# Principles of Computer Systems and Network Management

Dinesh Chandra Verma

# Principles of Computer Systems and Network Management

Springer

Dinesh Chandra Verma
IBM T.J. Watson Research Center
Yorktown Heights
NY 10598
USA
dverma@us.ibm.com

ISBN 978-0-387-89008-1 e-ISBN 978-0-387-89009-8
DOI 10.1007/978-0-387-89009-8
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2009928696

*To Paridhi*
*who stood by me through the entire process*
*of writing this book.*

# Preface

As computer systems and networks have evolved and grown more complex, the role of the IT department in most companies has transformed primarily to ensuring that they continue to operate without disruption. IT spending, as reported by a variety of studies, shows the trend that most of the expenses associated with IT are related to the task of operating and managing installed computer systems and applications. Furthermore, the growth in that expense category is outstripping the expense associated with developing new applications. As a consequence, there is a pressing need in the companies and organizations to find qualified people who can manage the installed base of computer systems and networks. This marks a significant shift from the previous trend in companies where the bulk of the IT department expenses were targeted on development of new computer applications.

The shift from developing new applications to managing existing systems is a natural consequence of the maturity of IT industry. Computers are now ubiquitous in every walk of life, and the number of installed successful applications grows steadily over the time. Each installed successful application in a company lasts for a long duration. Consequently, the number of installed applications is much larger than the number of projects focused on developing new applications. While there always will be new applications and systems being developed within companies, the predominance of managing and operating existing applications is likely to continue. A natural consequence of this is that the IT marketplace will continue to shift toward a technical population skilled at operating and managing existing computer applications and systems, as opposed to a technical population skilled at developing new computer applications and systems.

The education and training provided in our computer science courses, however, has not kept pace with the shift in the demand for computer scientists. While most computer science programs have a rich set of courses teaching the students the skills required to develop new systems, there is a paucity of courses that train them to manage and operate installed computer systems, both hardware and software. As a result, there is a mismatch between the demands of the IT marketplace and the supply of the technical talent from our universities. Fortunately, several universities have noticed this shift in the demand and have

started to offer courses in systems and network management. This book is intended to provide the supporting textbook to facilitate the teaching of such courses.

This book tries to define the fundamental principles that every student of systems management ought to be aware of. The algorithms, architectures, and design techniques used for different aspects of systems management are presented in an abstract manner. The technical knowledge required for different aspects of systems management is quite large, spanning mathematical domains such as queuing theory, time-series analysis, graph theory as well as programming domains such as configuration management. The book focuses on showing how the concepts from those domains are applied, rather than on the details of the specific domain – which can be found in many excellent related textbooks.

The abstract principles based approach requires decoupling systems management from the survey of the different management tools that exist currently – in both the open-source community and the commercial product offerings. This is not a book that provides a survey of existing management tools, nor tells the reader how to use such tools. This is a book that provides a survey of the techniques that are used to build those tools. However, this books enables the reader to compare the relative strengths and weaknesses of the different techniques used in these tools.

Apart from the students and teachers who are involved in learning or teaching a course on computer management, the book can also be used as a reference for understanding the basics of systems management. It would be useful for IT practitioners in companies developing systems or network management products. Such practitioners need to embody many of these principles in their products and offerings. Finally, the book should be a useful companion to practitioners involved in software development. Such developers are under an increasing pressure to deliver software that is more usable and manageable.

Structurally, the book is divided into 10 chapters. The first chapter provides an introduction and history of the field of systems management, along with an overview of three specific type of computing environments that are used as examples in subsequent chapters. It also provides an overview of the four stages in the life cycle of a computer system: planning, implementation, operations, and upgrade.

The second chapter discusses the principles involved in planning and implementation stage, i.e., how to design computer systems that can satisfy a variety of requirements ranging from performance and reliability requirements to power management requirements.

The third chapter provides an overview of the tasks required for operations management of computer systems. Operational computer systems management can be defined as consisting of two basic functions, discovery and monitoring, coupled with five analysis functions – fault management, configuration management, accounting management, performance management and security management. Chapters 4–9 deal with each of these functions respectively.

Chapter 4 discusses the subject of discovery. Discovery is the process of finding out the inventory of different components that exist in a computing environment, and the different ways in which the inventory of discovered assets can be maintained.

Chapter 5 discusses the different ways to monitor management data in computer systems, and to store them for analysis. Approaches to deal with scalability of data as well as techniques to deal with errors and data cleansing operations are discussed.

Chapters 6 and 7 discuss algorithms and approaches for fault management and configuration management, respectively.

