# 6

# Traffic Modeling for Wireless IP

Traffic modeling becomes increasingly important in future quality of service (QoS) wireless IP (Internet protocol) networks. It is indeed vitally important for any communication network to perform efficiently and to utilize the network resources more appropriately. The topic has been researched for many years in voice-based telephony networks but after the invention of the packet-switched networks and increasing the data applications over the Internet, it was not followed up accordingly. In continuation of our discussion in the previous chapter on QoS, we need to either design a perfect network by employing appropriate data traffic models or we need to rely on traffic management techniques, in order to provide QoS in data networks.

In this chapter, we look at the first approach, that is, to find appropriate traffic models for the future data networks and the wireless IP. In the next chapter, we will look at the traffic management techniques for wireless IP networks. We will describe several different characteristics of traffic in data networks and formulize the major models available. The discussion provided in this chapter will be sufficient for any researcher who is looking for a fundamental understanding and knowledge to start working on the important topic of traffic modeling in wired and wireless data networks.

## 6.1 INTRODUCTION

Telecommunication networks are evolving and they include more nonvoice traffic generated from Internet applications and other digital data sources. In a voice-centric network like most of the current telephony networks, the traditional traffic models, listed in

*The Wireless Mobile Internet: Architectures, Protocols, and Services.* Abbas Jamalipour
 ISBN: 0-470-84468-X

**Table 6.1** Traditional traffic models for voice-centric networks

| Probability distribution | Usage |
|---|---|
| Poisson | Packet and connection arrival |
| Exponential | Packet interarrival |

Table 6.1, could be sufficient to evaluate the network performance, such as queuing performance and congestion control as well as the designing process.

In addition, the well-known Erlang formulas have provided universal solutions to network problems for both wireline and wireless circuit-switched networks. Reference [1] is a classic reference for traffic modeling in cellular networks.

### 6.1.1 Emerging trend of the next-generation mobile traffic

Today, Global System for Mobile communications (GSM) is the most widely used second-generation digital cellular system. Although the current GSM is optimized for voice communications, the next-generation cellular mobile, that is, the third-generation mobile, will accommodate voice, data, and multimedia technology with a vast range of applications as illustrated in Figure 6.1. The main focus of the next-generation mobile will be anywhere-anytime communications for both voice and other types of data transmission.

Internet and multimedia traffic can be characterized by frequent transitions between active and inactive states, often called *ON/OFF patterns*. The ON period represents the file-downloading time and the OFF period is the user-reading time. If the present circuit-switched technique is used, the bandwidth of the dedicated circuit is wasted during the OFF period. However, the packet-switched technology allows higher data transmission rates and uses the bandwidth only within the ON period [2]. For the emerging future

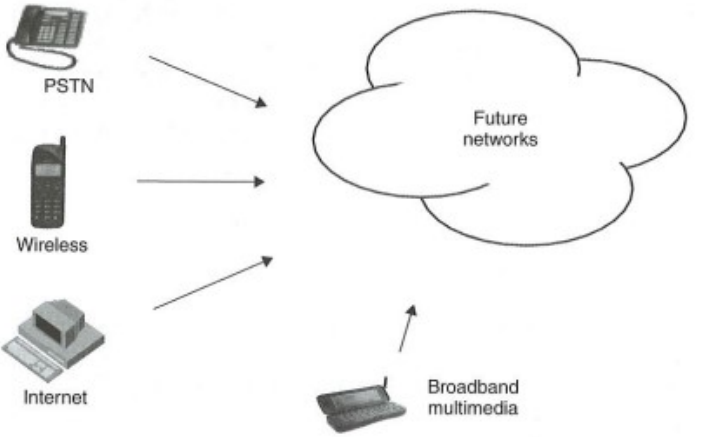



**Figure 6.1** Emerging trends in the third-generation mobile traffic

network traffic, the current circuit-switched technique and the Erlang formulas are no longer appropriate to use [3,4].

### 6.1.2 Importance of traffic modeling

Although traffic modeling can be a time-consuming and resource-intensive process, it is the basic tool for performance evaluation and resource provisioning [4–10]. Figure 6.2 shows how traffic models are used as the input for analytical and simulation studies of telecommunication networks [4].

Traffic models have a lot of important roles in planning and managing new and existing networks. Let us name a few of their important roles.

- They support efficient network-dimensioning procedures and traffic management functions.
- They assist in characterizing and modeling traffic behavior that is used for accessing QoS.
- They help estimate the resource utilization in a network environment.

### 6.1.3 Traffic modeling criteria

A good traffic model should be able to characterize the network dynamics with an acceptable level of accuracy. By doing so, it has to be [11]

1. *General* enough to provide a good approximation to the field data. It means that the proposed model should rely on a few parameters that can readily and reliably be estimated from measured observation.
2. *Simple* enough to obtain analytically tractable results for performance evaluation. It means that the proposed model should be simple in terms of
   - mathematical analysis,
   - programming,
   - computing (i.e. fast simulation and numerical analysis).

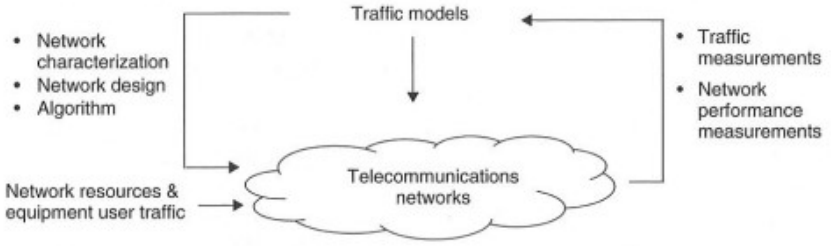



**Figure 6.2** The role of traffic modeling in telecommunication networks

**Table 6.2** Three important traffic characteristics in traffic modeling

| Traffic characteristics | Description |
|---|---|
| Queuing performance | Buffer size and parameters |
| Marginal distribution | Statistical multiplexing and source traffic control |
| Autocorrelation | Prediction of queuing behavior |

From an analytical point of view, a good traffic model should be able to capture three of the most important traffic characteristics of the measured data, that is, queuing performance, marginal distribution, and autocorrelation. These are described in Table 6.2.

The suitability of a traffic model is primarily determined by its ability to predict the queuing performance. More refined models predict a better marginal distribution and autocorrelation of the modeled traffic but usually at the cost of increase in model complexity.

## 6.2 POISSON AND MARKOV MODELS

Because of their theoretical simplicity, the Poisson and Markov Modulated Poisson Process (MMPP) are used extensively for packet-switched data networks. Although the self-similar nature of today's data traffic was noticed for some time, many practitioners ignored this phenomenon because of

1. inadequate physical explanation for the observed self-similar nature of measured traffic from today's packet networks,
2. lack of studies on its impact on the network, and protocol design and performance analysis.

Since the traditional traffic models are inadequate to capture today's network characteristics, the packet-switched data traffic models have been developed on the basis of measurements from actual data networks. However, their availability in wireless network modeling is still to be proven [12].

