Computação Distribuída

Funcionamento do BitTorrent

Vladimir Rocha (Vladi)

CMCC - Universidade Federal do ABC

Disclaimer

- Estes slides foram baseados no tutorial do professor Sukumar Gosh
  http://www.cs.uiowa.edu/~ghosh/bittorrent.ppt

Com permissão do autor

O problema

- A distribuição de um conteúdo estático grande, a partir de um emissor (fonte), para um grande número de usuários, tão rápido quanto possível

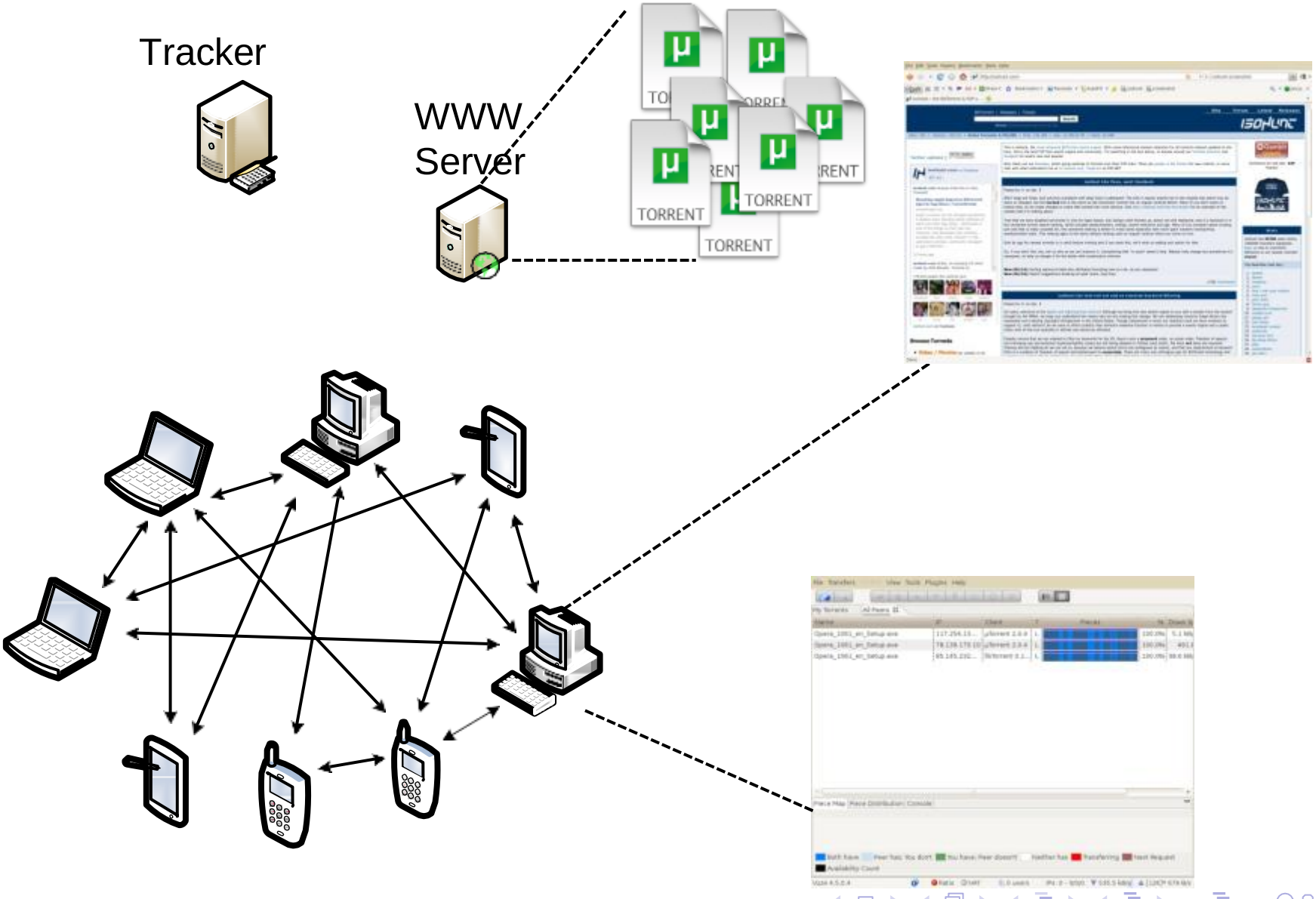
- Usar somente a largura de banda de upload do emissor é muito caro e as vezes inviável

- Soluções?

Ideia de solução

- Utilizar a capacidade de upload dos que baixam (que fazem download)
- Criar oportunidades para intercambiar os dados entre os que baixam

Sistema BitTorrent



Sistema BitTorrent

O sistema BitTorrent para distribuição de arquivos consiste em:

- Um arquivo estático de metadados (torrent file) do conteúdo a ser compartilhado.
- Um servidor WWW para publicar os arquivos torrent
- Um tracker
- Um peer que contém e publica o conteúdo (i.e.,original seed).
    - Ele poderá sair da rede se o conteúdo estiver disponível entre os outros peers (leechers & seeders) do sistema
- Um navegador WWW para encontrar o arquivo torrent
- Uma aplicação que permita gerenciar os arquivos torrent e baixar o conteúdo na rede BitTorrent

Arquivo Torrent

- Usado para encontrar o conteúdo (via tracker) e verificar a integridade
- Inclui hashes das peças
  Conteúdo particionado em peças com os hash calculados
- Outras infos: tamanho das peças, nome do arquivo, URL do tracker, etc.

```
<?xml version="1.0" encoding="UTF-8"?>
<tor:TORRENT
    xmlns:tor=http://azureus.sf.net
    /files
    xmlns:xsi="http://www.w3.org/XM
    LSchema>
<ANNOUNCE_URL>http://torrent.opera.com:6969/ann
ounce</ANNOUNCE_URL>
<CREATION_DATE>1281694453</CREATION_DATE>
<TORRENT_HASH>2EEAB52F02423D2F0C</TORRENT_HASH>
<INFO>
 <NAME encoding="utf8">Opera_Setup.exe</NAME>
 <PIECE_LENGTH>262144</PIECE_LENGTH>
 <LENGTH>10809256</LENGTH>
 <PIECES>
  <BYTES>54E6CCEC9EB3BEA12711023650D5AF</BYTES>
  <BYTES>D134F302346EFB60E57BD27A408C35</BYTES>
 </PIECES>
</INFO>
</tor:TORRENT>
```

Tracker

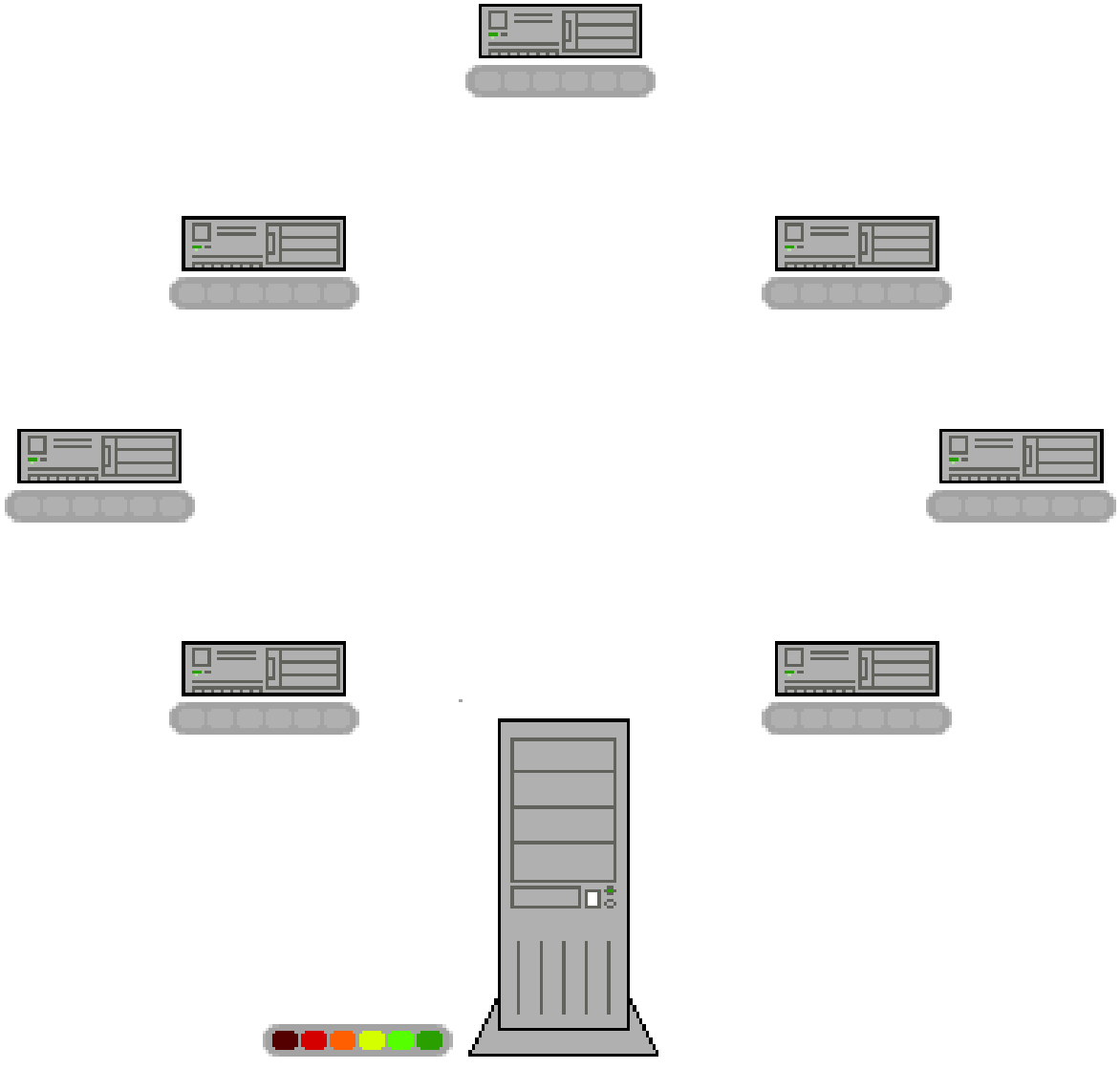
- Utilizado para encontrar outros peers interessados no mesmo torrent
  (a partir de aqui, conteúdo e torrent terão o mesmo significado)
- Armazena as informações de contato do peers
- Utiliza o protocolo HTTP para comunicar-se com os peers
- É contatado quando o peer começa, atualiza, para ou baixa completamente o conteúdo
  - Anúncios são realizados a intervalos regulares pelos peers
  - O tracker possui um timeout para remover os peers que saíram

- Ponto único de falhas (**Single Point of Failure – SPoF**)
  - Alvo de ações judiciais (tracker do pirate bay foi eliminado)

