

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

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

In this poster, we present the results of our extensive measurement study of Napster and Gnutella along with some preliminary analysis. We also wish to request suggestions on the ways in which the data we gathered could be used.

## 1 Motivation

After the enormous success of Napster, there has been a lot of recent research activity into peer-to-peer file sharing systems. Apart from Gnutella, some example systems include Freenet [2], Chord [6], Publius [7], Tapestry [8], CAN, [4] and PAST [3]. These file sharing systems differ significantly in the way the resources of the system (i.e., files being shared) are *named* and *located*. An evaluation of the effectiveness of these systems requires the knowledge of the fundamental characteristics of the peers of these systems, such as their bottleneck bandwidth, latency, and availability. However, barring a few studies [1, 5] which capture certain specific characteristics of these systems, there are no extensive measurement studies that attempt to fundamentally characterize the peers. This knowledge could help them evaluate the feasibility of the deployment of their 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 bottleneck link bandwidths between these peers and the Internet at large, the frequency with which these peers connect and disconnect from the system and how many files peers share, upload, and download.

## 2 Results of our Measurements

We plot graphs showing some of the data we gathered in the poster accompanying this draft. We outline some of our broad conclusions here. (1) The participants in these networks are very heterogeneous with respect to every aspect of their characterization. Hence, we conclude that attempts by these proposed peer-to-peer protocols [8, 3, 6, 4] to distribute functionality such as routing and file location uniformly among these heterogeneous peers are unlikely to succeed. (2) We noticed that a considerable percentage (around 25%) of Napster peers are either unaware of their link bandwidths or deliberately mis-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 peer

for reporting accurate information (e.g., prioritizing the query evaluations by the bandwidths of the issuing peers). (3) Previous studies [6] 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, the study claims that the Gnutella overlay fragments only when about 70% of the nodes break down. However, 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.

### 3 Future Work

We are yet to analyse our traces of mappings of file titles to sizes and locations for Napster peers, which could yield valuable information to evaluate the different indexing strategies proposed to search for files. We also plan to use the our data and analysis to evaluate the different distributed index schemes proposed in [8, 3, 6, 4].

### References

- [1] E. Adar and B. Huberman. Free riding on gnutella. In *First Monday, Volume 5, Number 10*, October 2000.
- [2] I. Clarke, O. Sandberg, B. Wiley, and T. Hong. Freenet: A distributed anonymous information storage and retrieval system. In *Proceedings of the ICSI Workshop on Design Issues in Anonymity and Unobservability*, Berkeley, CA, 2000.
- [3] P. Druschel and A. Rowstron. Past: A large-scale, persistent peer-to-peer storage utility. In *Proceedings of the Eighth IEEE Workshop on Hot Topics in Operating Systems (HotOS-VIII)*, Schoss Elmau, Germany, May 2001.
- [4] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. A scalable content-addressable network. In *Proceedings of the ACM SIGCOMM 2001 Technical Conference*, San Diego, CA, USA, August 2001.
- [5] K. Sripanidkulchai. The popularity of gnutella queries and its implications on scalability, 2001. <http://www.cs.cmu.edu/~kunwadee/research/p2p/gnutella.html>.
- [6] I. Stoica, R. Morris, D. Karger, F. Kaashoek, and H. Balakrishnan. Chord: A scalable content-addressable network. In *Proceedings of the ACM SIGCOMM 2001 Technical Conference*, San Diego, CA, USA, August 2001.
- [7] M. Waldman, A. Rubin, and L. Cranor. Publius: A robust, tamper-evident, censorship-resistant, web publishing system. In *Proceedings of the 9th USENIX Security Symposium*, August 2000.
- [8] B. Zhao, K. Kubiatowicz, and A. Joseph. Tapestry: An infrastructure for fault-resilient wide-area location and routing. Technical Report UCB//CSD-01-1141, University of California at Berkeley Technical Report, April 2001.