### 6.2.1 Limitation of the Poisson and Markov traffic models

When the traditional traffic modeling such as Poisson or MMPP is used in the framework of the ON/OFF pattern, the ON or OFF periods display either exponential or geometric distribution, that is, finite variance distribution. The traffic displays memory-less property, meaning that its correlation is of short-range-dependence (SRD). The aggregate traffic behaves like white noise and fails to capture any of the three most important traffic characteristics described in Table 6.2.

Recent traffic analyses on network traffic such as local area network (LAN) and wide area network (WAN) and application traffic such as World Wide Web (WWW) and variable bit rate (VBR) video traffic have revealed the prevalence of a long-range dependence

(LRD) on packet-switched networks [5–7,9,11,13–18]. This means that there is a high correlation of traffic over many timescales. Although the significance of traffic correlation on queuing performance was recognized, most of the studies were concentrated on SRD [5]. Table 6.3 shows some of the fundamental differences between traditional voice traffic and today's and emerging high-speed data traffic.

Apart from the differences described in Table 6.3, the next-generation mobile networks will offer many different applications and each application will have different QoS requirements. For example, in circuit-switched wireless networks, the network performance will be measured in terms of (1) continuous coverage and (2) high reliability of handovers. The failure of one of these two conditions results in dropped calls or inadequate QoS. However, in packet-switched networks, the nature of service is discontinuous and there is no strict restriction on delay requirements. Instead, packet error rates and loss rates are more important parameters to consider. Therefore, the network performance criteria have to be changed as well. So far, most of the work on self-similar traffic is concentrated on its impact on queuing performance [5,15,19–22]. However, its impact on admission and congestion control is rather neglected. Therefore, a close examination of these impacts on end-to-end QoS requirements of voice, data, and multimedia applications will be the focus of the next chapter on traffic management.

### 6.2.2 The need for new traffic models

Emerging high-speed network traffic displays new characteristics. Traditional traffic models fail to capture these characteristics and lead to an overly optimistic estimation of performance. The unexpected poor performance of asynchronous transfer mode (ATM) switches in the field may indicate that traditional traffic models are inappropriate for use in data-centric networks. With today's phenomenal increase in data traffic, it is essential

**Table 6.3** Comparison between traditional and emerging network traffic

| | Traditional traffic | Emerging network traffic |
|---|---|---|
| ON/OFF traffic distribution | *Exponential* or *geometric* distribution (i.e. finite variance distribution) | Heavy-tailed *distribution* (i.e. infinite variance distribution) |
| Burstiness | Multiplexing traffic streams tend to produce *'smoothed out'* aggregate traffic with reduced burstiness | Aggregate self-similar traffic streams can actually *intensify* burstiness |
| Aggregate traffic | Gaussian | LRD |
| Queuing performance | Queue length decreases *exponentially* with increase in buffer size | Buffer gain is *linear* so that queue length decreases linearly |
| Admission control | Extensive studies are done | Subject of future studies |
| Congestion control | Extensive studies are done | Subject of future studies |

to understand the characteristics of data traffic in order to utilize the network resources and to optimize the network performance.

In theory, the traffic should become more and more like the Gaussian processes in the future [5,23]. The prevalent effect of a single application will be less significant in terms of aggregate traffic. However, at the moment the network traffic is not anywhere close to the Gaussian model. Over the past 20 years, numerous attempts were made to find an Erlang-like formula for the traditional telephony for broadband (i.e. multimedia) traffic [23]. However, so far, there is no such model that fits the role. Presently, the level of aggregation is not sufficient enough for the bad behavior of one traffic stream to dominate the overall network traffic characteristics. The need for good traffic models are more acute than ever.

## 6.3 CHARACTERISTICS OF THE EMERGING TRAFFICS

It is reasonable to assume that (1) session arrival is Poisson with an arrival rate of $\lambda$ sessions per second and (2) the duration of each session is exponentially distributed. However, the packet-arrival patterns within the session depend on the application. Recent analysis on traffic measurements on packet-data networks such as LAN and WAN, show heavy-tailed, self-similar, fractal, and LRD characteristics. In this section, we have defined the terms that are frequently used to describe today's traffic and the emerging networks traffic.

### 6.3.1 Heavy-tailed

A distribution is heavy-tailed if the asymptotic shape of the distribution follows a power-law so that

$$P[X > x] \cong x^{-\alpha} \quad \text{as } x \longrightarrow \infty, 0 < \alpha < 2 \tag{6.1}$$

The parameter $\alpha$ describes the heaviness of the tail distribution so that as $\alpha$ gets smaller the distribution becomes more heavy-tailed. Figure 6.3 shows the effect of $\alpha$ in the heavy-tailed distribution. The asymptotic (i.e. tail) shape of the distribution is hyperbolic and converges slower than the exponential distribution. It appears to have a thicker tail distribution, and is therefore called a *fat-tailed* or *heavy-tailed distribution*. The heavy-tailed nature of the distribution comes from the fact that the larger portion of the probability mass may be present in the tail of the distribution. It differs from exponential, geometric, and Poisson distributions so that

- If $\alpha \leq 2$, the distribution has an infinite variance,
- If $\alpha \leq 1$, the distribution has an infinite mean.

Most commonly used examples of the heavy-tailed distributions are the Pareto and Weibull distribution.

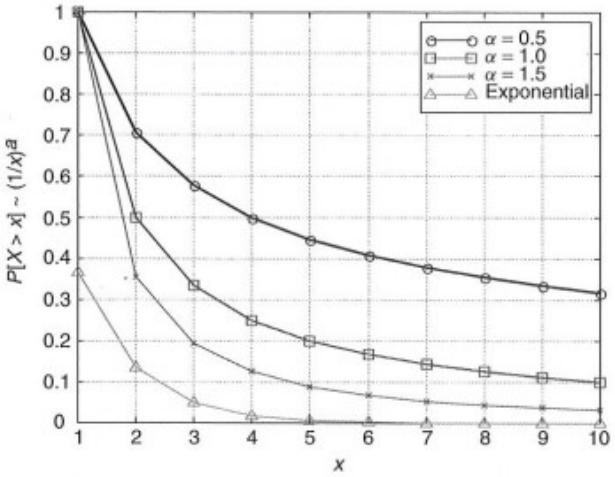



**Figure 6.3** The effect of $\alpha$ in a heavy-tailed distribution

#### 6.3.1.1 Pareto distribution

It is the simplest heavy-tailed distribution. Its distribution is hyperbolic over its entire range. Mathematically, the cumulative distribution function of the Pareto distribution, $F_p(x)$, is

$$F_p(x) = 1 - \left(\frac{k}{x}\right)^{\alpha} \tag{6.2}$$

where $k$ is the minimum value of $x$ and $\alpha$ is the heaviness of the tail distribution. Figure 6.4 shows the effect of $k$ in the Pareto distribution. $k$ is simply the scaling factor and does not affect the tail distribution. The effect of $\alpha$ is shown in Figure 6.3.

