Measurement-based Characterization of Peer-to-Peer Networks

Daniel Stutzbach and Reza Rejaie – University of Oregon

1. Motivation

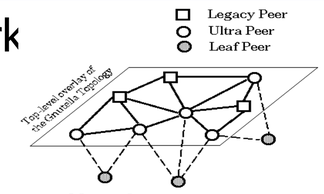
- Peer-to-peer (P2P) networks are becoming increasingly popular. P2P networks are inherently dynamic.
- Four key dimensions of dynamics in P2P networks are:
  - Peer participation (Churn)
  - P2P overlay topology
  - Query workload
  - Resource availability
- Characterization of these dynamics are essential for the design and evaluation of P2P protocols and applications.
- A few studies have conducted ***coarse*** characterization of dynamics, which is inadequate for design and evaluation.
- Proper characterization requires accurate *snapshots* of P2P networks. A common approach is to crawl the network which is slow and thus captures ***stretched*** snapshots.

2. Approach

- Developing a crawler and investigating key challenges in capturing the most accurate snapshots of P2P networks.
- Leveraging the snapshots to conduct characterizations.
- Focusing on Gnutella, the largest, open P2P network.

3. Crawling Gnutella

- Gnutella is a semi-structured P2P network
  - Top-level overlay (ultra-peers)
  - Leaf nodes



- We developed a fast, distributed crawler, called *Cruiser*.
- Cruiser contacts multiple top-level peers in parallel and retrieves their neighbor lists.
- The number of open connections is controlled adaptively.

➢ *Cruiser can capture a snapshot of the Gnutella network in less than 5 minutes, compared to the 30-60 minutes of previous studies.*

4. Results

Rate of Peer discovery

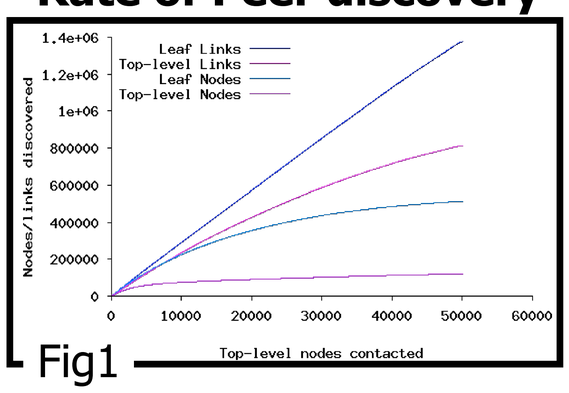


Fig1

Top-level Peer Dynamics

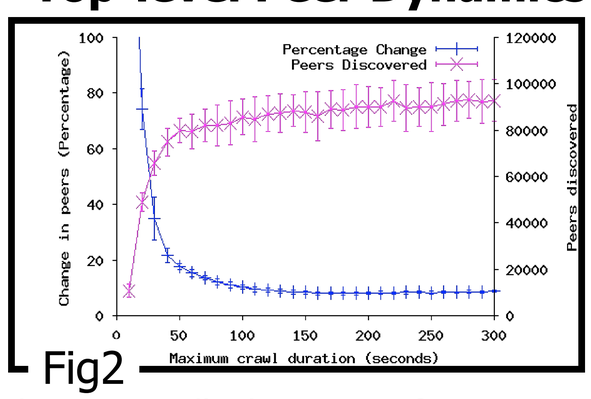


Fig2

- Fig 1, The number of newly-discovered peers rapidly decreases after a certain point. This shows the fundamental tradeoff of freshness vs. completeness.
- Fig 2, Capturing back-to-back snapshots with different crawling time reveals that around 8% of peer population is changing over a range of timescales. A two minute crawl time is sufficient to capture a complete top-level snapshot.

