KING'S College LONDON

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]

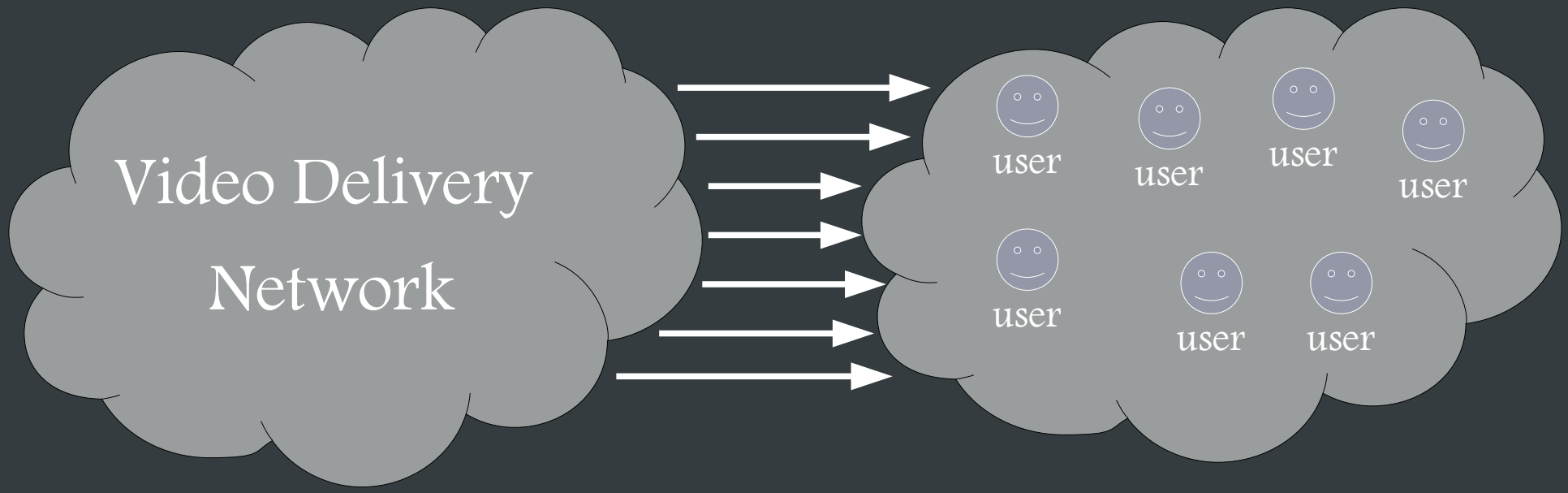
[1]King's College London, [2]BBC R&D

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



2nd after YouTube VoD source

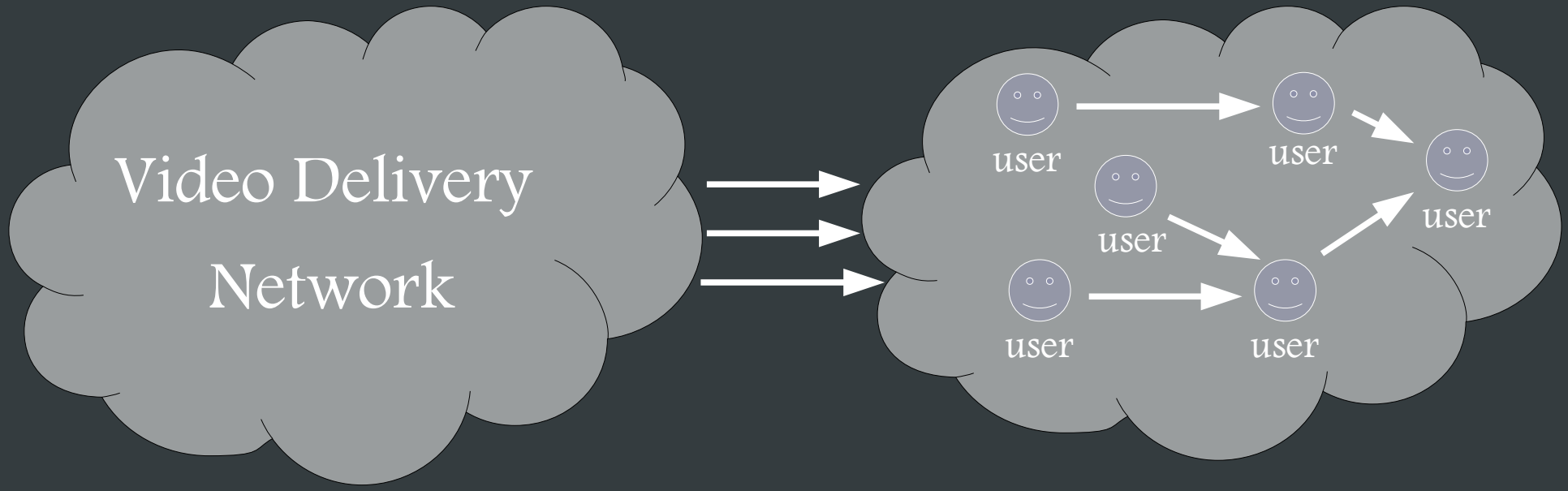
40% of UK's population*

