

VIUS Reports 9.1

Description of LionShare Architecture

Prototype Peer-to-Peer Service: LionShare

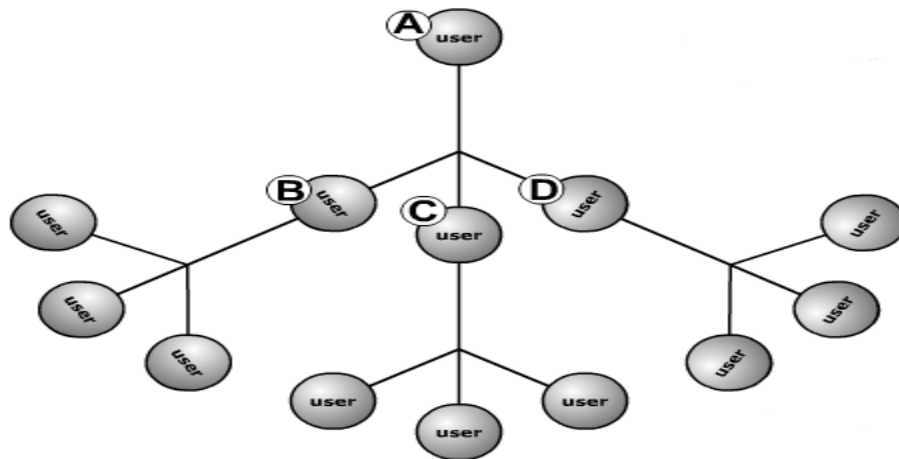
The popularity of file sharing and instant messaging applications on the Internet has led to the latest revolution in network computing, known as peer to peer (P2P). File sharing and IM are just one part of the P2P paradigm, distributed computing, web services and networked devices all fall under the realm of P2P. Simply put, P2P is a class of autonomous applications that take advantage of resources such as storage, CPU cycles, bandwidth, content, and human presence on computer networks. P2P on the Internet has grown remarkably, but what we've witnessed so far is only the beginning. P2P use will not be relegated to just the Internet, soon there will be a variety of applications for enterprise and academic environments. Once thought of as nuisance, P2P will evolve into an indispensable tool for organizations. We are starting to see this evolution with the growth of web services and enterprise IM, but we are only standing on the threshold of what's to come.

LionShare is an attempt to apply the P2P file sharing paradigm to an academic environment. LionShare is an academic P2P system which will assist in the distribution of academic materials through the University's network and beyond. The primary goal of LionShare is to provide a P2P network which has an unlimited potential for collaboration among faculty, students, departments, and even multiple universities. Most importantly, it provides a framework for harnessing the power of this once rogue technology, and transforming it into an academic tool for the future.

Gnutella

LionShare uses the Gnutella protocol and source code from the Limewire open source project, so before we can describe LionShare in some detail it's important to have an understanding of Gnutella and the history behind the protocol. On March 14th of 2000, news broke out that Nullsoft software was releasing a Napster clone. Nullsoft's most notable software at the time was Winamp, which even today is the most popular PC mp3 player. Nullsoft's P2P application which was available to download for only a few short hours was known as Gnutella. Nullsoft's parent company AOL, ordered the removal of the application, due to the media conglomerate's conflicting record company interests. Over ten thousand people downloaded Gnutella while it was available and in a short time there were many Gnutella clones which were compatible with the original protocol. There were many aspects that made the Gnutella protocol revolutionary, the most important of these being the fact that the Gnutella protocol is open and completely decentralized. When a user connects to the Gnutella network, the client connects to a specified number of other computers known as peers. A peer is any computer that is connected to the Gnutella network. Figure 1, below, is an example of a small Gnutella network with a client directly connected to three hosts.

Figure 1. A Small Gnutella Network



When peer A initiates a query with peers B, C, and D, they pass the query along to the other peers connected to them. When the query is matched with a result, the notification is sent directly to peer A. When a query is initiated on the Gnutella network it is passed along until there are no other peers left to receive it or no bandwidth left to pass on the query.

The Gnutella protocol is an open protocol giving anyone the ability to write software to connect to the Gnutella network. The openness of the Gnutella protocol has led to the development of many different clients which are all compatible with each other. Currently there are over fifty different programs that can connect to the Gnutella network. The most popular clients are Morpheus, Bearshare, and Limewire. Morpheus is a recent addition to the Gnutella network. Previously it relied on closed file sharing technology known as Fasttrack. To give a sense of how popular file sharing applications using Gnutella have become, here are the most recent figures from download.com. Please keep in mind that a large portion of the Morpheus downloads occurred when it used the Fasttrack protocol. Figure 2 shows how many clients were downloaded during the last week of April, 2002. This will give a sense of how fast file sharing is growing and the demand for such applications.

