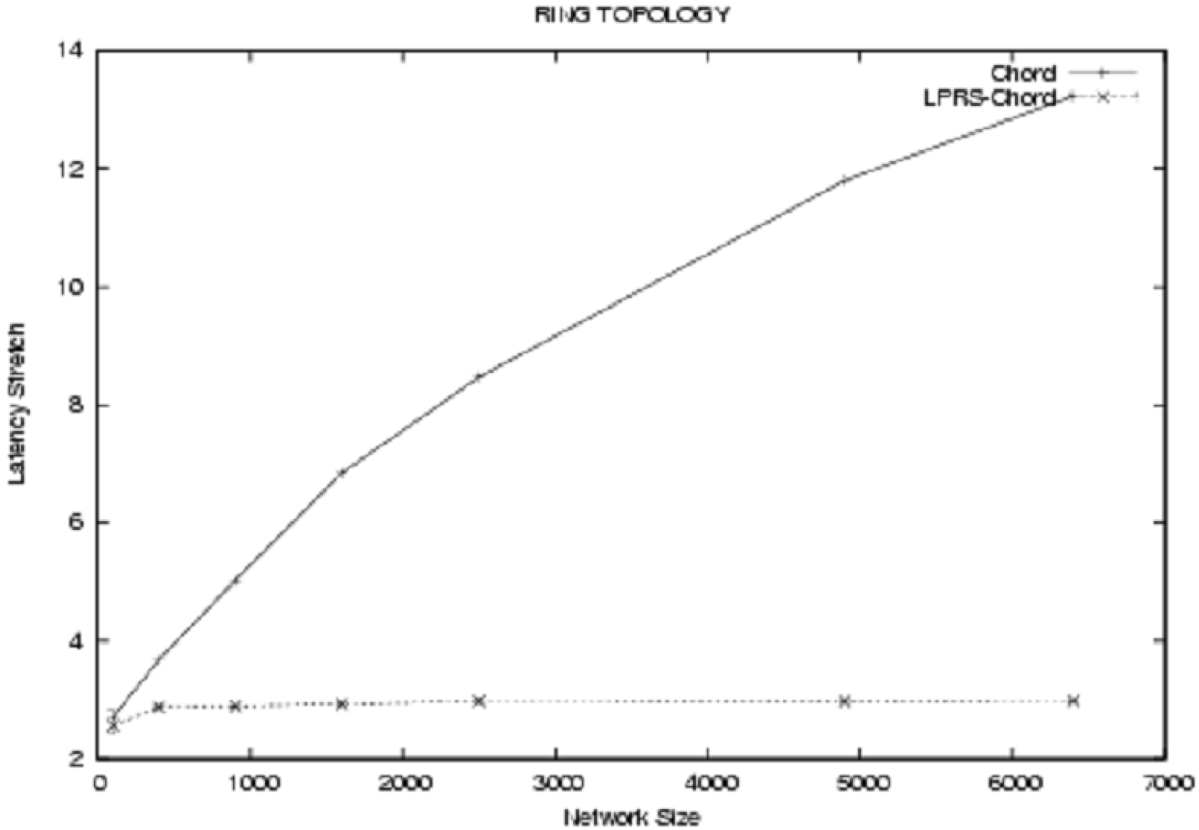
1. Recursive lookup.
2. Each intermediate hop appends its IP address to the lookup message.
3. When the lookup reaches its target, the target informs each listed hop of its identity.
4. Each intermediate hop then sends one (or a small number) of pings to get a reasonable estimate of the latency to the target, and update its routing table accordingly.

LPRS-Chord: convergence time



Convergence Time

LPRS-Chord: topology with power-law expansion



Ring Stretch
(at time 2logN)

What's the latency expansion characteristic of Internet?

