

Anatomia do BitTorrent

a Ciência da Computação por trás do protocolo

Paulo Cheadi Haddad Filho
Orientador: José Coelho de Pina

Trabalho de Conclusão de Curso



IME-USP

Universidade de São Paulo
São Paulo, 2013

Sumário

Capa	i
Sumário	i
Lista de Figuras	ii
Glossário	iii
Lista de tarefas pendentes	v
1 Napster, Gnutella, eDonkey e BitTorrent	1
1.1 Período pré-torrent	1
1.1.1 Napster	2
1.1.2 Gnutella	3
1.1.3 eDonkey	4
1.2 Nascimento do BitTorrent	5
1.3 Mundo pós-torrent	5
2 Bibliografia	6

Lista de Figuras

Glossário

anycast

método de endereçamento e roteamento de rede onde os datagramas de um único remetente são roteados para um membro de um grupo de receptores potenciais que estão definidos pelo mesmo intervalo no endereço de destino. Geralmente é usado para serviços que demandem alta disponibilidade. [3](#)

Audiogalaxy

rede P2P de compartilhamento de músicas MP3 criado em 1998. [1](#), [2](#)

DHT

. [4](#)

eDonkey

. [4](#)

Gnutella

software de compartilhamento P2P desenvolvido por 3 programadores da empresa Nullsoft, recém adquirida da AOL Inc., lançado em 2000 sob a licença GPL. No dia seguinte, a AOL ordenou indisponibilizar o software alegando problemas legais e proibindo a continuação do desenvolvimento. Alguns dias depois, o protocolo já tinha sido alvo de engenharia reversa e já havia softwares que o implementavam.. [3](#), [4](#)

hash

. [4](#)

ISP

do inglês *Internet Service Provider*; fornecedores de acesso a Internet, que são empresas que vendem serviço e equipamento que permitem o acesso de um computador pessoal acessar a Internet. [1](#)

Kademlia

. [4](#)

MP3

do inglês *MPEG-1/2 Audio Layer 3*; formato patenteado de compressão de dados de áudio digital que usa um método de compressão de dados com perdas. [1](#), [2](#)

P2P

do inglês *peer-to-peer*; redes de arquitetura descentralizada e distribuída, onde cada nó (*peer*) faz fornece e consome recursos. [2-4](#)

peer

do inglês *peer-to-peer*; como são chamados cada nó da rede desse tipo. [3](#), [4](#)

RIAA

do inglês *Recording Industry Association of America*; Associação da Indústria de Gravação da América, organização que representa as gravadoras musicais e distribuidores, e tem sido autora de ações judiciais devido a quebra de direitos autorais causada por compartilhamento indevido de música. [2](#)

swarming

. [4](#)

Lista de tarefas pendentes

■ falta apontar as refs	3
-----------------------------------	---

Capítulo 1

Napster, Gnutella, eDonkey e BitTorrent

Para entendermos como e por que o BitTorrent se tornou o que é hoje, devemos voltar um pouco no tempo e rever a história que precedeu à sua criação, que é o fim da década dos anos 1990.

1.1 Período pré-torrent

Entre o final dos anos 80 e o início dos 90 [14, 23], a Internet deixou de ser uma rede de computadores usada somente por entidades governamentais, laboratórios de pesquisa e universidades, passando a ter seu acesso comercializado para o público em geral pelos [fornecedores de acesso a Internet \(ISP\)](#) [17]. Com o advento do [formato de áudio MP3 \(MP3\)](#) [20] no final de 1991 e do seu primeiro reprodutor de áudio MP3 Winamp, o tráfego da Internet aumentou devido ao aumento da troca direta desse tipo de arquivo.

Entre 1998 e 1999, dois sites de compartilhamento gratuito de músicas foram criados: o MP3.com [19], que era um site de divulgação de bandas independentes, e o [Audiogalaxy.com](#) [9, 12]. Mais popular que o primeiro, o Audiogalaxy era um site de busca de músicas, sendo que o download e upload eram feitos a partir de um software cliente. A lista de músicas procuradas ia da página para o computador onde usuário tinha instalado o cliente, que então conectava com o do outro usuário, que era indicado pelo servidor. A lista possuía todos os arquivos que um dia passaram pela sua rede. Se algum arquivo fosse requisitado mas o usuário que o possuísse não

estivesse conectado, o servidor central do Audiogalaxy fazia a ponte, pegando o arquivo para si e enviando-o para o cliente do requisitante em seu próximo login.