Top-level Peer Lifetimes

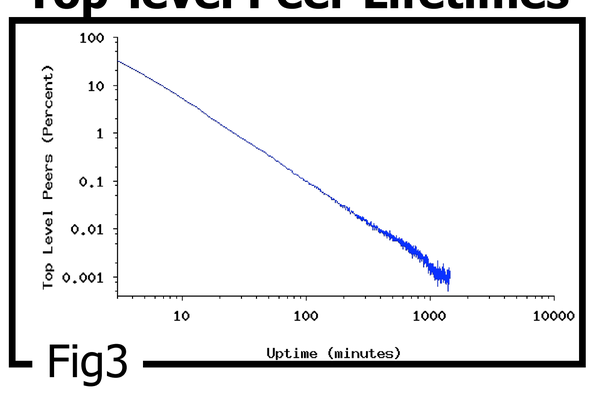


Fig3

Time-weighted Lifetimes

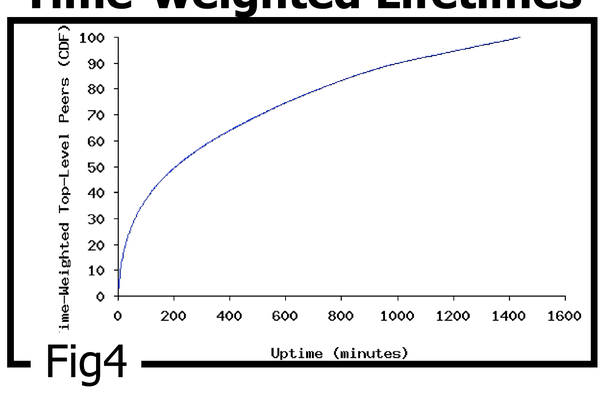


Fig4

- Fig 3, Peer lifetimes follow a *power-law distribution*, with an exponent around -1.76. No previous study has measured lifetime of all peers at this granularity.
- Fig 4, Although most top-level peers have short lifetimes (Fig3), most available top-level peers at any given time have long lifetimes.

A Week's Perspective

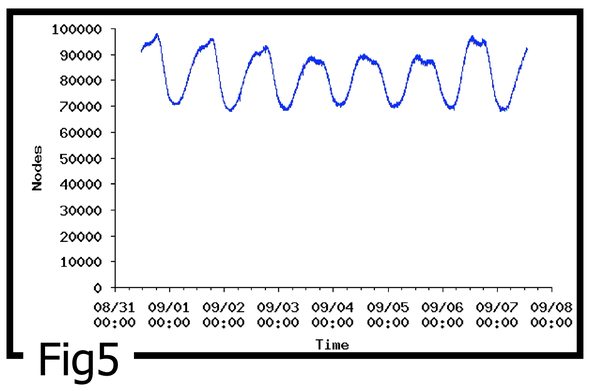


Fig5

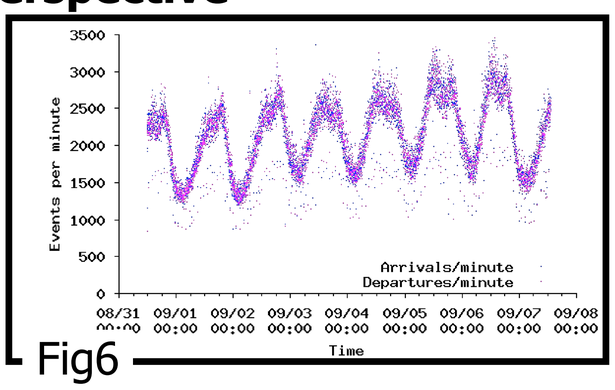


Fig6

- Fig 5, Back-to-back snapshots for one week show a diurnal effect, with two peaks, presumably for participating peers at the east and west coast of the US.
- Fig 6, Closer examination reveals a *close correlation between peer arrival & departure rates* over short (1 min) timescale.

Peer Degree Distribution

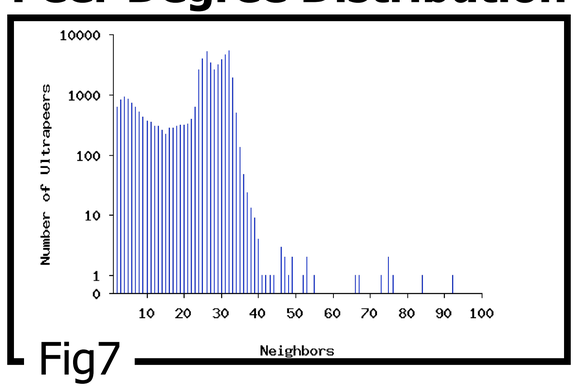


Fig7

Pair-wise Distance

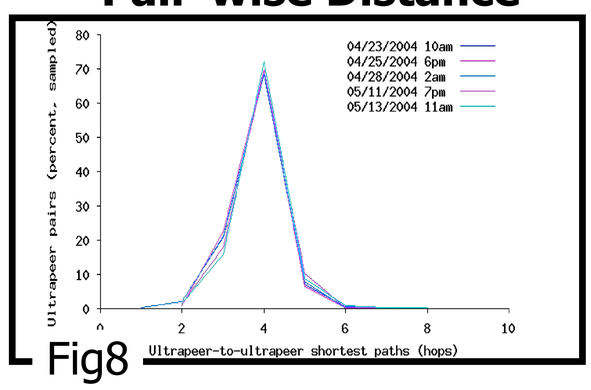


Fig8

- Fig 7, *Node degree is fairly homogenous*, with around 30 neighbors per Ultra-peer. This is different from prior studies that showed a power-law distribution.
- Fig 8, The network is still a *small-world* graph with low-path length and high clustering coefficients (0.01 compared to 0.0004 for same-size random graphs).

5. Conclusions & Future Work

- Gnutella Cruiser significantly reduces crawling time, thus improving accuracy of captured snapshots of P2P networks.
- Accurate snapshots enable us to characterize some aspects of dynamics in P2P networks at a finer granularity.
- We are currently characterizing churn, topology variations, and query workload in P2P networks, and any potential correlation among them.
- For further details visit *http://mirage.cs.uoregon.edu*