Study and Bridging of Peer-to-Peer File Sharing Systems

Georgios Portokalidis [1], Evangelos P. Markatos[2] Manolis Marazakis
Institute of Computer Science, Foundation for Research and Technology Hellas
PO BOX 1385 , Heraklion, Crete, GR71110, Hellas, markatos@ics.forth.gr

Technical Report 312, October 2002

***Abstract*—Peer-to-peer file sharing systems have become increasingly popular over the last few years, by attracting large numbers of Internet users, who share a continuously increasing volume of data. Since the launch of Napster, the first widely known peer-to-peer file sharing network, several other file-sharing systems have emerged. These systems are based on various architectures, which usually reflect different sharing policies and communities. Although current peer-to-peer systems facilitate the sharing of content among their users, they do not facilitate the sharing of content among themselves. In this paper we study two of the most popular open-source peer-to-peer file sharing systems in order to understand (i) whether their traffic patterns are similar, and (ii) whether there would be any benefit in facilitating the sharing of content between the users of different peer-to-peer systems.**

Keywords*— *peer-to-peer,systems, file-sharing, replication, traces

I. INTRODUCTION

File sharing peer-to-peer systems have become incredibly popular the last few years, by attracting huge numbers of Internet users with their continuously increasing volume of shared data. Since the launch of Napster many other file-sharing systems have emerged, based on various architectures and reflecting different policies and communities. Examples of such systems are Gnutella, Direct Connect, Kazaa, E-Donkey, etc. The existence of such a variety of systems, serving the same purpose, justifiably raises some questions concerning whether they contribute to each other and present similarities or differences between them.

In our study we will examine the Gnutella and Direct Connect systems. Its purpose is to gather information about the type of data the users of these systems prefer to exchange. Furthermore we will evaluate the benefits gained by peers, in the case than the systems were bridged.

The contributions of this paper are:

- We developed a software tool that performs trace logging about queries and responses of Gnutella and Direct Connect
- We developed a bridge between different peer-to-peer systems that allows queries and responses to be crossed from one system to another
- We evaluate the bridging to demonstrate the benefits gained
- We show that forwarding the queries of one peer-to-peer network into the other, (i) increases the number of responses to each query and (ii) increases the availability of files that are common in both networks. Actually, our results suggest that forwarding queries that originated from Direct Connect's clients into the Gnutella network, increases the number of responses to each query by a factor of 2. Similarly forwarding Gnutella's queries into the DC network, increases the number of results by as much as 40%.

The rest of the paper is organized as follows: Section II covers related work on peer-to-peer systems. In Section III we present the architectures of Gnutella and Direct Connect, we also describe the methodology we used to gather traces and bridge them. Finally in Section IV we present the measurements we conducted on the systems and the conclusions we reached.

II. RELATED WORK

The increasing popularity of peer-to-peer systems has lead researchers to devote a lot of effort to study their behavior and the impact they impose in Internet networks. Stefan Saroiu et. al. [1] made a study of two of

[1] G. Portokalidis is also with Leiden University

[2] E.P. Markatos and M. Marazakis are also with the University of Crete.

the most popular peer-to-peer systems Napster and Gnutella. Their study though, was focused on comparing the characteristics of the peers participating in these systems, in perspective with the system's architecture.

Markatos [2] used traces gathered from Gnutella to study the traffic of queries and responses. The information collected, was analyzed to propose an effective caching policy for queries responses.

Anderson [3] observed the traffic of Gnutella for a 35-hour period and reported several results, including the distribution of TTL values, the distribution of Hops for Queries, the distribution of Hops for all Packets, etc.

Adar and Huberman [4] have also studied the traffic of Gnutella for a 24-hour period. Their findings indicated that almost 70% of users shared no files and that 50% of all the responses they gathered was returned by only 1% of the hosts. These results were also confirmed by and independent study by Stefan Saroiu [1].

Ripeanu *et al.*, studied the topology of the Gnutella network [5] over a period of several months, and found several interesting properties. Among them is that the Gnutella network topology does not match well the underlying Internet topology leading to the inefficient use of the network bandwidth.

Jovanovic [6] studied several Gnutella connection graphs and identified significant performance problems, including *short-circuiting*, an effect that limits the reachability of the nodes in a Gnutella network. Jovanovic's experiments suggest that, due to short-circuiting, a typical Gnutella peer reaches only about 50% of the peers that it could typically reach

Besides file sharing, peer-to-peer systems have also been used for the efficient execution of highly parallel and distributed applications by capitalizing on the availability of idle cycles in home computers. Such applications range from systems that search for extra-terrestrial intelligence [7], to systems that seek a cure for AIDS [8].

