CHAPTER 1

FROM DAMS TO TELECOMMUNICATION – A SURVEY

N.U. PRABHU

School of Operations Research and Industrial Engineering
Cornell University, Ithaca, NY 14853-3801

Abstract: In 1954 P.A.P. Moran formulated a simple discrete time model for a finite dam. This model was extended in several directions by J. Gani and the author during 1956-63. The concepts underlying this model and the techniques used in its analysis are applicable in a wide variety of situations, as has already been demonstrated. Most recently, models for data communication systems have also been analyzed with these techniques. In this paper we survey some of this work.

Keywords and phrases: buffer content, dam, data communication, idle time, input, fluid input, Lévy process, Markov chain, Markov-additive process, packets, Poisson arrivals, queues, subordinator, unmet demand, workload

1.1 INTRODUCTION

In 1954 P.A.P. Moran formulated a simple discrete time model for the finite dam. The basic components of this model are inputs that are independent and identically distributed random variables, a constant demand for water and the release policy "meet the demand if physically possible." During 1956-63 J. Gani and the author extended this discrete time model to continuous time, where the input is described by a subordinator, the demand is at a unit rate and the release policy is the same as before. This continuous time model has several applications, in particular, to single server queues with Poisson arrivals and first come, first served discipline or priority discipline of the static or dynamic type. Because these models have several common features in regard to the underlying concepts and techniques of analysis, the author proposed the term *stochastic storage processes* to describe the processes that arise from the family of such models and presented a unified theory of these processes (see Prabhu (1998)). The most recent extension of this theory is to models for transmission of telecommunication data. Here the input of data is characterized as a Markov-additive process, the desired transmission (demand) rate depends on the Markov component of the input and the actual transmission (release) policy is to "meet the demand if physically possible." The resulting theory may be viewed as the Markov-modulated version of the theory of dams.

In this paper we survey some of this work, emphasizing only the modeling aspects in order to point out the common features of the models considered. For detailed results and recent references see Prabhu (1998). For historical references see Prabhu (1965).

In section (1.2) we describe Moran's discrete time model for the finite dam. The continuous time dam model is described in section (1.3), and its extension to the data communication model in section (1.4).

1.2 MORAN'S MODEL FOR THE FINITE DAM

Moran's discrete time model for a dam (water reservoir) is the following. A dam of finite capacity is designed to meet the demand for electric power (expressed in terms of the volume of water required to produce it) or for water to be supplied to a city. The demand for water at time n is m ($< c$) and this demand is met "if physically possible," that is, to the extent that this quantity is available in the dam at time n. The dam is fed by inputs of water such that if X_{n+1} denotes the input during the time interval $(n, n+1]$, then $\{X_n, n \geq 1\}$ is assumed to be a sequence of independent and identically distributed random variables. Because of this randomness the amount of water in the dam (the dam content) at time n is a random variable which we denote by Z_n ($n \geq 0$).

Since the capacity of the dam is finite there is a possibility of an overflow and the actual input during $(n, n+1]$ is therefore

$$\eta_{n+1} = \min(c - Z_n, X_{n+1}) \quad (n \geq 0). \tag{1.2.1}$$

The amount of water available for release at time $n+1$ is then $Z_n + \eta_{n+1}$

and the release policy implies that

$$Z_{n+1} = Z_n + \eta_{n+1} - \min(m, Z_n + \eta_{n+1}).$$

The sequence $\{Z_n, n \geq 0\}$ satisfies the relation

$$Z_{n+} = (Z_n + \eta_{n+1} - m)^+ \quad (n \geq 0). \tag{1.2.2}$$

To see how the dam operates subject to these assumptions we note that during a time interval $(0, n]$ there is a certain amount F_n of overflow from the dam, and an amount D_n of the total demand nm that is not met. Easy calculations show that

$$Z_n = Z_0 + (S_n - nm) - F_n + D_n \quad (n \geq 0) \tag{1.2.3}$$

where $S_n = X_1 + X_2 + \cdots + X_n (n \geq 1), S_0 = 0$ and $S_n - nm$ is the net input (input minus demand) during $(0, n]$.

The assumption on the inputs X_n implies that $\{Z_n, n \geq 0\}$ is a time-homogeneous Markov chain on the state space $\mathbb{R}_+$. The problems of practical importance that arise in the analysis of the model are the derivation of (i) the steady state distribution of $\{Z_n\}$ and (ii) the distribution of the random variable

$$T(Z_0) = \min\{n \geq 1 : Z_n = 0\} \tag{1.2.4}$$

which is the duration of the wet period in the dam whose initial content is $Z_0 > 0$. Although these problems are standard in the theory of Markov chains, general solutions are not known because of the presence of the constant c $(< \infty)$ in (1.2.2). However, solutions are available for some important special cases of the input distributions (see Prabhu (1965)).

