

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN CUỐI KÌ
MÔN KIẾN TRÚC INTERNET

MẠNG GNUTELLA

Người hướng dẫn: **ThS TRẦN TRUNG TÍN**

Người thực hiện: **ĐINH PHƯƠNG MY – 52100703**

Lớp: **21050401**

Khoá: **25**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG
KHOA CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN CUỐI KÌ
MÔN KIẾN TRÚC INTERNET

MẠNG GNUTELLA

Người hướng dẫn: **ThS TRẦN TRUNG TÍN**

Người thực hiện: **ĐINH PHƯƠNG MY – 52100703**

Lớp: **21050401**

Khoá: **25**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024

LỜI CẢM ƠN

Lời nói đầu tiên, em xin được gửi lời cảm ơn chân thành đến toàn bộ giảng viên Trường Đại học Tôn Đức Thắng nói chung cũng như toàn bộ giảng viên Khoa Công nghệ thông tin nói riêng vì đã tạo điều kiện cho em được học bộ môn Kiến trúc Internet.

Và đặc biệt, em xin gửi lời cảm ơn chân thành nhất đến thầy Trần Trung Tín – giảng viên giảng dạy và hướng dẫn cho đề tài môn Kiến trúc Internet. Trong suốt quá trình học tập và thực hiện bài báo cáo, thầy luôn giúp đỡ, chỉ bảo tận tình để nhóm có thể tìm ra cách giải quyết những vướng mắc gặp phải và hoàn thiện đề tài này một cách tốt nhất. Lượng kiến thức này em sẽ làm hành trang để áp dụng vào công việc sau này.

Không thể nói gì hơn nữa, một lần nữa, bằng cả tấm lòng, em xin chân thành gửi lời cảm ơn sâu sắc đến thầy - người đã dìu dắt lớp trong suốt chặng đường vừa qua!

ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Em xin cam đoan đây là sản phẩm đồ án của riêng bản thân và được sự hướng dẫn của ThS Trần Trung Tín. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đồ án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

Nếu phát hiện có bất kỳ sự gian lận nào nhóm xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình. Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do nhóm gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày 14 tháng 05 năm 2024

Tác giả

(ký tên và ghi rõ họ tên)



Đinh Phương My

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

Phần xác nhận của GV hướng dẫn

Tp. Hồ Chí Minh, ngày tháng năm
(ký và ghi họ tên)

Phần đánh giá của GV chấm bài

Tp. Hồ Chí Minh, ngày tháng năm
(ký và ghi họ tên)

TÓM TẮT

Báo cáo này tập trung vào việc nghiên cứu và phân tích về mạng Gnutella, một mạng ngang hàng phổ biến được sử dụng để chia sẻ tệp và tài nguyên trực tuyến.

Chương 1 cung cấp một tóm tắt và giới thiệu về nội dung của báo cáo, cùng với mục tiêu của nghiên cứu về mạng Gnutella. Chương này cũng giới thiệu về khái niệm của kiến trúc ngang hàng và vai trò của mạng Gnutella trong hệ thống này.

Chương 2 nhằm phân tích mục tiêu của việc thiết kế giao thức Gnutella và cách mà nó hoạt động. Mô tả cụ thể về cách mà giao thức Gnutella cho phép các nút trong mạng tương tác với nhau và chia sẻ tài nguyên.

Chương 3 tập trung vào quá trình thu thập dữ liệu, giới thiệu về trình thu thập thông tin được sử dụng để nghiên cứu mạng Gnutella bao gồm mô tả quy trình và phương pháp thu thập dữ liệu.

Chương 4 nhằm đánh giá và phân tích các thuộc tính và đặc điểm của mạng Gnutella dựa trên dữ liệu thu thập được. Nêu rõ những kết luận và nhận định quan trọng về cấu trúc và hoạt động của mạng Gnutella.

Chương 5 giới thiệu về ứng dụng Gtk-Gnutella và hướng dẫn cài đặt cách sử dụng cách tải và chia sẻ tập tin trên ứng dụng này.

Cuối cùng, tổng kết lại những điểm chính và kết quả quan trọng từ nghiên cứu và đưa ra những kết luận cuối cùng và đề xuất hướng phát triển hoặc nghiên cứu tiếp theo trong lĩnh vực này.

MỤC LỤC

LỜI CẢM ƠN	i
PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN	iii
TÓM TẮT	iv
DANH MỤC CHỮ VIẾT TẮT	3
DANH CHÚC NĂNG CÁC BẢNG BIỂU, HÌNH ẢNH	4
CHƯƠNG 1 – TỔNG QUAN.....	5
1.1 Đối tượng nghiên cứu	5
1.2 Mục tiêu nghiên cứu	5
1.3 Phương pháp nghiên cứu.....	5
1.4 Quy trình nghiên cứu	5
CHƯƠNG 2 – KIẾN TRÚC MẠNG NGANG HÀNG	6
2.1 Định nghĩa	6
2.2 Mục đích và vai trò	7
2.3 Định hướng nghiên cứu.....	7
CHƯƠNG 2 – GIAO THỨC GNUTELLA	9
2.1 Lịch sử Gnutella.....	9
2.1.1 Khởi nguồn và phát triển	9
2.1.2 Nguồn gốc tên gọi	10
2.2 Định nghĩa	11
2.3 Mục tiêu.....	11
2.4 Cách thức hoạt động	12
2.5 Các loại thông điệp	13
CHƯƠNG 3 – TRÌNH THU THẬP DỮ LIỆU	14
3.1 Giới thiệu chung	14
3.2 Mô tả quy trình	14
3.3 Phương pháp thu thập dữ liệu.....	14
CHƯƠNG 4 – PHÂN TÍCH MẠNG GNUTELLA.....	15
4.1 Xu hướng và hành vi tăng trưởng.....	15
4.2 Ước tính lưu lượng truy cập	17

4.3 Khả năng kết nối và độ tin cậy	19
4.4 Cơ sở hạ tầng internet và mạng Gnutella	20
CHƯƠNG 5 – ỨNG DỤNG GTK - GNUTELLA	24
5.1 Tải và cài đặt ứng dụng	24
5.2 Tìm kiếm và tải tập tin	25
5.3 Chia sẻ tập tin	27
5.4 Đánh giá	28
CHƯƠNG 6 – KẾT LUẬN	29
6.1 Kết quả đạt được	29
6.2 Hạn chế.....	29
TÀI LIỆU THAM KHẢO	31

DANH MỤC CHỮ VIẾT TẮT

CÁC CHỮ VIẾT TẮT

AOL	America Online
TLL	Time to Live
P2P	Peer to Peer

DANH CHỨC NĂNG CÁC BẢNG BIỂU, HÌNH ẢNH

DANH CHỨC NĂNG BẢNG

Bảng 4.1.1 Lưu lượng truy cập trong tháng 11/2000 và tháng 6/2001	17
---	----

DANH CHỨC NĂNG HÌNH ẢNH

Hình 2.1 Sự khác biệt giữa mạng dựa trên máy chủ và mạng ngang hàng.....	6
Hình 2.1.1 Justin Frankel và Tom Pepper	9
Hình 2.4 Cấu trúc liên kết của Gnutella 0.4.	12
Hình 4.1.1 Tốc độ tăng trưởng của mạng Gnutella.....	15
Hình 4.1.2 Tần suất thông điệp được gửi đi trong mạng	16
Hình 4.2.1. Phân phối đường đi ngắn nhất giữa các nút.	17
Hình 4.2.2. Số lượng kết nối trung bình.	18
Hình 4.3.1 Phân phối kết nối trong tháng 11/2000	19
Hình 4.3.2 Phân phối kết nối trong tháng 3/2001	19
Hình 4.4 Hai ánh xạ khác nhau của cấu trúc liên kết với cơ sở hạ tầng	20
Hình 5.1 Giao diện chính của ứng dụng GTK – Gnutella.....	24
Hình 5.2.1 Danh sách thông tin khi tìm kiếm và thông tin về server được chọn ...	25
Hình 5.2.2 Tiến độ tải của các bài hát đã chọn.....	26
Hình 5.2.3 Các tập tin đã được tải xuống thành công	26
Hình 5.3.1 Giao diện chọn thư mục	27

CHƯƠNG 1 – TỔNG QUAN

1.1 Đối tượng nghiên cứu

Báo cáo này tập trung vào việc nghiên cứu về hành vi của hệ thống mạng ngang hàng (P2P), với một trọng tâm đặc biệt vào mô hình P2P Gnutella. Đối tượng nghiên cứu là cấu trúc và hoạt động của mạng Gnutella và cách mà nó ảnh hưởng đến hiệu suất và khả năng mở rộng của hệ thống.

1.2 Mục tiêu nghiên cứu

Mục tiêu của nghiên cứu là phân tích cấu trúc liên kết của mạng ảo Gnutella và đánh giá lưu lượng mạng được tạo ra. Nghiên cứu cũng nhằm xác định các đặc điểm và hạn chế của mạng Gnutella và đề xuất các cải tiến có thể tăng hiệu suất và khả năng mở rộng của hệ thống.

1.3 Phương pháp nghiên cứu

Phương pháp nghiên cứu bao gồm việc sử dụng công nghệ phân tích và thu thập dữ liệu mạng để thúc đẩy sâu hơn trong quy mô và độ phức tạp của nghiên cứu. Phân tích cấu trúc và hoạt động của mạng Gnutella được thực hiện để hiểu rõ hơn về hành vi của hệ thống P2P.

1.4 Quy trình nghiên cứu

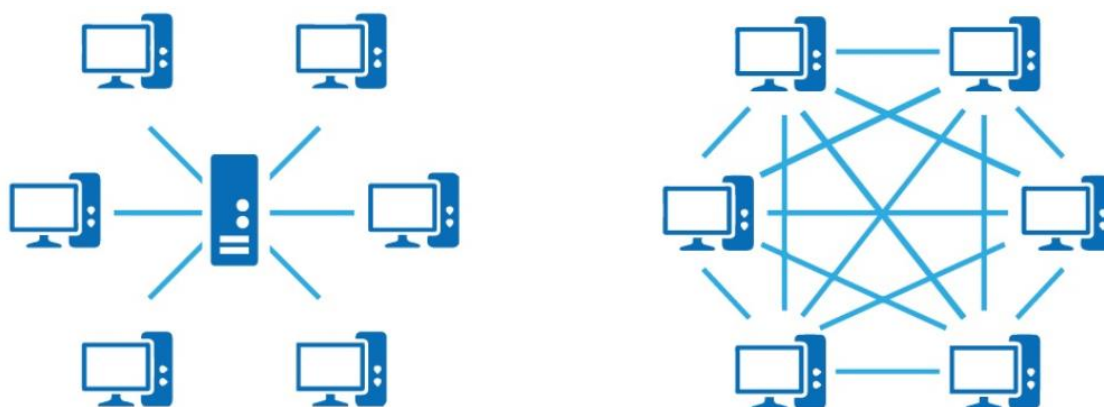
Quy trình nghiên cứu bao gồm các bước sau:

- Tiến hành thu thập dữ liệu về mạng Gnutella.
- Phân tích cấu trúc và hoạt động của mạng.
- Đánh giá hiệu suất và khả năng mở rộng của hệ thống.
- Xác định các hạn chế và đề xuất các cải tiến.
- Tổng hợp kết quả và đưa ra kết luận.