#### 6.3.1.2 Weibull distribution

The cumulative distribution function of the Weibull distribution, $F_w(x)$, is

$$F_w(x) = 1 - e^{-(x/a)^b} \tag{6.3}$$

Both parameters $a$ and $b$ affect the tail distribution. However, the heavy-tailed nature of the Weibull distribution is more sensitive to the value of $b$. Figure 6.5 shows the effect of $a$ and $b$ in the Weibull distribution.

Usually, a heavy-tailed distribution describes traffic processes such as packet inter-arrival times and burst length. If traffic is heavy-tailed, it is highly correlated. It means that the arrival rate is higher than the service rate. In the context of traffic modeling, it is often used to describe the burst individual source traffic distributions.

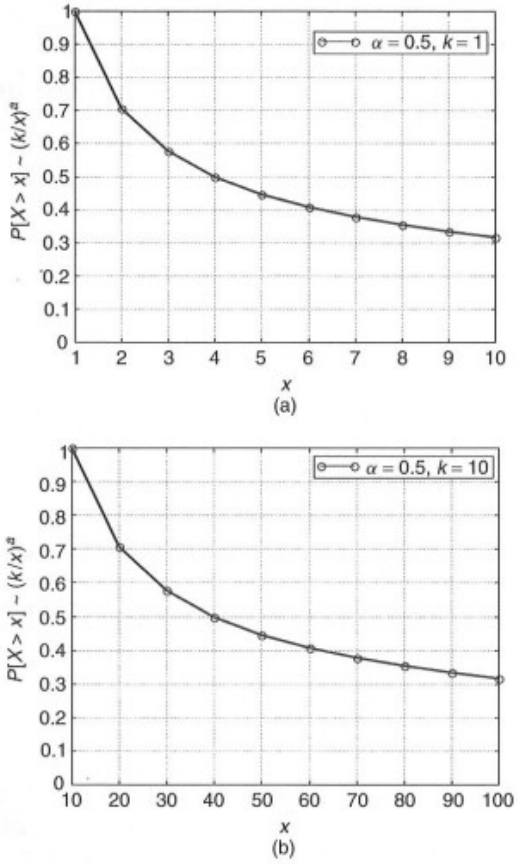



**Figure 6.4** The effect of $k$ in the Pareto distribution with (a) $k = 1$; and (b) $k = 10$

## 6.3.2 Self-similar

It is a scaling behavior of the finite dimensional distributions of a continuous- or discrete-time process. Traffic is *self-similar* if the aggregate traffic

- exhibits time correlation over a wide range of timescales, and
- can be characterized by a single parameter called *Hurst parameter* ($H$).

### 6.3.2.1 Self-similarity indicator

The Hurst parameter, $H$, is the measure of the degree of self-similarity of the aggregate traffic stream. As $H \rightarrow 1$, the degree of self-similarity increases. The Hurst parameter

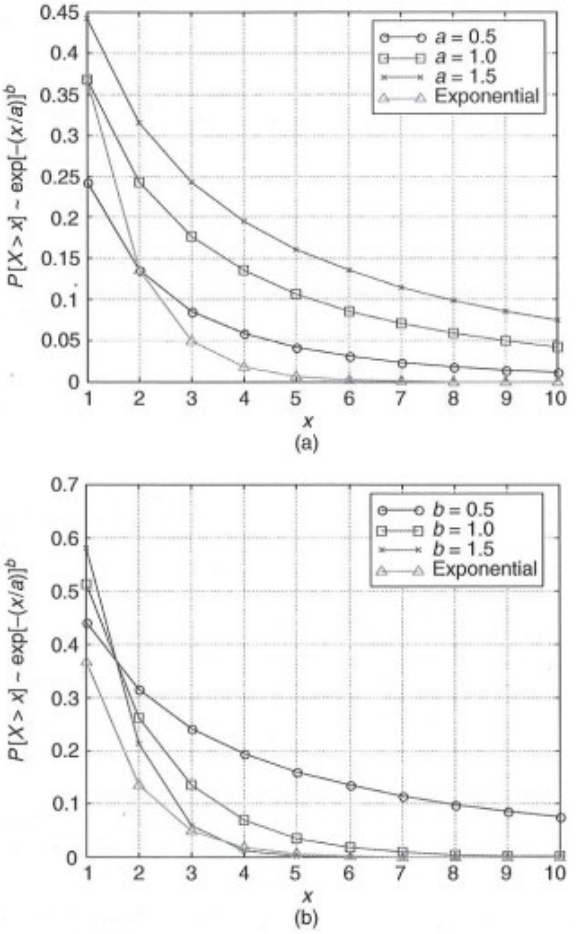



**Figure 6.5** The effect of (a) $a$; and (b) $b$ in Weibull distribution

can be measured in various ways. However, three of the most common methods are as follows:

1. *Variance versus Time*: If traffic is self-similar, then its slope $-\beta < -1$. For some historical reason, the relationship between the slope, $-\beta$, and $H$ is

$$H = 1 - \frac{\beta}{2} \tag{6.4}$$

Therefore, $H$ can be calculated by obtaining the slope of variance versus time graph.

2. *R/S plot*: Let $\{Y_k\}_{k=1}^{n}$ be an empirical time series with sample mean $\overline{Y}(n)$ and sample variance $S^2(n)$. The rescaled adjusted range, $R/S$ statistic, is given by $R(n)/S(n)$ with

$$R(n) = \max\left\{\sum_{i=1}^{k}(Y_i - \overline{Y}(n)) : 1 \le k \le n\right\} - \min\left\{\sum_{i=1}^{k}(Y_i - \overline{Y}(n)) : 1 \le k \le n\right\} \tag{6.5}$$

$$E\left[\frac{R(n)}{S(n)}\right] \cong n^H, \qquad \text{for large } n \tag{6.6}$$

Normally, the $H$ value of a self-similar process is $0.5 < H \le 1$ whereas that of the SRD process is $H \approx 0.5$.

3. *Whittle Estimator*: It provides the confidence interval but it requires some form of underlying stochastic process, which is a drawback. The most commonly used forms are
   - fractional Gaussian noise (FGN) with $0.5 < H < 1$,
   - fractional ARIMA $(p, d, q)$ with $0 < d < 1/2$ (to be discussed shortly).

#### 6.3.2.2 Description of self-similarity

- *Exactly self-similar* ($H = 1$): A distribution appears indistinguishable from one another but distinctively different from pure noise.
- *Asymptotically self-similar* ($0.5 < H < 1.0$): A distribution converges to a time series with nondegenerate autocorrelation structure.
- *Second-order self-similar*: For stationary sequences, whose aggregate processes possess the same nondegenerate autocorrelation functions as the original process.

This characteristic is often explained in terms of the high variability of individual connections that contributes to the aggregated traffic. Self-similarity is often used to describe individual application traffic.

### 6.3.3 Fractal

A fractal process is characterized by significant long bursts. These bursts are caused by downloading large files such as video files, long periods of high levels of VBR video, or intensive bursts of database activities. It is another term to describe the self-similarity of traffic. Current WAN traffic is often described as multifractal. Multifractal traffic can be considered as an extension of self-similar traffic, by considering properties higher than second-order characteristics so that it can capture more irregularities in the distribution.

