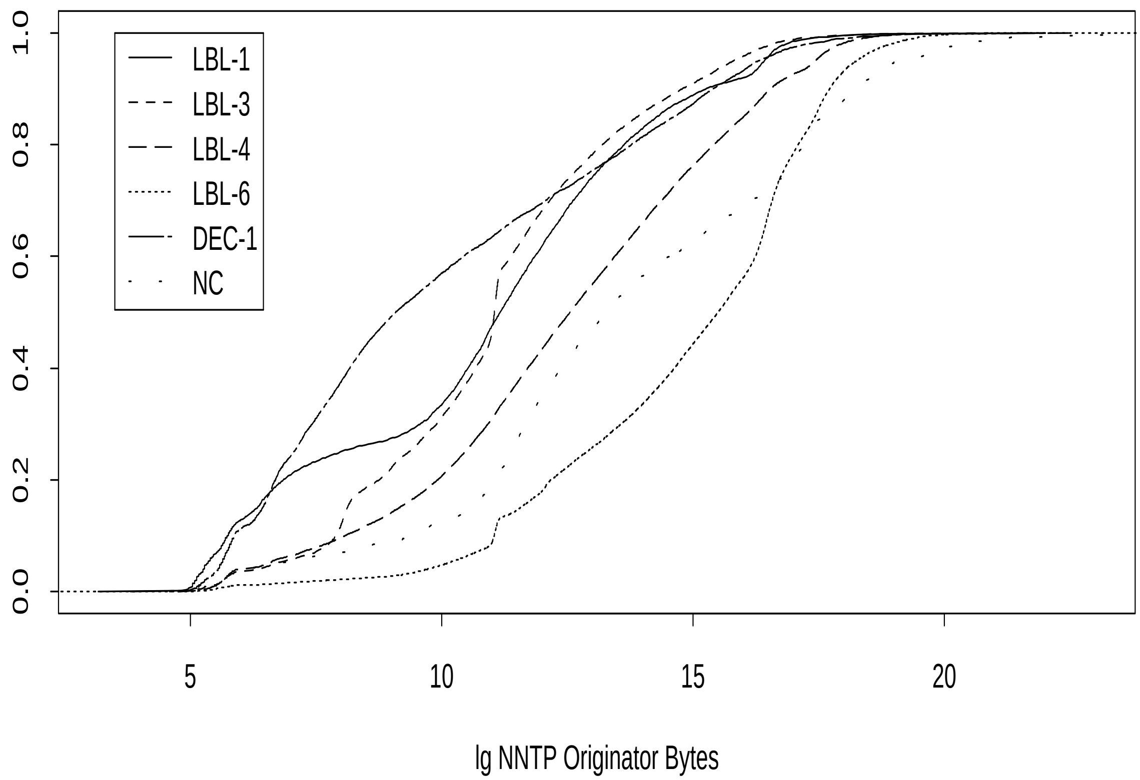

Figure 14: Distribution of NNTP Originator Bytes

(recall that the X axis is scaled logarithmically), suggesting that scaling is vital when modeling *nntp* traffic.

Given the great variation in originator bytes transferred, we decided to simply use a log-normal model to describe the connections, with the caveat that we do not expect the model to perform well in either scaled or unscaled forms (but we also do not expect empirical models to do well, either).

For the LBL test datasets, the log-mean ranged from 10.8 to 12.5, and the log-standard deviation from 2.6 to 3.3. For our fixed model we chose $\bar{x} = 11.5$ and $\sigma_x = 3$, respectively. We note that $\bar{x} \approx \log_2 3$ KB appears a bit higher than the median *nntp* article size of around 2 KB reported in [Adams92]. This difference probably means that, when an *nntp* server has a "fresh" article, it tends to have more than one.

Figure 15 shows the performance of the various models. As expected, none of the models does well due to the great variations from dataset to dataset, though scaling the models helps somewhat. In general, the analytic model performs acceptably only when scaled, and the empirical models only on the earlier LBL datasets and the scaled BC datasets. But the analytic model does as well as than either empirical model, indicating that the log-normal approximation is no worse than the inherent variation in the distributions.

Figure 29 in Appendix D shows the tail performance of the models. Again we show only the upper tail because with bulk transfer the lower tail is not of much interest. The unscaled models do quite badly, not surprisingly, given their poor overall performance. The upper 10% tail is only safe when scaling the models, and the upper 1% tail only when scaling empirical models. From this figure we conclude that the *nntp* models must always be scaled, and even then the otherwise somewhat successful analytic model is problematic due to grossly overestimating the upper 1% tail.

One final important point regarding modeling *nntp* originator bytes is that the distribution is not stationary but changes over the course of a day. Figure 16 shows the hourly $\bar{x}_{\text{orig}}$ for LBL-1 and LBL-4 non-failure *nntp* connections (this plot was made by constructing "superhours" as discussed in Section 3.11). We see considerable but not consistent variation. The peak-to-peak differences for both datasets is about a factor of 3.4; but LBL-1's connections tended to be largest in the middle of the night, with secondary peaks during "prime-time" work hours. LBL-4's connections peaked during working hours and were lowest at precisely the time when LBL-1's were highest. The test datasets also showed a weekly pattern, with LBL-1 and LBL-4 (and to a lesser extent LBL-2) having minimal $\bar{x}_{\text{orig}}$ during weekends (with a peak-to-peak variation of about a factor of 3), while LBL-3 had a *maximum* $\bar{x}_{\text{orig}}$ on Saturdays (peak-to-peak about a factor of 2).

The variation in the daily pattern may be due to the influence of key *nntp* gateways either propagating news as soon as it comes in (consistent with the LBL-4 case) or waiting till the