Figure 2. Downloads of Gnutella Clients

| Client | Total downloads |
|-----------|-----------------|
| Morpheus | 89,122,202 |
| Bearshare | 14,909,149 |
| Limewire | 11,716,916 |

| Client | Downloads (4/21-27/2002) |
|-----------|--------------------------|
| Morpheus | 1,130,635 |
| Bearshare | 414,228 |
| Limewire | 291,888 |

With the addition of other Gnutella clients and other download sites, there are over 2 million Gnutella clients downloaded every week. This is probably a conservative number considering we are only looking at the statistics of Windows platform downloads. Although other platforms make up a small percentage of Internet clients, its worth mentioning that Limewire is the most popular Macintosh and Linux application available for download. The Gnutella network is a community that is growing by the millions every week. File sharing programs have been the most popular applications available for download since the introduction of Napster over three years ago. The Record Industry Association of America (RIAA) popularized Napster by initiating a legal battle which ultimately led to the shutdown of Napster, but the publicity of the trial may have done more harm than good. The post-Napster P2P applications have

become more popular than Napster ever was during its heyday. The development of LionShare is an attempt to take the usefulness and popularity of file sharing, and forge it into tools to assist learning and encourage collaboration in academic environments.

LionShare: P2P in the Academic World

There is no question that technologies such as file sharing and instant messaging have reached a state of mass popularity on the Internet, but how do we transfer these technologies to the controlled environments such as a university? LionShare is an application that harnesses the power of Gnutella and transforms it from a technology which was once considered disruptive, into an indispensable tool for faculty and students alike. The proliferation of multimedia in the academic world has led to a revolution in learning. Instead of chalkboards and handouts, faculty members are making use of PowerPoint presentations, Flash animations and websites in the classroom. Students are submitting digital files instead of handing in papers. Digital imaging and video once relegated to early adopters, is becoming an everyday sight in the classroom. There is no question that digital media benefits both students and faculty. The real question is how can we extend the usefulness of this new media in the academic environment?

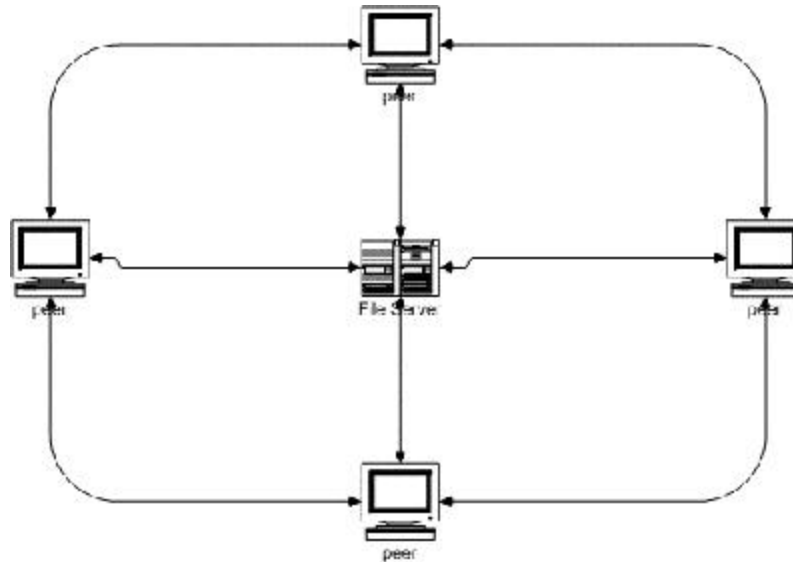
P2P technologies have quickly become the most efficient means of media distribution on the Internet. Single points of distribution such as web sites and FTP servers are no match for the scalability and ease of use P2P offers. One only needs to look to Napster as an example for ease of use. A great portion of Napster users had a limited knowledge of the Internet. Most did not know how to publish a web page or use a FTP site. File sharing gave the computer illiterate the power to acquire and publish digital media. With a certain modifications, the usefulness of file sharing can be transformed in to an organizational tool.

The LionShare Network

The LionShare client and file server are based upon the source code of the Limewire open source project. Limewire has become one of the most popular Gnutella clients available because of its intuitive interface, enhanced features, and availability on all popular platforms. The Limewire source code, written in Java, is well documented, stable, and available to the public for download under the GNU general public license (GPL). The Java language gives Limewire its multiplatform capabilities. The Limewire client is available for binary download on Windows, Macintosh, and Linux environments. With the availability of source code, Limewire code can be compiled on any computing platform that Java supports.

