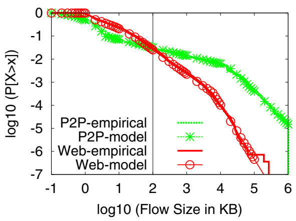
Table 5: Flow-level summary statistics of Web and P2P

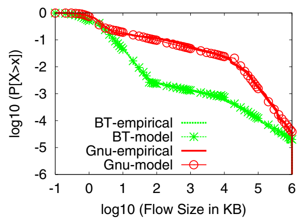
Characteristic	Web					P2P				
	Mean	Median	Std. Dev.	IQR	Skewness	Mean	Median	Std. Dev.	IQR	Skewness
Flow size (KB)	21.50	2.53	341.92	7.38	44.03	362.40	1.17	12470	1.89	192.13
Flow Inter-Arrival (sec)	0.11	0.007	3.53	0.016	26.05	1.77	0.18	17.21	0.39	48.69
Flow duration (sec)	13.32	0.40	56.71	1.80	14.48	123.54	24.80	274.37	93.30	7.61

Table 6: Flow-level summary statistics of Gnutella and BitTorrent

Characteristic	Gnutella					BitTorrent				
	Mean	Median	Std. Dev.	IQR	Skewness	Mean	Median	Std. Dev.	IQR	Skewness
Flow size (KB)	1159.40	1.89	15549	2.73	94.68	84.95	0.96	11189	2.10	292.31
Flow Inter-Arrival (sec)	2.30	0.21	22.22	0.51	30.15	2.46	0.42	20.25	0.99	49.78
Flow duration (sec)	89.35	9.70	386.22	25.60	8.12	135.43	33.20	221.41	180.90	3.03



(a) Web and P2P



(b) Gnutella and BitTorrent

Figure 2: CCDF of flow sizes

many mice and elephant flows, and possibly alters the mix of these flow types in today's IP networks. We elaborate on this phenomenon in Section 3.1.3.

We examined the tails of the flow size distributions using CCDF plots. Figure 2(a) presents the CCDF of flow sizes for Web and P2P. In the body of the distribution, P2P flows are smaller than Web flows, but in the tail (specifically, the upper 3.5% of flows after the "crossover" point) P2P flows are larger than Web flows. Also, the tail of the Web flow size distribution decays more quickly than the corresponding P2P distribution. These observations provide further evidence of P2P's large elephant-sized flows.

3.1.2 Gnutella and BitTorrent Flow Sizes

Table 6 indicates that Gnutella flow sizes are larger and more dispersed than BitTorrent flow sizes. The empirical CDF for the two P2P variants in Figure 1(b) shows that both applications generate a similar percentage of small-sized flows (e.g., 5 KB or less). Many of these smaller flows are the result of control information exchanged between peers, which is a byproduct of the distributed nature of P2P protocols. The ratio of large-sized to total flows for BitTorrent is, however, less than that for Gnutella. For example, approximately 5% of BitTorrent flows are larger than 10 KB, whereas 17% of Gnutella flows exceed this size. The characteristics of these large-sized flows are analyzed next.

Figure 2(b) shows the CCDF of flow sizes of Gnutella and BitTorrent applications. Gnutella appears to generate more large-sized flows than BitTorrent. BitTorrent uses *file segmentation* to split an object into multiple equal-sized "pieces" (256 KB each by default), and downloads these pieces from either the same or different peers using parallel flows. In contrast, Gnutella typically downloads the entire object from a single peer. As a result, we observe *fewer large flows in BitTorrent than Gnutella.*

3.1.3 Mice and Elephant Phenomenon

Table 7 shows the percentage of mice and elephant flows among the total flows contributed by different applications.

Table 7: Mice and elephant flow breakdown

Application	Mice		Elephants	
	% Flows	% Bytes	% Flows	% Bytes
Web	75.78	8.89	0.04	15.35
P2P	92.93	0.47	0.81	93.43
Gnutella	83.41	0.14	3.05	93.14
BitTorrent	94.96	1.94	0.08	94.87

We observe that both categories of application generate many mice flows. Although the mice flows originating from Web applications are less prevalent than those from P2P applications, Web mice flows account for a relatively higher proportion of the total Web bytes than P2P mice flows account for the total P2P bytes. For example, approximately 9% of total Web bytes are from Web mice flows, whereas only 0.4% of total P2P bytes are transferred by P2P mice flows.

Both applications generate a small proportion of elephant flows. Nevertheless, these few elephant flows contribute a significant fraction of the total bytes; the elephant-sized Web flows contributed about 15% of the total Web-generated bytes, while the elephant-sized P2P flows contributed as much as 93% of the total P2P bytes. Network operators may be interested in bandwidth-limiting these long-duration "elephant" flows, or may be interested in assigning these flows lower priority. As P2P applications become more popular, we can expect networks to carry increasingly more elephant flows. Our results also indicate that *P2P elephant flows are significantly larger than Web elephant flows.*

We next analyze mice and elephant flows generated by Gnutella and BitTorrent. While both P2P applications have a similar proportion of mice flows, the BitTorrent mice flows account for a much higher percentage of byte transfers than Gnutella mice flows; that is, Gnutella mice flows are smaller, on average, than BitTorrent mice flows. As mentioned earlier, signalling between peers is a major contributor to the pool of P2P mice flows. Our data suggests that BitTorrent applications have more intense signaling activities compared to Gnutella, resulting in relatively larger mice flows.

In our data, Gnutella has a much higher percentage of elephant flows than BitTorrent, even though both Gnutella and BitTorrent elephant flows account for a comparable proportion of byte transfers. Thus, on average, BitTorrent elephant flows are larger than Gnutella elephant flows. We believe that the type of files exchanged using these P2P systems can provide an explanation for our observation. A 2005 study by CacheLogic[5] showed that a majority of Gnutella users shared mostly audio files (70%), whereas BitTorrent users shared more video files (47%). Video files are, on av-

[5] CacheLogic. Peer-to-Peer File Type Study, `http://www.cachelogic.com/home/pages/research/filetypestudy.php`