* as of 2012

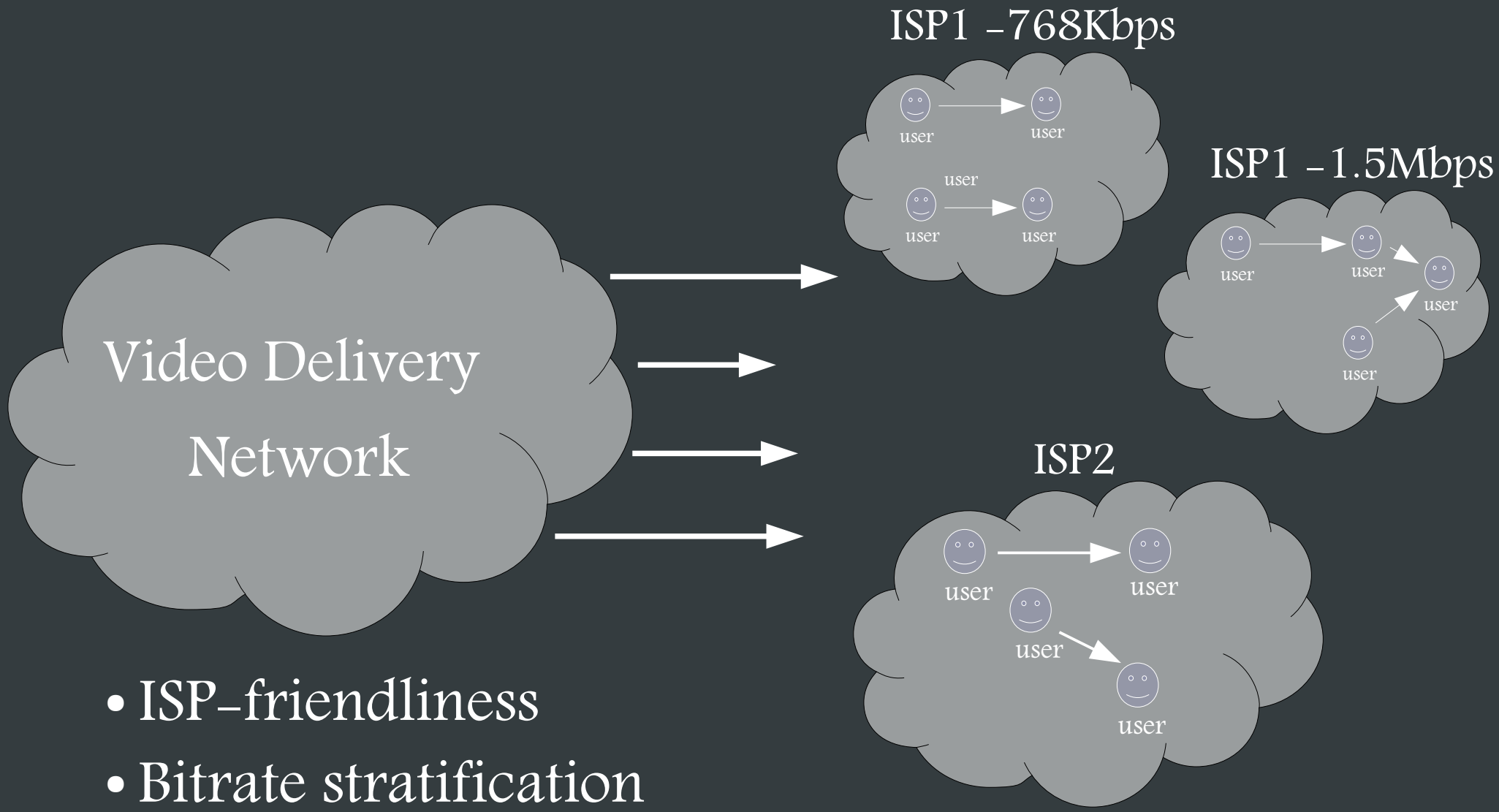
- most of the video content in iPlayer is long duration usually streamed in high bitrates
- iPlayer is a "catch-up" streaming service which allows VoD access to recently broadcast shows

Peer-assisted content delivery

Obstacles?