### 6.3.4 Long-range-dependence

A process with LRD has an autocorrelation function, $r(k)$, of:

$$r(k) \approx k^{-\beta} \quad \text{as } k \longrightarrow \infty \quad \text{where } 0 < \beta < 1, \text{ and } \Sigma\, r(k) \longrightarrow \infty \tag{6.7}$$

In other words, the autocorrelation function (1) decays hyperbolically and (2) is non-summable. For the conventional short-range dependence (SRD) process, an autocorrelation function decays exponentially. It is often used to describe the tail-end behavior of the auto-correlation function of a stationary time series. In traffic modeling, LRD is often used to describe the aggregate traffic such as WAN, whereas self-similarity is usually used in the context of LAN or individual application traffic.

Table 6.4 summarizes some typical traffic types and associated traffic distributions and models.

### 6.3.5 Suitability of self-similar and long-range dependence

After studying the terms that describe current and future network traffic characteristics, the next appropriate question would be "why does the traffic display these characteristics?' In References [9, 14, 24], it is pointed out that the heavy-tailed nature of ON and OFF periods has more to do with basic properties of information storage and processing. It is not a result of the network protocols or user preference. Therefore, changes in protocol processing and document display cannot remove the self-similarity of the web traffic. Also, it is shown that both the user's thinking or reading times and the file-size distributions are strongly heavy-tailed. In addition, Internet provides explicit support for multimedia formats; the file distribution is strongly heavy-tailed. Figure 6.6 shows the effect of multimedia files such as image and audio files on the file distribution. The values of $\alpha$ are taken from Reference [14].

Often, self-similarity in today's network traffic is explained in terms of application traffic. The burst data traffic and VBR real-time applications such as compressed video

**Table 6.4** Traffic distributions and frequently used traffic models

| Traffic types | Traffic distribution | Frequently used traffic models |
|---|---|---|
| Individual source traffic | Heavy-tailed ON/OFF distribution | • Pareto<br>• Weibull |
| Individual application traffic or LAN | Self-similar | • FGN<br>• FARIMA |
| Aggregate traffic | LRD Multifractal | • Fractional Brownian motion (FBm) model<br>• M/G/∞<br>• M/Pareto |

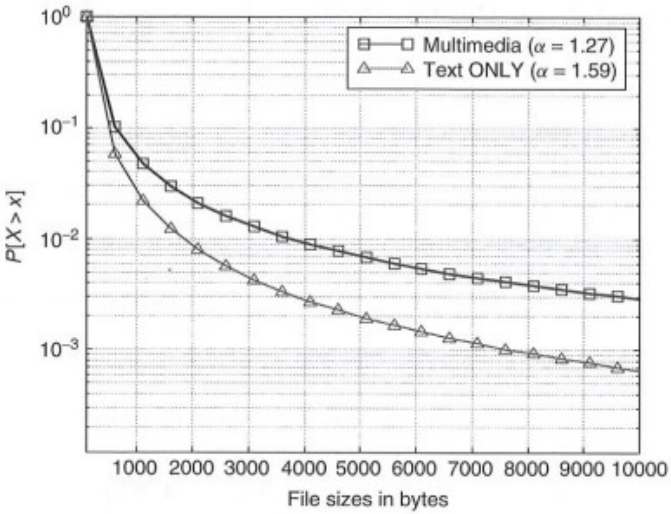



**Figure 6.6** The effect of multimedia files on file distributions

and audio display (1) a certain degree of correlation between arrivals and (2) slow LRD in time. As a result, the aggregate traffic is self-similar. Or, it could be the high variety of individual connections (i.e. infinite variance) that contributes to the aggregate traffic.

Overall, the factors, apart from application traffic itself, that contribute to the self-similar nature and the LRD behavior of the emerging network traffic are

- user behavior—user-reading time and user-induced delay,
- file-size distribution,
- set of files available in the server.

Table 6.5 summarizes different traffic distributions and their associated applications.

**Table 6.5** Traffic distributions and suitable applications

| Traffic distribution | Description |
|---|---|
| Poisson | Session arrival process |
| Exponential | Session duration |
| Heavy-tailed | Suitable for burst individual source traffic with ON/OFF patterns |
| • Pareto | File-transfer time distribution, user-reading (thinking) time, user-induced delay |
| • Weibull | Machine-processing time, file downloading time |

## 6.4 SELF-SIMILAR AND LRD TRAFFIC MODELS

As a conclusion, on the basis of the previous section, it appears that self-similar and LRD models are the most suitable models for future data networks, including the wireless IP. In this section, we explore these models further.

### 6.4.1 Traditional traffic models

Because of the long history of traditional telephony networks, there are plenty of traffic models available for voice-centric network traffic. The network traffic characteristics have been identified and extensive studies have already been completed to optimize the network resources.

#### *6.4.1.1 Poisson*

It is the oldest and one of the most elegant traffic models. The Poisson model is suitable for traffic applications that physically comprise a large number of independent traffic streams. Mathematically, the Poisson process is expressed as

$$P(n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \tag{6.8}$$

where $\lambda$ is the arrival rate per session and $n$ is the number of individual traffic streams. The interarrival times $\{A_n\}$ are exponentially distributed with

$$P\{A_n \leq t\} = 1 - e^{-\lambda t} \tag{6.9}$$

The Poisson model has some elegant analytical properties:

- The superposition of the independent Poisson process is a new Poisson process.
- It is a memory-less process.

However, the model fails to capture the autocorrelation of traffic as it vanishes identically for all nonzero lags. It is expected that burst data traffic will dominate the future network traffic. In that case, it is essential to capture the autocorrelated nature of the traffic for predicting the performance. In high-speed data networks, the Poisson process is no longer appropriate and has lost its merits.

#### *6.4.1.2 Markov*

Unlike the Poisson model, the Markov model introduces some dependency into the random sequence $\{A_n\}$; therefore, it capture the traffic 'burstiness'. The process $\{A_n\}$ is defined in terms of a Markov transition matrix $P = [p_{ij}]$. The Markov property introduces dependency into inter-arrival separation, batch sizes, and successive workloads.

However, any traffic modeling requires a multistate Markov and each state adds several free parameters. In practice, it is time consuming to estimate these parameters.

#### 6.4.1.3 Markov modulated

It introduces an explicit notion of state into the description of a traffic stream. Let $M = \{M(t)\}_{t=0}^{\infty}$ be a continuous-time Markov process with state space $\{1, 2, \ldots, m\}$. Assuming $M$ is in state $k$, the probability law for traffic arrivals is completely determined by $k$ and this holds for every $1 \leq k \leq m$. When $M$ undergoes a transition to state $j$, then a new probability law for arrivals takes effect for the duration of state $j$, and so on. The most commonly used form of the Markov modulated process is the MMPP.

MMPP combines the simplicity of the modulating Markov process with that of the modulated Poisson process. It is particularly suitable for use in a single traffic source with a variable rate, by quantifying the rate into a finite number of rates so that each rate gives rise to a state in some Markov-modulating process. For example, a simple two-state MMPP model has been widely used to model voice traffic sources.

