Patricia Morreale et al. "Voice and Data Communications"
The CRC Handbook of Modern Telecommunications
Ed. Patricia Morreale and Kornel Terplan
Boca Raton, CRC Press LLC. 2001

3.3.8 Lightweight Directory Access Protocol (LDAP)

Directory services are fast becoming the key to the enterprise, allowing applications to locate the resources they need and enabling network managers to authenticate end users. Corporate networkers need to be aware about what LDAP is capable of, where it is headed, and what it was never intended to do.

LDAP was intended to offer a low-cost PC-based front end for accessing X.500 directories. Due to high overhead and acceptance delays of X.500, LDAP has emerged to fill the gap, somehow expanding its role. It rapidly became the solution of choice for all types of directory services applications on IP networks. LDAP applications can be loosely grouped into three categories: those that locate network users and resources, those that manage them, and those that authenticate and secure them. Network managers who want to put the protocol to work need to go into detail, coming to terms with standard components and features. This protocol can save companies time and money. It can help network managers keep pace with the rising demand for directory services. New applications appear almost every day. But there are limits to what a protocol can do for distributed computing. It cannot store all the types of information that network applications need. Knowing the difference between LDAP facts and fiction is the only way to avoid potential pitfalls.

3.3.8.1 Attributes of LDAP

The current specification comprises eight features and functions (HOWE99):

- Information model: Organized according to collections of attributes and values, known as entries, this model defines what kinds of data can be stored and how that data behaves. For example, a directory entry representing a person named Jim Fox might have an attribute called sn (surname) with a value "Fox." The information model, inherited almost unchanged from X.500, is extensible: almost any kind of new information can be added to a directory.
- LDAP schema: Defines the actual data elements that can be stored in a particular server and how they relate to real-world objects. Collections of values and attributes — representing such objects as countries, organizations, people, and groups of people — are defined in the standard, and individual servers can define new schema elements as well.
- Naming model: Specifies how information is organized and referenced. LDAP names are hierarchical; individual names are composed of attributes and values from the corresponding entry. The top entry typically represents a domain name, company, state, or organization. Entries for subdomain, branch offices, or departments come next, often followed by common name entries for individuals. Like the LDAP information model, the naming model derives directly from X.500. Unlike X.500, LDAP does not constrain the format of the namespace; it allows a variety of flexible schemes.
- Security model: Spells out how information is secured against unauthorized access. Extensible authentication allows clients and servers to prove their identity to one another. Confidentiality and integrity also can be implemented, safeguarding the privacy of information and protecting against active attacks such as connection hijacking.
- LDAP functional model: It determines how clients access and update information in a LDAP directory, as well as how data can be manipulated. LDAP offers nine basic functional operations: add, delete, modify, bind, unbind, search, compare, modify distinguished name, and abandon. Add, delete, and modify govern changes to directory entries. Bind and unbind enable and terminate the exchange of authentication information between LDAP clients and server, granting or denying end-users access to specific directories. Search locates specific users or services in the directory tree. Compare allows client applications to test the accuracy of specific values or information using entries in the LDAP directory. Modify distinguished name makes it possible to change the name of an entry. Abandon allows a client application to tell the directory server to drop an operation in progress.

- LDAP protocol: Defines how all the preceding models and functions map onto TCP/IP. The protocol specifies the interaction between clients and servers and determines how LDAP requests and responses are formed. For example, the LDAP protocol stipulates that each request is carried in a common message format and that entries contained in response to a search request are transported in separate messages, thus allowing the streaming of large result sets.
- Application program interface (API): Details how software programs access the directory, supplying a standard set of function calls and definitions. This API is widely used on major development platfroms running C, C++, Java, Javascript, and Perl.
- LDAP data interchange format (LDIF): Provides a simple text format for representing entries and changes to those entries. The ability helps synchronize LDAP directories. LDIF and the LDAP API, along with scripting tools like Perl, make it easy to write automated tools that update directories.

LDAP directories and operating systems are melding to create intelligent environments that can locate network resources automatically. Examples include:

- Active Directory and Windows NT (Microsoft)
- HP-Unix and LDAP (Hewlett-Packard)
- Sun Solaris and LDAP (Sun Microsystems)
- Irix and LDAP (Silicon Graphics)
- Digital Unix and LDAP (Compaq)

In this new role as operating system add-on, LDAP furnishes a way to locate printers, file servers, and other network devices and services. LDAP makes these services standard, more accessible, and in many cases, more powerful and flexible. LDAP is also starting to play a critical role in network management, where it can be a great help to network administrators. Without LDAP, managers and administrators have to maintain duplicate user information in many specific and separate directories across the network. With LDAP, it is possible to centralize this information in a single directory accessed by all applications (Figure 3.3.12). Of course, replacing key legacy applications with LDAP-enabled ones takes time, but big changes are already underway.

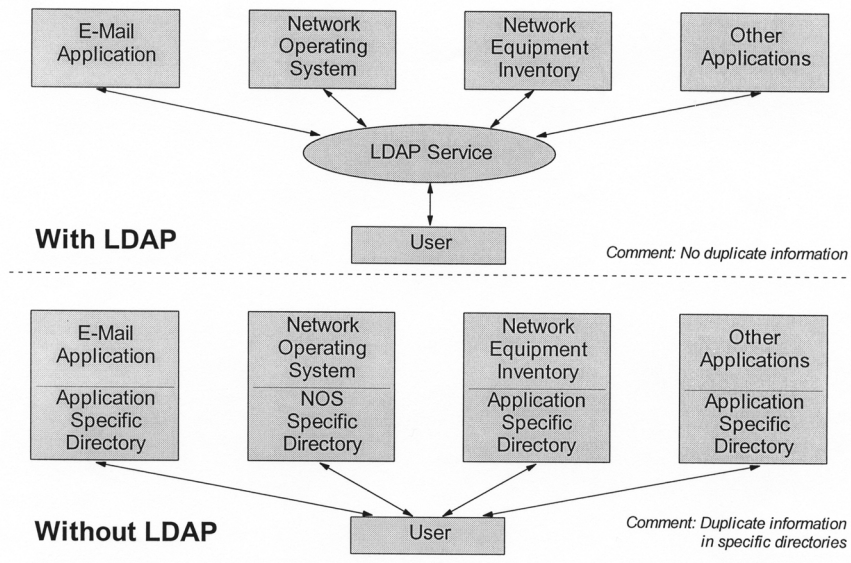

FIGURE 3.3.12 Directory centralization with LDAP.

LDAP also has an important role to play in tighter security, with the directory acting as gatekeeper and deciding who has access to what. In this capacity, LDAP performs two critical jobs. First, it serves as an authentication database. Second, once the identity of a user has been established, it controls access to resources, applications, and services using stored policies and other information. LDAP also permits corporate networkers to use their directories to implement PKI (public key infrastructure) security. From the user's point of view, LDAP provides the directory in which certificates of other users are found, enabling secure communication. From the administrator's point of view, LDAP directories are the way in which certificates can be centrally deployed and managed.

3.3.8.2 Limitations of LDAP

LDAP has three major limitations. First, the protocol cannot and will not make relational databases redundant. It lacks the heavy update, transaction processing, and reporting capabilities of these products. Nor does it offer two-phase committs, true relational structure, or a relational query language like SQL. Using LDAP to implement an airline reservation system would be a serious mistake. Second, it is not reasonable to expect that LDAP serves as a file system. Its information model is based on the simple pairing of attributes and values. Thus, it is not well suited to binary large object data that is managed by typical file systems. It is also not optimized for write performance and is unable to furnish byte-range access to values, both critical features of a file system. Finally, it does not have the locking semantics needed to read- and write-protect files. Third, LDAP is not a stand-in for DNS, which may well be the world's largest distributed database. Although LDAP's abilities are more or less a superset of DNS's — whose biggest job is translating names like home.netscape.com into IP addresses — there is a very good argument for not penetrating tasks of DNS; DNS is working fine. Also, LDAP cannot contend with the connectionless transport that DNS usually runs over. Ultimately, LDAP may have a role in managing and augmenting the information found in DNS. For example, it could link contact information to host information, but it cannot take the place of the DNS database itself.

In summary, LDAP has its place among the succesful network management tools.

3.3.9 Summary

There are multiple standards for network management. All of them have advantages and disadvantages, and, of course, also different application and implementation areas. Telecommunications suppliers and customers will have to live with multiple standards. The question is how these standards can seamlessly interoperate. There are basically three alternatives:

- Management gateway: The interoperability is realized by a special system responsible for translating management information and management protocols. Looking at the practical realization of such a gateway, it is important to target the use of OMA for both OSI- and Internet-based management. Many existing object specifications for management could be taken over by the OMA-based management.
- Platforms with multiple architecture: The interoperabiltity is realized by a multilingual platform, understanding multiple protocols. Protocol conversion is not necessary. Management information can be interpreted and transformed by the platform or by applications. Different architectures are supported simultaneously, but without deep integration.
- Agent with multiple architectures: The interoperability is realized at the agent level. In this case, the management agent understands multiple protocols and languages. It requires some intelligence for the agent. If selected, agent software must be implemented in many, practically in all, networking components. This number is considerably higher than in the case of management platforms.

There is a new group — Joint X/Open TeleManagement Forum Inter-Domain Management Group — that addresses in particular the interoperability between OSI–Management, Internet–Management, and OMG–OMA. This type of work takes a lot of time. In the meantime, practical solutions are absolutely

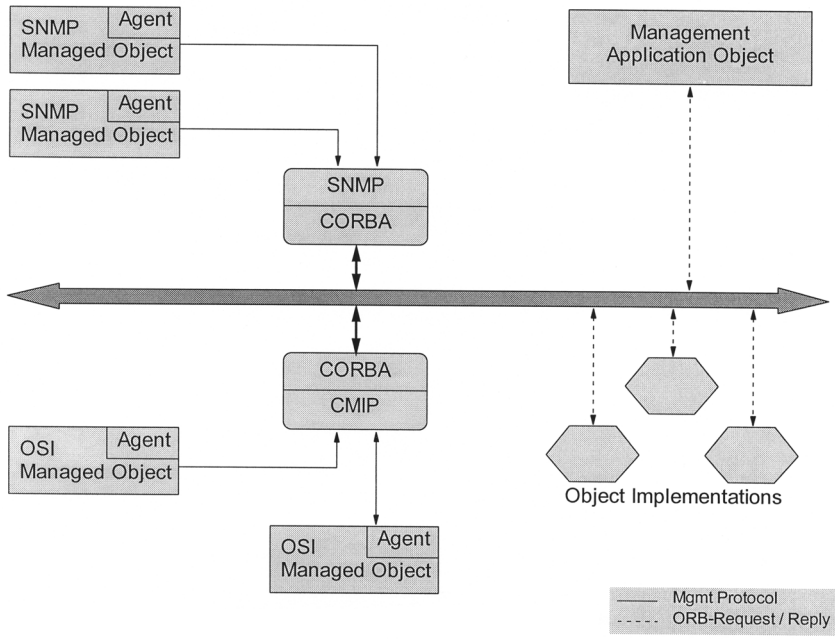

FIGURE 3.3.13 Using an object request broker to connect multiple management protocols.

necessary. In most cases, gateways deliver the quickest solutions. Such a possible solution with management gateways is shown in Figure 3.3.13. CORBA plays in both gateways, related to OSI–CMIP and Internet–SNMP, an important role.

Standardization is absolutely necessary to ensure interoperability of various components of communication systems. This chapter has laid down the basics. Management frameworks and platforms may support some of the standards, but there is no product which supports all of them.

Open database connectivity (ODBC) is an application programming interface (API) allowing a programmer to abstract a program from a database. When writing code to database, the user usually has to add code that talks to a database using a particular language. If the user wants his/her program to talk to an Access, Fox, and Oracle database, code the program with three different database languages. This can cause some problems.

When programming to interact with ODBC, the user only needs to talk the ODBC language (a combination of ODBC API function calls and the SQL language). The ODBC Manager will outline how to contend with the type of database the user is targeting. Regardless of the database being used, all of the calls will be to the ODBC API. All that the users need to do is install an ODBC driver specific to the type of database selected.

Directory enabled networking (DEN) is a specification to save information about network devices, applications and users in central directories. DEN addresses the integration of application and user-level directory information with network control and management, building on open Internet standards, such as LDAPv2 and WBEM/CIM. The CIM initiative is being extended to meet the needs of the DEN intiative. In the future, management applications will have access to authoritative information on the relationships among network elements, services, and individuals, so that preferences, privileges, and profiles can be enforced according to enterprise policies but with personal customization, and so that policies governing network applications can make use of directory-based information.

In summary, information will be consistent in DEN directories and in CIM management systems.

References

HOWE99 Howes, T.: LDAP: Use as Directed, *Data Communications*, February 1999, p. 95-103.

STAL99 Stalling. W.: *SNMP, SNMPv2, SNMPv3 and RMON 1 and 2*, Third Edition, Addison Wesley Longman, New York, 1999.

TERP96 Terplan, K.: *Effective Management of Local Area Networks*, Second Edition, McGraw-Hill, New York, 1996.

3.4 Telecommunications Management Network (TMN)

Endre Szebenyi

3.4.1 Introduction

A telecommunications management network (TMN) is a real, object-oriented, up-to-date, widely applicable network management model, defined by a number of standards and based upon the seven-layer OSI communications model. (Functions and architecture of a TMN are discussed in Paragraph 3.4.3.)

As a consequence of applying the open systems interconnection (OSI) communications protocol architecture, TMN is quite similar to the CMIP-based OSI network management described in Section 3.4.2.2.2. They are, however, not identical: TMN has been developed to be more future-oriented. (More or less, services of the OSI Management can be considered as a subset of TMN services.)

TMN network management standards have been and are conceptualized to satisfy the widest possible range of demands known today, taking streamlining requirements completely into consideration, and with the cooperation of a wide range of standard institutions involved. Institutions participating in standardization of TMN or being connected with these activities include:

- ITU/T
- ETSI
- ISO
- Network Management Forum
- ANSI, etc.

The idea of handling managed objects by network management systems in an object-oriented manner arose in the early 1980s, and later on, the conceptualization of the OSI management also began. Study of the TMN concept was started in 1985 by ITU. ITU Study Group IV aimed at working out a more comprehensive and standardized way of handling intelligent network elements in an object-oriented network management system. As the first formal result, Recommendation M.30 was published in 1988. M.30 summarized the fundamental principles of TMN. Later on, several revisions had been introduced, and M.30 was replaced by Recommendation M.3010 in 1991 [1]. The first recommendation defining the fundamentals was followed by several others discussing every important aspect of TMN, such as network management models, network management functions, standard interfaces and protocols, the way of managing standard network architectures (e.g., SDH), interconnecting networks and network management systems, etc. TMN standards have also been developed and approved by Committee T1 of ANSI.

At the moment, however, the scope of the TMN standards is not complete; their development is still in progress. New standards are planned for intelligent network management and for B-ISDN, etc. We especially have to note that a number of recommendations required to manage ATM networks are also currently being developed.

The series of TMN-related ITU-T Recommendations is illustrated in Table 3.4.1. The most important, relevant ANSI standards include: T1.204, T1.208, T1.210, T1.214, T1.214a, T1.215, T1.224, T1.227, T1.228, and T1.229. The TMN-related ITU-T and ANSI standards generally correspond to or rely on various ISO standards of the series ISO 9595, 9596, 10040, 10164 and 10165.

TABLE 3.4.1 The Series of TMN-related Recommendations

Number of Recommendations	Title of Recommendation
M.3000	Overview of Telecommunications Management Recommendations (supersedes M.30)
M.3010	Principles for a Telecommunication Management Network
M.3020	TMN Interface Specification Methodology
M.3100	Generic Network Information Model
M.3180	Catalogue of TMN Management Information
M.3200	TMN Management Service: Overview
M.3207.1	TMN Management Service: Maintenance Aspects of B-ISDN Management
M.3211.1	TMN Management Service: Fault and Performance Management of the ISDN Access
M.3300	TMN Management Facilities Presented at the F Interface
M.3400	TMN Management Functions
G.773	Protocol Suites for Q-interfaces for Management of Transmission Systems
G.774	Synchronous Digital Hierarchy (SDH) Management Information Model for the Network Element View
G.774.01	Synchronous Digital Hierarchy (SDH) Performance Monitoring for the Network Element View
G.774.02	Synchronous Digital Hierarchy (SDH) Configuration of the Payload Structure for the Network Element View
G.774.03	Synchronous Digital Hierarchy (SDH) Management of Multiplex Section Protection for the Network Element View
G.774.04	Synchronous Digital Hierarchy (SDH) Management of the Subnetwork Connection for the Network Element View
G.774.05	Synchronous Digital Hierarchy (SDH) Management of Connection Supervision Functionality (HCS/LCS) for the Network Element View
G.784	Synchronous Digital Hierarchy (SDH) Management
I.751	Asynchronous Transfer Mode (ATM) Management of the Network Element View
Q.811	Lower Layer Protocol Profiles for the Q3 Interface
Q.812	Upper Layer Protocol Profiles for the Q3 Interface
Q.821	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Alarm Surveillance
Q.822	Specifications of Signaling System No. 7. — Q3 Interface Stage 1, Stage 2, and Stage 3, Description for the Q Interface — Performance Management
Q.824	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Customer Administration — Common Information
Q.824.0	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Customer Administration — Common Information
Q.824.1	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Customer Administration — Common Information
Q.824.2	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Customer Administration — Integrated Services Digital Network (ISDN) Supplementary Services
Q.824.3	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Customer Administration — Integrated Services Digital Network (ISDN), Optional User Facilities
Q.824.4	Specifications of Signaling System No. 7. — Q3 Interface Stage 2 and Stage 3, Description for the Q Interface — Customer Administration — Integrated Services Digital Network (ISDN) — Teleservices

TMN network management systems are conceived to be capable of managing:

- Telephone networks
- LAN and WAN data communication networks
- ISDN networks
- Mobile networks
- Value added and intelligent network services
- Advanced broad-band digital networks, such as:
  - SONET/SDH networks
  - ATM networks
  - B-ISDN networks, etc.

In order to facilitate the operation, practical TMN implementations generally have a graphical user interface (GUI).

The scope of *practical implementations of TMN network management* systems is very limited *as yet.* TMNs are expected to become more widely used in the course of the next decade. In the case of SONET/SDH networks, however, due to the favorable circumstances and special requirements, TMN network management systems are likely to be predominant even in our days. Recent efforts aim at developing network management systems in the scope of which the TMN concept could also be applied to ATM, ISDN, B-ISDN, cellular mobile networks, and other types of communications systems [2–6].

3.4.2 Network Management Key Concepts

3.4.2.1 Evolution

Theory and practice of network management had emerged in the early days of telecommunication, long before sophisticated broadband data communication technologies, such as SONET/SDH, ATM, etc., were developed. Enforced by the growing demand on network quality and the increasing operational costs (more precisely: by the necessity to reduce these costs), management tools in local area (LAN) and wide area (WAN) networks increasingly came into use; moreover, in a relatively simple form, means of network management were applied in the analogue voice communication (telephone) networks as well.

With the increasing volume of computers and other intelligent devices, the interdependent system elements had to be interconnected by different communication tools (data communication lines, LANs, etc.). Large-scale networks with complex topology have evolved in this way. However, if a network does not operate reliably, if possible errors cannot be identified by simple methods, and if operational parameters of the network cannot be kept under control and be modified if necessary, users will not get much benefit from the network established at considerable expense and at great pains.

The development of network management systems was motivated by an intention to eliminate the problems listed above.

Eventually, network management is aimed at allowing authorized personnel to monitor or change important operational parameters of the network by using one or more management terminals. In practice, network management systems are made up of a suitable software, the scope of which basically extends to all the intelligent elements of the network and to some supplementary hardware elements located at the nodes of the network.

The advantage and result of applying a network management system can be summarized as follows:

- Increased safety (decreased time required for error detection, troubleshooting and error correction, increased efficiency of the efforts aimed at error correction, the possibility to reroute network traffic automatically if some parts of the network failed to operate correctly);
- Increased security (controlled and regulated network access pertaining to every user and according to the predefined authentications and authorizations);
- New services provided for the network users (acquiring billing/traffic information, etc.);
- Effective, computerized system monitoring (recording accounts and billing information, tracking and evaluating network load and performance, as well as error statistics, supporting network development strategies, etc.).

In conclusion, the objectives of network management systems are

- Decreasing operational costs
- Improving the quality of services

Achieving these objectives is equally in the interests of the

- Users
- Network operators

- Service providers
- Communications equipment manufacturers

Operating a network management system practically means performing the following essential steps:

a) Acquiring and collecting characteristic data about the elements of the network, about their operation and interrelation in the network (operating conditions, performance, fault conditions and types of the eventual malfunctions, relationship with the neighboring elements, traffic and load parameters, etc.);
b) Storing and evaluating the collected data by the appropriate data processing center of the system;
c) Controlling network operation (modifying functionality of some elements in the network if necessary) as a conclusion and result of the evaluation.

Insofar as a network management system is also required to perform tasks of service and/or business management (see Section 3.4.3.3), additional steps as part of the operation may be necessary, such as:

d) Recording service contracts and managing customer services;
e) Elaborating business plans and technical designs;
f) Modeling and simulating technical and/or financial processes, etc.

Note that a claim to perform operational steps listed under items a) to f) above necessitates the availability of a sophisticated network architecture, the presence of appropriately constructed network elements, and a suitable network management system as well. It means efficient network management techniques can be regarded as results of a joint and inseparable development of both LAN/WAN communications technology and the network management methods themselves. Network management is furnished with more and more effective tools by the evolving communications technology.

As actual network solutions concerns, not all of the above steps have been commonly realized up to now.

Chronology of evolution is represented by the sequence of items a) to f) at the same time. In the first stage of development, "network management" was limited to simply visualizing fault conditions (for example, by lighting appropriate error lamps on the equipment) and supervising them directly by the operating personnel. According to the present sense of the terminology, this type of operation cannot be considered as network management at all. The first significant step in the way to achieve efficient network supervision was by gathering fault conditions and other system parameters by computer(s) and storing them automatically. The availability of the stored data provided a convenient opportunity for their programmed evaluation. The increase in network size and volume of data resulted in the improvement of the evaluation algorithms.