Peer-assisted content delivery



- ISP-friendliness
- Bitrate stratification
- Partial participation

Analytical Model

Traffic Gain $G = 1 - \frac{T_s}{T_u}$ - T_s server traffic; T_u - useful traffic

Swarm's capacity $c_i = u_i r_i$ - average number of peers in a swarm

Gain for a single swarm

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

Gain across multiple swarms

$$G = 1 - \frac{\sum \frac{\beta_i 1_i r_i}{E[B_i] r_i + 1}}{\sum \beta_i 1_i r_i}$$

β_i - bitrate of a content swarm

r_i - 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

BBC iPlayer

 Users – 2.2 M

 IP address – 1.3 M

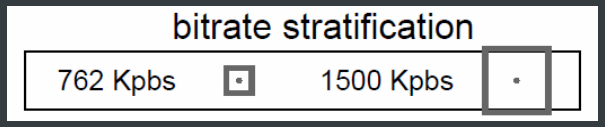
 Sessions – 15.9 M

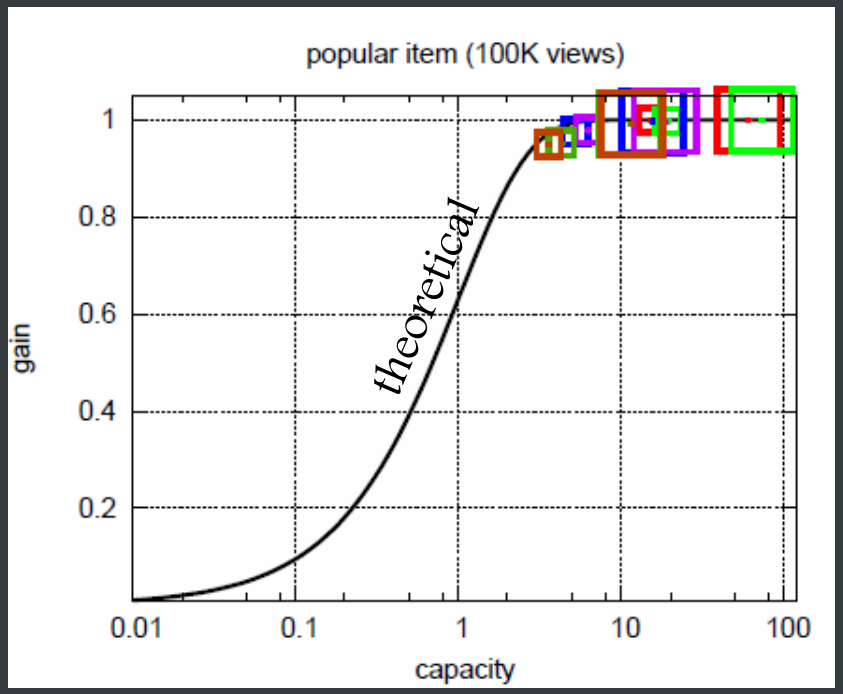
 Traffic – 6.5 Gbps

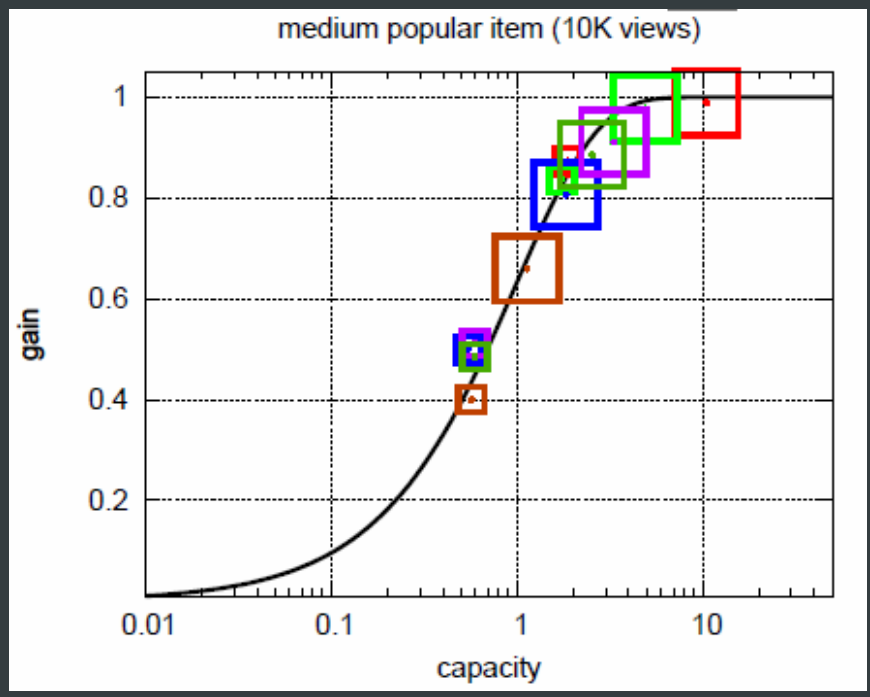
Records in the Dataset:

