

Problem:

- O problema está na cada vez maior utilização de protocolos P2P, como mostra o estudo da época.

Problem 2:

- Isto representa um problema porque o paradigma da comunicação modifica-se: **descentralizado**, cada um partilha o seu conteúdo.
- Descentralização gera muito mais tráfego por toda a rede e não apenas em determinados pontos (como acontecia com cliente-servidor) -> afecta qualidade de outros serviços
- Descentralização significa conteúdo espalhado pela rede, o que resulta em muita comunicação inter-ISP
- Descentralização é cara!

Problem 3:

- BitTorrent é o protocolo p2p que mais tráfego gera

Solution:

- Solução passa por analisar a localidade e o conteúdo.

Solution 2:

- Identificar padrões ao nível de comportamentos de peers, swarms e ao nível dos conteúdos partilhados

Objectives:

- Explicar o bittorrent muito por alto e depois ainda dizer que mais tarde e fazendo parte do swarm, os peers podem obter outros peers através de queries ao tracker ou a peers aos quais estejam ligados
- Dai a divisão em dois estudos: localidade e conteúdo

Bittorrent:

- Explicar pelas imagens
- Não esquecer infohash – identificador dos swarms/contéudo

Methodology for gathering data:

- Explicar através das imagens

Methodology for analysing data:

- Torrent database
- Cada estudo tem diferentes características e daí os diferentes critérios
 - Conteúdo ...
 - Localidade ...

Results:

- Os resultados obtidos do estudo foram divididos em três categorias

Content Analysis:

- Dividido em duas partes: pollution and repetition.
- Content pollution:
 - dos 3211 torrents analisados, 221 tinham mais de 70% das suas partes iguais
 - destes 221, quase 60% demoram menos de 24h a desaparecer (peers abandonam o swarm e não voltam)
- content repetition:

- dos torrents que têm pelo menos uma parte em comum, 50% partilham entre 0 e 16% e os restantes 50% partilham entre 95% e 100%
- de realçar que existe um numero significativo de torrents com 100% de partes comuns, para estes foram calculadas as infohash e descobrimos que nenhum partilhava a mesma infohash
- Maior parte dos conteúdos repetidos duas vezes. A partir da 4 vez são raros.
- Dois gráficos com a média de peers e seeders por conteúdo e a soma de todos para cada conteúdo. Facilmente se observa um ganho significativo na disponibilidade dos conteúdos.

Locality analysis:

- Primeiro foi analisada a localidade existente nos swarms a partilhar conteúdo repetido e de seguida os ganhos de localidade que existiriam se os swarms, que partilham mesmo ou similar conteúdo, fossem integrados num só.
- De seguida, foi estudada a localidade para todos os conteúdos e identificados os conteúdos regionais, conteúdos específicos a determinada região ou país.
- A localidade calculada anteriormente foi feita com base em dados agregados em períodos de 1 dia. No final, as análises foram repetidas mas para períodos mais pequenos, 2 horas. Isto foi feito tendo em conta menor tempo de download e consequentemente menor numero de fontes que possam partilhar os ficheiros.
- Repeated content:
 - Tabela mostra para um determinado conteúdo repetido, os valores XXX para cada torrent e para a agregação dos três
 - O gráfico mostra os ganhos para a agregação de dados para cada conteúdo por dia. A maioria tem ganhos significativos, no entanto, há alguns que têm ganhos negativos uma vez que ao agregar os torrents do mesmo conteúdo, o aumento no número de ISPs excedeu o aumento no número de peers.
 - Desta forma, é demonstrado que a junção dos vários swarms que partilham o mesmo ou similar conteúdo não só aumenta a disponibilidade do conteúdo como aumenta a localidade para aquele conteúdo.
- All content:
 - Media de peers por País e ISP relacionados com a media de diferentes peers obtida através das queries ao tracker. → nota-se uma clara tendência para aumento da localidade com o aumento do tamanho do swarm e tempo de download
 - Foi calculado para cada torrente, para todos os dias, a mediana da percentagem de peers que pertencia ao mesmo país ou ISP. Para todos os resultados foram calculados o mínimo, máximo, 25, 50 e 75 percentil. Era de esperar que os 25, 50 e 75 percentil aumentassem com o tamanho do swarm, no entanto, isso não acontece porque, ao aumentar o swarm, os peers ficam com uma melhor distribuição por todos os países e isps o que leva ao 25 e 75 percentil aproximarem-se um do outro. Estes resultados demonstram que a utilização de um mecanismo de localidade é capaz de criar clusters grandes o suficiente para conseguir reter o tráfego perto na rede e não afectar a experiencia do utilizador. Para swarms mais