1.1.1 Napster

Em maio de 1999 surgiu o Napster [21], um programa de compartilhamento de MP3 que inovou por desfigurar o usual modelo cliente-servidor, onde um servidor central localizava os arquivos nos usuários e fazia a conexão entre os usuários, onde ocorria a transferências. O Napster foi contemporâneo do Audiogalaxy e ambos fizeram muito sucesso por cerca de 2 anos, até que começaram as ações judiciais.

Não demorou muito tempo para a indústria da música entrar em ação contra a troca de arquivos protegidos por direitos autorais sem autorização pela Internet. Seu primeiro alvo foi o Napster, em dezembro de 1999, quando a [RIAA \(do inglês *Recording Industry Association of America*\)](#) entrou com processo representando várias gravadoras alegando quebra de direitos autorais [6]. Em abril de 2000, foi a vez da banda Metallica processar, como resposta à sua descoberto que uma música ainda não lançada oficialmente já circulava na rede [4, 5]. Um mês depois, outra ação, agora encabeçada pelo rapper Dr. Dre, que tinha feito pedido formal para a retirada de seu material de circular, também abriu processo [3]. Isso fez com que o Napster tivesse atenção da mídia, ganhando popularidade e atingindo o 20 milhões de usuários em meados de 2000 [7].

Em 2001, esses imbróglios judiciais resultaram numa liminar federal que ordenava a retirada de conteúdo protegido das entidades representadas pela RIAA. O Napster tentou, mas a juíza do caso não ficou satisfeita ordenando então, em julho, o desligamento da rede enquanto não conseguisse controlar o conteúdo que trafegava ali [21]. Em setembro, o Napster fez um acordo, onde pagou 26 milhões de dólares por danos já causados, uso indevido de música e também 10 milhões de dólares pelos danos futuros envolvendo royalties. Para pagar esse valor, o Napster tentou cobrar o serviço de seus usuários, que acabaram migrando de rede rede P2P (P2P), inclusive para o Audiogalaxy. Não conseguindo, em 2002, o Napster decreta falência e é forçado a liquidar seus ativos. De lá para cá, foi negociado algumas vezes e atualmente pertence ao site Rhapsody [8].

O sucesso do Napster, mesmo que por curto período tempo, mostrou o potencial das redes P2P poderiam ter, e com isso novos softwares e protocolos de redes foram sendo lançados, sempre tentando se diferenciar dos outros softwares a fim de não serem novos alvos de ações judiciais. A solução para isso foi tentar descentralizar os mecanismos de indexação e de busca, que foram

os pontos fracos do Napster.

falta apontar as refs

1.1.2 Gnutella

O tal sucessor foi o [Gnutella](#), em março de 2000, foi uma resposta de domínio público feita com “gambiarras” para os problemas que o Napster encontrou com relação às acusações de violação de direitos autorais. Enquanto o Napster possuía um servidor central como estrutura que, no julgamento, foi usado como prova de que o sistema encorajava a violação de direitos autorais, o Gnutella foi modelado como um sistema P2P puro, onde todos os peer (um nó da rede)s são completamente iguais, sendo responsáveis pelos seus próprios atos.

O Gnutella disponibiliza arquivos da mesma forma que o Napster, mas sem a limitação de ser de formato de música, ou seja, qualquer arquivo pode ser compartilhado. A diferença mais significativa entre os dois protocolos é o algoritmo de busca: a abordagem do Gnutella é baseada numa forma de [anycast](#). Isso envolve duas partes: a primeira é como cada usuário é conectado a outros nós e mantém a lista dessas conexões atualizada. A segunda parte é como ele trata as buscas e trabalha inundando de pedidos para todos os nós que estão a uma certa distância do usuário (nó-cliente). Por exemplo, se a distância limite for de 4, então todos os nós que estiverem a 4 passos a partir do cliente serão verificados, começando a partir dos mais próximos. Eventualmente, algum nó possuirá o arquivo requisitado e responderá, e assim será feita a transferência desse arquivo. Muitos softwares que implementam o protocolo vão além dessa funcionalidade básica de download simples tentando transferir de forma paralela partes diferentes do arquivo desejado de nós diferentes, tentando amenizar eventuais problemas de velocidade de rede.