CHƯƠNG 2 – KIẾN TRÚC MẠNG NGANG HÀNG

2.1 Định nghĩa

Hệ thống ngang hàng (P2P) là một kiến trúc mạng nơi các thiết bị máy tính kết nối trực tiếp với nhau để chia sẻ dữ liệu và tài nguyên mà không cần sự điều phối của một máy chủ trung tâm. Mỗi thiết bị trong mạng P2P có thể hoạt động như một client và server đồng thời. Trong mô hình này, mỗi nút trong mạng có thể tương tác và chia sẻ tài nguyên với nhau mà không cần thông qua một bên thứ ba.



Hình 2.1 Sự khác biệt giữa mạng dựa trên máy chủ và mạng ngang hàng

Hệ thống P2P mang đến nhiều lợi ích vượt trội so với mô hình mạng truyền thống, bao gồm:

- Hiệu quả về chi phí: Giảm thiểu chi phí quản lý và duy trì cơ sở hạ tầng do phân tán gánh nặng trên nhiều thiết bị.
- Tính sẵn sàng cao của tài nguyên: Tăng cường khả năng truy cập tài nguyên do sự dư thừa từ nhiều nguồn cung cấp khác nhau.
- Khả năng mở rộng mạnh mẽ: Dễ dàng mở rộng quy mô mạng bằng cách kết nối thêm các thiết bị mới mà không cần nâng cấp phần cứng hay phần mềm tập trung.
- Tính linh hoạt và chống lỗi cao: Hệ thống vẫn có thể hoạt động hiệu quả ngay cả khi một số nút bị lỗi hoặc gián đoạn kết nối.

Tuy nhiên, hệ thống P2P cũng tiềm ẩn một số nhược điểm cần được lưu ý:

- Sự phụ thuộc vào kết nối mạng: Hiệu suất có thể bị ảnh hưởng bởi chất lượng kết nối Internet của người dùng.

- Vấn đề an ninh: Nguy cơ cao về lỗ hổng bảo mật và truy cập trái phép do tính chất phi tập trung của hệ thống.
- Khó khăn trong quản lý: Việc quản lý và điều phối hoạt động của mạng trở nên phức tạp hơn do thiếu sự kiểm soát tập trung.

2.2 Mục đích và vai trò

Hệ thống P2P được ứng dụng rộng rãi trong nhiều lĩnh vực khác nhau, với các mục đích chính sau:

- Chia sẻ tài nguyên: Cung cấp nền tảng để chia sẻ tập tin, dữ liệu, chu trình CPU, không gian lưu trữ và các tài nguyên khác một cách hiệu quả.
- Hỗ trợ cộng tác: Tạo điều kiện cho cộng đồng cùng nhau làm việc trên các dự án chung, trao đổi thông tin và ý tưởng.
- Phát triển ứng dụng phi tập trung: Hỗ trợ xây dựng các ứng dụng phi tập trung (DApps) hoạt động trên nền tảng blockchain, mang lại tính minh bạch, tự chủ và an toàn cho người dùng.

Hệ thống P2P đóng vai trò quan trọng trong việc thúc đẩy sự phát triển của công nghệ hiện đại, góp phần tạo ra một môi trường mạng mở, tự do và dân chủ hơn.

2.3 Định hướng nghiên cứu

Báo cáo tập trung nghiên cứu chi tiết về cấu trúc liên kết và các giao thức hoạt động của mạng P2P phổ biến Gnutella. Nghiên cứu nhằm giải đáp hai câu hỏi chính:

- Cấu trúc liên kết của mạng Gnutella như thế nào?
- Mức độ tương thích giữa cấu trúc liên kết mạng ảo Gnutella và cơ sở hạ tầng Internet vật lý ra sao?

Kết quả nghiên cứu cho thấy mạng Gnutella sở hữu cấu trúc liên kết phi tập trung với khả năng chịu lỗi cao, ít phụ thuộc vào các nút trung tâm. Tuy nhiên, cấu trúc mạng ảo Gnutella không hoàn toàn tương thích với cơ sở hạ tầng Internet vật lý, dẫn đến áp lực lớn lên các Nhà cung cấp dịch vụ Internet (ISP).

Hệ thống P2P mang lại nhiều lợi ích và đóng vai trò quan trọng trong công nghệ hiện đại. Tuy nhiên, cần lưu ý đến những nhược điểm và giải quyết các vấn đề liên quan để đảm bảo hiệu quả và an ninh cho hệ thống. Nghiên cứu về các hệ thống

P2P như Gnutella có ý nghĩa quan trọng trong việc cải thiện hiệu suất, khả năng mở rộng và tính bảo mật cho mạng ngang hàng.

Định hướng nghiên cứu trong tương lai:

- Phát triển các thuật toán tối ưu hóa hiệu suất tìm kiếm và chia sẻ tài nguyên trong mạng P2P.
- Nghiên cứu các giải pháp nâng cao tính bảo mật và chống tấn công cho hệ thống P2P.
- Khám phá tiềm năng ứng dụng của hệ thống P2P trong các lĩnh vực mới như Internet vạn vật (IoT), trí tuệ nhân tạo (AI) và chuỗi khối (blockchain).

CHƯƠNG 2 – GIAO THỨC GNUTELLA

2.1 Lịch sử Gnutella

2.1.1 Khởi nguồn và phát triển

Gnutella, một trong những mạng ngang hàng đầu tiên, đã hình thành từ sự phát triển của khách hàng đầu tiên, được phát triển bởi Justin Frankel và Tom Pepper của Nullsoft vào đầu năm 2000, sau khi công ty được AOL mua lại. Gnutella đã nhanh chóng thu hút sự chú ý của hàng ngàn người sau khi được công bố trên Slashdot và có sẵn để tải xuống từ máy chủ của Nullsoft.



Hình 2.1.1 Justin Frankel và Tom Pepper

Mặc dù sau đó AOL đã ngừng cung cấp Gnutella do lo ngại pháp lý, nhưng điều này không ngăn cản sự phát triển của giao thức. Các bản sao mã nguồn mở và miễn phí của Gnutella bắt đầu xuất hiện sau đó. Sự phát triển song song của các khách hàng khác nhau bởi các nhóm khác nhau đã trở thành phương thức hoạt động của sự phát triển Gnutella.

Trong số những người tiên phong của Gnutella độc lập đầu tiên có Gene Kan và Spencer Kimball, người đã ra mắt cổng thông tin đầu tiên và phát triển chương trình mã nguồn mở “GNUbile”. Mạng Gnutella nhanh chóng trở thành một giải pháp thay thế được phân phối đầy đủ cho các hệ thống bán tập trung như FastTrack và Napster.

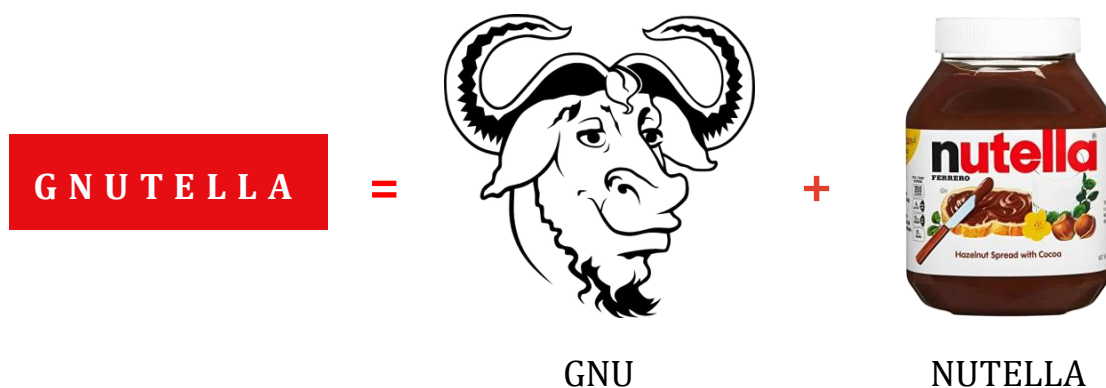
Mạng Gnutella đã trải qua các giai đoạn phát triển và thay đổi trong cách hoạt động và cấu trúc của nó. Các biến thể mới trên giao thức đã được phát triển để cải thiện khả năng mở rộng và hiệu suất của mạng. Mặc dù đã có những thách thức và biến động, Gnutella vẫn là một phần quan trọng của cộng đồng mạng ngang hàng và vẫn được sử dụng rộng rãi dưới dạng giao thức mở bởi các khách hàng khác nhau.

2.1.2 Nguồn gốc tên gọi

Tên “Gnutella” là một sự kết hợp độc đáo giữa hai từ: “GNU” và “Nutella”, mang lại một cái nhìn thú vị và dễ nhớ cho giao thức chia sẻ tập tin phân tán này.

- “GNU” là viết tắt của “GNU's Not Unix”, một dự án phần mềm mã nguồn mở do Richard Stallman khởi xướng vào những năm 1980. Tên “Gnutella” không liên quan trực tiếp đến dự án GNU hoặc mạng ngang hàng của GNU, GNUnet. Tuy nhiên, “GNU” cũng là tên của một trong hai loài linh dương châu Phi lớn, còn được gọi là linh dương đầu bò, có bờm và râu rủ xuống, đuôi búi dài và sừng cong ở cả hai giới tính. Linh dương GNU thường được tìm thấy ở các vùng đồng cỏ mở và thảo nguyên ở châu Phi.
- “Nutella” là tên thương hiệu của một loại phết hương vị hạt phỉ của Ý.

Sự kết hợp giữa từ “GNU” và “Nutella” tạo ra một cái nhìn độc đáo, vừa gợi nhớ đến phần mềm mã nguồn mở vừa mang lại cảm giác ngọt ngào và hấp dẫn của Nutella. Cách chơi chữ này thể hiện sự kết hợp giữa các phần mềm mã nguồn mở và sự ngọt ngào của Nutella, tạo ra một tên thú vị và dễ nhớ cho giao thức chia sẻ tập tin phân tán này.