### 6.4.2 Current and future models

#### 6.4.2.1 Fluid traffic model

In this model, traffic is considered as volume and is characterized by a flow rate. It is suitable to model the traffic where the individual traffic unit is insignificant, for example, individual cells in broadband ISDN (B-ISDN) ATM networks. Here, larger traffic units provide a simpler and better analysis of the network performance as well as saving, simulation, and computing resources. Fluid models are suitable for modeling burst traffic with ON/OFF patterns. For analytical tractability, the following assumptions are made:

- The ON-state traffic arrives deterministically at a constant rate $\lambda$.
- Traffic is switched off during the OFF state.
- The ON and OFF periods are exponentially distributed and mutually independent.

#### 6.4.2.2 Self-similar models

Here, we describe three commonly used models for the self-similar process.

- *Fractional ARIMA (FARIMA)*: For LRD modeling, FARIMA is one of the most commonly used models for the self-similar process. The main advantage of this model is that it can model both LRD and SRD processes simultaneously. In addition, FARIMA provides quick simulation. It is particularly useful to simulate the queuing performance of SRD and LRD traffic simultaneously. By changing the parameters that affect the degree of SRD and LRD, we can identify the parameters that are more or less sensitive to SRD or LRD.

- *Fractional Gaussian Noise (FGN)*: Together with the FARIMA, FGN is another most frequently used stochastic model for self-similar traffic modeling. It is suitable for burst data and multimedia application traffic modeling with a prevalence of LRD. It provides a good estimation of queuing performance for aggregate traffic.
- *Transform-Expand-Sample (TES)*: This is able to capture both the marginal distributions and the autocorrelations of the measured traffic. A good transform expand sample (TES) model should satisfy the following three requirements simultaneously:
  1. The histogram of measured traffic matches the model's marginal distribution.
  2. The model's autocorrelations should match the measured traffic up to a reasonable lag.
  3. Good correspondence exists between the sample paths of the simulated and the measured data.

#### 6.4.2.3 Long-range-dependence (LRD) models

Here, we describe three commonly used models for the LRD process.

- *Fractional Brownian Motion (FBm)*: It is a Gaussian process with a mean zero and stationary increments. Let us define $B_H$ as an FBm and its covariance function as:

$$B_H(s)B_H(t) = (1/2)\{s^{2H} + t^{2H} - |s-t|^{2H}\} \tag{6.10}$$

Its increments

$$Gj = B_H(j) - B_H(j-1) \quad j = 1, 2, \ldots \tag{6.11}$$

are called fractional Gaussian noise and

$$G_H(j)G_H(j+k) \approx H(2H-1)k^{2H-2} \text{ as } k \longrightarrow \infty \tag{6.12}$$

The power-law decay of the covariance characterizes long-range-dependence. As $H$ becomes larger, the decay becomes slower.

- *M/G/∞*: The M/G/∞ model is chosen to generate self-similar arrivals. The advantage of this model is that it introduces multifractal behavior at small/medium timescales without affecting the asymptotic self-similarity. It is considered to be more conservative than FBm as it predicts a stricter queuing performance.
- *M/Pareto*: The M/Pareto model is a particular type of the general M/G/∞ model. It is simple and particularly useful to estimate the queuing performance of a variety of realistic multimedia traffic streams. Another benefit of using M/Pareto is that the superposition of multiple independent M/Pareto processes is an M/Pareto process with a combined Poisson rate, $\lambda$. With an appropriate choice of $\lambda$, the M/Pareto process provides an accurate prediction of the queuing performance. Some of drawbacks are (1) there is no systematic way of calculating the appropriate value of $\lambda$ and (2) it is difficult to estimate the Hurst parameter, $H$, from a finite data set.

Table 6.6 summarizes characteristics of different traffic models.

**Table 6.6** Traditional, current, and future traffic models

| Traffic model | Applications | Mathematical complexity | Computing complexity | Advantages | Disadvantages |
|---|---|---|---|---|---|
| Poisson | • Voice<br>• Large number of independent traffic streams | Low | Low | • Oldest and commonly used model<br>• Superposition of Poisson process is a new Poisson<br>• Memory-less process | • Fails to capture autocorrelation<br>• Optimistic estimation of queuing performance for burst traffic |
| Markov | N/A | High | High | • Capable of capturing correlation of traffic (i.e. nonzero autocorrelations) | • Inflexible<br>• Complexity overshadows accuracy |
| MMPP | • A single traffic source with variable rates | Low | Low | • Simple and flexible<br>• Possible to capture some degree of correlation of traffic | • Inadequate autocorrelation<br>• Unsuitable for LRD traffic |
| Fluid | • ATM traffic<br>• Bursty traffic | Medium | Low | • Simple<br>• Fast simulation<br>• Suitable to model bursty traffic with ON/OFF patterns | • Unsuitable for variable rate traffic |
| Fractional ARIMA | • Voice<br>• Bursty data and multimedia traffic | Low | Medium—high | • Flexible<br>• Suitable for self-similar traffic with SRD and LRD | • High computing complexity |
| TES | • Broadband traffic streams<br>• Nonstationary traffic | Medium | Low | • Fast simulation<br>• Suitable to capture both marginal and autocorrelation functions of the traffic | • Requires high programming complexity |

| | | | | | |
|---|---|---|---|---|---|
| Gaussian | • Aggregated network traffic | • Low | • Low | • Simple<br>• Good representation of network traffic as more traffic is aggregated together | • Overly optimistic estimation of network performance if the aggregation level is low |
| FBm (continuous-time) | • Real-audio<br>• Real-video<br>• Aggregated network traffic | • Low | • Medium—high | • Flexible<br>• No need to select a sampling interval<br>• Simplest Gaussian model to capture today's network traffic | • Unsuitable for small timescales simulation<br>• Optimistic estimation of queuing performance |
| Fractional Gaussian noise (Discrete-time) | • Burst data & multimedia application traffic | • Medium | • Medium | • Flexible<br>• Good estimation of queuing performance for aggregated traffic | • Unsuitable for self-similar traffic with both SRD and LRD |
| Hyper-Erlang | • User mobility<br>• Self-similar traffic | • Low | • Low | • Simple and general<br>• Provides a good user mobility model in wireless and mobile networks | • Unsuitable in traffic management context |
| M/Pareto | • Broadband traffic streams (Ethernet, IP) | • Low | • Low—medium | • Simple<br>• Suitable for current network traffic where traffic is not Gaussian enough<br>• Good estimation of queuing performance | • Inadequate marginal distribution or autocorrelation function<br>• No simple formula to determine the appropriate value for $\lambda$ or $H$ |
| M/G/$\infty$ | • Aggregated network traffic | • Medium | • Medium | • Introduce multifractal behavior at small/medium timescales<br>• Good estimation of queuing performance | — |

### 6.4.3 Traffic models for the Internet applications

In Reference [18], four types of traffic profiles have been proposed, on the basis of the most frequently used wireless applications, e-mail, WWW, file transfer protocol (FTP), and telemetry traffic. Table 6.7 summarizes data traffic models and their respective numerical parameters. We will discuss the first three traffic types in this section. For the readers who are interested in other literature on the Internet data traffic model, Reference [24] and the references given therein provide some mathematical representations.

