Chapter 3

Survey of Traffic Measurements and Models

Data traffic measurement is an essential part to validate the assumptions made in analytical models of the networks. A model can not be accurate if assumptions do not reflect the reality. In order to reach some conclusions about packet process, the investigator made a survey of traffic measurement studies for LAN as well as WAN networks. This survey has led us to some important characteristics like long range dependence and hence self similarity (fractal behavior) was observed in LAN as well as WAN traffic. Thus it has put question about validity of previous simple models like Poisson models to capture correlations (Long Range Dependence) present in packet process over wide range of time scale.

It is also found that, to model the changed dynamics of packet process, models like Fractional Gaussian noise, FARIMA as explained in [10], Multiplexing ON/OFF sources with ON and OFF periods having heavy tailed distribution, M/G/∞ etc. are suggested. Further attempts are on to come up with structural model instead of black box techniques which will account for physical reasons for fractal nature of packet process.

As noted earlier, Internet traffic has shown exponential growth in terms of users, traffic
levels, topological complexity and heterogeneity. It will be unfair to apply traditional tele­
traffic theory, like the one developed for telephone networks, for its design and analysis. 
Thus measurement based empirical studies have become very important in understanding 
the dynamics of Internet behavior. This is not to say that, analytical models are irrelevant, 
rather that analytical models backed up by empirical and measurement studies will be nec­
essary in the characterisation of Internet behavior. Here we present a survey of various 
measurement studies of network traffic in LANs and WANs.

Section 3.1 discusses Local Area Network traffic measurements (mostly passive) taken 
by the various researchers and the related analysis. Similarly section 3.2 presents survey of 
Wide Area Network traffic.

3.1 Traffic Measurements on LANs

In this section we survey the traffic measurements studies on LANs. We first summarize 
the various studies and then we survey their findings for packet inter-arrival time, packet 
length distributions, utilisation behavior etc.

One of the earliest traffic measurements on LANs was by Shoch and Hupp [2] on the 
experimental Ethernet system developed at Xerox Palo Alto Research Center under normal 
load conditions. The data for this study was collected in 1979. Although the purpose of 
their study was to measure actual performance and error characteristics of Ethernet, their 
report of the utilisation, the packet length and packet inter-arrival distributions and traffic 
matrix characteristics is interesting to us in this study. The network traffic was primarily 
from applications like file transmission to printers, file transfer to one of the storage sys­
tems, multi-machine programs, access to shared database, remote diagnostic, down loading, terminal access to the time sharing machines etc. Amer, in [11] reports on the design 
of a measurement center for LAN traffic. The aim of this center was to report on the data 
packet size histograms, throughput distributions, the nature of the traffic matrix and statis­
CHAPTER 3. SURVEY OF TRAFFIC MEASUREMENTS AND MODELS

tics to study the behavior of the Medium Access Control Layer of the LAN. Unfortunately, no report on any of these statistics is made. [12] report on the measurements to characterise the delay process in the Ethernet including the delay at the MAC layer. We observe that the total delay in a LAN could be very large. Gusella in [1] reports the performance of an Ethernet LAN in an environment of disk-less work stations. In addition to reporting the statistics on packet length, inter-arrival times and utilisation, he also reports the contributions of the various protocols to these statistics. Jain and Routhier [3] also study the traffic on the ring network. Their measurements are from December 1984. They report packet inter-arrival time and packet length distributions and also develop models for packet inter-arrival times. Leland et al. in [4] collected a large amount of traffic traces from the Bellcore Morris Research and Engineering Center LAN and conducted extensive statistical tests on these traces and showed that they were self-similar. They also found that the aggregation of streams of traffic "increases" self-similarity rather than smoothing the traffic. This is in contradiction to the generally accepted argument for the use of Poisson models for tele-traffic that aggregate traffic becomes increasingly smooth with increasing number of sources. Thus these experiments proved that aggregate Ethernet LAN traffic is quite different both from the conventional telephone traffic for which Poisson and other processes derived from it have been successfully used. The conclusion from these experiments have prompted extensive research in tele-traffic models for the self-similar traffic.

We will now summarize the observations from each of the above experiments with respect to the various parameters.