Hình 2.1.1 Sự khác biệt giữa mạng dựa trên máy chủ và mạng ngang hàng

2.2 Định nghĩa

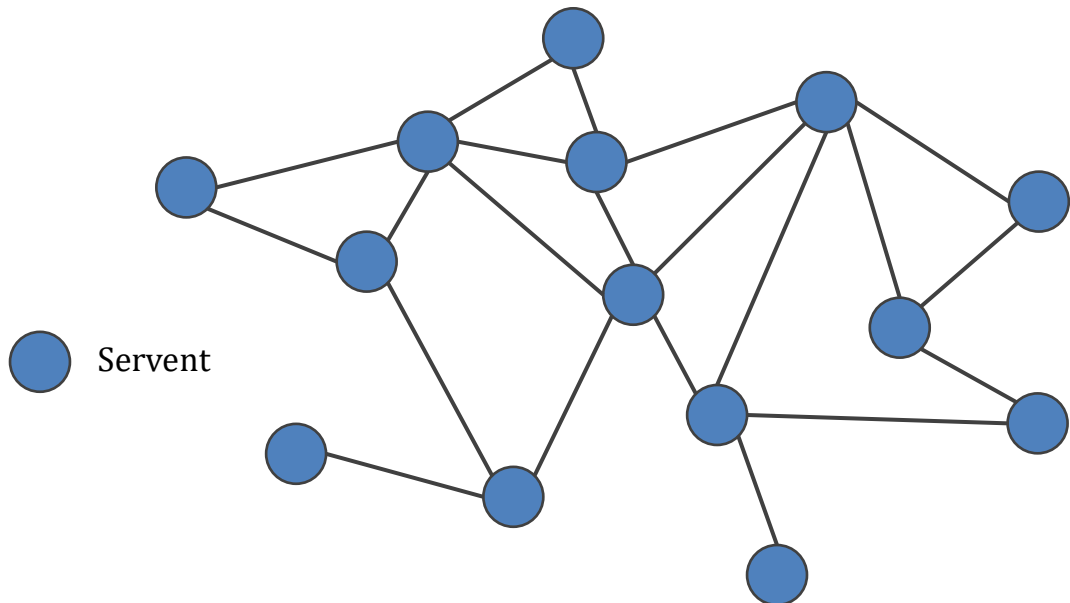
Gnutella là một giao thức mở, phi tập trung được sử dụng chủ yếu để tìm kiếm và chia sẻ tệp trên Internet. Thuật ngữ “Gnutella” có thể tham chiếu đến cả giao thức và mạng ngang hàng được tạo ra bởi các máy chủ chạy các ứng dụng Gnutella. Mạng Gnutella này bao gồm một số máy chủ và nút người dùng kết nối với nhau để chia sẻ tệp thông qua giao thức Gnutella. Nó cũng có thể bao gồm các mạng nhỏ hơn, thường là riêng tư và không kết nối với Internet, nhưng vẫn sử dụng giao thức Gnutella để tương tác nội bộ.

2.3 Mục tiêu

- Khả năng hoạt động trong môi trường năng động: Mạng Gnutella cần có khả năng linh hoạt để xử lý sự tham gia và rời đi của các máy chủ và người dùng một cách thường xuyên mà không làm gián đoạn hoạt động của mạng. Điều này đảm bảo rằng người dùng có thể kết nối và chia sẻ tệp một cách liên tục mà không gặp phải sự cản trở do sự thay đổi trong cấu trúc của mạng.
- Hiệu suất và khả năng mở rộng: Gnutella phải có khả năng mở rộng để có thể phục vụ một số lượng lớn người dùng cùng một lúc mà vẫn duy trì hiệu suất cao. Điều này đặc biệt quan trọng khi số lượng người dùng và lưu lượng truy cập vào mạng tăng lên. Mạng cần có khả năng mở rộng linh hoạt để đáp ứng nhu cầu ngày càng tăng của người dùng mà không làm giảm hiệu suất hoạt động của nó.
- Độ tin cậy: Mạng Gnutella được thiết kế để chịu được các cuộc tấn công từ bên ngoài mà không gây ra mất dữ liệu đáng kể hoặc suy giảm hiệu suất của mạng. Điều này đảm bảo rằng mạng vẫn hoạt động ổn định và đáng tin cậy trong mọi tình huống.
- Ẩn danh: Ẩn danh là một tính năng quan trọng trong mạng Gnutella, được coi là một phương tiện bảo vệ quyền riêng tư của người dùng. Điều này cho phép người tìm kiếm hoặc chia sẻ thông tin không phổ biến mà không cần tiết lộ danh tính của họ, tạo điều kiện thuận lợi cho sự tự do và quyền riêng tư của người dùng.

2.4 Cách thức hoạt động

Các nút Glutella, thường được gọi là servents (kết hợp giữa “serv” trong servers và “ents” trong clients) vì nó thực hiện các tác vụ được liên kết với cả máy chủ (servent) lẫn máy khách (client). Các nút này cung cấp một giao diện cho máy khách – nơi mà người dùng có thể đưa ra các truy vấn và xem kết quả tìm kiếm, chấp nhận các truy vấn từ các servents khác, kiểm tra các kết quả phù hợp với tập dữ liệu của họ và phản hồi lại kết quả tương ứng. Ngoài ra các nút Gnutella còn chịu trách nhiệm quản lý lưu lượng nền nhằm lan truyền các thông tin được sử dụng để duy trì tính toàn vẹn của mạng.



Hình 2.4 Cấu trúc liên kết của Gnutella 0.4.

Để tham gia vào hệ thống, một servent mới ban đầu sẽ kết nối với một trong số ít máy chủ được biết đến mà thường luôn sẵn có. Một servent mới (phía client) sẽ kết nối vào mạng bằng cách liên lạc với một hoặc nhiều máy chủ Gnutella (phía server) đã biết đến (ví dụ: gnutellahosts.com).

Sau khi kết nối vào mạng (có thể thông qua một hoặc nhiều kết nối mở với các nút đã tồn tại trong mạng), các nút sẽ gửi tin nhắn tương tác với nhau để thông báo sự hiện diện của mình trên mạng. Tin nhắn có thể được phát đi (tức là gửi đến tất cả các nút mà người gửi có kết nối TCP mở) hoặc đơn giản là được truyền ngược (tức là gửi trên một kết nối cụ thể trong một đường dẫn đã được thực hiện bởi một tin nhắn ban đầu, được phát đi).

2.5 Các loại thông điệp

Trong mạng Gnutella, các loại thông điệp chính được sử dụng để quản lý và truyền thông giữa các nút trong mạng bao gồm:

Thông điệp thành viên nhóm (PING và PONG): Khi một nút mới tham gia vào mạng, nó bắt đầu bằng việc gửi một tin nhắn PING để thông báo sự hiện diện của nó. Các nút khác trong mạng nhận được tin nhắn PING sẽ trả lời bằng tin nhắn PONG, đồng thời chuyển tiếp tin nhắn PING đến các nút khác. Thông điệp PONG chứa thông tin về nút như địa chỉ IP và thông tin về các tệp được chia sẻ.

Thông điệp Tìm kiếm (QUERY và QUERY RESPONSE): Khi một người dùng tìm kiếm một tệp cụ thể, một thông điệp QUERY chứa chuỗi tìm kiếm sẽ được phát đi trong mạng. Mỗi nút trong mạng sẽ kiểm tra xem liệu tên tệp này có khớp với các tệp được lưu trữ cục bộ hay không. Nếu có, nó sẽ gửi lại một tin nhắn QUERY RESPONSE chứa thông tin về tệp, để người dùng có thể tải xuống.

Thông điệp Truyền tệp (GET và PUSH): Các tin nhắn GET và PUSH được sử dụng để truyền tải tệp trực tiếp giữa các nút trong mạng. Khi một người dùng muốn tải xuống một tệp từ một nút khác, nó sẽ gửi một tin nhắn GET. Nếu nút đích đồng ý chia sẻ tệp, nó sẽ trả lời bằng một tin nhắn PUSH để bắt đầu quá trình truyền tệp.

CHƯƠNG 3 – TRÌNH THU THẬP DỮ LIỆU

3.1 Giới thiệu chung

Trình thu thập thông tin trong nghiên cứu về mạng Gnutella là một công cụ được phát triển để tự động tham gia vào mạng và thu thập dữ liệu về cấu trúc mạng. Nó được xây dựng dưới dạng một server, sử dụng giao thức thành viên (PING-PONG) để tương tác với các nút trong mạng và lấy thông tin từ chúng.

3.2 Mô tả quy trình

Quy trình thu thập thông tin bắt đầu bằng việc xác định danh sách các nút trong mạng. Trình thu thập sau đó khởi tạo kết nối TCP đến từng nút trong danh sách này và gửi thông báo PING để yêu cầu thông tin về các nút lân cận. Dựa trên các phản hồi PONG từ các nút lân cận, trình thu thập thông tin xác định và lưu trữ thông tin về mỗi nút, bao gồm địa chỉ IP, cổng, số lượng tệp và dung lượng được chia sẻ.

3.3 Phương pháp thu thập dữ liệu

Tiến hành thu thập dữ liệu bằng cách sử dụng hai phương pháp chính:

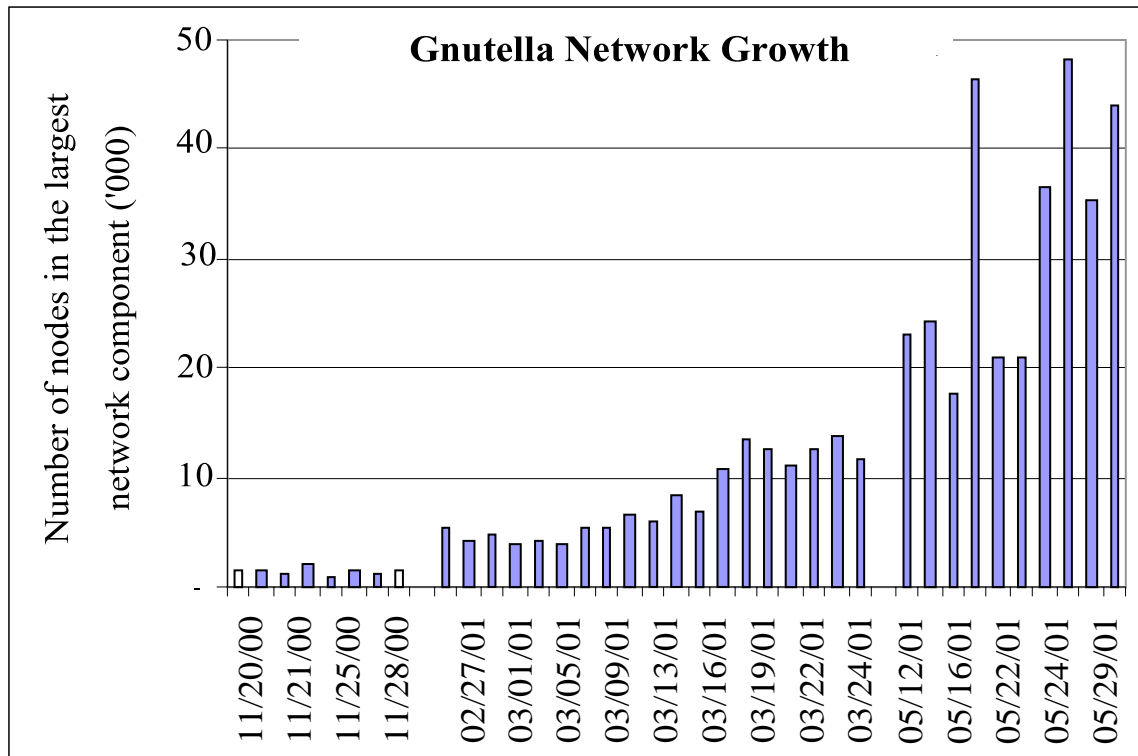
- Thu thập tuần tự: Ban đầu, chúng tôi đã phát triển một phiên bản tuần tự của trình thu thập thông tin. Tuy nhiên, phương pháp này đã cho thấy hiệu suất chậm do thời gian chờ kết nối và thời gian chờ phản hồi từ các nút. Điều này dẫn đến thời gian thu thập dữ liệu dài đối với một lượng nhỏ mạng.
- Thu thập phân tán: Để giảm thời gian thu thập dữ liệu, chúng tôi đã phát triển một chiến lược thu thập thông tin phân tán. Trong phương pháp này, máy chủ quản lý danh sách các nút cần liên hệ và phân công công việc cho các máy khách. Mỗi máy khách nhận một danh sách nhỏ các điểm ban đầu và khám phá cấu trúc mạng xung quanh các điểm này. Kỹ thuật này đã giúp giảm thời gian thu thập dữ liệu và tăng hiệu suất của quá trình.