Previous work doesn't examine both the Gnutella and Direct Connect systems, and doesn't answer the question of whether there is a benefit to be gained by bridging two p2p systems. Though it has given us the guides needed to carry out this study. We have also gathered traces, but we have extracted alternate measurements from them and have focused in the comparison of the available data offered by the systems to its' peers.

III. METHODOLOGY

A. The Gnutella and Direct Connect Architectures

The Gnutella and Direct Connect systems both serve the same purpose, the location and exchange of data between their users. The data usually shared are music and video files, as well as computer programs. In these systems the users or peers, can share the files stored in their computer with other peers using the same system. Each peer functions as a client, as well as a server in the same time, allowing users to connect directly to each other and exchange files. Even though the exchange of files is done in the same way in both systems, the file location policy and the communication protocol differ (Figure 1).

In Gnutella each peer maintains connections with a set of neighbor peers creating an overlay network. If a peer wants to initiate a query, it sends a query packet to every neighbor. A peer receiving a query checks it's local shared files database to return any matching results to the query originator and then forwards it to its neighbors. To prevent a query packet from being forwarded forever, the Gnutella protocol defines a time to live (TTL) for each packet in the network. The TTL of each packet is an integer number that is decremented every time it reaches a peer. The packet is no longer forwarded when its TTL becomes zero.

To maintain the overlay network, the Gnutella protocol also defines ping and pong messages that are used for peers to discover new neighbors. The connections with neighbors that fail to respond can be replaced with the newly discovered peers.

Direct Connect on the other hand is more centralized. Peers connect to a server called a Hub. To locate a file, a query is made to the hub. The hub itself doesn't hold any index of the shared files available, but it forwards the query to all the connected peers. The results returned are then forwarded back to the initiator of the query. To allow users to locate files at peers not connected to the same hub, queries can also be forwarded to remote hubs using UDP. The results gathered from the remote hub are returned to the originator peer, again using UDP.

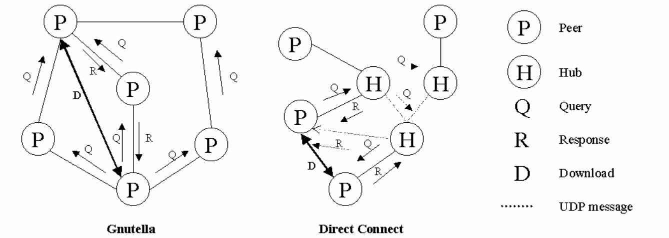

Figure 1. File location in Gnutella and Direct Connect

It is important to note, that Direct Connect hubs are maintained by users and not by a central authority unlike Napster. This has given the possibility to hub owners to create small communities (users sharing the same interests) by applying various rules to the connected peers. Such rules include restrictions on the number of concurrently connected peers to the hub, requirements on the minimum size of shared data from each peer, and even refusing to forward queries to other hubs. This has led to the creation of hubs specializing in distributing specific files like DivX movies, mp3 files or even iso images (files used by CD-burning software) of computer programs.

B. Tools Developed for this Study

To be able to study and eventually bridge the two systems, we had to develop a client application that would allow us to connect to both networks concurrently. To avoid development from the scratch, we chose to use open source clients, already available in the Internet and modify them to serve our purposes. The packages we selected were: Gnut ver. 0.4.28 (http://www.gnutelliums.com/linux_unix/gnut/) and DCTC ver. 0.72.0 (http://opendcd.sourceforge.net) terminal clients for Gnutella and Direct Connect respectively. Our choice was based in the fact that both packages were quite simple and were written in C.

1)Logging

We combined the packages mentioned above, to create a single client that implemented both the Gnutella and Direct Connect protocol. By using POSIX threads and altering the Gnut and DCTC main() functions we managed to run both clients concurrently as a single one. We left intact the protocol handling part of the clients. We intervened at the parts where query requests and responses were received and logged the information we required. The logging performed didn't impose any extra processing load, since the data were already provided by the protocol.

In this manner we collected traces of queries and responses from both systems. The information contained in these traces is described below.

The information logged about the incoming queries of Direct Connect, include besides the query string, the originator of the query and the timestamp at the moment the query is processed. The originator of a query in Direct Connect is specified by a nickname, that each peer provides to the Hub when connecting, or by the peer's IP if it is located on a remote hub. Furthermore the Direct Connect protocol allows users to specify the file type they 're interested in when making a query, thus we log this information too. In the Gnutella protocol the originator of an incoming query is the neighbor, from where we receive it.

Similar information is logged about the responses received for queries we generated. Again, we log the originator of the response, the timestamp, the filename contained in the response and the size of the file that the response references to.

2)Bridging

To bridge the two systems we had to propagate the incoming queries from one system to another and also forward the responses returned back to the originator system. In this manner a user using Gnutella gets responses from users using Direct Connect, and vice-versa. Of course a peer receiving such a response would perceive the bridging application as the originator of the response. To carry out such a task we had to extend the client we constructed to be able to exchange data between the two systems.