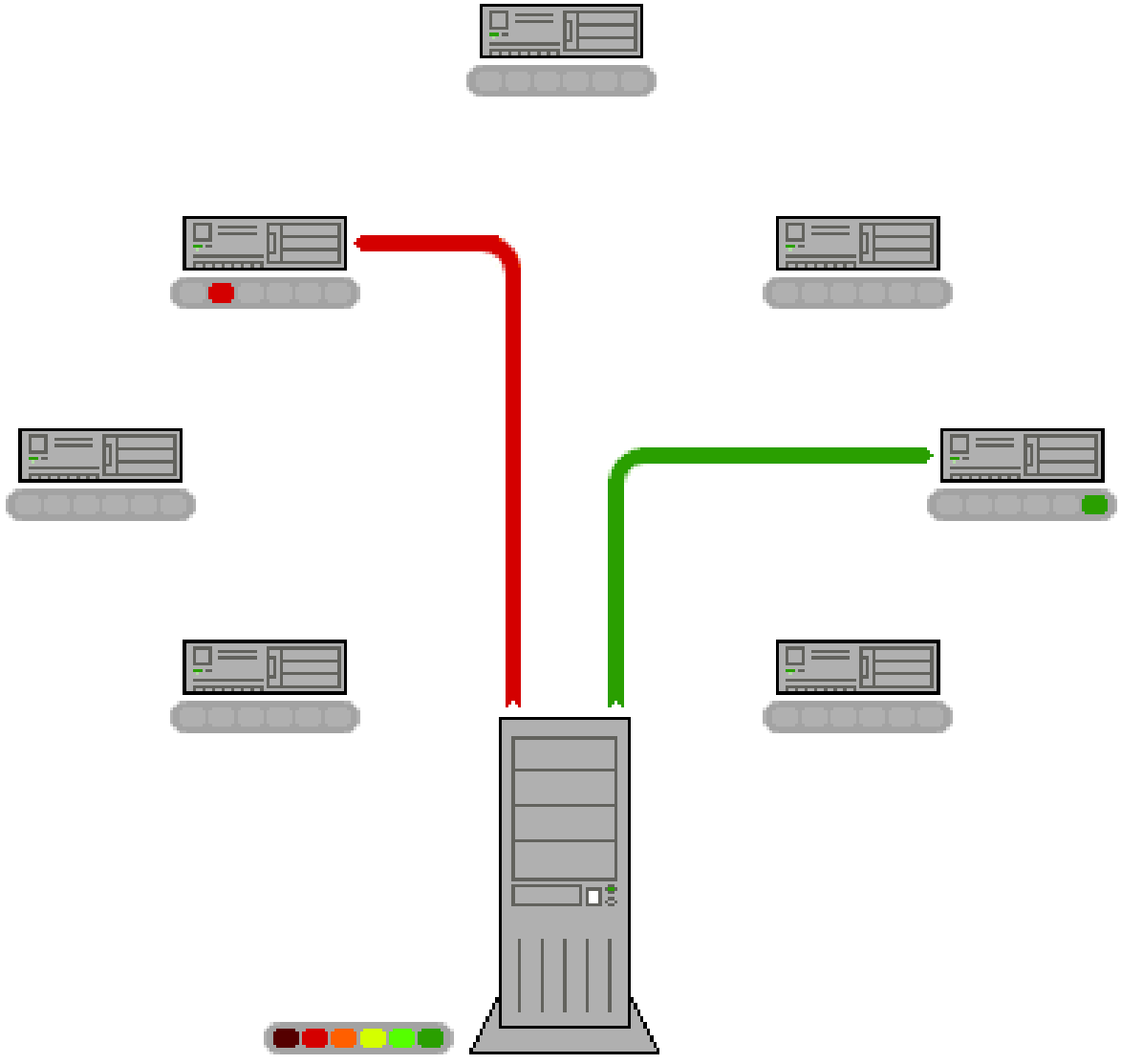
Protocolo para download

- Os peers se comunicam utilizando o protocolo BitTorrent acima do TCP
- Os peers interessados no mesmo torrent formam uma rede overlay independente denominada swarm
- Uma vez que o peer baixa o conteúdo completo pode manter-se como seed (altruistic) ou sair da rede (selfish)

