WILEY

SYSTEM MODELING AND SIMULATION

An Introduction

Frank L. Severance

SYSTEM MODELING and SIMULATION

An Introduction

Frank L. Severance, Ph.D.

Professor of Electrical and Computer Engineering
Western Michigan University

JOHN WILEY & SONS, LTD

Chichester • New York • Weinheim • Brisbane • Singapore • Toronto

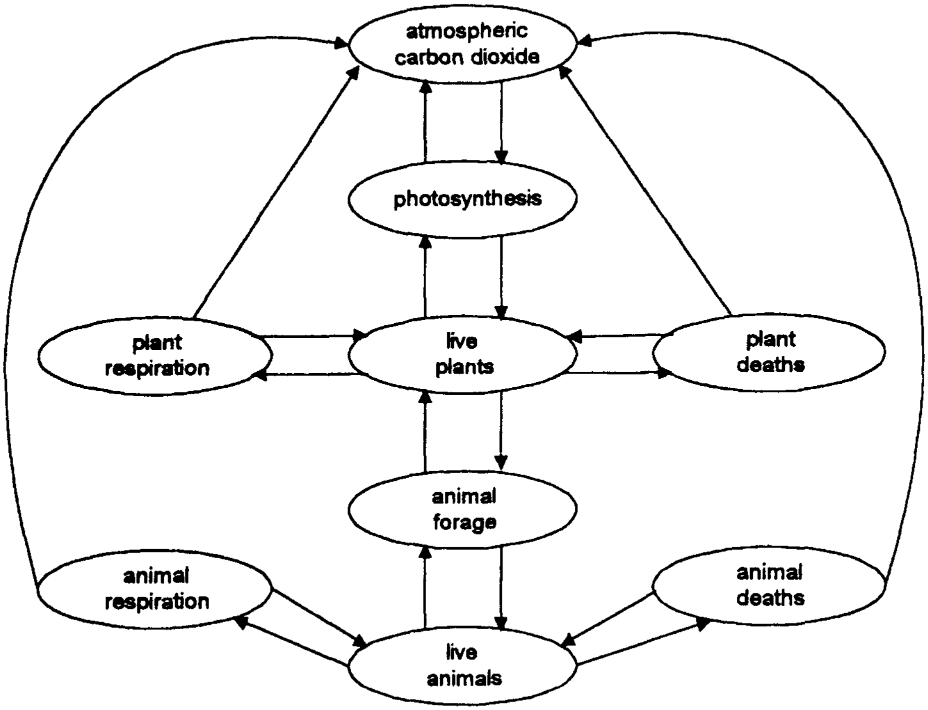
PREFACE

It is unlikely that this book would have been written 100 years ago. Even though there was a considerable amount of modeling going on at the time and the concept of signals was well understood, invariably the models that were used to describe systems tended to be simplified (usually assuming a linear response mechanism), with deterministic inputs. Systems were often considered in isolation, and the inter-relationships between them were ignored. Typically, the solutions to these idealized problems were highly mathematical and of limited value, but little else was possible at the time. However, since system linearity and deterministic signals are rather unrealistic restrictions, in this text we shall strive for more. The basic reason that we can accomplish more nowadays is that we have special help from the digital computer. This wonderful machine enables us to solve complicated problems quickly and accurately with a reasonable amount of precision.

For instance, consider a fairly elementary view of the so-called carbon cycle with the causal diagram shown on the next page. Every child in school knows the importance of this cycle of life and understands it at the conceptual level. However, "the devil is in the details", as they say. Without a rigorous understanding of the quantitative (mathematical) stimulus/response relationships, it will be impossible to actually use this system in any practical sense. For instance, is global warming a fact or a fiction? Only accurate modeling followed by realistic simulation will be able to answer that question.

It is evident from the diagram that animal respiration, plant respiration, and plant and animal decay all contribute to the carbon dioxide in the atmosphere. Photosynthesis affects the number of plants, which in turn affects the number of animals. Clearly, there are feedback loops, so that as one increases, the other decreases, which affects the first, and so on. Thus, even a qualitative model seems meaningful and might even lead to a degree of understanding at a superficial level of analysis. However, the apparent simplicity of the diagram is misleading. It is rare that the input–output relationship is simple, and usually each signal has a set of difference or differential equations that model the behavior. Also, the system input is usually non-deterministic, so it must be described by a random process. Even so, if these were linear relationships, there is a great body of theory by which closed-form mathematical solutions could, in principle, be derived.