The gathered data and their evaluation then allowed the proper network operation to be restored automatically in the case of certain system faults (e.g., by means of stand-by/hot stand-by system elements or through a predetermined reconfiguration of the system). Controlling network operation by a network management system was only applied in sophisticated LANs for a considerable period of time. Traditional WAN networks, however, based upon PDH technology widely used even today, offered a limited number of means for a suitable network control.

As an effect of the new communication technologies (such as SONET/SDH and ATM) the usual contrast between LANs and WANs, data and voice communication tends to vanish or disappear. This is an important moment promoting the evolution of network management methods and probably resulting in a homogeneous network management theory. Real network management products suitable to accomplish service management, technical as well as business planning, and modeling tasks are appearing nowadays.

TMN is an up-to-date management standard allowing construction of network management systems with the most comprehensive set of network management functions exploiting the facilities of the most up-to-date communications technologies.

3.4.2.2 Standard Network Management Systems

In general, different types of networks (LANs, WANs, ISDN, voice and data communication systems as well as public and private networks, etc.) set different requirements for the network management.

As seen before, in the course of the technical development, management systems became suitable for completing an increasing number of tasks. At the beginning, however, they comprised individual, manufacturer-specific hardware and software components, and, as a consequence, network management systems of different manufacturers developed in specific, diverse directions.

In order to avoid this confusing situation, in accordance with the philosophy of recent open systems, interworking of different networks should be ensured: networks and their management systems have to be operated in a multivendor environment. Consequently, in order to meet this requirement, substantial features of the network management systems (interfaces, protocols, system architecture, etc.) have to be defined by international standards. The importance of standardization is still growing.

The wide range of recommendations elaborated and/or accepted by the major international standards institutes (ANSI, CCITT and ITU-T, which became the successor of CCITT in 1993, ISO, IEC, ETSI, etc. and, especially, the Network Management Forum) aim at users' internationally approved independence from equipment manufacturers. Practically, in spite of the efforts to standardize crucial system characteristics (similar to some other fields of hardware/software solutions), there are several management systems competing with each other on the market showing a limited degree of standardization. Systems based on *de facto standards* are also often regarded as standard systems.

The most frequently used network management systems of the 1990s were as follows:

- SNMP-based management
- CMIP-based (OSI) management
- TMN

3.4.2.2.1 SNMP-based Management

The network management standard, Simple Network Management Protocol (SNMP) referring to the applied protocol, has been defined by the Internet Community for managing networks which implement TCP/IP communications protocol, such as Ethernet LANs and Internet segments. SNMP is a de facto standard, and has been available since 1990. Standardization of its second version (SNMPv2) was requested in 1993 and concluded in 1994 [7-11].

Actually, SNMP is a set of protocol standards defining the rules of information exchange between the entities of the network management software; that is, between the manager, the agent, and the management information base (MIB).

The *manager* is a software element running in the network management station and plays the role of a mediator between the human operator and the network management system itself.

The *agent* is the software element installed in the intelligent managed network elements and represents the managed resources of those elements.

MIB is the logical database containing local network management information residing in each of the agents. The manager is in command of a superset of the agents' MIBs. The MIB may include a number of standard objects. Types of standard objects are arranged (in accordance with their standard name syntax) in a hierarchic tree structure applied in TCP/IP management, making up the structure of management information (SMI). Each node of the tree represents a group of managed objects, and each group includes a number of objects. Each object may have values. The actual value of a managed object reflects the present state of the managed resource represented by the object in hand. Object values should comply with the type definition of the given object. The allowed object types are also defined, such as integer, bit string, etc. The involved standard objects are described with a standardized syntax, defined in ITU-T Recommendation X.208, known as ASN.1 (Abstract Syntax Notation One). It means SMI is similar to that of CMIP-based OSI management, but the objects defined in SNMP are different (see Section 3.4.2.2.2).

The manager may send requests to the agents periodically in order to collect actual information about their state (*polling*), or to modify variables in the agents' MIB. Accordingly, types of managers' requests are *get* and *set*. The agent may also send a *trap* to the manager automatically in order to notify it of an event, if necessary.

In spite of the availability of traps, SNMP is basically a polling-based, not event-driven protocol. It is principally devised to provide network element and network level management functionality. As opposed to OSI network management (CMIP) or TMN, the object model of SNMP is less flexible and less efficient. SNMP cannot be regarded as a truly object-oriented network management system. As the differences between SNMPv1 and SNMPv2 are of concern, one of the improvements is that SNMPv2 (contrasted with SNMPv1) can support communication between manager entities.

One of its advantages, however, is, that it can be easily operated and does not require sophisticated hardware resources. As a practical consequence, in spite of all its disadvantages, SNMP network management systems are also significant competitors of OSI network management, and have been more widely used up to now. As a trend in development of communications technology, it has to be mentioned that efforts are made to ensure interoperability of the various network management systems, such as SNMP, OSI/CMIP, and TMN.

SNMP network management systems are primarily applied for managing:

- LANs and corporate networks based upon TCP/IP protocol
- Internet network segments

Notwithstanding that TMN is expected to be the ideal management system of both SDH and ATM networks, existing ATM network products predominantly allow application of SNMP because of the lack of appropriate TMN standards at the moment.

3.4.2.2.2 CMIP-based Management

The CMIP-based OSI management defines a real object-oriented network management system based upon the seven-layer OSI communications protocol architecture. (The OSI Reference Model is defined in the CCITT/ITU-T X.200 series of recommendations, and in ISO Standard 7498 [12].) The standardization of the OSI protocol model began in the late 1980s, and has not been fully finished up to now. The first OSI management standards were defined by ISO; later they were adopted and developed by CCITT/ITU-T (X.700 series of recommendations) and other standards institutes [13, 14].

CMIP-based OSI network management systems can be applied to manage:

- Local area networks (LANs)
- Corporate networks and private wide area networks (WANs)
- National and international networks

The seventh (application) layer *protocol applied by the OSI management is the Common Management Information Protocol (CMIP).* In a CMIP-based OSI management environment, the user's application process (the operation of which is based upon the manager/agent principle) is provided with the Common Management Information Service (CMIS) as an application program interface (API) by the so-called Systems Management Application Entity (SMAE), which is implemented in the seventh (application) layer of the seven-layer ISO/OSI communications model [15, 16].

(Management operation is based upon the manager/agent principle. The functions of manager, agent and MIB are basically similar to those of SNMP, or more characteristically to those of TMN. See Sections 3.4.2.2.1 and 3.4.3.5, respectively.)

The following are major elements of the OSI management application program interface:

- Common management information service element (CMISE)
- Remote operation service element (ROSE)
- Systems management application service element (SMASE)

- Association control service element (ACSE)
- File transfer, access, and management (FTAM)

The *CMISE* is responsible for generating basic standard requests and processing answer messages as defined by the CMIS. CMIS may be used by an application process in a centralized or decentralized management environment to exchange information and commands for the purpose of systems management. (See ITU-T Recommendation X.710.)

The *ROSE* controls and supervises interactions between remote entities of a distributed application, where these interactions can be modeled and supported as remote operations. A remote operation is requested by one entity; the other entity attempts to perform the remote operation and then reports the outcome of the attempt. (See ITU-T Recommendations X.219 and X.229.)

The *SMASE* provides systems management services in support of specific management functions. (See ITU-T Recommendation X.750.)

The *ACSE* is responsible for performing initial negotiation in order to decide if data connection can be established and made available for data communication. According to the definition of Recommendation X.217: "ACSE provides basic facilities for the control of an application association between two application-entities. The ACSE includes two optional functional units. One functional unit supports the exchange of information in support of authentication during association establishment. The second functional unit supports the negotiation of application context during association establishment." (See ITU-T Recommendations X.217 and X.227.)

FTAM organizes and manages file access for application purposes, according to the specifications of ASN.1. (See ITU-T Recommendation X.209.)

In terms of the manager/agent principle, the manager as a software system element may send management operations in the form of CMIS requests to any of the software agents via the CMIP management protocol. The agent forwards these requests to the pertaining managed objects (MOs) which represent the physical and logical resources to be managed, and executes them on the appropriate MOs.

CMIP may also provide event-driven reports for the manager and can identify substantial events that influenced the state of the managed objects.

CMIP/CMIS requests, issued by the manager to the agent in order to initiate an operation, are the following:

- M-Get (gets an attribute value of one or more managed objects)
- M-Set (sets/modifies the attribute value of one or more managed objects)
- M-Action (performs a specific action on one or more managed objects)
- M-Create (creates a new managed object)
- M-Delete (deletes one or more managed objects)
- M-Cancel-Get (cancels a previously requested and currently outstanding M-Get operation)

In addition, the agent may send to the manager:

- M-Event-Report(notifies the manager of an event that occurred in the managed object)

Selection of managed objects that have to be affected by a given CMIP/CMIS operation is facilitated by scoping and filtering. Scoping entails the identification of the managed objects to which a filter is to be applied. Filtering entails a set of tests to each member of the group of the previously scoped managed objects to extract a subset of them.

Object model of the OSI management is based upon CCITT/ITU-T Recommendation X.722, widely known as GDMO (Guidelines for the Definition of Managed Objects). Structure of Management Information (SMI) is described in CCITT/ITU-T Recommendation X.720 [17, 18]. SMI is similar to that of SNMP; however, the involved standard objects and their attributes are different. The standard objects in OSI management are also described by using ASN.1 syntax, defined in ITU-T Recommendation X.208.

CMIP is a real event-driven protocol; the GDMO object model is more comprehensive than that of SNMP. It means that OSI management is more suitable for managing large and complex networks.

Application of CMIP-based OSI network management systems is spreading gradually. Their significance is expected to grow in the future (particularly for telecommunications service providers). However, TMN, as a more comprehensive, OSI-based, standardized management system seems to be a strong competitor of CMIP.

3.4.3 Functions and Architecture of a TMN

The functions and architecture of a TMN can be considered in several dimensions. Each dimension is defined (or will be defined) by existing (or developing) standards.

ITU-T Recommendation M.3010 discusses

- Functions associated with TMN
- Aspects of TMN architectures
- TMN Logical Layered Architecture

Functions associated with a TMN are classified in terms of the OSI Management; accordingly, five management functional areas are defined in Section 3.4.3.1.

As general TMN architecture concerns, three basic aspects are considered in ITU-T M.3010:

- TMN functional architecture
- TMN information architecture
- TMN physical architecture

In addition, M.3010 describes four plus one management layers.

All of these aspects will be discussed here in a didactic sequence, differently from their order in M.3010.

3.4.3.1 Functional Architecture

The TMN functional architecture is based upon a number of TMN function blocks. These represent the appropriate functions required by the TMN to fulfill its general function of network management. These are actually performed by the elements of the physical architecture of TMN (see Section 3.4.3.2). According to Recommendation M.3010, some of the function blocks are partly in and partly out of TMN. These function blocks are the workstation function block, the Q adaptor function block, and the network element function block. It means that these function blocks (besides those functions defined by the Recommendation M.3010) may include a range of functionality (and provide them for the TMN) not covered by TMN recommendations.

Recommendation M.3010 defines five function blocks:

- Operations systems function (OSF) block
- Network element function (NEF) block
- Workstation function (WSF) block
- Mediation function (MF) block
- Q adaptor function (QAF) block

Data communication requirements of these function blocks are satisfied by the

- Data communication function (DCF)

The *OSF block* processes information related to telecommunications management for the purpose of monitoring/coordinating and/or controlling telecommunication functions including management functions.

The *NEF block* communicates with the TMN for the purpose of being monitored and/or controlled. The NEF also includes telecommunications functions that are the subject of management, but not part of the TMN. These functions are represented to the TMN by the NEF.

The *WSF block* provides the means to interpret TMN information for the user, and vice versa. Parts of the WSF may also be located outside of the scope of TMN.

The *MF block* acts on information passing between an OSF and a NEF or QAF to ensure that the information should conform to the expectations of the function blocks attached to the MF.

The *QAF block* is used to connect a non-TMN compatible, NEF-like function block with an OSF, or a non-TMN compatible, OSF-like function block with a NEF. It means that information between a TMN-compatible function block and a non-TMN-compatible function block will be translated by the QAF. A given part of the functions of the QAF is definitely out of the TMN.

Each of the above function blocks is composed of basic functional components, which have been identified as the elementary building blocks of the TMN.

DCF is applied to transfer information between the TMN function blocks.

Pairs of the TMN functional blocks exchanging management information are separated by reference points; that is, reference points are boundaries between two management function blocks.

Three classes of reference points are defined in terms of TMN, namely:

- Reference points q
- Reference points f
- Reference points x

In addition, two further classes of non-TMN reference points are considered:

- Reference points g
- Reference points m

Definition of these reference points will be explained in Section 3.4.3.2, in connection with the physical architecture of TMN.

3.4.3.2 Physical Architecture

TMN physical architecture describes physical building blocks (physical elements) of a TMN network management system and contains exact rules for their relationships. The relating standard is CCITT/ITU-T Recommendation M.3010.

M.3010 defines the physical elements of TMN as:

- Operations system (OS)
- Workstation (WS)
- Network element (NE)
- Data communication network (DCN)
- Mediation device (MD)
- Q adaptor (QA)

Furthermore, M.3010 also defines rules of data interchange between physical elements as

- Standard interfaces

Standard interfaces define the

- Protocol suite
- Messages carried by the protocol

The *OS* performs operations system functions (OSFs, see Section 3.4.3.1). It practically is the heart itself, responsible for managing the network by controlling the operation of their elements. As its realization concerns, the OS is basically made up of one or more data processing centers, performing the task of gathering information from the network elements and processing them according to the functions of the network management system. (Suitably, the operations system meets open systems' requirements.)

Connected to one and the same network, more than one operations system may participate in the management process. Based upon the principle of task division, they may perform a predefined set of tasks and be interconnected through the data communication network.

Workstations perform Workstation Functions (WSFs). They are the physical representations of the necessary man-machine interfaces by means of which the operators can communicate with the TMN. Workstations are preferably computers themselves, equipped with efficient graphic capabilities in order to be able to meet the operators' requirements.

NEs perform Network Element Functions (NEFs). NEs are elementary, manageable telecommunication devices (e.g., switches, multiplexers, cross-connects) situated at the nodes of the network to be managed. They provide for the appropriate functions in network operation, and usually can be identified with a single address by the operations system of the TMN. The pieces of equipment forming network elements are intelligent devices controlled by their own microcomputers and equipped with standard interfaces. Through the standard interface the network element can transfer messages toward the operations station and inform it about the elements' actual state and receive control commands from it. Generally, network elements are devices meeting the requirements of TMN recommendations; however, network elements not conforming to TMN standards can also form part of the telecommunication network.

The *DCN* is a network that performs data communication function (DCF). It transmits messages required to perform management functions between the OS and the NEs. Information is exchanged through the DCN, using standard protocols as determined by standard interfaces. CCITT/ITU-T Recommendation M.3010 defines the data communication network of a TMN network management system in abstract terms. In practice, this network can be realized separately from the managed telecommunication network (e.g., it can be made up of leased lines, make use of a public data network etc.), or even (partly or entirely) by the managed network itself. If the DCN serving for management purposes is realized independently from the network managed, this will result in the advantage that troubles in the managed telecommunication network will not deteriorate functionality of the management system. An obvious disadvantage of this method as compared to the second one is higher investment and maintenance costs and a more complex overall system. A practical way of realization must always be determined with the requirements met in the possible highest degree. Management DCNs are often realized by the managed network (due to the relatively simple structure and the mostly small size) in the case of LANs, and it is almost the exclusive solution for networks built up of SDH/SONET technology. This latter is justified by the high reliability of the SDH/SONET networks (ensured by the reliable network elements and by redundant network topologies), and furthermore by the availability of communication channels reserved for management purposes in the SDH/SONET frame structure.

Mediation Devices perform mediation functions (MFs). They may adapt the information passing between the OS or the DCN and those TMN-compatible elements located in the network, which still require appropriate operations (storage, adaptation, filtering, etc.) to be performed on the exchanged data. It should be noted that confusion may often be observed concerning the interpretation of MF. Users are namely inclined to refer to the QA as a mediation device.

QAs perform Q adaptor functions (QAFs). They accomplish information exchange between the OS or the DCN applying standard protocols and the eventual non-standard NEs located in the network. For instance, mediation devices may establish interconnection between a standard DCN and the connected non-standard network elements with Q_x interfaces.

Standard interfaces define the ways and rules of information exchange that can be carried out through the reference points of a TMN network. Standard interface definitions include descriptions of the hardware interfaces, descriptions of the communication protocols as rules of data exchange, as well as the information model; that is, the system-level strategy of interworking between elements taking part in communication. Each of the interface definitions involves the definition of the appropriate *protocol families* and protocols.

Reference points are abstract concepts representing theoretical boundaries between the physical elements of a TMN. There are a number of well-defined standard reference points in a TMN network. They

are indicated by lower-case letters, whereas the appropriate capitals symbolize the corresponding interfaces. The reference points and interfaces defined in TMN are:

- Reference point "q" may be located between any two of the following function blocks: OSF, QAF, MF, and NEF. It means reference point "q" may be found between either the OS and a NE, or between the OS and a QA, or between the OS and a mediation device, or between any of the above elements and the DCN or between two OSs belonging to the same TMN. The interface "Q" allows data exchange through reference point "q." At present, two types of "Q" interfaces are distinguished; these are interfaces Q_3 and Q_x. The appropriate "Q" protocol suites are the Q_3 and Q_x protocols. "Q_3" protocols are complete implementations of the seven-layer OSI reference model, whereas "Q_x" protocols only comply with the three bottom layers of the OSI model. Q_3 is applied to connect functional elements being fully TMN-compatible, while Q_x can be used in special cases when the Q_3 interface cannot be implemented (due to the presence of any equipment not complying with TMN).
- A reference point "f" is located between function blocks OSF or MF and a WSF. It means that reference point "f" can be found between the DCN (or a mediation device) and a WS connected to it. The corresponding interface "F" allows data exchange through reference point "f" (that is, the interface "F" is applied in cases when the WS is not connected directly to the OS, but through the DCN). The appropriate protocol suite is the "F" protocol.
- A reference point "x" is located between OSFs of two TMNs or between the OSF of a TMN and the OSF-like functionality of another, non-TMN network. It means that a reference point "x" may be found between the DCNs of two TMNs, or it lies between a TMN and another managed system not complying with the TMN standards. A corresponding interface "X" allows data exchange between the two network management systems. The appropriate protocol suite is the "X" protocol.

In addition, Recommendation M.3010 defines two further reference points, lying outside of the area of the standard TMN elements:

- Reference point "g" is located between the human user and the WSF
- Reference point "m" is located between the QAF and an element that does not conform to TMN recommendations

The specifications of standard interfaces, primarily those of the family of interface "Q" and the corresponding protocols, are highly significant regarding the operation of TMN. A number of CCITT/ITU-T recommendations deal with the specification of TMN protocols (M.3020, M.3300, Q.811, Q.812, etc.). It should be noted, however, that the scope of standards specifying standard interfaces, including the "Q" interface, is not complete yet.

An example of a simplified physical architecture for a TMN as demonstrated in the ITU-T Recommendation M.3010 is shown in Figure 3.4.1. A more illustrative representation of the physical architecture of a TMN can be seen in Figure 3.4.2.

3.4.3.3 Logical Layered Architecture of a TMN

According to the TMN terminology, OSFs of the network management are broken down into four hierarchy layers. Each layer of the given hierarchy defines an appropriate group of management operations. The layers are built upon each other; they (and the appropriate operations) are closely interrelated.

TMN Standard M.3010 defines the following four layers of the OSF:

- Network element management layer
- Network management layer
- Service management layer
- Business management layer

OSFs in these layers interact with OSFs in the same or other layers within the same TMN through a reference point "q_3," and with ones of another TMN through reference point "x."

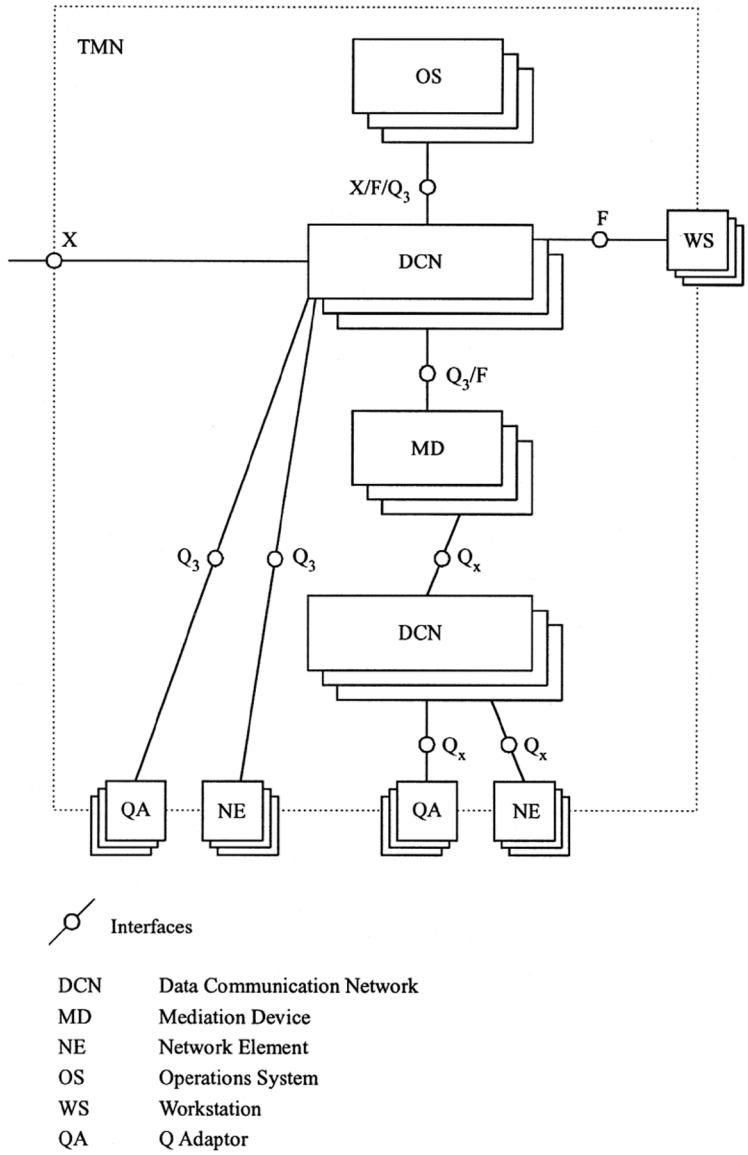

FIGURE 3.4.1 Example of a simplified physical architecture for a TMN, according to ITU-T M.3010.