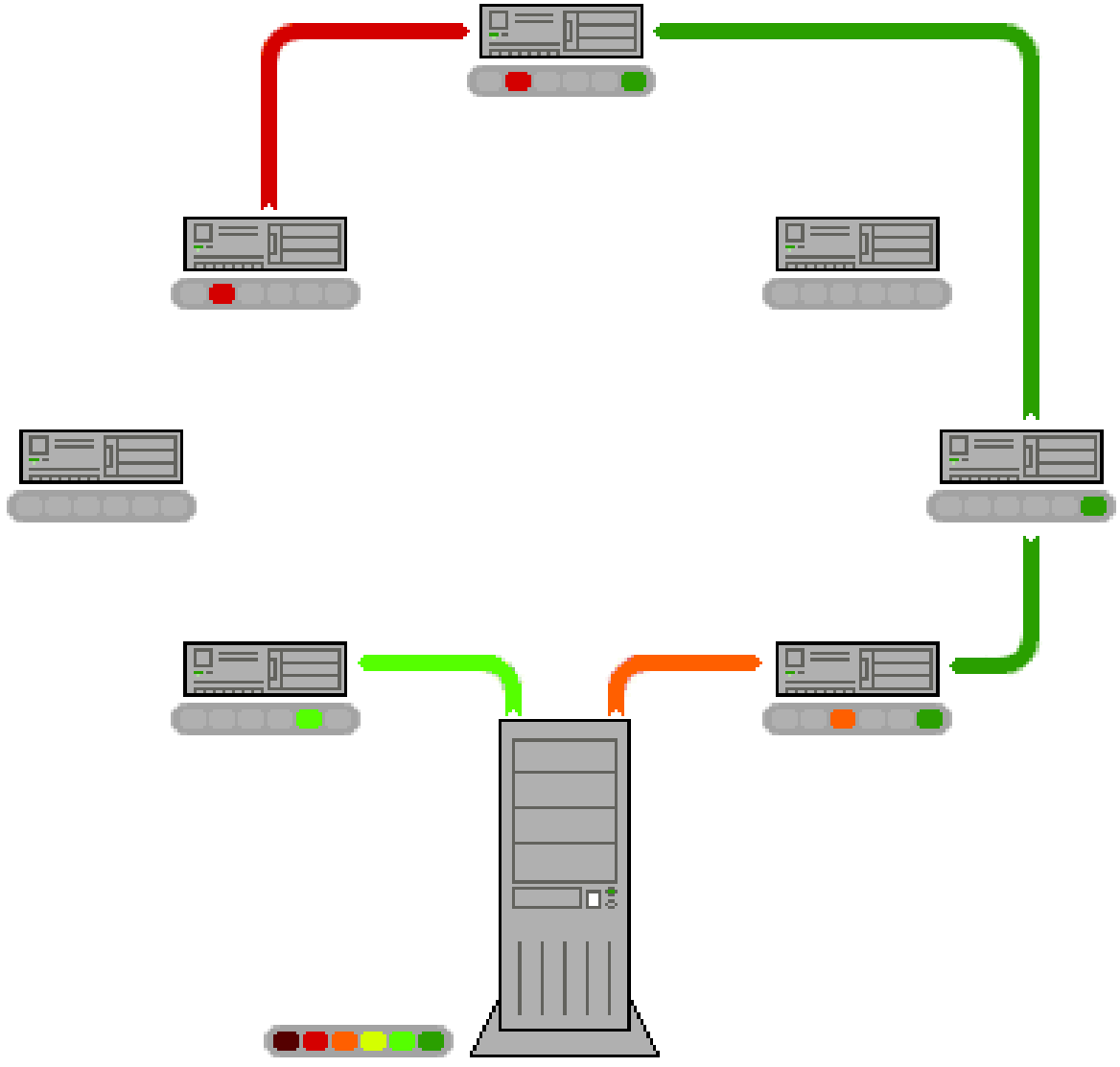
Funcionamento



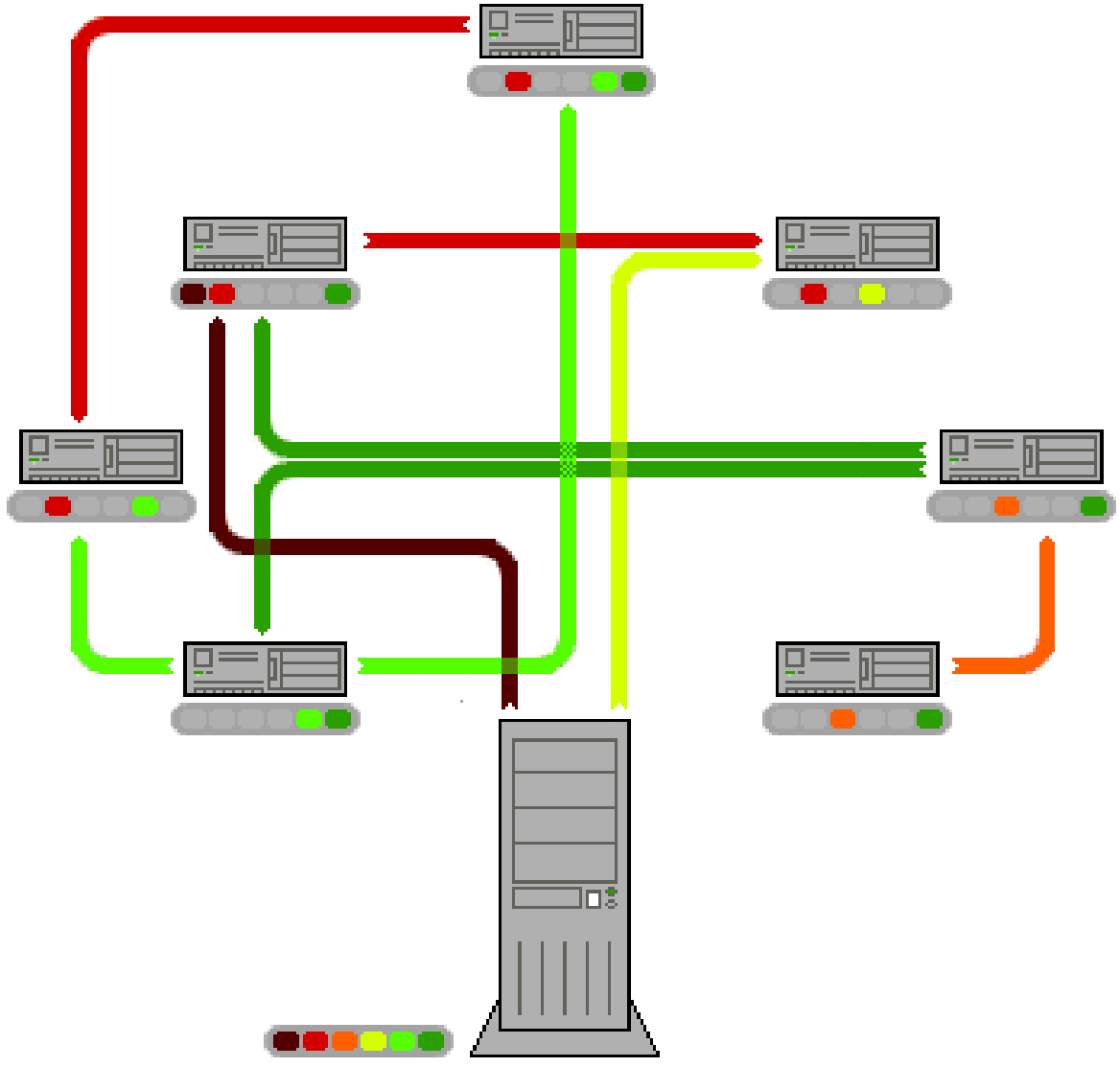
Funcionamento



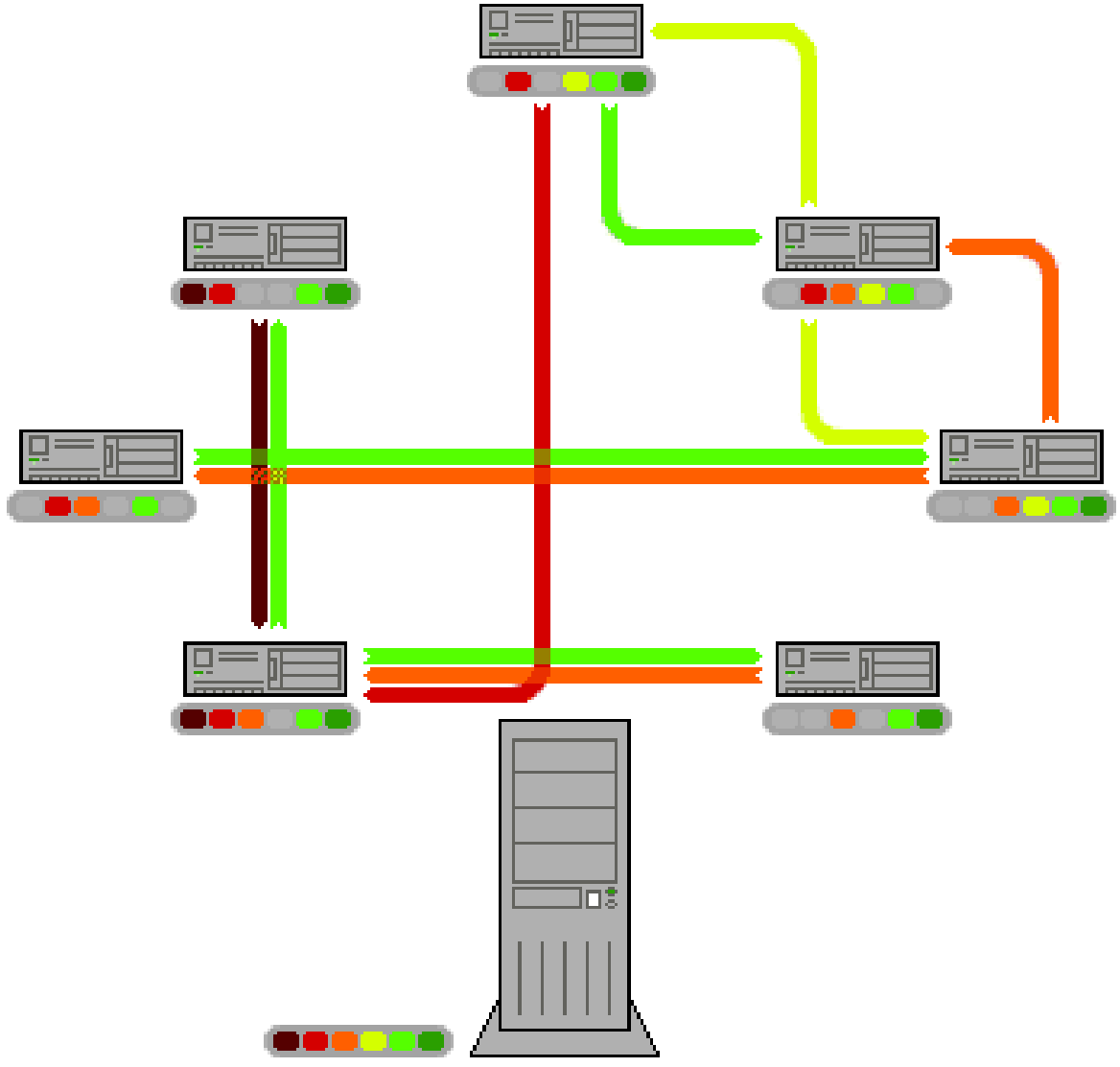
Funcionamento



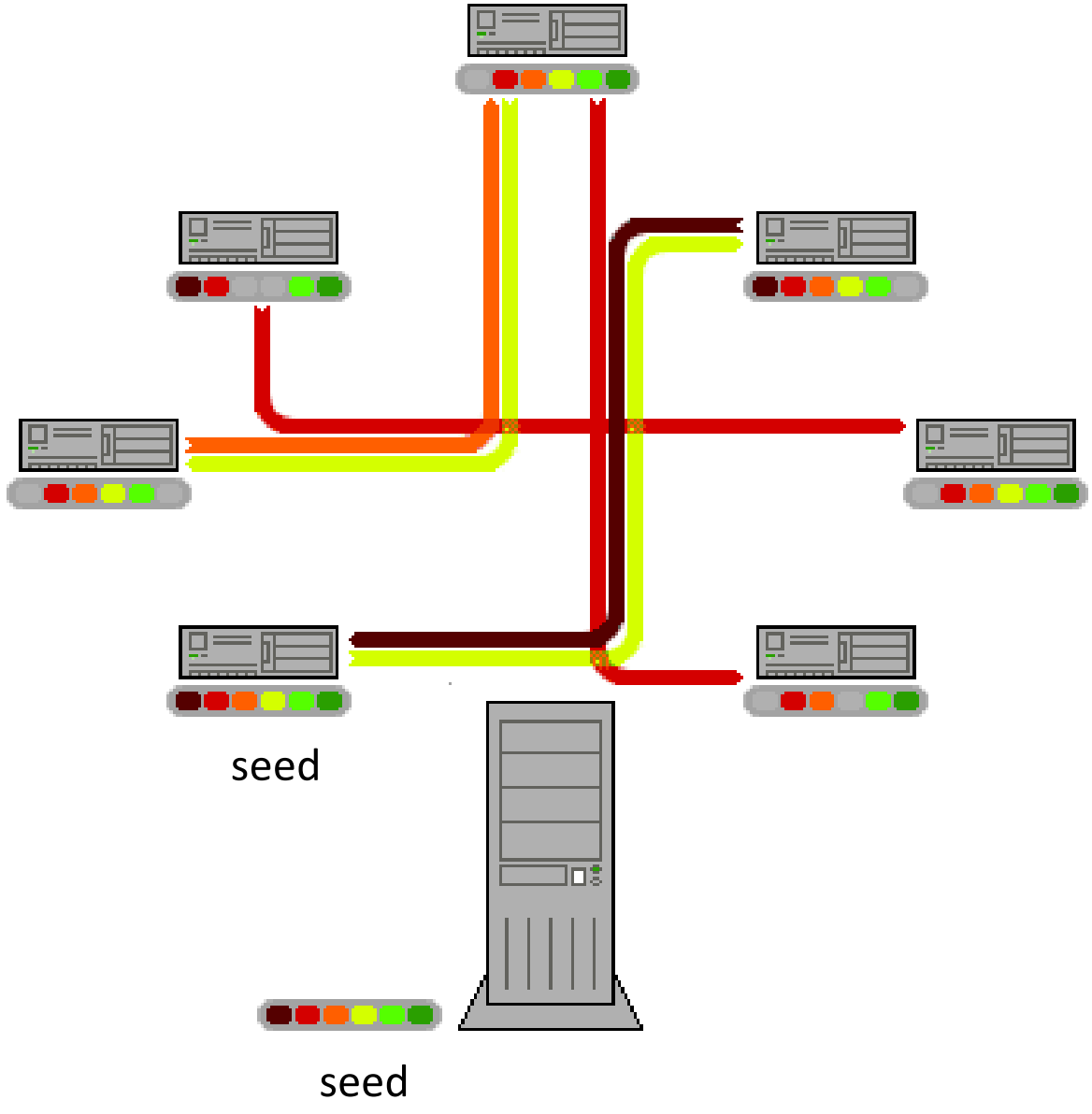
Funcionamento



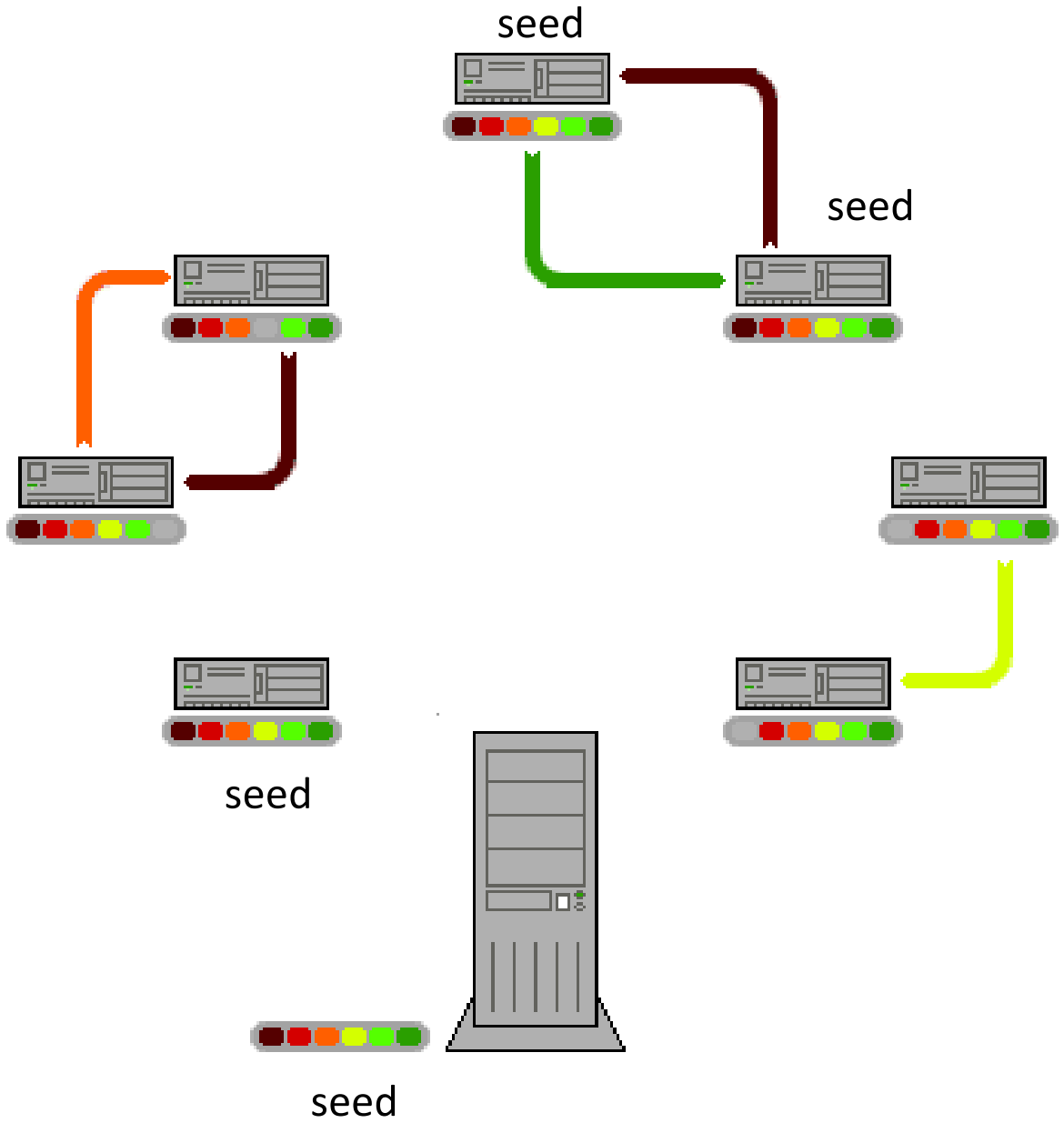