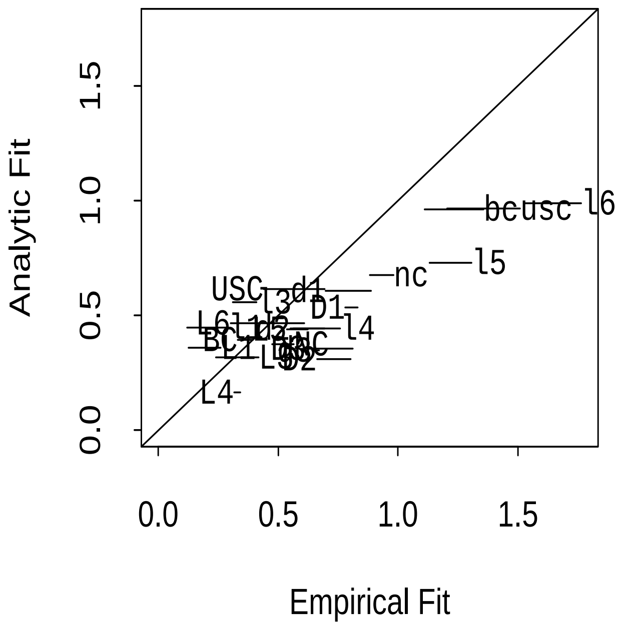

Figure 15: Empirical vs. Analytic Models for NNTP Originator Bytes

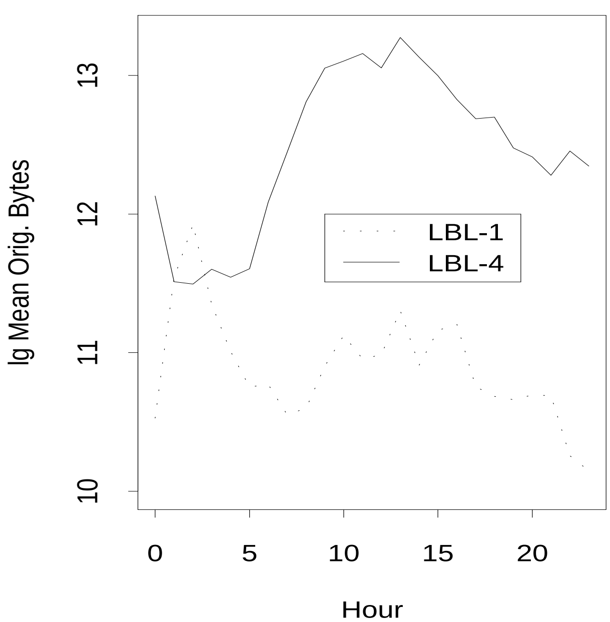

Figure 16: Daily Variation in $\log_2$-mean of LBL NNTP Originator Bytes

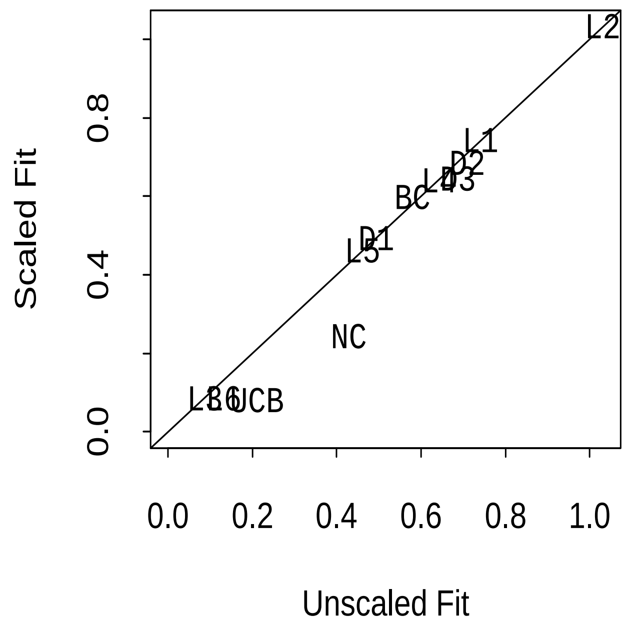

Figure 17: Interarrivals for NNTP

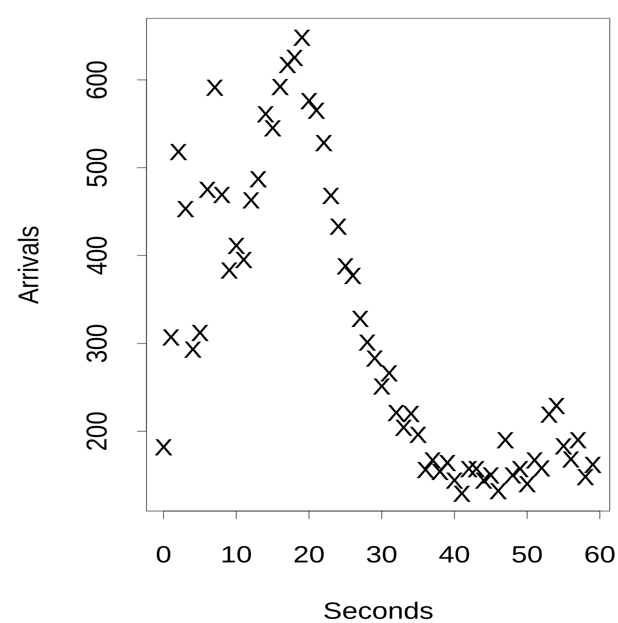

Figure 18: One-Minute Variation in DEC-2 NNTP Arrivals

late-night hours to take advantage of minimal loads (LBL-1). The weekly variation is more difficult to explain. We expect that most news articles are written during the week, and it seems unlikely that *nntp* gateways would queue a significant number of articles till the weekend. So the strong LBL-3 Saturday peak remains a puzzle.

5.3 NNTP Responder Bytes

As seen in Table 7 above, there is in general much less variation in the bytes sent by an *nntp* responder than by the originator. For all of the datasets except LBL-5, BC and NC, the responder sent fewer than 1500 bytes in 82% or more of the connections. For LBL-5 this value was 77%, for BC 65%, and for NC 63%. Thus we decided not to model *nntp* responder bytes, as in general the datasets do not show interesting variations. We did compute the correlation coefficients between $\log_2$ of the originator and responder bytes and found a range from 0.37 (NC and UCB) to about 0.8 (USC, BC). In general we would expect to find positive correlation since the more articles offered by the originator to the responder, the more replies the responder must generate.

5.4 NNTP Duration

Since *nntp* is a bulk-transfer protocol and not interactive, we do not model connection durations, because these are presumably dominated by networking latencies and not a fundamental aspect of the *nntp* protocols. Similarly, below we do not model *smtp* or *ftp* durations.

5.5 NNTP Interarrivals

Figure 17 shows the scaled and unscaled interarrival models.[8] The results appear puzzling. The UCB, LBL-3, LBL-6, and to some degree NC arrivals all appear well-modeled as Poisson processes with hourly rates corresponding to those in Figure 2. The other datasets, including the remaining LBL datasets, are poorly fitted both when scaled and when unscaled, indicating that they are *not* Poisson processes.

The poor fit to the BC data is in part due to its very low *nntp* arrival rate: less than two connections per hour on average. The other poorly-fitted datasets turn out to have interesting periodic behavior. In particular, *nntp* arrivals have a definite one-minute periodicity about them. Figure 18 shows the number of DEC-2 *nntp* connections that arrived during each second (i.e., ignoring minutes and larger units of the arrival time). Clearly, arrivals tended to show up at about 19 seconds past the minute, though some tended to arrive about 7 seconds past. All of the *nntp* datasets show this pattern to varying degrees except for LBL-3; LBL-4 shows two distinct spikes. With NC, UCB, and LBL-6, the spike is quite sharp. With the other datasets, it is broad, like in Figure 18. The sharp spikes mean that only a relatively small fraction of the interarrivals are skewed, evidently enough to preserve sufficient approximation to a Poisson process.

We also investigated one-hour variation. We found a three-minute pattern in LBL-1, five-minute patterns in LBL-5 and UCB, fifteen-minute patterns in LBL-6 and BC, and a less-strong twenty-minute pattern in DEC-2.

Figure 30 in Appendix D summarizes the tails corresponding to the scaled and unscaled models. Both models underestimate the lower 10% tail but overestimate the lower 1% tail; the upper 10% tail is well modeled but the upper 1% considerably underestimated. As in Figure 15 above, we see little difference between the scaled and unscaled models.

[8] The actual points for DEC-3 and LBL-4, and for LBL-3, LBL-6 and UCB, overlapped on this plot and became hard to distinguish, so we have added some horizontal bias away from the diagonal. These points all actually lie on the diagonal.

6 SMTP

6.1 Overview of SMTP Connections

Table 8 summarizes the *smtp* connections. Again, Appendix B summarizes the reasons for removing the connections marked as rejects. Based on the values for $\max_{\text{orig}}$ it is clear that *smtp* is sometimes used to transfer quite large files, though that is not its main purpose.

There is quite a bit of variation in $\bar{x}_{\text{orig}}$ (and just about none in $\bar{x}_{\text{resp}}$). In [WLC92] the authors note that the UK *smtp* data show a substantially higher (arithmetic) $\bar{x}_{\text{orig}}$ than for the LBL-1 and LBL-2 datasets reported in [Paxson91]. They attribute this difference to the fact that since the U.K. academic network (JANET) was not at that time fully connected to the Internet, U.K. users were more likely to use *smtp* to transfer files. The large UK σ_{orig} variance supports this hypothesis. The DEC traffic has similar σ_{orig} values, and Mogul also states in [Mogul92] that an "FTP-by-mail" facility is responsible for about 150 rather lengthy *smtp* messages at DEC-WRL each day. It is less clear whether this theory explains the large NC message sizes.

Another explanation is that perhaps the DEC, NC, and UK traffic tends to make more *smtp* "hops", each of which adds a `Received` header to the mail message [RFC822], pushing up the average number of bytes[9]. One would expect the greater number of hops to be correlated with "wider" wide-area traffic, presumably a property of the NC and especially the UK traffic, as these sites are at inter-network gateways. But this explanation does not address why the DEC traffic might tend to make more hops, unless due to the structure of DEC's internal mail gateways.

[9] A check of one of the author's mail folders revealed an average `Received` header length of more than 100 bytes.

Dataset	# Conn	# Rej	$\bar{x}_{\text{orig}}$	σ_{orig}	$\max_{\text{orig}}$	$\bar{x}_{\text{resp}}$	σ_{resp}	$\max_{\text{resp}}$
LBL-1	38,481	286	1.4KB	×2.8	2.1MB	331B	×1.2	1.9KB
LBL-2	51,240	572	1.5KB	×2.9	7.2MB	334B	×1.2	6.5KB
LBL-3	75,418	333	1.6KB	×2.6	1.6MB	334B	×1.2	2.9KB
LBL-4	92,694	1583	1.7KB	×3.0	1.2MB	335B	×1.3	2,980KB
LBL-5	123,741	446	1.7KB	×2.9	2.4MB	320B	×1.3	8.0KB
LBL-6	207,485	6,567	1.9KB	×3.0	37.0MB	321B	×1.3	9.5KB
BC	8,428	121	1.3KB	×2.8	1.1MB	324B	×1.3	10.2KB
UCB	16,929	61	1.3KB	×3.0	0.5MB	334B	×1.3	2.0KB
USC	3,498	3	1.4KB	×2.3	0.1MB	337B	×1.2	1.6KB
DEC-1	25,160	19	2.0KB	×3.1	2.5MB	340B	×1.2	4.7KB
DEC-2	10,777	5	2.1KB	×3.5	4.9MB	341B	×1.2	4.7KB
DEC-3	31,631	70	2.0KB	×3.2	5.1MB	338B	×1.2	3.5KB
NC	26,161	511	1.9KB	×2.9	1.8MB	340B	×1.4	10.6KB
UK	10,729	129	1.9KB	×3.3	4.6MB	319B	×1.3	6.0KB

Table 8: Summary of SMTP Connections

We see a definite trend in the LBL data indicating larger and larger mail messages. As discussed in [Paxson93], LBL's wide-area traffic did become "wider" during the 29 month period spanned by the LBL datasets, in agreement with the "hops overhead" explanation.

6.2 SMTP Originator Bytes

When modeling the number of bytes sent by the *smtp* originator, we found that nearly all connections transferred more than 300 bytes, while the connections transferring fewer bytes showed sporadic distributions. We hypothesize that the first 300 bytes of these connections constitute a more-or-less fixed overhead, and that connections with fewer total originator bytes correspond to "failures": either invalid email addresses or busy remote machines unable to accept mail at the moment. In constructing our models we therefore removed any connections of $\leq$ 300 bytes (anywhere from 0.6% to 2.3% of all connections) and subtracted 300 bytes from the remaining connections.

We found the distribution of *smtp* originator bytes to be bimodal, not surprisingly given that *smtp* is also used to transfer files. We model the distribution using two log-normal distributions, one (called f here) for the lower 80% of the data, and one for the remaining 20% (g). Figure 19 shows this model's fit to the LBL-3 test data after removing failures and subtracting 300 bytes; the horizontal line indicates the dividing line between using distribution f (below the line) and g (above).

For our fixed model, we found the mean of distribution f to range from 9.90 to 10.16, and chose $\bar{x} = 10$; the standard deviation ranged from 1.42 to 1.52, and we chose $\sigma_x = \log_2 2.75 \approx 1.46$. For g, the upper distribution, the means ranged from 8.43 to 9.06, and the standard deviations from 2.55 to 3.52. We chose $\bar{x} = 8.5$ (since three of the test datasets had means quite close to 8.5) and $\sigma_x = 3$. Note that, while the distribution g has a lower mean than distribution f, we use only g's upper 20% tail, which is larger due to g's significantly higher standard deviation.

