Future P2P Research Directions

Alex Valentine
alexvalentine@psu.edu

Summary

The LionShare project is a secure file sharing system designed to meet the needs of higher education and other learning institutions. The version 1.0 release of the LionShare project, due out in the fall of 2005, will provide a federated p2p network with support for authentication, authorization, advanced metadata, and various collaborative tools. Beyond the LionShare 1.0 release, we are looking at two different future development tracks.

The first track is to extend the capabilities of the LionShare network. Not only adding features, but also transforming LionShare from a research project to an academic computing service used in multiple institutions. The second track of development, looks at the possibility of working on new P2P projects. Either track is not mutually exclusive, research in to other areas does not necessarily mean that the further development of the LionShare network will stop. This document focuses on the second track, possibilities for new P2P development projects.

This is not a technical paper. This project description is a brief non-technical summary to be used for discussion and feedback purposes. The summary focuses on the possibility of extending the Bittorrent tracker concept to support academic content and centralized authentication.

Bittorrent Overview

In the past two years, much has changed in the area of P2P protocols and technologies. Many of the widely used protocols such as Gnutella and Fasttrack have added new features, increasing scalability and search accuracy. While some developers and companies work to extend older protocols, others have been working on entirely new P2P protocols. One such protocol is Bittorrent, which was first released in 2001. Bittorrent was designed to completely decentralize the distribution of large files over the Internet. The initial concept was born out of the popularity of Linux distribution CD image files.

When a new version of a Linux distribution is released on the Internet, it is generally posted to a handful of FTP and webservers. Those servers function as the only distribution point. Most users can't download an newly released cd image, because of the shear amount of traffic going to handful of distribution sites.

Bittorrent inventor Brahm Cohen realized that P2P would be an excellent solution for the problem of distributing large files to growing population of downloaders. The Bittorrent protocol is centered around the concept of swarm downloading. A Bittorrent publisher can seed a file to the network. The publishing user (Seeder) makes a special map of the file called a torrent, which also contains information for the client on how to download the file. The torrent file can be distributed to users through any means (www, e-mail, ftp). The downloading user (Peer), loads the torrent file in their client, which contacts the a

tracker. The trackers keeps Bittorrent organized, by keeping tabs on what peers and seeders are online and what portions of the file are available. The Bittorrent tracker acts as the matchmaker for Peers and Seeders.

A torrent file maps a large file or set of files in to a series of chunks. When the file set is seeded, peers can download portions of the file set. Once a peer starts downloading chunks, it immediately begins to share portions of the file. Bittorrent peers can download file chunks from multiple seeders and peers at the same time. Once a Peer downloads all the chunks of a file, it immediately becomes a Seeder. The popularity of a file only increases the amount of distribution points, thus increasing the efficiency of distribution across the network.

The Bittorrent concept brought a lot of new ideas to paradigm of P2P file distribution. The concept of a torrent file lets people download the torrent from an official website, thus insuring the contents and distributor of a file even though the actual file data could be downloaded from thousands of individual users.

Torrent files are usually downloaded from websites called torrent portals. For many of these portals, the website also serves as the tracker. Torrent portals can keep track of activity, provide discussion boards, and other various utilities to facilitate the file distribution process. Some torrent portals even provide basic authentication. Users login to a torrent portal site. When requesting a torrent for download, a username and ip address of the requester are stored in a database. The ip of the user is then authorized to contact the tracker to initiate transfer.

Bittorrent Possibilities

Bittorrent has grown from being a utility designed to distributing high demand large files, to an all purpose file distribution protocol. Bittorrent is now the fasting growing P2P protocol on the Internet. The LionShare concept took P2P file distribution using the Gnutella protocol and integrating federated authentication and authorization. The LionShare concept was designed under the goal of bridging the gap between decentralized P2P and centralized services to create accountable file sharing on academic networks.

The Bittorrent protocol has a lot of advantages for organizations looking to make use of P2P file distribution. Torrent files can be distributed from a centralized location such as a departmental web server. Generally, .torrent files are very small in size, averaging less than 100k per torrent. The centralized distribution of .torrent files lets an organization distribute official content through P2P.

Academic institutions can use the Bittorrent protocol as a distribution channel for e-learning materials in course management systems. The scalability and distribution model of the Bittorrent protocol makes it perfect for delivering large media files such as video and audio directly to CMS users. Besides Another potential use for the Bittorrent model is content distribution for learning object repositories.

In the library community, centralized learning object repositories are the preferred

method for content distribution. The centralized model requires a lot of start-up needs in order to support heavy bandwidth and centralized administration for adding new learning objects and metadata. The Bittorrent model of distribution can provide the best of both worlds, centralized control over content and P2P distribution.

Bittorrent portals have many set up options. The most popular configuration is where users post their own torrent files to a portal. Another configuration would be have librarian posted content on a portal, where librarians act as the gateway for quality control purposes. Because Bittorrent portals are easy to set up and require minimal resources, administrators can set up a variety of portals for particular communities of interest.

Development Possibilities

The current Bittorrent open source projects provide a great platform for developing a project to meet the needs of the Academic community. A project of this type would be broken down in to two main development categories: Security and Bittorrent development.

The security portion of this project would integrate centralized authentication in to the protocol, tracker, and the portal. The LionShare security platform built off the the Shibboleth concept, could be leveraged for the security integration. The portal would require authentication for torrent access. Once a user downloads a torrent to their client, the tracker could require authentication in order to to connect the the distribution network. Using digital signatures for content verification is another area where a Bittorrent project could leverage off the LionShare security infrastructure.

Besides integrating centralized authentication and security in the Bittorrent concept, there are also quite a few modifications that can be done at the application level to make the Bittorrent model more appealing to the academic and research communities. The current model has very little support for metadata. The integration of metadata support in to the torrent file and Bittorrent portal could make this a serious tool for academic content publication. Because the portal is a web application, web based community tools such as messages boards, chat, ratings and other community tools will be integrated in to the academic portal software.

Summary

The Bittorrent protocol is a powerful tool for online content distribution. The distribution model lets an infinite number of users access any particular resource, without clogging the content distribution medium. The concept of a client, tracker, and web based portal could become the next great model for online content distribution. The centralized nature of Bittorrent model, makes it a prime candidate for potential use in the library and e-learning spaces. The publishing model is very flexible, both centralized publishing or end user publication features are available. With the integration of security and support for organizational and collaborative features, the Bittorrent concept can be transformed in to a great academic content distribution system.