

Measurement of P2P: Gummadi, Dunn, Saroiu, Gribble, Levey, Zahorjan [Gummadi03a]

CSci551: Computer Networks
SP2006 Thursday Section
John Heidemann

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

1

Key ideas

- what does real p2p traffic look like?
 - important because lots of internet traffic (by bytes) is p2p
- compare p2p to web traffic (Kazaa)
 - sizes of files
 - download durations
- does this generalize to other p2p protocols?
 - BT parallelizes downloads
 - probably more connections; individual connections are shorter
 - probably not change total size, may change duration, ...
 - better to characterize what *users* want

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

5

Methodology

- take central trace at University (and anonymize it!)
 - long duration: 200 days
- look at
 - users
 - objects
 - and try and understand *why*
- develop model

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

6

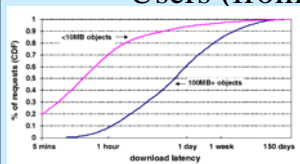
Users of p2p

- size (bytes)
 - large, MB (vs. KB)
- duration (time)
 - day or two (vs. seconds)
 - p2p are patient
- usage
 - stored and reused (vs. discarded and reloaded)
- popularity of individual files
 - different model (vs. Zipf)
 - strongly affects p2p traffic model: effectively the really popular stuff is retrieved less often in p2p because it's stored locally
- users slow down as they get older

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

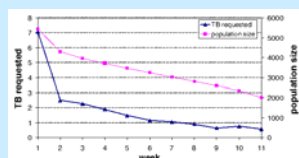
10

Users (from the paper)



[Gummadi03a, figure 1]

[Gummadi03a, figure 2]



15e_Gummadi03a: CSci551 SP2006 © John Heidemann

11

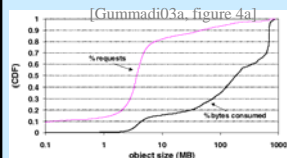
Objects

- if objects (files) are stored AND they don't change,
 - then there's no need to get them multiple times

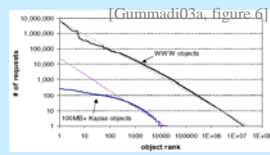
15e_Gummadi03a: CSci551 SP2006 © John Heidemann

15

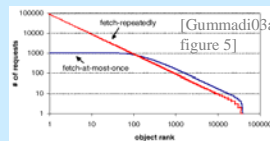
Dominance in Requests/Bytes



little objects have most requests,
but big objects have most bytes



fetch once is not like Zipf



and seems to hold
up in simulation

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

16

Caching

- caching in the network
- fetch-at-most once?
 - need for long-term caching is reduced
- and data is very age-dependent
 - new stuff is much more popular than old
- suggests that locality for temporary caching is important

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

20

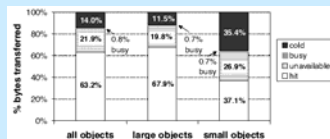
Exploiting Locality



locality can get much of
the same benefit as
caching (but beware
unavailability)

caching would help a lot
(wrt bytes transferred)

Q: what about wrt number
of connections?



[Gummadi03a, figure 13]

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

21

Other questions/observations?

- xxx

15e_Gummadi03a: CSci551 SP2006 © John Heidemann

23