GUIDE TO THE DESIGN OF REAL-TIME SYSTEMS

MICHAEL F. ROTHSTEIN

Vice President, Data Processing Services
Responsive Data Processing Corporation
Mt. Kisco, New York

Wiley-Interscience

A DIVISION OF JOHN WILEY & SONS
NEW YORK • LONDON • SYDNEY • TORONTO

10 9 8 7 6 5 4 3 2 1

Library of Congress Catalogue Card Number: 74-94914

SBN 471 73972 3

Printed in the United States of America

Preface

My professional experience has been predominantly in commercial real-time systems design. From the beginning I found it difficult to obtain information that would help me to do a better job—that is, to learn from the experience of others when designing a commercial real-time system. The reasoning behind major design decisions were nonexistent. Therefore, based on my lack of knowledge, I concluded that each system design was unique and that the type of information for which I was searching was part of a learning process that could be obtained only with experience. However, after having designed several commercial real-time systems I realized that what I was doing was similar to what I had done before. In each of the systems analyses I took the same recurring steps. My work progressed in the same sequence and involved basically the same type of analysis, judgments, and decisions. As far as I have been able to determine, these "steps" in the design of a commercial real-time system are fundamentally the same regardless of the manufacturer of the various system components.

This fact became more obvious to me while I was teaching a course on Real-Time Systems Design at New York University–Management Institute. To this course came students with varying levels of experience and types of background. As a result of teaching this course, I came to the following conclusions:

1. There is a great interest in commercial real-time systems. They are the systems of the future.
2. There is a need for a definitive work on the analysis required in the design of commercial real-time systems.
3. Little of the available information discusses the "why" of system design.

It was for these reasons that I wrote this book.

The purpose of the book is to describe what is required, and why, to design a commercial real-time system. Its goal is to develop an *awareness* of the effort, the knowledge, and the problems involved in the design of a commercial real-time system. This book is only a start, an introduction, a foundation on which the reader can build. It is intended for those who understand the basis of automated data processing, such as stored program concept, systems analysis, and the relationship between programming and systems analysis. It is a basic handbook for those who wish to obtain a more thorough understanding of commercial real-time systems and their design. It has been written for those who wish to learn.

As in many technical areas, there are certain aspects of real-time system design which require specialized treatment. It was for this reason that I invited my colleague, Bob Flood, to contribute the chapter on simulation. His extensive experience with simulation of on-line systems enabled a more authoritative discussion of this very important subject.

In the course of writing this book there have been several people who have helped, encouraged, and advised. Although the list is long, I wish especially to thank Michael B. Sichel for his encouragement, guidance, and ideas during the initial planning stages of the book. Thanks also to Selo Fisch, Elizabeth Thurlow, and Yaffa Yosiefya. It is with appreciation that I also wish to thank New York University–Management Institute and The Productivity Institute of the Government of Israel. It was in their classes, as teacher, that I developed and used much of the material resulting in this book.

Michael F. Rothstein

Mt. Kisco, New York
July 1969

GUIDE TO THE DESIGN OF REAL-TIME SYSTEMS

Contents

CHAPTER 1

Introduction

"KNOWLEDGE IS POWER AND CONTROL, PROVIDED IT IS TIMELY, AMPLE AND RELEVANT. ONLY A BUSINESSMAN WHO KNOWS WHAT IS HAPPENING INSIDE HIS COMPANY REASONABLY SOON AFTER IT HAPPENS CAN ADJUST HIS MEANS TO HIS AIMS. . . ." [1]

Timely information is essential for the proper functioning of any business. One of the most efficient ways of fulfilling many of the information requirements of a business is by use of a real-time system. A real-time system is a computerized method of simultaneously receiving information from many geographically dispersed locations and processing and returning it within a meaningful period of time, seconds if need be.[2]

What does this capability mean to a business? How can it be used? A review of how management guides and perpetuates a business, viewed on a functional basis, provides the answer to these questions. From this viewpoint a business is a continuous cycle, involving four major functions: planning, implementing, reviewing and modification. The relationship of these activities is shown schematically in Figure 1.1. It starts at point A with the formulation of new plans and goals. These consist of planning for such things as new product lines and services, budget for the coming year, developing three and five year plans, expansion (or contraction) of expenditures for facilities and capital equipment and the like. Review and analysis of these plans by management results in major policy decisions, the next step in the cycle. Those plans and policies that are approved by upper management move on to the next step where they are interpreted and broken down into the many detailed plans and requirements necessary for their completion. It is

[1] Gilbert Burch and the Editors of Fortune, *The Computer Age and Its Potential for Management,* Harper & Row, 1965, p. 27, Courtesy of Fortune Magazine.

