Elders Know Best: Lifespan-Based Ideas in P2P Systems

Yi Qiao and Fabián E. Bustamante

I. Introduction

The transiency of peer population and its implications on peer-to-peer (P2P) applications are increasingly calling the attention of the research community. As undesirable as unavoidable, peers transiency could negate many of the appealing features of the P2P approach.

We are exploring new P2P protocols and strategies that, by considering peers' lifespan a key attribute, can greatly boost the stability, efficiency and scalability of these systems. As part of our work, we performed a thorough study of peers' lifespan in a current and widely deployed P2P network. Through active probing of over 500,000 peers, we found that peer lifespan distribution can be modeled by a Pareto distribution of the form λT^k ($k < 0$) which, in this context, means that a peer's expected remaining lifetime is directly proportional to its current age.

II. Peer Lifespan and P2P Protocols

This last observation is the basis for a number of new protocols, first proposed in [1], for unstructured and loosely structured (ultra-peers) systems. In most P2P protocols there are a number of instances during a peer's life where it must choose among "acquaintances", such as when deciding whom to "befriend" (i.e. establishing a open connection) and when responding to a thir party's request for recommendation.

We have shown elsewhere [1], [2] that even simple lifespan-based P2P protocols can offer significant advantages, not only reducing control message processed by a node, but also greatly improving end-application performace such as caching efficiency, query hits and query resolution times. Trace-driven simulation results illustrate the potential benefits of lifespan-based protocols such as an over 42% reduction on the ratio of connection breakdowns to effective connections (a good indication of system stability) when compared with current deployed protocols. When using standard query and caching strategies such as parallel random walk query and PCX caching, lifespan-based protocols also help reducing query resolution time by 50%, while providing up to 60% higher query satisfaction rates. Moreover, even when relying on only 4 or 5 walkers, lifespan-based protocols can guarantee same query performance to that of 16 walkers over current unstructured and loosely structured P2P protocols, boosting system scalability by a factor of 3 to 4.

III. Elders Know Best

The reported improvements are a natural result of the *Used-Better-than-New-in-Expectation* property of peer lifespan distribution and our protocols' preference for older peers. We have started investigating the use of similar ideas as the basis for new query distribution, caching and replication strategies.

It seems obvious that a random-walk query strategy could significantly benefit from a bias toward old peers, as long-lived peers are likely to collect more shared files and have, thus, a higher likelihood to respond with a query hit. In addition, long-lived peers would have a better chance to route query and query-hit messages as well as remain up to honor subsequent requests for download.

Index caching and file replication could similarly benefit from a lifespan-based approach. One possibility we are exploring is the use of lifespan-based "region" (instead of "path") replication, making peers in the return path of a query push copies of files/index entries into potentially long-lived, nearby peers.

The advantages of a lifespan-based approach may not come for free. To begin with, our strategies' preference for long-lived peers could result in hotspots and their associated problems. We are currently addressing this by limiting the maximum connection number each peer can accept, an effective although implicit approach to load control.

A slightly less obvious problem is that of frequent collisions between query walkers. Due to the same preference for old peers, it is possible that lifespan-biased query walkers can run into some of the same nodes, reducing thus the effective coverage of queries. We are investigating this phenomena in our current query distribution strategy and have deviced a few possible responses.

Last but not least, we need to devise a viable approach to rightly determine peers' ages. We are currently considering a number of alternative reputation-based systems for age lookup and verification.

In addition to trace-based simulation, we have started to experiment with our ideas within PlanetLab, a wide-area experimental testbed. Our protocols and strategies are implemented as modifications to *Mutella* [1], a publically available Gnutella client.

References

[1] F. E. Bustamante and Y. Qiao. Friendships that last: Peer lifespan and its role in P2P protocols. In *8th WCW*, IBM T.J. Watson Research Center, Hawthorne, NY, USA, September 2003. To appear.

[2] Y. Qiao and F. E. Bustamante. The effects of lasting friendships in p2p protocols. Technical Report NWU-CS-23, Department of Computer Science, Northwestern University, 2003. Submitted for publication.

The authors are with the Department of Computer Science, Northwestern University, Evanston, IL 60201, USA. Emails: {yqiao,fabianb}@cs.northwestern.edu

[1] Mutella: *http://mutella.sourceforge.net*