Funcionamento



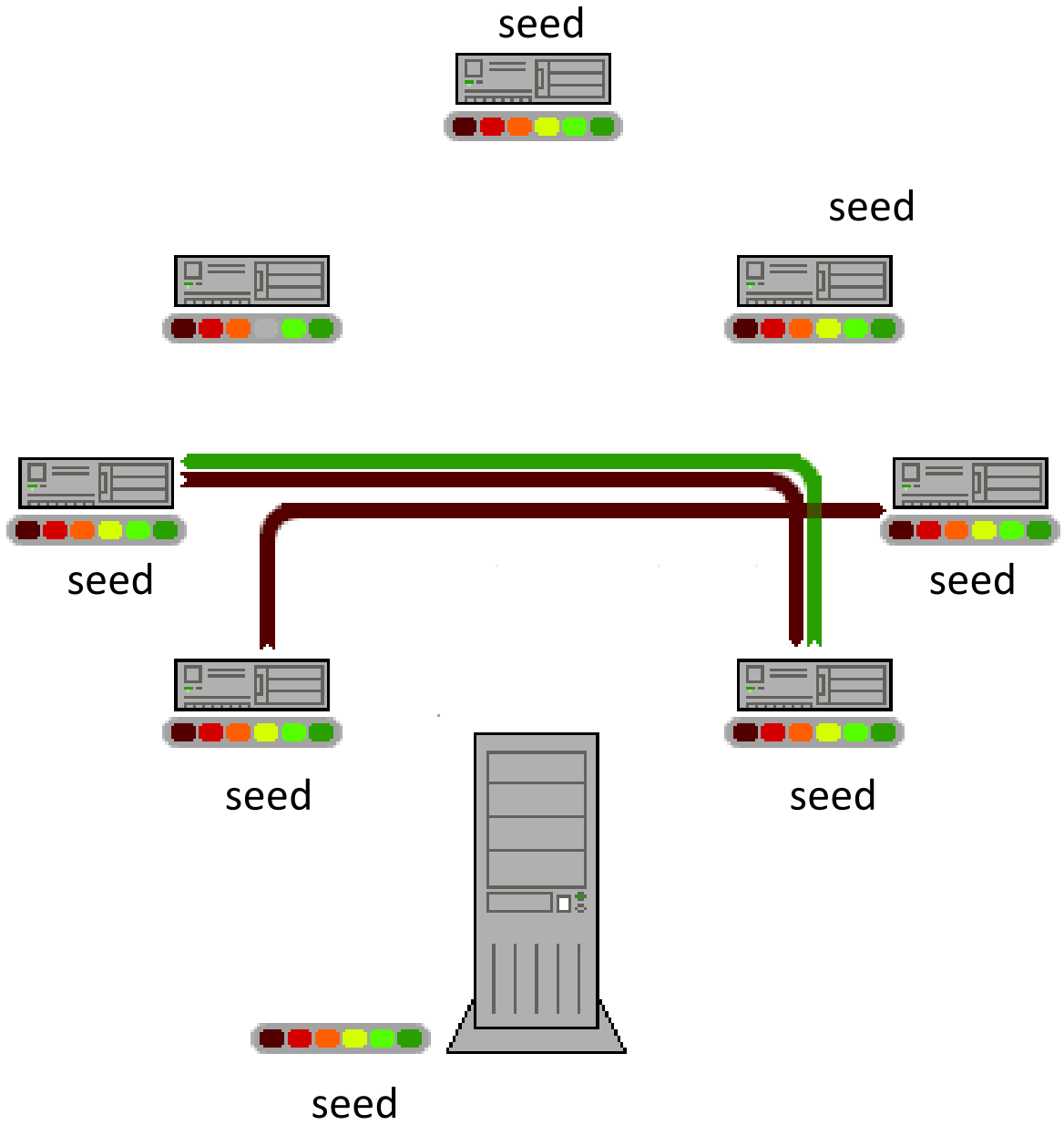
Funcionamento



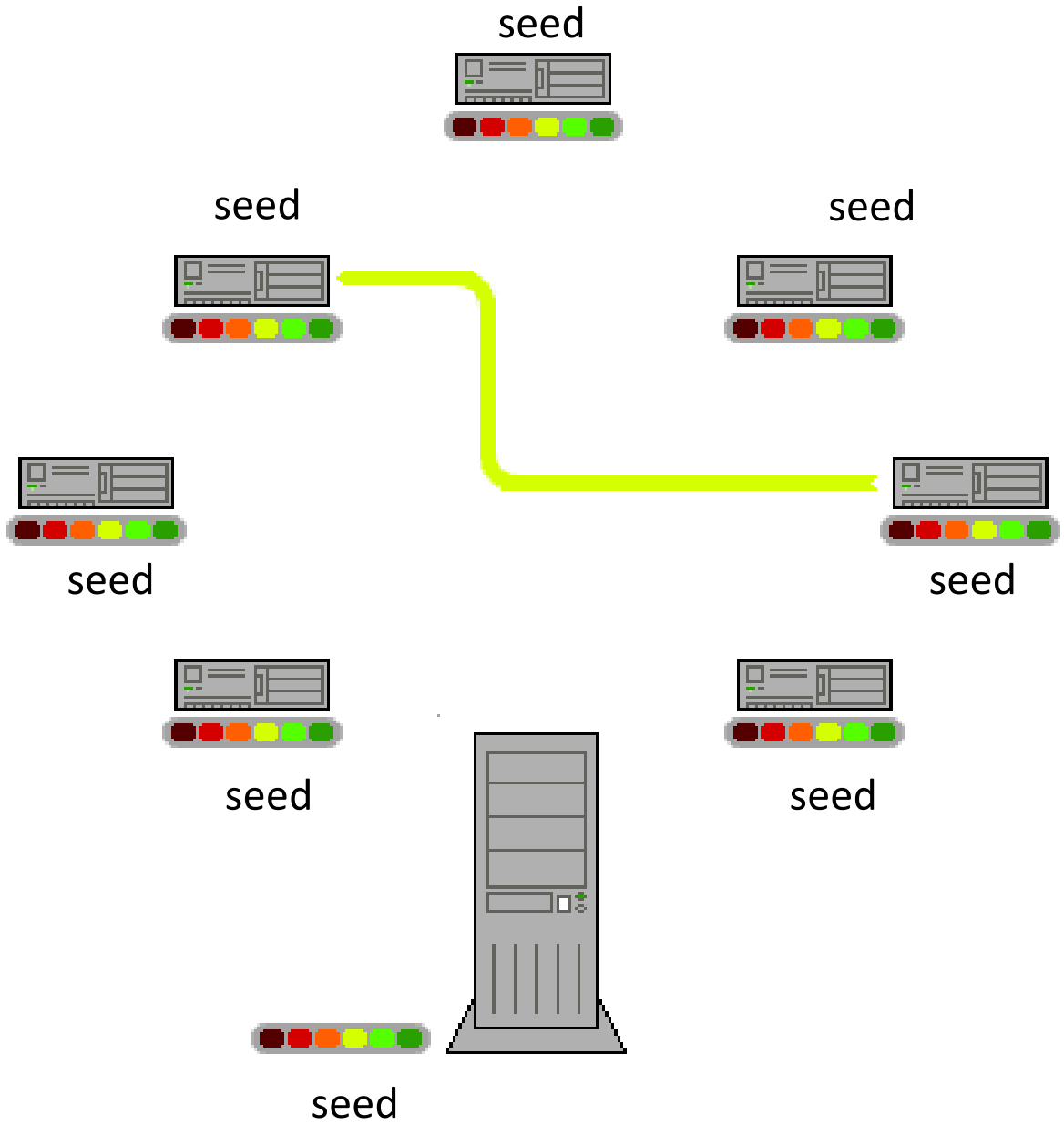
Funcionamento



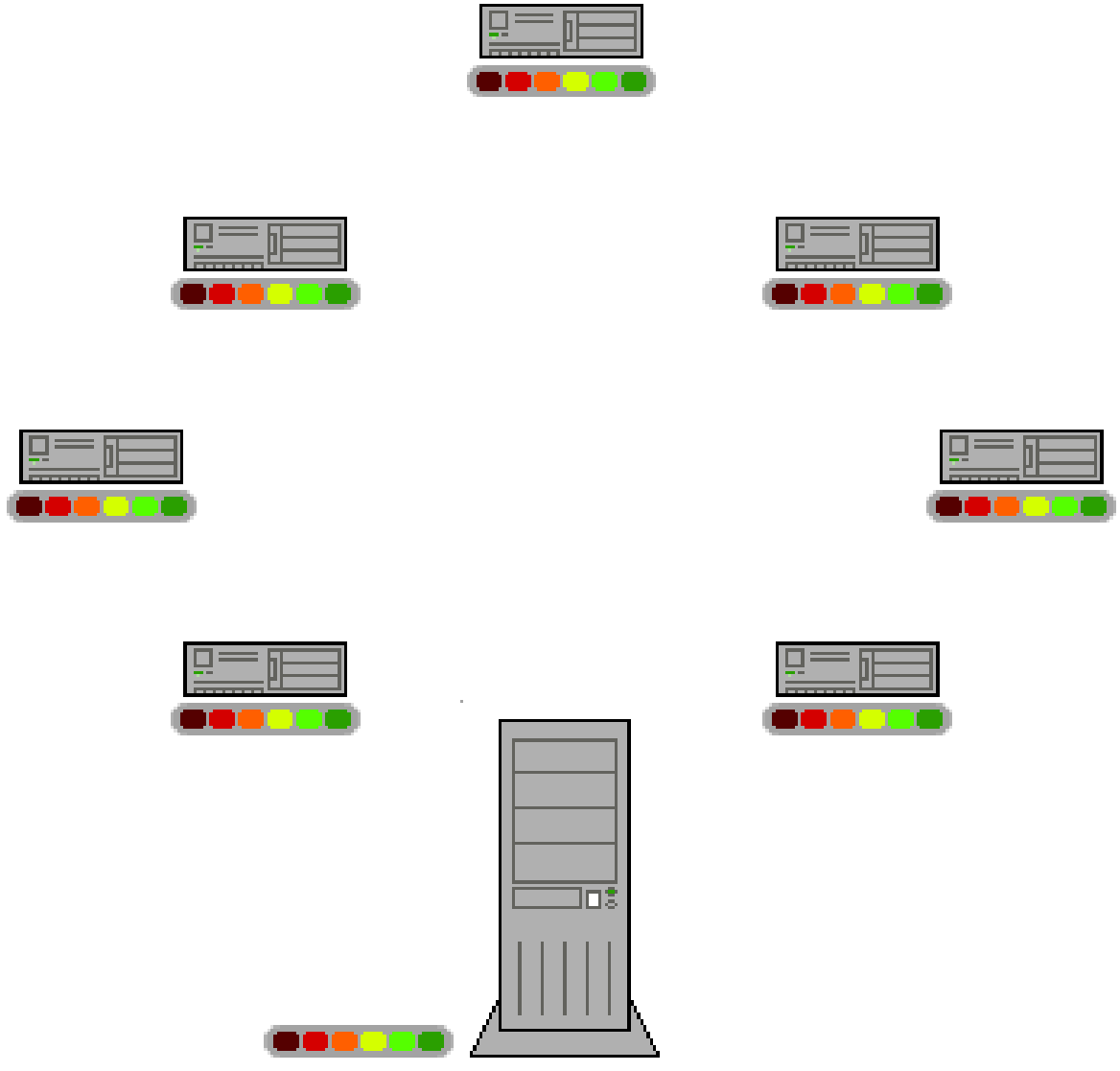
Funcionamento



Funcionamento



Funcionamento



BitTorrent em LA (2016)