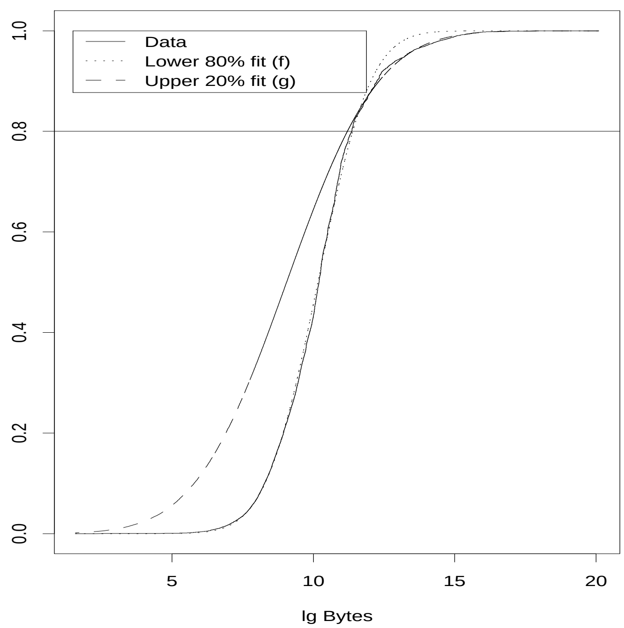

Figure 19: Log-Normal Fit to LBL-3 SMTP Originator Bytes

Figure 20 shows that this analytic model is highly successful compared to the empirical models. In virtually every case it performs as well or better than the empirical models, usually better. We also see that scaling consistently improves the performance of all three models, often substantially, indicating that there is considerable site-to-site variation in bytes transferred.

Figure 29 in Appendix D shows the tail distribution for this model. The models all do well, in general slightly underestimating the upper tails, except for the unscaled UCB model,

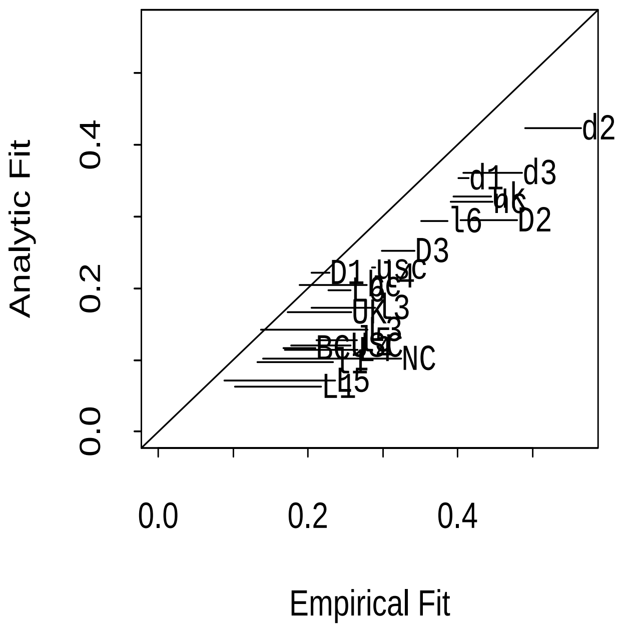

Figure 20: Empirical vs. Analytic Models for SMTP Originator Bytes

which is more severe in its underestimation.

As was the case for *nntp*, for *smtp* we found that the originator bytes distribution is not stationary. Figure 21 shows the hourly $\bar{x}_{\text{orig}}$ for LBL-1 and LBL-4 *smtp* connections after removing those less than 300 bytes and subtracting 300 bytes from the remainder. Unlike *nntp*, which suffered from inconsistent variations, here the pattern is more stable: connection sizes peak during off-hours, the evening and early morning, and reach minima during peak working hours. We conjecture that uses of *smtp* to transfer files and not messages typed in by users tend to happen off-hours and cause this pattern. Of the four test datasets, LBL-4 shows the greatest peak-to-peak variation, about a factor of 3.3, comparable to the *nntp* variation. The other datasets are closer to a factor of 2. Unlike *nntp*, we did not detect a noteworthy weekly pattern.

6.3 SMTP Responder Bytes

We did not model the distribution of the responder bytes in *smtp* connections, as the responder's role shows little variation. For the LBL test datasets, in about 75% (73% to 79%) of all connections the responder sent between 300 and 400 bytes, and more than 99% of the connections sent between 100 and 1000 bytes. Of all the datasets, LBL-5 had the lowest proportion of connections sending between 100 and 1000 bytes, still a very high 98%. We also found that the coefficient of correlation between $\log_2$ of the originator bytes and $\log_2$ of the responder bytes for the LBL datasets varied from .035 to .246; thus we found little interesting behavior to model in the responses. While reference [DJCME92] finds that *smtp*

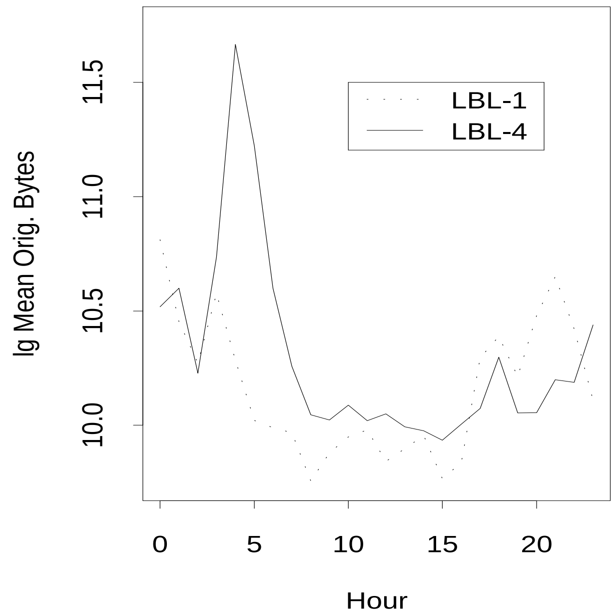

Figure 21: Daily Variation in $\log_2$-mean of LBL SMTP Originator Bytes

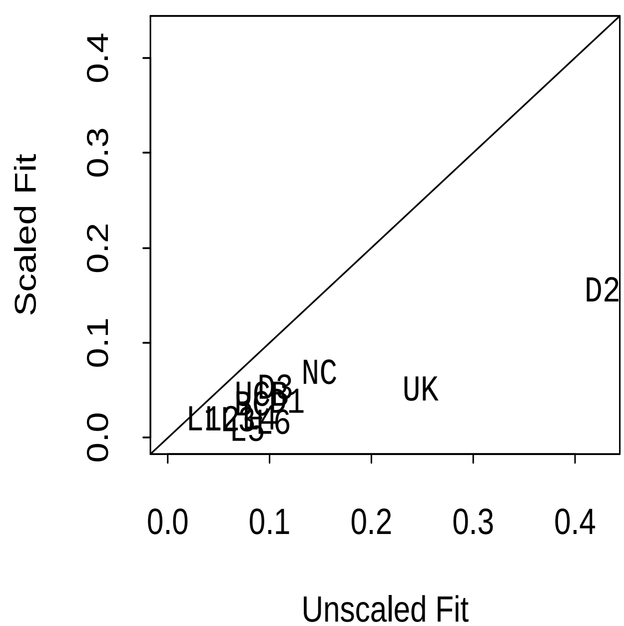

Figure 22: Interarrivals for SMTP

connections are strongly bidirectional, this finding must be interpreted with the rather fixed nature of the *smtp* responder in mind.

6.4 SMTP Interarrivals

Figure 22 shows the unscaled and scaled fits to the *smtp* interarrivals.[10] Both models do extremely well for almost all of the datasets, indicating these arrivals are well described by the pattern shown in Figure 2. We are somewhat puzzled that the BC interarrivals fared so well with the unscaled model, given the roughly three-hour shift between BC's arrival activity and LBL's as shown in Figure 2; evidently there is enough similar overlap during the busy 11AM-4PM times to bring the overall distributions into fairly close agreement. We do not have an explanation for the poor fit to DEC-2's interarrivals, though the traffic, which spanned the Thanksgiving holiday, is certainly atypical in one sense: only 6% of the DEC-2 connections originated from a DEC host, while for DEC-1 and DEC-3 the figure is 40-46%.

Figure 30 in Appendix D summarizes the tail behavior for the interarrival models. Both the scaled and unscaled models do quite well in the lower and upper tails, only underestimating the upper tail somewhat.

7 FTP

7.1 Overview of FTP Connections

Table 9 summarizes *ftpdata* connections. Each connection is unidirectional, with sometimes data flowing from the connection originator to the responder (corresponding to an *ftp* `get` command) and sometimes in the other direction. The "Get" column shows the percentage of connections that were `get` commands; the remainder were `put` commands. The next three columns show the (geometric) mean, standard deviation, and maximum for the number of bytes transferred. As before, Appendix B gives details regarding the connections we rejected.

Two rows are given for each of LBL-5 and LBL-6. As discussed in [Paxson93], a considerable portion of LBL traffic, particularly in LBL-5 and LBL-6, was generated by background scripts fetching weather maps from a remote anonymous ftp site. LBL-5* and LBL-6* show the *ftpdata* statistics with this traffic removed.[11]

Two rows are given for the UCB dataset. The first includes 2,315 connections of 74 bytes each, all but three of which were between the same two hosts, and 95% of which came between 30 and 45 seconds apart. The UCB* row summarizes the UCB data with this anomaly removed.

In testing our models below we used LBL-5*, LBL-6*, and UCB*.