Qua các phương pháp thu thập dữ liệu này, chúng tôi đã xây dựng một tập dữ liệu lớn với hơn 400.000 nút trong mạng Gnutella, giúp chúng tôi hiểu rõ hơn về cấu trúc và hoạt động của mạng này.

CHƯƠNG 4 – PHÂN TÍCH MẠNG GNUTELLA

Phân tích mạng gnutella dựa trên việc thu thập 400.000 nút trong mạng Gnutella trong 6 tháng gồm tháng 11 năm 2000, tháng 2/tháng 3 năm 2001 và tháng 5 năm 2001.

4.1 Xu hướng và hành vi tăng trưởng

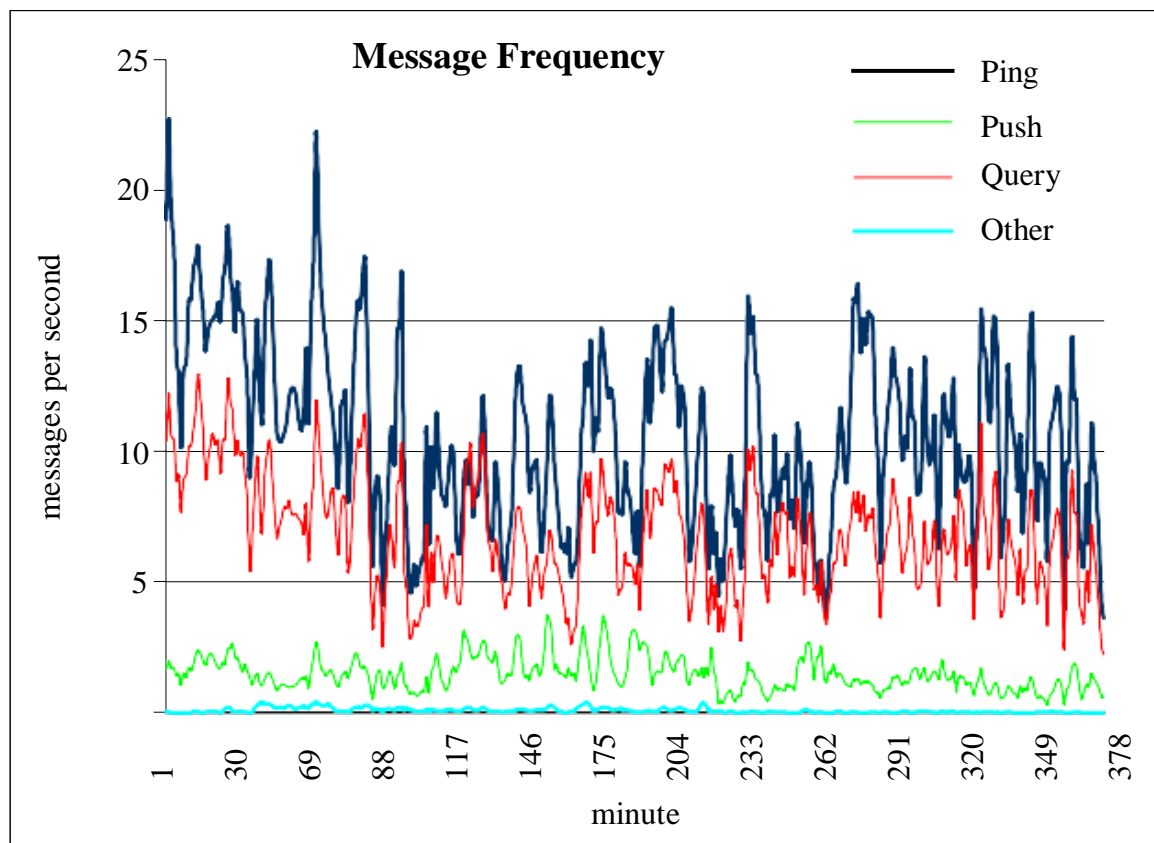


Hình 4.1.1 Tốc độ tăng trưởng ca mạng Gnutella

Biểu đồ hình 4.1.1 trình bày sự phát triển của mạng lưới Gnutella. Theo quan sát, ta thấy sự tăng lên đáng kể của số lượng máy chủ. ví dụ, từ 2.063 máy chủ vào tháng 11 năm 2000 lên đến 14.949 vào tháng 3 và 48.195 vào tháng 5 năm 2001. Mặc dù việc mở rộng quy mô của Gnutella gặp nhiều thách thức, số lượng nút trong thành phần lớn nhất của mạng đã tăng gấp khoảng 25 lần trong 6 tháng qua.

Nghiên cứu cũng đã phân tích cấu trúc động của biểu đồ theo thời gian dựa trên các bản ghi từ các lần thu thập dữ liệu liên tiếp. Kết quả cho thấy khoảng 40% các nút rời khỏi mạng trong vòng chưa đầy 4 giờ, trong khi chỉ có khoảng 25% các nút còn sống trong hơn 24 giờ. Điều này cho thấy sự thay đổi nhanh chóng của mạng Gnutella.

Để đảm bảo hiệu quả của việc thu thập dữ liệu, việc tìm ra sự cân bằng giữa thời gian khám phá và sự xâm lấn của trình thu thập thông tin là rất quan trọng. Tăng số lượng tác vụ thu thập dữ liệu song song có thể giúp giảm thời gian khám phá, nhưng cũng có thể tạo ra gánh nặng đối với hệ thống.



Hình 4.1.2 Tần suất thông điệp được gửi đi trong mạng

Trong tháng 11 năm 2000, các nhà nghiên cứu đã sử dụng phiên bản sửa đổi của trình thu thập thông tin để nghe lén lưu lượng truy cập của mạng Gnutella trong khoảng thời gian 376 phút. Trong biểu đồ hình 4.1.2, các nhà nghiên cứu đã phân loại lưu lượng truy cập theo loại tin nhắn và phát hiện rằng:

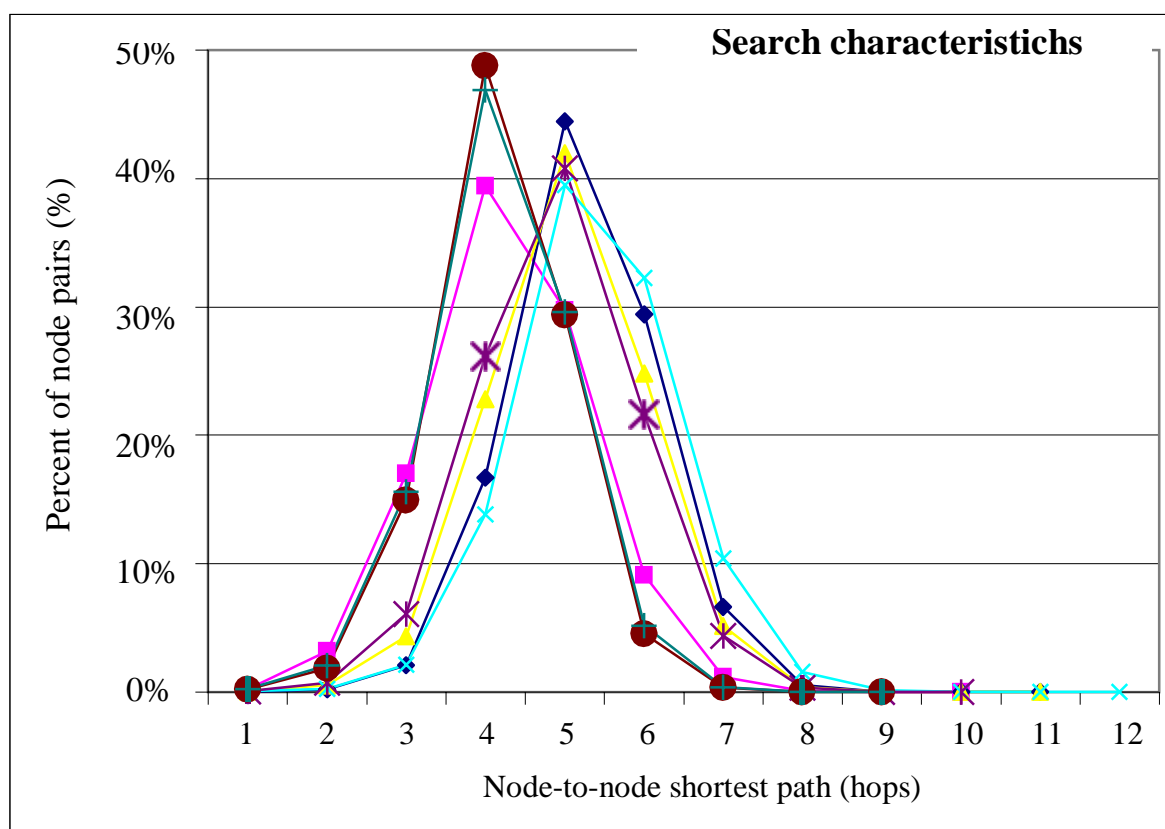
- Trung bình chỉ có khoảng 36% tổng lưu lượng truy cập là từ người dùng thông qua các thông điệp tìm kiếm (QUERY).
- 55% lưu lượng truy cập được sử dụng để duy trì tư cách thành viên nhóm thông qua tin nhắn PING và PONG.
- 9% chứa các thông điệp không chuẩn hoặc tin nhắn quảng bá PUSH từ các dịch vụ không tuân thủ giao thức.

Đến tháng 6 năm 2001, các vấn đề kỹ thuật này đã được giải quyết với sự xuất hiện của các triển khai mới hơn của Gnutella. Lưu lượng truy cập chủ yếu chứa 91% tin nhắn QUERY, 8% tin nhắn PING, và một số lượng không đáng kể của các loại tin nhắn khác. Mặc dù sự thất bại của mạng Gnutella trong việc mở rộng quy mô đã được dự đoán hết lần này đến lần khác, nhưng số lượng các nút trong thành phần mạng lớn nhất đã tăng khoảng 25 lần trong 6 tháng qua.