<network id, ISP, user id, request time, session duration, bitrate, content id>

Traffic Gains in a Content Swarm



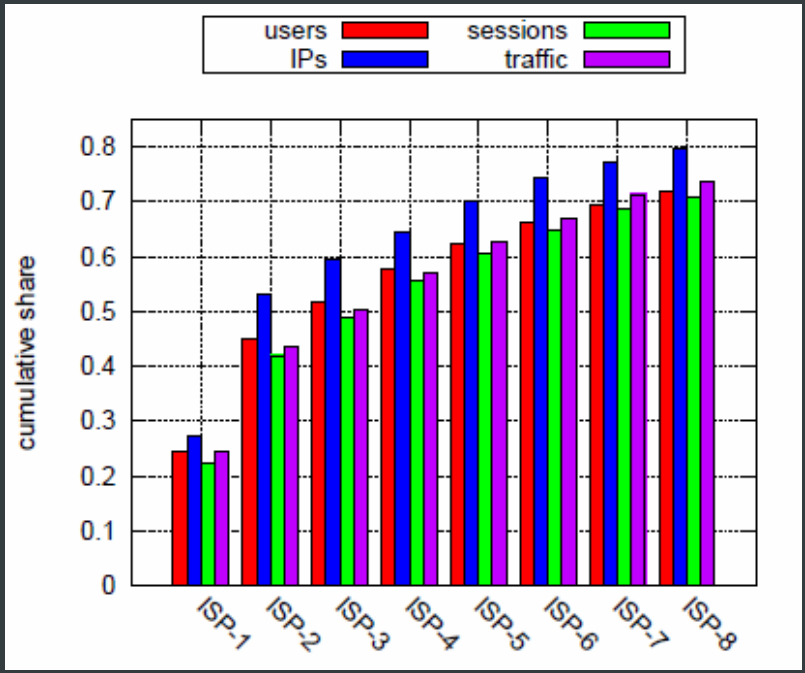




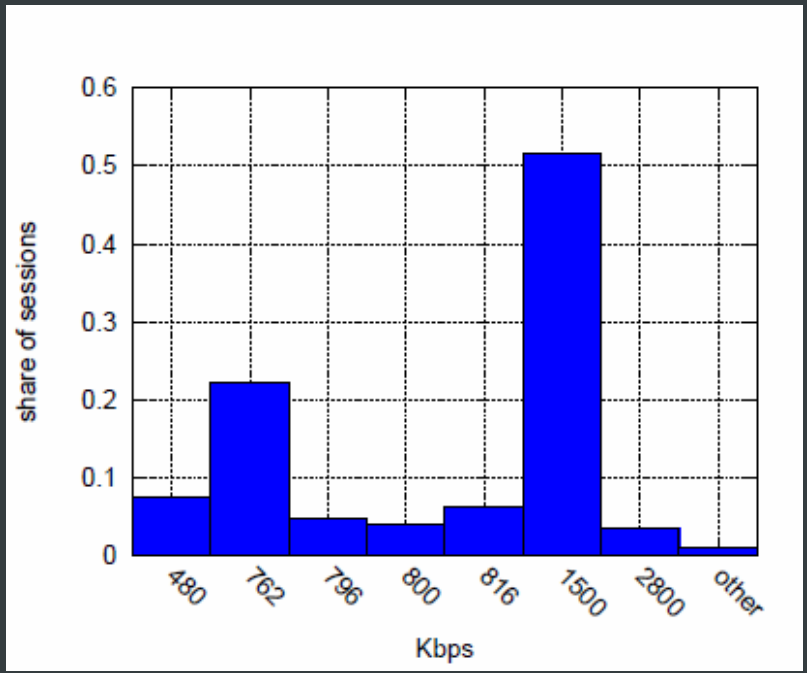
- model is in a good agreement with simulations
- significant traffic gains can be achieved despite constraints