Furthermore, as network element management is based on the data collected about the respective elements of the network, *network elements* themselves *can be considered as the lowest layer in the hierarchy.* As contrasted with the four OSF layers, layer of the network elements does not involve OSF, but it is concerned with the NEF.

Let us briefly review the functions included in the given layers of the hierarchy, from the bottom to the top.

Network elements are basic components of the managed network, installed as physical devices, specified by standard functions and interfaces, capable of delivering data on their operation, and providing means to be controlled in a specified way by the management system. The concept of NEs is clearly defined by TMN standards (see Section 3.4.3.2).

The network element management layer manages each network element on an individual or group basis. NE management includes gathering data on each of the network elements and controlling them

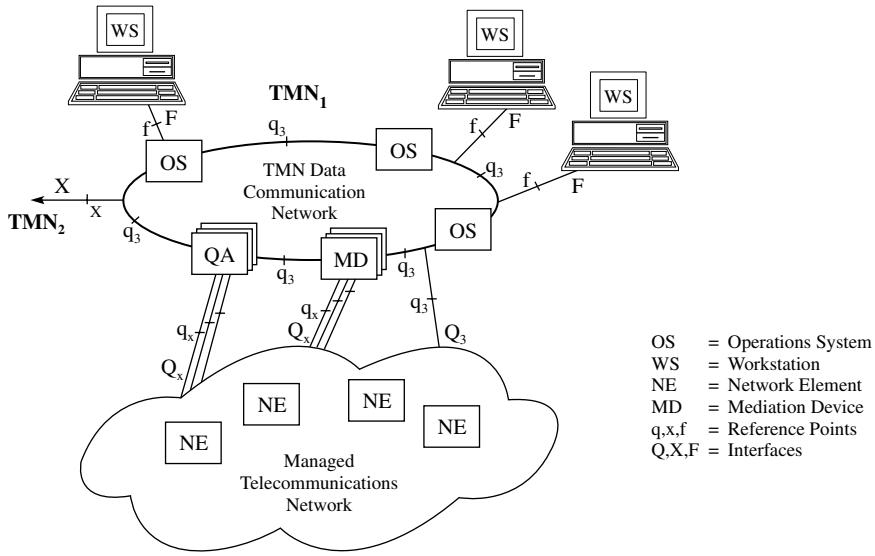

FIGURE 3.4.2 Illustrative representation of the physical architecture of a TMN.

individually. At this layer, decisions on changing the state of any individual NEs must rely on information about the same element, and cannot depend on the state of any other NEs or that of the entire network. Basic fault management as well as performance management operations, such as monitoring and displaying fault conditions or traffic performance of any single network elements, as well as taking elementary actions to eliminate an error (e.g., switching over to an auxiliary channel inside of the same network element, etc.) are performed by the NE management layer. (See also Sections 3.4.1, 3.4.2, and 3.4.3.)

The network management layer has the responsibility for the management of the whole network as supported by the element management layer. This layer transgresses competence of network element management and is responsible for the interconnection and cooperation of all the network elements in the managed system. Tasks of this layer include configuration management (both static and dynamic; see Section 3.4.3), event, fault, and performance management on the network level (all these employing a system approach, e.g., by applying error/performance correlation and evaluation algorithms), as well as security management (monitoring user requests and taking appropriate actions in order to prevent any unauthorized access to the network).

Service management is concerned with and responsible for the contractual aspects of services provided to customers or available to potential new customers. Service management aims at establishing relations between services provided by the network and requirements of the users or customers. Clients and service contracts are recorded, customers and the appropriate service parameters are related, quality of service is traced, clients' complaints are registered and reported, and new orders are accepted and processed, etc. at this management layer.

Business management has responsibility for the total enterprise. It deals with technical and business concerns as an organic complex of the network operators' activity, and has responsibility for the total enterprise. Functions included in this layer are billing and accounting, maintenance management, cost control, controlling spare parts inventory, designing new network elements and/or new network services, technical modeling and optimization, and planning and evaluating profitability of new investments, etc. The business management layer comprises proprietary functionality. In order to prevent access to their functionality, OSFs in the business management layer do not generally have "x" reference points and "X" interfaces. It is included in the TMN architecture to simply facilitate the specifications required of the other management layers. Nevertheless, it may still be necessary that the business management layer interacts with other management or information systems. The task of these interactions should be solved by special software solutions.

Logical layers of the management hierarchy are not defined by the existing standards in full detail. Tasks and processes of the different layers and rules of data exchange between them are not exactly determined at present.

The structure of layers is schematically covered by recommendations M.30 and M.3010. In technical literature, discussing functionality of network management is often confined to simply listing management functions without detailing their hierarchy.

In present practice, standard solutions are mostly restricted to realize network element management and network management layers. Frequently used, standard management functions (see Section 3.4.3.4) essentially correspond to the tasks of these two management levels. As yet, relatively few companies provide actual solutions for service management. Rather, network management products suitable for performing tasks of business management are now mostly under development.

As a consequence, existing tools promising to solve higher level tasks of network management can be regarded as manufacturer-specific, proprietary software products.

The layered architecture of the operations system functions as defined in ITU-T Recommendation M.3010 is shown in Figure 3.4.3.

In technical references, logical layers of the management functions are often illustrated by a pyramid (see Figure 3.4.4), indicating that the greatest amount of elementary data can be found at the bottom level, but the degree of their complexity (as they are processed) increases by going up through the layers.

It is perhaps not useless to note that a remarkable similarity may be established between the pyramid-shaped functional network management model and the structure of a typical business management information hierarchy. In general, this latter may also be divided into several layers between the operative and the top management controlling level. It is also obvious that a close interrelation may be required between business-level network management and the upper controlling level of the relevant business management information system. (Interrelating functions are billing, accounting, cost control, and inventory management, for example.) Establishing on-line interconnection between TMN network management systems and business management information systems may represent one of the most important future developing efforts on the side of software manufacturers of both product families (see Section 3.4.6).

3.4.3.4 Functions Associated with a TMN

As a management functional area of TMN concerns, Recommendation M.3010 lists five management functions and refers to OSI Recommendation X.700 [13]. TMN network management functions are classified by ITU-T Recommendation M.3400 in more detail (see Table 3.4.1). (Remember that streamlined methods of network management are results of a natural evolution, and obviously most items defined as standard TMN functions can also be found in network management systems and standards established before the appearance of TMN. Still, as a new aspect of structuring systems architecture, management functionality is divided into hierarchic layers in terms of TMN.)

Referring to Section 3.4.3.3, we have to point out that existing TMN standards with respect to the layer definitions are rather schematic — especially regarding functions belonging to the higher management layers. Regardless, all data gathered in the scope of the management functions defined by ITU-T M.3400 and listed below may also be related to the functions of the higher management layers, at the least in a condensed and evaluated form. Still, they represent tasks that basically have to be implemented just at the network element and network management layers. As an exception, accounting management may be strictly related to the higher management layers in the architecture.

Standard TMN management functions according to the definitions of Recommendation ITU-T M.3400 are:

- Performance management
- Fault (or maintenance) management
- Configuration management
- Accounting management
- Security management

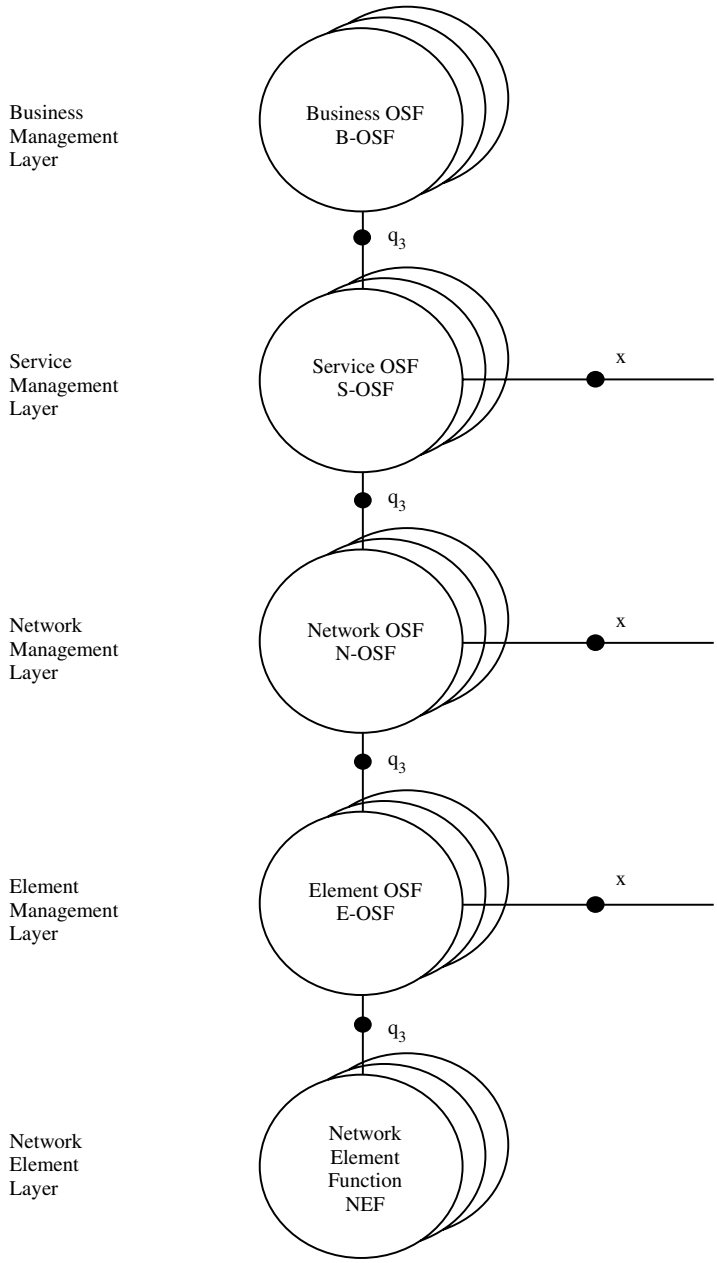

FIGURE 3.4.3 Layered architecture of the operations system functions (OSF) as defined in **I**TU-T M.3010.

Although not considered as an independent management function in itself, and TMN standards do not cover it at the moment, the process of downloading programs from the network management system to the intelligent network elements through the data communication network is closely related to network management functions.

3.4.3.4.1 Performance Management

Performance management (sometimes also referred to as traffic and performance management) provides functions to evaluate and report upon the behavior of telecommunication equipment and the effectiveness of the network and/or network elements. It may involve measuring the intensity of data flow along

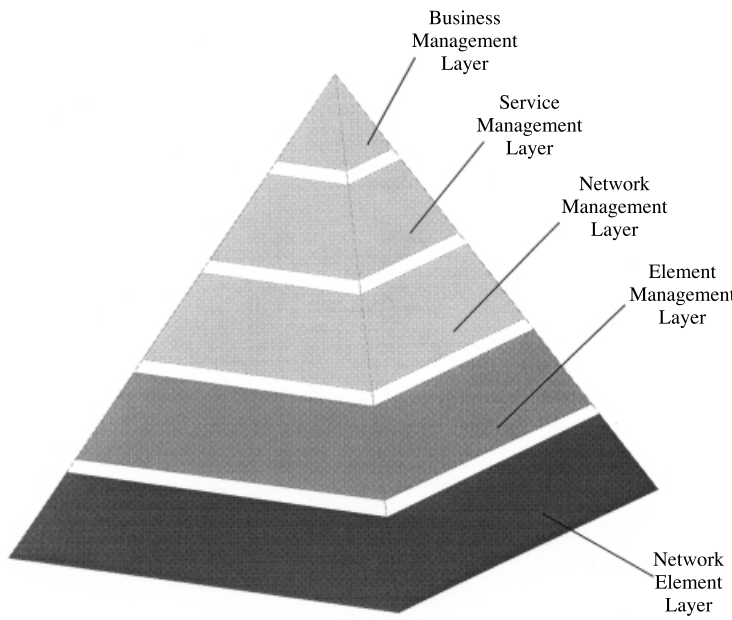

FIGURE 3.4.4 Pyramidal illustration of TMN Logical Layered Architecture.

the different routes of the network, collecting, evaluating, and displaying data measured in this way, as well as determining efficiency indices and calculating trend analysis. Information gathered and evaluated in this process can be utilized similarly to and in connection with data gained in the scope of the fault management. On the basis of these data, the level of traffic load can be established and it can be determined whether a given network complies with the necessary performance requirements. (If any congestion occurs, overloaded network routes may be relieved by system reconfiguration or by altering the actual routing strategy. Automatic intervention in network operation can be performed by the network management system. If a permanent lack of network capacity has been observed, the decision should be made to start new investments and increase network capacity (see also Section 3.4.2).

3.4.3.4.2 Fault Management

The *fault (or maintenance) management* (sometimes also referred to as event and error management) is a set of functions which enables the detection, isolation, and correction of abnormal operation in the telecommunication network and its environment. Its purpose is to detect and record events that have occurred in different parts of the network, then to establish the cause of these events with the most possible details and accuracy. All these are aimed, first of all, at exploring errors and being able to repair them in the shortest time possible.

Fault management may be confined to simply recording alarm states originating from the separate network elements and generating appropriate error messages in order to inform the operator. A more sophisticated and effective method is to correlate these individual events and evaluate their correlation by means of appropriate algorithms. Correlation algorithms may be needed to identify causes of malfunctions and locate their sources precisely in complex situations. That is, one elementary error may generate a number of error reports — and, oppositely, one error report may refer to several events or alarm states (equipment breakdown, cable rupture, traffic congestion, etc.). If necessitated by the amount or type of failures, automatic loopback tests or other special test routines will be started in order to locate faulty system elements. Analyzing interrelations of the elementary alarm states and evaluating results of the test routines may result in a precise diagnosis even if individual methods have been unsuccessful.

As a result of applying event and error management, a great part of malfunctions can be discovered and eliminated before they may cause any noticeable problems for the users. In some cases, the impact

of the emerging failure can be eliminated by an automatic and dynamic reconfiguration of network routes (see also Section 3.4.3). Automatic reconfiguration may generally be effective in meshed networks and (if just a single cable rupture occurred) in a ring network configuration or in case of traffic congestion, etc.

Fault management is also concerned with keeping statistics. Statistics may assist the operator in deciding if a network complies with the appropriate requirements (reliability, performance, etc.) or not.

3.4.3.4.3 Configuration Management

Configuration management provides functions to exercise, control, identify, and collect data from or provide data to Network Elements.

In practice of managing up-to-date, broadband telecommunication networks, *configuration management* generally includes two essential, logically different functions. In particular: static and dynamic configuration management.

Static configuration management involves assigning and unassigning network elements to or from the network logically ("attach" and "detach"), as well as recording, indicating, displaying, and reporting network topology and lists of network equipment along with their system parameters, such as type, topological location, symbolic and physical addresses, etc. In case of small and simple networks (along with recording equipment parameters), static configuration management may also involve *inventory control*; however, in the case of complex networks, this function should be handled separately.

Dynamic configuration management involves establishing the actual routes for the required interconnections via the network. It implies network reconfiguration by establishing a new possible route, if the actual route breaks down, or taking down the route if a request arrived requiring that the connection has to be canceled. In terms of dynamic configuration management, displaying network topology has to reflect the actual routes and connections in the network.

3.4.3.4.4 Accounting Management

Accounting management (sometimes also referred to as billing and accounting management) is a set of functions that enables the use of the network service to be measured and the costs for such use to be determined. Accounting management should provide facilities to collect accounting records and to set billing parameters for the usage of the service.

In the scope of accounting management, time and other characteristics of users' network access is measured and charging data are calculated on the base of several parameters (price lists, subscriber contracts, time of use, utilized services, etc.). Billing and accounting information is collected, classified and recorded. On the base of charging, data bills can be prepared and sent to the customers, income can be calculated and recorded, etc.

3.4.3.4.5 Security Management

Security management involves establishing classes of authentication, checking users' authorization to network access, controlling passwords, and taking other possible measures to prevent the network from any unauthorized access. It may be a special task to protect the management terminals from unauthorized interventions according to the given security requirements. Depending on the special requirements given in accordance with the purpose of a network, functions contained within security management may differ from application to application.

3.4.3.4.6 Downloading

Downloading new software versions, routing tables, cross-connect tables, or other program segments from the management center to the intelligent network elements constitutes an important task of a network management system. Software downloading provides some means to perform remote control of network elements as well. By way of software downloading, for instance, the operation of a network element can be modified, a new software version can be put into use, or the configuration table stored in a network element can simply be modified. One of the practical ways to solve configuration management may be to modify configuration tables by software downloading. While specifications of software downloading in LANs are included in IEEE Standard 803.1, there is no general recommendation for this

way of remote control of managed TMN network elements at the moment. (For management of SDH, ITU-T Recommendation G.784 assigns this item for further study.)

3.4.3.5 Information Architecture

The general characteristics of the information model of TMN are discussed by CCITT/ITU-T Recommendations M.3010 and M.3100. Further recommendations deal with the information models of managing SDH and PDH networks (e.g., Recommendation G.774).

Essentially, network management involves the exchange of information between management processes. The TMN network management information model relies, to a great extent, on the network management model OSI/CMIP. The information architecture of TMN is based upon an object-oriented model, applies transaction-oriented information exchanges, and utilizes the so-called agent/manager principle.

The basic concepts used in the definition of the TMN information architecture are similar to those applied in SNMP and OSI/CMIP (see Sections 3.4.2.2.1 and 3.4.2.2.2). They are:

- Managed object (MO)
- Agent
- Manager
- Management information base (MIB)

MOs are abstractions of the *physical or logical* resources to be managed in order to monitor operation of the network and to prevent disorders in network operation. A managed object does not generally correspond to a single network element as defined to be a part of the physical architecture (see Section 3.4.3.2). In most cases, a managed object represents one of the important components of a physical network element (it may be the control unit of a given circuit) or a logical resource (e.g., the status of a basic physical element). Every managed object must have a single unique name; its actual condition is a function of time.

A managed object is defined by

- Its attributes visible at its boundary
- The management operations that may be applied to it
- The behavior exhibited by it in response to management operations
- The possible notifications or messages it may emit during operation

A *manager* is a system element whose task is sending management requests toward the agents for control, coordination, and monitoring purposes, performing operations on the agents by the aids of these requests and receiving messages emitted by the managed objects and forwarded by the agents to the manager. In practice, the manager's part is played by a workstation (that is, the pertaining network management software running in the workstation) which in turn is a part (physical element) of the network management system (see Section 3.4.3.2).

An *agent* is a system element toward which management commands are directed for control, coordination, and monitoring purposes. Agents perform operations on managed objects according to the manager's requests, and forward messages emitted by the managed objects to the manager. In practice, the agent's function is performed by an intelligent network element of the network management system (that is, by the program running in it).

By definition, a many-to-many relation between Managers and Agents may exist. It means that one manager may be involved in the information exchange with several agents, and one agent may exchange information with several managers.

The *management information base* (MIB) is a database containing data pertaining to the managed objects of the system. Similar to the concept applied in the case of SNMP and OSI/CMIP, each of the agents has its own MIB, the manager is in command of a superset of the agents' MIBs (see Section 3.4.2.2.1). The set of rules describing the structure of a management information base is called structure of management information (SMI).

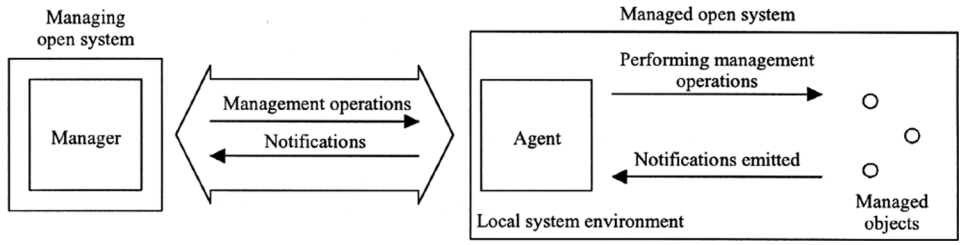

FIGURE 3.4.5 The interaction between manager, agent, and managed objects according to ITU-T M.3010.

The interaction between an agent, a manager, and the managed objects according to ITU-T M.3010 is shown in Figure 3.4.5.

Manager and Agent communicate using standard "Q" protocols built up according to the seven-layer OSI communication model.

The components of a Q protocol are:

- Application interface (command/answer structure)
- Application protocol (the seventh layer of the OSI model)
- Support protocols (layers 4–6 of the OSI model)
- Network protocols (layers 1–3 of the OSI model)

Essential elements of the TMN application interface are highly similar those of the OSI/CMIP management (see Section 3.4.2.2.2). They are:

- Common management information service element (CMISE)
- Remote operation service element (ROSE)
- Systems management application service element (SMASE)
- Association control service element (ACSE)
- File transfer, access and management (FTAM)

The arrangement of Q protocol layers in the seven-layer OSI communication model is shown in Figure 3.4.6.

3.4.3.5.1 The OSI-based Object Model

The object model gives formal description of standard objects applied in the system and defines their relationships. Objects are grouped together to form object classes: members belonging to the same class will have the same characteristics as the given respect of classification concerns. Furthermore, object classes can also be grouped together to form more general object classes. As a consequence, standard objects are arranged in a hierarchical tree structure. Each node of the tree represents a group of managed objects, and each group includes a number of objects.