Shoch and Hupp observed that under the normal load conditions there is modest use of the network, between 0.60% to 0.84%. However the maximum utilisation in the busiest interval was found to be about 3.6% over the busiest hour, 17% over the busiest minute and 37% over the busiest second. Gusella in [1] noted that the network utilisation was low at about 6.5% averaged over 24 hours, but the burst generated short term peak utilisation above the one third of Ethernet raw bandwidth. Further he showed that character traffic (
like that from telnet applications) generated many small packets but no substantial network utilisation. File access to the remote file servers generated burst of traffic lasting several seconds which might demand bandwidths of the order of 120 KBytes/sec or about 10% of the Ethernet raw bandwidths over intervals of a few seconds. Observe that the network utilisation is a function of length of the interval over which it is averaged and it is highly dependent on it. Network delay is a function of the network utilisation and the perceivable delay is of the order of a second. Thus we believe that higher network utilisation over intervals of the order of a second is an indication of the congestion in the network and hence the quality of service is provided by the network.

Shoch and Hupp in [2] have reported that the packets on their experimental Ethernet have a bimodal distribution which is similar to that reported for the ARPANET. Gusella in [1] (ref. Figure 3.1) also reports the similar results from his measurements. He found

![Figure 3.1](image)  
Figure 3.1: Histogram of the packet lengths from [1]

that the mean packet size was 578 bytes and the median was 919 bytes. Small packets of less than 50 bytes together transport 30.4% of total packets, but only 2.9% of total bytes. Packets larger than 576 bytes comprise 50.2% of total packet count and contributed to 93.1% of byte traffic. From the above measurements it is clear that packet traffic on LAN has always been bimodal - a peak at small packet lengths corresponding to character and
acknowledgment traffic and another peak at around 600 bytes corresponding to packets from data transfer applications involving larger volumes of data like FTP.

We now look at the inter-arrival time distributions reported in the above measurements. Shoch and Hupp in [2] found that mean inter-arrival time is 39.5ms. Figure 3.2 shows histogram of the packet inter-arrival time reported by them and Figure 3.3 shows the cumulative distribution. It can be seen from this figure that 50% of packets are followed by next packet within 10ms, 90% of packets are followed by next within 64ms, and 99% within 183ms.

Jain and Routhier in their study [3], were the first to study correlations in the inter-arrival times and to suggest models for the same. Treating the packet inter-arrival times as a time series, they observed the autocorrelation between them and found that ACF at lags 1, 2 and 3 were respectively 0.015, 0.046 and 0.043. The mean inter-arrival time was seen to be 65.8ms with the standard deviation of 2835.3 and coefficient of variance as 43.1. They concluded that with the ACF being small, arrivals were independent. However, with

![Figure 3.2: Histogram of inter-arrival time for Ethernet from [2]](image-url)
the coefficient of variation was very high, the inter-arrival times were not exponentially distributed, and the packet arrivals did not form a Poisson process. However, the marginal inter-arrival distribution may be approximated as a “piece-wise exponential” distribution. This has been shown in Figure 3.4. The above studies expose the considerable difference in the assumptions made on the packet processes in the analysis of LAN protocols like, for example, those by Tobagi and Hunt in [13], Almes and Lazowaska in [14], Coyle and Liu in [15] and Bertsekas and Gallager in [16] and their characteristics in actual networks. For example, it has been consistently shown that the packet lengths are multi-modal and that the inter-arrival times of the packets in the network is not exponential. More importantly, Jain and Routhier observed that the inter-arrival times were not independent, which was usual assumption made in most of the analysis of LAN protocols. Interestingly, we note from Boggs et al. in [17] that realistic throughput models do match with that of measured ones. For example the empirical result reported by Boggs et al showed throughput around 0.97. In their experimental set up, packet lengths were 512 Bytes and the maximum distance in the network was 910 meters corresponding to $a = 0.01$. For analytical model given by Takagi in [18] the corresponding throughput is around 0.95.
Now we will look at delay model. we note from Boggs et al. [17] that average transmission delay increases linearly with number of nodes. Thus empirical results suggest that the bandwidth penalty is low with increasing number of nodes. Analytical models with reasonably realistic assumptions on topology like fixed end-to-end propagation delay give results like in Figure 3.5 (from [18]). It shows linear increase in delay up-to throughput of 0.85. After that there is drastic increase in delay, indicating heavy penalty in bandwidth. Let us compare this delay with that of delay observed in M/M/1 system. As given in [19], for M/M/1 system we have following equation 3.1:

$$T = \frac{N}{\lambda} = \frac{\rho}{(\lambda(1-\rho))} = \frac{1}{(\mu(1-\rho)))} = \frac{1}{(\mu)(1 + N)}$$