Fragmentation of traffic

Eight biggest ISPs account for over 70% of traffic

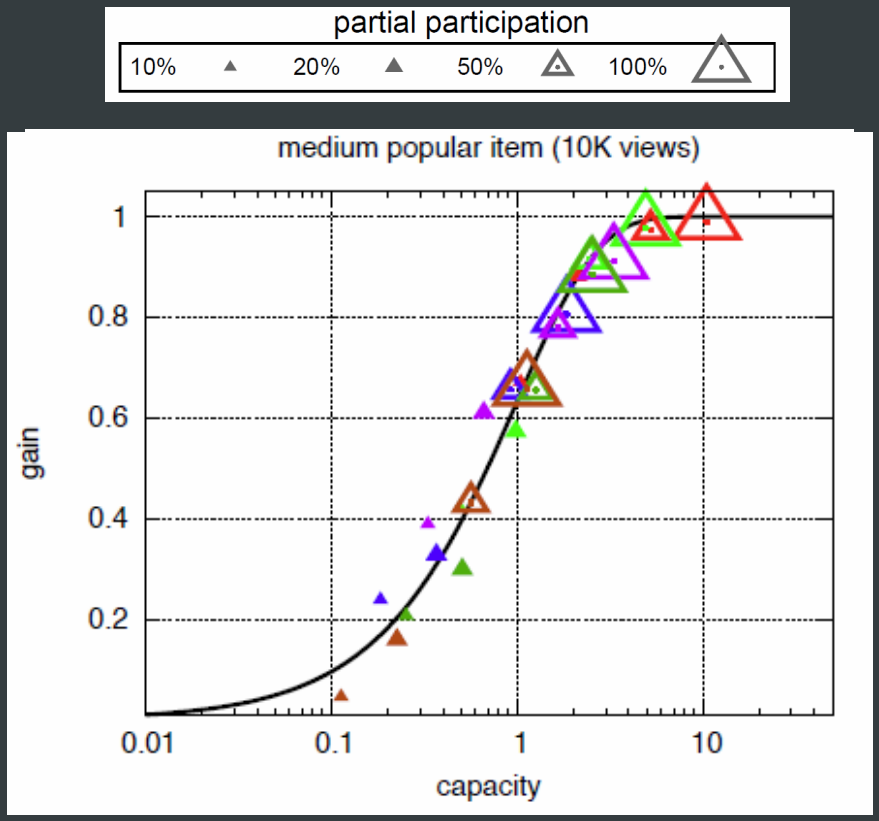


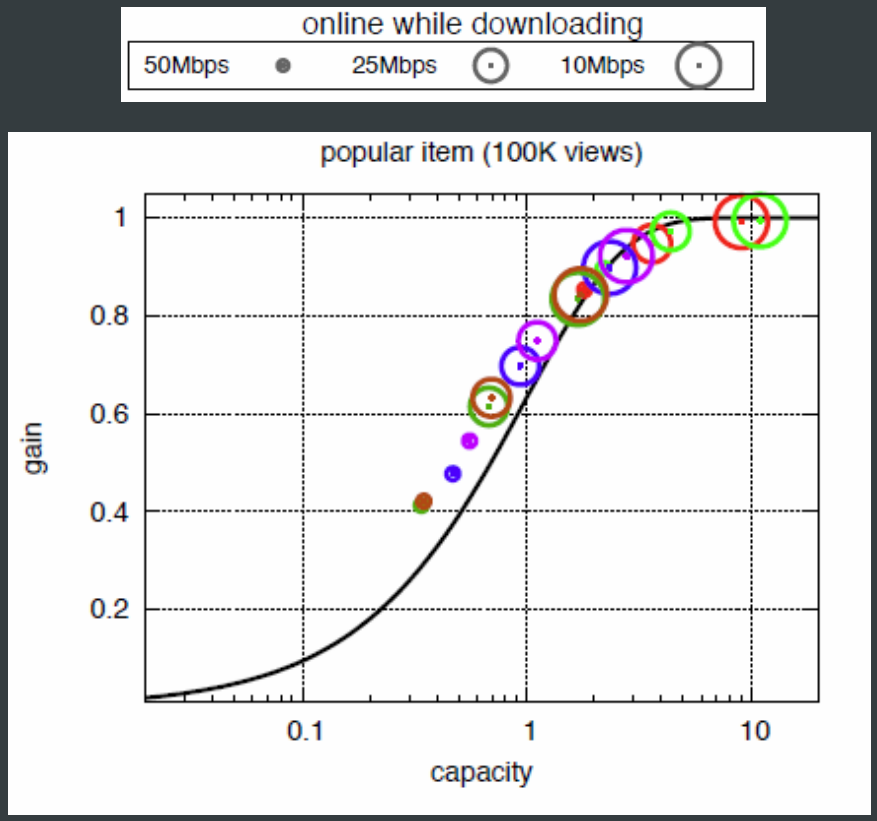
Two main bitrate formats dominate in over 70% of sessions