[2] This type of system is also called "fast-response," "on-line," "teleprocessing" as well as Real-Time.

at this point that first-line management and those concerned with the daily operations of the business become involved, for control of these detailed operations is their responsibility. Operating departments hire, buy, and produce in accordance with the specific instructions they have received. As each order is put into execution, reports and modifications based on problems and deviations from the planned schedule (represented by the smaller inner circle B) are channeled to first-line managers who, in turn, modify the schedule and do whatever is required to conform to and maintain an orderly "on-schedule" sequence of operations.

Trends, progress-to-date, and the overall status of daily operations make up the feedback information received by middle and upper management. It is used to review present performance and compare it with the previously made plans. Deviations cause management to take corrective action, which may include modification to previously approved intermediate and long-range objectives. Those modifications that are approved continue to implementation of management decisions, and the cycle is repeated.

The faster the cycle is completed the better the control that can be exercised by management. The more often that management is able to review the status of each plan and operation, the better (more quickly) nonprofitable and out-of-control situations can be detected and adjusted. Management makes its decisions based upon information it receives (feedback) concerning the efforts of the operating personnel and first-line management. This phase of the business cycle, daily operations and first-line management control, is shown as the shaded area in Figure 1.1. That is, daily operations personnel receive information from management (its requirements), act on it, and return to management information concerning the results it was able to achieve. Because of the expanding size and increasing costs and complexity of today's business, management needs to receive more information *more* rapidly than ever before. Unfortunately, the expansions that have created this need for more information, have the effect of delaying its return to management. Thus management is being forced to work in an increasingly demanding business environment with either less timely information or with timely data which costs more to prepare.

Computerized data processing systems have been an aid to management in attempting to solve this problem. Commercial data processing systems have evolved because they are more efficient. They do jobs faster, more accurately, at a lower cost than was previously possible. From the time of their initial acceptance, newer, cheaper, and faster devices and techniques have been developed—so that today there are many different

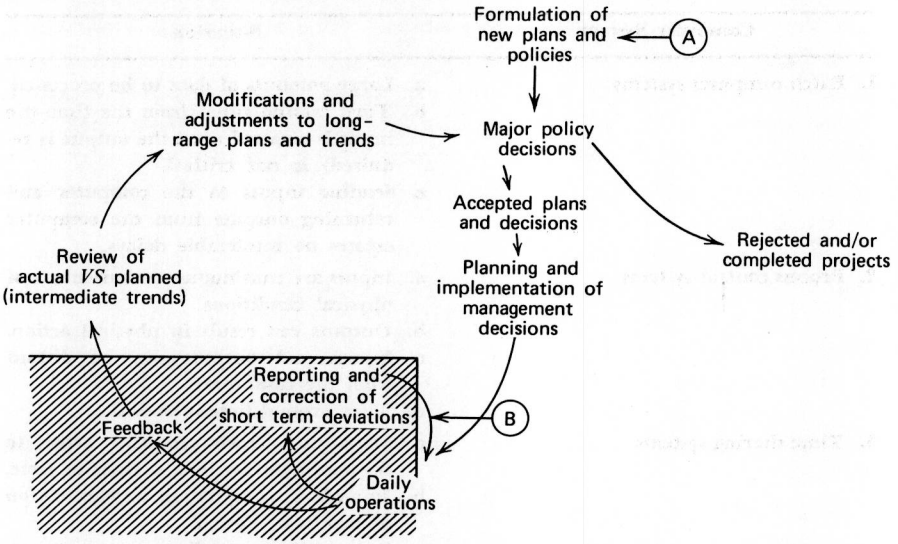

Figure 1.1. The business organization cycle.

types of systems—each depending on the data processing requirements of the user. The result is, of course, many different types of systems. These varied systems can be classified into five major categories. They are shown in Table 1.1. The column on the right shows the data processing environment for which the system is best suited.

BATCH COMPUTER SYSTEMS

Batch systems represent the more basic type of computerized data processing. Information is sent to the computer for processing, and returned in the form of reports or other types of printed documents. Before being processed by the computer the information must first be put into machine intelligible form. Batch processing's main advantage is that it manipulates information at electronic speeds while utilizing comparatively inexpensive data processing equipment.