There clearly is quite a range in $\bar{x}_{\text{bytes}}$, even day-to-day as shown in the DEC data (though the low-point there, DEC-2, includes Thanksgiving, and might therefore be uncharacteristic). We might be tempted to declare a trend towards increasing file sizes with time in the LBL datasets, save for the LBL-6 dataset, which shows a sharp drop. We do not know whether LBL-6 was atypical, or whether the mean file size simply fluctuates a great deal. The uniformly large values of σ_{bytes} shows that in general file sizes vary widely.

Finally, computing the coefficient of variation (i.e., $\sigma / \bar{x}$) for the *ftpdata* interarrival times gives values from 2.4 to 8.0, significantly higher than for the other protocols. If the arrivals came from a homogeneous Poisson process, then for $\sigma / \bar{x}$ we would get 1, and if they were perfectly periodic, 0. These high values show that the traffic is quite bursty. This result is not surprising, as a "multiple-get" file transfer results in a rapid succession of *ftpdata* connections, sometimes quite large (see below).

Table 10 summarizes the *ftpctrl* connections. We have not shown statistics for bytes transferred and duration of the *ftpctrl* connections themselves since the primary use of *ftpctrl* connections is to spawn *ftpdata connections*, either for file transfer or to list remote directories. Instead, we grouped with each *ftpctrl* connection its associated *ftpdata* connections. We considered an *ftpdata* connection to belong to a *ftpctrl* connection if it occurred during the span of the *ftpctrl* connection and was between the same two hosts.

The starred LBL rows summarize the LBL datasets with the weather-map traffic removed. Unlike with *ftpdata*, here we include LBL-3 and LBL-4 in the filtering, since weather-map traffic had a substantial influence on their *ftpctrl* connections. Again, we use the starred datasets in our analysis below.

The "Orphans" column lists the percentage of *ftpdata* connections for which we could not associate an *ftpctrl* connection. High percentages of orphans were often due to *ftpctrl* connections that were terminated by RST packets instead of FIN packets, which, as explained in Section 2.1 above, were not included in our analysis. The authors of [EHS92] reported about 3% of *ftpctrl* connections were terminated by RST packets. As seen in the Table there appears to be considerable variation in this value.

The "# Overlap" column lists the number of overlapping *ftpctrl* connections between two hosts. For our analysis we merged such overlaps into a single conversation. The large number of LBL-4 overlapping connections is almost all due to overlapping connections to one of the weather-map sites, as can be seen by the appreciably lower value for LBL-4*. In LBL-4 we observed up to five overlapping connections (typically four), all virtually identical in bytes transferred and

[10] Here again we have added some horizontal space between the points in the lower-left corner to aid legibility. The LBL points all have unscaled $\mu \leq 0.05$; the excursion up to $\mu = 0.1$ is an artifact of the plot.

[11] Weather-map *ftp* traffic comprised 3.2% of LBL-3 connections and 8.5% of LBL-4 connections. Excluding this traffic lowered LBL-3's mean to 3.2KB and raised LBL-4's mean to 4.0KB. The corresponding standard deviations rose to 18.0 and fell to 14.0, respectively.

Dataset	# Conn	# Rej	Get	$\bar{x}_{\text{bytes}}$	σ_{bytes}	$\max_{\text{bytes}}$
LBL-1	23,555	287	80 %	2.3KB	×15.3	54.0MB
LBL-2	27,917	335	92 %	2.4KB	×17.4	124.3MB
LBL-3	39,552	349	91 %	3.3KB	×17.7	61.6MB
LBL-4	65,860	335	86 %	3.8KB	×14.7	67.2MB
LBL-5	82,025	344	83 %	5.1KB	×16.0	176.5MB
LBL-5*	66,411	344	80 %	4.5KB	×16.0	176.5MB
LBL-6	123,773	464	89 %	3.8KB	×16.2	291.6MB
LBL-6*	86,464	464	91 %	2.1KB	×14.9	291.6MB
BC	5,199	58	97 %	2.5KB	×12.6	16.1MB
UCB	6,844	77	98 %	0.4KB	×11.5	22.1MB
UCB*	4,529	77	96 %	1.0KB	×13.5	22.1MB
USC	1,870	29	93 %	1.3KB	×14.5	4.7MB
DEC-1	7,970	6	100 %	2.2KB	×16.5	4.9MB
DEC-2	4,013	13	100 %	1.3KB	×17.1	7.1MB
DEC-3	6,775	25	99 %	1.9KB	×16.7	12.8MB
NC	19,076	183	98 %	1.8KB	×19.0	43.8MB
UK	10,018	58	97 %	3.4KB	×14.2	6.8MB

Table 9: Summary of FTP Data Connections

Dataset	# Conn	# Rej	Orphans	# Overlap	0 xfer	$\bar{x}_{\text{xfers}}$	σ_{xfers}	$\max_{\text{xfers}}$	$\bar{x}_{\text{bytes}}$	σ_{bytes}
LBL-1	3,757	51	5 %	57	19 %	3.3	×2.9	1,006	28KB	×15.2
LBL-2	5,312	72	5 %	49	25 %	3.2	×2.8	388	27KB	×17.0
LBL-3	7,920	93	5 %	135	19 %	2.7	×2.8	612	24KB	×17.8
LBL-3*	6,916	90	5 %	135	21 %	3.1	×2.9	612	30KB	×18.4
LBL-4	11,587	191	6 %	1,012	15 %	2.8	×2.7	1,951	24KB	×17.5
LBL-4*	7,941	189	7 %	112	17 %	3.3	×3.0	1,951	33KB	×17.6
LBL-5	18,501	1,227	5 %	160	15 %	2.2	×2.5	975	28KB	×12.3
LBL-5*	9,968	1,227	6 %	108	26 %	3.0	×3.0	975	31KB	×16.7
LBL-6	31,734	535	7 %	212	21 %	2.2	×2.5	2,996	30KB	×12.5
LBL-6*	12,470	535	10 %	196	24 %	3.1	×2.9	2,996	31KB	×16.8
BC	669	19	40 %	2	32 %	3.3	×2.7	426	13KB	×14.2
UCB	756	19	15 %	7	26 %	3.9	×2.6	350	12KB	×14.9
USC	272	6	26 %	2	22 %	3.8	×2.8	133	20KB	×14.5
DEC-1	727	8	6 %	18	26 %	5.4	×3.2	961	36KB	×15.6
DEC-2	491	8	6 %	14	13 %	5.0	×3.0	106	36KB	×17.8
DEC-3	811	17	14 %	18	25 %	4.8	×2.9	232	36KB	×15.3
NC	2,500	59	7 %	49	31 %	5.0	×2.9	392	26KB	×18.6
UK	1,733	35	5 %	133	24 %	3.4	×3.0	368	22KB	×16.0

Table 10: Summary of FTP Control Connections

duration, repeating every half hour for days on end. Evidently a number of weather-map scripts were run in the background on the same host and managed to synchronize. The large number of overlapping UK connections, on the other hand, is due to the high frequency of connections between pairs of popular hosts, such as one vendor's main Internet site in the U.K. connecting to the anonymous ftp archives of Washington University in Missouri. The authors of [WLC92] noted the Missouri site as the single most popular U.S. *ftp* site (and, indeed, Missouri was the most popular state in general for U.K.-U.S. traffic).

The next four columns in Table 10 show statistics regarding the number of *ftpdata* connections that occurred during each *ftpctrl* connection. The "0 xfer" column lists the percentage of all *ftpctrl* connections that did not have any associated *ftpdata* connection. These numbers are lower than the 44% reported in [EHS92] because the authors of that paper were able to distinguish between file transfers and remote directory listings; we consider any *ftpdata* connection to be a "file transfer". Presumably a large proportion of the "0 xfer" connections are due to failed attempts to provide log-in information to the remote host. Still, the rates are surprisingly high.

The $\bar{x}_{\text{xfers}}$ and σ_{xfers} columns give the geometric mean and standard deviation for the number of files transferred, given that at least one file was transferred. That the mean is substantially higher than one is not surprising since we classify remote directory listings as file transfers, and probably the most common use of *ftp* is to connect to a remote archive site, do several listings to find the file or files of interest, and then transfer those files. We did find the values for $\max_{\text{xfers}}$ surprising, as we would expect that large sets of related files would be grouped together into a single archive file, unless the archive would be too large. This latter hypothesis turns out to be the case: the 1,951 files transferred at one time during LBL-4 together totaled 93MB, certainly too much to conveniently pack into an archive. Similarly, the 961 transferred during DEC-1 totaled 50MB.

The $\bar{x}_{\text{bytes}}$ and σ_{bytes} columns show the geometric mean and standard deviation for the total number of bytes transferred via *ftpdata* connections during each *ftpctrl* connection (for those connections with at least one *ftpdata* transfer). We note that these means are 5-10 times greater than those for *ftpdata*, an increase larger than that due simply to the multiplying effect of $\bar{x}_{\text{xfers}}$. We suspect that this disparity is due to a typical *ftpctrl* connection including at least one true file transfer. As files will tend to be significantly larger than directory listings, the mean number of transferred bytes will approach the mean file size, and not be held down, as are the *ftpdata* connection summaries, by a large number of smaller directory listings. The σ_{bytes} values are quite large, again showing a wide range in transfer sizes.

7.2 FTP Connection Bytes

We model the bytes transferred during an *ftpdata* connection using a log-normal distribution. Figure 23 shows this model fitted to the first half of the LBL-4 dataset. While the model appears to match the overall shape, a number of clumps and spikes make the actual distribution rather irregular. For example, LBL-4 has a spike of 1,269 connections, each transferring 1,856 bytes. For the most part, unfortunately, these spikes do not occur in predictable locations, making it difficult to incorporate them into our analytic model. Such unpredictability also impairs the ability of empirical models to fit other datasets. One spike stands out, however, being present in all the DEC datasets, the NC dataset, LBL-4, and LBL-5 (but not LBL-6). This spike occurs at 524,288 bytes (i.e., exactly 2^{19} bytes), a size often used when splitting a large distribution archive into manageable pieces.