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

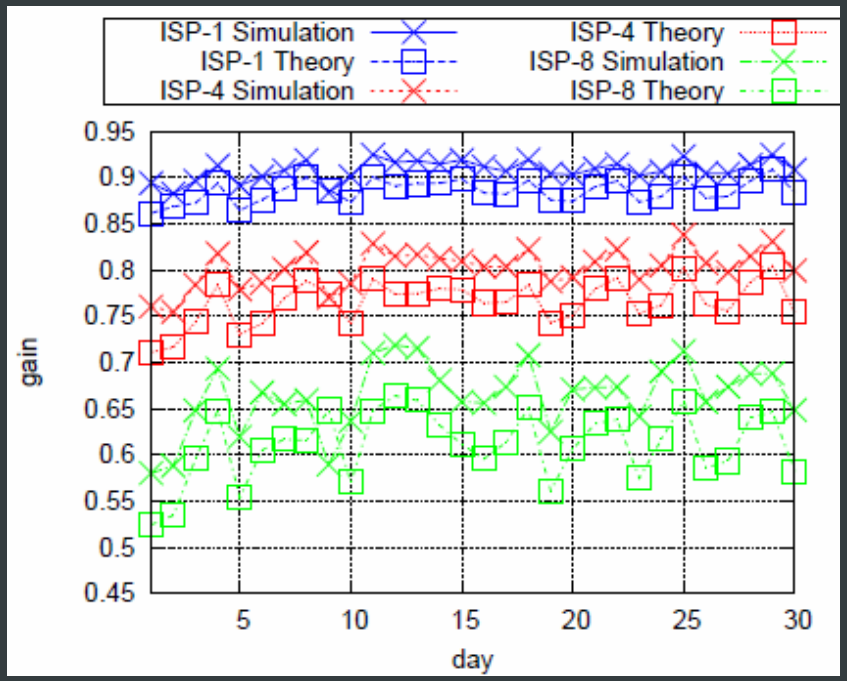
	No stratification		Stratification	
ISP	G_{sim}	G_{theo}	G_{sim}	G_{theo}
ISP-1	0.91	0.89	0.81	0.78
ISP-2	0.90	0.87	0.77	0.75
ISP-3	0.80	0.75	0.65	0.59
ISP-4	0.80	0.77	0.66	0.63
ISP-5	0.79	0.75	0.65	0.61
ISP-6	0.75	0.71	0.63	0.60
ISP-7	0.78	0.73	0.63	0.60
ISP-8	0.66	0.61	0.49	0.44

Content Popularity

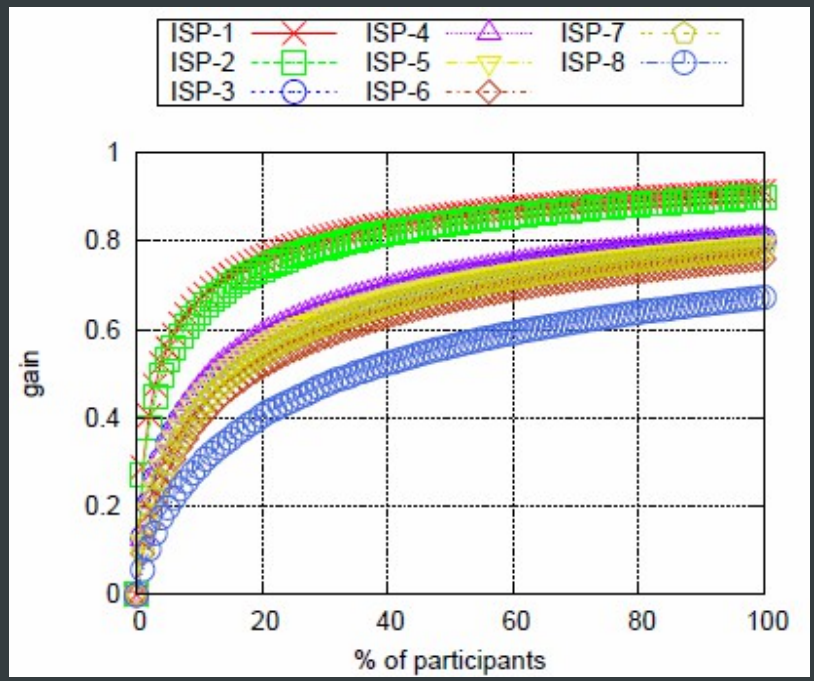
- gains remain high when the whole content corpus is considered
- most of the traffic is generated by few very popular content items

Content popularity

Daily patterns



Partial Participation



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

Ways to improve traffic gains

Content Bundling

Historic Caching

Pro-active pushing VS reactive pulling?