Loại message	Nov 2000	Jun 2001
User (Query)	36%	91%
Group membership (Ping/Pong)	55%	8%
Other Non-Standard	9%	1%

Bảng 4.1.1 Lưu lượng truy cập trong tháng 11/2000 và tháng 6/2001

4.2 Ước tính lưu lượng truy cập



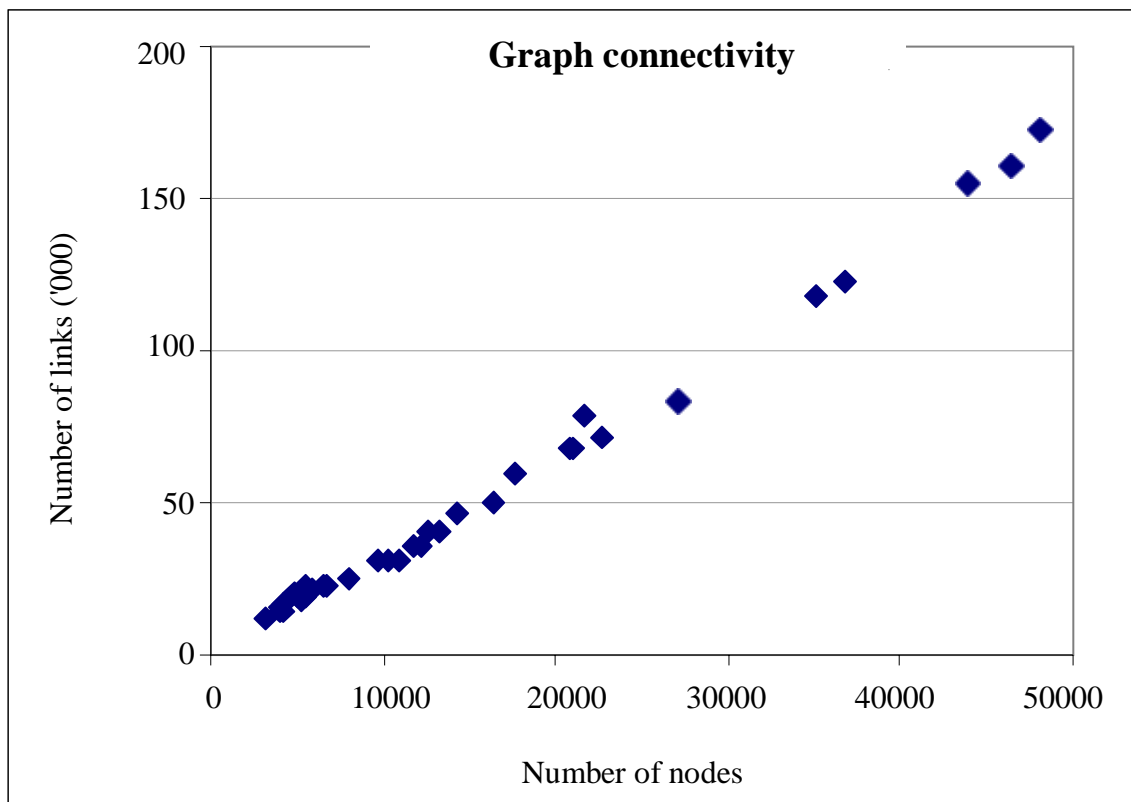
Hình 4.2.1. Phân phối đường đi ngắn nhất giữa các nút.

Trong biểu đồ hình 4.2.1, nghiên cứu trình bày phân bố của đường dẫn ngắn nhất từ một nút đến một nút khác trong mạng Gnutella. Đường dẫn ngắn nhất được

xác định là số lượng liên kết mà một thông điệp phải đi qua để đi từ một nút này sang một nút khác.

Với đường kính nhỏ của mạng (trung bình hai nút cách nhau ít hơn 7 bước), thường thì thời gian tồn tại của thông điệp được sử dụng và thuật toán định tuyến được áp dụng.

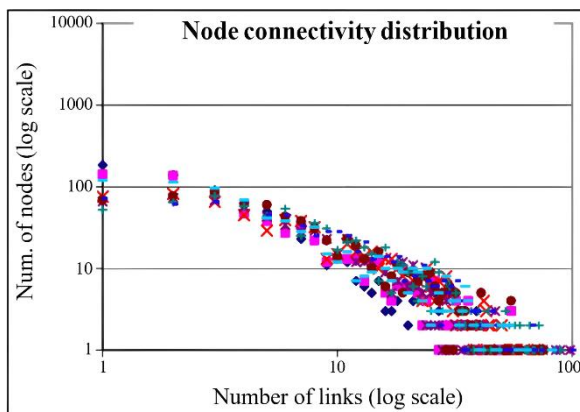
Nhận thấy rằng hầu hết các liên kết hỗ trợ lưu lượng truy cập một cách tương đồng. Điều này đã được xác minh thông qua việc đo lường truy cập tại nhiều nút được chọn ngẫu nhiên. Kết quả cho thấy tổng lưu lượng truy cập tạo ra bởi Gnutella tỷ lệ thuận với số lượng kết nối trong mạng. Đặt lưu lượng truy cập này vào bối cảnh, nhận thấy rằng nó chiếm khoảng 1,7% tổng lưu lượng truy cập trên Internet của Hoa Kỳ vào tháng 12 năm 2000. Các nhà nghiên cứu đã kết luận rằng khối lượng lưu lượng truy cập được tạo ra là một rào cản quan trọng đối với sự mở rộng và việc sử dụng hiệu quả cơ sở hạ tầng mạng cơ bản là rất quan trọng để mở rộng và triển khai rộng lớn hơn.



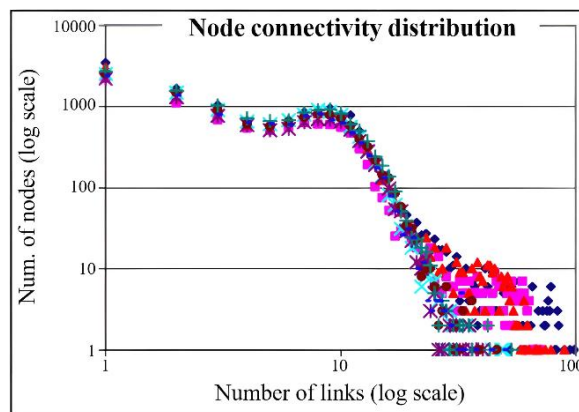
Hình 4.2.2. Số lượng kết nối trung bình.

Biểu đồ hình 4.2.2 trình bày một mô hình về kết nối nút trung bình trong mạng Gnutella. Mỗi điểm trên đồ thị đại diện cho một lần thu thập dữ liệu từ mạng Gnutella. Điều đáng chú ý là, khi mạng phát triển, số lượng kết nối trung bình trên mỗi nút không đổi và được ước tính là khoảng 3,4 kết nối trên mỗi nút.

4.3 Khả năng kết nối và độ tin cậy



Hình 4.3.1 Phân phối kết nối trong
tháng 11/2000



Hình 4.3.2 Phân phối kết nối trong
tháng 3/2001

Gnutella là một mạng lưới tự tổ chức, nơi người dùng giới hạn tối đa sự kết nối và các nút tự quyết định kết nối với nhau. Tương tự như các mạng lưới luật quyền lực, Gnutella có những đặc điểm sau:

- Rất ổn định và đàn hồi, nhưng dễ bị hỏng hóc thường xuyên
- Rất ít nút có khả năng kết nối cao và nhiều nút có kết nối thấp
- Cực kỳ mạnh mẽ đối với lỗi nút ngẫu nhiên, nhưng dễ bị tấn công theo kế hoạch.

Những đặc điểm này đều ảnh hưởng đến tính linh hoạt và sự ổn định của mạng Gnutella, và cung cấp cơ sở cho việc nghiên cứu và cải tiến hiệu suất của nó.

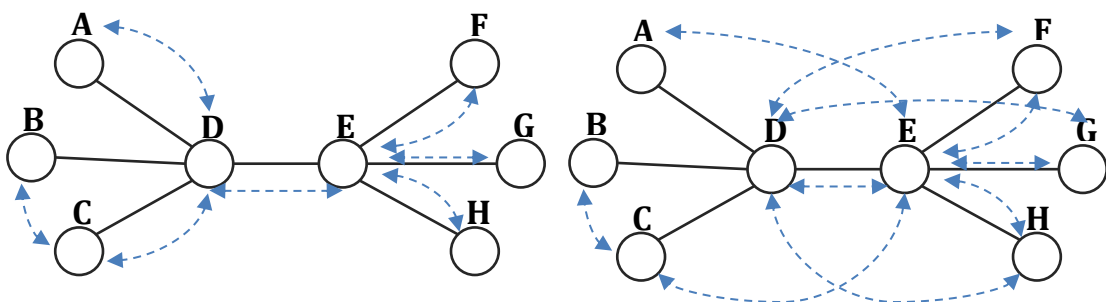
Biểu đồ hình 4.3.1 và biểu đồ hình 4.3.2 trình bày phân bố kết nối trong mạng Gnutella vào tháng 11 năm 2000 và tháng 3 năm 2001. Mỗi chuỗi điểm đại diện cho một cấu trúc liên kết mạng Gnutella được phát hiện vào các thời điểm khác nhau trong từng tháng. Ở cả hai hình biểu đồ, chúng ta có thể thấy một biểu hiện rõ ràng của tỷ lệ nhất ký trên cả hai trục. Điều này ám chỉ rằng các nút trong mạng Gnutella tự tổ chức thành một mạng lưới luật quyền lực, trong đó một số nút có số lượng kết nối lớn hơn so với các nút khác. Biểu đồ hình 4.3.1 cho thấy phân phối gần

như tuyến tính, trong đó ít phụ thuộc vào các nút được kết nối cao. Điều này có nghĩa là mạng có khả năng chịu lỗi cao hơn. Tuy nhiên, biểu đồ hình 4.3.2 không cho thấy đặc điểm của mạng lưới luật quyền lực cho các nút có ít hơn 10 liên kết, mà thay vào đó, nó hiển thị một biểu đồ đuôi nặng cho các nút có hơn 10 liên kết. Lý do là rất ít nút có kết nối cao, hoạt động giống như máy chủ và có khả năng được kết nối bởi nút mới cao hơn 50%. Vì thế, trong biểu đồ hình 4.3.2, mạng phụ thuộc nhiều hơn vào các nút được kết nối cao, dẫn đến sự giảm độ tin cậy của mạng.

Sự tự tổ chức này có thể là kết quả của cơ chế tự tổ chức của giao thức Gnutella, nơi các nút tự động kết nối với nhau để tạo thành một mạng phân phối không cần sự điều phối từ máy chủ trung tâm. Mô hình này có thể giúp mạng Gnutella tồn tại và hoạt động một cách hiệu quả, vì các nút có thể giao tiếp trực tiếp với nhau mà không cần thông qua trung gian, từ đó tạo ra một môi trường chia sẻ tệp đáng tin cậy và linh hoạt.

4.4 Cơ sở hạ tầng internet và mạng Gnutella

Mạng ngang hàng cho phép các máy tính ở rìa mạng hoạt động ở cả vai trò máy khách lẫn máy chủ. Dẫn đến các ứng dụng P2P thay đổi hoàn toàn lượng băng thông mà người dùng internet trung bình tiêu thụ. Điều này khiến cho các nhà cung cấp dịch vụ internet (ISP) có thể phải chỉnh lại mô hình kinh doanh và lên một kế hoạch định giá khác do họ đang sử dụng một mức giá cố định dựa trên lượng băng thông để tính chi phí internet cho khách hàng của họ.