To forward queries from one system to another, we had to be able to support many concurrent active queries. The Gnutella protocol gives us this possibility by specifying a unique request id number for each query a peer initiates. Therefore we just had to extend the structures holding information for active queries, to include its originator in DC. The responses are matched by using their id.

In Direct Connect on the other hand the queries cannot be distinguished, so we had to create our own mechanisms to support multiple queries. We stored each query sent to DC in a list, along with all the necessary information to construct a response packet later for Gnutella. The responses received are compared with all the queries in the list. If a response contains all the keywords of a query string, it is sent to the query originator in Gnutella.

3)Trace Analysis

To analyze the traces we gathered in our study and produce some useful statistics, we also developed a set of scripts that use our traces as input. From the incoming queries traces, we needed to extract statistics concerning the file types the users of the two systems seek. To classify the queries we used the classes already provided by the Direct Connect protocol, which are: any, audio, compressed, document, binary, picture, video. Class *any* specifies that the query doesn't target a specific file type and all matching results should be returned.

Because Gnutella doesn't provide any classification on queries, we post-processed the traces scanning the query strings for well-known file extensions, such as used in the DCTC client for results classification. We also applied the same processing to Direct Connect traces on queries classified as *any* to get more accurate results.

The traces we gathered were used as input to our client for the systems to gather responses. Similarly to the incoming queries, we classified the responses we got based on file type. Results that didn't have a well-known extension were classified as *any*. Furthermore we extracted statistics concerning the number of responses collected, the size of the files referenced and the number of unique files found on the systems.

In the study of the results, we distinguished files based on filename. So we consider two files sharing the same filename are the same, despite any differences in size.

IV. EXPERIMENTAL DATA

Our measurements were collected at the Computer Science Department of the University of Crete during the period of March-June 2002. The workstation we used, was a personal computer on a 10MBit LAN running Linux.

A. Distribution of Queries According to File Type

With our first series of measurements on the two systems, we tried to gather information about the type of queries submitted by the users. As we mentioned earlier, the classification of the queries was based on the mechanism already provided by Direct Connect. The script we implemented enabled us to apply the same rules to Gnutella queries, and additional to DC queries that were classified as requesting *any* file type.

Table I describes the dates and duration of the query traces we collected from Gnutella, as well as the number of queries contained in them. Table II respectively, describes the traces collected from Direct Connect, along with the IP address of the hub we connected to. The difference between the two systems is obvious. The number of queries collected from Gnutella is 10 times or more, higher than the number of queries collected from Direct Connect in the same amount of time.

TABLE I. GNUTELLA TRACES

Date	Duration	Queries Collected
3/4/02	2 hours	189038
9/4/02	2 hours	124576
16/4/02	4 hours	348392
19/4/02	30 minutes	26435
18/5/02	2 days	1281921

1) Gnutella

Since Gnutella doesn't provide originating users, with a way to specify the file type they 're querying, the results presented here are based in scanning the query strings for well known file extensions that are usually given by users to narrow the number of returned results. As a consequence the percentage of queries classified as *any* is quite large in our measurements. In Figure 2 is shown the percentage of queries requesting each file type in traces taken at different dates and of variable duration.

As expected about 45% of the queries of each trace are classified as requesting *any* file type. It is noticeable that the distribution of the queries in all the traces we used is similar. This allows us to presume that even though almost half of the queries weren't classified, this distribution reflects what the users of Gnutella are searching for. It is noticeable that about 30% of the queries are targeted in video files, while another 18% in audio files. These two file types are dominating the user's preferences in Gnutella and it would be very interesting to compare this behavior with the one expressed by Direct Connect users in the following section.

2) Direct Connect

As discussed in Section III,A, Direct Connect peers are less likely to be able to search the entire network, since most hubs don't cooperate with each other. To get an overall appreciation of the Direct Connect network we connected to various hubs and calculated the distribution of the queries gathered from each one. Figure 3 shows the distribution of queries between file types, with the addition of the *folder* type defined only in DC to allow querying for whole directories.

The classification of the queries was better than in Gnutella, since it was already provided by the DC system. With the exception of one hub (isopub.no-ip.com) the queries classified as *any* were less than 30%.