PROCESS CONTROL SYSTEMS

Operations such as the monitoring of temperatures, the regulation of flows of liquids (oil, water, and chemicals), and electricity can be con-

TABLE 1.1

Computer System	Situation
1. Batch computer systems	a. Large amounts of data to be processed. b. Turn around time (from the time the input is createtd until the output is required) is not critical. c. Sending inputs to the computer and returning outputs from the computer creates no intolerable delays.
2. Process control systems	a. Inputs are continuous measurements of physical conditions. b. Outputs can result in physical action. c. Inputs and outputs can be distant from the computer. d. Turn around time is critical.
3. Time sharing systems	a. Many users, each requiring access to the computer for long periods of time. b. Immediate compilation and execution of programs desired. c. Comparatively small amounts of output. d. Inputs are geographically distant from the computer. e. Users can be from different companies.
4. Data transmission systems ("Batches" of inputs are transmitted for subsequent batched processing and/or the transmission of groups of processed output for subsequent printing.)	a. Large amounts of data b. Turn around time not critical c. The origin of the inputs and/or the destination of the outputs are geographically distant from the computer. d. Preferrable to maintain one set of data files.
5. Real-Time systems	a. Generally dedicated to one user. b. Turn around time is important. c. Inputs and outputs are distant from the computer.

trolled effectively by computers. Process control systems are designed to accept information from sensing devices. These sensors can be temperature gauges and other measuring devices which are continuously determining the status of one or more of the many factors that can affect the quality, or end result, of an operation. In a process control system this information is fed into a computer where it is compared with prespecified constants. These constants represent the acceptable limits within which the variable being monitored can fluctuate. Once any limit is exceeded

corrective action is triggered by the computer. It can consist of an automatic readjustment of temperature, pumping, etc. (a closed loop system) or the transferral of information to those operators who will take the steps necessary to bring the process once again under control (an open loop system). Computers are ideal for monitoring these types of operations since they are able to take readings of the many variables affecting a process more often than is possible with strictly human monitoring.

TIME SHARING SYSTEMS

In order to reduce turn around time, time sharing systems utilize remote terminals for entering programs and tests. The environment of time sharing systems can consist of the following:

1. Many users, each of which are able to write and test their programs on-line. Time sharing systems have been primarily used by scientists and engineers who require immediate answers to a one-shot type of problem rather than running and rerunning of data-manipulating programs.
2. Users working in a "conversational mode." The time sharing system analyzes each line of the program being entered for correct syntax. After each line of coding, errors are typed out by the computer system.
3. Users entering programs on-line which are stacked for subsequent compiling and testing.

In order to service a large number of users, the amount of time made available to each is controlled by the central processor. This is called time slicing. The computer controls the input locations in such a way that it appears, to the user, that he has continuous access to the computer. In reality the computer is leaving and returning to each terminal. With this technique one computer is able to service a large number of input locations.

A time sharing system is similar to a real-time system in that it is a general purpose on-line system designed to fill *one type of data processing need for many users.* A real-time system is a specially designed on-line system used to fulfill a full range of data processing requirements, most often for a single organization.

DATA TRANSMISSION SYSTEMS

These represent an intermediate stage of data processing, combining remote data transmission with batch processing. Information can be

collected at locations distant from the main computer complex. Based on a predetermined schedule a telephone-type connection is made and data is transferred into the main computer complex in one of two ways:

1. Onto tape and into the computer. Note that this method does not, at this stage, require a computer. The data tape is subsequently transferred to the computer for processing.
2. Into the computer and onto tape or disk for subsequent batch processing.

The transfer of processed data involves basically the same steps but in the opposite direction: from the computer complex to geographically dispersed locations. One additional difference is that the data could be transferred directly to tape for subsequent printing and/or punching or printed immediately and/or punched as it is being transferred.

Data transmission system can be subdivided into two major categories:

1. Off-line data transmission systems.
2. On-line data transmission systems.

Off-line data transmission systems

Under certain circumstances although there are remote data processing needs, there are neither enough transactions nor a continuing need for immediate response to warrant a "permanent" communication link with the computer. In these situations information is processed by

(a) batching inputs,
(b) establishing a communication connection (on a prescribed time schedule) for transferring the information,
(c) transferring the information onto a device in a computer room distant from the source of the input, and
(d) manually removing the transferred data to a computer for batch processing.

Once processed, the updated information can be returned to its source or routed to other locations.

The sequence of operations for an off-line system is shown schematically in Figure 1.2.

The advantages of this type of approach are twofold. Computer costs are decreased since only one central computer complex is required and one centralized set of files is maintained.

