On the Cost of Participating in a Peer-to-Peer Network

Nicolas Christin and John Chuang
University of California, Berkeley
School of Information Management and Systems
{christin,chuang}@sims.berkeley.edu

Problem statement

- Evaluate the amount of resources each peer contributes for being part of an overlay network
- Evaluate the benefits associated with participation
- Related work generally only considers connectivity, and is not concerned with maintenance or routing overhead
- We'll focus on *geometries*, i.e., set of nodes and edges (*topology*) associated with a routing algorithm (shortest path routing unless otherwise noted)

Motivation

- Allows us to predict potential disincentives to collaborate
- Allows us to identify hot spots
- Allows us to help design load balancing algorithms
- Possible benchmark to characterize efficiency of network as a whole
- Can be used to distinguish between proposals (e.g., DHTs)

Cost model

- A given node u requests an item, serves a request, or route requests between other nodes
- Latency cost

$$L_u = \sum_{v \in V} \sum_{k \in K_v} l_{u,k} t_{u,v} \Pr[Y =$$

Benefits enjoyed

- Service cost

$$S_u = \sum_{k \in K_u} s_{u,k} \Pr[Y = k]$$

- Routing cost

$$R_u = \sum_{v \in V} \sum_{w \in V} \sum_{k \in K_w} r_{u,k} \Pr[X = v$$

- Maintenance cost

$$M_u = m_u \deg(u)$$

Price paid

Analysis assumptions

- Homogeneous peers and homogeneous links (i.e., for any u and k, $l_{u,k} = l$, $s_{u,k} = s$, $r_{u,k} = r$ and $m_u = m$.)
- Steady-state regime
- Sources of requests uniformly distributed over the set of nodes (i.e., $\Pr[X = u] = 1/N$)
- Destinations of requests uniformly distributed over the set of nodes (implies $S_u = s/N$)
- Quite idealistic!

Social optimum

- Total cost for node u

$$C_u = L_u + S_u + R_u + M_u$$

- Total cost (for the whole network)

$$C = \sum_u C_u$$

$$m \leq l/N + r/N^2$$

- Social optimum: geometry that minimizes C
  - If number of nodes N is small and/or maintenance operations come cheap (i.e., m is small), social optimum is the **full mesh**
  - Otherwise, adding or removing links from a **star network** always **increases** C (local optimum)

Application to DHT geometries

- Closed form expressions can be derived

$$\rho_{u,D} = 1 + \quad N^{\frac{D-1}{D}}\left(-N^{\frac{1}{D}} + D\left(N^{\frac{1}{D}} - 1 + \left(\left\lfloor\frac{N^{\frac{1}{D}}}{2}\right\rfloor - 1\right)\left(\left\lceil\frac{N^{\frac{1}{D}}}{2}\right\rceil - 1\right)\right)\right)$$

4

- $R_u = r\frac{\rho_{u,D}}{N^2}$
- $M_u = 2mD$

R trees (as used in

try, Tapestry, …):

$$L_u = l\frac{D\Delta^{D-1}(\Delta - 1)}{N}$$

- $R_u = r\frac{\Delta^{D-1}(D(\Delta - 1) - \Delta) + 1}{N^2}$
- $M_u = mD(\Delta - 1)$

(same results for Chord rings with Δ=2)

$$L_{\min} = \frac{l}{N}\left(D\Delta^D + \frac{D}{\Delta - 1} - \frac{\Delta(\Delta^D - 1)}{(\Delta - 1)^2}\right)$$

$$L_{\max} = l\frac{D\Delta^{D+1} - (D+1)\Delta^D + 1}{N(\Delta - 1)}$$

- Different nodes have different latency costs

$$L_{\min} \le L_u \le L_{\max}$$

- Different nodes have different routing costs

$$0 \le R_u \le r\rho_{\max}/N^2$$

- Different nodes have different maintenance costs

$$M_u =$$

$$\rho_{\max} = \frac{(D-1)(\Delta^{D+2} - (\Delta - 1)^2) - D\Delta^{D+1} + \Delta^2}{(\Delta - 1)^2}$$

- Not really clear if it can be made symmetric

Asymmetry in de Bruijn graphs (cnt'd)

