# Peer-to-Peer Architecture Case Study: Gnutella Network

Matei Ripeanu
*Computer Science Department,*
*The University of Chicago*
*matei@cs.uchicago.edu*

## Abstract

*Despite recent excitement generated by the P2P paradigm and despite surprisingly fast deployment of some P2P applications, there are few quantitative evaluations of P2P systems behavior. Due to its open architecture and achieved scale, Gnutella is an interesting P2P architecture case study. Gnutella, like most other P2P applications, builds at the application level a virtual network with its own routing mechanisms. The topology of this overlay network and the routing mechanisms used have a significant influence on application properties such as performance, reliability, and scalability. We built a 'crawler' to extract the topology of Gnutella's application level network, we analyze the topology graph and evaluate generated network traffic. We find that although Gnutella is not a pure power-law network, its current configuration has the benefits and drawbacks of a power-law structure. These findings lead us to propose changes to Gnutella protocol and implementations that bring significant performance and scalability improvements.*

## 1. Introduction

Peer-to-peer systems (P2P) have emerged as a significant social and technical phenomenon over the last year. P2P systems provide infrastructure for communities that share CPU cycles (e.g., Entropia, SETI@Home) and/or, storage space (e.g., FreeNet, Gnutella), or that support interpersonal collaborative environments. Two factors have fostered the recent explosive growth of such systems: first, the low cost and high availability of huge computing and storage resources, and second, increased network connectivity. As these trends continue, the P2P paradigm can be expected to become more popular.

Peer-to-peer networks allow individual computers to communicate directly with each other and to share information and resources without using specialized 'servers'. A common characteristic of this new breed of applications is that they build, at the application level, a virtual network with its own routing mechanisms. The topology of this virtual network and the routing mechanisms used have a significant influence on application properties such as performance, reliability, and, in some cases, anonymity. The virtual topology also determines the communication costs incurred when running the P2P application.

These considerations have motivated us to conduct a detailed study of the topology and protocol of a popular P2P system, Gnutella. In this study, we benefited from Gnutella's large existing user base and open architecture, and, in effect, use the public Gnutella network as a large-scale, if uncontrolled, testbed. We capture the network topology, the generated traffic, and the resources' dynamic behavior. We use our findings to evaluate costs and benefits of the P2P approach and to investigate possible improvements to the routing protocol that would allow for better scaling and increased reliability.

*Related work*: Distributed Search Solutions (DSS) group publishes partial results of their Gnutella research [4]. A number of other research projects build on DSS's data: [2] analyzes Gnutella user behavior while [6] focuses on efficient search protocols in power-law networks. T. Hong [5] uses medium scale simulations to explore Gnutella network properties. However, our network crawling and analysis technology (developed independently of this work) goes significantly further in terms of scale (both spatial and temporal) and sophistication.

## 2. Gnutella Network Analysis

Gnutella is an open, decentralized, P2P search protocol [3] that is mainly used to find files. In addition, the term Gnutella designates the virtual network of Internet accessible hosts running Gnutella-speaking applications We used this network for our measurements over a 6 month period.

In a Gnutella network, each maintains open TCP connections with at least one other node, thus creating a virtual network of servents at the application level. Query and group maintenance messages are propagated using a flooding technique while query reply messages are back-propagated. We have developed a *crawler* [7] that joins the network uses the membership protocol [3] to collect topology information.

Using the records of multiple successive crawls, we investigated the dynamic graph structure over time. We discover that about 40% of the nodes leave the network in less than 4hours while only 20% of the nodes are 'persistent' (they are alive after 24h). We also used the crawler to eavesdrop the traffic generated by the network. Figure 1 categorizes generated traffic according to message type. We detected that the volume of generated traffic is the main obstacle for better scaling and wider deployment.