Chapter 8 discusses the task of performance management and accounting management, including a discussion of capacity planning for computer systems.

Chapter 9 discusses the subject of security management, including a discussion of operational aspects such as security policy specification.

Chapter 10 discusses topics that are related to the principles of systems management, including subjects such as IT Service Management, ITIL, and helpdesk systems.

Overall, this book provides a holistic approach covering all aspects of systems management, and will prepare a student of computer science to take on a career dealing with systems operation and management in the new IT industry.

# Contents

# Chapter 1
# Introduction

## 1.1 Introduction

Computer systems and networks are indispensable components of the modern industry. Regardless of the size of the business, be it a small shop operated by a single family or a large international conglomerate, the profitability of any business operation has a critical dependency upon proper functioning of the computer systems supporting that business. Even on the consumer side, households having multiple computers are the norm in most industrialized nations. For many of our daily needs, from browsing the news to printing out an assignment, we depend upon computers and network connectivity. Computers enable us to enjoy the fruits of modern civilization in all walks of life. However, the benefits provided by computers are attainable only when they are operating flawlessly.

Systems management is the branch of computer science that deals with the techniques required to ensure that computer systems operate flawlessly. The definition of a flaw in the computer system is any kind of problem which prevents users from attaining the benefits that the computer system can offer. A flawless operation is one in which the user is able to obtain such benefits.

A flawless operation does not mean that the computer system needs to operate at its optimal configuration, but simply that the system is operating in a manner that the user is satisfied. A computer link that is configured to operate at only a fraction of the capacity may be operating flawlessly if the total bandwidth demand on the link is an even smaller fraction. Flawless operation can be attained by properly designing and operating any computer system.

Flawless operations cannot be expected to happen automatically in a computer system without human intervention. Human beings need to be involved in a variety of roles in order to ensure that the computer system is operating well. As a matter of fact, the involvement of a human being needs to happen even before the computer systems may have been obtained or procured. In order for most computer systems to operate as desired, a human being must plan the system so that it can meet the anticipated needs when it is operational.

The exact functions that are performed by the human operator to ensure flawless operation of the system depend upon the nature of the computing

D.C. Verma, *Principles of Computer Systems and Network Management*,
DOI 10.1007/978-0-387-89009-8_1, © Springer Science+Business Media, LLC 2009

environment the operator is managing. Nevertheless, there are some common responsibilities that need to be conducted by the operators. There are a set of common principles, techniques, and approaches that can be used to deal with most of the challenges that may arise in ensuring a flawless operation of the network. The goal of this book is to provide a compilation of those common principles, techniques, and approaches.

An understanding of these common principles would enable people in charge of ensuring flawless operation of computer systems and networks to operate these systems more efficiently. On many occasions, the task of flawless operation can be aided by means of software systems – or more precisely systems management software. An understanding of the common principles would aid in development of better management software.

In this book, we will use the term computer system to refer to a collection of electronic devices that are collectively used to provide a set of useful services to human users. The computer system in this sense is a broad term which includes a collection of personal computers, workstations, mainframes, network routers, firewalls, etc. In technical publications, the computer system defined in this context is also referred to by other terms such as IT system, IT infrastructure, and distributed systems.

In order to ensure a flawless operation, we need to look at the entire life cycle of a computer system – beginning from when the plans for the system are conceived to when the system is taken off from active service. The life cycle of the computer system is described in the next section, with a discussion of the different management activities that need to be performed during each of the different stages of the life cycle. The next sections of this chapter describe some common computer systems in different environments and discuss the system management functions that need to be done in those environments throughout the life cycle. Finally, this chapter provides a brief overview of how systems management has evolved over the years.

## 1.2 Computer System Life Cycle

Any computer system can be viewed as progressing through a four-stage life cycle as shown in Fig. 1.1. The four stages are planning, implementation, operation, and upgrade/termination. Each of these stages is an important aspect in the life cycle of the computer systems and has its own unique requirements for the flawless operation of the computer systems.

