

Resumo do artigo

Unraveling the BitTorrent Ecosystem

Tiago Funk¹

¹Programa de pós-graduação em computação aplicada
Universidade do Estado de Santa Catarina (UDESC)
Joinville – SC – Brasil

tiagoff.tf@gmail.com

Perguntas rápidas

Qual o objetivo do artigo?

Coletar estatísticas sobre a rede BitTorrent afim de descobrir como tudo funciona. A intenção é analisar arquivos, usuários, infraestrutura, ferramentas que são utilizadas para manter a rede funcionando.

Qual a justificativa do autor para a sua proposta?

A rede do BitTorrent é distribuída, isso dificulta fazer um estudo que seja abrangente, por isso não existia um trabalho que fez um mapeamento completo da rede até então. O artigo coletou informações sobre tantos quesitos para justamente ter a visão mais geral possível sobre o ecossistema BitTorrent.

Qual a contribuição do autor em relação aos trabalhos relacionados da área?

Trabalhos anteriores não fizeram uma análise tão completa quanto este artigo. Geralmente buscavam analisar um conceito apenas. Em contra partida, este artigo foi mais abrangente, coletou dados de mais *sites*, coletou dados sobre arquivos, infraestrutura, usuários, etc. Os autores também realizaram uma análise de impacto da infraestrutura do Pirate Bay no ecossistema.

Qual a proposta do autor? O que ele realmente está propondo no artigo?

O artigo tem a proposta de levantar informações de como o ecossistema funciona. Desenvolveram uma ferramenta que visitava os sites de indexação de *torrents* e fazia a coleta dos dados. A partir dessas informações, foi feita a análise dos dados. Informações sobre distribuição geográfica, motivação de usuários, características sobre os arquivos, etc, foram apresentadas.

Quais os materiais e métodos usados pelo autor para elaborar sua proposta?

Os autores desenvolveram uma ferramenta de coleta de dados. Quando trabalharam sobre os dados, separam em dados sobre *trackers*, pares *peers* e conteúdo. Os dados são apresentados em gráficos, tabelas.

Quais os resultados, discussões e conclusões apresentados no artigo?

Além da apresentação dos dados sobre as características do ecossistema, descobriu-se que existe uma grande diversidade em vários aspectos. Porém existe grande centralização da rede em torno da plataforma Pirate Bay. Se por acaso esta plataforma for

desativada, o resto da rede pode não conseguir absorver a nova demanda. Mas ferramentas distribuídas podem ser uma solução que esse evento catastrófico vir a acontecer.

Críticas ao artigo

O artigo apresenta os objetivos de forma bem clara, é bem escrito e claro na proposta. Apresenta os dados bem em gráficos e tabelas. Separa os dados em categorias para facilitar a apresentação. É um artigo bom e bem escrito. O único defeito foi que algumas siglas não foram apresentadas, por exemplo, a sigla CDF não foi apresentada e aparece várias vezes no artigo e em gráficos.

Resumo do artigo

O artigo tem intenção de estudar o ecossistema do *BitTorrent*. ou seja, analisar todas as ferramentas que dão suporte, como descobridores de torrent, trackers, peers, comportamento de uso e dados do conteúdo da rede. O artigo foca em seis pontos:

- Desenvolveram um *multitracker crawler* que pode simultaneamente visitar centenas de *trackers*.
- Estudaram cinco sites para coletar informações sobre o ecossistema. Além disso, descobriram características e motivações dos usuários que subiram os arquivos e como os sites adquirem os arquivos.
- Estudaram o ecossistema de *trackers*. Identificaram 39000, dos quais apenas 728 eram ativos.
- Estudaram o ecossistema dos pares. Muito dos *torrents* são muito pequenos (82 % não tem mais que 10 pares). Em contra partida, os maiores tem mais 10000 pares simultâneos. Estudaram também a motivação das pessoas que criaram o arquivo.
- Estudaram o ecossistema do conteúdo. Classificaram o conteúdo em 10 categorias e realizaram a análise geográfica.
- Por mais que o ecossistema seja distribuído, mais de 90 % dos arquivos recebem *tracking* do Pirate Bay's. Então analisaram a fragilidade do sistema se o Pirate Bay for removido.

Uma rede BitTorrent possui três componentes básicos. O primeiro é o par (peer). Quando uma coleção de pares participam da distribuição de um arquivo por um tempo determinado, isto é chamado de *torrent*. Em qualquer momento, pares podem participar sendo um receptor ou semeador. Quando é um semeador, o par possui o arquivo inteiro e envia partes dele para outros pares. Quando é um receptor, o par possui apenas uma parte do arquivo e vai recebendo partes de outros.

O segundo componente é o mecanismo de descoberta de pares. O mais comum é o *tracker*. Ele é uma ferramenta que conecta vários pares (peers). Quando um par quer enviar ou receber um arquivo pelo sistema *BitTorrent*, se inscreve em um *tracker*. A ferramenta então começa a enviar informações para que pares possam trocar pedaços do *torrent*. Quando um usuário quer iniciar um novo *torrent*, precisa semear o conteúdo do arquivo e registrar o torrent em um *tracker*, assim criando um arquivo .torrent e este pode ser disponibilizado na internet para terceiros o acessarem.