Upstream		Downstream		Aggregate	
BitTorrent	30.03%	YouTube	28.48%	YouTube	25.91%
YouTube	9.30%	HTTP - OTHER	11.66%	HTTP - OTHER	11.12%
HTTP - OTHER	7.59%	SSL - OTHER	9.76%	BitTorrent	10.06%
Facebook	6.72%	Netflix	8.31%	SSL - OTHER	9.28%
SSL - OTHER	6.19%	BitTorrent	6.96%	Netflix	7.45%
Ares	5.27%	Facebook	5.10%	Facebook	5.32%
Skype	2.53%	MPEG - OTHER	2.28%	MPEG - OTHER	2.10%
Netflix	1.97%	RTMP	1.79%	RTMP	1.66%
Dropbox	1.16%	Google Market	1.69%	Google Market	1.52%
MPEG - OTHER	0.92%	Flash Video	1.60%	Flash Video	1.46%
	71.69%		77.63%		75.87%

sandvine

Table 3 - Top 10 Peak Period Applications - Latin America, Fixed Access

BitTorrent no mundo

Upstream

2011: 52% (poucas plataformas de streaming)

2015: 27% (Netflix)

2018: 32% (Netflix, HBO, Amazon, Disney)

2022: 10% (... + TikTok)

O custo envolvido em assinar todas as plataformas aumentou o uso do BitTorrent, mas diminuiu novamente – mudança comportamento.

https://entretenimento.uol.com.br/noticias/redacao/2019/02/09/streaming-netflix.htm
https://torrentfreak.com/bittorrent-is-still-the-king-of-upstream-internet-traffic-but-for-how-long-220304/

Estratégias do BitTorrent

- Tentar que a taxa de download de cada peer seja proporcional a sua taxa de upload
    - Ajuda a evitar os free-riders
- Usar um método eficiente para a distribuição do arquivo (Tit-for-Tat)

Estratégias do BitTorrent

- Dilema do Prisioneiro
  Houve um assassinato
  Prisioneiro A e B em celas separadas (i.e. não sabe o que o outro falará).
  Policiais não sabem quem é o culpável (ou ambos)

	Prisioneiro "B" nega	Prisioneiro "B" delata
Prisioneiro "A" nega	Ambos são condenados a 6 meses	"A" é condenado a 10 anos; "B" sai livre
Prisioneiro "A" delata	"A" sai livre; "B" é condenado a 10 anos	Ambos são condenados a 5 anos

O que você faria?
O "insight" genial, como usar isso no BitTorrent?

Tit for Tat

- A melhor estratégia para o dilema do prisioneiro iterativo.

  Basicamente o agente:
  - a menos que seja provocado, coopera
  - se provocado, retalia
  - perdoa (esquece) rapidamente
  - sabe que terá várias chances de encontrar-se novamente com outros

    por isso o "iterativo"

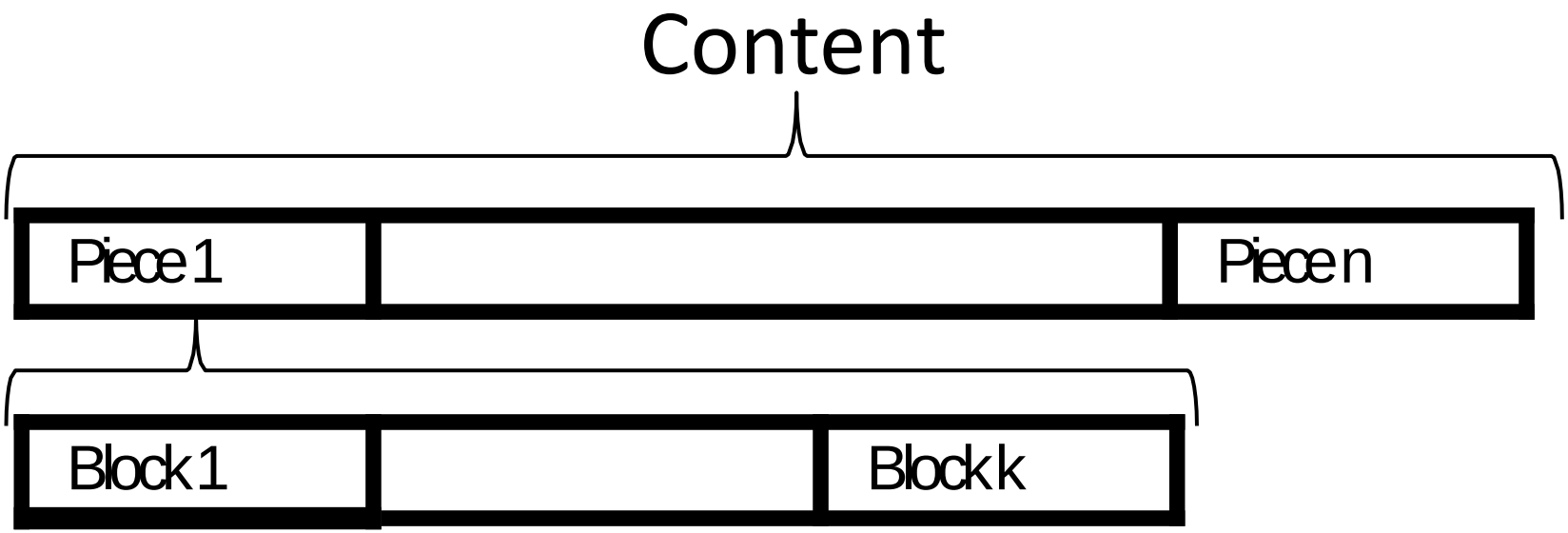
https://cs.stanford.edu/people/eroberts/courses/soco/projects/1998-99/game-theory/axelrod.html

Publicação do conteúdo

- Divide o conteúdo em peças, calcula os hashes para cada uma, e cria o arquivo torrent com os metadados
- Registra o torrent no tracker
- Inicia o aplicativo BitTorrent atuando como um seed
- Publica o arquivo torrent em um servidor Web

Estratégias do BitTorrent

- O conteúdo é dividido em peças (256KB-2MB)
- Cada peça é dividida em blocos (16KB)



Join no swarm

- O peer encontra o arquivo de metadados torrent no servidor WWW
- O tracker lhe devolve uma lista com os peers que estão em um swarm (50 peers aleatórios)
    - Corresponde ao Neighbor Set (vizinhos)
- Note que o tracker é um componente (servidor) centralizado mas não está envolvido na transferência do conteúdo

Peers Vizinhos (Neighbors)

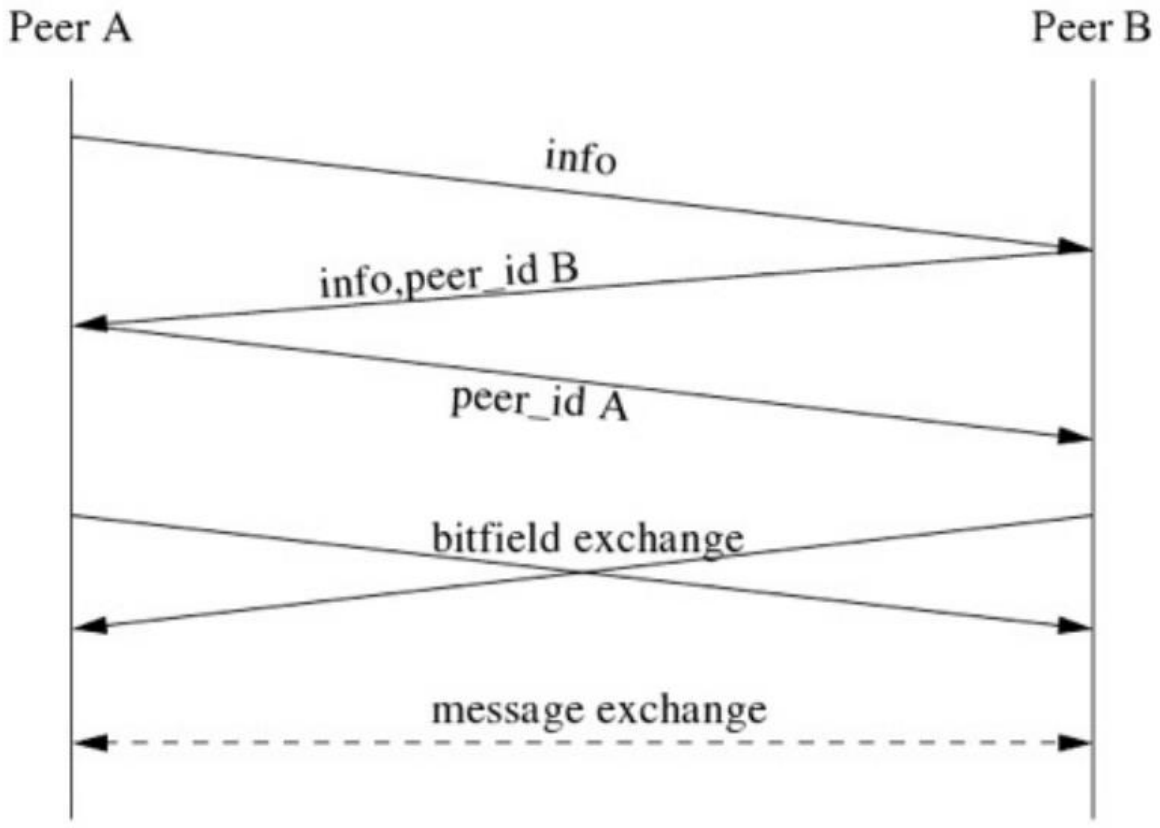
- Cada peer possui um neighbor set
  - Inicialmente obtido do tracker
  - Com tamanho máximo de 80
- Cada peer abre uma conexão TCP com os vizinhos
  - Se quantidade de vizinhos < 20, pede ao tracker mais peers
  - Inicialmente abre 40 conexões
  - As outras 40 serão conexões advindas de outros peers