On-line data transmission systems

Once the connection between two points is made, transfer of data to (or from) the computer is possible with no manual intervention. It is

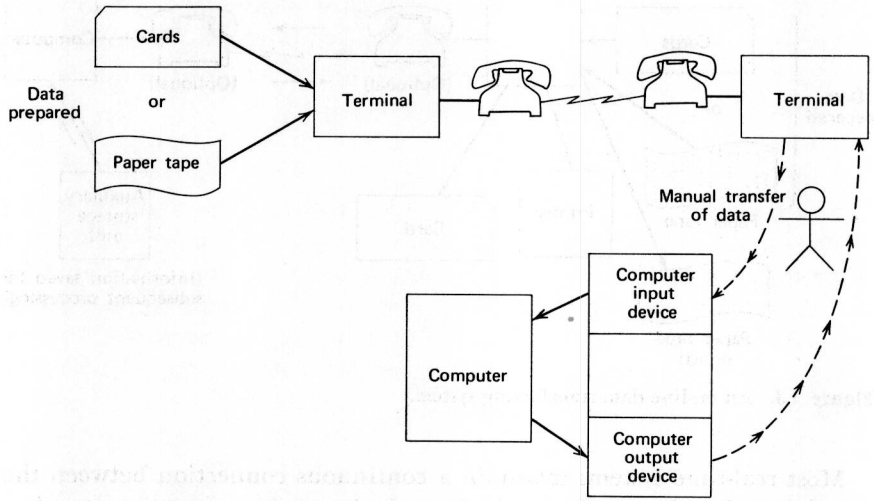

Figure 1.2. A basic data-transferring (off-line) system.

also possible to maintain a permanent communication link with the main computer complex, thus eliminating the need to manually establish a connection when information is to be transferred. This type of system is advantageous in situations where there is a slow processor, a large number of locations transferring information, and/or a large volume of data to be transferred. The transferred information is stored on an auxiliary device (tape or disk) for subsequent processing. This system does not overcome the costliness of the manual handling of data at the computer site, as the number of people required as well as the general confusion that can arise from trying to control manually the flow of such large volumes of information has not been reduced. The form for this type of system is shown schematically in Figure 1.3.

REAL-TIME SYSTEMS

Batch and data transmission systems operate in a comparatively unconstrained environment. Commercial real-time systems, however, must be able to process information in highly demanding situations (that is, fast turn around time, large number of inputs, and outputs to be processed within short intervals of time).

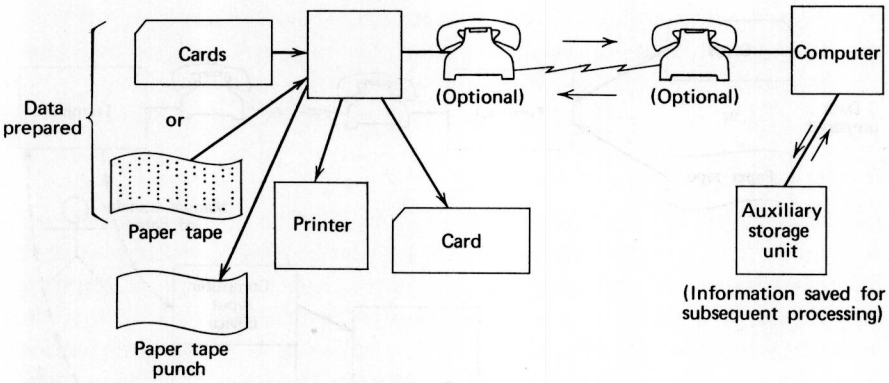

Figure 1.3. An on-line data-transferring system.

Most real-time systems maintain a continuous connection between the geographically dispersed terminals and the main computer complex. The input transaction triggers

(a) immediate processing of the input with the answer returned within seconds;
(b) updating of the central data files with only receipt of the message being returned by computer system; and
(c) a routing of the message to another terminal location, either on an immediate or whenever desired basis.

As with other types of systems, on-line systems can also vary in degree of sophistication. There is the on-line data processing system with "direct" file accessing capability. The more complete on-line data processing system includes an intricate internal data file crossreferencing capability, and is called a Management Information System.[3]

The on-line data processing system

Whenever immediate response (within an interval of time which could not be met by the previous types of systems) is needed, a system can be specified in which:

1. Information can be entered whenever desired.

[3] This type of processing of information should not be confused with Management Reporting by Exception. Reporting by exception is the automatic analysis of data for determining and reporting of *pre-specified* out of limits conditions.