The life cycle of the computer system in any business begins as soon as the owners of the business decide that they need a computer system to take care of some aspects of their business. Prior to the beginning of the life cycle, some key business decisions may be made such as a cap on the total amount of money to be spent on the implementation and rollout of the system, the upper limit on monthly operating expenses, the function to be performed by the computer system, the selection of the entity doing the implementation or managing

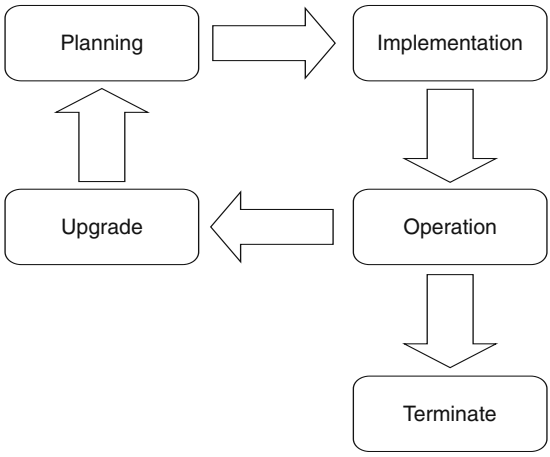


**Fig. 1.1** Life cycle of computer systems

operation of the system – either an in-house entity or an external vendor, and expected revenues or cost savings or another measure of business utility derived from the computer system. When purchasing a personal computer for the home, similar types of decisions need to be taken.

The business aspects of computer systems' life cycle are outside the scope of this book, which focus primarily on the technology required to ensure the computer system's flawless operation once the business decision to acquire the computer system has been made. Nevertheless, many decisions made during the business planning stage, e.g., the budget limits, have a significant impact on the smooth operation of the system.

Once the business decision to acquire a computer system is made, the life cycle begins with the *planning phase*. The planning phase requires development of detailed plans and the design for the computer system. During the planning phase, decisions regarding how the computer system would look like are made. These include decisions about the number and types of physical machines needed to implement the systems, the network connectivity that needs to be obtained, the applications that need to be procured and installed on these systems, the configuration to put in place for the applications and machines, and the type of systems management infrastructure that ought to be put in place to maximize the probability of a flawless operation.

After the plans and designs are completed and approved, the *implementation phase* of the life cycle begins. In the implementation phase, the planning decisions are put into practice. During the implementation phase, the different types of physical machines are obtained, the applications installed, customized applications developed, and testing undertaken to validate that the computer system will perform properly once it is operational and that the implemented system conforms to the specifications that are put forth in the planning phase.

Planning and implementation for systems management purposes are very different than planning and implementation for development of new software or applications. Systems management is viewing these activities from the perspective of Information Technology department of a corporation. Such departments are usually responsible for acquiring, installing, and managing computer systems. This role is quite distinct from that of the development organizations within the same company that are responsible for developing software applications. Planning and implementation, from a systems management perspective, are determining the set of existing software and hardware components to acquire and assemble to create a computer system. In this sense, these planning and implementation are radically different from software engineering and development practices needed for developing new application software.

Once implemented, the system enters the *operation phase* of the life cycle. In the operation phase, the system is live and performing its functions. The bulk of system management functions is performed during this phase. A management system is required during the operation phase to monitor the performance and health of the computer system, to diagnose the cause of any failures that happen within the system, to ensure the security of the system, and to provide any information needed for accounting and bookkeeping purposes.

Each system has a finite operational lifetime. After the expiration of this lifetime, the system would enter either the *upgrade* or *terminate* phase. In the upgrade or terminate phase, an operational system needs to be modified or changed. An upgrade may be required for many reasons, e.g., the performance of the operational system may be below the desired level, new technology may allow for a cheaper mode of operation, or merger with another company may require changes to the operational computer systems. An example of such a change occurring due to mergers or acquisitions would be the switching of the domain names of the computers. The upgrade phase requires activities similar to the planning and implementation phase, except with the additional twist that the planning and implementations have to take cognizance of the fact that they need to be applied to an existing system, as opposed to being developed for a new installation. In the trade, it is common to refer to new installations as green-field while upgrades are referred to as brown-field.

When a system needs to be terminated, special considerations need to be taken into account that appropriate information and data have been transferred from the existing system, and that a smooth transition be made from the old system to the new system that will replace it.

The operational stage in the life cycle is the most dominant one from the systems management perspective. Some practitioners in the field take the position that the operational stage is the only part where systems management is truly needed. Nevertheless, a student of the subject should study the principles and concepts that are useful in other stages of the life cycle, since decisions made at those stage have a significant impact on how the system works during the operational phase.

Having looked at the life cycle of the computer systems, let us now look at some typical computer systems that need to be managed, how the life cycle applies to those systems, and what type of management issues arise in each type of computer system. The systems presented below are simplified representations of real computer systems, and their primary purpose is to illustrate the life cycle and management needs during different stages of the life cycle. In actual businesses, the systems deployed in real life tend to be much more complex and irregular than the simplified examples we will be discussing. Computer systems we discuss include a shared hosting data center, a network, and an enterprise computing infrastructure.