TABLE II. DIRECT CONNECT TRACES

Hub IP	Date	Duration	Queries Collected
isopub.no-ip.com	2/4/02	2 hours	3439
194.108.145.2	19/4/02	2 hours	3843
Borg.i989.net	23/4/02	30 minutes	2686
billdal.no-ip.com	19/5/02	2 hours	3029
tralala.dns2go.com	25/5/02	4 hours	11681
tralala.dns2go.com	26/5/02	4 hours	25905
tralala.dns2go.com	27/5/02	1 hour	4900

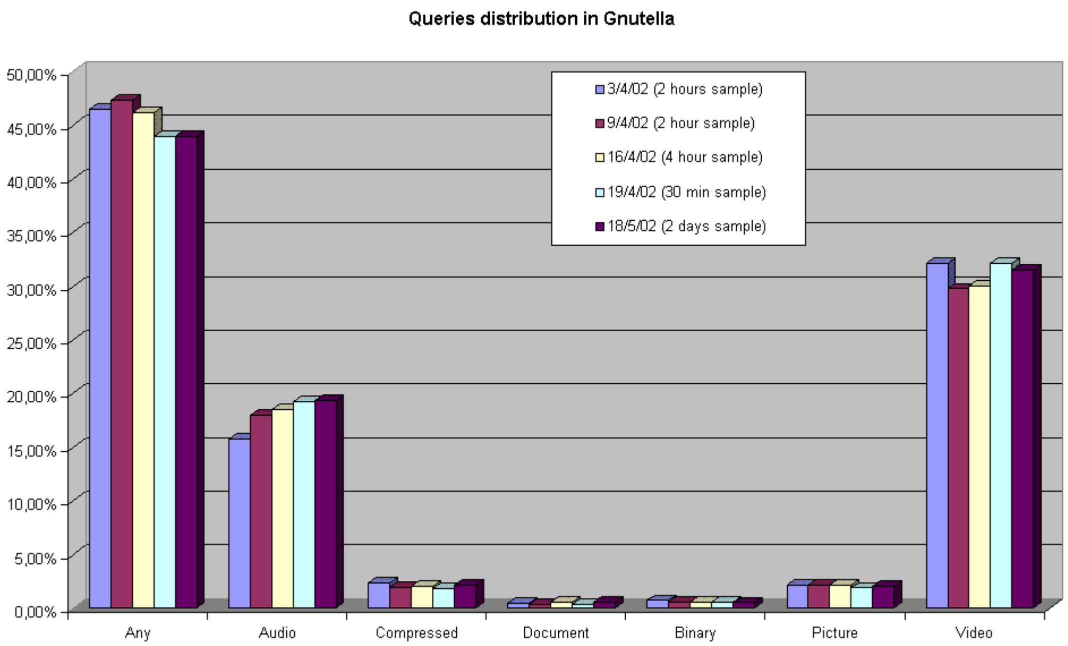

Figure 2. Queries Distribution in Gnutella

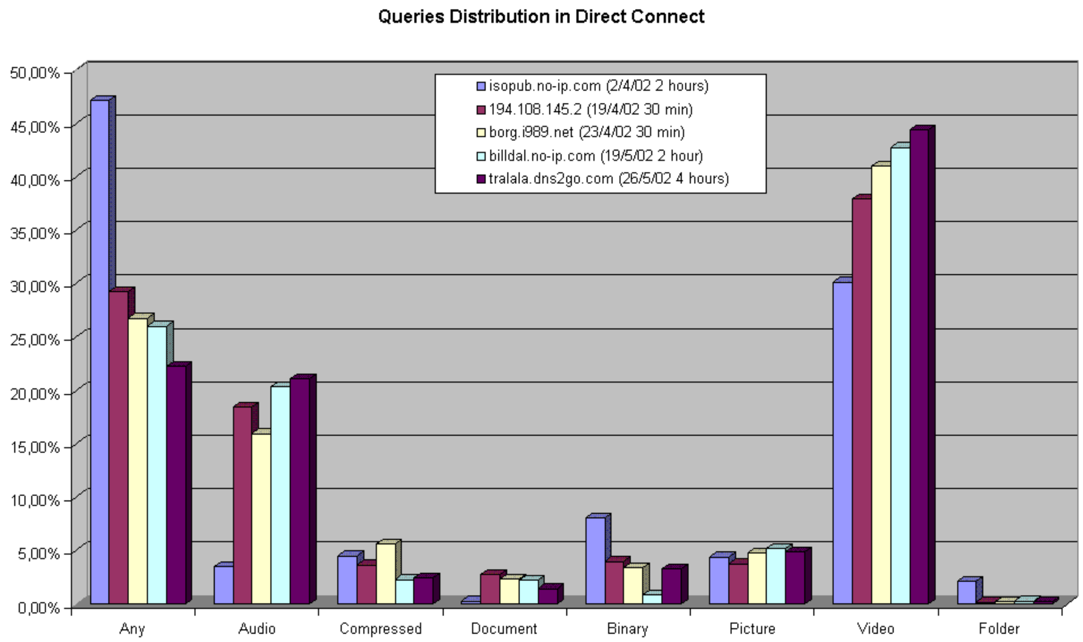

Figure 3. Queries Distribution in Direct Connect

The distribution in DC is quite similar with the one in Gnutella, with the video and audio file types getting the most queries, but as we notice there is an increase in the document and binary types, especially in the isopub.no-ip.com hub. By taking a closer look in this hub we concluded that it attracted users exchanging iso binary files. This very unusual specialization could be responsible for the hub's distribution.