(3.1)

Where,

$N$ : Average number of packets in the system

$T$ : Average packet time in the system
CHAPTER 3. SURVEY OF TRAFFIC MEASUREMENTS AND MODELS

Figure 3.5: Throughput plot from [4]

\[ \lambda : \text{Average packet arrival rate} \]
\[ \mu : \text{Average service time} \]
\[ \phi = \frac{\lambda}{\mu} : \text{Utilisation factor.} \]

From equation 3.1 we note that for small values of \( N \) average time \( T \), for which a message is in system, is linear, but for higher values it increases rapidly. We note that difference is the penalty due to collisions occurring at higher loads.
CHAPTER 3. SURVEY OF TRAFFIC MEASUREMENTS AND MODELS

Thus from above discussion it can be observed that empirical results do not match with simplistic analytical models but do match with realistic analytical models.

However much of these analyses were done when network traffic was dominated by transaction type applications like telnet and occasional burst applications like an NFS file access or an ftp file transfer. With increasing networked applications, input process assumptions of independent Poisson packet arrivals, and exponential or fixed length packets were bound to fail. And also the corresponding analytical models based on these assumptions. Recall that much of the empirical studies discussed so far, (Shoch [2], Jain and Routhier [3], ) are essentially aimed at characterising the marginal distributions of the inter-arrival times. Except for Jain and Routhier, none of these studies even investigate the presence of correlations in either the inter-arrival times or the packet lengths. Leland et al. [4] were the first to make a detailed investigation of the correlations in the packet process on a LAN. The packet process considered for measurements can be described as follows:

The time axis is divided into intervals of length T to find the number of packets in the ith interval. In this way we get packet count for each interval. Thus we get sequence of packet count versus time, which is a time series. They characterise this time series to get its statistics.

While statistically analysing the collected LAN traffic data they observed that, the autocorrelation plot has a long tail giving a non summable autocorrelation function, thus implying long range dependence (LRD) in the traffic departure process from the Ethernet LAN. A plot is shown in Figure 3.6, which is sample variance time plot (one of the graphical method to find self-similarity) as given by Leland et al. in [4]. In this figure, the slope of fitted straight line gives $H$, Hurst parameter value to be 0.80, confirming self-similarity in the traffic data. Thus even though the studies by Boggs et al. [17] and Leland et al. [4] were done independently, we note that the reasoning for Boggs's observations are given in Leland's study. It is important to note that the observations made by Leland et al. in [4] is the output process on the Ethernet and not the traffic arrival process to the network. The
input process to the network is not known and would be hard to characterize empirically. Thus the implications of the output process from the Ethernet being self-similar, will be on the inter networking devices like bridges, switches, and routers.

### 3.2 Wide Area Network Traffic Measurements

In this section we present the survey of measurement studies for Wide Area Network (WAN) traffic. After going through different studies on LAN traffic measurements, it would be important to survey WAN measurement studies from network modeling point of view. In the following part we mention few WAN traffic measurement studies followed by their statistics.

Caceres’s study [20] is one of the earliest efforts to measure the WAN traffic param-
CHAPTER 3. SURVEY OF TRAFFIC MEASUREMENTS AND MODELS

eters. Data for his measurements was collected at AT & T Bell laboratories, Murray Hill, New Jersey in July 1989. His study gives protocol-wise statistics corresponding to packet count, byte count and packet length frequencies. In recent past Thompson *et al.* in [21] studied WAN traffic measurements in terms of packet size, flow duration, volume and composition of traffic by MAC layer protocol as well as by application. For this purpose they have used MCI’s OCMON monitor at all vBNS supercomputer sites and on the Internet MCI backbone.

