

Instituto de Matemática e Estatística
Universidade de São Paulo

**Anatomia do BitTorrent:
a Ciência da Computação por trás do protocolo**

Paulo Cheadi Haddad Filho

Orientador: José Coelho de Pina

São Paulo, 2013

Sumário

Glossário	iv
Lista de Figuras	v
1 Introdução	1
2 Histórico	2
2.1 Período pré-torrent	2
2.2 Nascimento do BitTorrent	4
2.3 Mundo pós-torrent	4
3 Anatomia do BitTorrent	5
3.1 Busca por informações	5
3.2 Fontes de arquivos	5
3.3 Jogo da troca de arquivos	5
4 BitTorrent e a Ciência da Computação	6
4.1 Estruturas de dados, listas ligadas e árvores	7
4.2 Funções de hash	7
4.3 Criptografia	7
4.4 Bitfields	7
4.5 Protocolos de redes	7
4.6 Multicast	7
4.7 Roteamento de pacotes	7
4.8 Retomada de downloads	7
4.9 Conexão com a Internet	7
4.10 Threads	7
4.11 Engenharia de Software	7
5 Comentários Finais	8
6 Referências Bibliográficas	9

Glossário

Linux is a generic term referring to the family of Unix-like computer operating systems that use the Linux kernel. [1](#)

Lista de Figuras

Capítulo 1

Introdução

Aqui vou explicar o objetivo do trabalho e o que será mostrado ao longo dele.

[Linux](#)

Capítulo 2

Histórico

Vou contar a história do BitTorrent, desde o aumento do uso de transferências de dados via internet, passando pela criação de protocolos e softwares usados para baixar arquivos de forma ilegal, até a sua criação e o boom de usuários que o utilizam para essa motivação. Em paralelo, discutirei os motivos legais pelos quais os softwares anteriores foram descontinuados, que influenciaram diretamente na sua criação.

Separarei em 3 períodos: o antes, a criação e as consequências disso no mundo.

2.1 Período pré-torrent

Para entender como e por que o BitTorrent se tornou o que é hoje, devemos voltar um pouco no tempo e rever a história recente do compartilhamento de arquivos pela Internet.

Primeiro vieram as BBSs (1978) e a Usenet (1979), depois o protocolo FTP (1985) e seus servidores. O IRC (1988) e o Hotline (1997) ainda permitiam que seus usuários se comunicassem e enviassem arquivos. Até o fim dos anos 1990, esses métodos usados para transferir arquivos seguiam o modelo cliente-servidor. Neste modelo, o cliente é um computador que faz uma requisição de serviço ou recurso para um servidor, que se dedica a aguardar pedidos de clientes.

Com o advento da codificação MP3, os arquivos de música em formato digital passaram a ser largamente utilizados na segunda metade dos anos 1990, devido à grande redução de tamanho de arquivos de áudio nesse formato quando comparado aos outros formatos contemporâneos.

Então, em 1998, dois sites de compartilhamento de músicas foram criados: o AudioGalaxy.com e o MP3.com, ambos sites de busca de MP3 sobre arquivos que seus usuários faziam upload. Ambos os sites ajudaram na popularização do MP3 como forma de se escutar músicas no dia a dia, que foi catalisado quando indústrias de equipamentos eletrônicos anunciaram seus sucessores do toca-fitas: os tocadores de mídia portáteis, popularmente conhecidos como MP3 players.

Não demorou muito tempo para a indústria da música, representada pela RIAA (Recording Industry Association of America, Associação da Indústria de Gravação da América) começar a enxergar essa popularização do formato MP3 como um perigo real de diminuição de seus lucros com as vendas de discos, e então passou a encabeçar a frente de combate contra a troca de arquivos protegidos por direitos autorais pela Internet. Pouco tempo depois, durante o governo Clinton, o Senado americano aprova uma lei de proteção a conteúdos com direitos autorais, estabelecendo regras mais específicas sobre esses direitos autorais e se tornando obstáculo àqueles sites de compartilhamento por penalizar os usuários por infringir esses direitos.

No segundo semestre de 1999 surgiu o Napster, um serviço de compartilhamento de MP3 que inovou por desfigurar o modelo cliente-servidor ao possuir conexões entre usuários e não a servidores, criando assim a primeira rede P2P.

O sucesso foi rápido, e em dezembro de 1999 já enfrentava sua primeira ação judicial, vinda de várias grandes gravadoras. Sua popularidade foi crescendo, atingindo o auge de 13.6 milhões de usuários em fevereiro de 2001. A sua vida começou a complicar em julho de 2001, quando foi iniciada uma ação judicial por responsabilidade na distribuição de conteúdo protegido. O caso foi finalizado em setembro de 2001, obrigando a desligar o serviço de indexação, que tornou a rede P2P indisponível, e ainda a pagar 26 milhões de dólares por danos já causados e também pelos futuros, fazendo com que pedisse falência.

O sucesso do Napster, mesmo que por curto período tempo, mostrou a eficiência que redes P2P poderiam ter, e com isso novos softwares e modelos de redes foram sendo lançados, porém tentando contornar o ponto fraco do antecessor a fim de não serem novos alvos de medidas judiciais. A solução para isso foi tentar descentralizar o mecanismo de indexação e busca, que foi o calcanhar de Aquiles do Napster.

A pioneira nessa tentativa foi a rede Gnutella, que foi lançada em 2003 mas que em sua primeira versão não conseguiu manter o bom desempenho do Napster. A busca era demorada e inconsistente, pois era repassada aleatória e finitamente de peers para seus vizinhos, o que podia terminar em buscas sem resultados mesmo quando um arquivo estava sendo compartilhado por alguém conectado à rede.

continua...

2.2 Nascimento do BitTorrent

Aqui contarei a história do surgimento do protocolo.

2.3 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 3

Anatomia do BitTorrent

Aqui mostrarei como funciona o BitTorrent de forma linear, como se fosse um humano o utilizando.

3.1 Busca por informações

Falarei quais informações um arquivo torrent pode conter e como o Transmission as utiliza, tentando mostrar detalhes da troca de mensagens dele com servidores de rastreamento.

3.2 Fontes de arquivos

Mostrarei o processamento dos dados adquiridos na seção anterior e como ele organiza a lista das fontes de arquivos usando a tabela hash DHT Kademlia.

3.3 Jogo da troca de arquivos

Explicarei o algoritmo tit-for-tat padrão do protocolo BitTorrent, que vem da Teoria dos Jogos, e como o Transmission o implementa.

Capítulo 4

BitTorrent e a Ciência da Computação

Aqui mostrarei detalhes técnicos sobre as partes coadjuvantes do BitTorrent e do Transmission.

4.1 Estruturas de dados, listas ligadas e árvores

4.2 Funções de hash

4.3 Criptografia

4.4 Bitfields

4.5 Protocolos de redes

4.6 Multicast

4.7 Roteamento de pacotes

4.8 Retomada de downloads

4.9 Conexão com a Internet

4.10 Threads

4.11 Engenharia de Software

Capítulo 5

Comentários Finais

Capítulo 6

Referências Bibliográficas

Capítulo 7

Visão Pessoal