Furthermore we collected more measurements from the tralala.dns2go.com hub, which provided us with the largest DC trace, to validate the correctness of our results. Figure 4 shows the distribution of the queries collected in three different dates from this hub. There is regularity between the traces, with the one of 25/5/02 having the largest variation. That can be justified by the percentage of the files classifies as *any*, which is higher than in the rest of the measurements. We believe that the small variation in our results indicates their validity.

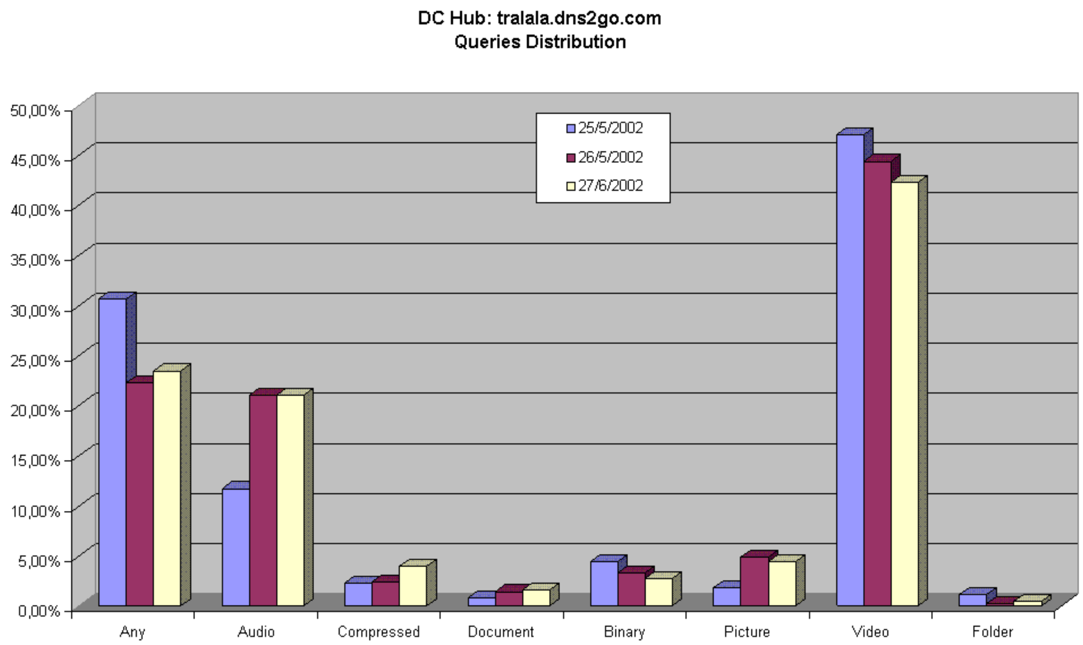

Figure 4. Queries Distribution in tralala.dn s2go.com Hub

B. *Responses Comparison*

To complete our view of the two systems, we generated queries and compared the responses gathered. We used the traces of 18/5/02 and 26/5/02 from Gnutella and Direct Connect respectively. These had the longest duration and included the most queries. The DC hub we chose to connect to was tralala.dns2go.com.

Due to the large number of queries, we chose the 500 most popular of each trace. Our decision was based on our intention for responses to represent the main volume of shared files in the systems. A 1-minute timeout was used between the submissions of queries in Direct Connect, while a larger timeout of 2 minutes, was used in Gnutella due to the greater latencies observed.

1) *Responses Using a Gnutella Trace*

The statistics in Table III were extracted from the responses we received from the two systems, when we propagated the 500 most popular queries of the 18/5/02 trace to them. As expected the results we gathered from Gnutella are three times more than Direct Connect, but in the contrary the size of the files returned is similar ($1.508x10^7$ and $1.049x10^7$ MBs respectively). We also calculated the number of unique files returned, based on the different filenames we collected. We calculated that only 116867 and 59343 of the responses received by Gnutella and DC respectively, referenced different files.

We also extracted the information that these unique files exist in more than one copy. This fact brings forth the replication behavior of peer-to-peer file sharing sy--stems that is accomplished by distributing a file to multiple peers. A measurement of this behavior is the number of copies per file that are present in the system.