Now we will present measurement statistics for above two studies. From Caceres’s study [20] we note that protocol level traffic is dominated by TCP. Almost 80% of the WAN traffic is constituted by TCP packets against 20% traffic which is using UDP packets. Application level statistics, in TCP- protocol, marks SMTP as dominant application in terms of packet count. It is followed by FTP, TELNET etc. In terms of byte, FTP slightly overtakes NNTP followed by TELNET etc. About packet length distribution we note that TCP as well as UDP traffic is multi-modal. From the given data it is clear that TCP has peaks at 1 to 10 bytes, 512 bytes, 576 bytes length. UDP packets has peaks in the range from 1 to 80 byte lengths. It should be noted that above statistics is corresponding to the traffic traces which were collected before 1989.

Here, we present a bit recent measurements given by Thompson in [21] and comment appropriately. Thompson *et al.* in [21] analysed 240,000 traffic flows over 24 hour period. These flows are dominated by small packets with peaks of packets occurring at 44, 552, 576 and 1500 sizes. Thus we note that multiplexed traffic is multi-modal which is in agreement with Caceres’s study in [20]. Small size packets can be confirmed from cumulative distribution plot for the packets. It shows that almost 75% packets are smaller than the TCP MSS of 552 bytes. Protocol level statistics is dominated by TCP with an average of 95% of bytes, 90% of the packets, 75% to 85% of the flows on the link, the average TCP packet size is 300 bytes, average number of packets per flow is 16 to 20 and average flow duration is 12 to 19 seconds. UDP traffic is second highest level at 5% of bytes, 10% of packets, 20% of
flows, 5 to 15 packets per flow, 1 to 2 KBytes per flow, and 10 to 18 seconds in duration on
an average over 24 hour. Among the others ICMP comprise 2% of overall packets, 0.5%
of overall bytes and 1.5% of total flows. The other IP protocols measured are IPV6, IP in
IP etc. but they make negligible contribution to the traffic composition. Application level
statistics is dominated by web client/server traffic followed by FTP data and NNTP traffic
in TCP protocol. Thus we note that at application level web client/server traffic is likely to
be major contributor to future WAN traffic. Besides this, we do not observe any substantial
change from protocol-wise contribution point of view in WAN traffic from 1989 to 1997.
TCP continues to be dominant protocol to contribute WAN traffic. These studies suggest
that for better model of traffic we should carefully model the dominant protocols like TCP
and applications like FTP.

Above studies presents significant WAN traffic dynamics but do not talk about modeling
aspects of their findings as well as do not give implications of their findings on network
performance. Also we note that above studies do not talk about traffic parameters like
inter-arrival times, correlations in packet process which has definite effect on performance
parameters like packet delay, loss ratio, congestion etc. Much of discussion about these
parameters is given in following studies.

Paxson in [22] suggested few analytic models for various random variables like origi-
nator bytes, responder bytes, data burst bytes etc. which describe TELNET, NNTP, SMTP
and FTP connections associated with TCP traffic. Thus he attempted to put TCP traffic
into model form. To validate these analytic models he used Tcplib model for various
random variables. (Tcplib is an empirical model of wide area traffic. With the help of
Tcplib, distribution of random variables associated with different application protocols can
be modeled.)

It is observed that except originator bytes in TELNET and data burst bytes in FTP all
other random variables are very well modeled using log 2 normal model. It is found that
upper 1% tail and lower 20% tail in TELNET originator and responder bytes respectively
do not match with that of the Tcplib model. NNTP originator bytes distribution shows considerable variation over the measured interval. This variation is due to propagating news as soon as it comes or defers its transmission to take advantage of minimal loads. It is observed that SMTP originator bytes distribution is bimodal. Similarly FTP data burst shows large tail, which can be modeled using Pareto distribution. Thus it can be observed that heavy tail in FTP data burst contributes to long range dependence which intuitively suggests that FTP may contribute to self-similarity in wide area network traffic. This is an important observation to make. Also TELNET traffic model (analytic) do not obey upper 1% tail given by Tcplib. Thus FTP and TELNET, two dominant applications of TCP protocol traffic need to be carefully modeled.