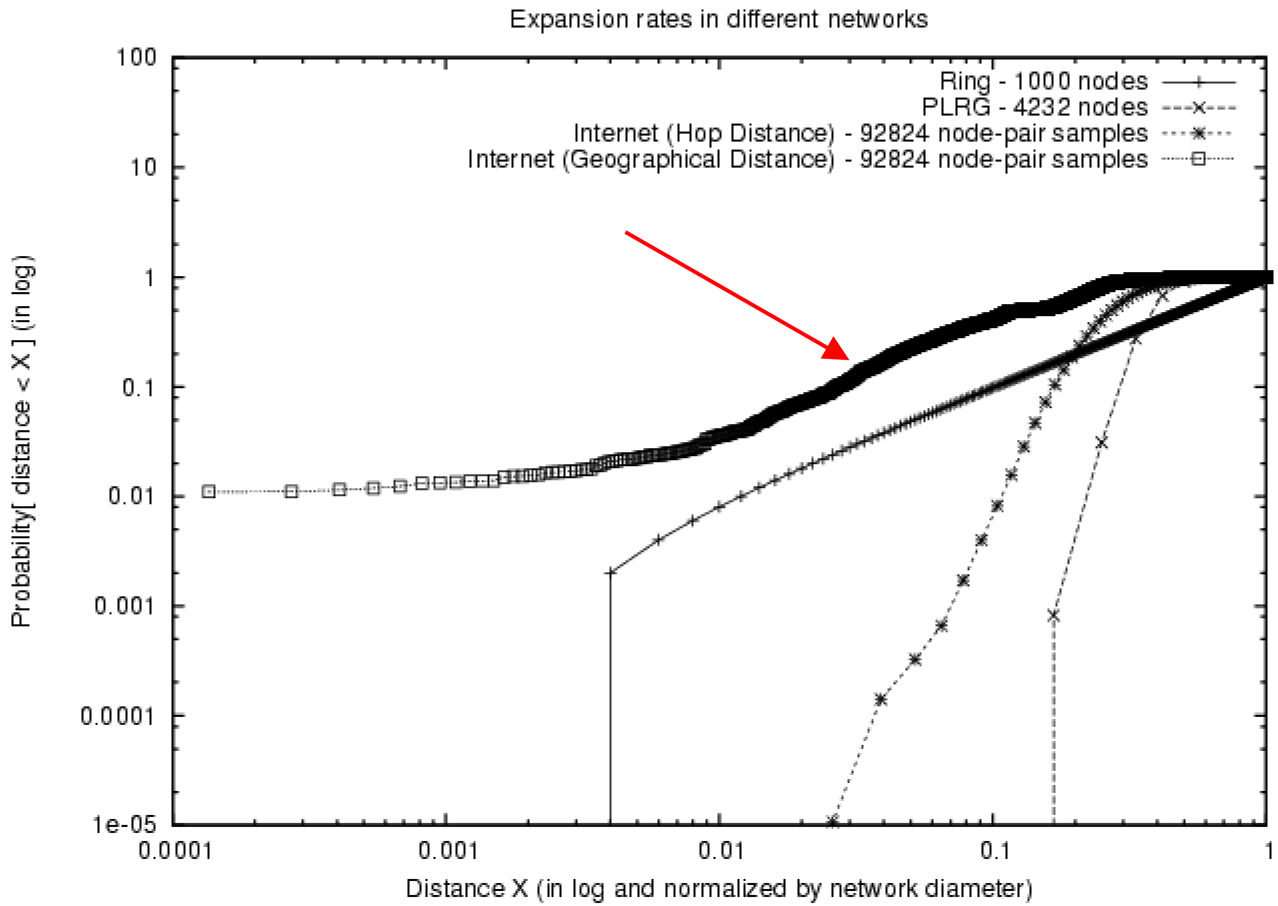
on the end-user level

on the router level

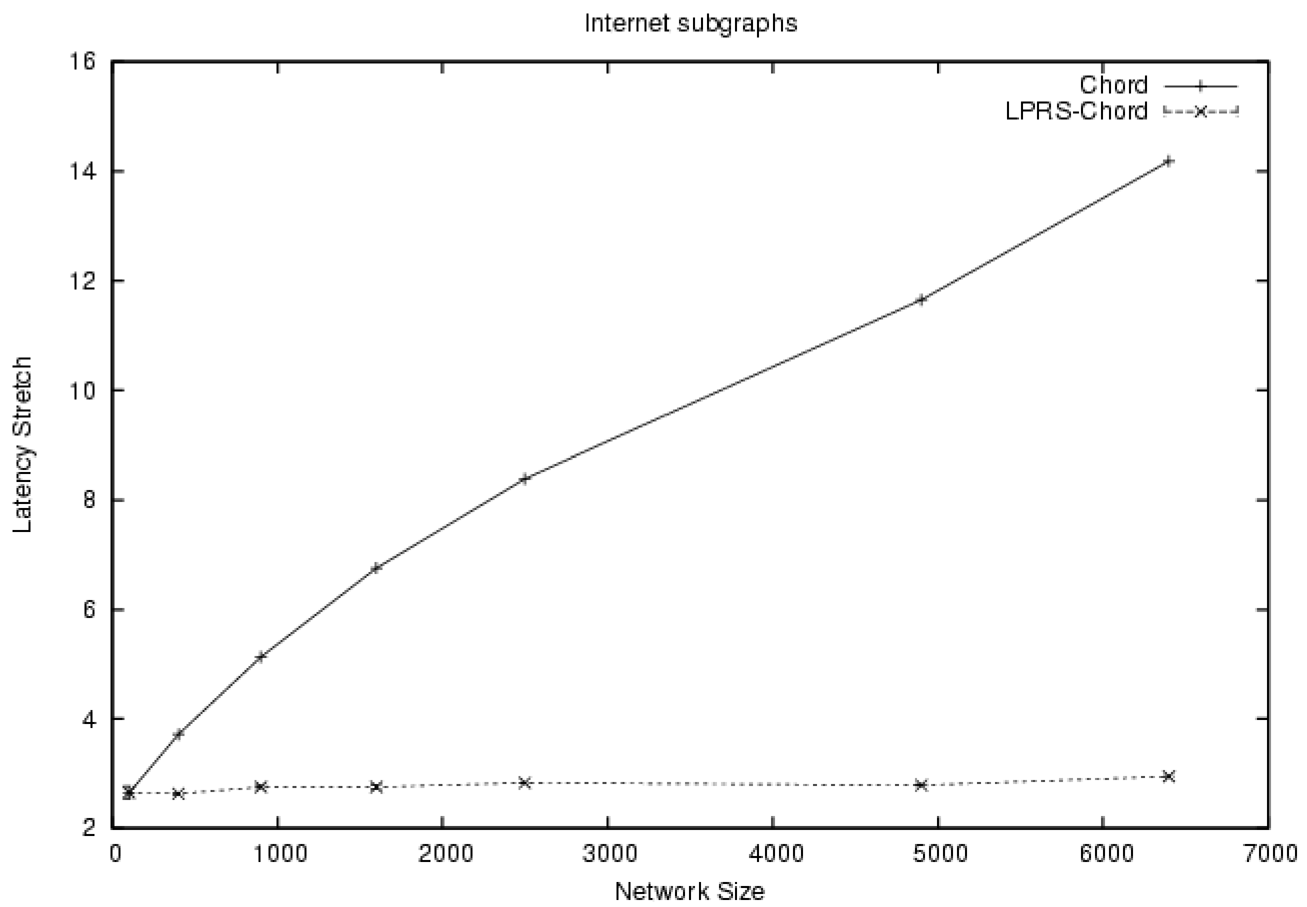
Internet router-level topology: latency measurement

- Approximate link latency by geographical latency
  - assign geo locations to nodes using Geotrack[Padmanabhan2001].
- A large router-level topology dataset
  - 320,735 nodes, mapped to 603 distinct cities all over the world.
  - 92,824 node pairs are sampled to tractably compute the latency expansion of this large topology.

Internet router-level topology: latency expansion



LPRS-Chord on router-level topology



Stretch on the router-level subgraphs (at time 2logN)

Conclusion

- LPRS has significant practical applicability as a general latency reduction technique for frugal DHT systems.

- Future work

  - Studying the interaction of LPRS scheme with the dynamics of P2P systems.

Thank you!