Basics of TMN object model and information model are described in ITU-T Recommendation M.3010; a catalogue of TMN object classes is given in ITU-T Recommendation M.3180 (see Table 3.4.1). Together with basic object class definitions, M.3010 defines the agent/manager relationship as shared management knowledge (SMK).

In practice, object models, applied in presently used TMN management systems highly rely upon OSI management principles, namely on CCITT/ITU-T Recommendations X.722 (GDMO) and X.208 (ASN.1). See also Section 3.4.2.2

3.4.3.5.2 Distributed Object Models and Service Management

The evolution of communications technologies is reflected by the evolution of TMN and further by the evolution of its object model. Focusing on the second one and considering the hierarchical layers of network management described in Section 3.4.3.3 (representing the most illustrative dimension of this

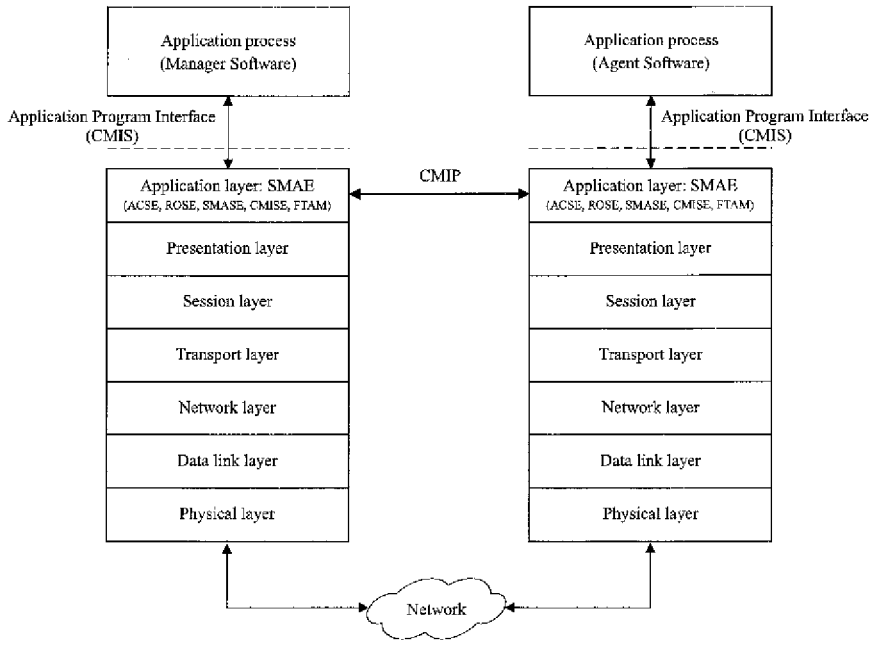

FIGURE 3.4.6 Q protocol layers in the seven-layer OSI model.

evolution), the first steps going up this hierarchy were modeling simple network elements then defining whole network models. The next grade that should now be mastered is service management. One of the main difficulties in this developing phase is that service management sets have different requirements for the object model than management on the element and element management level.

From the point of view of the object model, the main differences between network elements and telecommunication services are that (not intending to be exhaustive): the number of NEs is generally much higher than that of network elements; services are, however, more complicated, more dynamic than NEs, and are of a distributed character while network elements are not.

According to the first two developing steps discussed above, the current OSI-based object model (GDMO) is static, built up using a centralized system approach, and basically preconditions central intelligence. Most of the operations performed on GDMO objects using CMIS/CMIP are relatively simple; these operations generally relate to a group of objects and their extent is determined by the given filter and scope conditions. However, they are very effective in handling a large number of relatively simple objects as required by network element and network level management, of whose requirements they have been designed to meet.

Recent telecommunications technology, taking end-to-end customer service for example, represents a complex, integrated environment which necessitates the availability of service management, and as such, requires interoperability, flexibility, and distributed object processing of the network management system. Service management involves solving the problem of how to integrate and manage different services like telephone, video, Internet, and CATV, and how to manage services passing through the domains of different telecommunications network and/or service providers being geographically dispersed, probably having different service technologies, applying various business and traffic policies, etc.

GDMO is not really suitable for modeling distributed management objects and for realizing the appropriate management functions that are necessary for service management.

To override this problem and to meet the requirements of a distributed system approach, ITU-T defined the Reference Model for Open Distributed Processing (RM-ODP) and Open Distributed Management Architecture (ODMA) in Recommendations X.901-4 and X.703, respectively [19, 20]. RM-ODP lays down the principles and gives basic definition of a distributed system; it does not involve directives

for practical implementation. ODMA is aimed at developing OSI management toward a distributed processing environment.

The TeleManagement Forum also tries to standardize an object model for service management. One of the results of this developing effort was OMNIPoint (Open Management Interoperability Point), a standard solution providing object class definition and specifying an integrated management architecture [21, 22]. Recently, the Forum turned its attention to the latest solutions in the field of network management technology, such as integrating CORBA and TMN.

CORBA (Common Object Request Broker Architecture) is a solution for distributed, object-oriented data processing, developed by the Object Management Group [23]. CORBA 1.1 was introduced in 1991, whereas CORBA 2.0 was adopted in 1994. By now, CORBA has become a *de facto* standard in this field.

CORBA defines a general framework for distributed, object-oriented program development. (Network management is one of its possible applications.) CORBA includes an object model; the possible operations on them are furthermore defined and the way these operations may be performed is specified. CORBA also defines a programming language-independent Interface Definition Language (IDL) and the appropriate programming interfaces. The communication mechanism meeting the requirements of the managed objects (as a counterpart of CMIS/CMIP in CORBA) is provided by the Object Request Broker (ORB).

As proven both in theory and by some realized software models, CORBA may effectively be applied for communications service management. It does not mean that traditional OSI object models, such as GDMO (together with CMIS/CMIP) would have to simply be replaced by CORBA at any time. Rather, they should be integrated — GDMO being used for network element- and network-level management — while CORBA is used for the management of services. Principal reasons of this argument are that despite all the advantages of CORBA in service management, GDMO is more effective on the element level, and furthermore, economic reasons require that new developments do not disadvantageously influence utilization of existing implementations.

Integration of CORBA and TMN, however, cannot be solved without any difficulties. The CMIP protocol, the appropriate services defined by CMIS and the GDMO object model, are well different from those protocols, services, and object models represented by CORBA. In order to build facilities of CORBA into TMN, the two different object models, the appropriate management functions and the related operations should be combined. Publications report on special software solutions in this field, and simultaneously, the NMF makes efforts to elaborate standard methods to meet this requirement [24–26].

Nevertheless, recent CORBA applications have to be regarded as proprietary solutions, since neither the way of TMN-CORBA integration, nor service-level TMN objects have been standardized up to now.

One of the most recent results in this respect is Telecommunications Information Networking Architecture (TINA), being developed by the TINA Consortium. TINA specifies a technology-independent information model for telecommunications systems, based upon distributed operations. According to all indications, TINA will not get away from being integrated with TMN [27–29].

3.4.4 Interconnecting Managed Networks

3.4.4.1 Interworking of Different Network Management Systems

As a result of recent technical development, telecommunications networks have become globalized; private and public networks, LANs, and WANs cannot be sharply separated. It means that (in most cases) public network services will actually be realized by a number of network and/or service providers, and by using a number of network elements originating from different vendors and manufacturers. A single end-to-end interconnection, satisfying the users' actual demands, may pass through ISDN, SDH, ATM, mobile network, LAN, WAN, and public and private network paths.

This situation drives both private and public network operators to integrate their network management systems to the highest possible degree. Lately, demand on establishing end-to-end management has also arisen.

As technical aspects of the TMN concerns, increase interworking of different network management systems has to be realized.

According to ITU-T Recommendation M.3010, TMN hierarchies may interact for many reasons, such as:

- To manage the interactions required to provide value-added services
- To manage a number of geographical/functional TMNs as a single TMN
- To provide end-to-end services

As described in Section 3.4.3.2, OSFs located in one and the same TMN may exchange information via the "q" reference point, using a "Q" interface, while different standard TMN systems can be interconnected via the "x" reference points and may exchange information by using standard "X" interfaces. Should we have a network management system not complying with the TMN standards, interworking of a TMN and a non-TMN system may theoretically be established through the "X" interface as well. In the latter case, the non-TMN system must be able to communicate through this "X" interface, handle the appropriate protocol, and process data formats according to the MIB of the TMN. In order to do this, however, sophisticated software solutions should be applied.

In practice, efforts have been made to solve the problem of integrating OSI and SNMP, or TMN and SNMP management. Solutions aimed at this purpose involve the appropriate protocol conversion and the translation of management information [30].

3.4.4.2 Virtual Networks, Customer Network Management

TMN network management systems also enable us to establish virtual networks within the communication routes of the network as a whole.

The users of a virtual network will have an experience similar to using a separate network, being independent of the entire communication system by the aid of which the virtual network is realized. The nodes and the domain of authorized users of a virtual network can be determined in accordance with the user or customer requests, independently of the actual geographical location of the required virtual nodes and that of the user access points (of course, they should be covered by the topology of the entire network) and independently of the relation between the network hierarchy and the locations of the virtual nodes.

A separate management center may be assigned to the virtual network; however, it has to be subordinated to the entire network management system. This means that performing management functions (configuration management, event management, performance management, etc.), with respect to the operation of the whole network, remains the task of the central network management. The management center of the virtual network will have rights to monitor the actual operating parameters of the virtual network, and (by chance) will also have restricted competence to exert active influence on some partial operating parameters as the scope of the virtual network concerns.

Virtual networks will obtain particularly great importance in those cases, when services of public communications networks are used to build up a private network in part or in its entirety. The constituents of the public network, designated to form part of the given private network at the same time, can be regarded as a virtual private network (VPN). In such cases, operators of the private network would like to manage their own communications system as a whole, including those parts being utilized as a VPN.

Standard network management services that can be provided for the customers by Public Data Networks are called Customer Network Management (CNM). The framework of CNM is defined in ITU-T Recommendations X.160–X.162 [31–33]. The given CNM-related ITU-T standards define a set of services and the way of their implementation that may enable the user to manage those parts of its communications network which are actually made available by public service providers. ITU-T Recommendation X.160 defines the architecture for CNM for public data networks. Recommendation X.161 defines services for CNM, whereas management information for CNM service is specified in Recommendation X.162. Furthermore, ITU-T Recommendation M.3010 provides an opportunity for private network operators to realize the concept of CNM in the scope of TMN.

As described in Recommendation M.3010, TMN functionality offered by service providers may be accessed through CNM services, these services being realized as management interactions between users and TMN, or between TMNs. In both cases, reference point CNM (the reference point between the

customer and the service provider) defined in Recommendation X.160 has to be realized by TMN reference point x, as defined in M.3010.

Recently, efforts have been made to realize the CNM concept in association with high-speed public data networks [34–36].

3.4.5 Management of SDH/SONET

3.4.5.1 A Brief Survey of SDH and SONET Communications Technologies

Synchronous Digital Hierarchy (SDH) implies the concept of a powerful broadband telecommunications system, defined by a number of international standards, issued by CCITT/ITU-T. The first recommendations, including principles of SDH, were issued by CCITT in 1988 (CCITT G.707, G.708, G.709). SDH is the counterpart of Synchronous Optical Network (SONET), the predecessor of SDH. SONET was originally developed and proposed by Bellcore (U.S.). It is now standardized by ANSI, the base standard of SONET being ANSI T1.105. The concepts and architectures of SDH and SONET are highly similar (SONET may now practically be regarded as a subset of SDH), but there are also differences, first of all concerning standard transmission bit rates defined in the respective signal hierarchies. Recently, close coordination exists between U.S. SONET and international SDH standard bodies.

SDH and SONET are advanced telecommunications technologies — one of their most important features is that they effectively support network management. They have significant advantages over traditional PCM-based, Plesiochronous Digital Hierarchy (PDH) digital data transmission systems.

3.4.5.2 SDH/SONET and TMN

The principles of managing SDH/SONET networks by means of a TMN do not differ from those having been described elsewhere in Section 3. Notwithstanding, regarding SDH/SONET network management, it is important to emphasize the following significant circumstances:

- SDH/SONET are the first communications technologies that have been designed with particular attention to network management aspects;
- In the course of developing TMN standards and among the most important, up-to-date communications technologies, SDH/SONET are the first to have reached an appropriate degree of standardization, allowing a comprehensive TMN network management system capable of meeting the users' actual requirements (of course, all these are prevalent within the unquestionable limits of the present state of TMN standards; see also Sections 3.4.1, 3.4.3.3, and 3.4.3.5.2);
- Application of TMN network management is not only supported by the streamlined systems architecture of SDH/SONET and the existing TMN standards, but also by the fact that in practice, SDH/SONET system devices (add-drop multiplexers, cross-connects etc.) are actually equipped with the appropriate TMN interfaces, hence they are now capable of handling Q protocols and of implementing TMN network management;
- The complexity and high performance of SDH networks obviously require the facilities of TMN network management, and as a consequence, TMN network management is almost the only technology being applied to manage SDH networks in practical implementations [29, 37, 38].

3.4.5.3 Transmitting Management Information Through an SDH/SONET Network

As already mentioned in Section 3.4.3.2, in SDH/SONET networks, management information is usually transferred via the managed network itself; that is, DCN of the management system is actually a part of the network as a whole.

The most important factors of the SDH/SONET technology with respect to TMN network management are:

- The SDH frame structure allows management information to be transferred in data bytes D1, …, D12 of the Section OverHead (SOH in the STM-1/OC-3 frame) alongside the network;
- The characteristic features of SDH/SONET networks (high reliability of the network elements in itself, the availability of rapid and automatic "self-healing" protection mechanisms in redundant

ring topologies and the applicability of error recovery in meshed networks, allowing network communication most likely to be restored in the event of failures) make SDH/SONET networks highly suitable for realizing the DCN of the management system in the managed network itself, without impacting safety of operation (see also Section 3.4.3.4.2).

Getting a nearer view of the SDH/SONET frame, we can find that the Section OverHead (SOH) area of the STM-1/OC-3 SDH/SONET frame contains (and transfers) a number of data bytes, being characteristic of network operation and playing a significant role in network management.

In the SOH, parity control bytes B1 and B2 indicate bit errors of the generator and the multiplexer section, respectively, whereas bytes D1, …, D12 hold processed management status information. (Each of the bytes in the STM-1/OC-3 SDH/SONET frame represents a communication channel with a bit rate of 64 kbit/sec. It means bytes D1–D12 ensure a total communication capacity of 768 kbit/sec for network management.) The communication channel represented by management bytes D1–D12 can transfer management information from node to node (e.g., from add-drop multiplexer to add-drop multiplexer) along the network itself; furthermore (after the necessary protocol conversions), this information can be "tapped" at the standard Q interface of any node equipment and forwarded to the operations system of the network management center.

The status and alarm messages carried by the SOH can be accessed at the "S" monitoring points of the multiplexers and regenerators. The information made available at points "S" does not appear in an adequately structured form. Therefore, it has to be converted into object-oriented messages by a conversion function included in the given device. The object-oriented messages will then be returned to the STM-N aggregate transfer module and transferred in the SOH from node to node in the network. In order to get local access to these messages, it is necessary to have a standard "Q" interface that performs the necessary protocol conversion and signal interfacing. Output of the "Q" interface may then be connected for example to the local operations system of the network management center. Note, that multiplexers and cross-connects generally do have "Q" interfaces, but regenerators do not.

Figure 3.4.7 illustrates the scheme of the management data flow in the multiplex section of an SDH network.

3.4.5.4 Structure of Managed SDH/SONET Networks

3.4.5.4.1 General Considerations

The physical architecture of a managed SDH network accords with the principles described in Section 3.4.3.2.

It stands to reason, however, that designing a SDH/SONET network management system should be performed, in particular, consideration of the design aspects and architecture of the managed SDH/SONET network itself.

Furthermore (basically as the opposite of the requirement stated above), it is easy to see that, in practice, safety and reliability of the designed SDH/SONET network will essentially be determined by the structure of the pertaining network management system.

From the arguments stated above we can conclude that designing a SDH/SONET communications network and the related network management system should be performed in a close interrelation.

In the case of a simple network architecture, the network management system is also relatively simple. As an example, Figure 3.4.8 shows the scheme of a TMN-managed SDH ring consisting of four network nodes. Operation of the illustrated network management system can briefly be described as follows.

The SDH ring is managed by a TMN. The DCN of the management system is realized by the available (SOH) management channels of the SDH ring, the OS (i.e., the network management center) is represented by a single computer. (This computer also includes the operator's workstation.) The management center is connected to one of the four nodes of the SDH ring through a single "Q" interface.

Taking a more complex network architecture, the structure of the TMN network management system will also be more complex, and its operation will require more complex algorithms. To this topic, the following remarks can be added.

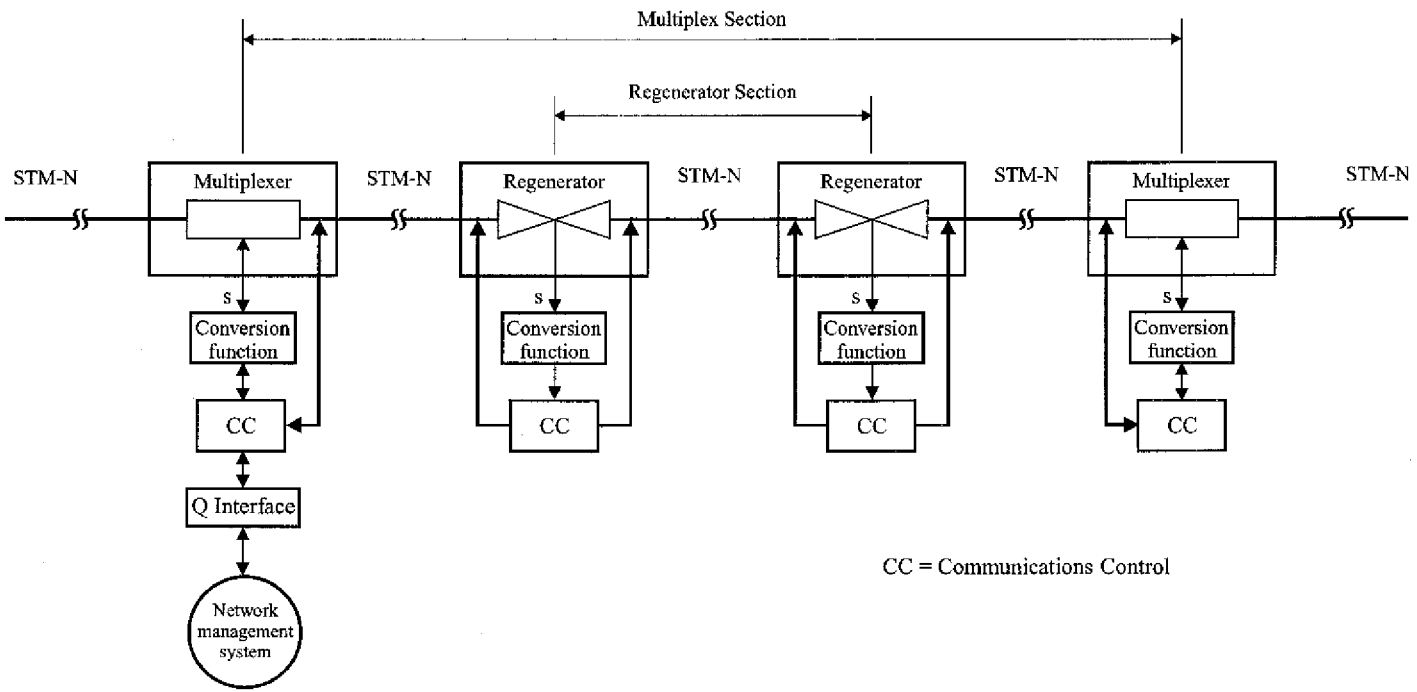

FIGURE 3.4.7 Management data flow in a SDH multiplex section.

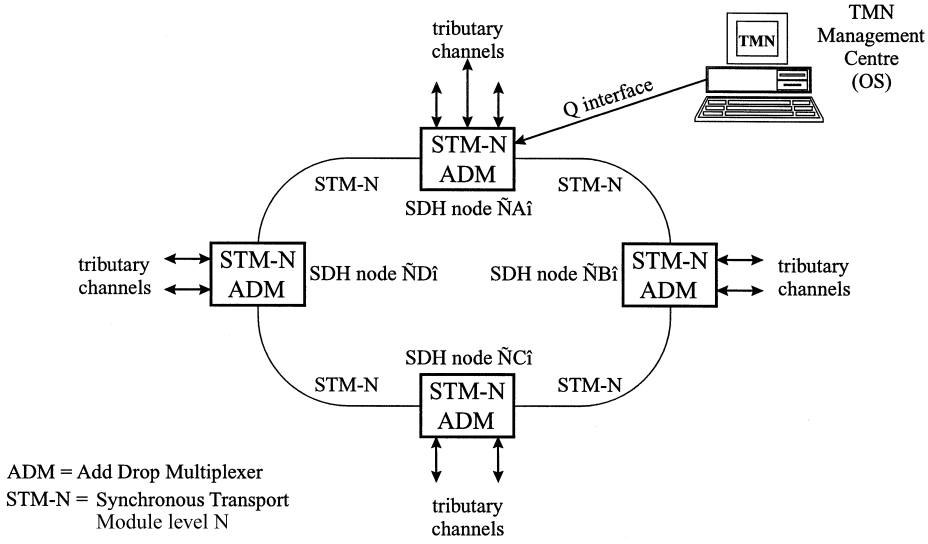

FIGURE 3.4.8 Scheme of a managed SDH ring.