Interestingly this point has been rigorously dealt by Paxson and Floyd in [23]. In addition to this, they also studied the large scale correlations in collected traces. For their measurements they used two data sets collected at Lawrence Berkeley Laboratory’s Internet gateway and at primary access point for Digital Equipment Corporation. As we noted in Leland’s et al. study [4], that LAN traffic can be better modeled as self-similar process instead of Poisson process, which indicate that packet process inter-arrival time is not exponentially distributed. Similarly, adequacy of Possion process for packet process in wide area network traffic is verified by Paxson in [23]. It is found that Poisson process assumptions are only valid for TELNET connection arrivals and FTP control connection arrivals. Other WAN traffic arrivals like FTP data, NNTP, SMTP, packet process within TELNET connection are found to deviate from Poisson process. Further it is observed that these connections possess burstiness over wide range of time scales. Failure of Poisson process to capture burstiness over 1hr as well as 10 minutes duration has been exposed. Moreover burstiness in packet process is confirmed using time variance plot. Also it is observed that FTP data connection spacing within single session shows heavy upper tail. Thus Poisson model and hence exponential inter arrival time distribution gives considerable deviation from actual traffic.
CHAPTER 3. SURVEY OF TRAFFIC MEASUREMENTS AND MODELS

In this study we also find that WAN traffic exhibits large scale correlations in packet process of different protocols. We call such a traffic as long range dependent (LRD) (explained in [24]) traffic with non summable autocorrelation function i.e. \( \Sigma r(k) = \infty \). We note that long range dependence is contributed by TCP connection sizes and transmission rates and to certain extent is independent of network dynamics. This implies that multiplexed data networks will also exhibit long range dependence. Further we find in the literature that self-similar process is simplest model exhibiting long range dependence behavior, rather bursty traffic over wide range of time scale.

Long range dependence in NSFNET wide area traffic has been studied in detail by Klivansky in [25]. For this purpose Klivansky et al. in [25] studied NSFNET WAN traces collected at different NSFNET core switches in US. For carrying analysis, the collected traffic data was grouped as Mode A (FTP Ctrl,NNTP, and SMTP), Mode B (FTP DATA, TELNET, and RLOGIN), and Mode D (all packets). It is found that Hurst parameter \( \hat{H} \), a measure of self-similarity for mode A is in range .78 to 0.87, for mode B it is 0.79 to 0.89 and for mode D it is 0.69 to 0.84. From these figures we note that mode A and B (TCP traffic) have larger variations of \( \hat{H} \) as compared to mode D(all packets). This implies that TCP traffic contribution to self-similarity is more as compared to other protocols like UDP.

This is in agreement with observations made by Leland et al. in [4]. Further Sahinoglu et al. in [26] had observed that the degree of self-similarity varies linearly with the traffic load. In Sahinoglu’s study we note that decrease in buffer capacity results in increased packet loss for fixed degree of self-similarity. Also the studies like Feldmann in [27] have exploited the cascade of computer networks to show the multiplicatively generated multifractal nature of WAN traffic. Thus we conclude WAN section with the remark that TCP traffic in WAN also exhibits self-similarity as has been seen for LAN traffic. As observed by Thompson and Caceres, TCP is major contributor to overall WAN traffic, and hence overall WAN traffic will also exhibit self-similarity. This implies that WAN as well as LAN traffic inherently exhibits the self-similar nature.
CHAPTER 3. SURVEY OF TRAFFIC MEASUREMENTS AND MODELS

3.3 Conclusion

we note that studies like [2], [3], [4], [22], [23] etc. have consistently shown that traffic sources do exhibit burstiness and that there has been significant correlations in the observed traffic. As a consequence of this we receive heavy traffic over longer period which may result in degrading performance of network. As an example it may result in buffer overflow. It has also been observed in [28] that burstiness over different time scales results into long range dependence (LRD) leading to self-similar traffic (same statistics over different time scales). In number of other studies too it is observed that traditional traffic models like Poisson can not capture burstiness in traffic. Studies like [29] show that Markov and Markov modulated traffic models are able to capture burstiness in network traffic. In addition Markov modulated models, long tail Pareto models, time series models also captures autocorrelations in network traffic.