We should be so lucky! Realistic systems are usually nonlinear, and realistic signals are noisy. Engineered systems especially are often discrete rather than continuous. They are often sampled so that time itself is discrete or mixed, leading to a system with multiple



time bases. While this reality is nothing new and people have known this for some time, it is only recently that the computer could be employed to the degree necessary to perform the required simulations. This allows us to achieve realistic modeling so that predictable simulations can be performed to analyze existing systems and engineer new ones to a degree that classical theory was incapable of.

It is the philosophy of this text that no specific software package be espoused or used. The idea is that students should be developers new tools rather than simply users of existing ones. All efforts are aimed at understanding of first principles rather than simply finding an answer. The use of a Basic-like pseudocode affords straightforward implementation of the many procedures and algorithms given throughout the text using any standard procedural language such as C or Basic. Also, all algorithms are given in detail and operational programs are available on the book's Website in Visual Basic.

This book forms the basis of a first course in System Modeling and Simulation in which the principles of time-driven and event-driven models are both emphasized. It is suitable for the standard senior/first-year graduate course in simulation and modeling that is popular in so many modern university science and engineering programs. There is ample material for either a single-semester course of 4 credits emphasizing simulation and modeling techniques or two 3-credit courses where the text is supplemented with methodological material. If two semesters are available, a major project integrating the key course concepts is especially effective. If less time is available, it will likely be that a choice is necessary – either event-driven or time-driven models. An effective 3-credit course stressing event-driven models can be formed by using Chapter 1, the first half of Chapter 3, and Chapters 7–9, along with methodological issues and a project. If time-driven models are to be emphasized, Chapters 1–6 and 10 will handle both deterministic and non-deterministic input signals. If it is possible to ignore stochastic signals and Petri nets, a course in both time-driven and event-driven models is possible by using Chapters 1 and 2, the first half of Chapter 3, Chapter 4, and Chapters 8–10.

ACKNOWLEDGEMENTS

As with any project of this nature, many acknowledgments are in order. My students have been patient with a text in progress. Without their suggestions, corrections, and solutions to problems and examples, this book would have been impossible. Even more importantly, without their impatience, I would never have finished. I thank my students, one and all!

Frank L. Severance
Kalamazoo, Michigan

Contents

CHAPTER 1

Describing Systems

1.1 THE NATURE OF SYSTEMS

The word "system" is one that everyone claims to understand – whether it is a physiologist examining the human circulatory system, an engineer designing a transportation system, or a pundant playing the political system. All claim to know what systems are, how they work, and how to explain their corner of life. Unfortunately, the term system often means different things to different people, and this results in confusion and problems. Still there are commonalities. People who are "system thinkers" usually expect that systems are (1) based on a set of cause–effect relationships that can be (2) decomposed into subsystems and (3) applied over a restricted application domain. Each of these three expectations require some explanation.

Causes in systems nomenclature are usually referred to as inputs, and effects as outputs. The system approach assumes that all observed outputs are functions only of the system inputs. In practice, this is too strong a statement, since a ubiquitous background noise is often present as well. This, combined with the fact that we rarely, if ever, know everything about any system, means that the observed output is more often a function of the inputs and so-called white noise. From a scientific point of view, this means that there is always more to discover. From an engineering point of view, this means that proposed designs need to rely on models that are less than ideal. Whether the system model is adequate depends on its function. Regardless of this, any model is rarely perfect in the sense of exactness.

There are two basic means by which systems are designed: top-down and bottom-up. In top-down design, one begins with highly abstract modules and progressively decomposes these down to an atomic level. Just the opposite occurs in bottom-up design. Here the designer begins with indivisible atoms and builds ever more abstract structures until the entire system is defined. Regardless of the approach, the abstract structures encapsulate lower-level modules. Of course there is an underlying philosophical problem here. Do atomic elements really exist or are we doomed to forever incorporate white background noise into our models and call them good enough? At a practical level, this presents no problem, but in the quest for total understanding no atomic-level decomposition for any physically real system has ever been achieved!

