Characterizing the Two-Tier Gnutella Topology

Daniel Stutzbach, Reza Rejaie
University of Oregon
{agthorr,reza}@cs.uoregon.edu

ABSTRACT

Characterizing the properties of peer-to-peer (P2P) overlay topologies in file-sharing applications is essential for understanding their impact on the network, identifying their performance bottlenecks in practice, and evaluating their performance via simulation. Such characterization requires accurate snapshots of the overlay topology which is difficult to capture due to the large size and dynamic nature. Previous studies characterizing overlay topologies not only are outdated but also rely on partial or potentially distorted snapshots. In this extended abstract, we briefly present the first characterization of two-tier Gnutella topologies based on recent and accurate snapshots.

Categories and Subject Descriptors

C.2.4 [**Computer-Communications Networks**]: Distributed Systems; C.4 [**Performance of Systems**]

General Terms

Measurement

Keywords

Peer-to-peer, Gnutella, Topology

1. INTRODUCTION

The Internet has witnessed an explosive growth in the popularity of Peer-to-Peer (P2P) file-sharing applications (*e.g.*, FastTrack, eDonkey, Gnutella) which in turn has led to an astounding increase in network usage by these applications [1, 2]. In these applications, participating peers form an ad-hoc overlay topology that dynamically changes as peers join or leave the overlay. Characterizing the properties of P2P overlay topologies is important for *(i)* understanding their impact on the network, *(ii)* identifying their performance bottlenecks and design anomalies in a realistic setting, and *(iii)* properly evaluating their performance through simulation.

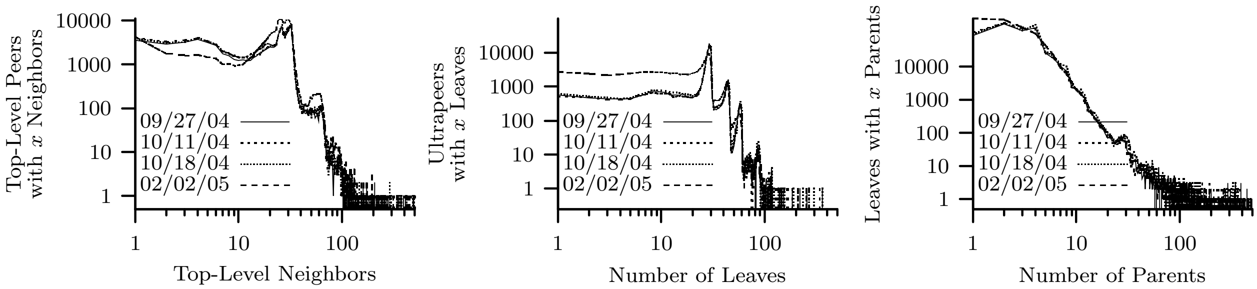
The characterization of P2P overlay topologies is a difficult task. Such a characterization is usually conducted by capturing a snapshot of the overlay with a crawler. The crawler progressively discovers participating peers (as nodes) and their pairwise connections (as edges) to present a snapshot of the overlay (as a graph). Examination of a single snapshot reveals graph-related properties of the overlay whereas comparison of subsequent snapshots presents dynamics of the overlay. Therefore, the accuracy of such characterizations directly depends on the accuracy and granularity of captured snapshots. In practice, P2P crawlers are often slow and thus unable to capture accurate snapshots of a large and dynamic P2P overlay. More specifically, since the overlay is changing during a crawl, captured snapshots by a crawler can be *distorted* where the degree of distortion depends on the speed of crawling relative to the rate of change in the overlay. Furthermore, the accuracy of captured snapshots might be affected by the non-negligible fraction of discovered peers that are unreachable by the crawler.

SIGMETRICS'05, June 6–10, 2005, Banff, Alberta, Canada.
ACM 1-59593-022-1/05/0006.

Despite its importance, little attention has been given to the characterization of P2P overlay topologies. Ripeanu et al. [4] mapped the Gnutella topology using captured snapshots from a slow crawler (250 peers/minutes). Saroiu et al. [5] captured partial snapshots of the Gnutella overlay topology and briefly examined its resiliency in the face of attack. These studies are outdated; the Gnutella network not only has grown from a few thousand nodes to more than one million but also has adopted a two-tier network architecture to improve its scalability[1]. More importantly, these studies have derived characterization of the Gnutella overlay topology from distorted or partial snapshots without any examination of snapshot accuracy. Therefore, the accuracy of their findings is unclear [7].

To conduct accurate characterizations of Gnutella topologies, we developed a parallel crawler, called *Cruiser* [6], which significantly increases crawling speed in two ways. First and foremost, it couples a master-slaver architecture with parallel crawling. Second, Cruiser leverages the two-tier architecture and only crawls the top-level overlay to discover the entire topology. Furthermore, a special handshaking feature in Gnutella is used to improve efficiency of data collection. Finally, Cruiser implements a load adaptation mechanism to effectively run on a distributed set of heterogeneous nodes. *Overall, Cruiser can capture the P2P overlay at the rate of 140,000 peers/minute which is orders of magnitude faster than any previously reported crawler.* Thus, Cruiser enables us to capture a snapshot of the Gnutella network with more than one million peers in around 7 minutes. We carefully examined the status of unreachable peers and considered their impact on the accuracy of captured snapshots. Our prior work [6] presents the design

[1]Modern Gnutella, as well as other popular P2P file-sharing applications, adopted such a two-tier topology. In this architecture, a small fraction of peers, called *Ultrapeers*, form an ad-hoc top-level overlay whereas the remaining peers, called *Leaves*, each connect to the overlay through a small number of ultrapeers. In modern Gnutella, ultrapeers have a high degree (*i.e.*, maintain 30 neighbors) in order to keep short path lengths between participating peers.