When $c = \infty$ (the case of the infinite dam) the equations (1.2.2) and (1.2.3) reduce to

$$Z_{n+1} = (Z_n + X_{n+1} - m)^+ \quad (n \geq 0) \tag{1.2.5}$$

and

$$Z_n = Z_0 + (S_n - m_n) + D_n \quad (n \geq 0). \tag{1.2.6}$$

These lead to the expressions

$$Z_n = \max\{Z_0 + S_n - nm, S_n - nm - m_n\} \tag{1.2.7}$$

$$D_n = (Z_0 + m_n)^- \tag{1.2.8}$$

where m_n is the minimum functional of the random walk $\{S_n - nm, n \geq 0\}$, namely

$$m_n = \min_{0 \leq k \leq n} (S_k - km) \quad (n \geq 0). \tag{1.2.9}$$

The equation (1.2.5) arises in queueing theory, specifically for waiting times Z_n in the single server queue with constant interarrival times ($= m$) and general service times $X_n (n \geq 1)$. The quantity D_n in (1.2.6) is the total idle period during (0,n], while the random variable $T(Z_0)$ defined by (1.2.4) is the number of customers served during the busy period initiated by a waiting time $Z_0 > 0$. Thus the results for the infinite dam are applicable to queueing theory.

1.3 A CONTINUOUS TIME MODEL FOR THE DAM

In developing a continuous time model for the dam we first assume that its capacity is ∞. For the input we postulate a nonnegative process with stationary independent increments, that is, a Lévy process $\{X(t), t \geq 0\}$ with nondecreasing sample functions (also called a subordinator) and zero drift. The demand for water occurs at a rate $d \circ Z(t)$, where $Z(t)$ is the dam content at time $t \geq 0$. As in the discrete time case, this demand is met "if physically possible". These assumptions lead to the integral equation

$$Z(t) = Z(0) + X(t) - \int_0^t d \circ Z(s) 1_{\{Z(s)>0\}} ds. \qquad (1.3.10)$$

We can rewrite this is

$$Z(t) = Z(0) + X(t) - \int_0^t d \circ Z(s) ds + \int_0^t d \circ Z(s) 1_{\{Z(s)=0\}} ds. \qquad (1.3.11)$$

Here on the right side of (1.3.11) the first integral represents the total demand during $(0, t]$ and the second integral is the part of this demand that is not met. The equation (1.3.11) is the continuous time analogue of (1.2.6).

The most extensively studied special case of (1.3.10) is the one with unit demand rate (that is, $d(x) \equiv 1$), which arises also in the queueing system $M/G/1$ and single server queues with Poisson arrivals and static or dynamic priorities. In the queue $M/G/1$, the input $X(t)$ of workload is a compound Poisson process, and $Z(t)$ represents the remaining workload (virtual waiting time) at time t. In dam models the special cases of input include the gamma process, stable process with exponent $1/2$ and the inverse Gaussian process. The integral equation (1.3.11) reduces in the case of unit demand rate to

$$Z(t) = Z(0) + Y(t) - \int_0^t 1_{\{Z(s)=0\}} ds. \qquad (1.3.12)$$

where $Y(t) = X(t) - t$ (the net input) and the integral

$$I(t) = \int_0^t 1_{\{Z(s)=0\}} ds \tag{1.3.13}$$

represents the amount of unmet demand (dry period in a dam or idle time in the queue $M/G/1$).

As formulated above, the integral equation (1.3.12) does not have a unique nonnegative solution. However, if we modify it by writing

$$Z(t) = Z(0) + Y(t) - \int_0^t 1_{\{Z(s)\leq 0\}} ds \tag{1.3.14}$$

then the unique nonnegative solution of (1.3.14) is given by

$$Z(t) = \max\{Z(0) + Y(t), Y(t) - m(t)\} \tag{1.3.15}$$

where $m(t)$ is the minimum functional

$$m(t) = \inf_{0\leq s\leq t} Y(s). \tag{1.3.16}$$

It follows from (1.3.14) that

$$I(t) = \int_0^t 1_{\{Z(s)=0\}} ds = [Z(0) + m(t)]^- \tag{1.3.17}$$

on account of the nonnegativity of $Z(t)$. The results (1.3.15) and (1.3.17) are the continuous time analogues of (1.2.7) and (1.2.8) for the discrete time case.