## 1.3 Shared Hosting Data Center (SHDC)

Many companies perform the task of hosting applications and data systems for other companies. Web-hosting companies are ubiquitous on the Internet. Many companies in industries other than information technology frequently choose to hire hosting companies to run, manage, and operate their data centers and applications. As the first example of a computer system, let us examine the infrastructure of a shared hosting data center provider. The collection of the different hardware and software components that are used to provide the hosting data center is an example of a computer system. The techniques to manage such a computer system are the subject of this book.

A typical computing environment one may encounter in a shared hosting data center is shown in Fig. 1.2. The shared hosting data center is operated by a data center operator. The data center operator supports multiple customers who are hosting their servers and applications at the center. The data center connects to the public Internet through an access router. Typically the access link to the Internet would be replicated to provide resiliency. The data center

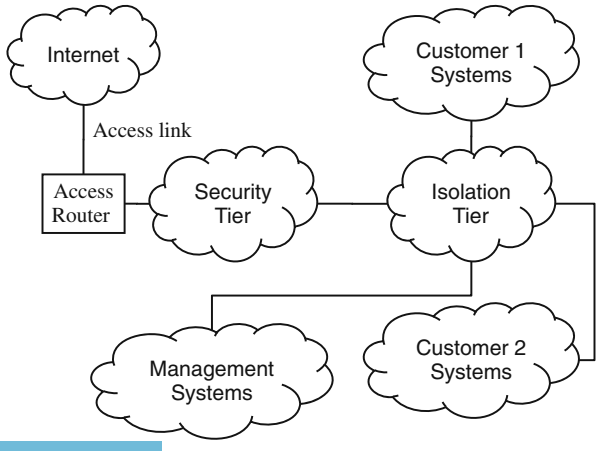


**Fig 1.2** A shared hosting data center

would have service agreements with one or more telecommunication companies to provide access to the data center with assurances about the bandwidth and resiliency of the provided access links.

Immediately following the access router, a shared hosting data center would have a security tier consisting of devices such as firewalls, intrusion detection systems, and honeypots. The devices in this tier are designed to protect against any security threats that one may experience from the Internet. These security threats may be malicious person trying to gain unauthorized access to the computers of the data center, someone trying to launch a denial of service attack, someone trying to inject a virus into the computers of the data center, or any number of other possible threats.

After the security tier, a shared hosting data center would have a tier of devices designed to provide isolation among the different customers of the shared data. Such isolation may be provided using a variety of techniques, including firewalls, virtual private networks (VPN), and packet filters. The isolation tier prevents one customer from accessing the machines belonging to another customer.

The next tier consists of segments dedicated to individual customers and for internal operations of the data center. One of these computer segments would be the common management system that the shared hosting data center operator uses to manage the entire infrastructure. Internal to each customer segments there may be several tiers of computing infrastructure embedded. For example, if a customer is hosting a web site it is common to have a three-tier structure consisting of caching proxies, web application servers, and database systems. Each of these tiers may be preceded by load balancers that distribute incoming workload among different machines.

Let us now look at the different stages of the life cycle of the computer system in an SHDC. During the planning phase one needs to decide the number and capacity of Internet connections required for the data center. The SHDC operator may want to select two or three different telecommunication companies to get greater resiliency. It needs to decide whether or not to use a wide area load-balancing solution to spread traffic around the three companies providing the Internet connection. It may also opt to use one of the telecom companies as the primary connection provider and others as backup connection providers to deal with failures. It also needs to decide the type and number of security devices to install in order to implement the desired security and isolation tier functions. It will need to select the access routers and number of servers in each customer segment to meet the combined traffic load of the customers.

During the implementation phase the primary task is of acquiring and installing the machines that make up each tier of the data center. Proper attention has to be paid to physical separation between the customer segments and for efficient power management of the systems.

During the operation stage, administrators need to make sure that the Internet connection is up and the servers assigned to each customer are live and operational. These administrators would use applications running in the

management systems tier to monitor and administer the data center. Helpdesk support may be required to address any problems a customer may have. Fault detection systems may be needed to continuously monitor the infrastructure for any down or malfunctioning machines. In the event of a fault, steps need to be taken to replace or repair the malfunctioning or failed system with minimal disruption of service. Depending on the terms of the contract signed by the customer the operator may need to provide periodic performance reports demonstrating that they have met the desired performance and availability requirements of the hosted application.

If the data center is being terminated, the customer data and applications may need to be transferred to a new data center. Care must be taken to close out the accounts and clean out any customer data from the servers which are being decommissioned.

Thus, each stage of the life cycle has its own set of requirements and a common set of principles that can be applied to ensure the flawless operation of the computer systems of an SHDC operator.