The power of the systems approach and its wide acceptance are due primarily to the fact that it works. Engineering practice, combined with the large number of mathematically powerful tools, has made it a mainstay of science, commerce, and (many believe) western culture in general. Unfortunately, this need for practical results comes at a price. The price is that universal truth, just like atomic truth, is not achievable. There is always a restricted range or zone over which the system model is functional, while outside this application domain the model fails. For instance, even the most elegant model of a human being's circulatory system is doomed to failure after death. Similarly, a control system in an automobile going at 25 miles per hour is going to perform differently than one going at 100 miles per hour. This problem can be solved by treating each zone separately. Still there is a continuity problem at the zone interfaces, and, in principle, there needs to be an infinite number of zones. Again, good results make for acceptance, even though there is no universal theory.

Therefore, we shall start at the beginning, and at the fundamental question about just what constitutes a system. In forming a definition, it is first necessary to realize that systems are human creations. Nature is actually monolithic, and it is we, as human beings, who either view various natural components as systems or design our own mechanisms to be engineered systems. We usually view a system as a "black box", as illustrated in Figure 1.1. It is apparent from this diagram that a system is an entity *completely isolated from its environment* except for an entry point called the *input* and an exit point called the *output*. More specifically, we list the following system properties.

P1. All environmental influences on a system can be reduced to a vector of m real variables that vary with time, $x(t) = [x_1(t), \ldots, x_m(t)]$. In general, $x(t)$ is called the *input* and the components $x_i(t)$ are *input signals*.

P2. All system effects can be summarized by a vector of n real variables that vary with time, $z(t) = [z_1(t), \ldots, z_n(t)]$. In general, $z(t)$ is called the *output* and the components $z_i(t)$ are *output signals*.

P3. If the output signals are algebraic functions of only the current input, the system is said to be of *zeroth order*, since there can be no system dynamics. Accordingly, there is a state vector $y(t) = [y_1(t), \ldots, y_p(t)]$, and the system can be written as

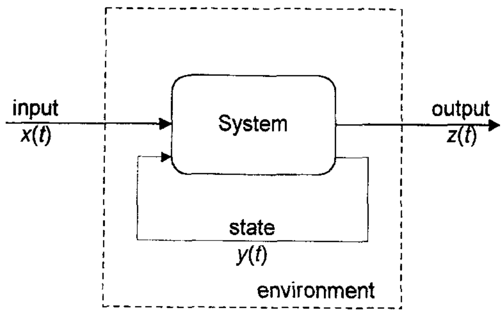


FIGURE 1.1 System block diagram.

two algebraic equations involving the input, state, and output:

$$\begin{aligned} y(t) &= f_1(x(t)), \\ z(t) &= f_2(x(t), y(t)), \end{aligned} \tag{1.1}$$

for suitable functions f_1 and f_2. Since the state $y(t)$ is given explicitly, an equivalent algebraic input–output relationship can be found. That is, for a suitable function g,

$$z(t) = f_2(x(t), f_1(x(t))) \equiv g(x(t)). \tag{1.2}$$

P4. If the input signal depends dynamically on the output, there must also be system memory. For instance, suppose that the system samples a signal every $t = 0, 1, 2, \ldots$ seconds and that the output $z(t)$ depends on input $x(t-1)$. It follows that there must be two memory elements present in order to recall $x(t-1)$ and $x(t-2)$ as needed. Each such implied memory element increases the number of system state variables by one. Thus, the state and output equations comparable to Equations (1.1) and (1.2) are dynamic in that f_1 and f_2 now depend on time delays, advances, derivatives and integrals. This is illustrated diagrammatically in Figure 1.2.