Content Bundling

Server traffic for a bundle

$$T_s^b = \Omega^b R^b P^b = \sum_{i=1}^{k} [\beta_i 1_i] \sum_{i=1}^{k} [r_i] \prod_{i=1}^{k} [p_i]$$

Delta server traffic

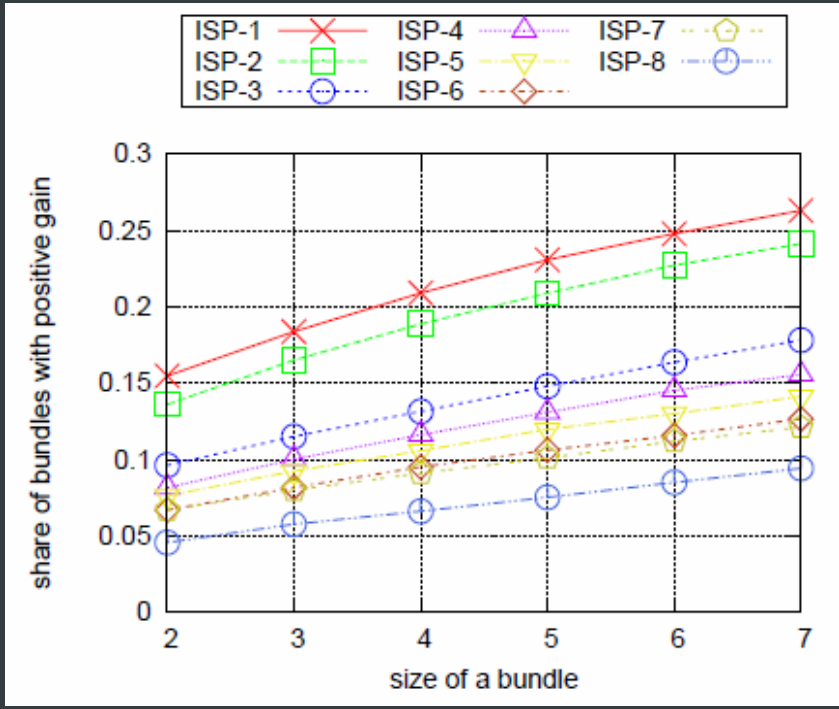
$$\Delta T_s = \sum_{i=1}^{k} T_s - T_s^b = \sum_{i=1}^{k} [\beta_i 1_i r_i p_i] - \sum_{i=1}^{k} [\beta_i 1_i] \sum_{i=1}^{k} [r_i] \prod_{i=1}^{k} [p_i]$$

Delta traffic gain

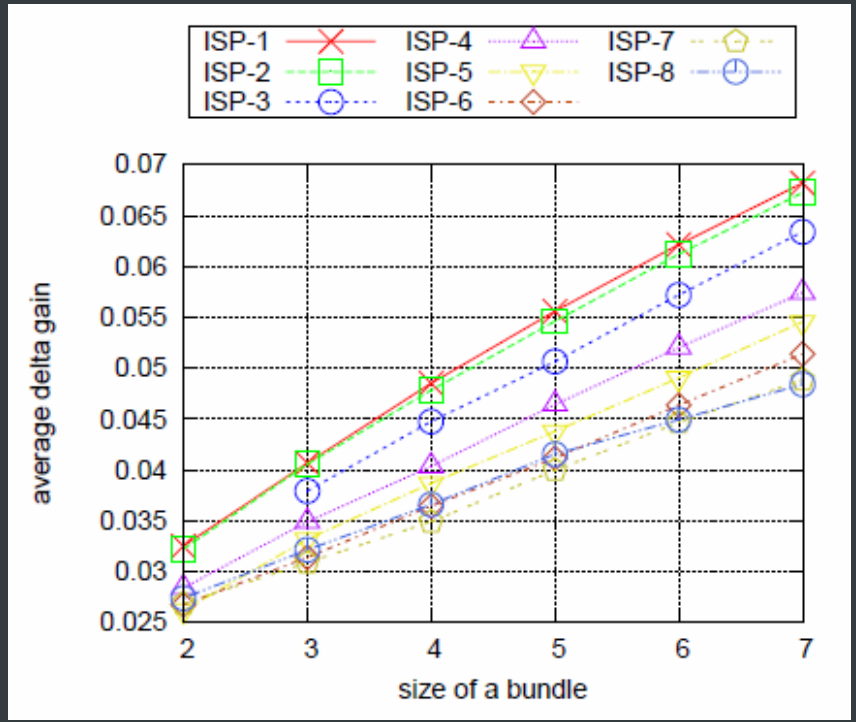
$$\Delta G = G - G^b = \frac{\Delta T_s}{T_u} = \frac{\sum_{i=1}^{k} [\beta_i 1_i r_i p_i] - \sum_{i=1}^{k} [\beta_i 1_i] \sum_{i=1}^{k} [r_i] \prod_{i=1}^{k} [p_i]}{\sum_{i=1}^{k} [\beta_i 1_i r_i]}$$