Mensagens do protocolo BitTorrent

- Handshake
- Bitfield
- Keep_alive
- Interested / Not_interested
- Unchoke / Choke
- Request
- Piece
- Have / Cancel

Informação da Peça (Bitfield)

- Após estabelecer a conexão, os peers realizam o shake hands e compartilham as peças que possuem através do BITFIELD



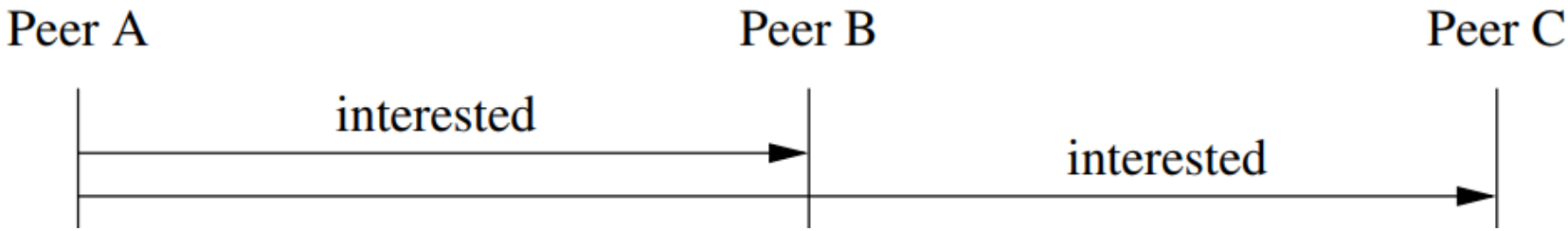
Conexões do peer

- Para evitar o custo do handshake na abertura de uma conexão, o peer mantém a conexão aberta o máximo possível
- A mensagem KEEP_ALIVE é enviada cada 2 minutos

Informação da peça

- Após o intercâmbio de bitfield, ambos peers conhecem quais peças o outro possui
  - Peer A está interessado INTERESTED no peer B se A não tem peças de B
  - Peer A está NOT INTERESTED no peer B se A já tem as peças de B

- Quando um peer adquire uma nova peça, avisará a todos seus vizinhos enviando a mensagem HAVE



Mecanismo de Choking

- Uma das ideias mais poderosas do BitTorrent é assegurar que os peers cooperem, eliminando os free-riders

- Para isso, cada peer *p* sempre fará o unchoke de um número fixo de peers.
  - Unchoke é o permisso temporal de upload
  - Choke é a recusa temporal de upload

- Mas a qual deles realizar o unchoke?

- O peer *p* fará o unchoke (upload) aos quatro leechers dos quais *p* baixou exitosamente peças
  - A decisão de choke/unchoke é realizada cada 10 segundos
  - Baseada na taxa de download avaliada nos últimos 30 segundos

Mecanismo de Choking

Na visão de um peer p

- Passo 1: ordenar todos os peers interessados (INTERESTED) que fizeram o upload de ao menos 1 bloco nos últimos 30 segs.
    - Elimine os peers snubbed, i.e., que não enviaram nada durante os 30 segs.

Mecanismo de Choking

Na visão de um peer p

- Passo 2: os três peers mais rápidos serão unchoked por *p*.
    - Esses peers são denominados de Regular Unchoked (RU) peers
    - Envia a mensagem UNCHOKED para os RUs

- Passo 3: escolha aleatoriamente um peer que não está em RU para ser o peer optimistic unchoked (OU)
    - 3ª: se o OU está interessado, envie a mensagem UNCHOKED

Mecanismo de Choking

Na visão de um peer p

- Passo 2: os três peers mais rápidos serão unchoked por *p*.
    - Esses peers são denominados de Regular Unchoked (RU) peers
    - Envia a mensagem UNCHOKED para os RUs

- Passo 3: escolha aleatoriamente um peer que não está em RU para ser o peer optimistic unchoked (OU)
    - 3ª: se o OU está interessado, envie a mensagem UNCHOKED

Para quê?

Mecanismo de Choking

- O Optimistic Unchoking permite:
    - Encontrar peers potencialmente mais rápidos
    - Permitir aos peers que iniciaram (portanto com poucas peças) obter peças

Estratégias de seleção da peça

- Um unchocked peer A envia uma mensagem REQUEST ao peer B por uma peça
- O peer B envia uma mensagem PIECE com o bloco/peça como payload
- Quando o peer A receber a peça, envia:
    - uma mensagem HAVE a todos os peers vizinhos
    - uma mensagem CANCEL a todos os peers vizinhos

Tudo bem, mas qual peça o peer A deveria requisitar?

Estratégias de seleção da peça

- Random first
  - No começo o peer precisa de qualquer peça completa
  - Usado quanto tiver menos de 4 peças completas
  - Importante para ter algo a ser usado no Tit-for-Tat
- Strict priority
  - Outros blocos da mesma peça e do mesmo peer (reuso do canal TCP)
  - Ou seja, permite obter a peça completa o mais rápido possível
- Rarest first (muito importante)
  - Obter peças menos replicadas, deixando as mais comuns para depois
- Endgame mode
  - Broadcast por todos os blocos restantes
  - Usado no final do download do conteúdo

Estratégias de seleção da peça: rarest-first

Na visão de um peer *p*

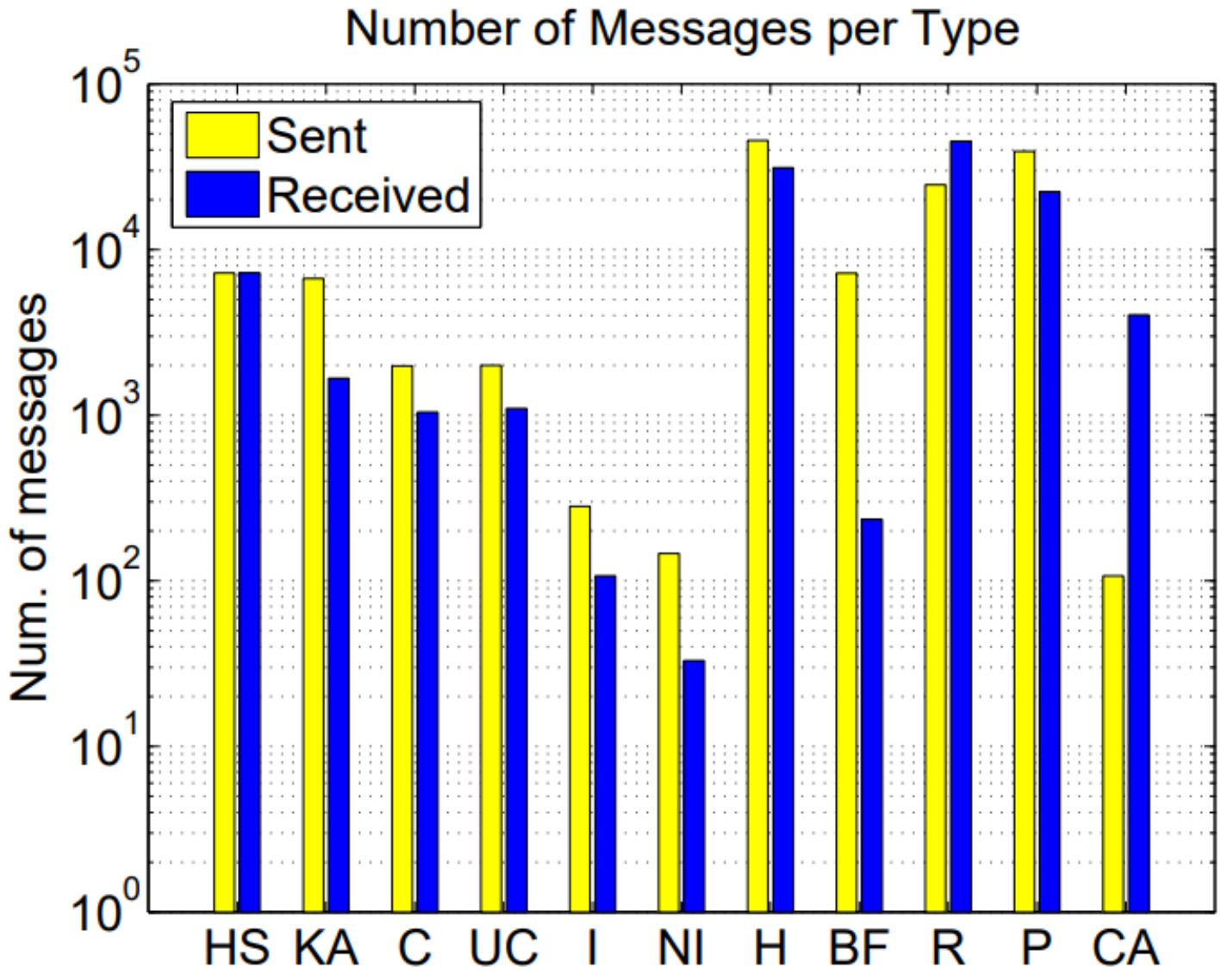
- Conhece quais peças seus vizinhos possuem
  - Via as mensagens HAVE e BITFIELD
- Pode calcular a disponibilidade de cada peça
  - Quantas vezes a peça está replicada nos vizinhos
- Assuma a peça *m* com menor disponibilidade
  - Essa peça *m* será a rarest-piece
- Ao pedir a peça *m*, prolonga-se a vida do conteúdo, reduzindo o risco que a peça vire extinta