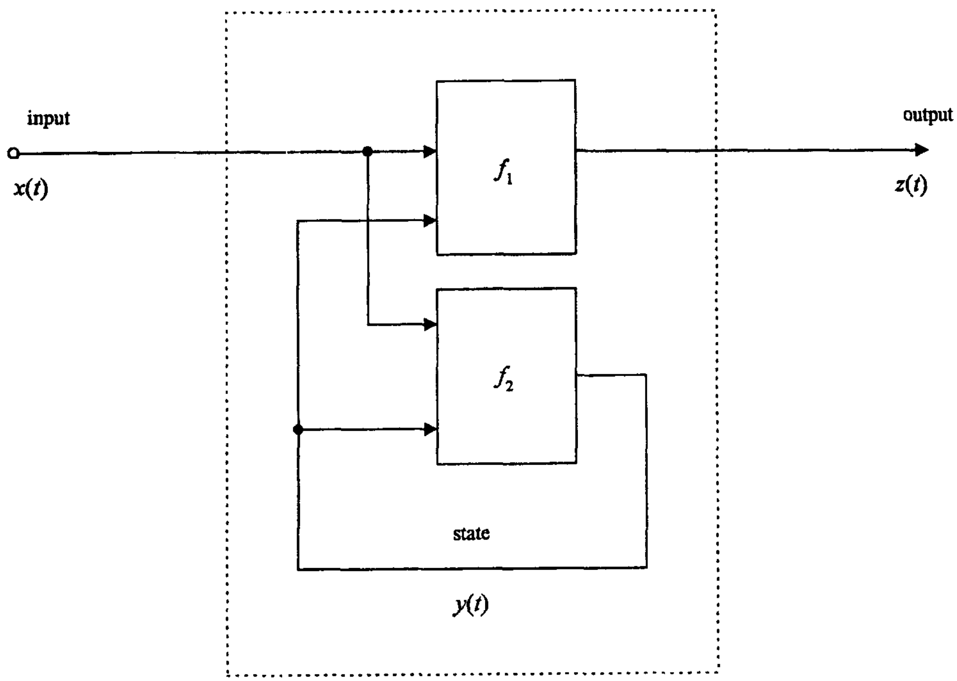


FIGURE 1.2 Feedback system.

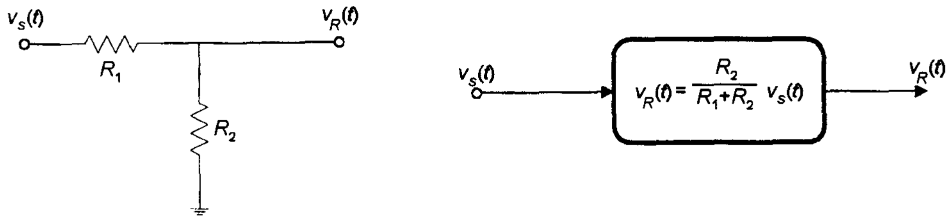


FIGURE 1.3 Electrical circuit as a SISO system, Example 1.1.

EXAMPLE 1.1

Consider the electrical resistive network shown in Figure 1.3, where the system is driven by an external voltage source $v_S(t)$. The output is taken as the voltage $v_R(t)$ across the second resistor R_2.

Since there is a Single Input and a Single Output, this system is called SISO. The input–output identification with physical variables gives

$$\begin{aligned} x(t) &= v_S(t), \\ z(t) &= v_R(t). \end{aligned} \tag{1.3}$$

Since the network is a simple voltage divider circuit, the input–output relationship is clearly not dynamic, and is therefore of order zero:

$$z(t) = \frac{R_2}{R_1 + R_2} x(t). \tag{1.4}$$

In order to find the state and output equations, it is first necessary to define the state variable. For instance, one might simply choose the state to be the output, $y(t) = z(t)$. Or, choosing the current as the state variable, i.e., $y(t) = i(t)$, the state equation is $y(t) = x(t)/(R_1 + R_2)$ and the output equation is $z(t) = R_2 y(t)$.

Clearly, the state is not unique, and it is therefore usually chosen to be intuitively meaningful to the problem at hand. ○

EXAMPLE 1.2

Consider the resistor–capacitor network shown in Figure 1.4. Since the capacitor is an energy storage element, the equations describing the system are dynamic. As in Example 1.1, let us take the input to be the source voltage $v_S(t)$ and the output as the voltage across the capacitor, $v_C(t)$. Thus, Equations (1.3) still hold. Also, elementary physics gives $RC dv_C/dt + v_C = v_S$. By defining the state to be the output, the state and output relationships corresponding to Equations (1.1) are

$$\begin{aligned} \dot{y}(t) &= \frac{1}{RC}[x(t) - y(t)], \\ z(t) &= y(t). \end{aligned} \tag{1.5}$$