Δ: alphabet size
D: network diameter

$$L_{\max} = \max_u L_u$$

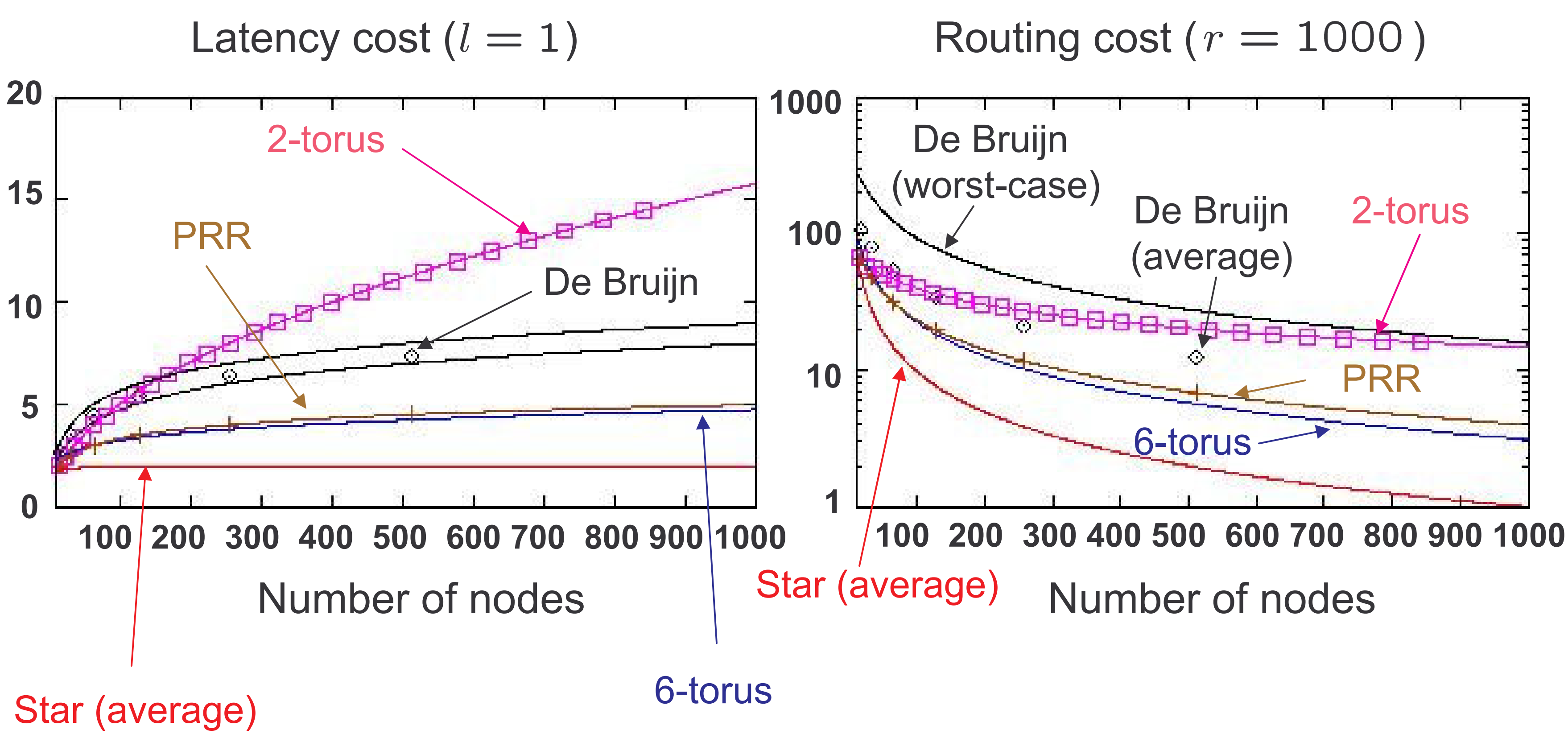
$$L_{\min} = \min_u L_u$$

$$R_{\max} = \max_u R_u$$

$$R'_{\min} = \min_u \{R_u : R_u > 0\}$$

(Δ, D)	$\frac{L_{\max}}{L_{\min}}$	$\frac{R_{\max}}{R'_{\min}}$
(2,9)	**1.11**	**4.51**
(3,6)	**1.04**	**4.41**
(4,4)	**1.03**	**2.71**
(5,4)	**1.02**	**2.78**
(6,3)	**1.01**	**1.86**

Routing and latency costs



Discussion

- Full meshes
  - Low overall resource usage
  - Low scalability
- Star networks
  - Low overall resource usage
  - Low scalability
  - Low resiliency
  - Very asymmetric
- Distributed designs (e.g., DHTs)
  - High resiliency, high scalability, good symmetry
  - Higher overall resource usage
- Best of both worlds: hierarchical networks?

Open problems

- Remove assumptions
  - Measurement studies?
- Incentive mechanisms in asymmetric networks
  - How much knowledge should the peers possess?
  - Use asymmetries to match heterogeneity?
- How to determine a meaningful set (l, s, r, m) for a class of applications?
- Network formation