Using the LBL test datasets, we found a range for the log-mean of 11.27 to 11.88, and chose $\bar{x} = \log_2 3000 \approx 11.55$ for our fixed model. The log-standard deviation varied from 3.83 to 4.21; we chose $\sigma_x = 4$.

Figure 24 shows the comparison between analytic and em-

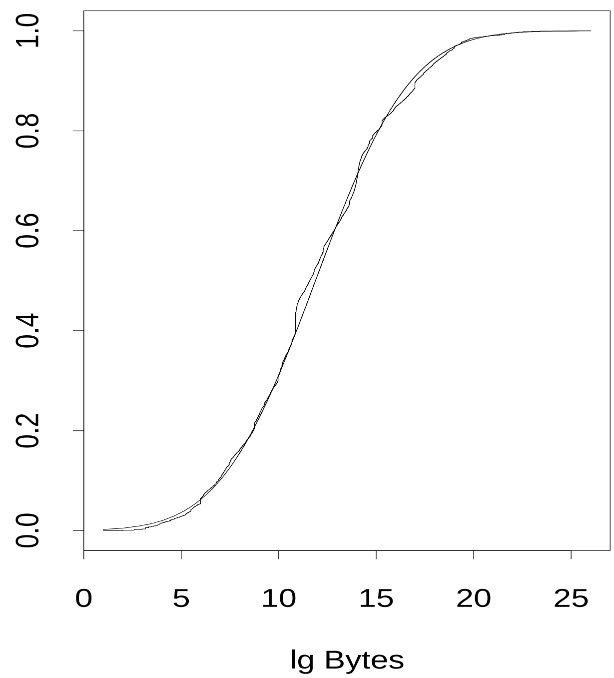

Figure 23: Log-Normal Fit to LBL-4 FTP Data Bytes

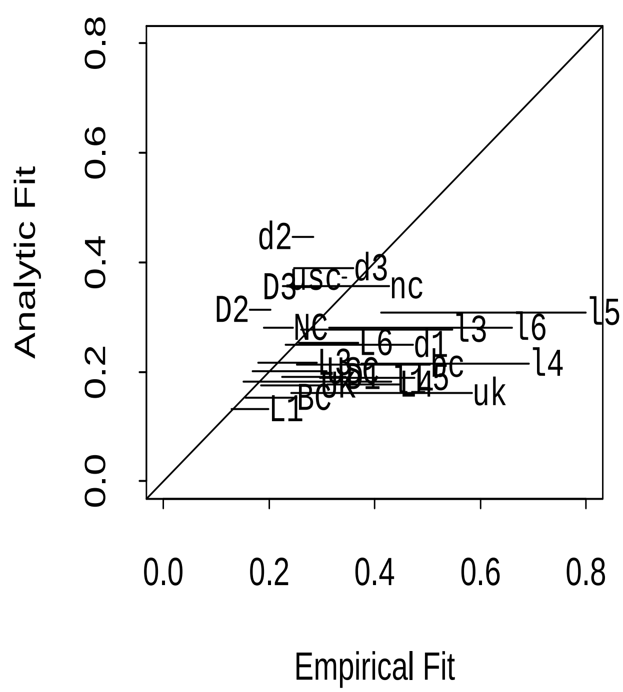

Figure 24: Empirical vs. Analytic Models for FTP Data Bytes

pirical models for the bytes transferred during an *ftpdata* connection. For the most part, scaling considerably improves the fit, as we might expect given the wide range in $\bar{x}_{\text{bytes}}$ shown in Table 9. The LBL-2 model almost always fits better than the UCB model, even though the UCB model has had its anomalous spike at 74 bytes removed. In general the analytic model does well, though it suffers somewhat when fitted with the DEC datasets, which are very noisy.

For modeling *ftpdata* data bytes we are not particularly concerned with the degree of fit in the lower tail. The upper tail, on the other hand, is of particular importance because large file transfers can consume a tremendous amount of network resources. Figure 29 in Appendix D summarizes the tail performance of the various models. While most of the models fit the upper 10% tail fairly well (with the exception of the unscaled UCB model), all except the scaled LBL-2 model fare poorly in the upper 1% tail, each overestimating the tail except, again, for unscaled UCB. Thus all of these models must be used with particular care concerning their predictions for large *ftpdata* connections. The overestimation of the analytic model might be understood at least in part by the tendency to split huge files into several pieces (see discussion of the 2^{19} byte spike, above). In this case we would expect such large files to be fetched together as a group, and hope that models of the total bytes transferred during an entire *ftp* conversation might more accurately predict the upper tail. Unfortunately those models actually prove worse; see below.

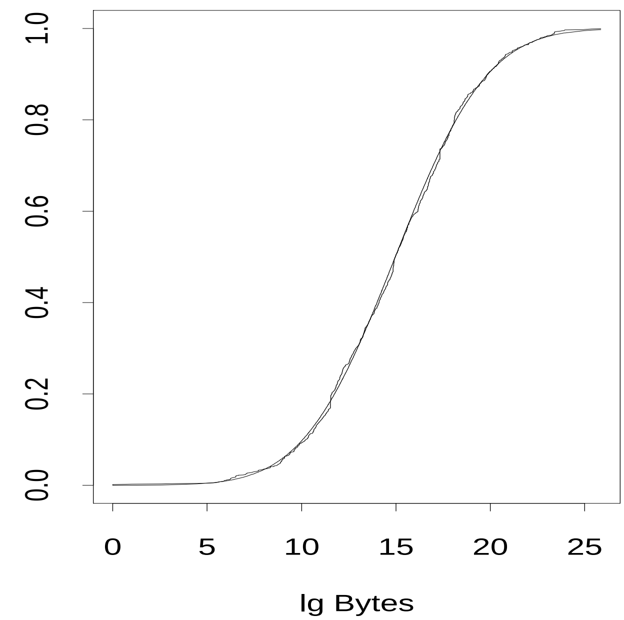

Figure 25: Log-Normal Fit to Bytes in LBL-1 FTP Conversations

7.3 FTP Conversation Bytes

Perhaps more important than an *ftpdata* bytes-transferred model is a model for the total number of *ftpdata* bytes transferred due to an *ftpctrl* connection. Such a model gives an indication of the total impact of each file transfer conversation. Figure 25 shows this distribution for the LBL-1 test dataset, again fitted to a log-normal model. In this case the fit is visually quite satisfying, and indeed an A^2 test indicates this fit is valid at the 1% level. Unfortunately this fit is also the best of those to the LBL test datasets; the others fail validity.

For the LBL test datasets, the log-mean ranged from 14.85 to 15.20, and the log-standard deviation from 3.82 to 4.18. For the fixed analytic model we took $\bar{x} = 15$ and $\sigma_x = 4$.

Figure 26 summarizes the performance of the models. The overall variance of the analytic model is fairly low; ignoring the BC datasets, the fits all fall in or quite close to the range $0.2 \le \mu \le 0.3$. Scaling has only a minor effect on the caliber of the fits (except for BC), and in some cases worsens them. Since scaling is beneficial for modeling *ftpdata* connection bytes but not *ftp* conversation bytes, we conjecture that the "mix" between short directory listings and larger file transfers varies considerably from site-to-site. Such variation would mean that scaling would aid *ftpdata* bytes considerably more than *ftp* conversation bytes, since the latter are dominated by the actual files transferred and are relatively unaffected by

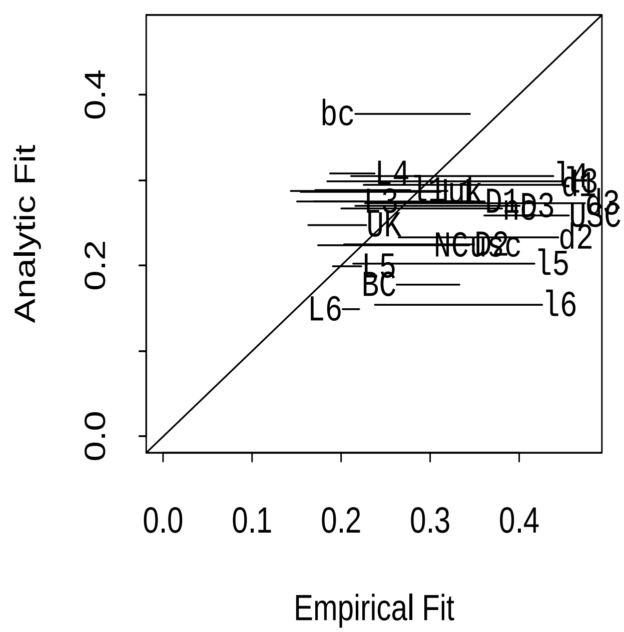

Figure 26: Empirical vs. Analytic Models for FTP Conversation Bytes

directory listings.

The analytic model performs noticeably better than the UCB empirical model, but not as well as LBL-2, which overall does quite well. We note that the authors of [DJCME92] reported that 80% of *ftp* conversations transfer less than 10 KB. But once we remove the 20-30% of conversations that did not transfer any data, half of the remainder transfer more than 32 KB, and a sixth transfer more than 500 KB. Thus if a file transfer conversation is not a “failure”, it should not be assumed small.

As with *ftpdata* connections, we again are most concerned with the behavior of the models in the upper tails, summarized in Figure 29 of Appendix D. Each model except unscaled UCB does well in the upper 10% tail, but the analytic model greatly overestimates the upper 1% tail, even with scaling, and only scaled UCB performs well in both regards.