Hình 4.4 Hai ánh xạ khác nhau của cấu trúc liên kết mạng Gnutella với cơ sở hạ tầng

Kiến trúc lưu trữ và chuyển tiếp của Gnutella làm cho cấu trúc liên kết mạng ảo trở nên cực kỳ quan trọng. Sự không tương thích giữa cơ sở hạ tầng mạng và cấu

trúc liên kết ảo của ứng dụng P2P càng lớn thì “áp lực” đối với cơ sở hạ tầng càng lớn. Trong phần này, ta sẽ xem liệu cấu trúc liên kết mạng Gnutella tự tổ chức có ánh xạ tốt với cơ sở hạ tầng vật lý hay không.

Trong Hình 7, tám máy chủ tham gia vào một mạng giống như Gnutella. Các đường màu đen, liền mạch để đại diện cho cơ sở hạ tầng mạng bên dưới và các đường màu xanh lam, chấm chấm để biểu thị cấu trúc liên kết ảo của ứng dụng. Trong hình bên trái, cấu trúc liên kết ảo phản ánh chặt chẽ cơ sở hạ tầng, cho phép một thông điệp được phân phối từ nút A đến tất cả các nút khác chỉ thông qua một liên kết vật lý. Tuy nhiên, trong hình bên phải, mặc dù cấu trúc liên kết ảo có chức năng tương tự, nhưng không phản ánh chính xác cơ sở hạ tầng, dẫn đến cùng một phân phối thông điệp yêu cầu sáu lần truyền qua cùng một liên kết.

Tuy vậy, việc tính toán chính xác ánh xạ của Gnutella lên cấu trúc liên kết internet là cực kỳ tốn kém, do khó khăn trong việc trích xuất cấu trúc liên kết internet và quy mô tính toán của vấn đề. Thay vào đó thì hai thử nghiệm cấp cao làm nổi bật sự không phù hợp giữa các cấu trúc liên kết của hai mạng đã được tiến hành.

Internet là một tập hợp các hệ thống tự trị (AS) được kết nối bởi các bộ định tuyến. Các AS là tập hợp các mạng cục bộ dưới một cơ quan quản lý kỹ thuật duy nhất. Theo quan điểm của ISP, lưu lượng mạng qua biên giới AS đắt hơn lưu lượng mạng cục bộ. Các nhà nghiên cứu nhận thấy rằng chỉ có 2-5% kết nối Gnutella liên kết các nút nằm trong cùng một AS, mặc dù hơn 40% các nút này nằm trong top mười AS hàng đầu. Kết quả này chỉ ra rằng hầu hết lưu lượng do Gnutella tạo ra đều vượt qua biên giới AS, do đó làm tăng chi phí một cách không cần thiết.

Trong thử nghiệm thứ hai, các nhà nghiên cứu giả định rằng tổ chức phân cấp của tên miền phản ánh cấu trúc phân cấp của cơ sở hạ tầng Internet. Ví dụ: có khả năng chi phí giao tiếp giữa hai máy chủ trong miền “uchi Chicago.edu” sẽ nhỏ hơn đáng kể so với giữa “uchi Chicago.edu” và “sdsc.edu”. Giả định cơ bản ở đây là tên miền thể hiện một số loại phân cấp tổ chức và các tổ chức có xu hướng xây dựng mạng khai thác cục bộ trong phân cấp đó.

Để nghiên cứu mức độ ánh xạ của cấu trúc liên kết ảo Gnutella với phân vùng Internet được định nghĩa bởi các tên miền, các nhà nghiên cứu chia sơ đồ cấu trúc

liên kết ảo Gnutella thành các cụm (cluster), tức là các đồ thị con có kết nối nội bộ cao. Với thuật toán định tuyến flooding-like được sử dụng bởi Gnutella, chính trong các cụm này, phần lớn lượng dữ liệu trong mạng được tạo ra. Do đó, các nhà nghiên cứu muốn xem các cụm này ánh xạ tốt như thế nào trên phân vùng được xác định bởi lược đồ đặt tên miền.

Một thuật toán phân cụm đơn giản được sử dụng dựa trên phân phối kết nối được mô tả trước đó: các nhà nghiên cứu định nghĩa là các đồ thị con của các cụm được tạo thành bởi một trung tâm với các nút lân cận của nó. Nếu hai cụm có hơn 25% nút chung thì hợp nhất chúng lại. Sau khi phân nhóm được thực hiện:

- (1) Chỉ định các nút có trong nhiều hơn một cụm vào thành một cụm lớn nhất.
- (2) Tạo thành một cụm cuối cùng với các nút không còn lại không thuộc trong bất kỳ cụm nào khác.

Ta xác định entropy của một tập C , chứa $|C|$ host, mỗi máy được gán nhãn bằng một trong n tên miền riêng biệt như sau:

$$E(C) = \sum_{i=1}^n (-p_i \log(p_i) - (1 - p_i) \log(1 - p_i)) \quad (1)$$

Với p_i là xác suất chọn ngẫu nhiên một máy chủ có tên miền i . Sau đó, ta xác định entropy của một nhóm đồ thị có kích thước $|C|$, được nhóm lại trong k cụm C_1, C_2, \dots, C_k có kích thước $|C_1|, |C_2|, \dots, |C_k|$ với $|C| = |C_1| + |C_2| + \dots + |C_k|$ như sau:

$$E(C_1, C_2, \dots, C_k) = \sum_{i=1}^n \frac{|C_i|}{|C_1| + |C_2| + \dots + |C_k|} \times E(C_i) \quad (2)$$

Dựa trên những lý do đã trình bày ở trên thì $E(C) \geq E(C_1, C_2, \dots, C_k)$ bất kể cụm C_1, C_2, \dots, C_k được chọn như thế nào. Nếu phân cụm khớp với phân vùng của miền thì ta thấy rằng: $E(C) \gg E(C_1, C_2, \dots, C_k)$. Ngược lại, nếu phân cụm C có cùng mức độ ngẫu nhiên như trong tập C ban đầu, thì entropy phần lớn sẽ không thay đổi. Về cơ bản, hàm entropy được sử dụng ở đây để đo lường mức độ phù hợp của hai phân vùng được áp dụng để thiết lập các nút: phân vùng đầu tiên sử dụng thông tin có trong tên miền, trong khi phân vùng thứ hai sử dụng phương pháp phân nhóm.

Các phân tích này được thực hiện trên 10 đồ thị cấu trúc liên kết được thu thập trong tháng 2/2001 và 3/2001. Các nhà nghiên cứu không phát hiện thấy entropy giảm nghiêm trọng sau khi thực hiện phân nhóm (tất cả các mức giảm đều nằm trong khoảng dưới 8% so với giá trị entropy ban đầu). Do đó, ta kết luận rằng các nút Gnutella tập hợp theo cách hoàn toàn độc lập với cấu trúc internet. Giả sử rằng cấu trúc tên miền Internet gần như khớp với cấu trúc liên kết cơ bản (chi phí gửi dữ liệu trong một miền nhỏ hơn chi phí gửi dữ liệu giữa các miền), có thể kết luận rằng mạng Gnutella tự tổ chức không sử dụng một cách hiệu quả cơ sở hạ tầng vật lý bên dưới.

CHƯƠNG 5 – ỨNG DỤNG GTK - GNUTELLA

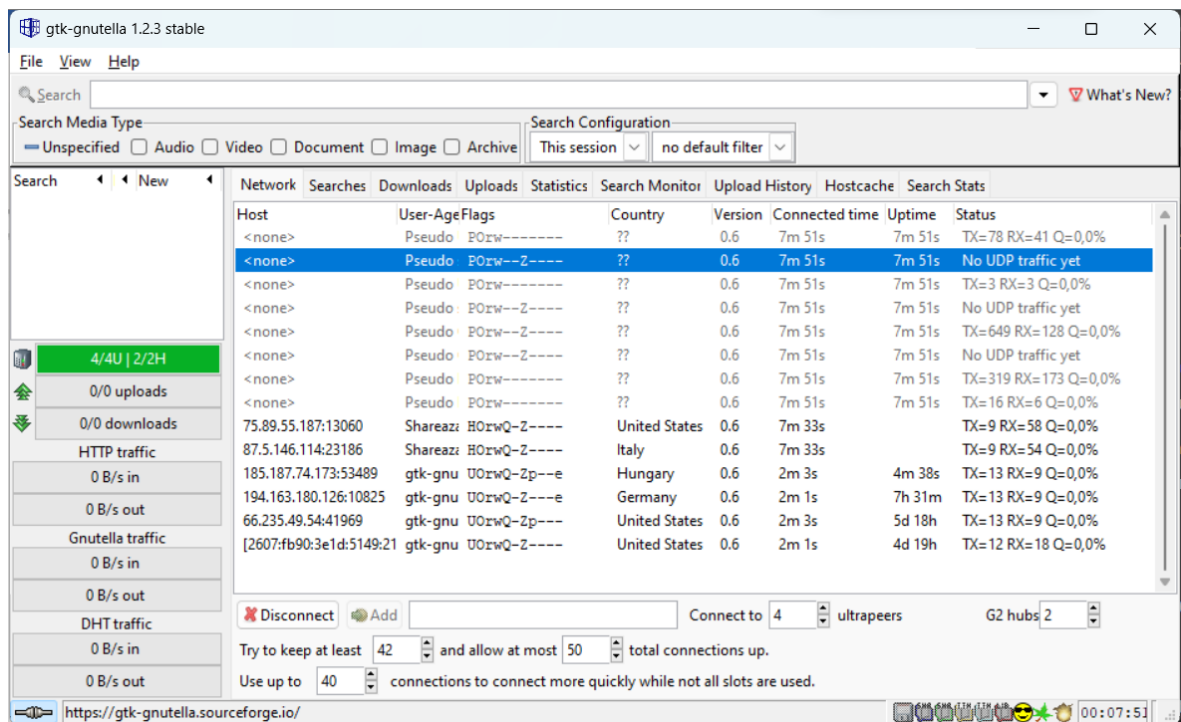
Gtk-gnutella là một ứng dụng miễn phí được viết bằng ngôn ngữ C, hoạt động trên đa nền tảng và hiện vẫn còn hoạt động trong mạng Gnutella. Ứng dụng này cho phép chia sẻ các loại tập tin khác nhau như sách, phim ảnh, âm nhạc và trò chơi trên mạng Gnutella.

Trong chương này, chúng ta sẽ tìm hiểu cách cài đặt gtk-gnutella trên hệ điều hành Windows và cách sử dụng ứng dụng này để có thể đánh giá một cách khách quan về nó.

5.1 Tải và cài đặt ứng dụng

Trước tiên, tải phiên bản mới nhất của gtk-gnutella (1.2.3) từ liên kết: [link](#)

Sau khi tải về, tiến hành cài đặt và khởi động ứng dụng. Giao diện chính của ứng dụng sẽ hiển thị bên dưới đây.