O terceiro componente é o *site* de descoberta de *torrents*. São sites que fornecem os arquivos .torrent. Existem centenas de *sites* que fornecem esse serviço, como o Pirate Bay, Mininova, Isohunt, BTmonster, etc. Existem dois tipos, os públicos, onde qualquer

pessoa podem baixar o arquivo .torrent e receber o arquivo. Os privados implementam um controle de acesso de pessoas e geralmente é necessário ser convidado para entrar. Importante dizer que *sites* de descoberta não necessariamente fornecem *trackers*.

Os autores, ao criarem o *multitracker crawler*, realizaram a extração de dados de cinco dos mais populares *sites* de *torrents*. Como cada site usava diferentes tipos para representar o formato de dados, foram implementados diferentes *parsers* para extração de metadados dos arquivos. O Artigo também tinha a intenção de coletar informações sobre os pares, calculando uma estimativa da quantidade de conexões. Devido a grande quantidade de requisições simultâneas que precisariam ser atendidas para encontrar o número de pares, foi implementado um modelo assíncrono de I/O em um modelo de *thread-pool*.

A ferramenta descobriu 4,6 milhões de arquivos únicos, dos quais 1,2 milhões eram ativos. Ao analisar os dados, descobriu-se que uma grande porcentagem dos arquivos .torrent são compartilhados entre sites. Os sites recebem novos arquivos por *upload* de usuários (Alguns sites utilizam mais esta forma) ou por mineração de arquivos de sites similares (Outros sites utilizam mais esta forma).

Ao analisar os dados de *uploaders* de arquivos, descobriu-se que 55 % dos usuários subiram apenas um arquivo. 90 % não subiram mais que 10 arquivos. Porém existem usuários com mais que 1000 arquivos. A principal motivação para subir os arquivos é o altruísmo de compartilhar. Outra análise foi a de número de novos arquivos adicionados por dia. Alguns *sites* possuem dias que recebem a adição de arquivos muito acima do normal dos outros dias, indicando os possíveis dias que ocorreu a mineração dos arquivos .torrent pelo site.

As estatísticas coletadas sobre os *trackers* mostram que existem 38996, dos quais 728 são ativos. A maioria dos *trackers* são criados por usuários em suas máquinas locais com seus clientes locais. Ao analisar as maiores organizações de *trackers*, o Pirate Bay lidera com larga vantagem nos quesitos de números de pares trackeados e número de arquivos trackeados. Analisando apenas os *trackers*, existe uma grande diversidade, com 40 % dos *trackers* com no máximo 4 arquivos e 26 % possui mais que 1000 arquivos. Os Estados Unidos possuem o maior número, enquanto os maiores ficam na Suécia.

Para coletar estatísticas sobre os pares, no dia 22 de abril de 2009, durante 12 horas, foram coletados dados com a ferramenta. O resultados são:

- O total de pares únicos foi de 5085217.
- 82 % dos arquivos não tinha mais que 10 pares, enquanto apenas 1 % tinha mais de 100 pares. 22 arquivos tinha mais que 10000 pares e um tinha mais que 150000 pares.
- Os arquivos apresentam tamanhos diversos.
- 56 % dos pares apenas se juntam em apenas um *torrent*. Enquanto 44 % se juntam em mais que um (4 % se juntam em mais que 10 arquivos).
- Os Estados Unidos foi o país com mais pares. A maioria dos países com mais pares estão na América do Norte ou Europa. Nem a China e nem a Índia aparecem.
- Os países com a proporção de mais usuários de BitTorrent por usuários de Internet são o Emirados Árabes Unidos, Singapura e Canadá.
- O cliente uTorrent é o mais utilizado pelos pares.

- O cliente uTorrent é o mais utilizado para criar arquivos.

Ao realizar uma análise sobre a popularidade versus idade, descobriu-se que arquivos com até 40 semanas de idade, a sua popularidade via caindo em uma taxa constante. Após as 40 semanas, a popularidade não se altera muito.

O artigo também fez a classificação de acordo com o conteúdo do *torrent*. Filmes, música e *shows* de TV são as categorias mais comuns. As categorias Livros, Jogos, Aplicações e Pornografia também são comuns. Os Emirados Árabes Unidos é o país com mais pares em arquivos de Filmes e Pornografia. A Noruega aparece na frente na categoria Música.

A última análise feita pelo artigo é determinar a importância do Pirate Bay no ecossistema. A pergunta é a seguinte: se o sistema do Pirate Bay sofrer um ataque ou cair, o ecossistema vai continuar funcionando?

No quesito do site de descoberto de torrents, apenas 3,6 % dos arquivos que estão indexados no Pirate Bay não estão em outros *sites*. Não gerando um grande impacto no ecossistema. Porém, a remoção dos *trackers* é um problema mais sério. 90 % dos arquivos ativos são trackeados pelo Pirate Bay, mas apenas 23 % são exclusivos. Outros sistemas poderiam continuar trackeando os arquivos, mas é difícil dizer se a infraestrutura daria conta de absorver todos os pares. A notícia boa é que a maioria dos clientes suporta *trackers* distribuídos como o DHT e PEX.

Para finalizar, foram coletadas estatísticas sobre as ferramentas necessárias para o ecossistema BitTorrent funcionar. As conclusões sobre o artigo são que existe uma diversidade grande entre pares, trackers, características dos arquivos, países que utilizam, etc. Porém, a plataforma do Pirate Bay desempenha um grande papel, que se ele for removido, pode impactar negativamente toda o ecossistema.