Resultados

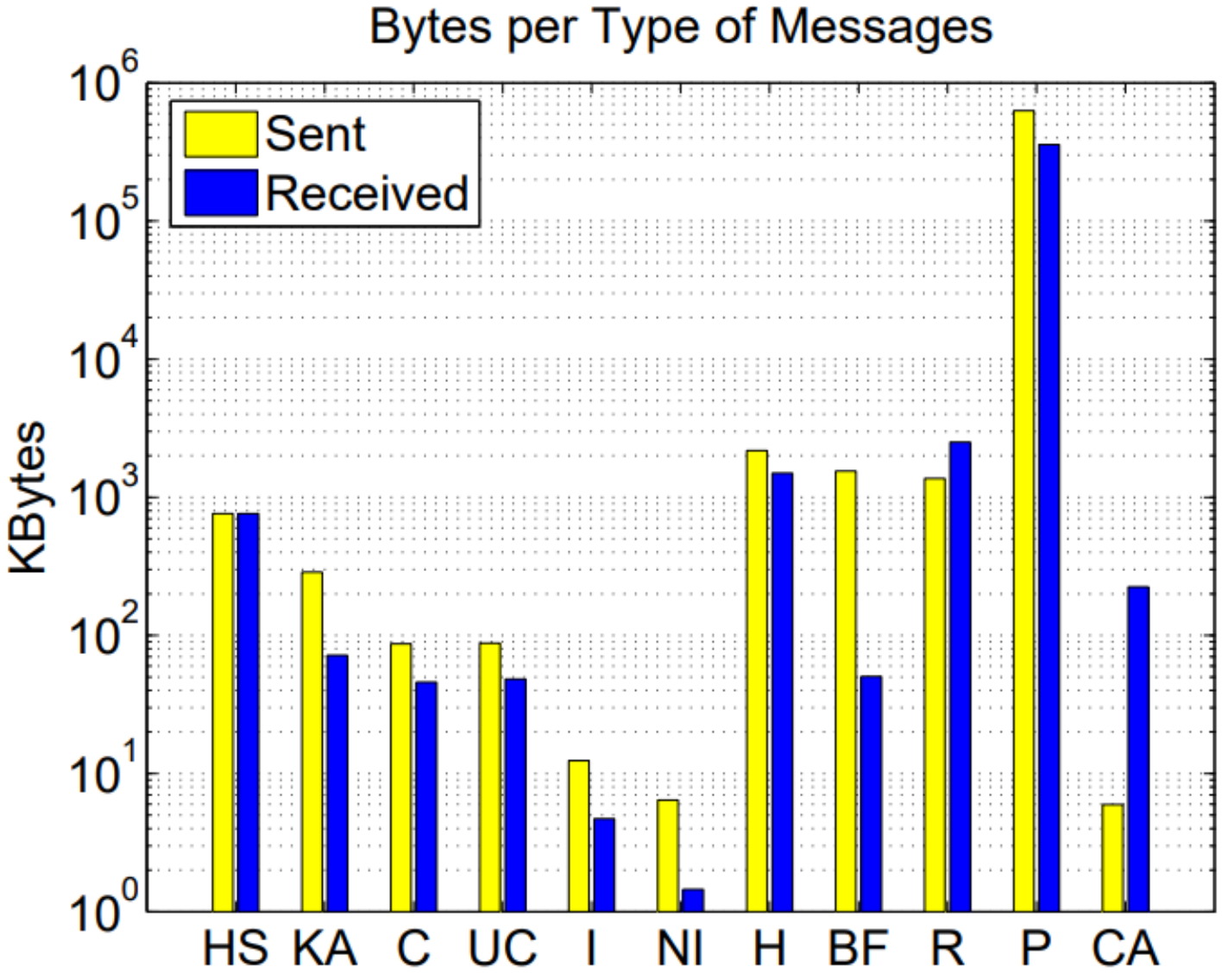
- Overhead do protocolo muito baixo (< 2%)

Understanding BitTorrent: An Experimental Perspective
[Legout et al., INRIA-TR-2006]

Resultados

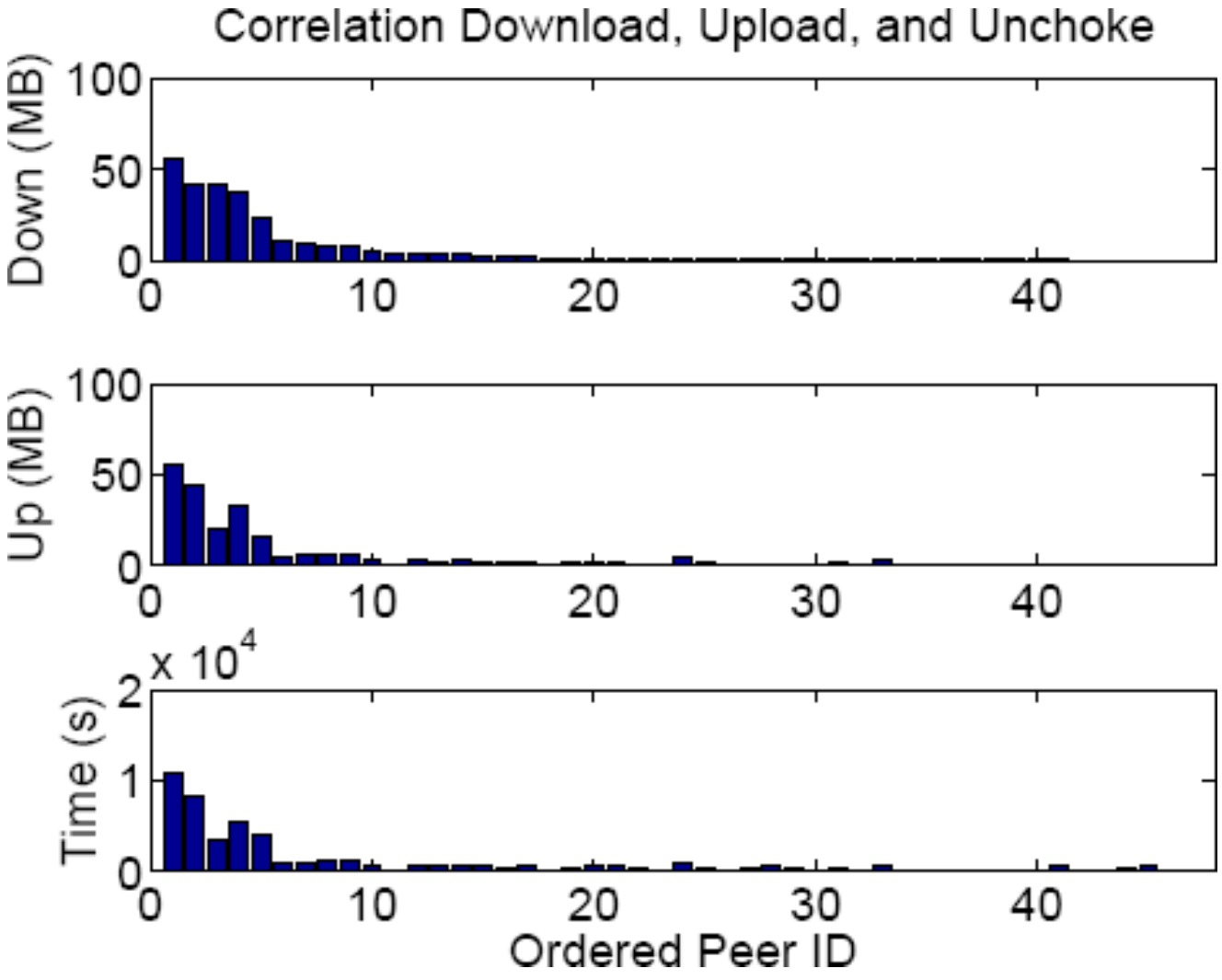

Number of Messages per Type
Num. of messages
10^5
10^4
10^3
10^2
10^1
10^0
Sent
Received
HS
KA
C
UC
I
NI
H
BF
R
P
CA

Resultados


Bytes per Type of Messages
Sent
Received
KBytes
10^6
10^5
10^4
10^3
10^2
10^1
10^0
HS
KA
C
UC
I
NI
H
BF
R
P
CA

Resultados

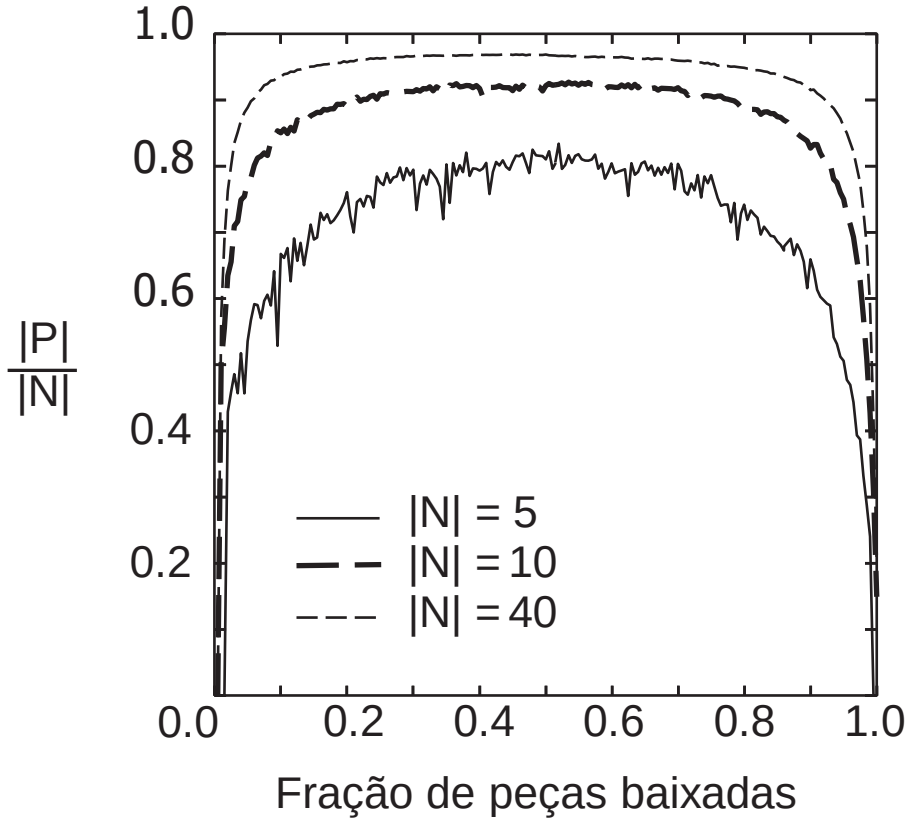
- Algoritmo de choke consegue reciprocidade no envio/recebimento de dados



Resultados

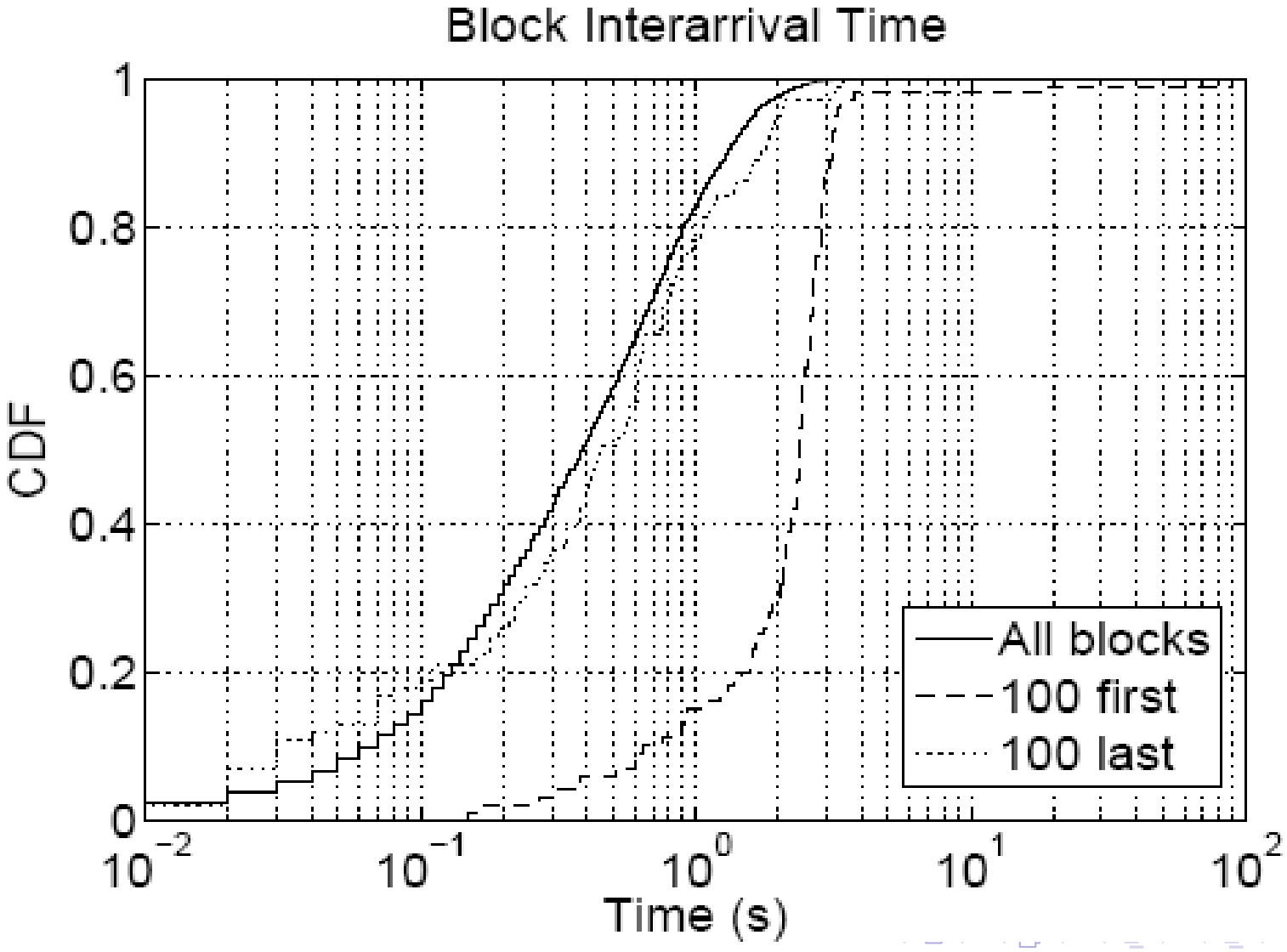
- Seja P um nó potencial a enviar uma peça e N o total de nós no swarm.