pequenos, e motivados pela quantidade de localidade encontrada em alguns, decidimos aprofundar a análise para uma procura por conteúdo regional. Em relação a isp, não é possível ver nenhum padrão, no entanto, também foram identificadas propriedades de localidade.

- Torrents regionais:
 - Torrents regionais são torrents que representam conteúdo específico a determinado país, região ou língua.
 - Para um torrente ser considerado regional, decidimos que tinham de ter 60% dos seus peers a pertencer ao mesmo país pelo menos 75% das vezes.
 - Dos resultados anteriores, procuramos os que tinham 30% dos peers a pertencer ao mesmo isp, 75% das vezes.
 - Mínimos devem-se a queries ao tracker que retornaram poucos peers e muito espalhados pelos diferentes isps
 - Estes resultados sugerem grande localidade existente num número significativo de torrents. No entanto, 80% dos torrents tinham associados swarms até 500 peers. Isto demonstra que muitos dos pequenos swarms, em vez de serem pouco populares, são apenas populares em determinadas regiões. Estes torrents regionais representam quase 5% do total de swarms com tamanhos entre 50 e 500 peers.
- Two hour period:
 - Esta análise foi feita para perceber o quanto o tempo de download afecta a localidade existente. Só foi feita para swarms acima dos 5000 peers.
 - Média de peers por isp e país. Valores mais baixos do que os obtidos para períodos de 1 dia, como esperado. Quanto menor o tempo de download, menos peers se obtêm, e tendo em conta a utilização por parte do tracker de um algoritmo de selecção aleatória para a escolha dos peers a enviar para o cliente, é normal que os peers estejam mais distribuídos pelos diferentes países e isps.
 - Apesar dos resultados mais baixos, ainda é possível observar grandes quantidades de localidade para ser explorada.
 - Conclui-se que o tempo de download tem um grande impacto na localidade existente. Desta forma, um mecanismo de localidade conseguirá melhores resultados se modificar, entre outros, o algoritmo de selecção aleatória de peers do tracker. Utilização de DHTs e PEX também é importante para conseguir mais fontes em downloads mais rápidos.

Peer and tracker behavior:

- Tracker distribution:
 - Dos 3211 torrents, 1094 tinham mais de um tracker a responder a queries. No entanto, reparamos que os peers não se distribuíam uniformemente entre os diferentes trackers.
 - Figura mostra 5 trackers para o mesmo ficheiro. É esperado que a localidade e disponibilidade sejam diferentes de tracker para tracker e que seja mais acentuada em trackers com diferentes tamanhos
 - Os picos são quebras do tracker, no entanto, são de rápida recuperação e com tecnologia tipo PEX e DHTs não existe qualquer problema com isto.
- Day-night behavior:

- A figura mostra claramente o enxame a aumentar e diminuir de tamanho num período aproximado de 1 dia.
- Mudanças foram de 5% até 60%
- Comportamento mais evidente em swarms grandes.
- Free-riding:
 - Figura mostra que com o aumento do tamanho do swarm, este numero tende a ser igual ao numero de seeders, ou seja, os peers à medida que acabam o download vao ficando online a partilhar o que obtem da rede.
 - Free-riding não é um problema no bittorrent, significativo pelo menos.

Conclusao:

- Muita localidade em grandes swarms ao nível de distribuição de peers por pais e ISP mas também ao nível dos pequenos swarms – conteúdo regional
- Muitos padrões observados para uma rede p2p descentralizada:
 - Conteúdo partilhado repetido
 - Comportamentos do peer como o day-night
- Provado que a performance do bittorrent pode ser melhorada, aumento de disponibilidade para conteúdos repetidos, e o trafego gerado mantido perto na rede, propriedades de localidade muito fortes.