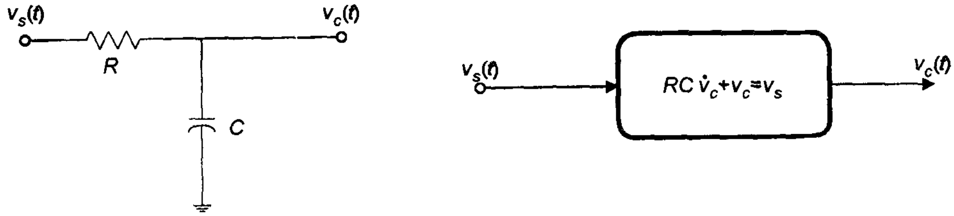


FIGURE 1.4 Electrical *RC* circuit as a first-order SISO system.

As will be customary throughout this text, dotted variables denote time derivatives. Thus,

$$\dot{y}=\frac{d}{dt}y(t), \qquad \ddot{y}=\frac{d^2}{dt^2}y(t), \qquad \ldots, \qquad y^{(n)}=\frac{d^n}{dt^n}y(t). \qquad \bigcirc$$

Electrical circuits form wonderful systems in the technical sense, since their voltage–current effects are confined to the wire and components carrying the charge. The effects of electrical and magnetic radiation on the environment can often be ignored, and all system properties are satisfied. However, we must be careful! Current traveling through a wire does affect the environment, especially at high frequencies. This is the basis of antenna operation. Accordingly, a new model would need to be made. Again, the input–output signals are based on abstractions over a certain range of operations.

One of the most popular system applications is that of control. Here we wish to cause a subsystem, which we call a *plant*, to behave in some prescribed manner. In order to do this, we design a *controller* subsystem to interpret desirable goals in the form of a *reference signal* into plant inputs. This construction, shown in Figure 1.5, is called an *open-loop* control system.

Of course, there is usually more input to the plant than just that provided by the controller. Environmental influences in the form of *noise* or a more overt signal usually cause the output to deviate from the desired response. In order to counteract this, an explicit *feedback* loop is often used so that the controller can make decisions on the basis of the reference input and the actual state of the plant. This situation, shown in Figure 1.6, is called a *feedback control* or *closed-loop control system.*

The design of feedback control systems is a major engineering activity and is a discipline in its own right. Therefore, we leave this to control engineers so that we can concentrate on the activity at hand: modeling and simulating system behavior. Actually, we

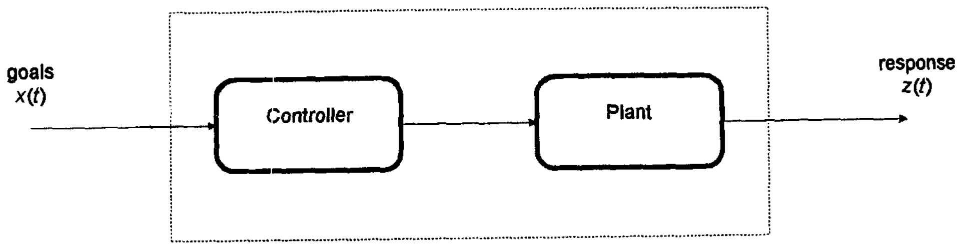


FIGURE 1.5 An open-loop control system.

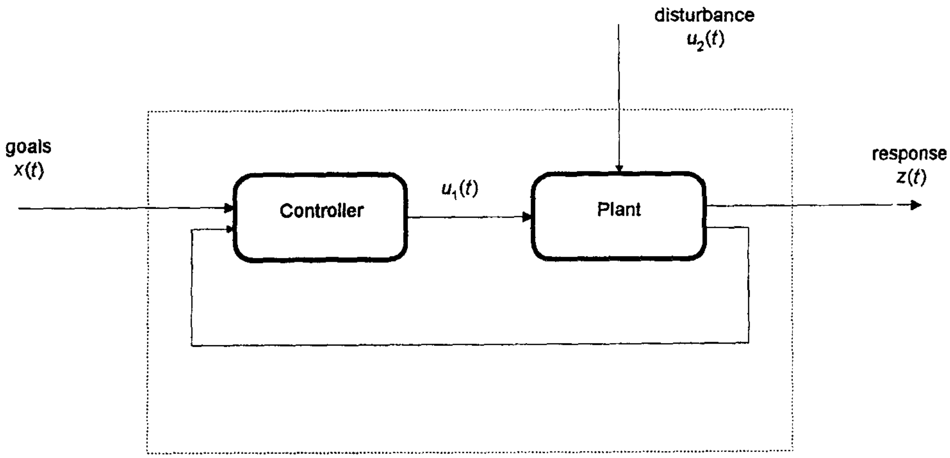


FIGURE 1.6 A closed-loop control system.