Hình 5.1 Giao diện chính của ứng dụng GTK – Gnutella

Khi giao diện hiện lên, ứng dụng sẽ ngay lập tức gửi thông điệp Ping để tìm các server xung quanh. Sau khi nhận được phản hồi (Pong), danh sách các server phản hồi sẽ hiển thị trên giao diện như trong Hình 12. Từ đây, người dùng có thể xem thông tin chi tiết từng server và tiến hành các hoạt động tiếp theo.

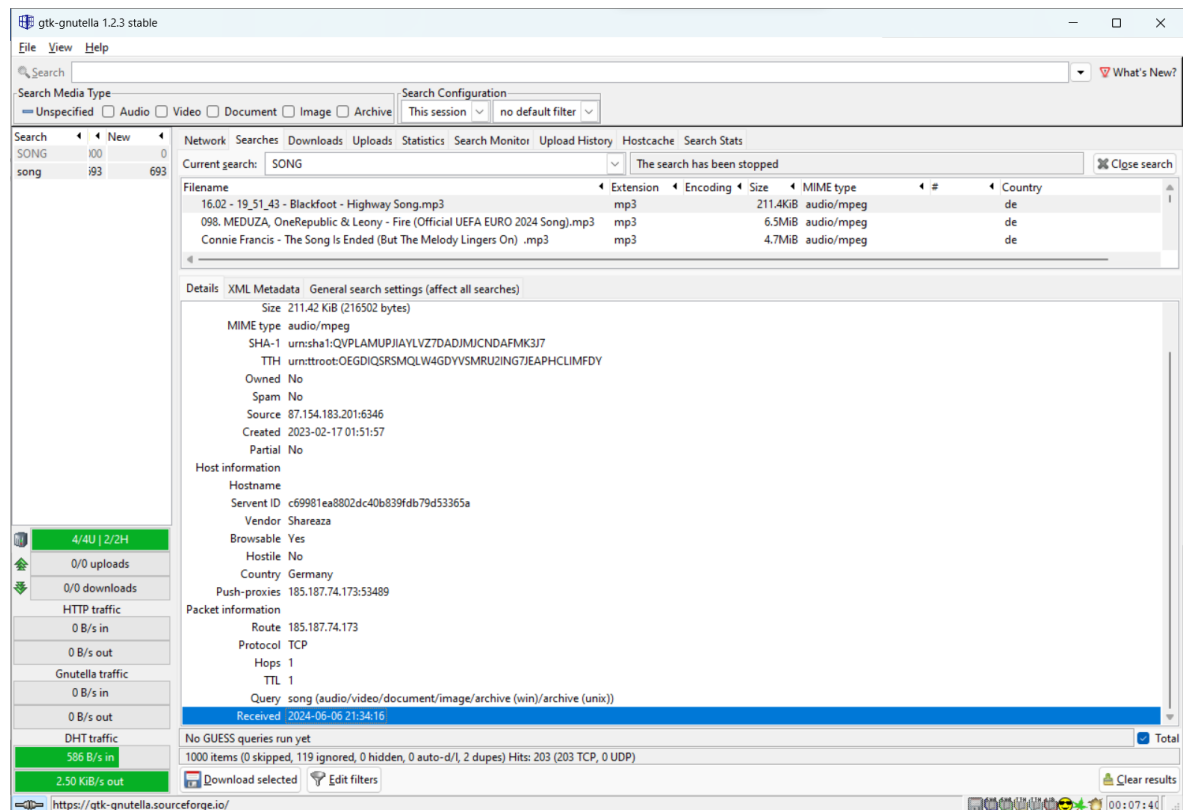
Thông tin chi tiết các server:

- **Host:** Địa chỉ của các server mà ứng dụng tìm thấy. Các host màu nhạt là địa chỉ IPv4 của máy tính đang sử dụng ứng dụng.
- **User-Agent:** Tên ứng dụng mà các server khác sử dụng, ví dụ như gtk-gnutella hoặc Shareaza, cho thấy khả năng kết nối với mọi server trong mạng Gnutella.
- **Flags:** Trạng thái của server, ví dụ “U” là ultrapeer, “O” là loại kết nối outgoing.
- **Quốc gia và Phiên bản Gnutella:** Thông tin về quốc gia và phiên bản Gnutella (đa số các server hiển thị là phiên bản 0.6).

5.2 Tìm kiếm và tải tập tin

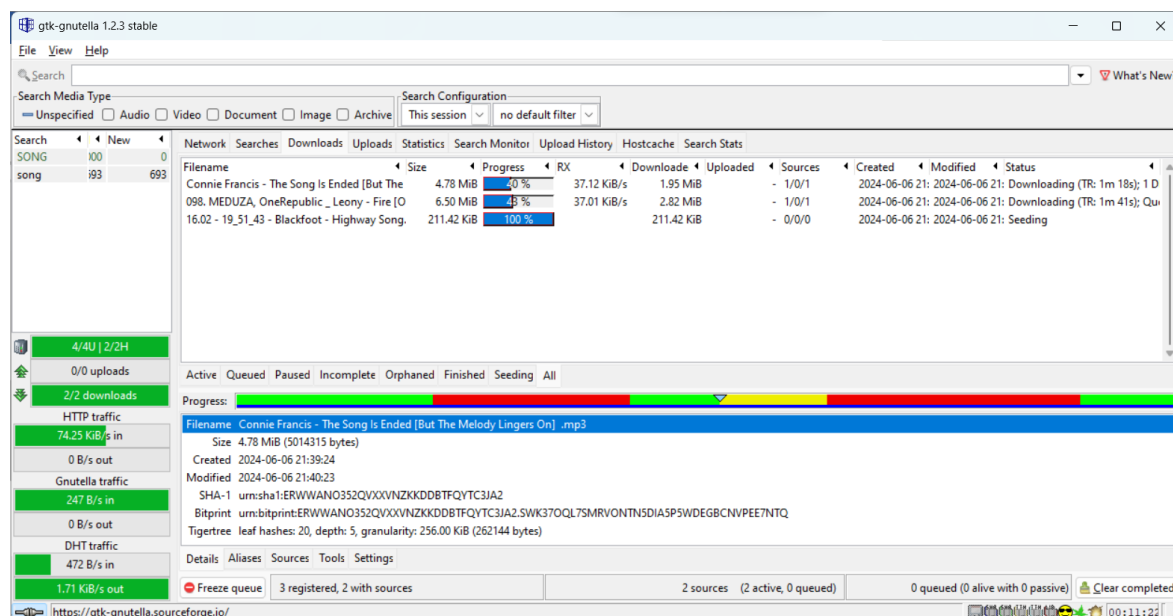
Gtk-gnutella xoay quanh hai chức năng chính: tìm kiếm và chia sẻ tập tin.

Khi sử dụng chức năng tìm kiếm và tìm bài hát tên “song”, một thông điệp truy vấn (query) sẽ được gửi đến các server trong mạng. Các nút chứa thông tin khớp sẽ gửi lại một truy vấn ngược (query hit) bao gồm thông tin của server và tập tin.



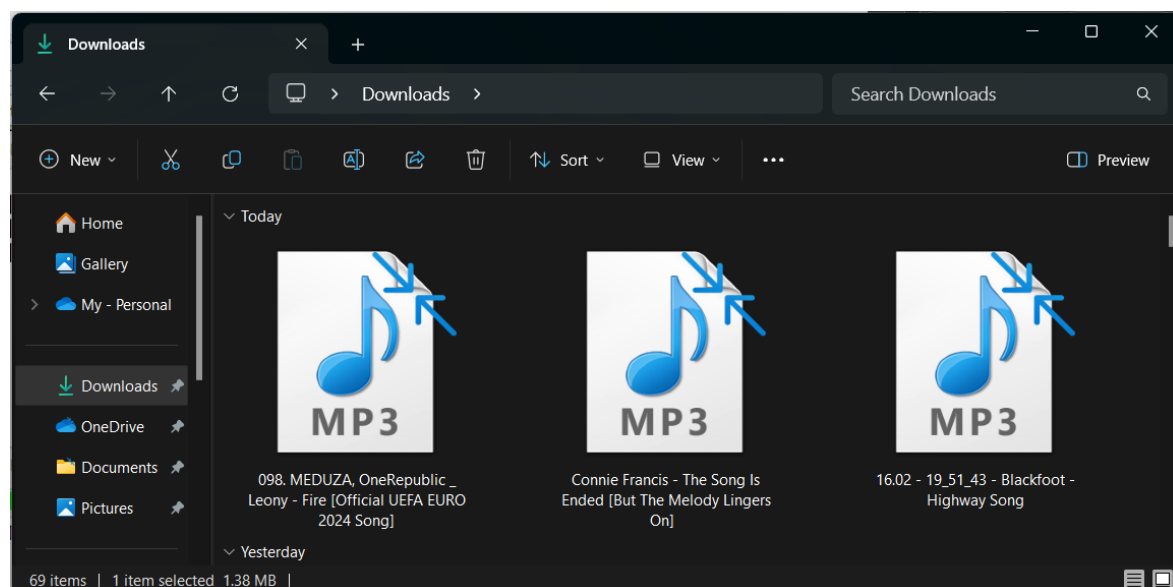
Hình 5.2.1 Danh sách các thông tin khi tìm kiếm và thông tin về server được chọn

Sau khi có danh sách các tập tin khớp với từ khóa tìm kiếm, bạn có thể chọn một tập tin và tiến hành tải về. Tốc độ tải có thể tương đối chậm, chỉ 12.37 KiB/s. Với một bài hát khoảng 4 MB, cần hơn 30 giây hoặc 1 phút để tải xong. Việc tải nhiều tập tin cùng lúc sẽ chia đều tốc độ tải.



Hình 5.2.2 Tiến độ tải của các bài hát đã chọn

Sau khi tải thành công, các tập tin sẽ hiển thị tại thư mục cài sẵn.