# Items	Range	$Range_{LBL}$
0	13-32 %	17-26 %
1	10-24 %	20-24 %
2	8-15 %	12-15 %
3	5-11 %	8-10 %
4	6-8 %	6-7 %
5	4-6 %	4-5 %
6	3-5 %	3-4 %
7-10	7-13 %	7-8 %
11-20	5-14 %	5-7 %
21-99	3-9 %	3-5 %
100+	0-1 %	0-1 %

Table 11: FTP Data Items Per FTP Conversation

7.4 Data Items Per FTP Conversation

While *ftpdata* connection bytes and total *ftp* conversation bytes are closely related, we were unable to produce a good model of the number of *ftpdata* connections in each *ftp* conversation, both because the distribution varies considerably from dataset to dataset and because of the heavy upper tail in the distributions. Table 11 lists the range over all of the datasets for the distribution. For example, for one dataset 8% of all conversations transferred 2 items, while for another 15% did. Site-to-site variation is considerable. Furthermore, there is too much mass in the upper tail to accommodate a geometric distribution, our most likely candidate for modeling.

The third column lists the same ranges for just the six LBL datasets, with the weather-related *ftp* conversations removed. Here the variation is substantially less, indicating that the mix at a particular site is fairly stable over time. Looking at the Table one might wonder whether the LBL data is “holding down” the lower-end of the ranges in the second column for 4 or more items. This turns out not to be the case; removing the LBL data from the tabulation does not change any of the ranges for 4 or more items except to narrow the 11-20 range from 5-14% to 6-14%. Thus the LBL data is not particularly atypical.

7.5 FTP Interarrivals

Figure 27 shows the fits of the scaled and unscaled arrival models for *ftp* conversations. Overall both models perform quite well, with the maximum μ for the unscaled model about 0.3 and for the scaled model 0.2. We found periodicity in the DEC datasets, with arrivals peaking on the hour and the half hour, too great an interval to much affect individual interarrivals.

The deviation in the tails is similarly quite low, as shown in Figure 30 in Appendix D, with almost no distortion in the 10% tails except for the lower tail of the unscaled model, and only moderate distortion in the 1% tails.

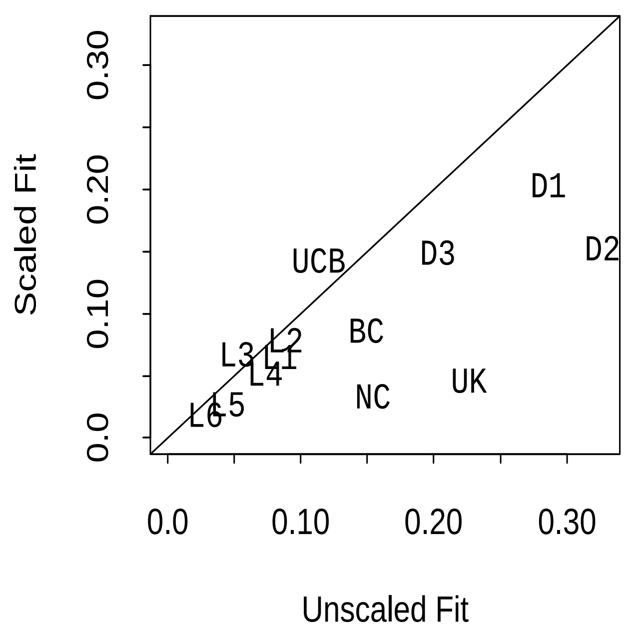

Figure 27: Interarrivals for FTP

8 Summary

We have presented a number of analytic models for describing the characteristics of *telnet*, *nntp*, *smtp*, and *ftp* connections, drawn from wide-area traces collected from seven different sites, comprising more than 2.5 million connections. While these models are rarely exact in a statistical sense, we developed a methodology for comparing their effectiveness to that of other models, and found that in general they capture the essence of the connections as well or better than the empirical *tcplib* library. We also compared the models to an empirical model derived from a one-month trace of traffic at the

Protocol	Variable	Model	Parameters	Abs.	Rel.	Tails
telnet	orig. bytes	lg-extreme	$\alpha = \log_2 100; \beta = \log_2 3.5$	0	0	u: over
	resp. bytes	lg-norm, upper 80%	$\bar{x} = \log_2 4500; \sigma_x = \log_2 7.2$	+	0	u: over
	duration secs.	lg-norm	$\bar{x} = \log_2 240; \sigma_x = \log_2 7.8$	0	0	u: Over
	resp. / orig.	lg-norm	$\bar{x} = \log_2 21; \sigma_x = \log_2 3.6$	+	0	u: good/okay; l: good/okay
	resp. / dur.	exp., 0-90% resp.	$\bar{x} = 30$	0	+	both good
	resp. / dur.	lg-norm, 90-100% resp.	$\bar{x} = 5.3; \sigma_x = 1.5;$	0	0	u: okay; l: OVER/good
nntp	orig. bytes	lg-norm	$\bar{x} = 11.5; \sigma_x = 3;$	−/0	0	u: Over
smtp	orig. bytes	lg-norm + 300B, lower 80%; lg-norm + 300B, upper 20%	$\bar{x} = 10; \sigma_x = \log_2 2.75;$ $\bar{x} = 8.5; \sigma_x = \log_2 3;$	0/+	+	u: good
ftp	conn. bytes	lg-norm	$\bar{x} = \log_2 3000; \sigma_x = 4;$	0/+	0	u: over
	conv. bytes	lg-norm	$\bar{x} = 15; \sigma_x = 4;$	0	0	u: over

Table 12: Summary of Analytic Models of Connection Characteristics

Protocol	Abs. Fit	Scaling Helpful?	Lower Tail	Upper Tail
telnet	+	Sometimes	good	good
nntp	−	No	1% over, 10% under	1% under
smtp	+	Sometimes	over	under
ftp	+	Sometimes	okay/over	okay/good

Table 13: Summary of Analytic Interarrival Models

Lawrence Berkeley Laboratory, which we found in general to be slightly better at modeling the traffic than the analytic models.

Table 12 summarizes the models characterizing the different protocols' individual connections. The "Variable" column lists the random variable being modeled, where "orig." stands for the bytes sent by the connection originator, "resp." for those sent by the responder, "conn. bytes" the total number of bytes transferred during the connection (for *ftpdata*), and "conv. bytes" the total bytes transferred during a *conversation* (an entire *ftp* session). For *telnet*, we also modeled ratios between some of these variables, to capture their interdependence.

The "model" column lists the models used. Almost all first apply a $\log_2$ transformation to the data. One model is log-extreme, where the extreme distribution is defined by Equation 4; one is exponential; and the remainder are log-normal. Four of the models have restrictions. The *telnet* responder bytes model describes only the upper 80% of the responses. The *telnet* "resp. / dur." models describe the ratio of the responder bytes to the connection's duration. The first such model does so for those connections whose number of responder bytes fell into the lower 90% of all connections. The second model describes this ratio for those connections in the upper 10% of all responses. Finally, the *smtp* originator model uses parts of two different log-normal distributions in its description. The lower 80% of the originator distribution is modeled using the lower 80% of the first log-normal distribution; similarly, the upper 20% is modeled using the upper 20% of the second log-normal distribution.

The "Parameters" column gives the parameters we used for the fixed (i.e., unscaled) version of the model. The "Abs." column summarizes the quality of the model's fit in absolute terms: how well we assess the model as describing the random variable's distribution. A "0" indicates the model describes it adequately, a "+" that the model describes it well, and a "−" that it does poorly. When two values are given, the first is for the unscaled version of the model and the second for the scaled version. When only one value is given, scaling did not significantly improve the model's fit. The Table shows that we assess the models as being at least adequate for every random variable except when modeling *nntp* originator bytes; in that case the scaled version of the model is required for an adequate description.

The "Rel." column compares the model's performance to that of the two empirical models, one constructed from the UCB dataset and corresponding to *tcplib*, and one constructed from the LBL-2 dataset. A "0" indicates that the analytic model performs about equally and a "+" that it performs better. In all cases we found the analytic model does overall at least about as well as the empirical models, though for some of the "0" entries it did somewhat better than the UCB model and somewhat worse than the LBL-2 model.

The final column summarizes the analytic model's performance in modeling the tails. A "u:" entry gives the fit to the upper tail and an "l:" entry to the lower tail. A value of "over" indicates the model substantially overestimates the 1% tail; "Over" that it also somewhat overestimates the 10% tail; and "OVER" that it grievously overestimates both the 1% and 10% tail. Similarly for "under". For models that do well describing their tails, we chose subjective evaluations of "okay" and "good". Some models have two evaluations reported for one of their tails; in this case the first is for the unscaled model and the second for the scaled model.

Table 13 summarizes the interarrival models for each protocol. Here "Abs. Fit" summarizes the absolute fit of the model using the same notation as before.

Since we did not compare the analytic interarrival models to empirical ones (due to difficulties in constructing such empirical models), we omit the "relative" fit. The "Scaling Helpful?" columns indicates whether scaling substantially improved the model. The lone "No" entry, for *nntp*, reflects our finding that the *nntp* connection arrival process is not Poisson. The other "Sometimes" entries indicate that for many of the datasets scaling was not needed to produce good fits, and the arrivals can be modeled as a non-homogeneous Poisson process with hourly rates given by Figure 2. For those datasets requiring scaling, the arrivals can also be better modeled as a non-homogeneous Poisson process with different hourly rates than those in Figure 2.

The last two columns summarize the arrival models fit in the lower and upper tails. When two values are given, such as for *ftp*, then the first is for the unscaled model and the second for the scaled model.