will still need to analyze control systems, but we will usually just assume that others have already designed the controllers.

The electrical circuit systems described in Examples 1.1 and 1.2 are cases of *continuous time-driven models.* Time-driven models are those in which the input is specified for all values of time. In this specific case, time t is continuous since the differential equation can be solved to give an explicit expression for the output:

$$v_C(t) = v_C(t_0) + \frac{1}{RC}\int_{t_0}^{t} v_S(\tau)e^{(\tau - t)/RC}\, d\tau, \qquad t \geqslant t_0 \tag{1.6}$$

Where $v_C(t_0)$is the initial voltage accross the capacitor at time $t = t_0$. Thus, as time "marches on", successive output values can be found by simply applying Equation (1.6).

In many systems, time actually seems to march as if to a drum; system events occur only at regular time intervals. In these so-called discrete-time-based systems, the only times of interest are $t_k = t_0 + hk$ for $k = 0, 1, \ldots$. As k takes on successive non-negative integer values, t_k begins at initial time t_0 and the system signal remains unchanged until h units later, when the next drum beat occurs. The constant length of the time interval $t_{k+1} - t_k = h$ is the step size of the sampling interval.

The input signal at the critical event times is now $x(t_k) = x(t_0 + hk)$. However, for convenience, we write this as $x(t_k) = x(k)$, in which the functional form of the function x is not the same. Even so, we consider the variables t and k as meaning "continuous time" and "discrete time", respectively. The context should remove any ambiguity.

EXAMPLE 1.3

Consider a continuous signal $x(t) = \cos(\pi t)$, which is defined only at discrete times $t_k = 3 + \frac{1}{2}k$. Clearly the interval length is $h = \frac{1}{2}$ and the initial time is

$t_0 = 3$. Also,

$$x(t_k) = \cos[\pi(3 + \tfrac{1}{2}k)]$$
$$= -\cos(\tfrac{1}{2}\pi k)$$
$$= \begin{cases} 0, & k = \text{odd}, \\ (-1)^{(k+2)/2}, & k = \text{even}. \end{cases} \qquad (1.7)$$

Thus, we write

$$x(k) = \begin{cases} 0, & k = \text{odd}, \\ (-1)^{(k+2)/2}, & k = \text{even}, \end{cases} \qquad (1.8)$$

and observe the significant differences between the discrete form of $x(k)$ given in Equation (1.8) and the original continuous form $x(t) = \cos(\pi t)$. ○

EXAMPLE 1.4

Consider a factory conveyor system in which boxes arrive at the rate of one box each 10 seconds. Each box is one of the following weights: 5, 10, or 15 kg. However, there are twice as many 5 kg boxes and 15 kg boxes as 10 kg boxes. A graphic of this system is given in Figure 1.7. How do we model and simulate this?