Remarks.
1. When $Z(0) = 0$, the solution (1.3.15) reduces to

$$Z(t) = Y(t) - m(t). \tag{1.3.18}$$

In current literature (1.3.18) is referred to as reflection mapping. This term does not give credit to the pioneering 1958 paper by E. Reich, who derived (1.3.15) for the virtual waiting time in $M/G/1$. Furthermore, the identification of the idle time with the minimum functional does not follow from the reflection mapping.
2. The joint distribution of $Z(t)$ and $I(t)$ can be obtained directly from (1.3.12). For the compound Poisson input the older technique of analysis is based on the forward Kolmogorov integro-differential equation for the distribution of $Z(t)$. □

1.4 A MODEL FOR DATA COMMUNICATION SYSTEMS

A buffer of infinite capacity receives inputs of data represented as a Markov-additive process $\{X(t), J(t), t \geq 0\}$ on the state space $\mathbb{R}_+ \times \mathcal{E}$ in which the additive component is a compound Poisson process. Specifically

$$X(t) = X_0(t) + \int_0^t a \circ J(s)ds. \tag{1.4.19}$$

Here $X_0(t)$ is a compound Poisson process in which the rate at which jumps occur as well as the jump sizes depend on the state of the Markov process J on a countable state space $\mathcal{E}$, these jumps representing the arrivals of packets. In addition X has a drift that occurs at a rate $a(j)$ when J is in state j, and the integral in (1.4.19) represents the amount of data that arrive in a fluid fashion. The desired transmission (demand) rate is $d(j)$ when J is in state j and the transmission (release) policy is to meet the demand "if physically possible." Let $Z(t)$ denote the buffer content at time $t \geq 0$. The above assumptions lead to the integral equation

$$Z(t) = Z_0(t) + X(t) - \int_0^t r \circ (Z(s), J(s))ds \tag{1.4.20}$$

where the release rate r is given by

$$\begin{aligned} r(x,j) &= d(j) \text{ if } x > 0 \\ &= \min(d(j), a(j)) \text{ if } x = 0. \end{aligned} \tag{1.4.21}$$

Comparison with (1.3.10) show that (1.4.20) is indeed an extension of the (now classical) dam model. The presence of J is to be understood with reference to specific models. We first consider two special cases.

A Fluid Model for Data Communication. If the arrival of data is only in a fluid fashion, then $X_0(t) \equiv 0$ and the integral equation (1.4.20) reduces to

$$Z(t) = Z(0) + \int_0^t x \circ J(s)^+ ds - \int_0^t x \circ J(s)^- 1_{\{Z(s)>0\}} ds \tag{1.4.22}$$

where $x(j)$ is the net input rate

$$x(j) = a(j) - d(j). \quad \square \tag{1.4.23}$$

A Model with Packet Arrivals. In the presence of packet arrivals we need to assume that the desired transmission rate $d(j)$ exceeds the rate of fluid arrival $a(j)$. The integral equation (1.4.20) then reduces to

$$Z(t) = Z(0) + X_0(t) - \int_0^t d_1 \circ J(s) 1_{\{Z(s)>0\}} ds \tag{1.4.24}$$

where $d_1(j) = d(j) - a(j) > 0$. □

The integral equation that describes each of the above models is of the form

$$Z(t) = Z(0) + X(t) - \int_0^t r \circ (Z(s), J(s)) ds \tag{1.4.25}$$

where $\{X(t), J(t)\}$ is a Markov-additive process and

$$r(x, j) = d(j) 1_{\{x>0\}}. \tag{1.4.26}$$

Comparing (1.4.25) with the integral equation (1.3.10) we see that the data communication models described here are extensions of the continuous time dam model of section (1.3). The unique nonnegative solution of (1.4.25), modified as in (1.3.14), is formally the same as (1.3.15), where the net input $Y(t)$ given by

$$Y(t) = X(t) - \int_0^t d \circ J(s) ds \tag{1.4.27}$$

and it should be noted that $\{Y(t), J(t)\}$ is a Markov-additive process.

The following are two fluid models that have been investigated in the literature. The presence of the Markov component J will be clear from these models.

a. A Multiple Source Data Handling System. There are N sources of messages, which are "on" or "off" from time to time. A switch receives messages at a unit rate from each source and transmits them at a fixed maximum rate c ($1 \leq N < \infty, 0 < c < \infty$). Messages that are not transmitted are stored in a buffer of infinite capacity (see Figure 1.1). Denoting by $J(t)$ the number of "on" sources at time $t \geq 0$, we assume that $\{J(t), t \geq 0\}$ is a birth and death process on the state space $\{0, 1, 2, \ldots, N\}$. Of interest is the buffer content $Z(t)$. It is seen that $Z(t)$ satisfies the integral equation

$$Z(t) = Z(0) + \int_0^t J(s) ds - \int_0^t r \circ (Z(s), J(s)) ds \tag{1.4.28}$$

where

$$\begin{aligned} r(x, j) &= c \text{ if } x > 0 \\ &= \min(j, c) \text{ if } x = 0. \end{aligned} \tag{1.4.29}$$