- The "self-healing" protection mechanisms, frequently applied in SDH/SONET rings (aimed at retaining error-free communication in the case of a single failure) are based upon automatic reactions of the network elements and will function without the intervention of the network management system. The task of the TMN is simply to record the occurred events in such cases.
- In meshed networks, it is the TMN's responsibility to select and set up alternative routes in order to eliminate the possible consequences of any network malfunctions. Consequently, management algorithms required in meshed networks are generally more complex than those applied in simple self-healing rings.
- In the course of establishing management systems for hierarchical and/or interconnected networks, the TMN system(s) and the management algorithms must be designed with great care. Very likely, applying more than one OS will be necessary in this case. If so, tasks, access rights, control privileges, and ways of interworking the different OSs (management centers) should be carefully and clearly defined.
- Designing a network management system will be even more difficult if several networks of different topologies, those built up using different network technologies (SDH-PDH) or containing different vendors' equipment should be managed at the same time and in coherence.

3.4.5.4.2 Hierarchical Networks

TMN enables us to build up network management systems hierarchically, according to the eventually layered architecture of the network to be managed.

Consequently, a network management system may involve separate network management centers for (as an example) each of the

- Backbone
- Regional
- Sub-regional or local

network layers.

More than one management center may belong to one and the same hierarchy layer (e.g., as a stand-by safety reserve, if necessary); the centers may connect with the DCN at more than one point in order to increase reliability. The number of the operators' workstations may also be more than one. Moreover,

some workstations can be installed at remote locations, separated from the network management center. In this case, the workstation has to be connected to the DCN through a standard "F" interface.

The tasks and privileges designated to the management centers of the different hierarchy layers (rights to access data, privileges to control network elements, and to alter network configuration in the different network layers) as well as the rules of how stand-by management centers should take over the functions of network management in case of emergency have to be determined during system design.

Figure 3.4.9 shows the scheme of a hierarchical TMN system managing a layered SDH network.

The illustrated network consists of three layers: the national, regional, and local network layers. Each of them is managed by a separate management center; these in turn are connected to their respective network layers at two network nodes, using two separate communication channels and standard Q_3 interfaces. (Taking a practical example, one of the duplicated interconnections may be realized through the managed network and the second one, in order to increase safety, through an independent communication channel.)

Furthermore, the management centers are interconnected via appropriate communication channels as well, using standard X interfaces. (These interconnections may also be realized through the managed network via its available channels reserved specially for this purpose.)

3.4.5.4.3 Managing PDH Network Elements

In practice, network management often has to be extended to some non-SDH system components, such as elements or segments based upon PDH technology, being interfaced with the managed SDH/SONET network. This may be more necessary so that PDH payload may be mapped into the virtual containers of the SDH multiplex structure.

However, it will raise the difficulty that manageability of PDH network elements is fairly limited, both in principle and in practice. This is partly because there is a poor channel capacity available for transferring management information in a PDH network, partly because PDH systems technology does not support dynamic configuration management functions, and partly because PDH network elements cannot handle Q protocols. Management of PDH is not even supported by the existing TMN standards at present. According to Recommendation M.3180, plans exist to cover PDH transmission equipment concerning network-level information model of TMN.

Nevertheless, as up-to-date network management systems are capable of processing element-level network management information in different, non TMN-compatible formats (e.g., in ASCII code), managing PDH network segments with limited functionality is still feasible in practice.

3.4.6 TMN and GIS (Geographic Information System)

The benefits of GIS could be applied to support business processes of telecommunications service providers. The benefits of use are obvious considering that telecommunications services are end-to-end related and include a number of geographical domains, and multiple providers with different networking infrastructures. Managing physical and logical infrastructures requires very accurate documentation. State-of-the-art documentation systems work with digital maps that can be acquired from third-party vendors.

Network management tasks can be defined and supported in various TMN layers. The level of details and the type of information required differ by the layer. The service and business management layers may be using three-dimensional applications that may not be required in the network and element management layers.

The applicability of GIS will be evaluated for three TMN layers. The functional relation of TMN and GIS is shown in Figure 3.4.10.

3.4.6.1 Network Management Layer

The graphical presentation and documentation of physical networks belong to the traditional tasks of CAD/CAM applications. Principal documents include cables (earth and air), cable ducts, vaults and shafts, telco buildings, location of equipment, allocation of ports to wire pairs, and wire pairs to cables,

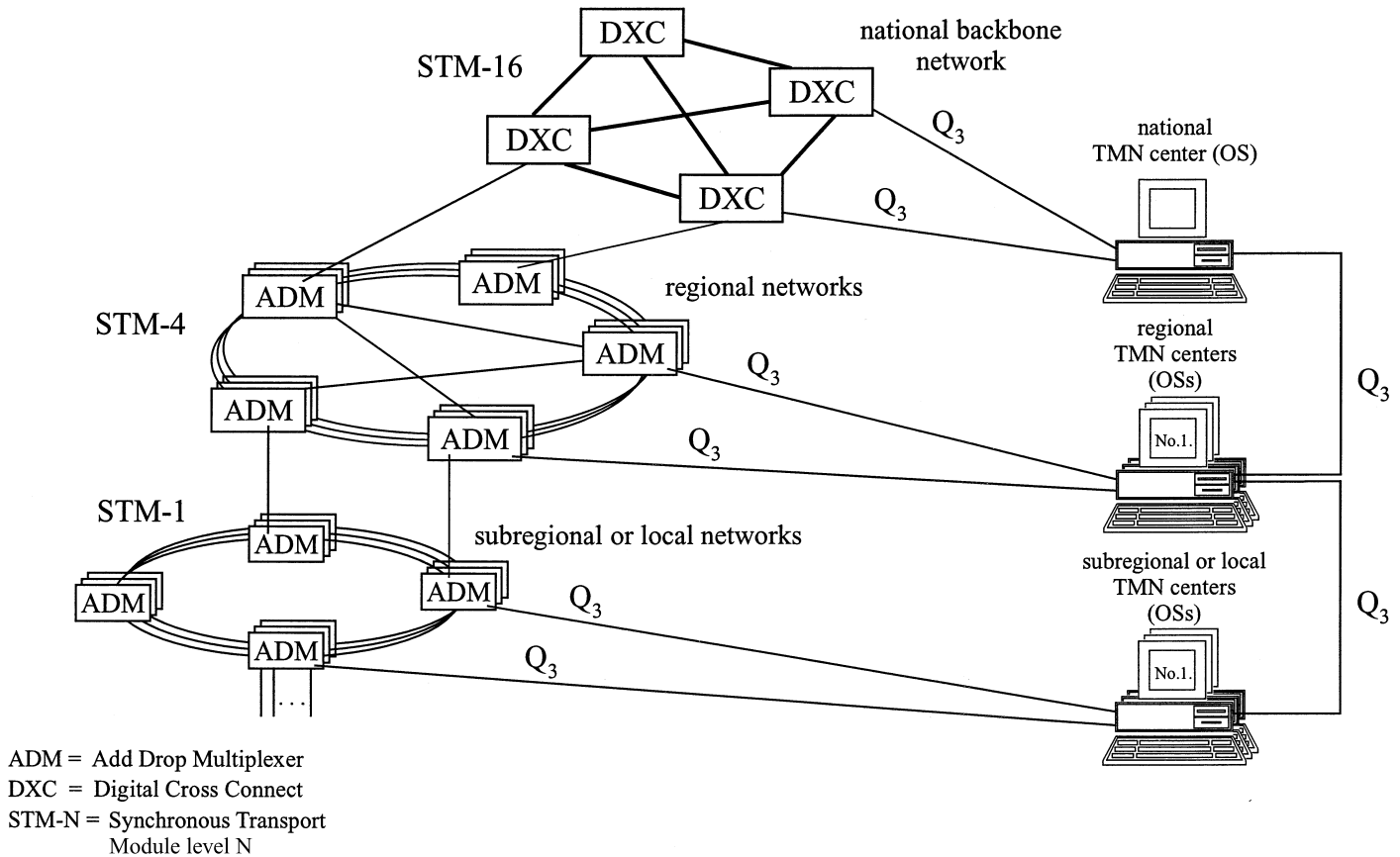

FIGURE 3.4.9 Management of a hierarchical SDH network.

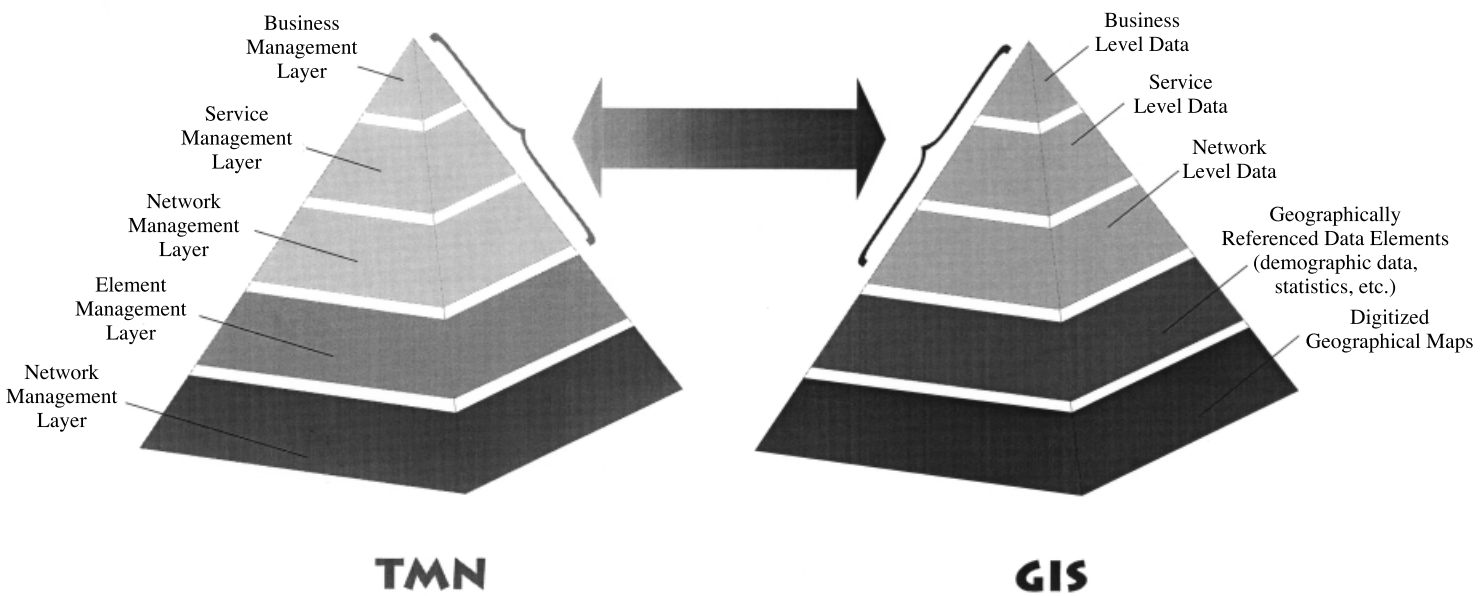

FIGURE 3.4.10 Functional relation of TMN and GIS.

respectively; in addition, other technical attributes of components have to be maintained as well. Details of component attributes are expected to be maintained in the Network Management Layer.

GIS applications are extremely important for wireless communication networks, such as mobile phone systems, URH, and satellites, in analyzing wave diffusion models. Models also include area of coverage and geography of selected areas. Inventory items are part of static configuration management.

Basically, there are two types of configuration management solutions. Dynamic configuration management means on-the-fly configuring of routes on the basis of actual real-time status and performance data of facilities and equipment. It is obvious that GIS applications cannot support this management function by its capabilities of three-dimensional presentation services. Static configuration management includes the logical configuring and changing of routes, equipment configurations, or both; maintaining types and location of equipment; maintaining both physical and symbolic names and addresses; also, other attributes may be maintained and information displayed. In case of relative stability of configuration parameters, GIS applications may be successfully used for displaying logical and physical routes of the network on top of various digital maps.

GIS applications contribute successfully to improve performance, fault, and security management. Performance metrics, load profiles, eventual faults and outages or security violations can be positioned and displayed on top of accurate digital maps. This kind of presentation helps to estimate and quantify impacts on customers and on service areas in real time with the results of expedited problem resolution, load balancing, and security protection.

A GIS is also a great help with preventive and reactive maintenance functions. GIS can pinpoint the location of faults with high accuracy, can highlight the optimal route for workforce to the site, and can search for and assemble the appropriate support documentation.

3.4.6.2 Service Management Layer

Managing services requires accurate data on service configuration, customer level quality metrics and actual metrics agreed upon in Service Level Agreements (SLA). GIS can assist to retrieve information on load profiles, fault summaries, network extensions, and visualization of service areas very fast; also, the correlation of this information with subscribers, contracts, subcontracts, service level agreements, bills, and bill presentment can be done in near real time. The result is that the quality of customer care is improved and customer satisfaction grows.

3.4.6.3 Business Management Layer

Business management requires a number of strategic and tactical decisions on behalf of top management. This is the area where Enterprise Resource Planning (ERP) can be utilized very effectively. It can support decision on complex and inter-related issues, visualization on the basis of GIS may help. Visualization examples may include the view of network segments, the position and size of existing service areas, peering with other service providers income statistics by service areas, statistical background data to support marketing, demographic data by service areas, and expected level of consumption by regions. Most of these data are not available in traditional management and documentation systems. GIS and data warehousing in combination can significantly offer better support to top management of service providers and also of enterprises.

In summary, the meaningful collaboration between TMN, GIS, and ERP (Figure 3.4.11) will ensure a higher quality of information exchange and presentation for service providers.

3.4.7 Trends of Evolution of TMN

The next steps of the evolution of TMN will most likely be driven by the following factors, being characteristic of the global situation of the telecommunications industry early in the 21st century:

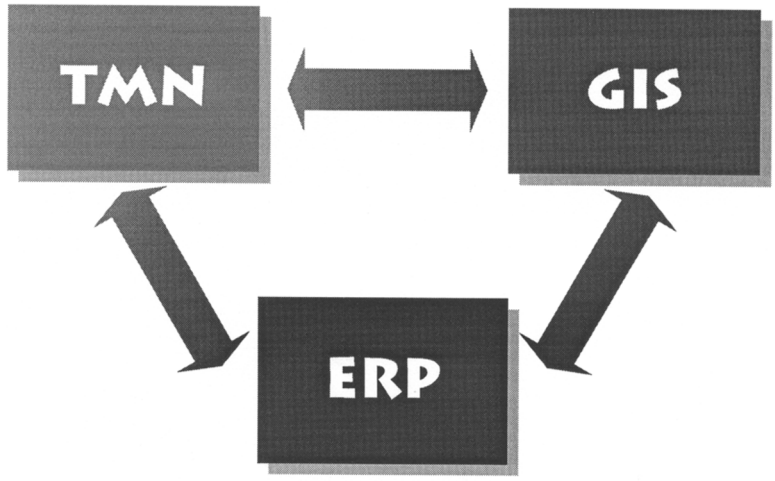

FIGURE 3.4.11 Possible interrelation of TMN, GIS, and ERP systems.

- The growing importance of business aspects of telecommunications network operation, such as rentability of investments, costs, profitability and competitiveness of network services, etc. [39];
- The evolution of the broadband telecommunications technology and the growing prevalence of the most up-to-date network architectures, such as ATM, intelligent networks (IN) [40–42] etc., in both the public and corporate or private network domains;
- The proliferation of the available hardware/software tools in the world of telecommunications;
- The perpetually growing sphere of telecommunications services and their applications, such as mobile services, video and multimedia applications, the increasing area of Internet utilization, such as on-line shopping, etc.;
- The globalization of the world of telecommunications, the integration of LANs and WANs, vanishing of the borders between public and private network domains, voice, data, and video transmission systems, etc.

As concluded from all these general aspects, the following changes probably will occur in the evolution of TMN, concerning its features and functionality:

- Focus of network management will shift from the network element and network-centric view to the service management first, and thereafter to the business management level — it means the relevant standards will be elaborated and the appropriate solutions will gradually appear on the market [39];
- The importance of the end-to-end management and customer management will increase (see Section 3.4.4.2).

Reflecting these functional requirements on the platform of software technology, actual tasks will arise (and have arisen) which have to be completed, such as:

- Defining a standard distributed object model meeting the requirements of service management for TMN (see Section 3.4.3.5.2);
- Defining standards and developing actual solutions for managing up-to-date telecommunications technologies presently not covered by TMN (first of all, as ATM concerns);
- Integrating TMN with the most important foreign platforms in the field of network management such as SNMP (see Section 3.4.4.1);
- Mapping TMN into some more recent telecommunications architectures (such as TINA and IN) theoretically being able to be integrated with TMN (see Section 3.4.3.5.2);

- Integrating TMN with some external information systems being operated outside the scope of telecommunications having however close functional connections with it, such as integrated corporate business management systems (see Section 3.4.3.3) or (representing one of the most recent challenges for TMN) Geographic Information Systems (GIS) [43].

As a result of present and future developments, the field of possible TMN solutions allowed by the whole set of the relating standards will be more and more complex. This situation sets a great task on network operators and/or system designers having to cope with the problem of specifying an appropriate network management system to effectively meet the requirements of their special application. It means design and specification of an actual TMN solution should be performed by selecting the relevant TMN functions from the "endless" set of standard possibilities, not simply declare or require compliance with "standard TMN" [39].

References

[1] CCITT Recommendation M.3010 (1996), *Principles for a Telecommunications Management Network.* International Standard, May, 1996.

[2] Glitho, R. H., Hayes, S., Telecommunications Management Network: Vision vs. Reality, *IEEE Communications Magazine,* 47, March 1995.

[3] Sidor D. J., Managing Telecommunications Networks Using TMN Interface Standards, *IEEE Communications Magazine,* 54, March 1995.

[4] Towle, T., TMN as Applied to the GSM Network, *IEEE Communications Magazine,* 68, March 1995.

[5] Fowler, H. J., TMN-Based Broadband ATM Network Management, *IEEE Communications Magazine,* 74, March 1995.

[6] Mader, W., Ferraris, G., Huber, M. N., Lehr, G., Müllner, W., Rotolo, S., Management of Optical Transport Networks, *NOC '97 (European Conference on Networks and Optical Communication),* Vol. 3., 111, 1997.

[7] Rose, M., McCloghrie, K., *Structure and Identification of Management Information in TCP/IP Based Internets,* RFC1155, May, 1990.

[8] Case, J., Fedor, M., Schoffstall, M., Davin, J., *Simple Network Management Protocol,* Internet RFC1157, May 1990.

[9] Case, J., McCloghrie, K., Rose, M., Waldbusser, S., *Structure of Management Information for version 2 of the Simple Network Management Protocol (SNMPv2),* Internet RFC1442, April, 1993.

[10] Stallings, W., *SNMP, SNMPv2 and CMIP — The Practical Guide to Network Management Standards,* Addison-Wesley, 1993.

[11] Terplan, K., *Effective Management of Local Area Networks,* McGraw-Hill Inc., 1992. Chap. 4.

[12] ITU-T Recommendation X.200 (1994), ISO/IEC 7498-1, *Information Technology — Open Systems Interconnection — Basic Reference Model: The Basic Model.* International Standard, July 1994.

[13] CCITT Recommendation X.700 (1992), *Data Communication Networks — Management Framework for Open Systems Interconnection (OSI) for CCITT Applications.* International Standard, September 1992.

[14] CCITT Recommendation X.701 (1992), ISO/IEC 10040, *Information Technology — Open Systems Interconnection — Systems Management Overview.* International Standard, January 1992.

[15] CCITT Recommendation X.710 (1991), *Common Management Information Service Definition for CCITT Applications.* International Standard, March 1991.

[16] CCITT Recommendation X.711 (1991), *Common Management Information Protocol Specification for CCITT Applications.* International Standard, March 1991.

[17] CCITT Recommendation X.722 (1992), ISO/IEC 10165-4, *Information Technology — Open Systems Interconnection — Systems Management Guidelines for the Definition of Managed Objects.* International Standard, January 1992.

[18] CCITT Recommendation X.720 (1992), ISO/IEC 10165-4, *Information Technology — Open Systems Interconnection — Structure of Management Information: Management Information Model.* International Standard, January 1992.

[19] ITU-T Recommendation X.901-4, *Information Technology — Open Systems Interconnection — Open Distributed Processing — Reference Model.* International Standard.
[20] ITU-T Recommendation X.703, *Information Technology — Open Systems Interconnection — Open Distributed Management Architecture.* International Standard, January 1992.
[21] Network Management Forum, Forum Library, Volume 4: *OMNIPoint 1, Definitions,* Issue 1.0, August 1992.
[22] Network Management Forum, *OMNIPoint Integration Architecture — Delivering a Management System Framework to Enable Service Management Solutions,* Technical Report, Issue 1.0, August 1994.
[23] OMG Group, *The Common Object Request Broker: Architecture and Specification.* Technical Report, December 1993. Revision 1.2.
[24] Manley, A. Thomas, Evolution of TMN Network Object Models for Broadband Management, *IEEE Communications Magazine,* 60, October 1997.
[25] Hashida, Y., Towards Service Operation and Management Technology, *IEICE Transactions on Communications,* E-80-B, 790, June 1997.
[26] Kong, Q., Chen, G., Integrating CORBA and TMN Environments, *NOMS '96 (IEEE Network Operations and Management Symposium),* Vol. 1, 86, 1996, Kyoto.
[27] Strick, L., Wittig, M., Paschke, S., Meinköhn, J., Development of IBC Service Management Services, *NOMS '96 (IEEE Network Operations and Management Symposium),* Vol. 2, 424, 1996, Kyoto.
[28] Larsson, R., Ferrari, L., Use Case Driven Integration of TMN Interface Specification and Application Design, *NOMS '96 (IEEE Network Operations and Management Symposium),* Vol. 2, 434, 1996, Kyoto.
[29] Dr. Ku, B. S., Use of TMN for SONET/SDH Network Management, *NOMS '96 (IEEE Network Operations and Management Symposium),* Vol. 2, 454, 1996, Kyoto.
[30] Motomura, K., Nakamura, N., Aibara, T., Integrated Platform for CMIP-Based and SNMP-Based Management, *IEICE Transactions on Communications,* E-80-B, 861, June 1997.
[31] ITU-T Recommendation X.160 (1994), *Data Networks and Open System Communications Public Data Networks, Maintenance — Architecture for Customer Network Management Service for Public Data Networks.* International Standard, July 1994.
[32] ITU-T Recommendation X.161 (1995), *Data Networks and Open System Communications Public Data Networks, Maintenance — Definition of Customer Network Management Services for Public Data Networks.* International Standard, April 1995.
[33] ITU-T Recommendation X.162 (1995), *Data Networks and Open System Communications Public Data Networks, Maintenance — Definition of Management Information for Customer Network Management Service for Public Data Networks to be used with the CNMc Interface.* International Standard, April 1995.
[34] Aronsheim-Grotsch, J., Berkom D. T., Customer Network Management, CNM, *NOMS '96 (IEEE Network Operations and Management Symposium),* Vol. 2, 339, 1996, Kyoto.
[35] Park, J. T., Lee, J. H., Hong, J. W. K., Customer Network Management System for Managing ATM Virtual Private Networks, *IEICE Transactions on Communications,* E-80-B, 818, June 1997.
[36] Yamamura, T., Tanahashi, T., Hanaki, M., Fujii, N., TMN-based Customer Network Management for ATM Networks, *IEEE Communications Magazine,* 46, October 1997.
[37] Kunieda, T., A Synchronous Digital Hierarchy Network Management System, *IEEE Communications Magazine,* November 1993.
[38] Yamagishi, K., Sasaki, N., Morino, K., An Implementation of a TMN-Based SDH Management System in Japan, *IEEE Communications Magazine,* 80, March 1995.
[39] Willets, K. J., Adams, E. K., TMN 2000: Evolving TMN as Global Telecommunications Prepares for the Next Millennium, *IEICE Transactions on Communications,* E-80-B, 796, June 1997.
[40] Mizuno, O., Shibata, A., Okamoto, T., Niitsu, Y., Models for Service Management Programmability in Advanced Intelligent Network, *IEICE Transactions on Communications,* E-80-B, 915, June 1997.
[41] ITU-T Recommendation I.312 (1992), *Principles of Intelligent Network Architecture.* International Standard, October 1992.