By comparing the responses we gathered from the two systems, we saw that some of the files were present in both systems (described as *Common files* in Table III). However the number of these files is small relative to the number of unique files returned. Considering the number of common files, the bridging of Gnutella and Direct Connect would increase the number of responses that a peer receives and furthermore would increase the number of unique files returned to him. Specifically a Gnutella a peer would gain a 43.63% increase in unique files returned to him.

Finally, by calculating the percentage of common files over the number of unique files in each system, we see that there is a small overlap between the files shared by peers of Gnutella and Direct Connect. Just 2.79% of the files returned by Gnutella are also present in DC. This percentage rises to 14.07% when we examine the files returned by DC.

To explain the large volume of files returned by Direct Connect, we calculated the distribution among file types, for the responses and the volume they represented. Fig. 5 shows that almost 70% of Gnutella responses are audio files, while most of Direct Connect responses are video files. Furthermore by looking at Table IV we can see that video and binary files consti-

TABLE III. RESPONSES STATISTICS USING THE 18/5/02 GNUTELLA TRACE

	Gnutella	Direct Connect
Number of results	320910	125135
Queries with no results	22(4.40%)	2(0.40%)
Results per query	641.82	250.27
Results per query (not including queries with no results)	671.36	251.28
Size of results in Mb	1.508×10^7	1.049×10^7
Unique files	116867	59343
Unique copies	299064	76325
Copies per file	2.559	1.286
Common files	8350	
Copies of common files	82579	17188
Increase in unique files	43.63%	
Increase in results	38.99%	
Copies per file on both systems	2.24	
Overlapping percentage on unique files	2.79%	14.07%

-tute the main volume of files returned by DC, while audio files do that for Gnutella. This behavior of Direct Connect can be accounted to its community profile that attracts groups of users who wish to exchange illegal pirated software and movie files, which both imply large file sizes. The large percentage of responses classified as *any* in DC, may also seem surprising at first place. Though it can be explained, by taking in account the policy of many hubs to require a minimum size of shared data by their peers, forcing them in this way, to share all the files located in their hard disks, thus sharing a lot of useless operation system files.

TABLE IV. RESPONSES STATISTICS USING THE 18/5/02 DIRECT CONNECT TRACE

	Gnutella	Direct Connect
Any	1.573×10^5	9.414×10^4
Audio	3.951×10^6	3.195×10^5
Compressed	4.324×10^5	1.414×10^5
Document	4.150×10^5	1.760×10^4
Binary	3.273×10^4	1.581×10^6
Picture	1.050×10^6	3.405×10^4
Video	9.043×10^6	9.730×10^6

2)Responses Using a Direct Connect Trace

To validate the conclusions we have reached and to calculate the increase in results that DC peers gain, we propagated the 500 most popular queries, located in the trace from the DC hub tralala.dns2go.com at 26/5/02, to both networks. In Table V are shown the statistics we extracted by the collected responses. Using input queries originating from Direct Connect we notice that the number of the results returned, as well as the number of unique files is similar between the two systems, unlike we have seen in Table III. Despite that, the size of the returned results remains on the same level as before, for both systems.

The increase in unique files that DC gains is much larger than the one Gnutella had, reaching 92.60%.

We also notice that the drop in the number of returned results by Gnutella, affects the copies per file mean that we calculated, reducing it to 1.932. While in the same time the increase in the number of results in DC had negligible effect to it.