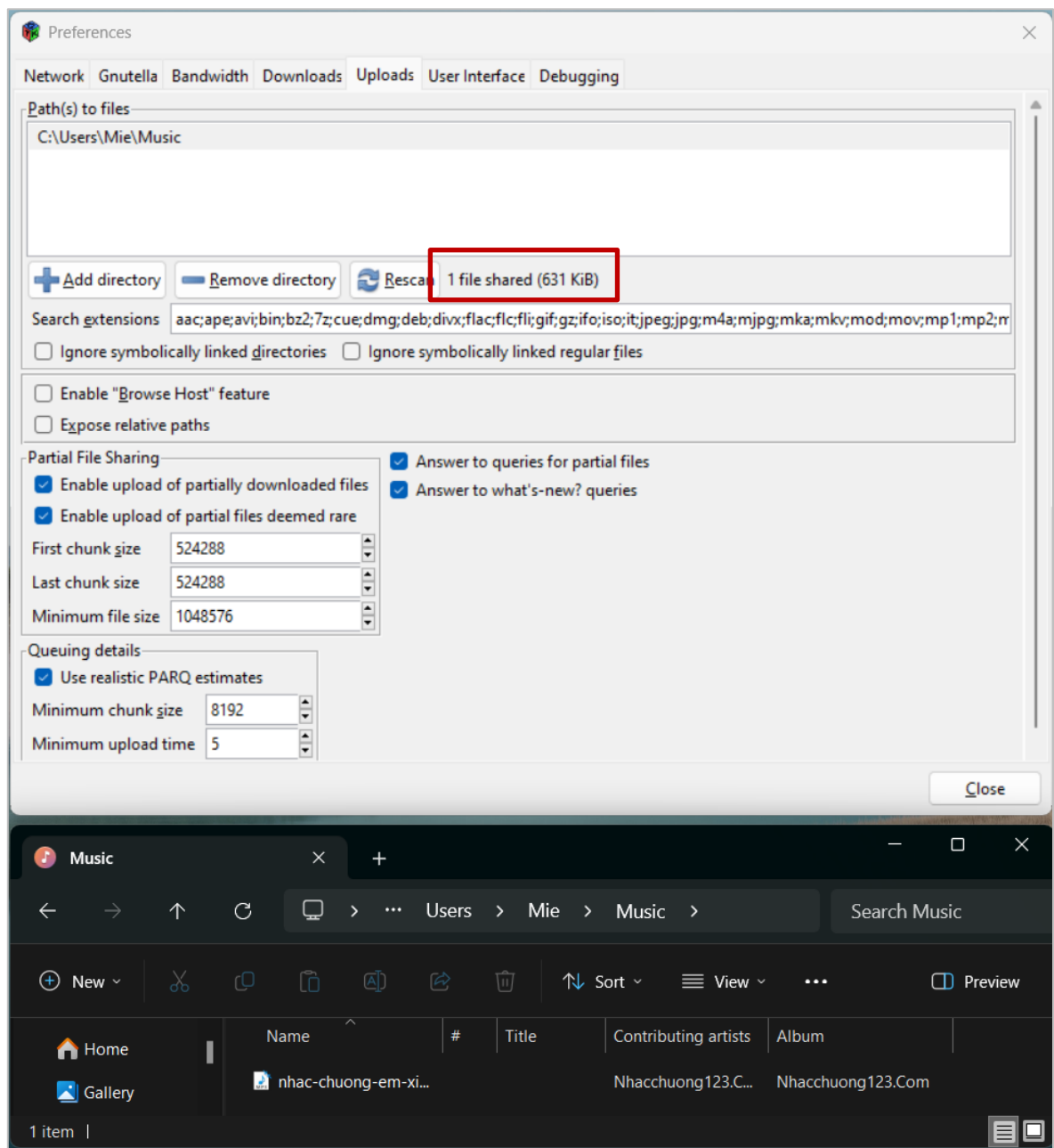


Hình 5.2.3 Các tập tin đã được tải xuống thành công

5.3 Chia sẻ tập tin

Ngoài khả năng tìm kiếm và tải tập tin, Gtk-Gnutella còn cho phép chia sẻ dữ liệu từ máy tính của bạn. Để chia sẻ tập tin, thực hiện các bước sau:

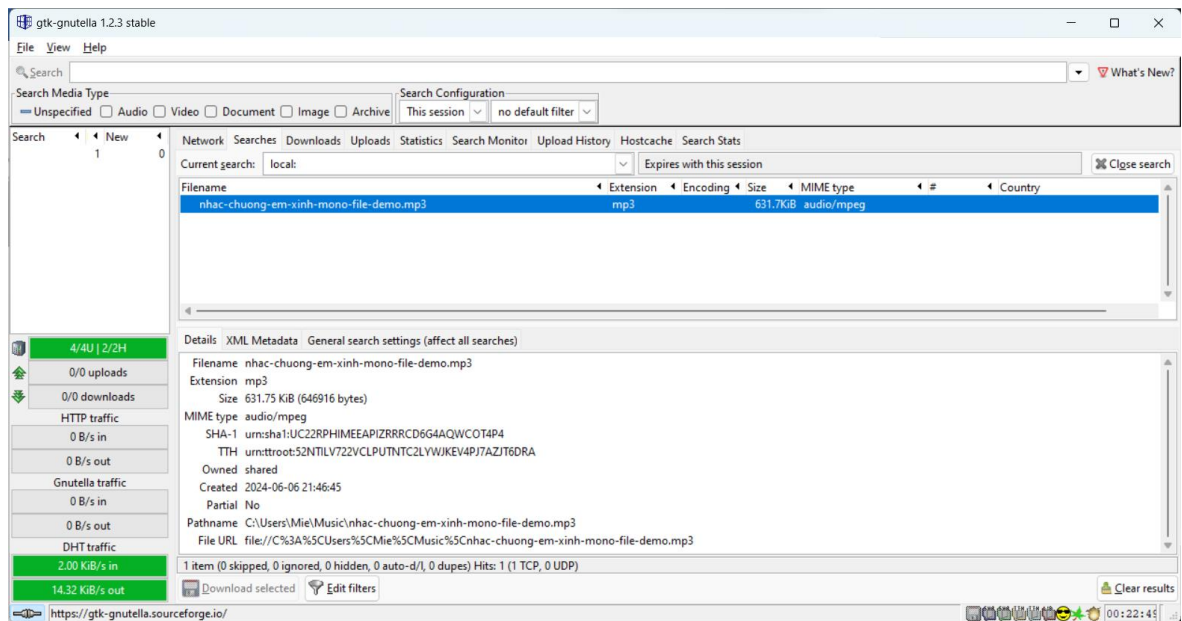
- Nhấn File > Preferences.
- Chọn Add directory.
- Chọn thư mục chứa các tập tin bạn muốn chia sẻ, ví dụ như thư mục Music.



Hình 5.3.1 Giao diện chọn thư mục

Kết quả cho thấy có 1 file đã được tải lên mới kích thước 631 KB. Để giới hạn kiểu dữ liệu của tập tin thông qua phần **Search extensions** hoặc điều chỉnh lượng dữ liệu có thể chia sẻ trong **Partial File Sharing**.

Để kiểm tra các tập tin được chia sẻ, gõ lệnh “local:” trên thanh tìm kiếm. Các tập tin đã hiện chính là các tập tin có thể được chia sẻ trên mạng Gnutella. Bây giờ thì các server khác đều có thể tìm các tập tin trên miễn là họ search đúng từ khóa có trong các tập tin.



5.4 Đánh giá

Nhìn chung, gtk-gnutella rất dễ cài đặt với giao diện đơn giản và dễ sử dụng. Tuy nhiên, chức năng của ứng dụng chỉ gói gọn trong ba tính năng chính: tìm kiếm, tải và chia sẻ tập tin. Tốc độ tải tương đối chậm, chỉ đáp ứng được các tập tin có dung lượng nhỏ. Tốc độ không đều còn phụ thuộc vào chất lượng đường truyền của các nút tham gia mạng.

Vấn đề bảo mật cũng đáng lưu ý, khi chỉ hiển thị tên và dung lượng của tập tin mà không cho xem trước nội dung. Điều này tiềm ẩn nguy cơ tập tin có thể là mã độc giả dạng thành định dạng an toàn.

CHƯƠNG 6 – KẾT LUẬN

6.1 Kết quả đạt được

Giao thức Gnutella đã chứng tỏ khả năng hoạt động đơn giản và hiệu quả. Khi một nút mới kết nối tới một host có sẵn, nó ngay lập tức trở thành thành viên của mạng và bắt đầu truyền đi các thông điệp tới các nút khác đã tham gia.

Trong các mạng lưới quy mô vừa và nhỏ, quá trình tìm kiếm diễn ra nhanh chóng. Không cần phải thông qua một máy chủ trung gian, chỉ cần gửi một thông điệp đi, nó sẽ tự động lan rộng đến các nút xung quanh và tiếp tục truyền đi cho đến khi đạt $TTL = 0$.

Mỗi nút lưu trữ thông tin về các tệp tin của mình, giúp quá trình tìm kiếm hiệu quả hơn. Khi một nút muốn tìm kiếm tệp tin, nó có thể trực tiếp truy xuất đến nút chứa tệp tin đó mà không cần phải gửi thông điệp qua nhiều nút trung gian như lần đầu tiên.

Mạng Gnutella không cần máy chủ, nên có thể duy trì hoạt động liên tục. Chỉ cần một nút trong mạng hoạt động, các nút khác có thể tham gia bất cứ lúc nào.

6.2 Hạn chế

Tuy nhiên, việc tìm kiếm tệp tin trong một mạng lưới lớn với hàng ngàn nút tham gia có thể rất mất thời gian, đặc biệt với các tệp tin hiếm chỉ xuất hiện ở một vài nút trong mạng. Mỗi thông điệp gửi đi đều có giới hạn số lần truyền (TTL), do đó có thể bị “chết” trước khi đến được đích, buộc nút tìm kiếm phải gửi lại thông điệp với TTL lớn hơn.

Tính không ổn định của các server (online - offline) trong mạng cũng gây gián đoạn khi các thông điệp đang được gửi. Nếu nút đang nhận thông điệp đột ngột offline, tệp tin cần tìm có thể không được tìm thấy, làm quá trình tìm kiếm khó khăn hơn. Hơn nữa, các thông điệp không thành công vẫn tiêu tốn băng thông mạng, gây lãng phí tài nguyên.

Tính bảo mật của mạng Gnutella kém, vì các dữ liệu tải về từ mạng này có thể chứa phần mềm độc hại và không có cơ chế kiểm tra từ các nút tham gia. Vấn đề bản quyền cũng là một hạn chế lớn. Các tệp tin như bài hát mới phát hành có thể được

chia sẻ và tải về miễn phí mà không cần sự cho phép của tác giả, vi phạm bản quyền. Đây là một trong những nguyên nhân chính ngăn chặn sự phát triển của Gnutella do các vấn đề pháp lý liên quan đến bản quyền từ các hãng ghi âm.

TÀI LIỆU THAM KHẢO

Tiếng Việt

1. Nguyễn Đại Thọ (2007), “*Công nghệ mạng ngang hàng*”, Bộ môn Mạng & Truyền thông Máy tính Khoa Công nghệ Thông tin, trường Đại học Công nghệ - Đại học Quốc gia Hà Nội.
2. Nguyễn Hoài Sơn, Hồ Sĩ Đàm (2008), “*Tìm kiếm thông tin theo các giá trị thuộc tính trên mạng ngang hàng có cấu trúc*”, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Tiếng Anh

3. M. Faloutsos, P. Faloutsos, C. Faloutsos, “*On Power-Law Relationships of the Internet Topology*”, SIGCOMM 1999
4. A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener, “*Graph structure in the web*”, 8 th International WWW Conference, May 15-19 Amsterdam
5. A. Barabasi and R. Albert, “*Emergence of scaling in random networks*”, Science, 286(509), 1999.
6. R. Albert, H. Jeong, A. Barabási, “*Attak and tolerance in complex networks*”, Nature 406 378 (2000).
7. A. Barabási, R. Albert, H. Jeong, G. Bianconi, “*Power-law distribution of the World Wide Web*”, Science 287, 2000.
8. L. Adamic, R. Lukose, A. Puniyani, B. Huberman, “*Search in Power-Law Networks*”. Phys. Rev. E, 2001.