LionShare will use a heavily modified version of the Limewire client, adjusted to the needs of the academic environment. Gnutella is famous for being a completely decentralized network. In the structured environment of a University a certain degree of centralization is necessary for security, administration, and user convenience. One feature that differentiates LionShare from other file sharing networks is the ability for users to share files via the client-server model, as well as the tradition P2P method for sharing files. Users will have the ability to publish files to a file server on the network. The publish function is necessary for users who want to share certain files via LionShare but don't want to leave the client running on their computer.

Figure 3. LionShare's Persistent Server



LIONSHARE

Figure 3 shows how the file server will fit in to the LionShare network, please keep in mind this is not an accurate depiction of Gnutella traffic flow, the illustration in the previous section dealing with Gnutella gives an accurate representation of Gnutella traffic flow.

Published files on the LionShare will have the same capabilities of files shared locally on a user's hard drive. Users will be able to remove published files and control access at the user, group, or department level. This function is explained with further detail in the next section. The publish functionality combines the usefulness of the client/server architecture with the flexibility of P2P. One of the key factors in determining the success of a P2P network is user participation. Internet file sharing networks don't have participation problems because of user interest, specifically in attaining often illegally shared media. Organizational file sharing does not have the same high degree of interest so with the hybrid topology, users can share files without being connected to the LionShare network.

The LionShare network also makes use of Limewire's ultrapeer technology. In the early days of the Gnutella network, modem connections would slow queries to crawl because there was no real hierarchical structure to the network. Ultrapeer organizes the Gnutella network by connecting lower bandwidth peers to higher bandwidth peers, which are known as ultrapeers. With ultrapeer technology, modem connections stay where they belong, at the edges of the network. The use of ultrapeer technology with LionShare will be less significant than the actual Gnutella network because most traffic will occur from within Penn State's high speed network.

Another feature which LionShare makes extensive use of is metadata. Simply put, metadata is data which describes other data. An example of a commonly used metadata format is the id3 tag which is used to describe MP3 files. MP3 filenames alone are not very good at describing what is on a specific file. With the use of ID3 tags, music players and organizers automatically present relevant information embedded in the ID3 tag such as: the Artist, Album, and year published. Metadata helps extend the usefulness of digital files by providing criteria from which to search, organize, and categorize digital media. The latest version of Limewire includes metadata capabilities which are new to the Gnutella network. LionShare will incorporate these new metadata features in the network, including the ability to search for specific media such as: video, audio, software, pictures, and e-books. LionShare will also give users the ability to annotate media files using XML. The annotation feature will provide an easy to use interface for users to add descriptions to existing digital media.

It is important to support metadata descriptions that have a standardized structure to achieve the greatest interoperability of content for a variety of academic applications. Thus, LionShare will support the IMS Metadata specification. This specification has been specifically developed to support academic uses of technology and the description of learning objects. The IMS Metadata specification has been employed by two PSU projects. The first is the Committee for Institutional Cooperation Educational Resources On-line (CICERO) project – a learning object metadata repository and the second, the Mellon Foundation Visual Image User Study (VIUS). The reliance on a common metadata schema will allow the two projects to exchange common descriptive information about the materials in either repository.

Security

File sharing networks are typically associated with anonymous transfer of files. Users don't know who they are downloading from or who is downloading their shared files. Most P2P networks don't even require users to authenticate before logging on to the network. There is no degree of access control on file sharing networks except to shared or unshared. Shared files are available to everyone; unshared files are available to none. In the structured environment of a university network, authentication, access control, and logging are mandatory. In order for a file sharing network to thrive in an academic environment, security must be the highest priority.

LionShare will require users to authenticate upon initialization of the client. Authentication is accomplished via the Kerberos network authentication protocol which protects sensitive login data with secret key cryptography. Alternative authentication methods such as LDAP are also a possibility. Users will log in with their standard Penn State userid. The authentication server keeps track of user information, which can be used to classify userids. LionShare users can be classified by function (student, faculty), department, or by group. User information is critical for the access control function. Access control can be specified by user type, department, and users will have the ability to create custom groups with individual userids. When files are published or downloaded, the poster's userid will follow the file wherever it goes. This feature will help prevent unauthorized use and it will also serve as a convenience when someone needs to get in contact with the original poster or author of a file.

