A Measurement Study of Napster and Gnutella as Examples of Peer-to-Peer File Sharing Systems

P. Krishna Gummadi Stefan Saroiu Steven D. Gribble
Department of Computer Science and Engineering
University of Washington, Seattle, WA, USA, 98195-2350
{gummadi,tzoompy,gribble}@cs.washington.edu
http://www.cs.washington.edu/homes/gummadi/p2pfilesharing

In this paper, we present and analyze an extensive measurement study of Napster and Gnutella.

Motivation

After the enormous popularity of Napster and Gnutella, there has been a lot of recent research activity into peer-to-peer file sharing systems. Some example systems include Freenet, Chord, Publius, Tapestry, CAN, and PAST. An evaluation of the effectiveness of these systems requires the knowledge of the characteristics of the peers that choose to participate in these systems, such as their bottleneck bandwidth, latency, and availability. However, barring a few studies that capture certain specific characteristics of these systems, there are no extensive measurement studies that attempt to fundamentally characterize the peers. This knowledge could help evaluate the feasibility of deploying these theoretical models in the Internet. Hence, through our study we seek to precisely characterize the population of peers that participate in these networks with respect to the speed of their Internet connections, the frequency with which these peers connect and disconnect from the system and how many files peers share, upload, and download.

Results of our Measurements

In [1], we plot graphs generated from the data we gathered. We outline some of our broad conclusions here. (1) The participants in these networks are very heterogeneous with respect to all their characteristics. Hence, we believe that in order to be successful, the proposed peer-to-peer protocols should be modified to delegate functionality such as routing and file location to peers based on their characteristics. (2) We noticed that a considerable percentage (around 25%) of Napster peers are either unaware of their link bandwidths or deliberately misreport-report it to discourage other peers from choosing to download files from them. Hence, we believe that any peer-to-peer system should either have built in mechanisms to measure and verify the characteristics of the peers (for example, the infrastructure could directly measure bottleneck link bandwidths with high accuracy) or provide proper incentives to the peers for reporting accurate information (e.g., prioritizing the query evaluations by the bandwidths of the issuing peers). (3) Estimates from our data show that the median session duration of the peers (i.e., the difference between timestamps taken when a peer joins the network and when it leaves the network) to be around an hour (Figure 1), while the average session duration is much smaller. This necessitates the incorporation of quick network stabilization mechanisms into these protocols to handle entry and departure of nodes. (4) Previous studies confirm that vertex connectivity in overlays such as Gnutella tend to obey power law distributions and are very resilient to random failures of the nodes in the network. In fact, our study shows that the Gnutella overlay fragments only when about 70% of the nodes break down. On the other hand, we found that an orchestrated attack (such as a DOS attack) against the top 4% of high degree nodes is enough to shatter the overlay into hundreds of disjoint fragments.

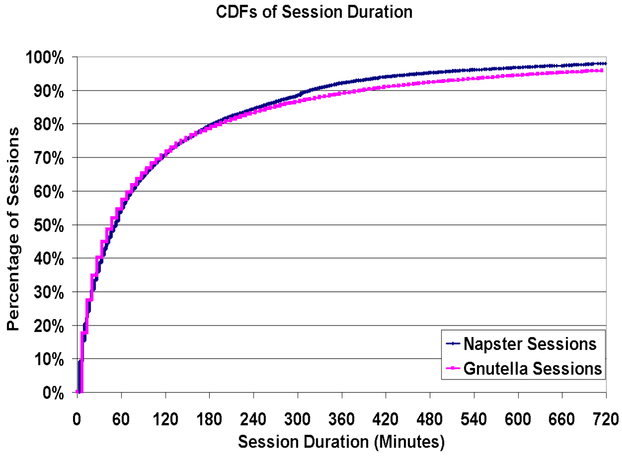

Figure 1. Distributions of Napster/Gnutella session durations.

References

[1] S. Saroiu, K. Gummadi, and S. Gribble. A measurement study of peer-to-peer file sharing systems. In *Proceedings of Multimedia Computing and Networking 2002*, San Jose, CA, USA, January 2002.