Content Bundling

Number of Bundles



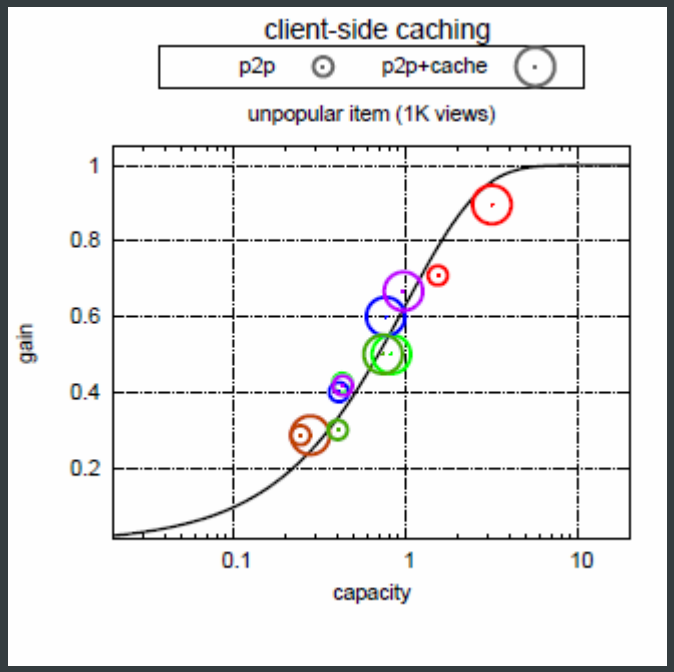
Gain from Bundles



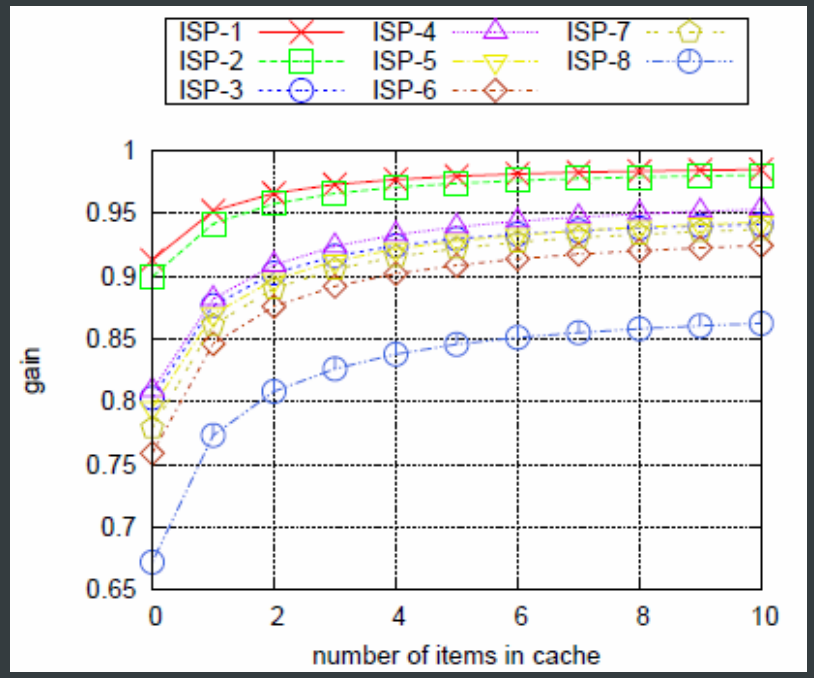
- only a very small portion of bundles leads to traffic savings
- an average delta gain from a bundle is 2–7%

Historic Caching

Improv. in a single swarm



Improv. for the content corpus



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

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

KING'S College LONDON

Thank you for your attention!

Dmytro Karamshuk

King's College London

follow me on Twitter: **@karamshuk**