(a) Top-level Degree Distribution (b) Ultrapeers to Leaves (c) Leaves to Ultrapeers

Figure 1: Different angles on degree distribution in two-tier Gnutella topologies

of Cruiser, explores the accuracy of its captured snapshots, and examines the effect of various crawling tradeoffs on the accuracy of captured snapshots.

In this extended abstract, we briefly present a few characterizations of the two-tier Gnutella topology based on recent and accurate snapshots captured by Cruiser. Further characterizations of Gnutella can be found in a related technical report [8]. To our knowledge, this is the first measurement-based characterization of any two-tier overlay topology in a P2P file-sharing application.

2. CHARACTERIZATION OF GNUTELLA

During the past year, we have captured more than 18,000 snapshots of the Gnutella network with Cruiser, and conducted various analysis on the captured snapshots. Our main findings can be summarized as follows:

- The node degree in the top-level overlay does not exhibit a power-law distribution as reported in previous studies [3, 4]. More importantly, power-law distributions can be an artifact of distorted snapshots [6].
- The Gnutella network has dramatically grown over the past couple of years. Despite this increase, the diameter of the topology remains low because of the high degree of connectivity among ultrapeers. Thus, the overall topology has become denser and clearly exhibits small-world properties.
- Despite variations in the total number of peers with time of day, a large number of peers are available at any time, and the two-tier structure remains balanced.

In the remainder of this section, we present some of our results in further details.

Degree Distribution: We examine different angles of degree distribution in the two-tier overlay. As shown in Figure 1(a), the distribution of node degree in the top-level overlay has a significant spike around 30 which is the default maximum degree in the most popular Gnutella implementation (LimeWire). There are significantly fewer peers with a degree greater than 30. This clearly contrasts with the power-law degree distribution reported in previous studies [3, 4]. In our prior work [6], we showed that as the crawling speed decreases, the captured snapshot becomes more distorted and the degree distribution looks more similar to a power-law distribution. This suggests the power-law distribution for node degree in previous studies could be the result of measurement artifacts.

Figure 1(b) presents the degree distribution of connections from ultrapeers to leaves. Distinct spikes at a degree of 30, 45 and 75 are visible. The first two spikes are due to the corresponding parameters used in LimeWire and BearShare implementations, respectively. The third spike is due to a less common implementation. This figure shows that a significant minority of ultrapeers are connected to less than 30 leaves, which indicates the availability of open slots for new leaves to join the overlay. As Figure 1(c) depicts, most leaves connect to only a few ultrapeers, a small fraction of leaves connect to several ultrapeers, and few leaves connect to an extremely large number of ultrapeers.

Reachability: We also explore the distribution of shortest paths between peers in the top-level overlay. Despite dramatic growth, path lengths remain very short. More specifically, more than 60% of paths are exactly 4 hops in length while more than 99.5% of paths are at most 5 hops in length. Furthermore, there is little variation in the path length distribution across snapshots captured over the course of several months. This result reveals that the increased node degree effectively compensates for the growth in population to keep the pair-wise distance short. We also computed the clustering coefficient for the top-level overlay (0.012) and found it to be much larger than a corresponding random graph (0.00038). Thus, the Gnutella overlay is a small world graph.

3. REFERENCES

[1] slyck.com. http://www.slyck.com, 2005.

[2] T. Karagiannis, A. Broido, N. Brownlee, K. Claffy, and M. Faloutsos. Is P2P dying or just hiding? In *Globecom*, Nov. 2004.

[3] Q. Lv, P. Cao, E. Cohen, K. Li, and S. Shenker. Search and Replication in Unstructured Peer-to-Peer Networks. In *International Conference on Supercomputing*, 2002.

[4] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the Gnutella Network: Properties of Large-Scale Peer-to-Peer Systems and Implications for System Design. *IEEE Internet Computing Journal*, 6(1), 2002.

[5] S. Saroiu, P. K. Gummadi, and S. D. Gribble. Measuring and Analyzing the Characteristics of Napster and Gnutella Hosts. *Multimedia Systems Journal*, 8(5), Nov. 2002.

[6] D. Stutzbach and R. Rejaie. Capturing Accurate Snapshots of the Gnutella Network. In *Global Internet Symposium*, pages 127–132, Mar. 2005.

[7] D. Stutzbach and R. Rejaie. Evaluating the Accuracy of Captured Snapshots by Peer-to-Peer Crawlers. In *Passive and Active Measurement Workshop*, Extended Abstract, Mar. 2005.

[8] D. Stutzbach, R. Rejaie, and S. Sen. Characterizing Two-Tier Overlay Topologies in Modern P2P File-Sharing Systems. Technical Report CIS-TR-2005-01, University of Oregon, Feb. 2005.