## 1.4 Large Enterprise

As another example, let us consider the computing systems that one can encounter in a typical enterprise. An enterprise is a company, usually a commercial business, which has premises at several geographically dispersed sites. The IT environment of a large enterprise would be similar to the IT environment of a large university or a government organization, and they can also be considered as enterprises. A typical enterprise's IT environment consists of three major components: computer servers, client machines (including personal computers, laptops, and personal digital assistants), and an enterprise intranet. The typical structure of the enterprise system would be as shown in Fig. 1.3.

The enterprise consists of several sites that are connected together by an intranet. The intranet is typically a private network owned and operated by the enterprise IT department or a subcontracted company. One or more of these sites may be connected to the public Internet. Some of these sites serve as data centers, i.e., they contain several servers which run applications that are required to be operational and available to any user within the enterprise. Some of the servers at a data center may be accessible to users within the enterprise as well as users outside the enterprise. Other servers may be accessible only to users within the enterprise. Firewalls positioned between the Internet, the enterprise intranet, and the data centers ensure the separation and access control restrictions among the external and internal users.

In addition to the data centers, the enterprise has many sites where the primary users are employees with personal computers or laptops. These users access the servers and the Internet using their personal computers. Furthermore, some of the employees may be using personal digital assistants and

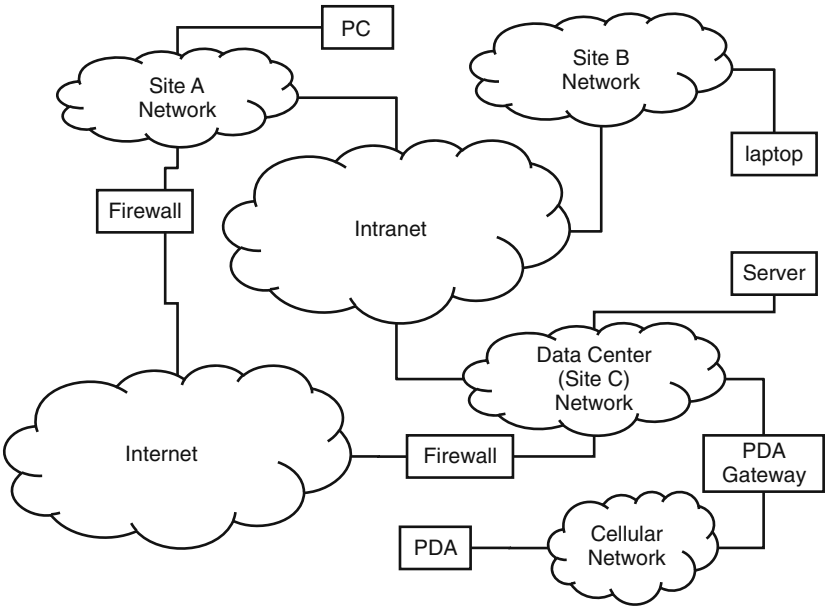


**Fig. 1.3** Structure of an enterprise IT environment

accessing some of the servers in the enterprise through a cellular network or a wireless network. The enterprise needs to run the servers which support access to those applications through the wireless networks.

Thus, the system administrator in the enterprise environment would need to worry about the applications running on personal computers, laptop machines, and servers. Furthermore, the enterprise needs to support the connectivity of the various types of devices by supporting the local intranet, access to the public Internet and in some cases access to a wireless or cellular network for some applications such as e-mail access through a personal digital assistant.

Systems management in an enterprise is a complex affair with many different types of computer systems that have their own specific requirements. As a result, many enterprises have different organizations in their information technology department in charge of managing different portions of their infrastructure. Typically, the personal computers and laptops are going to be managed by a PC support division, the servers managed by a server support division, and the networks by a network support division. Some of the more complex applications may have their own support divisions, e.g., one may find separate support people for IP telephony if it is an implemented application in the enterprise. The task of the systems management in each of the support division is to ensure that the problems arising in their portion of the IT infrastructure are resolved quickly, and that the applications in the enterprise are running properly with the desired level of performance.

Many enterprises have a large computer infrastructure, and this infrastructure grows continuously due to addition of new branches, acquisition of new companies, and similar additions to existing infrastructure. Let us consider the life cycle of an enterprise system when a new branch of the enterprise is opened. During the planning phase of the life cycle, the planners need to determine if the branch site will host servers of any type, or will it be primarily client machines with network access to the intranet and Internet. They need to decide whether the site ought to be connected to the public Internet or not, and the amount of bandwidth needed on the links to the intranet and the Internet. They also need to determine the structure of the networking that will be required internally within the site. During the implementation phase, the requisite machines and applications need to be installed at the site and tested for correct operation.

