

Bulk Data P2P: Cost *Shifting*, not Cost *Savings*

Nicholas Weaver

**All opinions are my own
Not of my employer or funding institutions**

Bulk Data P2P is Cost *Shifting* Not Cost *Savings*

- A P2P system which used only “unused” bandwidth would need to be ***friendlier***-than-TCP
 - But such a P2P system would be considered slow by the users: the incentive in building P2P systems is the opposite
 - You want to view the movie now, not tomorrow
 - Thus protocols like Joost: 300k download, 100k upload UDP
- What bulk-data P2P does is shift the upload bandwidth costs from the content provider to the recipients
 - The content provider has a large incentive: he’s dealing with a lot of data
 - \$.1/GB x 100k users at 10 GB/month is 100 k\$/month
 - The recipients have little incentive in the flat-rate pricing model
 - The ISPs bear the costs
 - And the ISP has a lot of incentives to prevent this cost-shifting

Bulk Data P2P is an *Inefficient* Cost Shifter

- Bulk-Data P2P greatly increases **aggregate** costs
 - Content provider bandwidth is cheap: \$.17-\$.10/GB (Amazon S3)
 - ISP bandwidth is expensive: \$100/Mbps/month (Iariat.net's wholesale price to buy bandwidth)
 - At 8 hours/day utilization -> \$1.00/GB
 - At 100% utilization -> \$.33/GB
 - And will always be more expensive: it will always be cheaper to connect 100 fibers to 1 location instead of 1 fiber to 100 locations
- Bulk-Data P2P is transport inefficient, even on highly popular files compared to HTTP
 - HTTP is transparently cached in many (?most?) ISPs:
 - 1 copy across the ISP's boundary, n-1 copies from the cache to clients
 - Bulk-Data P2P which is perfectly topologically aware:
 - 2 copies across the ISP's boundary, n-1 copies from client to client
 - Anything less than perfect increases the load
 - 50%-perfect means ~n copies must cross the ISP boundary, which is uncached http!
 - Lone files are always 2x the bits compared with HTTP

Conclusions

- Continued Conflict?
 - Bulk data P2P is **bulk data** and **P2P** under traffic analysis:
Creates an advantage for the ISP
- Usage based pricing and/or bandwidth caps?
 - Kill bulk data P2P **business model**:
Users hit their limits twice as quickly, and why would users spend \$1 to save the content provider \$.1?
- Caches? Probably not
 - Always less efficient than HTTP caches
 - For BitTorrent, asking for a lawsuit
 - Even if you win, the costs mean you lose
 - For legal commercial content: “I’m saving \$1 by costing you \$50, but this can reduce your costs to \$10...”

Appendix: Is Conflict Inevitable?

Plausible Outcomes

- Bulk-data P2P's cost shifting creates a conflict between content providers and ISPs
- Plausible outcome: the ISPs fight P2P
 - The ISPs have an inherent advantage, as bulk-data P2P may always be detectible, as it is **bulk data** and **peer to peer**
- Plausible outcome: usage-based pricing
 - Bulk-data P2P becomes a profit center for the ISPs
 - Bulk-data P2P ceases to be effective for content providers
- Plausible outcome: bandwidth caps
 - Bulk-data P2P (and all other heavy users) can only cost the ISP a bounded amount
 - If money can release the cap, it becomes usage-based pricing
- My hopeful outcome: User-fairness
 - Enforce fairness, **only within the ISP's network to the ISP border, with shapers located only at the point of user attachment, without changes to the routers or end hosts**, measured across ms to hours
 - Light users are happy, so the ISP is happy
 - If the heavy users are still happy, cool!
 - If the heavy users are unhappy, the ISP says "Go someplace else"
 - Open research/engineering problem to develop

Appendix:

What about caches?

- Can easily build BitTorrent caches today
 - Use (pick-your-IDS) to recognize tracker requests and send the messages to a local modified client
 - Local client/cache communicates with the tracker and all local peers
 - Use currently unused addressees to look like a large group
 - Probably a couple man-weeks for prototype, a man-year or two for production quality code
- Less effective than HTTP caches
 - Even with perfection, bulk-data P2P requires twice the transfers: one up and one down, unless the ISP creates a “leech-only” policy
- For pirated material: asking for a lawsuit
 - Even if the ISP wins, it is still costly to fight
 - So no generic BitTorrent cache
- For legitimate material: why would the ISP tolerate the content provider’s cost-shifting?
 - “Here, I’m going to save myself \$1 in a way that costs you \$20 to \$50, but you deploy our cache that only works for our special proprietary variant of the protocol and it will only cost you \$10...”