Table 1, at the beginning of the paper, states our major conclusions. Here we summarize our additional findings:

- The ratio between bytes sent by a user in a remote-login session and those sent back by the remote computer is about 1:20.
- Of *ftp* conversations that are not "failures" (no data transferred), half transfer more than 32 KB, and a sixth transfer more than 500 KB.
- *smtp* and *nntp* connections show variations in size over the course of the day, with the largest *smtp* connections coming during evening and early morning hours, while the peaks of *nntp* varied considerably.
- *rlogin* traffic can be described by models for *telnet* traffic (see Appendix C), but requires scaling for acceptable fits, and even then does not in general fit as well as *telnet* traffic.
- We believe the site-to-site and month-to-month variations in network traffic characteristics are in part responsible for the success of the analytic models: the inter-site differences are large enough that analytic models tend to be just as good a compromise among the varying datasets as empirical models.

The essence of the argument presented in this paper is that while wide-area traffic cannot be easily modeled exactly, if we can abide some inexactness then we can reap the benefits of using analytic models instead of empirical ones, without any relative loss of accuracy. We believe the approach discussed in this paper will prove beneficial for developing future analytic models and for gauging their effectiveness.

9 Acknowledgments

This work would not have been possible without the support and patience of Van Jacobson and Domenico Ferrari. I am also much indebted to Peter Danzig and his coauthors for the UCB, USC, and BC datasets; Jeff Mogul for the DEC datasets; Ian Wakeman and Jon Crowcroft for the UK dataset; Wayne Sung for the NC dataset; and especially to Ramón Cáceres and Sugih Jamin, who between them made all of the non-LBL datasets available.

The LBL traces were gathered with the help of Craig Leres and Steve McCanne. Craig was also helpful in understanding *nntp*-related phenomena.

The Bellcore traces were gathered by D. V. Wilson.

I also want to thank Terry Speed and particularly Sally Floyd for valuable discussions on both modeling the data and presenting the results; and Domenico Ferrari, Sally Floyd, Van Jacobson, and Jon Crowcroft for their many helpful comments on earlier drafts of this paper.

This work was supported by the Director, Office of Energy Research, Scientific Computing Staff, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

References

[Adams92] R. Adams, *USENET Readership Summary Report for Oct 91*, Internet Society News, 1(1), Winter, 1992.

[AD54] T. W. Anderson and D. A. Darling, *Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes*, Ann. Math. Statist. 23, pp. 193-212, 1954.

[Bryan67] G. E. Bryan, "JOSS: 20,000 Hours At A Console", Proc. of the Fall 1967 AFIPS Conference, Vol. 31.

[Càceres89] R. Cáceres, "Measurements of Wide Area Internet Traffic", *Report UCB/CSD 89/550*, Computer Science Division, University of California, Berkeley, California, 1989.

[CW91] J. Crowcroft and I. Wakeman, "Traffic Analysis of some UK-US Academic Network Data", Proceedings of INET'91, Copenhagen, June, 1991.

[DS86] R. B. D'Agostino and M. A. Stephens, editors, "Goodness-of-Fit Techniques", Marcel Dekker, Inc., 1986.

[DJ91] P. Danzig and S. Jamin, "tcplib: A Library of TCP Internetwork Traffic Characteristics", *Report CS-SYS-91-01*, Computer Science Department, University of Southern California, 1991.

[DJCME92] P. Danzig, S. Jamin, R. Cáceres, D. Mitzel, and D. Estrin, *An Empirical Workload Model for Driving Wide-area TCP/IP Network Simulations*, Internetworking: Research and Experience, 3 (1), pp. 1-26, 1992.

[EHS92] D. Ewing, R. Hall, and M. Schwartz, "A Measurement Study of Internet File Transfer Traffic", *Report CU-CS-571-92*, Department of Computer Science, University of Colorado, Boulder, Colorado, 1992.

[FJ70] E. Fuchs and P. E. Jackson, *Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models*, Communications of the ACM, 13(12), pp. 752-757, December, 1970.

[Gupta52] A. K. Gupta, *Estimation of the mean and standard deviation of a normal population from a censored sample*, Biometrika 39, pp. 266-273, 1952.

[Heimlich90] S. Heimlich, "Traffic Characterization of the NSFNET National Backbone", Proceedings of the 1990 Winter USENIX Conference, Washington, D.C.

[JS69] P. E. Jackson and C. D. Stubbs, "A study of multiaccess computer communications", Proc. of the Spring 1969 AFIPS Conference, Vol. 34.

[JLM89] V. Jacobson, C. Leres, and S. McCanne, *tcpdump*, available via anonymous ftp to `ftp.ee.lbl.gov`, June, 1989.

[Knuth81] D. Knuth, "Seminumerical Algorithms", Second Edition, Addison-Wesley, 1981.

[MM85] W. T. Marshall and S. P. Morgan, *Statistics of Mixed Data Traffic on a Local Area Network*, Computer Networks and ISDN Systems 10(3,4), pp. 185-194, 1985.

[Mogul92] J. C. Mogul, "Observing TCP Dynamics in Real Networks", Proceedings of SIGCOMM '92, Baltimore, Maryland, August 1992.

[MT77] F. Mosteller and J. W. Tukey, "Data Analysis and Regression", Addison Wesley, 1977.

[MJ93] S. McCanne and V. Jacobson, "The BSD Packet Filter: A New Architecture for User-level Packet Capture", Proceedings of the 1993 Winter USENIX Conference, San Diego, CA.

[Pawlita89] P. Pawlita, "Two Decades of Data Traffic Measurements: A Survey of Published Results, Experiences and Applicability", Teletraffic Science for New Cost-Effective Systems, Networks and Services, ITC-12, M. Bonatti (Editor), Elsevier Science Publishers B.V. (North-Holland), 1989.

[Paxson91] V. Paxson, "Measurements and Models of Wide Area TCP Conversations", *Report LBL-30840*, Lawrence Berkeley Laboratory, Berkeley, California, 1991.

[Paxson93] V. Paxson, "Growth Trends in Wide-Area TCP Connections", in submission to *IEEE Network*. Available as *WAN-TCP-growth-trends.ps.Z* via anonymous ftp to `ftp.ee.lbl.gov`.

[RFC822] D. Crocker, "Standard for the Format of ARPA Internet Text Messages", RFC 822, Network Information Center, SRI International, Menlo Park, CA, 1982.

[RFC977] B. Kantor and P. Lapsley, "Network News Transfer Protocol", RFC 977, Network Information Center, SRI International, Menlo Park, CA, 1986.

[Ross87] S. Ross, "Introduction to Probability and Statistics for Engineers and Scientists", John Wiley & Sons, 1987.

[SC92] A. Schmidt and R. Campbell, "Internet Protocol Traffic Analysis with Applications for ATM Switch Design", *Report No. UIUCDCS-R-92-1735*, Department of Computer Science, University of Illinois at Urbana-Champaign, May, 1992.

[WLC92] I. Wakeman, D. Lewis, and J. Crowcroft, "Traffic Analysis of Trans-Atlantic Traffic", Proceedings of INET'92, Kyoto, Japan, 1992.

A Details of filtering non-WAN traffic

We filtered the traffic datasets as followed:

- As mentioned in Section 2.3, we removed all connections between LBL and U.C. Berkeley.

  The LBL datasets did not include any internal or transit traffic.

- The BC dataset consisted of about 65% internal and transit traffic, which we removed, keeping only traffic between a Bellcore host and an external host.

- Similarly, 10% of the UCB dataset was transit or internal traffic, and another 2% was traffic with LBL. We removed these connections.

- The USC dataset had no transit or internal traffic.

- The DEC traffic consisted of about 3% transit traffic, virtually all (e.g., 99.8% of the DEC-1 transit traffic) of which was *smtp*. We left this traffic in the dataset. We did not identify any internal traffic in the DEC datasets.

- The NC dataset had no transit traffic. We removed internal traffic, comprising 2% of the connections.

- About 2.5% of the UK traffic was internal, which we removed. Another 3.2% of the traffic was transit (source and destination both outside of the United Kingdom); as the UK traffic represents a truly wide-area link, we felt it appropriate to include this traffic in our analysis.

B Outliers removed from datasets

As noted in Section 3.4, we removed from our analysis connections that did not transmit at least 1 byte in each direction (or, for *ftp* data transfers, 1 byte total). In addition, we removed connections that were clearly the result of protocol errors. These latter connections were identified by their high purported data volume, transmitted over time scales that required unlikely data rates:

- For *telnet*, the additional removals were an LBL-4 connection of 4GB at 600KB/s, four LBL-5 connections of 4GB at rates from 480KB/s to 6.2MB/s, and three NC connections, two of more than 100MB at rates greater than 1.2MB/s, and one of 33MB at 42KB/s. It is possible (though improbable) that this last was a legitimate connection.

  The very large value given in Table 6 for LBL-4's $\max_{\text{resp}}$ appears to be due to a legitimate connection; it lasted just over 18 hours, for an average data rate of 1.3KB/s. We removed it as an extreme outlier.

  Each DEC dataset had fewer than 100 *telnet* connections due to the DEC firewall, so we did not include the DEC traffic in our *telnet* study.

- For *nntp*, we removed four LBL-5 connections of 3.8GB or more at rates of 5.9MB/s or higher.

  Two connections were removed from DEC-3 since they purported to have transferred in excess of a 1GB at rates of 1MB/s or higher, along with two USC connections of 3GB or more at rates of 40KB/s or higher.

  We omitted the UK dataset because it contained only four connections.

- For *smtp*, we removed one DEC-3 connection as it purported to have transferred 2GB in 20 seconds.

- For *ftpdata*, we discarded three LBL-4 connections due to each involving 500MB or more at rates exceeding 400KB/s.

- For *rlogin*, we rejected one USC connection, purporting 4GB transferred in 40 seconds, as a protocol error. We removed the DEC and UK connections because in each case there were fewer than 100 connections.

C Modeling *rlogin* traffic using *telnet* models

Table 14 summarizes the *rlogin* traffic in the same manner as Table 6 does for *telnet*.

