CD-GAIN: Measuring Traffic Gains from Peer-assisted Content Delivery of Long Duration Video-on-Demand content

Dmytro Karamshuk$^1$, Nishanth Sastry$^1$, Jigna Chandaria$^2$, Andy Secker$^2$

$^1$King's College London, $^2$BBC R&D

ESPRC project in collaboration with BBC R&D
On-demand video delivery – BBC iPlayer

10% of UK's net traffic*
55.1% of the internet TV

Video Delivery Network

2\textsuperscript{nd} after YouTube VoD source

40% of UK's population*

* as of 2012

- most of the video content in iPlayer is long duration usually streamed in high bitrates
- iPlayer is a “catch-up” streaming service which allows VoD access to recently broadcast shows
Peer-assisted content delivery

Obstacles?

Video Delivery
Network
Peer-assisted content delivery

- ISP-friendliness
- Bitrate stratification
- Partial participation
Analytical Model

Traffic Gain

\[ G = 1 - \frac{T_s}{T_u} \]

- server traffic

- useful traffic

Swarm's capacity

\[ c_i = u_i r_i \]

- average number of peers in a swarm

Gain for a single swarm

\[ G = 1 - \left( \frac{c_i \left( 1 + e^{c_i c_i^{-m}} \left( m \Gamma(m) - \Gamma(1 + m, c_i) \right) \right) - 1}{m} \right) \]

Gain across multiple swarms

\[ G = 1 - \frac{\sum \beta_i 1_i r_i}{\sum \beta_i 1_i r_i} \]

- bitrate of a content swarm

- arrival rate of a content swarm

\[ E[B_i] = \frac{c_i \left( 1 + e^{c_i c_i^{-m}} \left( m \Gamma(m) - \Gamma(1 + m, c_i) \right) \right)}{r_i m} \]
Dataset of accesses in London, Sep. 2013

Records in the Dataset:

<network id, ISP, user id, request time, session duration, bitrate, content id>
Traffic Gains in a Content Swarm

- model is in a good agreement with simulations
- significant traffic gains can be achieved despite constraints
Fragmentation of traffic

Eight biggest ISPs account for over 70% of traffic

Two main bitrate formats dominate in over 70% of sessions

Fragmentation of traffic by ISP and bitrate formats yields several large content swarms
Partial Participation and Online Model

- Partial participation of users might be a critical obstacle.
- "Online while watching" model is important.
Traffic Gains for a Content Corpus

Traffic Gains across ISPs

<table>
<thead>
<tr>
<th>ISP</th>
<th>No stratification</th>
<th>Stratification</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISP-1</td>
<td>G_{sim} 0.91, G_{theo} 0.89</td>
<td>G_{sim} 0.81, G_{theo} 0.78</td>
</tr>
<tr>
<td>ISP-2</td>
<td>G_{sim} 0.90, G_{theo} 0.87</td>
<td>G_{sim} 0.77, G_{theo} 0.75</td>
</tr>
<tr>
<td>ISP-3</td>
<td>G_{sim} 0.80, G_{theo} 0.75</td>
<td>G_{sim} 0.65, G_{theo} 0.59</td>
</tr>
<tr>
<td>ISP-4</td>
<td>G_{sim} 0.80, G_{theo} 0.77</td>
<td>G_{sim} 0.66, G_{theo} 0.63</td>
</tr>
<tr>
<td>ISP-5</td>
<td>G_{sim} 0.79, G_{theo} 0.75</td>
<td>G_{sim} 0.65, G_{theo} 0.61</td>
</tr>
<tr>
<td>ISP-6</td>
<td>G_{sim} 0.75, G_{theo} 0.71</td>
<td>G_{sim} 0.63, G_{theo} 0.60</td>
</tr>
<tr>
<td>ISP-7</td>
<td>G_{sim} 0.78, G_{theo} 0.73</td>
<td>G_{sim} 0.63, G_{theo} 0.60</td>
</tr>
<tr>
<td>ISP-8</td>
<td>G_{sim} 0.66, G_{theo} 0.61</td>
<td>G_{sim} 0.49, G_{theo} 0.44</td>
</tr>
</tbody>
</table>

Content Popularity

- gains remain high when the whole content corpus is considered
- most of the traffic is generated by few very popular content items
Content popularity

Daily patterns

- performance sparks on Weds reflecting TV schedules
- with 20% participation gains remain quite high

Partial Participation
Ways to improve traffic gains

Content Bundling

Historic Caching

Pro-active pushing VS reactive pulling?
Content Bundling

Server traffic for a bundle

\[ T_s^b = \Omega^b R^b P^b = \sum_{i=1}^{k} \left[ \beta_{i} \right] \sum_{i=1}^{k} \left[ r_i \right] \prod_{i=1}^{k} \left[ p_i \right] \]

Delta server traffic

\[ \Delta T_s = T_s - T_s^b = \sum_{i=1}^{k} \left[ \beta_{i} r_i p_i \right] - \sum_{i=1}^{k} \left[ \beta_{i} \right] \sum_{i=1}^{k} \left[ r_i \right] \prod_{i=1}^{k} \left[ p_i \right] \]

Delta traffic gain

\[ \Delta G = G - G^b = \frac{\Delta T_s}{T_u} = \frac{\sum_{i=1}^{k} \left[ \beta_{i} r_i p_i \right] - \sum_{i=1}^{k} \left[ \beta_{i} \right] \sum_{i=1}^{k} \left[ r_i \right] \prod_{i=1}^{k} \left[ p_i \right]}{\sum_{i=1}^{k} \left[ \beta_{i} \right] \left[ r_i \right]} \]
only a very small portion of bundles leads to traffic savings
an average delta gain from a bundle is 2–7%
Historic Caching

- client-side caching increases swarm's capacity by an average 10x
- historic caching with up to 10 items can boost gains for 15–30%

Improv. in a single swarm

Improv. for the content corpus
Takeaways

- a few high popular items can obtain gains of nearly 99% in the best case, and are hardly effected by the obstacle factors

- partial participation is the most critical obstacle, as others lead to forming few large swarms

- gains across the content corpus remain high due to a skewed popularity distribution

- content bundling can hardly improve performance, while a simple historic caching can add 15–30% to traffic gains
Thank you for your attention!

Dmytro Karamshuk
King's College London
follow me on Twitter: @karamshuk