[42] ITU-T Recommendation Q.1200 (1993), *Q-series Intelligent Network Recommendation Structure.* International Standard, March 1993.
[43] Edwards, G., Examining GIS In Relation To The TMN Model, *presented at GIS In Telecoms & Utilities (Conference arranged by IIR Telecoms & Technology)*, September 1997, London.

3.5 TINA

Takeo Hamada, Hiroshi Kamata, and Stephanie Hogg

3.5.1 Introduction[1]

The increasing demand for broadband and sophisticated services, such as universal personal communications and mobile multimedia services, calls for a more flexible and distributed information architecture for advanced telecommunication services and management. Telecommunication Information Network Architecture (TINA) is a software architecture which has been designed by TINA-C since 1995 [1], and a set of architecture and interface specifications has been published.[2] Historically, TINA evolved from two predecessors: IN and TMN. As a service architecture, TINA inherited the concept of network support for telecommunication services from IN. As a management architecture, TINA inherited the concept of distributed, object-oriented representation of network resources from TMN. As its design methodology, TINA adopted ODP principles [2], and also developed its own specification languages for the description of information [3] and computational viewpoints [4,5].

As a result of these architectural heritages, TINA can be understood both from the network service and the service management points of view. For the users of TINA, TINA is a distributed object-oriented system, into which multimedia network services and user applications can be plugged. In IN-like network services, e.g., multimedia version of 800 service, the service components (objects) are provided by the network or the service providers. In collaborative workspace applications, the basic call model and the multi-party session control are provided by the TINA service layer. In either case, TINA services are supported by built-in service management functions in its service architecture [6], and by the service components supporting the service architecture [7].

In this chapter, we focus mainly on the management aspect of TINA, due to the following two reasons; (1) built-in support for management is essential for the integration of call control and management in TINA, and (2) support for telecommunication management needs distinguishes TINA from other enterprise data-communication-oriented network architectures. The TINA management architecture has many facets. It is a set of common goals, principles, and concepts covering the management of services, resources, and parts of the DPE. Due to its broad nature, management is one of the most challenging areas in TINA. TINA imposes the following design goals on the management architecture:

- *Object oriented and distributed:* TINA management architecture takes advantage of a distributed processing environment (DPE), which offers a natural distributed object-oriented environment for both resource and service management. It builds on the object-oriented and distributed approaches of telecommunication management network (TMN) [8] and Open Distributed Management Architecture (ODMA) [9].
- *Service-oriented:* TINA uses a more goal-driven design approach rather than a resource-driven one. In particular, the service management should guarantee quality of service (QoS) as an integrated part of the service.
- *Dynamic and flexible:* The service-oriented nature of TINA requires that resources need to be more dynamically assigned or configured, to support the flexibility required by TINA services.
- *Integrated:* Management is an integrated part of service provision, inherent in the various TINA layers (service, resource, and DPE).

[1]This article is a revised and updated version of a paper, An Overview of the TINA Management Architecture, *Journal of Network and Systems Management*, Vol. 5, No. 4, 1997, Plenum Press.

[2]All the TINA-C public documents are available from the TINA-C home page, http://www.tinac.com.

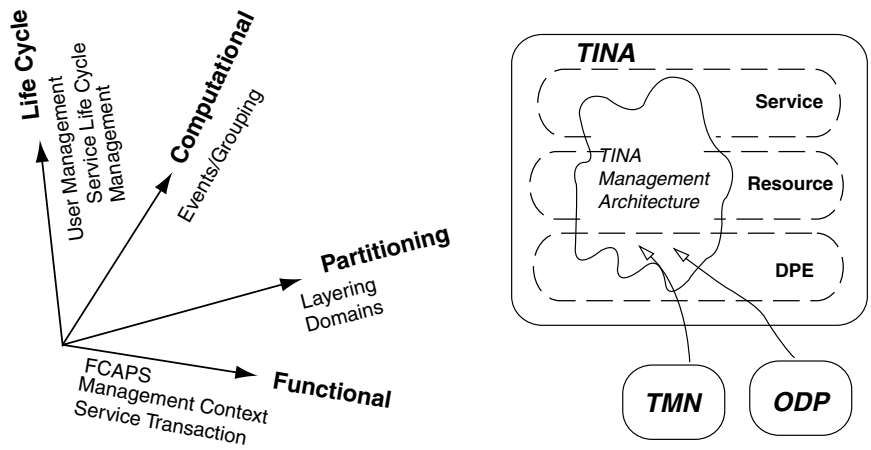

FIGURE 3.5.1 TINA management architecture and conceptual axes. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements,* 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

The TINA management approach consists of several related disciplines and builds upon other management approaches, such as TMN, ODP, and OmniPoint [10]. The session-oriented, multiparty, multidomain nature of TINA service architecture poses unique problems in its service management. To illustrate the TINA management architecture, this paper is organized around the following four conceptual axes (Figure 3.5.1):

- *Partitioning axis* (Section 3.5.2) represents layering and domain concepts. TINA is partitioned into three layers: service, resource, and DPE. The management architecture is likewise partitioned, into service, resource, and DPE management. It also supports the concept of domain, the management of domains.
- *Functional axis* (Section 3.5.3) represents FCAPS (fault, configuration, accounting, performance, security) functions. To support the FCAPS integrity of a service session, constructs such as management context and service transaction are provided;
- *Computational axis* (Section 3.5.4) represents computational support for management needs. These computational supports are mostly offered by DPE, but we will consider some TINA refinements to event management and grouping concepts; and
- *Life cycle axis* (Section 3.5.5) represents the life cycle issues, including service life cycle (SLC) management and user management, i.e., the life cycle management of consumers.

This chapter intends to give an overview of TINA management architecture, with the main focus on service management and supporting concepts. As a result, some major topics had to be left out. In particular, connection management, which is an important topic in TINA, is not discussed. Pointers to other important topics in TINA are given at the end of this chapter.

3.5.1.1 Relationship between TINA, TMN, and ODP

The relationship between TINA, TMN, and ODP deserves special attention. As we noted earlier, TINA inherited some of its architectural concepts from both TMN and ODP. In fact, there is often more resemblance than difference in TINA to TMN, ODP, and ODMA [9], an ODP-based management architecture. For example, TINA Network Resource Information Model (NRIM) [11] used many network element concepts originated in related TMN documents, such as G.803 [12] and G.805 [13]. There are, however, significant differences between TINA and TMN, as discussed below.

- *Session Concept:* TINA is session-oriented, i.e., many of the constructs and service components are dynamically bounded within the lifetime of service sessions. As a consequence, much of service management activities are also service session-oriented, which is in stark contrast to TMN.

- *Separation between information and computational viewpoints:* By adopting ODP principles, TINA maintains clearer separation between information and computational viewpoints than TMN. In TMN, managed objects (MOs) correspond to an information viewpoint, and a computational viewpoint is represented by manager-agent roles and associated interfaces known as Q, X, and F interfaces. In terms of information-to-computational mappings, therefore, many-to-one mapping is dominant in TMN. In TINA, information objects can be mapped more freely to computational objects. For example, one information object may be mapped to one computational object (one-to-one mapping), or may be mapped to a collection of computational objects (one-to-many mapping). For further discussions on this issue, please see 3.5.4.1, *Computational Aspects of Distributed Management.*
- *Different computational model:* In TMN, GDMO and Common Management Information Protocol (CMIP) in effect defines the computational viewpoint of the TMN system. In TINA, CORBA-based DPE provides its computational viewpoint. Computational objects in TINA are described in CORBA IDL and TINA ODL, an extended version of CORBA IDL, which allows multiple interface objects.
- *Different engineering model:* The above difference in computational model results in different engineering platforms. Usually, TMN systems are CMIP-based, whereas TINA systems are CORBA-based.

Interoperability and migration between TMN and TINA are of serious practical concern. Interoperability and migration issues between TMN and TINA, and between IN and TINA have been extensively studied in the Eurescom P508 project [15,16]. For a recent development on the migration issues, the reader is referred to [14]. For the recent developments of TMN, a recent report [17] gives an excellent survey.

3.5.2 Partitioning Axis

The partitioning axis considers how management may be broken into more manageable problems. TINA supports two types of partitioning: layering and domains. These two concepts are independent. If layering is considered a "horizontal" partitioning, domains may be considered a "vertical" one.

3.5.2.1 Layering

Layering corresponds to the traditional distinction between the resource management and the service management. Resources are network capabilities and equipment. Resource management relates to the management of network resources. We define service management as all the management activities within the service layer. Service management is an inherent part of the service layer.

3.5.2.1.1 Service Management

Service management is becoming increasingly more important as the telecommunication industry evolves toward information and communication services. TINA views service management as a number of management activities, either inherent in its service architecture or provided by a set of management services, allowing the provision of services and their management by users and providers. Management services will use the same service structure as normal services or be based upon DPE and object services. Service management is therefore an intrinsic part of the service layer. TINA service management has the following common requirements:

- *Access management:* User access should provide security, allow customization, and other management capabilities. TINA service architecture [6] entails both access session and subscription management (Section 3.5.5.2.1) to provide required access management.
- *User service management:* Allow users to negotiate management requirements with service providers. A management context approach (Section 3.5.3.1) addresses this issue.
- *Flexibility:* TINA addresses an open market. It is expected that the market and price structure dynamically changes, reflecting today's competitive market. Service management must be flexible to satisfy those needs. This requirement is fulfilled by the management context concept (Section 3.5.3.1).

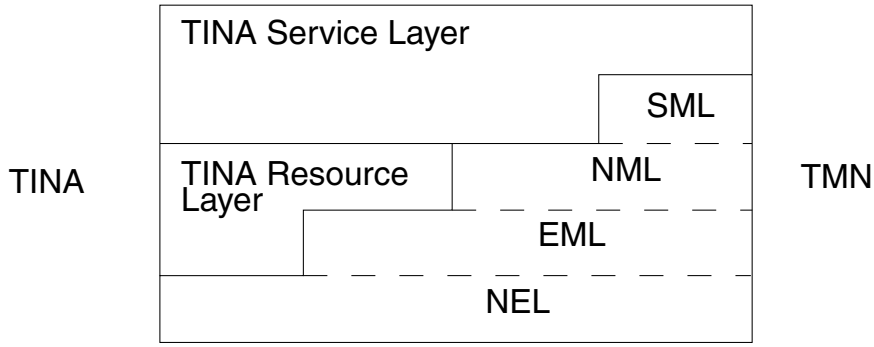

FIGURE 3.5.2 TINA Service Management Principle. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

- *Service provision:* Providers must be able to deploy and withdraw their services.This requirement is answered in the SLC management (Section 3.5.5.1).
- *Controllability:* Providers need to be able to enforce their management policies and to exercise controls over service instances in their domains. For instance, a provider may need to monitor or control the status of service instances to maintain and manage their quality. Management policy (Section 3.5.2.2.3) and life cycle management (Section 3.5.5.1) help satisfy these needs.
- *Multiple business domains:* TINA service management must interoperate across multiple business domains. It must offer a consistent and guaranteed service management quality across these domains. The service management architecture needs to be relatively independent from the business model and the relationships between business entities, as these may vary. The service transaction concept (Section 5.5.3.2) fulfills this requirement.
- *Different QoS levels:* TINA has to offer a guaranteed and more consistent service management quality, such as accountability and fault tolerance, than Internet. Internet technology is very accessible and constantly evolving; on the other hand, its resource usage tends to be wasted and its quality varies. It is the telecommunication provider's ideal to support different QoS requirements of the user in a guaranteed manner as an integrated part of service management (Section 3.5.3.2).

TINA takes a service-oriented approach, starting with service and business-related concerns with the aim of achieving an information and communication service architecture upon a DPE. This provides a natural basis for considering an integrated approach to services and management. In comparison, TMN takes a resource-oriented, bottom-up approach, where service and business requirements are considered last. TMN service management corresponds to a fixed set of functionality on top of the network management layer. Figure 3.5.2 compares the TINA and TMN management models.

3.5.2.1.2 Resource Management

TINA resource management is based on the following principles and requirements:

- *Layers:* TINA borrows layering concepts from TMN, but uses them differently. It does not consider the network element management layer, and combines the network and element management into a single layer that may be used recursively.
- *Domain management capability:* Providers need to be able to enforce their management policies and to exercise management control over resources in their domain (Section 3.5.2.2.3).
- *Policy based:* Unlike service management, resource management tends to be driven by the provider's management policy (Section 3.5.2.2.3) rather than management context (Section 3.5.3.1).

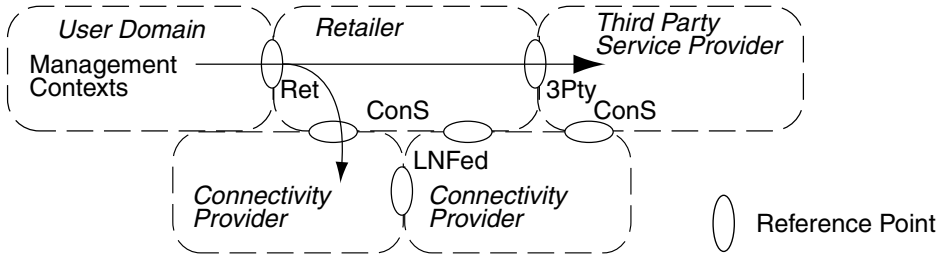

FIGURE 3.5.3 TINA business model and management context. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements,* 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

- *Multiple business domains:* The interoperability between multiple businesses in the resource layer is considered by management policy and federation (see Sections 3.5.2.2.3 and 3.5.2.2.4). The role of service transaction here needs further consideration.
- *Service layer support:* Provides management services to the service layer which allow information and communication services to set up connections, control their quality, and specify management support.This is part of the service and resource architectures.

3.5.2.2 Domains

TINA assumes that there will be a number of stakeholders acting in various business roles. Stakeholders are real-world commercial entities that provide services or communication resources. Business roles arise from analysis of a business model [6]. The NMF has performed similar analysis [19–21], though it focuses on separate business processes such as billing and trouble ticketing, rather than enterprise level requirements as they are expressed in the TINA business model.

3.5.2.2.1 Administrative Domains

An administrative domain is a portion of a TINA system which is submitted to a single stakeholder's ownership. TINA presumes that each stakeholder will manage and administer the policies and resources of its domain — giving rise to discrete domains that will be required to interoperate. For clarity, the following discussions treat each business role as occupying a separate domain associated with a single stakeholder. In reality, a stakeholder may participate in multiple roles and may administer one or more domains which are based on organizational considerations and need not be coincident with a single business role.

Figure 3.5.3 illustrates the relationship between the TINA business model and management context. Each stakeholder in the diagram constitutes a separate administrative domain, which contact each other through the reference points [18]. The recursive nature of management in TINA is apparent. In this example, the user domain contacts the retailer through the Ret reference point. Similarly, the retailer contacts connectivity providers through the ConS reference point. When a TINA service session is established across multiple administrative domains, management contexts are passed through reference points along with requests for the service session creation. Each context is specific to the service and the domain. For instance, the management context passing through ConS has to contain requirements for resource management.

3.5.2.2.2 Management Domains

A management domain can be modeled by an information object associated with certain management functionality such as accounting, security, or DPE management. Domains are similar to object groups ([22], Section 4.1), in that they both represent a set of objects. Unlike object groups, a domain does not usually provide explicit membership operations, since domain boundaries are usually based upon natural affinities between objects, such as the network resource topology, business stakeholder, or geographical area.

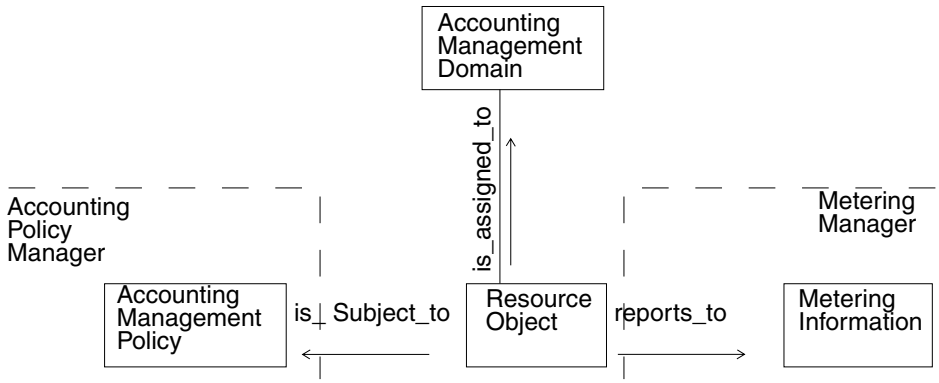

FIGURE 3.5.4 Accounting Management Domain information model. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

The domain concept allows a hierarchical composition of domains. The generic management domain can be specialized with a set of management functions (policies) to become a specific management domain, e.g., an accounting management domain. Domains are by no means disjoint. Not only may an object belong to two or three different management domains, these domains may also relate to each other in different ways. For example, two accounting management domains may belong to two respectively different security management domains, but they both may belong to a single DPE management domain. Each management domain may have its own information model. The information model will usually consist of policy, resource, and manager objects (Figure 3.5.4).

3.5.2.2.3 Management Policy

A management policy is a set of rules governing a particular management function in the domain that is associated with the policy. For example, an accounting management policy may prohibit accounting events from being transported across the domain boundary, due to security or enterprise considerations. This rule effectively forbids on-line billing. A management policy can address many issues, including:

- *Managers:* Management policy determines the number and types of managers deployed within the domain and which resources they manage. Managers responsible for ensuring management activities are in accordance with management policy.
- *Event management policy:* Event management policy dictates acceptable event management options, such as translation, forwarding, and duplication. Policy may limit options, even though they are physically supported, for security or performance reasons. Event management options may limit federation or interoperability between domains.
- *Security features:* Since management actions and events may govern resource access and billing, they are natural targets of attack. As TINA service and resource layers are built on open DPEs, security features are mandatory. For instance, non-repudiation security options may be required for accounting events.
- *Policy applicability rules:* A policy is a set of rules. Rule applicability may vary depending on how the management domain is operated. Each rule is associated with one of three different applicability values: *mandatory (hard)*, *optional (soft)*, or *negotiable*. These applicability values are also used to resolve conflicts when two management domains overlap.

Management policy governs a management domain. By contrast, a service transaction and its associated management context (Section 3.5.3) set management requirements for the delivery of an instance of a service. A management context can take effect only for the duration of the service transaction to which the management context is bound.

3.5.2.2.4 Federation of Management Domains

Federation is one use of the management domain concept. With the management policy and its applicability rules, two domain managers can determine whether policy rules can be resolved and the two domains federated. The federation of management domains can be seen as an extension of the third-party call control, giving extra freedom to the management services. There are two major benefits of federation:

- *Sharing of Management Responsibility* means that two federated domains share a part of the management responsibility, such that one of the domains can cover the other's task, adding some fault tolerance.
- *Sharing of Management Resources* means that two federated domains share some management resources which may reduce management cost overheads and allow third party management, for instance a retailer may provide management for small content providers.

3.5.3 Functional Axis

This axis deals with the functional aspect of the management architecture. The key concept in this axis is FCAPS management, which is a well-known acronym of OSI functional classification. The TINA management architecture uses management context and service transactions to support FCAPS management. In this section, we introduce management context and service transactions.

- *Management Context* is a (service) management contract between the user and the provider, which drives (service) management activities in the provider domain.
- *Service Transaction* is a construct to guarantee the integrity of the service session with respect to its FCAPS management. There is some similarity to the database transaction concept, though the purpose of service transactions is to manage the integrity of the service session, not the integrity of data.

Finally, general context configuration management (GCCM) is also explained to show how context can be set-up and negotiated in a TINA environment.

Relationship between management context and service level agreement (SLA) [23] deserves special attention. In essence, management context and SLA services the same purpose, that is, to set up an end-to-end service management guarantee. The difference is in their lifetimes and dynamics. Since TINA is session-oriented, management contexts are dynamically bound to service sessions, thus instances of management contexts and their lifetimes are within those of associated service sessions. Service level agreement, in contrast, is a longer term, more static agreement between operators.