We expect to find the *rlogin* characteristics similar to those of *telnet* connections, since the two protocols are very similar in purpose, though *rlogin* will usually involve a pair of Unix hosts, which would eliminate one source of variance in the characteristics. Indeed, the LBL *rlogin* traffic shown in Table 14 is quite similar to that of the other *rlogin* datasets, suggesting that the difference in *telnet* traffic shown in Table 6 is due to LBL *telnet* connections tending to be with different types of computers than the other *telnet* datasets.[12]

We tested all of the analytic and empirical *telnet* models discussed in Section 4 on the *rlogin* datasets as well. In general we found that scaling improved fitting considerably and was required to achieve adequate fits. From Tables 6 and 14 we see that *rlogin* connections have smaller σ_{orig}'s, larger $\bar{x}_{\text{resp}}$'s, and shorter $\bar{x}_{\text{dur}}$'s, so it is not surprising that the *telnet* models must be scaled. Performance in the tails also often improved when the models were scaled. For the ratio models, however, as with *telnet*, scaling sometimes improved the *rlogin* fits and sometimes did not.

Even after scaling, though, the *rlogin* fits tend not to be as good as those for *telnet*. The arrival fits, however, were

[12] Perhaps because the scientific community still favors mainframe and VMS machines

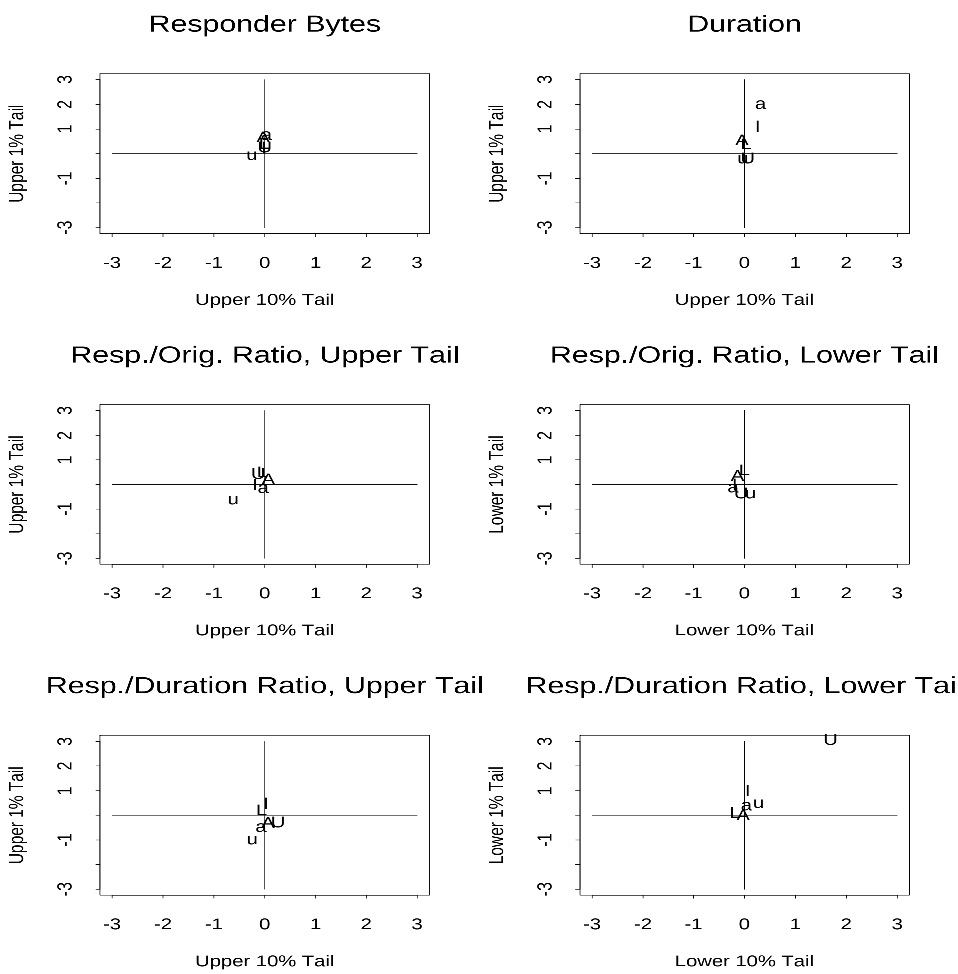

Figure 28: Tail Summaries for TELNET Models

Dataset	# Conn	# Rej	$\bar{x}_{\text{orig}}$	σ_{orig}	$\max_{\text{orig}}$	$\bar{x}_{\text{resp}}$	σ_{resp}	$\max_{\text{resp}}$	$\bar{x}_{\text{dur}}$	σ_{dur}	$\max_{\text{dur}}$
LBL-1	1,436	2	170B	×3.2	49KB	3.3KB	×6.6	0.4MB	251 s	×6.5	89.3 h
LBL-2	1,617	5	199B	×3.8	70KB	4.9KB	×7.5	2.1MB	287 s	×7.0	143.9 h
LBL-3	3,009	5	211B	×3.6	44KB	5.8KB	×6.0	1.4MB	308 s	×6.5	185.4 h
LBL-4	3,305	13	240B	×3.8	51KB	6.1KB	×6.1	17.3MB	306 s	×7.2	109.3 h
LBL-5	4,303	29	202B	×3.4	37KB	6.0KB	×6.9	2.4MB	302 s	×7.3	170.4 h
LBL-6	4,424	9	201B	×3.7	246KB	6.3KB	×7.1	5.7MB	303 s	×8.6	257.3 h
BC	307	1	177B	×3.3	7.9KB	5.8KB	×7.3	0.4MB	462 s	×6.8	36.8 h
UCB	340	2	185B	×4.3	23KB	3.0KB	×10.3	0.7MB	225 s	×6.8	6.1 h
USC	144	0	208B	×3.8	9.0KB	4.6KB	×9.6	0.4MB	308 s	×6.9	21.2 h
NC	201	0	134B	×3.2	5.8KB	3.7KB	×8.1	0.5MB	283 s	×6.1	4.3 h

Table 14: Summary of RLOGIN Connections

quite good for the LBL datasets, not requiring scaling, and acceptable for BC, UCB, and NC, with scaling required for good fits.

D Tail Summaries

Figure 28 shows tail summaries for the various *telnet* models. The text associated with Figure 5 in the main body of the text explains how to read the summaries.

Figure 29 shows similar tail summaries for the *nntp*, *smtp*, and *ftp* models.

Figure 30 shows tail summaries for the various connection arrival models. A lower-case "u" indicates the fit for the unscaled model in the lower tail; upper-case "U" the same model's fit in the upper tail; and "s" and "S" the same for the scaled model.

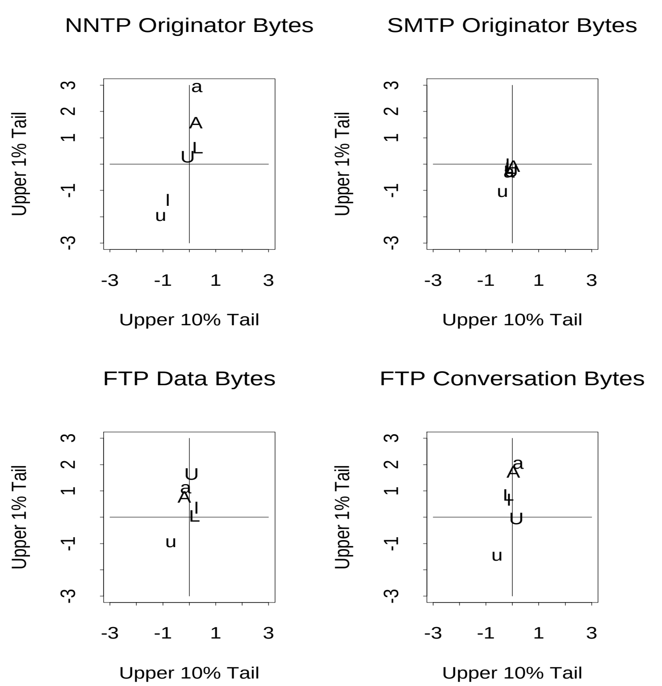

Figure 29: Tail Summaries for NNTP, SMTP, and FTP Models

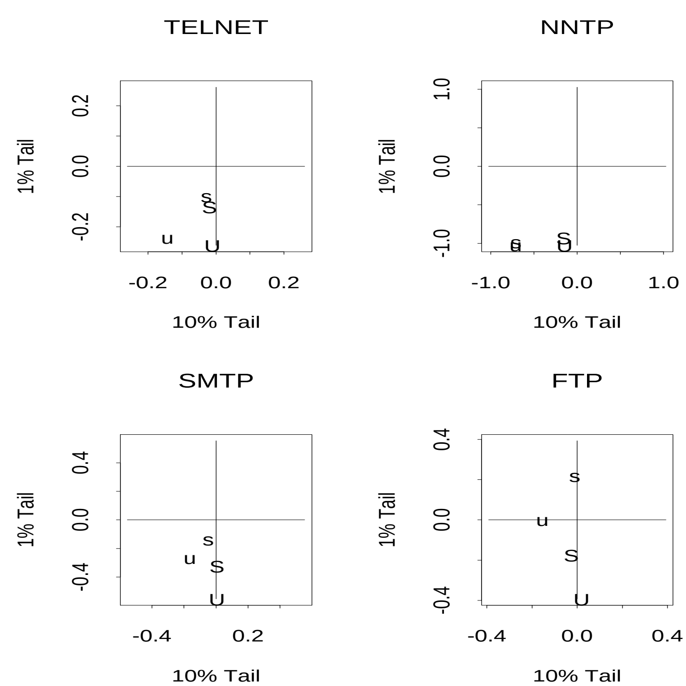

Figure 30: Tail Summaries for Arrival Models