Experiências sugerem assim que o sistema escala para um tamanho maior, o mecanismo de anycast se torna extremamente caro e em algumas vezes até proibitivo. O problema ocorre nas buscas por arquivos menos populares, onde será necessário um maior número de nós perguntados.

O Gnutella ainda teve uma segunda versão, no final de 2002, onde utilizou o mesmo protocolo que o original, porém organizando a rede de peers em *leafs* (folhas, em inglês) e *hubs*. Um *hub* poderia ter centenas de conexões de folhas outras 7, em média, a outros *hubs*, enquanto uma folha se conectaria apenas a 2 *hubs*. Essa nova topologia, somada com uma nova tabela de índice de arquivos das folhas mantida pelos *hubs* onde estavam conectados, melhorou o desempenho das buscas, que era ruim na versão antiga.

1.1.3 eDonkey

Lançado em 6 de setembro de 2000, o protocolo [eDonkey](#) foi inaugurado juntamente com o software que o utilizava, o eDonkey2000, mas inúmeros softwares cliente para diferentes plataformas surgiram nos dias seguintes ao lançamento. Porém, não foi o cliente, mas sim o protocolo, que inovou em muitos aspectos de seus precursores, tendo papel fundamental na história das redes P2P.

O eDonkey implementou o primeiro método de download por [swarming](#), que é chamado o método onde peers fazem downloads de diferentes partes de um arquivo e de peers diferentes, utilizando de forma efetiva a largura de banda de rede para todos os peers ao invés de ficar limitado somente à banda de um único peer.

Outra melhoria foi a busca: no seu lançamento, os servidores eram separados entre si, porém nas versões seguintes permitiu que eles formassem uma rede de buscas. Isso permitiu que os servidores repassassem buscas de seus clientes conectados localmente a outros servidores, facilitando a localização de peers conectados em qualquer servidor da rede de buscas, aumentando a capacidade de download do enxame.

Uma terceira diferença com o Napster foi o uso de [hashs](#) de arquivos nos resultados de busca ao invés dos simples nomes dos arquivos. As buscas geradas pelos usuários eram baseadas em palavras-chave e comparadas com a lista de nomes de arquivos armazenada no servidor, mas o servidor retornava uma lista de pares de nomes de arquivos com seus respectivos valores hash. Enfim, quando o usuário selecionasse o arquivo desejado, o cliente iniciaria o download do arquivo usando o seu valor hash. Desse modo, um arquivo poderia ter muitos nomes entre os diferentes peer e servidores, mas seria considerado idêntico para download se possuísse o mesmo hash.

A arquitetura da rede em dois níveis usando cliente e servidor alcançou um meio termo entre as redes centralizadas, como o Napster, e as descentralizadas, como o Gnutella, já que o servidor central no primeiro era um alvo estável para ações legais, enquanto o segundo rapidamente mostrou-se inviável devido ao tráfego massivo de buscas entre peers.

Por fim, a inovação mais importante foi o uso de [DHT](#), em específico o [Kademlia](#), como algoritmo de indexação e busca nos servidores centrais dos arquivos através da rede eDonkey. Além de ser uma das causas da melhora no desempenho nas buscas, DHTs possuem ainda outras características como tolerância a falhas e escalabilidade. O Kademlia em específico ainda oferece outras vantagens como armazenamento de dados eficiente; anonimato; segurança de rede,

conteúdo e usuário; e autenticação.

1.2 Nascimento do BitTorrent

Aqui contarei a história do surgimento do protocolo.

1.3 Mundo pós-torrent

Aqui discorrerei sobre as consequências do uso do BitTorrent, desde as influências acadêmico-tecnológicas da sua difusão, bem como explicarei as questões jurídicas envolvidas no seu uso.