Because of the great potential for the misuse of file sharing, all LionShare activity will be logged in order to deter misuse. Users will be informed upon client initialization that all LionShare activity is logged. There will also be several warnings imbedded in the client to prevent the sharing of copyrighted materials. The client could be modified with pop up warnings when users attempt to share file formats that are commonly used to store copyrighted materials such as MP3 and video files. A FAQ explaining what files can be shared and an unauthorized use policy, need to be prominently displayed and clearly written in order to prevent inadvertent illegal sharing. Every possible step needs to be in place to avoid the use of LionShare for the distribution of copyrighted and illicit files. With the proper controls in place, file sharing networks can be efficient and legal on closed networks and even the Internet. A recent success story is the Furthur network, a file sharing network for trading music. Certain bands allow the trading of live concert tapes recorded by taping enthusiasts. With very few controls in place the Furthur network has remained 100% legal with over 1,500 unauthenticated anonymous peers connected simultaneously on any given day. The success of the Furthur network shows that if legal file sharing is attainable in the wilds of the Internet, file sharing in the controlled environment of a University is possible.

Additional Features

File sharing is only the beginning for LionShare. The application has the potential to evolve into a system which serves as a convergence for a variety of technologies and applications. Instant messaging, chat, and media organization tools can be incorporated into the LionShare client. File sharing may end up being just one part of the LionShare system, used as a foundation to incorporate multimedia and communications tools. The exponential growth of digital media has led to a new problem, how do users keep track of their digital files? Media that was once kept on a bookshelf, in a jewel case, on a VHS tape or inside a photo album, now resides on hard drives. The proliferation of digital files with the popularity of digital cameras and digital music, has led to an organizational conundrum. How do you organize a digital file collection? Traditional file systems don't provide enough functionality and information for the average

PC user. People want to be able to search and organize their data in a manner in which they see fit. Using LionShare as a foundation and metadata as the solution, LionShare can become a digital media hub.

Media Organization. There are already many different media organization tools on the Internet. Some of the more popular packages are ACDsee and Apple's iphoto. LionShare will give users the ability to organize their collections by different criteria. Subject matter, date, media type, folder, and custom descriptors are all examples of media criteria. Some file types such as MP3 files, already have built in metadata to categorize the file. LionShare will have the ability to read this data and incorporate it into a media database. Functionality of the media organizer will extend beyond just organization, users will be able to view, launch, and even play certain media types from within the LionShare client. Media organization capabilities incorporated in the LionShare client, will give users the ability to share, play, and organize digital media from one application. The convenience of file sharing and media organization is further enhanced with the inclusion of communications tools.

Jabber. As mentioned earlier in this paper, Jabber is an open source IM protocol which features interoperability with a host of different protocols. There are already plans to incorporate a jabber server with the Penn State portal, so the inclusion of a jabber client with LionShare is a possibility. Jabber authentication via LDAP will occur during the standard LionShare authentication. Jabber userids will be in the familiar userid@psu.edu format. The inclusion of a jabber client with LionShare will allow for collaboration among users who are sharing files. Have quick question for a user sharing a photo? Want to collaborate with department members on a PowerPoint presentation? The capabilities of Jabber can extend LionShare beyond the realm of ordinary file sharing.

Besides the standard IM features that would be expected, Jabber has a lot of features not seen in ordinary IM clients. The Jabber protocol can be encrypted via SSL for secure communications making it a safe platform for private communication. Jabber has its own file transfer features that would be separate from the LionShare network. This feature will be convenient for one time user-to-user file transfer. User-to-User IM is just one feature of the Jabber protocol, Jabber also has the ability to be used as groupware for live chats with multiple users in private rooms. Need to keep record of online meetings and chats? Another feature that makes Jabber an enterprise strength IM client is the ability to log discussion. The power of file sharing combined with media organization and instant messaging can transform LionShare in to a one-stop media hub.

Summary

Peer to Peer technologies hold much promise for academic environments. Because most of these technologies are still in their infancy, organizations have been hesitant to adopt P2P strategies to traditionally client-server networks. In the wilds of the Internet, we have witnessed the migration towards 100% digital media, the mass adoption of instant messaging, and the exponential growth of file sharing. Organizations have been slow to accept these new technologies. The adoption of P2P in organizations has been met with skepticism of whether or not this once rouge technology can be tamed for enterprise environments. The promise of P2P is so great that institutions will eventually be forced to adopt it. With LionShare, Penn State has a chance to take an early lead in the race to implement the promise of P2P.

Michael Halm
6/2002
rev. 10/2003