During the operation phase, the system managers need to address problems that are reported by the employees of the enterprise. The problems may be diagnosed automatically by monitoring systems deployed by the system managers, or may be reported by employees reporting a problem or their inability to access a given server or application, forgetting their passwords or providing access to new employees who have joined the organization. Other functions performed by the system managers during the operation phase may include upgrading software versions of the installed applications, dealing with security incidents such as the attack of a new computer virus, reporting performance and problem resolution reports, as well as assisting other employees with any issues in using specific applications.

The proper operation of the applications in the enterprise requires coordination among all the different support departments. If a user is not able to access an application, the problem may lie with the configuration of his/her laptop computer, an issue in the intranet, or a problem with the server side of the application. The different support teams need to be aware of the outages, scheduled downtimes, and status of other parts of the enterprise if they need to be able to provide a smooth operation of the computer enterprise.

## 1.5 Network Service Provider

In this section, we describe the structure of computer systems that can be found in a company which supports only one type of service to its customers, providing them with connectivity to other networks, either providing them with network connectivity to other sites on their private intranet or providing access to the public Internet. Most network service providers provide a menu of connectivity services much richer than the simple connectivity model we are describing, but the example will suffice to illustrate the issues encountered in the field of systems management.

The structure of the network service provider network can be seen in Fig. 1.4. The provider would have a core network consisting of several routers. These

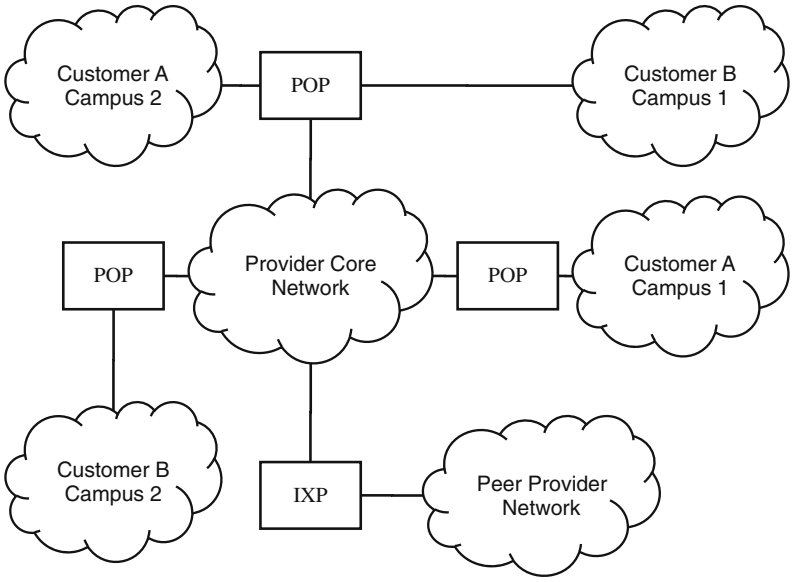


**Fig 1.4** Structure of a service provider's network

routers will be operational on a set of links which may be deployed using technologies such as SONET or high-speed ATM links.

The network service provider would have multiple points of presence (POPs) at various cities. POPs are sites used to access the provider network by its customers. Each POP consists of several routers which provide access to one or more customers and connect the POP to the provider core network. Customers may connect to POPs using leased lines. For some customers, the ISP may place an access router on the customer's premises, and connect it to the POP using a metropolitan area network or a leased line.

In addition to the POPs, the network provider needs to partner with other network providers in order to connect to the public Internet. The public Internet is nothing but the collection of the networks of all the network providers in the world. These peering points are known variously as Internet exchange points or (IXPs), NAP (network access point), MAE (metropolitan area exchange), or FIX (federal Internet exchange). An IXP can connect a regional network provider to a national provider or act as conduit among networks belonging to several large providers. Different network providers would have peering agreements among themselves as to which traffic they would accept from other ISPs at an exchange point. Very large service providers also have private peering arrangements with each other.

The points of presence and exchange points of a provider are interconnected by its core network. The ISP's core network of the ISP consists of several routers connected by means of high-bandwidth circuits that may be owned by

the provider or leased from other bandwidth vendors. The core network would have a network segment which will be used exclusively for managing the operation of the network.

Let us now consider the decisions that need to be made during the different stages of the network life cycle. The decisions taken by the provider can significantly impact the operational nature of the network. The provider needs to select the cities to have a point of presence in and the locations to establish exchange points with other peer networks. It also needs to determine which of the other peer networks it ought to connect to and choose the right bandwidth for the various links, both the ones connecting it to the customers and peer network providers, as well as the ones that make its core network.