Capítulo 2

Bibliografia

- [1] Kennon Ballou. *R.I.P. Audiogalaxy*. [Online; accessed 30-September-2013]. 21 de jun. de 2002. URL: <http://www.kuro5hin.org/story/2002/6/21/171321/675>.
- [2] Kenneth P. Birman. *Reliable Distributed Systems: Technologies, Web Services, and Applications*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005, pp. 532–534. ISBN: 0387215093.
- [3] *Dr. Dre Raps Napster*. [Online; accessed 30-September-2013]. 18 de abr. de 2000. URL: <http://www.wired.com/techbiz/media/news/2000/04/35749>.
- [4] Daniela Hernandez. *April 13, 2000: Seek and Destroy – Metallica Sues Napster*. [Online; accessed 30-September-2013]. 13 de abr. de 2012. URL: <http://www.wired.com/thisdayintech/2012/04/april-13-2000-seek-and-destroy-metallica-sues-napster/>.
- [5] Christopher Jones. *Metallica Rips Napster*. [Online; accessed 30-September-2013]. 13 de abr. de 2000. URL: <http://www.wired.com/politics/law/news/2000/04/35670>.
- [6] David Kravets. *Dec. 7, 1999: RIAA Sues Napster*. [Online; accessed 30-September-2013]. 7 de dez. de 2009. URL: <http://www.wired.com/thisdayintech/2009/12/1207riaa-sues-napster/>.
- [7] *Napster: 20 million users*. [Online; accessed 30-September-2013]. 19 de jul. de 2000. URL: <http://cnfnf.cnn.com/2000/07/19/technology/napster/index.htm>.
- [8] *Rhapsody.com*. [Online; accessed 30-September-2013]. URL: <http://www.rhapsody.com>.
- [9] “Um pouco de história: redes P2P de compartilhamento de arquivos”. Em: *Revista PnP* 10 (out. de 2008). [Online; retirado de http://www.thecnica.com/artigos/PnP_10_02.pdf], p. 12. URL: http://www.revistapnp.com.br/pnp_10.php.

- [10] Wikibooks. *The World of Peer-to-Peer (P2P)* — *Wikibooks, The Free Textbook Project*. [Online; accessed 3-October-2013]. 2012. URL: [http://en.wikibooks.org/w/index.php?title=The_World_of_Peer-to-Peer_\(P2P\)&oldid=2316492](http://en.wikibooks.org/w/index.php?title=The_World_of_Peer-to-Peer_(P2P)&oldid=2316492).
- [11] Wikipedia. *Anycast* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 2-October-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=Anycast&oldid=574680440>.
- [12] Wikipedia. *Audiogalaxy* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 29-September-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=Audiogalaxy&oldid=560950036>.
- [13] Wikipedia. *EDonkey network* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 3-October-2013]. 2013. URL: http://en.wikipedia.org/w/index.php?title=EDonkey_network&oldid=568576016.
- [14] Wikipedia. *File sharing* — *Wikipedia, The Free Encyclopedia*. [Online; acessado em maio de 2013]. 2013. URL: http://en.wikipedia.org/w/index.php?title=File_sharing&oldid=556034682.
- [15] Wikipedia. *Gnutella2* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 2-October-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=Gnutella2&oldid=556794729>.
- [16] Wikipedia. *Gnutella* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 2-October-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=Gnutella&oldid=574304390>.
- [17] Wikipedia. *Internet service provider* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 29-September-2013]. 2013. URL: http://en.wikipedia.org/w/index.php?title=Internet_service_provider&oldid=573549991.
- [18] Wikipedia. *Kademlia* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 3-October-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=Kademlia&oldid=569941159>.
- [19] Wikipedia. *MP3.com* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 29-September-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=MP3.com&oldid=571025541>.
- [20] Wikipedia. *MP3* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 29-September-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=MP3&oldid=574123988>.
- [21] Wikipedia. *Napster* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 30-September-2013]. 2013. URL: <http://en.wikipedia.org/w/index.php?title=Napster&oldid=573999770>.

- [22] Wikipedia. *Recording Industry Association of America* — *Wikipedia, The Free Encyclopedia*. [Online; accessed 30-September-2013]. 2013. URL: http://en.wikipedia.org/w/index.php?title=Recording_Industry_Association_of_America&oldid=574192660.
- [23] Wikipedia. *Timeline of file sharing* — *Wikipedia, The Free Encyclopedia*. [Online; acessado em 28 de setembro de 2013]. 2013. URL: http://en.wikipedia.org/w/index.php?title=Timeline_of_file_sharing&oldid=571061187.