

## 1.1 DISTRIBUTED DATA PROCESSING

The term *distributed processing* (or *distributed computing*) is probably the most abused term in computer science of the last couple of years. It has been used to refer to such diverse systems as multiprocessor systems, distributed data processing, and computer networks. This abuse has gone on to such an extent that the term *distributed processing* has sometimes been called "a concept in search of a definition and a name." Here are some of the other terms that have been used synonymously with *distributed processing*: distributed function, distributed computers or computing, networks, multiprocessors/multicomputers, satellite processing/satellite computers, backend processing, dedicated/special-purpose computers, time-shared systems, and functionally modular systems.

Obviously, some degree of distributed processing goes on in any computer system, even on single-processor computers. Starting with the second-generation computers, the central processing unit (CPU) and input/output (I/O) functions have been separated and overlapped. This separation and overlap can be considered as one form of distributed processing. However, it should be quite clear that what we would like to refer to as distributed processing, or distributed computing, has nothing to do with this form of distribution of functions in a single-processor computer system.

A term that has caused so much confusion is obviously quite difficult to define precisely. There have been numerous attempts to define what distributed processing is, and almost every researcher has come up with a definition. In this book we define distributed processing in such a way that it leads to a definition of what

a distributed database system is. The working definition we use for a *distributed computing system* states that it is a number of autonomous processing elements (not necessarily homogeneous) that are interconnected by a computer network and that cooperate in performing their assigned tasks. The “processing element” referred to in this definition is a computing device that can execute a program on its own.

One fundamental question that needs to be asked is: What is being distributed? One of the things that might be distributed is the *processing logic*. In fact, the definition of a distributed computing system given above implicitly assumes that the processing logic or processing elements are distributed. Another possible distribution is according to *function*. Various functions of a computer system could be delegated to various pieces of hardware or software. A third possible mode of distribution is according to *data*. Data used by a number of applications may be distributed to a number of processing sites. Finally, *control* can be distributed. The control of the execution of various tasks might be distributed instead of being performed by one computer system. From the viewpoint of distributed database systems, these modes of distribution are all necessary and important. In the following sections we talk about these in more detail.