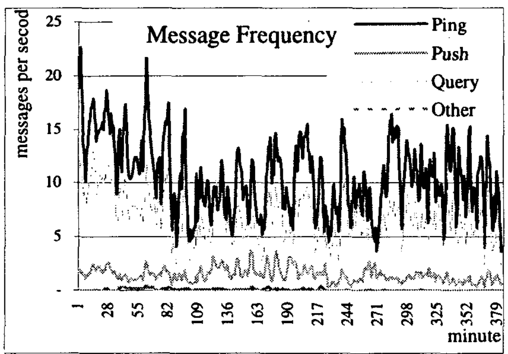


**Figure 1: Generated traffic split by message type** over a 376 minute period. Note that overhead traffic (PING messages serve only to maintain network connectivity) forms more than 50% of traffic. For backward compatibility reasons, flooding is also used to deliver some file download requests (PUSH messages). The only 'true' user traffic is QUERY messages.

Recent research [1,8] shows that networks ranging from natural networks such as molecules in a cell or people in a social group to the Internet organize themselves so that most nodes have few links and a small number of nodes have many links. In [9], Albert finds that networks following this organizational pattern (power-law networks) display an unexpected degree of robustness when facing random node failures: the ability of the network to communicate is unaffected by high failure rates. However, error tolerance comes at a high price: these networks are extremely vulnerable to attacks, i.e., to the selection and removal of a few nodes that play the most important role in assuring the network's connectivity. We show (in detail in [7]) that Gnutella is similar to a power-law network, thus being able to operate in highly dynamic environments.

## 3. Summary

Gnutella is an open, decentralized, P2P search protocol that is mainly used to find and share files. Computers running Gnutella protocol-compatible software form an application-level network. We have developed tools to discover and analyze this network. Our analysis shows that Gnutella node connectivity follows a multi-modal distribution, combining a power law and a quasi-constant distribution. This property keeps the network as reliable as a pure power-law network when assuming random node failures, and makes it harder to attack by a malicious adversary. Gnutella takes few precautions to ward off potential attacks. For example, the network topology information that we have obtained here is easy to obtain and would permit highly efficient denial-of-service attacks. Some form of security mechanisms that would prevent an intruder from gathering topology information appears essential for the long-term survival of the network (although it would make global network monitoring more difficult if not impossible).

We see two directions for improvement. First, we observe that the application-level topology determines the volume of generated traffic, the search success rate, and application reliability. We could design an agent that constantly monitors the network and intervenes by asking servents to drop or add links as necessary to keep the network topology optimal. Agents (or nodes) could embed some information about the underlying physical network and build accordingly the virtual application topology. Note that implementing this idea requires only minor protocol modifications.

The second direction is to replace flooding with a smarter (less expensive in terms of communication costs) routing and group communication mechanism. We have collected a large amount of data on the environment in which Gnutella operates and shall use it in simulations to investigate the appropriateness of these and various other alternatives.

## 3. References


[1] M. Faloutsos, P. Faloutsos, and C. Faloutsos, On Power-Law Relationships of the Internet Topology, SIGCOMM 1999.

[2] E. Adar and B. Huberman, Free riding on Gnutella, First Monday Vol. 5, Number 10, - Oct. 2nd 2000.

[3] Gnutella protocol specification v4.0. Available at: http://dss.clip2.com/GnutellaProtocol04.pdf

[4] DSS group, Gnutella: To the Bandwidth Barrier and Beyond, available at: http://dss.clip2.com, Nov. 6, 2000.

[5] T. Hong, Performance, In Peer-to-Peer: Harnessing the Power of Disruptive Technologies, ed. A. Oram. O'Reilly and Associates: Sebastopol, CA, 2001.

[6] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman, Search in Power-Law Networks

[7] M. Ripeanu, Peer-to-Peer Architecture Case Study: Gnutella Network, University of Chicago Technical Report TR-2001-26.

[8] A. Barabasi and R. Albert. Emergence of scaling in random networks, Science, 286(509), 1999.

[9] R. Albert, H. Jeong, and A. Barabási, Attak and tolerance in complex networks, Nature 406 378 (2000).