Solution
From the description, the weight distribution of the boxes is

w	$\Pr[W = w]$
5	0.4
10	0.2
15	0.4
	1.0

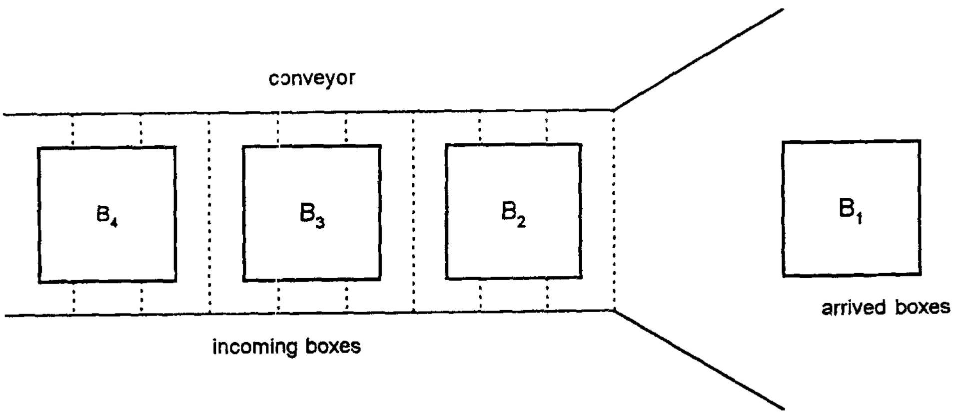


FIGURE 1.7 A deterministic conveyor system, Example 1.4.

where W is a "weight" random variable that can take on one of the three discrete values $W \in \{5, 10, 15\}$. The notation $\Pr[W = w]$ is read "the probability that the random variable W is w". The set $\{5, 10, 15\}$ is called the *sample space* of W, and is the set of all possible weights.

According to the description, these boxes arrive every 10 seconds, so $t = 10k$ gives the continuous time measured in successive k-values, assuming the initial time is zero. However, how do we describe the system output? The problem statement was rather vague on this point. Should it be the *number* of boxes that have arrived up to time t? Perhaps, but this is rather uninteresting. Figure 1.8 graphs $N(t)$ = number of boxes that have arrived up to and including time t as a function of time t.

A more interesting problem would be the weight of the boxes as they arrive. Unlike $N(t)$, the weight is a non-deterministic variable, and we can only hope to *simulate* the behavior of this variable $W(k)$ = weight of the kth event. This can be accomplished by using the RND function, which is a hypothetical random number generator that provides uniformly random distributed variates such that

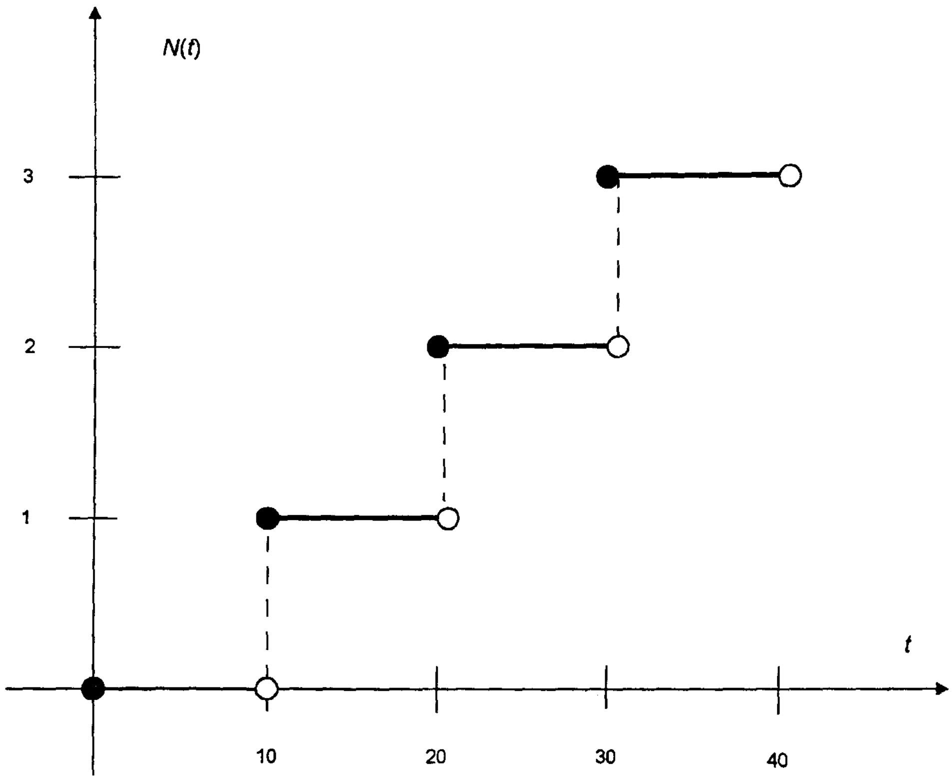


FIGURE 1.8 State $N(t)$ for constant inter-arrival times.