During the implementation phase, the network service provider will need to obtain the physical premises and the equipment it needs to provide its services at various locations, and test their proper operation. The provider also needs to validate that the implementation has been done in accordance with the provided specifications, and that the system is performing to the specifications that are desired.

During the operation phase, the provider needs to monitor its network for any performance or availability problems, track its usage for planning any capacity updates, and resolve any problems raised by the customers regarding the operation of the network. Customers may face problems related to connectivity as well as those related to the performance or usability of the network. Another key function that needs to be performed during operation is the collection of appropriate information to bill customers for their network usage.

When the network needs to be upgraded, the provider needs to take into account the issues related to network planning and implementation, as well as determine how best to upgrade the network without disrupting the operations and services provided to the current users.

All the above functions need to be performed for one of the many services that a provider may offer. Commercial networking service providers offer much more than basic Internet Protocol connectivity to their customers. The networking system of a commercial provider usually includes many different technologies, e.g., they may use SONET over optical fibers to obtain point-to-point connectivity in part of the network, use ATM switches on top of SONET links to create a physical network, then use IP to create another network on top of ATM. In another portion of the network, the provider may provide a DSL service to its residential customers and provide IP on top of a DSL network. While it is beyond the scope of the book to explain the alphabet soup of networking protocols like SONET, ATM, IP, and DSL, each of these layers of the networking stack has its own idiosyncrasies which require specialized technical expertise. As a result, many commercial networking service providers would operate separate network management system for each such technology. Additionally, for each application that runs on the network stack, e.g., an Internet Telephony service may need its own management infrastructure. In order to work properly, these management infrastructures need to be linked

together. Resolution of problems in an application often requires frequent interactions between the different management systems of the provider.

As can be seen from a brief examination of three typical computer environments, computer systems management is a complex and challenging task with many different aspects. Although there are many systems management problems that are unique to the different environments, there are many common requirements and needs that cut across a variety of computer systems.

Studying those common requirements and the approaches that can be used to address the problems in different contexts is the scope of the book. In subsequent chapters of the book, we will look at the systems management challenges at each stage of the life cycle of a computer system, and explore the principles that can be used to address the challenges during that stage of the life cycle.

## 1.6 History of Systems Management

As we look into the different principles underlying systems and network management, it is useful to get a historical perspective of how the field of systems management has evolved over the years. Systems management has morphed itself in various ways as new technologies and architectures have emerged in computer systems over the years.

In the 1960s, the dominant form of computer systems in most companies were large servers and mainframes. Each server supported many users by means of different terminals. These servers permitted management via dedicated operator consoles [1] which was a special terminal usually set in a privileged area. Each manufacturer provided its own version of a management console. The management console typically consisted of a text-oriented display of current status of the system parameters and allowed changes in the configuration files of the mainframe. There was little interoperability among the management systems provided by different manufacturers. Each server had a different management console which would not be connected together.

In the 1980s, developments in networking technology and open standards such as the Internet Protocol enabled machines from different manufacturers to communicate with each other. This led to the development of large networks of computers and open standards were needed to manage the networks. Simultaneously, the development of personal computers and the advent of client–server model for computing led to a rethinking of the ways in which systems management ought to be done. The need for standards drove to specifications such as SNMP (Simple Network Management Protocol) and MIB (management information base) format from the IETF [2], as well as the Common Management Information Protocol [3] from ITU. SNMP subsequently eclipsed CMIP as the preferred management scheme for computer networks.

Both SNMP and CMIP mimicked the dominant client–server computing paradigm of the 1980s and had an agent–manager model. In the client – server paradigm, a server application is used to provide services to many client applications. In the context of management, a *manager* provided management function to many different *agents*. An agent in each managed device provided access to the information required for management, and centralized manager software would display the information to an operator. The agent–manager architecture is the predominant management architecture in most deployed systems today.

Another key innovation of the SNMP management model was the standardization of management information through standard formats (the MIBs). Prior to the development of the standards, management information used to come in many different formats and it was difficult to compare the information from two different devices. Representing the management information in a common standard format reduced the complexity of management significantly.