Clearly, this is a fluid model with $a(j) = j$ and $d(j) = c$. □

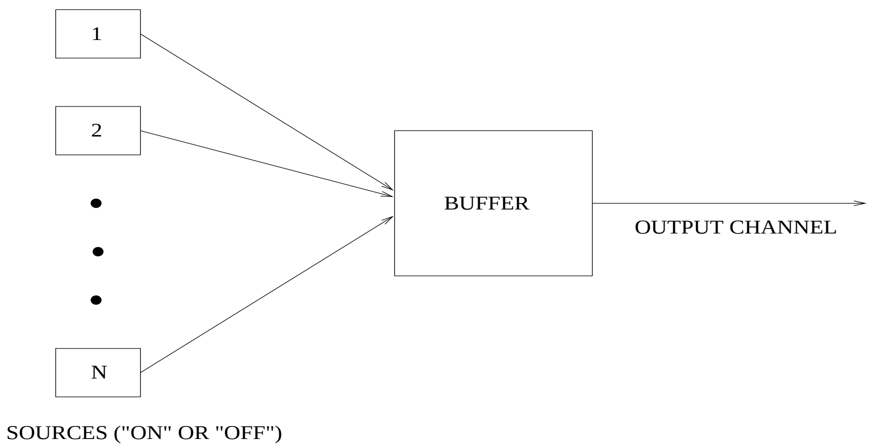

Figure 1.1
A Multiple Source Data Handling System

b. An Integrated Circuit and Packet Switching Multiplexer. A buffer of infinite capacity receives voice calls as well as data. There are $s+u$ output channels, of which u channels are reserved for data transmission, while the remaining s channels are shared by data and voice calls, with calls having preemptive priority over data and calls that find all s channels that serve them being lost (see Figure 1.2).

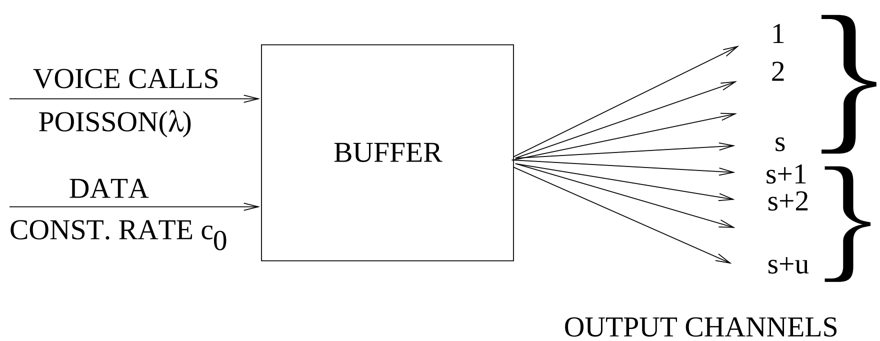

Figure 1.2
An Integrated Circuit and Packet Switching Multiplexer

Voice calls arrive in a Poisson process and their service times have an exponential density. Data arrive continuously at a constant rate c_0 and are transmitted at a rate $c_1(< c_0)$. At time $t \geq 0$, let $Z(t)$ denote the amount of data in the buffer and $J(t)$ the number of channels available for data transmission. It is clear that $s + u - J(t)$ represents the queue length in an $M/M/s$ loss system, and $Z(t)$ satisfies the integral equation

$$Z(t) = Z(0) + \int_0^t c_0 ds - \int_0^t r \circ (Z(s), J(s)) ds \tag{1.4.30}$$

where

$$\begin{aligned} r(x, j) &= c_1 j \text{ if } x > 0 \\ &= \min(c_1 j, c_0) \text{ if } x = 0. \end{aligned} \tag{1.4.31}$$

This is a fluid model with $a(j) = c_0$ and $d(j) = c_1 j$.

Remarks.
1. Some authors take (1.3.18) as the starting point of their investigation of data communication models. Such an approach neglects the modeling aspects that are important in any area of applied probability. In particular it does not emphasize the role of Markov-additive inputs.
2. The forward Kolmogorov equation (in the matrix form) can be used to derive the joint distribution of $Z(t)$ and $J(t)$. However, as in the case of the dam model it is much more straightforward to derive the joint distribution of $Z(t), I(t)$ and $J(t)$ directly from (1.4.25), $I(t)$ being the amount of the unmet demand.
3. It is hoped that this brief survey has made it clear that all of the models described in sections (1.3) and (1.4) are indeed*storage models.* The use of the term *fluid queue*, currently in fashion, is obviously based on lack of familiarity with earlier literature in this subject area. This term is both unnecessary and unpleasant, and the author hopes that discriminating researchers will not use it in the future. □

References

1. Prabhu, N.U. (1965). *Queues and Inventories: A Study of Their Basic Stochastic Processes*, John Wiley & Sons, New York.

2. Prabhu, N.U. (1980). *Stochastic Storage Processes.* Springer Verlag, New York. (1998): 2nd Edition.