2. The data is immediately automatically updated and stored on one central set of files.
3. The information required is transferred, again immediately and automatically, to the location entering the request and/or any other locations linked to the computer system.

This type of real-time system is shown schematically in Figure 1.4.

Management Information Systems

The state of the art has progressed both in types of equipment and functions that can be performed by a real-time system. Techniques have been developed and implemented for supplying all types of information to the user with no restrictions placed on his requirements, save that the answers exist in, or can be derived from, his data files. If the information is on the file, he can ask for such information as "for all customers who have ordered more than $10,000 in the last two months, tell me the number of times they have ordered item X2314, and the quantity of each order."

The procedure used for answering such types of inquiries includes the following:

1. An on-line real-time data processing system for accepting and returning of information.
2. A special file of indexes specifying the name and location of each *field* of data. These indexes and related data records can and should include most of the data requirements of the organization (a data bank or a data base).
3. A set of macros whose sequence of execution is automatically ordered based upon what has been requested.

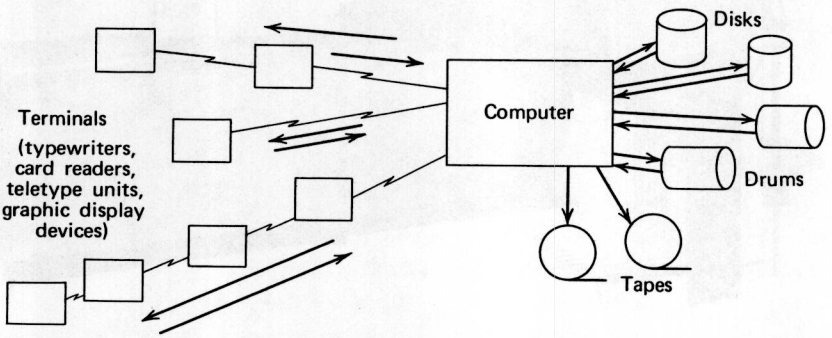

Figure 1.4. An on-line real-time data-processing system.

REAL-TIME DATA PROCESSING EQUIPMENT

The equipment available for performing real-time data processing are terminals, communication lines, computers and auxiliary storage devices.

TERMINALS

These are the devices which make available to the operating personnel the power of the computer system. The terminals are the input and output devices of the system. When such a device is physically located anywhere other than in a computer room, it is called a terminal. Inside the computer room it is called an Input/Output device. Their purpose is to

(a) facilitate the entry of data *into* the system, and
(b) receive information *from* the computer.

Some terminals are very similar to the card readers, card punches, and printers (see Figure 1.5) that one sees in the computer room of a batch type system.

Another type of terminal is the graphic display (see Figure 1.6). This unit is similar to a television set except that letters, numbers, or diagrams rather than pictures are displayed. Another terminal is the typewriter, similar to those used in any office. Although these, and others which are

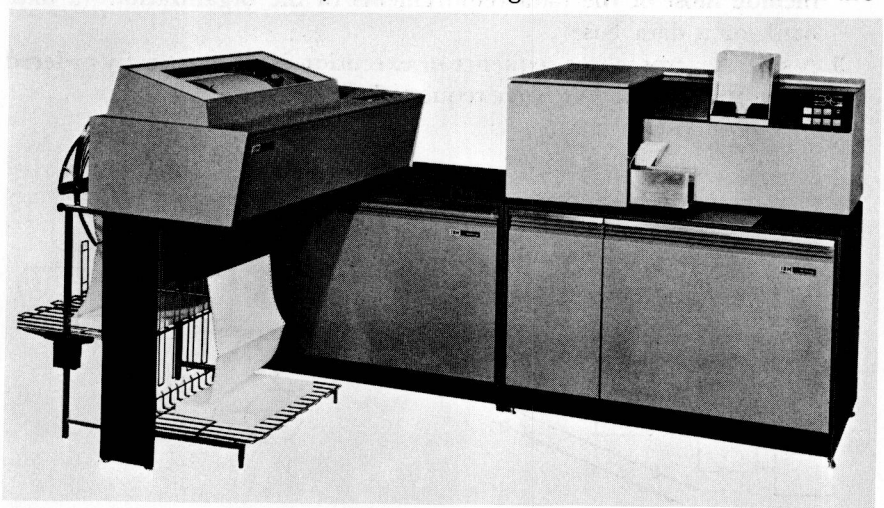

Figure 1.5. An IBM 2780 data transmission (input-output) terminal.