#### 6.4.3.1 E-mail traffic

E-mail traffic models are summarized in Table 6.8. The message is downloaded from the mail server to the mobile terminal during the ON period. The length of the ON period depends on the message size and the instantaneous throughput available to the user. The OFF period is the reading time taken by the user.

The OFF period distribution of the e-mail is Pareto. The minimum OFF period (i.e. $k_e$) is the minimum time required by a user to read an e-mail message. From the given parameters, the e-mail OFF time distributions are illustrated in Figure 6.7 for $k_e = 30$ s and 60 s. Comparing the OFF time distributions of $k_e = 30$ s and $k_e = 60$ s, it is reasonable to assume that $k_e = 30$ s, since most users will finish reading an e-mail message in 2 to 3 min, provided there is no attachment.

#### 6.4.3.2 WWW traffic

Typical WWW traffic models are summarized in Table 6.9. For the WWW traffic, the ON and OFF patterns are still clear but we have also started observing active and inactive OFF patterns (see Figure 6.8). As in e-mail, the file is transferred on the downlink during the ON period and its period depends on the file size, $\alpha$, and the available downlink bandwidth.

The ON period distribution is based on the file size. Here, $k_w$ is the minimum file size in bytes during the ON period. However, the Inactive OFF period distribution is based on the user reading time; therefore, $k'_w$ is the minimum time required by the user to read a web page.

##### *Active OFF time*

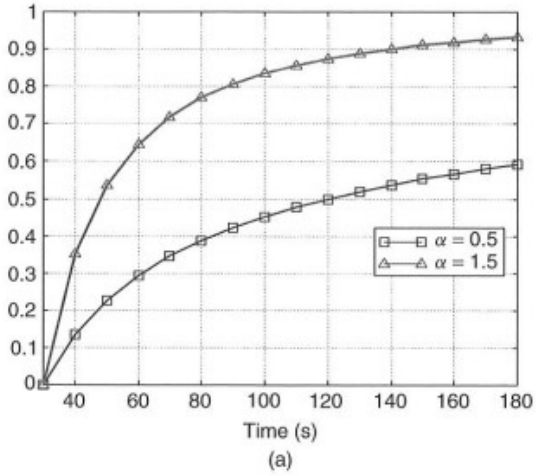
By definition, active OFF represents the time needed to process transmitted files such as interpret, format, and display a document component. Generally, if the OFF time is less than 1 s, it is assumed to be the machine-processing and display time for data items that are retrieved as part of a multipart document. However, some embedded components require more than 30 s to interpret, format, and display. As a rule, if the OFF time is greater than 30 s, it is considered as user-initiated delay (i.e. reading time). Therefore, the minimum value of $k'_w$ for the inactive OFF-time distribution should be at least 30 s (i.e. $k'_w = 30$ s). Figure 6.9 shows the active OFF time distribution based on the parameters given in

**Table 6.7** Wireless packet data traffic models and parameters [18]

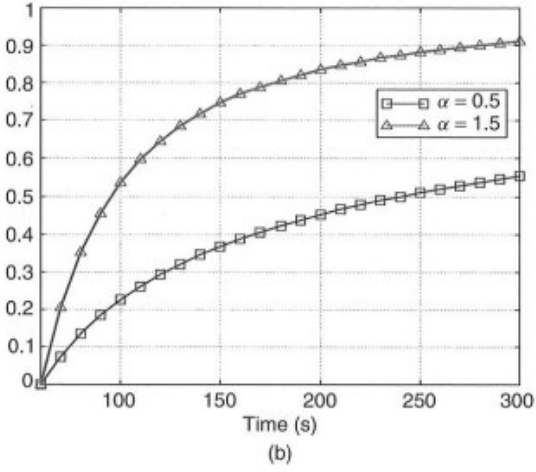
| Application | Period | Distribution | Formula | Parameters |
|---|---|---|---|---|
| E-mail | Packet arrival | Poisson | $P(m_e = n) = \frac{(P_e\lambda_e T_e)^n}{n!} e^{-p_e\lambda_e T_e}$ | |
| | ON | Weibull | $F_e(x_e) = \begin{cases} 1 - e^{-e^{k1} x_e^{c1}} \\ 1 - e^{-e^{k2} x_e^{c2}} \end{cases}$ | $C_1 = 1.2 - 3.2(m = 2.04)$, $C_2 = 0.31 - 0.46(m = 0.37)$ $k_1 = 14.0 - 21.0(m = 17.64)$, $k_2 = 2.8 - 3.4$ |
| | OFF | Pareto | $\Gamma_e(t_e) = 1 - \left(\frac{k_e}{t_e}\right)^{\alpha_e}$ | $k_e = 30 - 60$ s, $\alpha_e = 0.5 - 1.5$ |
| WWW | ON | Pareto | $f_W(x_W) = 1 - \left(\frac{k_W}{x_W}\right)^{\alpha_w}$ | $k_w = 1000$ bytes, $\alpha_w = 1.1 - 1.5$ |
| | Active OFF | Weibull | $\Gamma_W(t_W) = 1 - e^{-\left(\frac{t_w}{a}\right)^b}$ | $a = 0.328$, $b = 1.46$ |
| | Inactive OFF | Pareto | $\Gamma'_W(t'_W) = 1 - \left(\frac{k'_W}{t'_W}\right)^{\alpha'_w}$ | $k'_w = 1$ sec, $a'_w = 1.5$ |
| FTP | ON | Pareto | $F_f(t_f) = 1 - \left(\frac{k_f}{t_f}\right)^{\alpha_f}$ | $0.9 \le \alpha_f \le 1.1$ (by Crovella [14]) |
| | OFF | Weibull | $\Gamma_f(t_f) = 1 - e^{-\left(\frac{t_f}{b}\right)^a}$ | |

**Table 6.8** E-mail traffic models

| E-mail | Traffic models |
|---|---|
| ON period | Weibull distribution |
| OFF period | Pareto distribution |




(a)




(b)

**Figure 6.7** E-mail OFF time Pareto distribution with (a) $k_e = 30$ s; (b) $k_e = 60$ s

**Table 6.9** WWW traffic models

| WWW | Traffic models |
|---|---|
| ON | Pareto distribution |
| Active OFF | Weibull distribution |
| Inactive OFF | Pareto distribution |

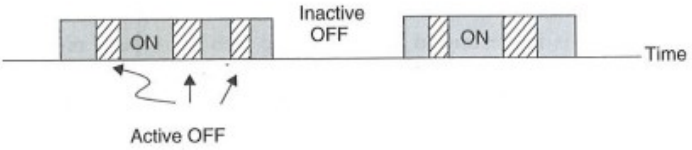



**Figure 6.8** Active and inactive OFF patterns in WWW traffic

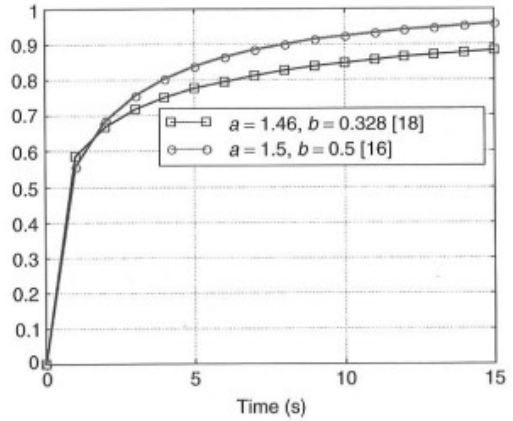



**Figure 6.9** WWW active OFF period with different parameters

References [16,18]. The active OFF-time distribution provided by Reference [16] is more heavy-tailed. If the web page is text-intensive, the parameters provided by Reference [18] are more suitable. However, today, there is a lot of image, audio, and even video files in the web page; therefore, the parameters provided by Reference [16] seem more appropriate to use.