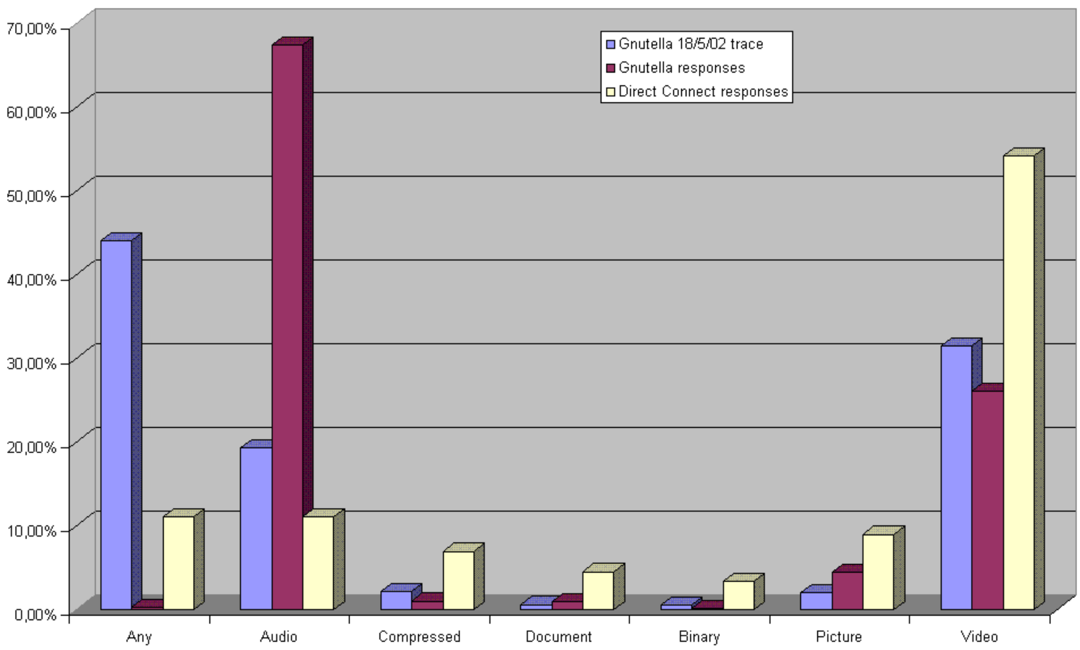

Figure 5. Responses Distribution Using the 18/5/02 Gnutella Trace

Figure 6 shows the distribution of the responses among file types. It is obvious that the distribution is similar to the one in Fig. 5 where we used a Gnutella trace to generate queries. The only variation we notice is that Gnutella returned more video files than DC. That is reflected in Table VI, where we can see that the volume of the video files returned is similar for both systems contrary to our previous measurement shown in Table IV.

TABLE V. RESPONSES STATISTICS USING THE 26/5/02 DIRECT CONNECT TRACE

	Gnutella	**Direct Connect**
Number of results	194963	171494
Queries with no results	108(21.73%)	41(8.25%)
Results per query	392.28	345.06
Results per query (not including queries with no results)	501.19	376.08
Size of results in Mb	1.215×10^7	1.49×10^7
Unique files	95466	95503
Unique copies	184515	123962
Copies per file	1.932	1.298
Common files	7030	
Copies of common files	35347	15585
Increase in unique files		92.60%
Increase in results		113.69%
Copies per file on both systems	1.68	
Overlapping percentage on unique files	3.81%	7.36%

It is important to notice that in this measurement as well, the overlap between the files of Gnutella and Direct Connect was very small. We also believe that the drop in the number of results of Gnutella is due to the nature of the queries used that fitted more to the files shared by Direct Connect peers.

3) Benefits from bridging

Table VII summarizes the benefits we have gained by bridging the two systems. We have accomplished to increase the number of unique files that are available to peers of the Gnutella and Direct Connect systems, as well as the number of copies of the files that were present in both systems.

The Gnutella system seems to benefit less than Direct Connect from this connection, since is consists of a larger number of peers. When we used a Gnutella trace as input, we managed to get just a 43.63% increase in unique files; nevertheless Direct Connect gained over 90% in unique files.

TABLE VI. RESPONSES STATISTICS USING THE 18/5/02 DIRECT CONNECT TRACE

	Gnutella	**Direct Connect**
Any	1.573×10^5	9.414×10^4
Audio	3.951×10^6	3.195×10^5
Compressed	4.324×10^5	1.414×10^5
Document	4.150×10^5	1.760×10^4
Binary	3.273×10^4	1.581×10^6
Picture	1.050×10^6	3.405×10^4
Video	9.043×10^6	9.730×10^6