Ter mais nós no swarm aumenta a chance de um nó P ter a quem enviar



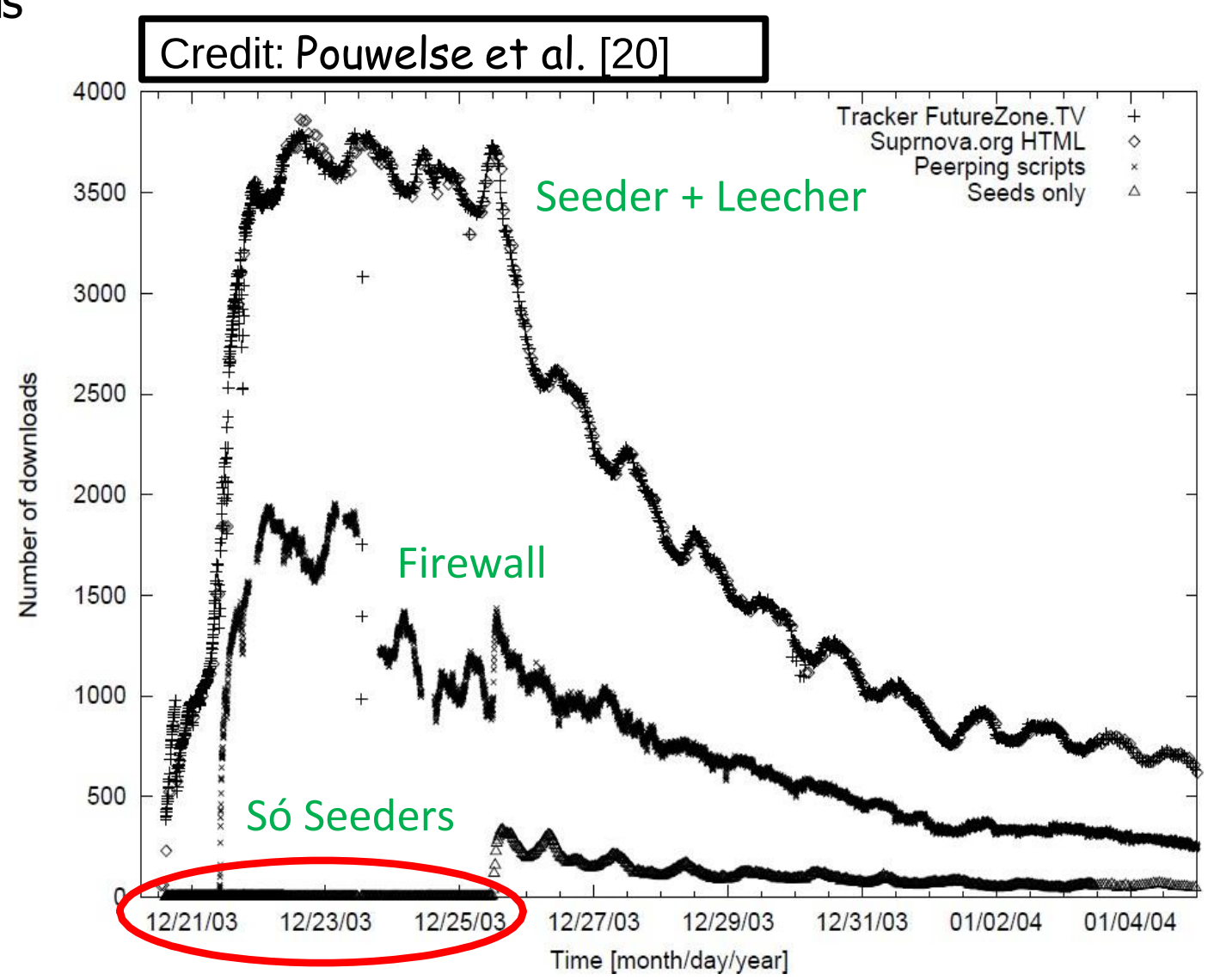
Resultados

- O tempo para adquirir os primeiros blocos é subestimado



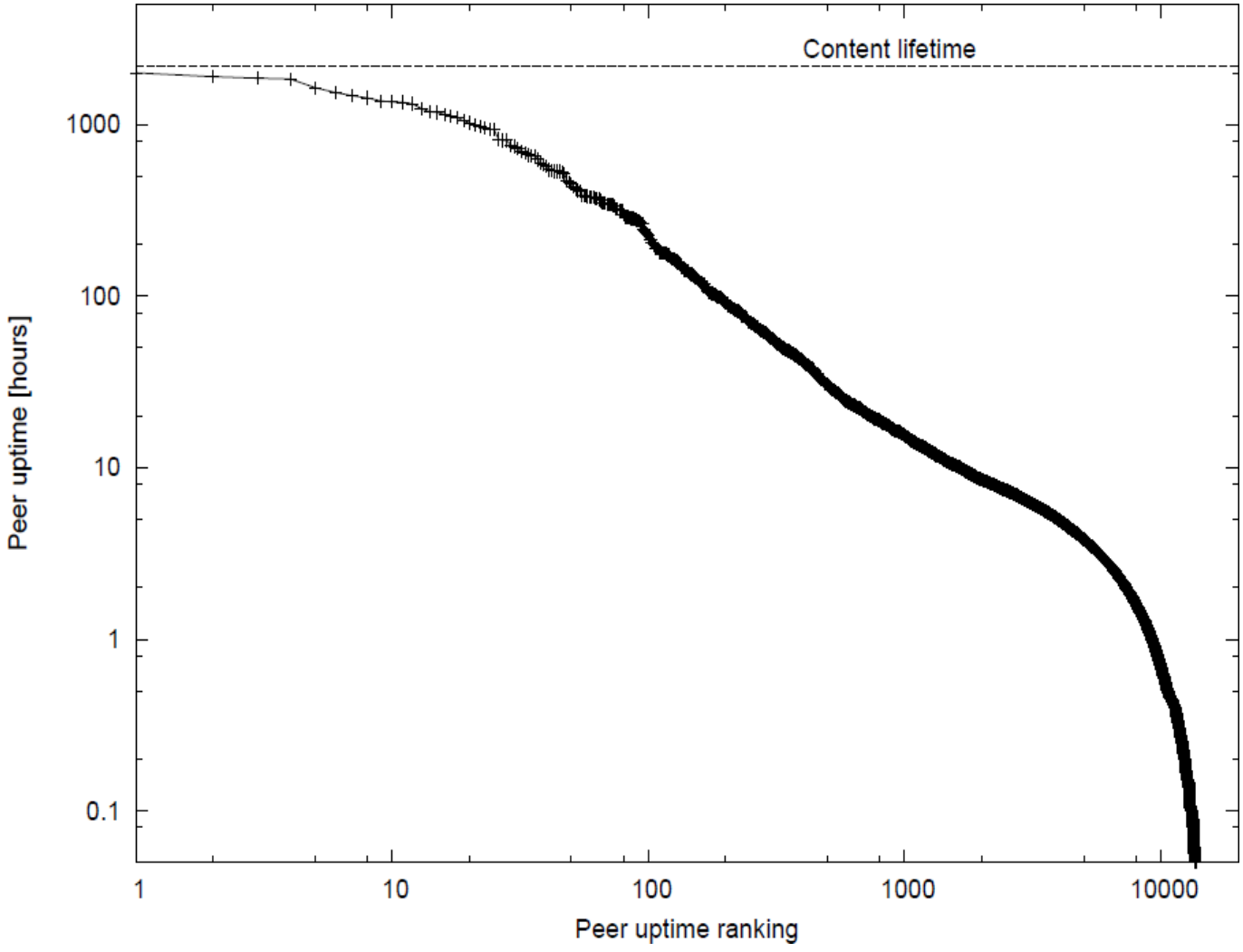
Resultados

- Flash crowds



Resultados

- Disponibilidade do peer no estado seed (i.e., uptime em modo altruísta)



- 17%~9200 têm uptime > 1 h
- 3.1%~1600 têm uptime > 10 h
- 0.34% ~183 têm uptime > 100 h

1 seeder com ~ 2000 h.

Trackers e servers

- Qual é a capacidade e infraestrutura de um tracker?

http://www.living-torrent.com/2018/12/top-torrent-tracker-hits-record_18.html


Top Torrent Tracker Hits Record Breaking 30 Million Peers

amine belotmani December 18, 2018

"At this moment we have a tendency to square measure hosting the huntsman on 3 medium-sized dedicated servers …"

"Across the 3 servers this adds up to spill 2 billion connections per day…"

Trackers e servers

- Qual é a capacidade e infraestrutura de um tracker?

Meta (anteriormente Facebook), apresentou em 2022 a modificação do protocolo **BitTorrent** para compartilhamento de arquivos entre seus servidores multimídia.

A ideia foi ter vários trackers, cada um responsável por um conjunto de peers, controlando exatamente o que cada peer deve baixar, quando baixar e de quem baixar.

A comunicação entre Tracker e Peer é realizada via **RPC**.

O sistema atende aproximadamente 10 milhões de clientes concorrentes e transfere 1 exabyte de informação por dia. Cada Tracker gerencia aproximadamente 100 mil peers.

1 exabyte = 1 milhão de terabytes.

Conceitos adquiridos

- BitTorrent tracker, seeder e leecher.
- SPoF – Single Point of Failure.
- Estratégia Tit-for-Tat.
- Estratégia Rarest-first.