### *Inactive OFF time*

The inactive OFF time is the user reading time. According to References [14,18], the ON period is more heavy-tailed (i.e. smaller $\alpha$) than the OFF period. However, in Reference [16], it is the inactive OFF time that contributes more to the heavy-tailed behavior. Assuming that the minimum reading time of a web page (i.e. $k'_w$) is 30 s, the parameters

given in References [13,17] provide a reasonable inactive OFF-time distribution. Reference [16] has also assumed that $k'_w = 30$ s and the OFF-time distribution assumed there is more appropriate for the text-intensive web pages, where it takes more time for a user to finish reading. However, $k'_w = 1$ s is too short a time and provides an inappropriate OFF-time distribution. This is shown in Figure 6.10. For the sake of comparison, Figure 6.11 also illustrates the WWW ON period file size distribution for different $k_w$.

### *Web file size*

In Reference [14], it is interesting to notice that the web file system prefers documents in the 256–512 byte range, while with the UNIX file system the file sizes are more commonly in the 1000–4000 byte range. Also, UNIX files show heavier tail distribution

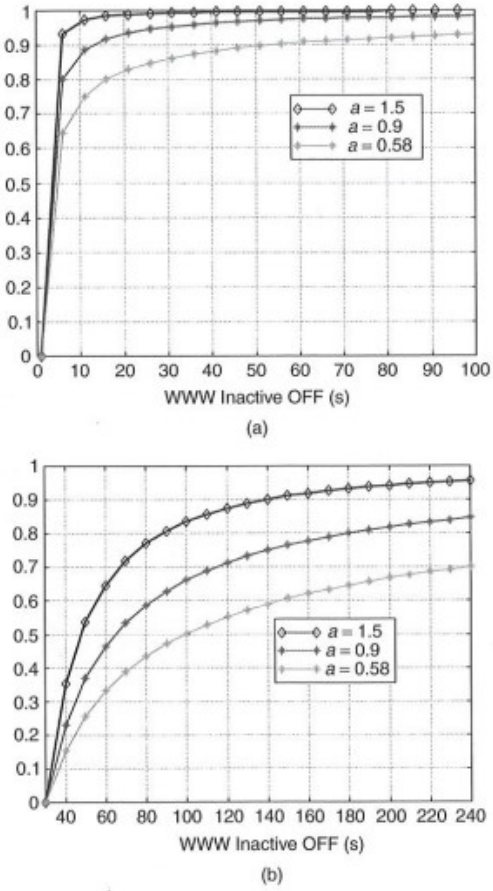



**Figure 6.10** WWW inactive OFF-time distribution with (a) $k'_w = 1$ s; and (b) $k'_w = 30$ s

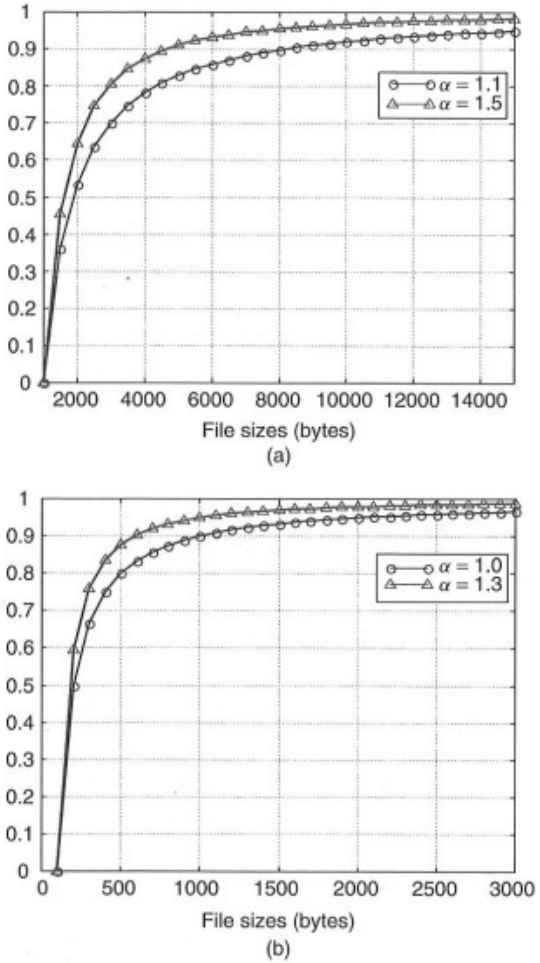



**Figure 6.11** WWW ON period file size distribution with (a) $k_w = 1000$ bytes; and (b) $k_w = 100$ bytes

**Table 6.10** File types and their sizes in bytes

| File sizes (bytes) | File types |
|---|---|
| <1000 | Text |
| 1000–30 000 | Image |
| 30 000–300 0000 | Audio |
| 300 000 | Video |

(i.e. smaller $\alpha$) than web files despite the emphasis on multimedia in the web. The web file systems are currently more biased toward small files than UNIX systems. Some typical Internet file sizes are listed in Table 6.10.

Table 6.11 provides a comparison between the traffic models and numerical parameters proposed for WWW applications in References [14,16,18].

#### 6.4.3.3 FTP traffic

The behavior of the FTP sessions is similar to e-mail but with larger file sizes and longer ON periods. A summary of appropriate FTP traffic models is given in Table 6.12.

In case of an FTP session, the OFF periods may be shorter than the ON periods. The OFF-periods distribution rather depends on the user-induced delay such as user think time and typing speed. As pointed out in Table 6.5, the Weibull distribution is more appropriate to describe the machine-processing, interpret, and display times. Therefore, the Pareto distribution will provide a better fit for the OFF-period distribution.

## 6.5 SHORT-RANGE AND LONG-RANGE DEPENDENCE MODELS

Recent studies on high-speed traffic such as Ethernet packets show not only LRD but also strong short-range dependence (SRD) as well. As the network gets larger and carries traffic from more independent sources, future traffic will be more and more Gaussian-like. However, the current network traffic is not anywhere near Gaussian. Therefore, in order to capture the current and emerging network traffic characteristics, it is necessary that the traffic models are able to represent both LRD and SRD simultaneously.