Systems management technology has always adopted the dominant computing paradigms at any given time. In the early 1990s, CORBA (Common Object Request Broker Architecture) [4] gained popularity as an approach to implement distributed systems. Several system management products were developed using an agent–manager model where CORBA was used as the communications protocol instead of SNMP. While SNMP was used primarily for network devices and MIB specifications made it clumsy to represent structured information, CORBA objects could represent arbitrarily complex management information. SNMP also had a relatively weak security mechanism. A dominant telecommunications network in the 1990s, TINA (Telecommunications Information Networking Architecture) [5] based its management specifications on CORBA. CORBA-based management systems are common for managing servers and systems within an enterprise.

As the Internet based on a new protocol HTTP (HyperText Transfer Protocol) grew in prominence in the second half of the 1990s, it exposed some of the deficiencies with CORBA as management architecture. Enterprises tended to have several tiers of firewalls to guard against security threats from outside and within their networks, and CORBA protocols needed to get holes in the firewalls to communicate across each other. CORBA provided a mechanism for the manager and agents to communicate with each other, but did not provide a standard for management information, i.e., CORBA as a management protocol did not have an analogue of SNMP MIBs, and each management product vendor defined its own specifications for the same.

The Desktop Management Task Force (DMTF) [6] was formed in the 1990s to develop a common information model for general systems management – an analogue of MIBs for network management. The leading standard from that body, CIM (common information model) used an object-oriented approach to represent systems management information in a standard way. The other standard from DMTF, Web-Based Enterprise Management (WBEM), provides a transport protocol like SNMP except based on top of web-based

management. The acceptance of CIM as well as WBEM has been relatively spotty in the field, even though most leading computer companies have been participating in the standard definition process.

In addition to the developments in protocols and management information representations, the field of systems management has been quick to adopt practices and benefits from emerging technologies in the field of computing. As relational databases [7] grew in prominence, the nature of the manager in most implementations took the form of a database where management information was stored, and management functions were essentially operating and manipulating that database. Subsequently, the popularity of the Internet browser has led to many management consoles being redefined to be browser-based, where the user interface is a browser, and all management operations are done by means of scripts run at a web server. As the concept of usability and human factors have taken hold of the industry, new initiatives such as policy-based management and autonomic computing have been attempted with the goal of improving the usability of management.

Some attempts to adopt new technology for systems management have not proven very successful, e.g., initiatives to use mobile codes and active networking for management purposes, economic theory-based management paradigms, peer-to-peer management paradigms, and delegation based management approaches have not had any significant traction in the marketplace till the time of the writing of this book. A survey of such techniques can be found in [8]. Research is continuing on exploiting other emerging technology, e.g., web services, system virtualization, and business process modeling, for the purpose of systems management.

## 1.7 Summary

This chapter provides an introduction to the life cycle of a computer process and outlines the management needs of the computer system in each of the life cycle stages. It introduces three idealized environments for the computer systems, the shared hosting data center, the enterprise system, and the networking service provider, and discusses the system management needs in each environment during various stages of the computer system life cycle. The primary purpose of this chapter is to introduce the basic concepts and to set the stage for the other chapters.

## 1.8 Review Questions

1. What is computer systems management? What is the primary goal of computer systems management?
2. What are the key stages in the life cycle of a computer system? What are the principle problems addressed at each stage of the life cycle?

3. What type of computer systems do you have at your home? What are the steps you need to take to ensure that the computers are operating without any problems?
4. What are the implications of decisions made at any one stage of the life cycle on the other stages of the computer life cycle? List some of the impacts a decision made during the planning stage can have during the operation stage of the system.
5. *Project:* Do a survey of the computing systems installed at your university. To what extent does the system resemble the simplified model of enterprise network? What are the differences from the simplified model presented in this book?

## References

1. A. Westerinen and W. Bumpus, The Continuing Evolution of Distributed Systems Management, IEICE Transactions on Information and Systems, E86-D, (11): 2256–2261, November 2003.
2. J. D. Case et al., Simple Network Management Protocol, Internet Engineering Task Force Request for Comments RFC 1157, May 1990.
3. ISO/IEC JTC1/SC21/WG4 N571, Information Processing Systems – Open Systems Interconnection, Systems Management: Overview, July 1988.
4. D. Slama, J. Garbis, and P. Russell, Enterprise CORBA, Prentice Hall, March 1999.
5. Proceedings of Telecommunication Information Networking Architecture Conference, Oahu, HI, April 1999.
6. Desktop Management Task Force, URL http://www.dmtf.org.
7. C. Allen, C. Creary, and S. Chatwin, Introduction to Relational Databases, McGraw Hill, November 2003.
8. R. Boutaba and J. Xiao, Network Management: State of the Art, In Proceedings of the IFIP 17th World Computer Congress – Tc6 Stream on Communication Systems: the State of the Art, Deventer, Netherlands, August 2002.