3.5.3.1 Management Context

Each service management context (MgmtCtxt) represents a set of functionality requirements, associated with a particular FCAPS. In general, a typical service transaction is associated with a number of (FCAPS) management contexts. The details of the management context depend on the individual FCAPS management objective. For example, the accounting MgmtCtxt [24] consists of the following information:

- *EventManagementConfiguration* specifies delivery timing and the event channel to be used for accounting event delivery.
- *TariffStructure* is essentially a function. Provider calculates charge/charging rate from accounting event sequence.
- *BillingConfiguration* contains types of billing options such as on-line or shared billing.
- *RecoveryConfiguration* specifies recovery actions to take when the service transaction is aborted for some reason such as network failure.

Figure 3.5.5 illustrates the role of management context in the TINA service architecture. Whenever a service session is joined or started, a management context can be bound to it. The PA could request a preset management context, or the UA could use a default management context, or the user domain (via the PA) can request to negotiate or configure a management context. The resulting context is then bound to a service session, which corresponds to the service transaction explained in the following section.

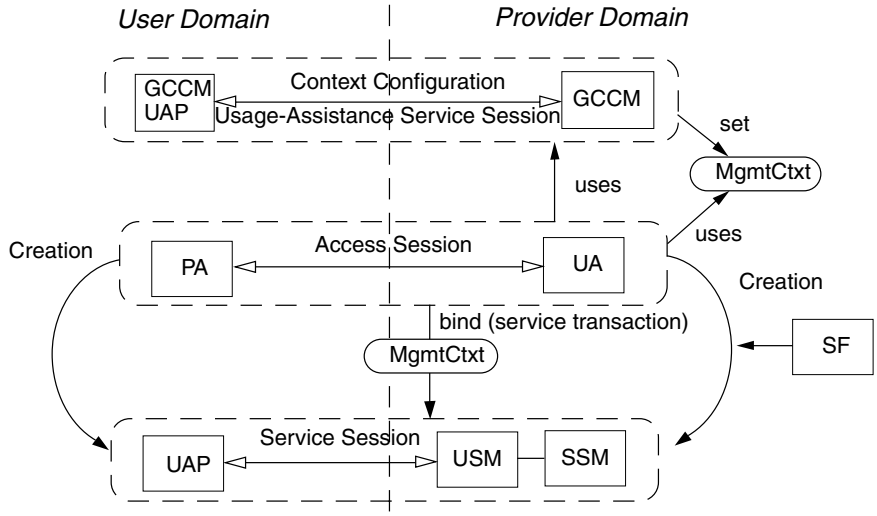

FIGURE 3.5.5 Role of management context in service architecture. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

A service session may be started without a preceding context configuration step. In many cases, a default context specified in the user profile, which is configured within the subscription management, may be sufficient. In all cases, the management context must be compatible with the provider's management policies.

3.5.3.2 Service Transaction

A service transaction is a construct which guarantees that its associated service session satisfies management requirements prescribed by the management contexts. With the service transaction distributing the management contexts, a consistent level of service management throughout multiple administrative domains is achieved.

Figure 3.5.6 illustrates the relationships between service session and service transaction. The service session has two participants, represented by the two respective user session managers (USMs) in the provider domain. Each service transaction is associated with a management context, MgmtCtxt 1 and MgmtCtxt 2, respectively. When service transaction 1 executes, MgmtCtxt 1 is interpreted and translated into resource level operations. The service session manager (SSM) activates the communication session manager (CSM), which activates metering activity for the following communication service session. The USM corresponding to the service transaction passes its notification interface to the metering manager, and the metering manager reports notifications to the USM as required. Requirements for service transaction 2 are met similarly. Note that the two service transactions share the SSM.

The service transaction represents a view of the service session local to the concerned user–provider relationship. The management context in particular represents the contractual agreement between the two parties within the service session. The service transaction is an information object, covering part of the user application (UAP), USM, and SSM. In a sense, it is a virtual framework, which guarantees the completeness of the management requirements.

Execution of a service transaction consists of three phases. Here is an example, using an accounting management context.

1. *Set-up* — The respective management contexts are interpreted by the SSM and the associated USM of the service session and necessary resources such as the accounting log record and event channels are reserved or assigned.
2. *Execution* — The UAP starts running. Following the accounting/billing specified in the accounting management context, accounting events may be logged or reported. If the service guarantees QoS,

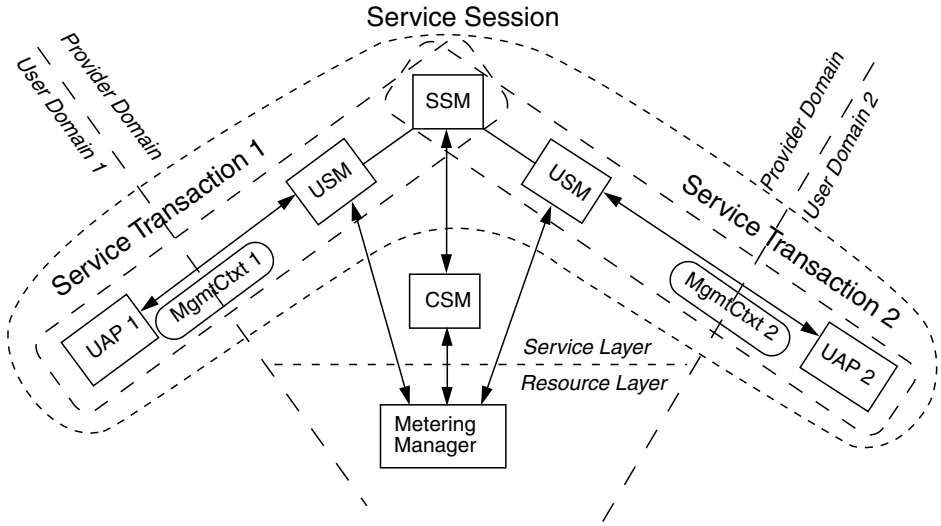

FIGURE 3.5.6 Relationships between service session and service transaction. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements,* 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

notifications from the performance management may be reported during this execution phase. If the QoS does not to meet the guaranteed quality, it may lead to an early termination of the service transaction.

3. *Wrap-up* — The UAP stops running. Reports from performance management or fault management can be summarized, and are then checked to see if the reported QoS has satisfied the guaranteed level. If it has, the service transaction concludes successfully. If it hasn't, some recovery actions, such as billing compensation, may be considered and actions may be taken by USM and SSM.

3.5.3.3 General Context Configuration Management

The concept of context is widely used in TINA service architecture in various forms. Examples include usage context, accounting management context (AccMgmtCtxt), and security management context (SecMgmtCtxt). The major purpose of these contexts is to associate additional information, represented by a context, with a service instance, so that particular service objectives are achieved. For example, the usage context is used to realize terminal mobility [25]. The management contexts (AccMgmtCtxt, SecMgmtCtxt, etc.) are used to achieve service management goals such as guaranteed accountability and quality of protection.

In any case, despite the difference in their usage and purposes, all these contexts are strikingly similar in their structure and the way they are configured in the service architecture. They need to be agreed between the user and the provider, therefore negotiation is often necessary. We introduce *general context*, which is a generalized class of context in the service architecture. The general context is conceived as a super-class of all the contexts used in the TINA service architecture.

3.5.3.3.1 The General Context Configuration Management Service

General context configuration management (GCCM) is positioned as a usage-assistance service in TINA service architecture. As such, GCCM service should be available across the service level reference points, including the Ret reference point. The objectives of the GCCM service are:

- *Semantics independent*: The GCCM service is semantics independent, which means that GCCM service is available for any type of context, but the semantics of the context needs to be interpreted by other semantic-dependent components used by the GCCM.
- *Negotiation*: The GCCM offers necessary and sufficient negotiation capability for general context management.

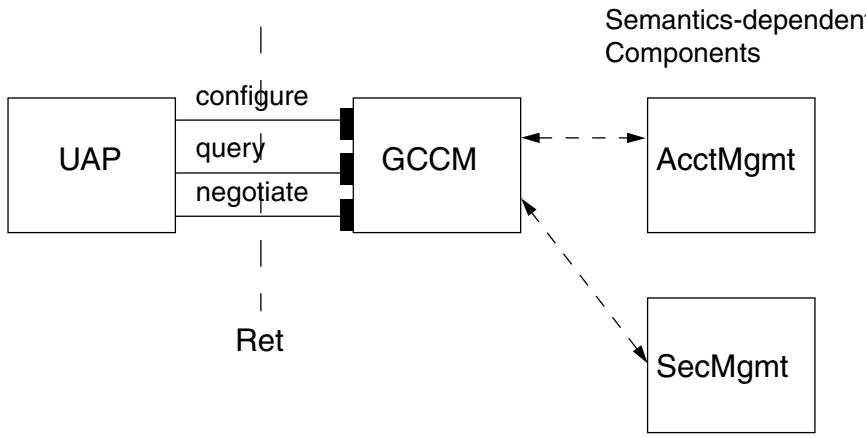

FIGURE 3.5.7 GCCM computational model. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

Figure 3.5.7 illustrates the computational model of GCCM. The GCCM object provides three interfaces:

- *Configure*: This interface is used when the UAP is allowed to set a context without negotiation, or when UAP asks the GCCM to set up a predefined default context. It can be used if a user is invited to a session, or if plug-and-play is preferred and negotiation is not possible.
- *Query*: This interface is used when UAP asks GCCM for the current contents of the contexts, or the life cycle stage (live, dead, when modified, etc.) of the contexts.
- *Negotiate*: This interface is used when UAP and GCCM must negotiate the context.

Although the semantic part of the context is not considered here, three distinctive categories can be recognized, as they are evolutionary steps of the general context concept.

- *Static:* The context consists of a fixed set of attribute–value pairs. The scheme of the context is thus statically fixed. Negotiation has a clear meaning in this case, since the goal of the negotiation is a set of values agreeable to both sides.
- *Dynamic:* The context is still mostly made of the attribute–value pairs, but the scheme is dynamically expandable, i.e., it is allowed to add a new attribute pair on the fly.[3] The negotiation process can be more complex, and more intelligence on both sides (UAP and GCCM) is necessary.
- *Executable:* The context is not only dynamic, but now it can be executable in the provider domain. It may resemble a program more than data, like a shell script or Java program. The context may contain an intelligent mobile agent, which is capable of negotiating for system resources directly with other service components in the provider domain.

Though the contexts conceived so far in TINA service architecture are mostly at the first two levels, it is envisaged that the mobile or intelligent agent type configuration management will be more prevalent in the future [30].

3.5.3.3.2 GCCM Service Usage

The GCCM service is used in a number of places, such as at service instantiation. The following are typical uses of the GCCM service:

- *Prior service subscription:* If a GCCM service is activated prior to the service subscription (or during service subscription modification), the resulting configured context will prevail as the default context, which all the subsequent service instances can use.

[3]For example, CORBA Context object (CORBA 2.0, Section 4.5) [9] belongs to this category.

- *At service instantiation:* A GCCM service may be accessed just before a service instantiation, to set a context effective only for the service instance. Once configured, the context can be bound to the associated service session.
- *During a service session:* If a GCCM service is activated during a service session, it may be used to modify the context of the service session dynamically.

The third case may not always be possible, since modification of the context may trigger additional operations, e.g., resource management, to satisfy requirements of the modified context. As a usage assistance service, it can be triggered using normal service requests between domains. The general context can be seen as a contract. To ensure that the contract is agreed upon and valid, it should be possible for the contract to be time-stamped, signed, and made nonreputable. By using security features available from DPE security [26] and service layer security [27, 28], options such as *Time stamp, Nonrepudiation,* and *Persistency* should be available to the user of GCCM service.

3.5.4 Computational Axis

The computational axis considers computational aspects of management. The computation viewpoint considers the decomposition of a system into a set of interacting objects that are candidates for distribution.

As TINA is built on DPE, the computation aspect is very important. It provides the basis for the structure of managed and managing systems. The DPE infrastructure provides basic services upon which a more complex management framework can be based.

3.5.4.1 Computational Aspects of Distributed Management

Previous attempts at considering the computation aspects of management have been limited. In TMN, information and computation viewpoints are not clearly distinguished. It effectively uses the Manager and Agent roles as its computational model. ODMA [9] builds on this approach by defining both managers and managed objects as computational objects. It tends to view this as a one-to-one mapping from information to computation objects.

TINA builds from the ODMA approach, but maps from information objects to computational interfaces (rather than objects) and allows more flexible mappings. A managed object may be mapped to one or more computational interfaces and an interface may represent one or more managed objects. Note that a TINA computational object can support both control (or usage) interfaces and management interfaces.

As managers, managed objects, or agents representing managed objects, computational objects can use DPE services, which may be based on generic CORBA common object service (COS) [29]. Interworking between DPE systems and legacy management systems based on Simple Network Management Protocol (SNMP) and Common Management Information Protocol (CMIP) is important, and NMF [31] and X-Open [32] are working on GDMO to IDL translations. These form the basis of interworking between DPE-based and OSI systems, enabling vertical integration by building TINA service management on top of legacy network management systems. For recent developments on interworking and migration issues see [14].

Computational objects themselves have to be managed, and the computational object management is supported by DPE services (Section 3.5.4.2), service life cycle (Section 3.5.5.1), and object groups. An object group is a "software package," or "subsystem," that contains one or more objects or object groups. It is built, installed, maintained, and managed as a unit. Object groups [22] can be used to structure and control usage and management functions. Unlike an object, however, an object group has internal structure that may be visible and can be distributed.

3.5.4.2 DPE Support Services

Typically a DPE offers a wide range of support services to aid system management. These include various life-cycle management services that support object creation and deletion; security services, which could

include encryption, authentication, and authorization; and various basic facilities such as event management and logging.

In general, TINA has similar requirements to the OMG when it comes to considering DPE services and generally has adopted OMG services as the basis of its management services. In some cases, TINA can build on basic DPE services for particular management service requirements. For instance, to locate managed objects in a distributed system, a locator service is required. A locator service could be based on the general CORBA trading service.

As TINA considers telecommunication needs extending from resources to services, including the DPE and DPE services supporting resource and service layers, TINA places particular requirements on its support services. Instead of discussing the many possible services that a TINA system could require, we will focus on event management services and TINA's requirements of them.

3.5.4.2.1 Event Management

Event management is a common issue in both resource and service layers and is common to most functional aspects. What must event management support in a TINA system? Let us consider an accounting management example. Accounting event management is a challenge, since almost all resources are considered accountable, thus generating accounting events.

- *DPE event management service:* we assume that the DPE event management services such as the notification service would guarantee the delivery of accounting events.
- *Interworking issues:* TINA Network Resource Architecture (TINA NRA) accounting event management should be able to interwork with X.734 compliant TMN accounting event management. This requirement is a part of interworking and migration issues in general, but adding this requirement at the event management level would allow flexibility and versatility to migration scenarios.
- *Context-driven event management:* A service management context acts on objects in its scope for the duration of the service transaction with which it is associated. We call this a service management scope, which designates a set of objects that share common service management interests. It is natural that accounting events are collected within this scope, so that correlation and analysis is more efficient as the events should be more relevant semantically.
- *Federation:* The federation in the accounting event management is part of federation issues in general. This level of federation, however, can be done without assuming a complete federation of the accounting management domains, but still allowing the accounting events to cross the domain boundaries.

The underlying event management mechanisms are to be provided by a DPE with event management facilities based on the specifications such as CORBA COS Event Service [33], X/Open Notification Service [34], DPE Notification Service [35], and CORBA Notification Service [36].

3.5.4.2.2 TMN-style Event Management Facilities

TMN provides event management facilities as a set of managed object classes.

- *EventForwardingDiscriminator (EFD)* is defined in X.734/ISO 10164-5. The EFD discriminates on events: that is, the EFD uses given filtering conditions to decide which events are to be forwarded to a particular destination and which are not to be forwarded (and eventually discarded).
- *Log-control function* is defined in X.735/ISO 10164-6 and enables the user of the log-managed object to control its logging activities and to update its logging records.
- *log* is a managed object and is defined in X.721/ISO 10165-2. The log-managed object is a container of logging records with a management interface.

TINA NRA [37] only supports the minimum necessary functionality of its TMN counterparts. There are two reasons for introducing TMN-style event management facilities:

- *Fine-grain interworking/migration:* In general, migration can be characterized as coarse-grained or fine-grained, or as loosely coupled or tightly coupled. In fine-grain interworking, two accounting systems (TINA and TMN) interwork with each other, using these TMN-compliant components.
- *Domain-based management:* TMN-style event management is domain based and independent of service management scope (i.e., it is not service transaction driven).

3.5.4.2.3 Event Management Ladder

How can service level events be correlated with resource level events? This question is of particular importance in TINA, where services can be dynamically configured on demand from the user. Since some of the resource-level accounting activities are triggered by a service transaction and are sustained only for the duration of the service transaction, it is natural that the context for the accounting event management be established for those service-oriented accounting management events. The context is established at the beginning of the service transaction and is resolved at the end of the service transaction.

Figure 3.5.8 illustrates the event management ladder concept. The ladder represents a context for the accounting event management encompassing the accountable objects involved in the service session. Accountable objects are any resources whose use is to be charged. They support some basic management functionality and can generate accounting events. The control messages are passed from the top of the ladder which is a USM within the service layer. Accounting events are passed from one rung of the ladder to the next, starting with the element management layer (EML) in the resource layer at the bottom of the ladder and moving toward the top, so that service-level billing information can be synthesized from the resource-level accounting events.

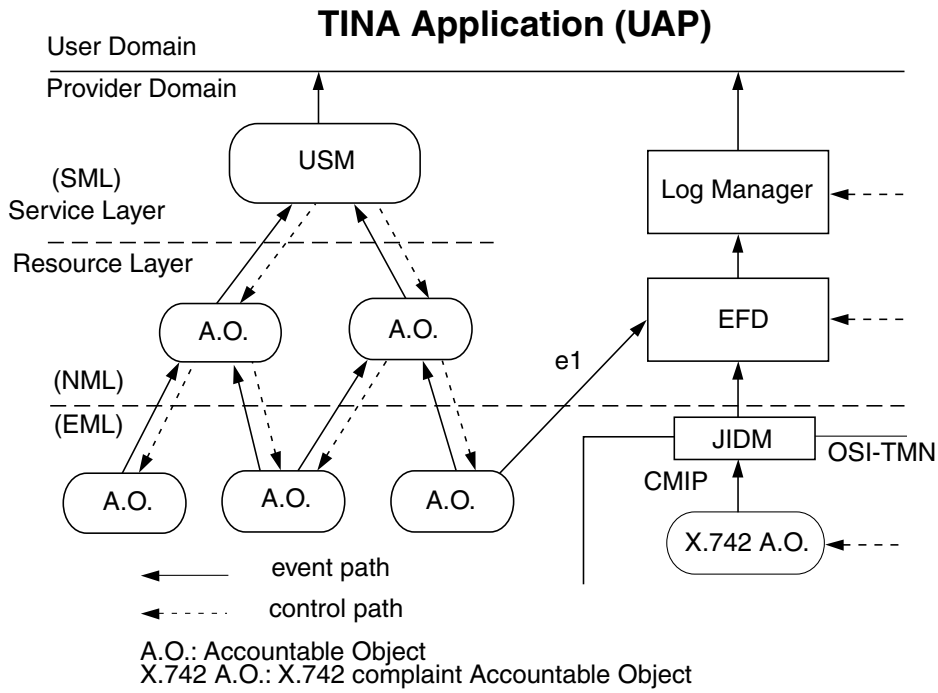

FIGURE 3.5.8 Event Management Ladder (EML). (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

The two accounting systems can interwork. For example, an accountable object in the ladder can send an event to the EFD as shown by e1 in the figure. This type of interworking between the two accounting systems is the key to the fine-grain migration from TMN-style accounting management to a TINA-based accounting management system using the event management ladder.

3.5.5 Life Cycle Axis

The life cycle axis deals with service and user life cycles. Service life cycle management considers how providers can deploy services, make them available, and control instance creation and deletion. TINA service life cycle management builds on software life cycle ideas with an emphasis on deploying services in a distributed, object-oriented environment across multiple domains. In a TINA service environment, user starts its life cycle by creating a new account. User life cycle management considers how users are introduced in the TINA service environment, and how they are supported and managed.

3.5.5.1 Service Life Cycle Management

Figure 3.5.9 summarizes the TINA service life cycle model. Every TINA service goes through the states described in the figure. The model focuses on the deployment/withdrawal phases, where three life cycle states corresponding to software states are shown. Further analysis of the service life cycle may reveal more states, as in the Service Instance Life Cycle (SILC). There are finer states within the utilization phase, such that service instances have their own life cycle, separate from the service type.

The following states are identified in the service life cycle:

- *Conceived, not planned:* The service has been conceived but no details about its implementation are known.
- *Not installed:* The service has been planned and does not exist in a TINA environment (although it might have existed in the past).
- *Installed:* The constituent parts of the service exist in a TINA environment but the service cannot be instantiated.
- *Activatible:* The service has the potential for being instantiated.
- *Instantiated:* An instance of the service has been instantiated.