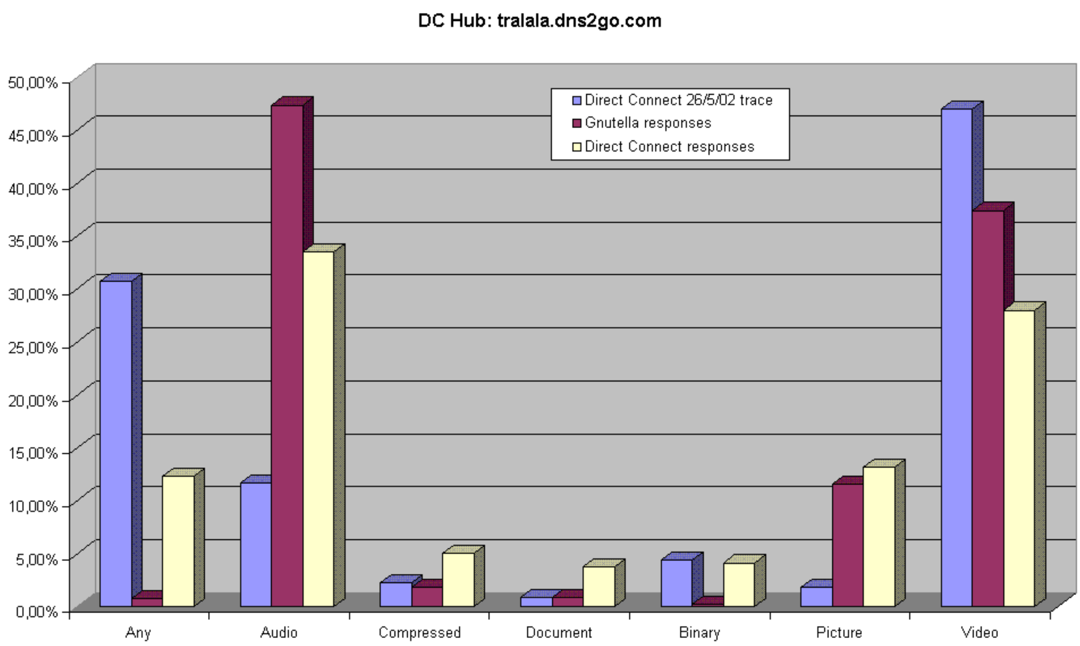

Figure 6. Responses Distribution Using the 26/5/02 DC Trace

TABLE VII. BENEFITS OF BRIDGING GNUTELLA AND DIRECT CONNECT

	INPUT TRACE	
	18/5/02 Gnutella trace	**26/5/02 Direct connect trace**
	Gnutella	***Direct Connect***
Increase in the number of unique files	43.63%	92.60%
Increase in the number of copies of common files	20.81%	226.80%

Furthermore we achieved an increase in the number of copies of the systems common files. This way we improve the replic ation of the popular files that are located in both systems. Again Gnutella gets benefited less than Direct Connect, since it offers most of the copies of common files as seen in Tables III and V.

4) Other Measurements

To further enforce our conclusions we collected responses using all the queries in a trace. To perform that we connected to the tralala.dns2go.com DC hub for 10 minutes and logged all the queries received. We collected 558 queries and we used them to gather responses from both systems in the same time. The responses gathered were less than our measurement when we used the 500 most popular queries of DC trace. Despite that, the behavior observed was similar to the one presented in Section IV,B,2 confirming our previous results.

In addition the paper of Markatos [2] intrigued us to check the flow of queries in the traces we had gathered. We calculated the mean queries per second for time intervals of 5, 30 and 60 seconds and confirmed that there are bursts in the traffic of queries in both systems.

C. Summary of Results

We obtained traces of queries propagating through Gnutella and Direct Connect. We studied the traces we gathered to discover the type of queries that the peers of these systems generate, and compared them to check for differences between the systems. We concluded that the patterns concerning the file types requested by peers are similar in both systems and are mainly concentrated in audio (approx. 20%) and video files (30%-40%).

To evaluate the advantages of a bridge between the two systems, we generated queries based on the traces we had already collected. By analyzing the results we collected we reached the conclusion that there are significant differences between the files shared by the peers of the systems we are examining. Furthermore by calculating the number of files present in both systems we notice that there is a high level of independence between them. According to our calculations the percentage of files that are present in both systems doesn't exceed 15% of its files.

Finally the small variations between our two measurements on responses, indicate that the queries made to Gnutella and Direct Connect, despite their similarities concerning the file types requested, do differ and they return better results to their originating peer.

V. CONCLUSION

We have performed a study on the Gnutella and Direct Connect peer-to-peer file sharing systems. We gathered traces to examine the behavior of their peers relatively to their file type preferences and concluded that they are similar.

We bridged the two systems and conducted measurements to conclude that there is indeed a high level of independence between them. We show that the peers of each system would benefit from this bridging, by gaining access to more files.

Our study can be further extended in the future to include other peer-to-peer file sharing systems as well. Another matter that is worthwhile to be looked into is to examine, if there are users that participate in more than one system concurrently.

REFERENCES

[1] Stefan Saroiu, P. Krishna Gummadi, Steven D. Gribble, "A measurement study of peer-to-peer file sharing systems," In *Proceedings of Multimedia Computing and Networking (MMCN) 2002*, 2002.

[2] Evangelos P. Markatos, "Tracing a large-scale peer to peer system: an hour in the life of Gnutella," 2nd IEEE/ACM Int. Symp. on Cluster Computing and the Grid 2002

[3] K. Anderson, "Analysis of the traffic on the Gnutella network", 2001. http://www-cse.ucsd.edu/classes/wi01/cse222/projects/reports/p2p-2.pdf

[4] E. Adar and B. Huberman, "Free riding on Gnutella," First Monday 5(10) 2000

[5] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. IEEE Internet Computing Journal, 6(1), 2002.

[6] M. Jovanovic, "Modeling lareg-scale peer-to-peer networks and a case study of gnutella," Master's thesis, University of Cincinnati, 2001.

[7] E. Korpela, D. Werthimer, D. Anderson, J. Cobb, and M. Lebofsky. Seti@home-massively distributed computing for seti. Computing in Science & Enginering, 3(1):78-83, 2001.

[8] FightAIDS@home. http://www.fightaidsathome.org/.