### 6.5.1 Self-similar traffic models

Both FGN and Fractional ARIMA (FARIMA) are the most commonly used stochastic self-similar traffic models. However, FARIMA is the preferred model as it can be used to model both SRD and LRD simultaneously. In addition, although there is no extensive work done in TES modeling, it can also be used to model both SRD and LRD and promises to give the three important traffic characteristics described previously in Table 6.2. On the basis of these conclusions, the most preferable self-similar traffic models are FARIMA, TES, and FGN in that order. Table 6.13 illustrates this conclusion.

### 6.5.2 Long-range-dependence traffic models

Although it is hard to determine the sufficient aggregation level where short-range dependence (SRD) effects can be ignored, if the traffic is aggregated enough, SRD would be

**Table 6.11** Comparison of the WWW traffic models and corresponding parameters

| Application | Period | Reference [17] | Reference [15] | Reference [13] |
|---|---|---|---|---|
| WWW | ON | Pareto<br>$f_w(x_w) = 1 - \left(\frac{k_w}{x_w}\right)^{\alpha_w}$<br>$x_w$: WWW file size<br>$k_w$: minimum file size<br>$\alpha_w = 1.1 - 1.5$<br>$k_w = 1000$ bytes or larger | Weibull<br>$\Gamma_{w_{ON}}(t_{w_{ON}}) = 1 - e^{-\left(\frac{t_{w_{on}}}{\theta}\right)^k}$<br>$\theta : e^{4.4} - e^{4.6}$<br>$k : 0.91 - 0.77$ | Pareto<br>$f_w(x_w) = 1 - \left(\frac{k_w}{x_w}\right)^{\alpha_w}$<br>$x_w$: WWW file size<br>$k_w$: minimum file size<br>$\alpha_w = 1.0 - 1.3$<br>$k_w = 100$ bytes or larger |
| | Active OFF | Weibull<br>$\Gamma_w(t_w) = 1 - e^{-\left(\frac{t_w}{a}\right)^b}$<br>$a = 0.328, b = 1.46$ | Weibull<br>$\Gamma_w(t_w) = 1 - e^{-\left(\frac{t_w}{a}\right)^b}$<br>$a = 1.5, b = 0.5$ | Weibull |
| | Inactive OFF | Pareto<br>$\Gamma'_w(t'_w) = 1 - \left(\frac{k'_w}{t'_w}\right)^{\alpha'_w}$<br>$k'_w = 1, \alpha'_w = 1.5$ | Pareto<br>$\Gamma'_w(t'_w) = 1 - \left(\frac{k'_w}{t'_w}\right)^{\alpha'_w}$<br>$k'_w = 30, \alpha'_w = 0.9 - 0.58$ | Pareto<br>$\Gamma'_w(t'_w) = 1 - \left(\frac{k'_w}{t'_w}\right)^{\alpha'_w}$<br>$k'_w = 30, \alpha'_w = 1.5$ |

**Table 6.12** FTP traffic models

| FTP | Traffic model |
|---|---|
| ON period | Pareto distribution |
| OFF period | Weibull distribution |

**Table 6.13** Self-similar traffic modeling preferences

| Preference | Traffic models | Traffic types | Applications |
|---|---|---|---|
| 1 | FARIMA . | Self-similar traffic with both SRD and LRD | • Ethernet traffic modeling<br>• LAN<br>• Cooperate network |
| 2 | TES | Self-similar traffic with both SRD and LRD | • LAN<br>• Cooperate network traffic modeling |
| 3 | FGN | Self-similar traffic with LRD only | • WAN |

**Table 6.14** Traffic models for LRD modeling

| Preference | Traffic models | Traffic types | Applications |
|---|---|---|---|
| 1 | M/Pareto | • LRD | • Multimedia traffic<br>• Broadband traffic in general |
| 2 | M/G/∞ | • Multifractal LRD traffic | • WAN |

averaged out. We only need to consider the LRD properties. On the basis of the mathematical and computational complexity, the M/Pareto and M/G/∞ models are appropriate for LRD modeling, as shown in Table 6.14.

A traffic model should match most of the measured traffic characteristics. However, a model is a tool for decision-making. Its quality depends on the quality of the decisions it leads to rather than on its closeness to physical reality [12].

## 6.6 SUMMARY AND CONCLUSIONS

In this chapter, we have examined the most suitable traffic models for the communication networks, with an emphasis on current and future data networks and wireless IP networks. The future networks will have the Internet applications as their primary sources of traffic and, similar to the requirement of an appropriate model in the traditional telephony networks, we will need to come up with respective traffic models for the future wireless data network, in order to design them more appropriately.

Modeling of data traffic loads, however, is not an easy task and not comparable with the voice-centric telephony networks. There are many different multimedia traffics coming

from current and future applications, and having a single model to illustrate the characteristics of all these traffic would be a complex research task in the years to come.

Considering the exponential increase in the traffic load of the data networks and the necessity of designing these systems by using precise traffic models, there are not many available models. This could be because of the complexity involved in finding these models or the lack of feeling the requirement for such a traffic model at this time. Soon, the wireless data technology will find the need to investigate more on this important issue, the fundamentals of which we have described in this chapter. The materials presented here can provide the required knowledge for the traffic engineering as well as for researchers in the field.

When a good traffic model is not available during the design process of a communication network or when applying an available traffic model, it makes the network design too complicated, and we need to search for other alternatives. Traffic management techniques are considered as appropriate partial replacements to precise traffic modeling, and we thus discuss this topic in the following chapter.

## REFERENCES


1. Hong D & Rappaport SS, Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures, *IEEE Transactions on Vehicular Technology*, **VT-35**(3), 1986.
2. Ho J Zhu Y & Madhavapeddy S, Throughput and buffer analysis for GSM general packet radio services, *Proceedings of IEEE WCNC '99*, New Orleans, September 1999.
3. Wirth P, Teletraffic implications of database architectures in mobile and personal communications, *IEEE Communications Magazine*, **33**(6), 54–59, 1995.
4. Wirth P, The role of teletraffic modeling in the new paradigm, *IEEE Communications Magazine*, **35**(8), 86–92, 1997.
5. Addie R, Zukerman M & Neame T, Broadband traffic modeling: simple solutions to hard problems, *IEEE Communications Magazine*, **36**(8), 88–95, 1998.
6. Dahlberg TA & Jung J, Teletraffic modeling for mobile communications, *Proceedings of IEEE ICC '98*, Atlanta, Ga., June 1998.
7. Fiorini P, On modeling concurrent heavy-tailed network traffic sources and its impact upon QoS, *Proceedings of IEEE ICC '99*, Vancouver, Canada, June 1999.
8. Lam D, Cox D & Widom J, Teletraffic modeling for personal communications services, *IEEE Communications Magazine*, **35**(2), 79–87, 1997.
9. Lazar A, Programming telecommunication networks, *IEEE Network*, **11**(5), 8–18, 1997.
10. Sahinoglu Z & Tekinay S, On multimedia networks: self-similar traffic and network performance, *IEEE Communications Magazine*, **37**(1), 48–52, 1999.
11. Fang Y, Hyper-Erlang distributions and traffic modeling in wireless and mobile networks, *Proceedings of IEEE WCNC '99*, New Orleans, September 1999.