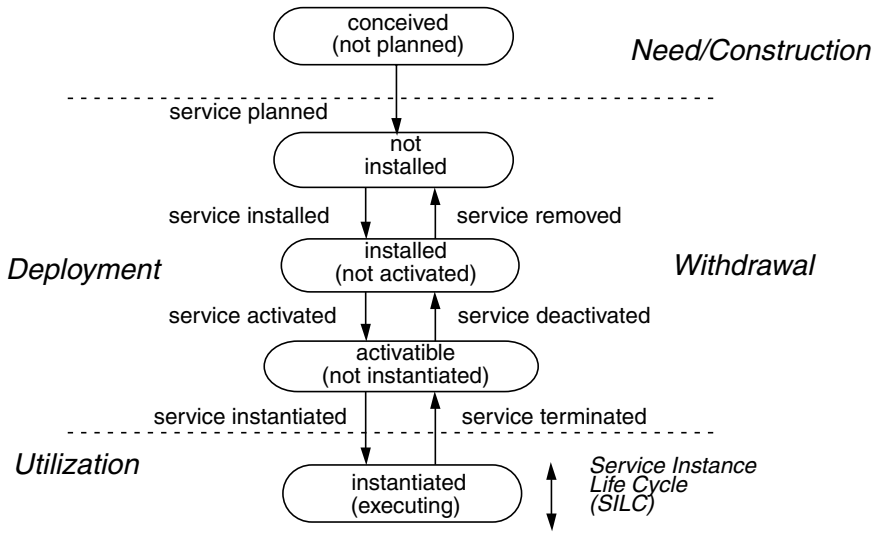

FIGURE 3.5.9 Service Instance Life Cycle (SILC) state transitions. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

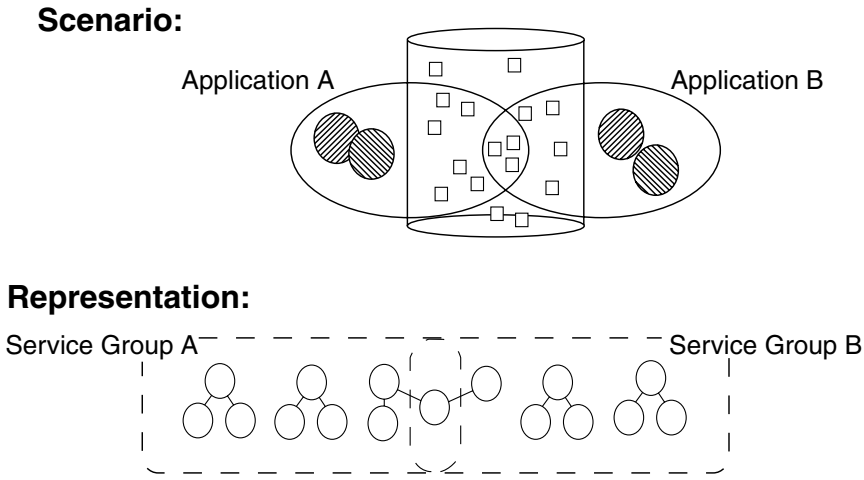

FIGURE 3.5.10 Two Service Groups (SGPs) sharing an Object Group. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements,* 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

3.5.5.1.1 Service Group

Since deployment and withdrawal of a service implies that a group of objects constituting the service needs to be deployed or withdrawn, it is natural to bundle these as a Service Group (SGP) [38]. Entities may include objects, executable files, data bases, and other resources required for a service to achieve an activatible state. The advantages of this grouping are best illustrated by an example. Consider a database which serves a number of applications, as shown in Figure 3.5.10. There is data in the database used by Application A. A subset of this data is also used by Application B. This situation is shown by the intersection of application data in the figure.

The two applications are each modelled as service groups as shown in the lower part of the figure. Shared data and objects are modelled by the intersection of the two service groups. Should Application A be withdrawn, then it is necessary to remove the data used solely by Application A but to leave those objects and data used by both Application A and Application B. Thus when withdrawing SGP A, the object group in the intersection of the two SGPs must remain.

In the view of the life cycle management, a TINA service is represented by a set of SGPs. Each SGP can be associated with a SGP template, which is the computational representation of service, encompassing the binaries of contained objects and service groups and any configuration information. A SGP template describes how a new service group can be instantiated and installed on a deployment request. A service group or a service group template provides the following properties:

- *Containment:* One or more SGPs may share an entity (e.g., an object group or database). A SGP may contain one or more other SGPs. These complexities are hidden from clients.
- *Scope:* A SGP may span one or more DPEs.
- *Dependencies:* A SGP knows the dependencies between contained entities, such as object groups or other SGPs. A SGP is dependent on the prior installation and activation of contained SGPs and SGPs with which it shares entities.

SGP is similar to the object grouping concept at DPE level. In particular, both support a similar containment hierarchy concept. They are distinguished by two of the properties listed earlier: their scope — a SGP may extend across several DPEs — and the concept of deployment dependency, which is not considered by object groups. The last feature is particularly important in deploying services. For example, a certain object group may need to be installed before others to avoid generating false alarms. Deployment dependency may also be relevant in the withdrawal phase.

3.5.5.1.2 Service Type Management

Service type management (Figure 3.5.11) handles control and management of service types, service type aliases, and service type templates. A service type template gives a generic template of service type, from which individual service instances are created. A service type must be maintained for the active lifetime of a TINA service, and the service type is also used for subscription and service instance creation. Therefore, service type management is an indispensable part of service life cycle management in TINA.

The service type manager is responsible for maintaining a service type during a service's active lifetime, and withdrawing it when the service is turned off. Although some functionality of the service type manager is similar to that of a DPE Trader [39] and CORBA trading service [41], it entails obtaining and distributing unique service type identifiers, registering and distributing service type templates, version control, and sub-domain control, as well as making particular services available.

3.5.5.1.3 Service Instance Management

Service type management and service instance management are both part of the utilization phase of service management. However, instance management covers the dynamic aspects, while service type management covers the static aspects. Service instance management includes creation and deletion of service instances, and monitoring and control of service instances (e.g., stop, resume) during their lifetimes.

The service factory is responsible for creation and deletion of each service instance for a specific service type. A service factory contains a service template which is a service-specific information object that determines how to create (or delete) a service instance. The service instance management operations are applied to the service session through the management interface of their corresponding SSMs.

3.5.5.2 User Life Cycle Management

Since consumer care and support are part of the retailer's business, it is important to consider consumers and their management. TINA has a clear view of consumers in its business model. A consumer role is a business role assumed by a stakeholder, which can be associated with a number of session roles in a service session [6]. Session roles can be associated with access and service usage. Access session roles determine how a person is recognized by the provider, and how a person's access privileges are given from the provider. TINA has a subscription model in which subscribers represent a stakeholder to negotiate contractual relations. An end-user can access TINA services based on subscription information and the user's service profile information created from the subscription management. We separate user life cycle management into two parts: end-user management and subscription management. Figure 3.5.12 shows the relationship between subscriber management, end-user management, and service type management.

3.5.5.2.1 Subscription Management

Subscription management allows a subscriber to establish a contractual relation between a user and a retailer, and it sets up a user service profile and user account information such that the user has access to the subscribed service. It involves the following information objects:

- *Subscription contract* represents the relationship established between consumer and retailer.
- *Subscriber profile* contains all relevant information regarding the subscriber.
- *Subscriber service profile* describes a particular service and its customization for the subscriber.
- *Service assignment group (SAG)* binds a service to one or more end users. SAGs specify the services available to a subscriber's associated users. An end-user may belong to multiple SAGs.

3.5.5.2.2 End-User Management

The end-user management allows end users to customize their services within the bounds set by their subscriber. It involves the following information objects:

- *End-user profile* represents all relevant information of the given end user, which includes registration, session description, usage context, and end-user service profile.
- *End-user service profile* describes a particular service and its relation to the end user, including customizing information. This is similar to a subscriber service profile except that it relates to a single end user.

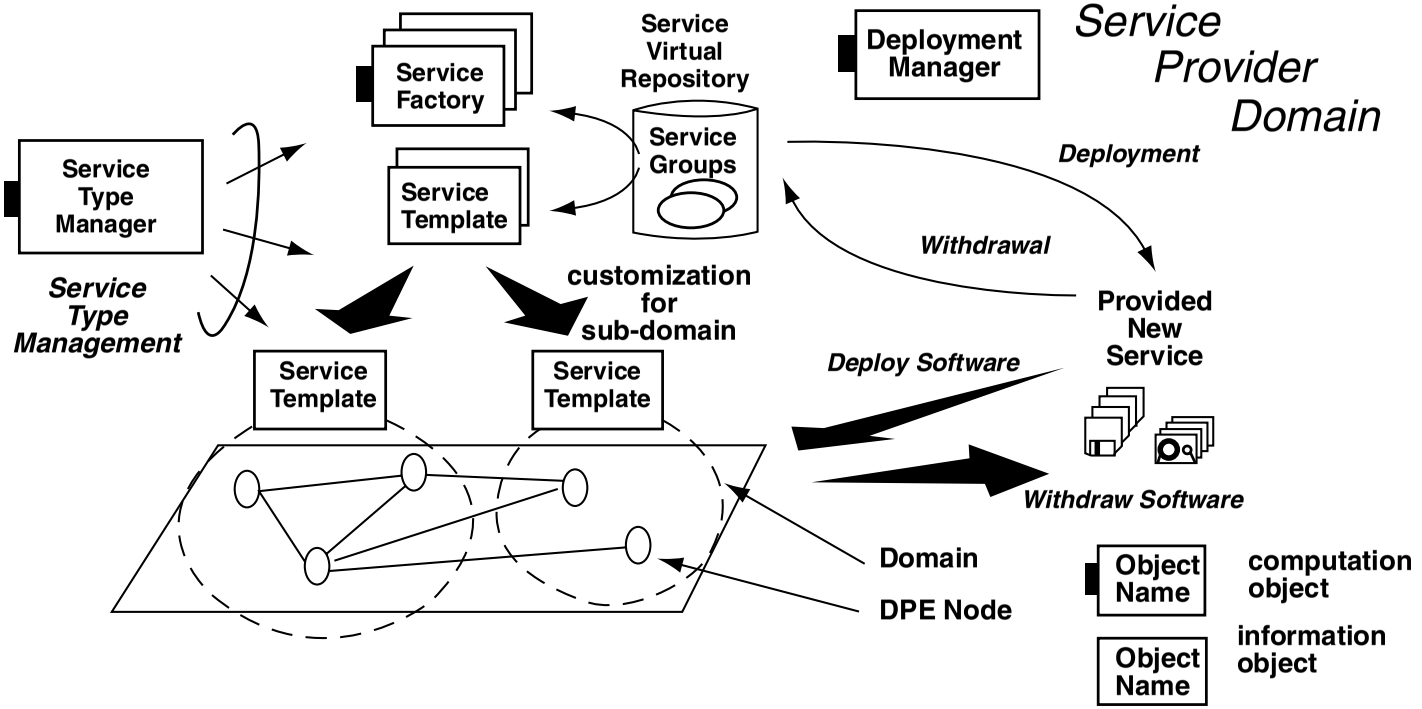

FIGURE 3.5.11 Service deployment and type management fragments. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements*, 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

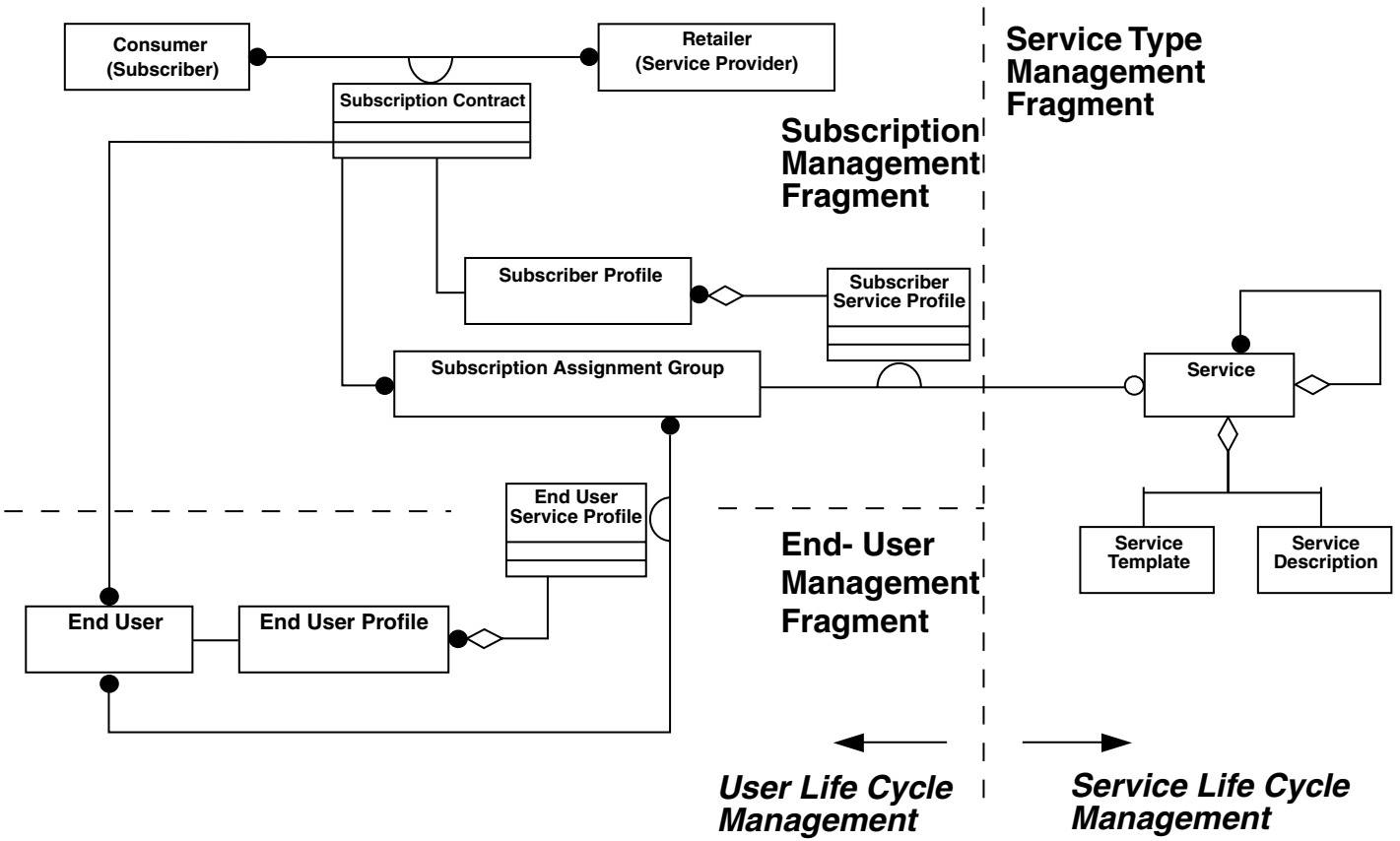

FIGURE 3.5.12 Life cycle management information model fragments. (From Hamada, T., Kamata, H., and Hogg, S., An Overview of the TINA Management Architecture, *Journal of Network and Systems Managements,* 5, 411–435, 1997. Used with permission of Plenum Publishing Corporation.)

3.5.6 Summary and Further Work

TINA imposes a high degree of user control and flexibility in the service layer. It also emphasizes service requirements in a multiprovider environment, and their impact on the resource layer. TINA uses domains to support multiple providers and introduces management policy as a means of controlling domain management. To meet service and functional requirements, the concepts such as management context and service transactions have been devised. The result is a flexible management framework, aimed at supporting consistent management across multiple domains.

TINA also considers service life cycle management and user life cycle management. Service grouping concepts enhance flexibility in deployment and withdrawal of services, and they allow providers to manage the complexity of service life cycle management, particularly when services are interrelated and service components are reused across multiple services. Subscription and subscriber control allows an entity (organization or household) to place controls on user access to services, while users can customize their services, allowing the services to meet individual needs and preferences. It provides retailers with a framework for supporting consumer relations at an enterprise level, which may span multiple services and end users.

Although the current framework presented in this chapter can offer considerable management support to TINA services, there are still many important research issues demanding further study. For example, we consider that service configuration management should be studied as an integrated part of service life cycle management. To derive a practical architecture, however, more use case analyses based on real-world examples are necessary. The service session graph [38] was originally designed as a generic tool for session control, but it can also be used for management support. A security management framework in TINA has yet to be fully developed. Some progress has been made in this direction [28] by using role-based control in conjunction with service session graph.

Acronyms

3Pty	3rd party service (reference point)
CMIP	Common management information protocol
ConS	Connectivity service (reference point)
CORBA	Common object request broker architecture
CSM	Communication session manager
DPE	Distributed processing environment
EFD	Event forwarding discriminator
EML	Element management layer
FCAPS	Fault, configuration, accounting, performance, security
GCCM	General context configuration manager
GDMO	Guideline for the definition of managed object
JIDM	Joint inter-domain management
LNFed	Layer network federation (reference point)
MgmtCtxt	Management context
NEL	Network element layer
NMF	Network management forum
NML	Network management layer
NRA	Network resource architecture
ODL	Object definition language
ODMA	Open distributed management architecture
ODP	Open distributed processing
OMG	Object management group

OSI	Open systems interconnection
PA	Provider agent
QoS	Quality of service
QoP	Quality of performance
Ret	Retailer (reference point)
SAG	Subscription assignment group
SF	Service factory
SGP	Service group
SILC	Service instance life cycle
SLC	Service life cycle
SML	Service management layer
SS	Service session
SSM	Service session manager
TINA [-C]	Telecommunications Information Networking Architecture [Consortium]
TMN	Telecommunications management network
UA	User agent
UAP	User application
USM	User service session manager

References

[1] Barr, W. J., Boyd, T., and Inoue, Y., The TINA Initiative, *IEEE Communication Magazine*, March, 70, 1993.

[2] ISO/IEC 10746-1, *Information Technology — Basic Reference Model of Open Distributed Processing — Part 1: Overview*, 1995.

[3] TINA-C, *Information Modelling Concepts, Ver. 2.0*, April 1995.

[4] TINA-C, *Computational Modelling Concepts, Ver. 3.2*, May 1996.

[5] TINA-C, *TINA Object Definition Language Manual, Ver. 2.3*, July 1996.

[6] TINA-C, *Service Architecture, Ver. 5.0*, June 1997.

[7] TINA-C, *Service Component Specification — Computational Model and Dynamics*, January 1998.

[8] ITU-T, *Principles for a Telecommunication Management Network*, ITU-T Recommendation M.3010, 1992.

[9] ITU-T, *Open Distributed Management Architecture*, ITU-T Recommendation X.703, 1997.

[10] Network Management Forum, *OMNIPoint Technical Architecture, Delivering a Management System Framework to Enable Service Management Solutions*, 1994.

[11] TINA-C, *Network Resource Information Model Specification, Ver. 2.2*, November 1997.

[12] ITU-T, *Architectures of Transport Networks Based on the Synchronous Digital Hierarchy (SDH)*, ITU-T Recommendation G.803, June 1992.

[13] ITU-T, *Architectures of Transport Networks*, ITU-T Recommendation G.805, June 1995.

[14] Pavon, J., Tomas, J., Bardout, Y., and Hauw, L., CORBA for Network and Service Management in the TINA Framework, *IEEE Communications Magazine*, March, 72, 1998.

[15] Eurescom Project P508, *Annex 1 Migration Paths Towards TINA*, ref P508.BT.PL.006, December 1996.

[16] Eurescom Project P508, *CORBA as an Enabling Factor for Migration from IN to TINA*, ref P508.FT.INCWP.4.1, December 1996.

[17] Sidor, D., TMN Standards: Satisfying Today's Needs While Preparing for Tomorrow, *IEEE Communications Magazine*, March, 54, 1998.

[18] TINA-C, *TINA Reference Points, Ver. 3.1*, June 1996.

[19] Network Management Forum, *Network Management Detailed Operations Map — Detailed Process Models to Support the Telecommunications Operations Map*, NMF GB908 Draft 0.9, March 1998.

[20] Network Management Forum, *NMF Technology Map*, NMF GB909 Issue 1.0, March 1998.
[21] Network Management Forum, *NMF Telecom Operations Map*, NMF GB910 Draft 0.2b, April 1998.
[22] Parhar, A. and Handegard T., TINA Object Groups: Patterns in Chaos, *Proc. of TINA'96*, 13, Heidelberg, Germany, September 1996.
[23] ITU-T, *Terms and Definitions Related to the Quality of Telecommunication Services*, ITU-T Recommendation E.801, October 1996.
[24] Hamada, T., TINA Accounting Management Architecture, *Proc. of TINA'96*, 193, Heidelberg, Germany, September 1996.
[25] Eckardt, T. et al., Personal Communications Support in the TINA Service Architecture — A new TINA-C Auxiliary Project, *Proc. of TINA'96*, pp. 55-64, Heidelberg, Germany, September 1996.
[26] Staamann, S., Wilhelm, U. et al., Security in the Telecommunication Information Networking Architecture — the CrysTINA Approach, *Proc. of TINA'97*, Santiago de Chile, November 1997.
[27] Hamada, T., Hoshiai, T., Yates, M., A Perspective on TINA Service Security Architecture, *Proc. of 5th Workshops on Enabling Technologies (WET ICE'96)*, Stanford, California, 74, June 1996.
[28] Hamada, T., Dynamic Role Creation from Role Hierarchy — Security Management of Service Session in Dynamic Service Environment, *Proc. of TINA'97*, Santiago de Chile, November 1997.
[29] OMG, *Common Object Request Broker: Architecture and Specification, Revision 2.0*, July 1995.
[30] Appleby S. and Steward S., Mobile software agents for control in telecommunications networks, *BT Technology Journal*, Vol. 12, No. 2, 104, April 1994.
[31] NMF, *NMF Component Sets: CORBA/CMIP/SNMP Interworking*, Draft, CS342, 1996.
[32] X/Open, *Inter-Domain Management Specifications: Specification Translation (JIDM)*, 1995.
[33] OMG, *The Common Object Request Broker: Common Object Service Specifications, Revision 1.0*, OMG, 94-1-1, January 1994.
[34] X/Open, *Generic Security Service*, X/Open Publication Set T402, 1994.
[35] OMG, *The Common Object Request Broker: TINA Notification Service Description*, telecom/96-07-02, July 1996.
[36] OMG, *The Common Object Request Broker: Notification Service Request for Proposal*, Draft 4, telecom/96-12-01, December 1996.
[37] TINA-C, *Network Resource Architecture Ver. 3.0*, February 1997.
[38] TINA-C, *Service Architecture Annex, Ver. 5.0*, June 1997.
[39] TINA-C, *Engineering Modelling Concepts (DPE Architecture), Ver. 2.0*, December 1994.
[40] Berndt, H., Darmois, E., Dupuy, F. et al., *The TINA Book — Cooperative Solution for Competitive World*, Prentice-Hall, 1998.
[41] OMG, *Trading Object Service*, OMG RFP 5 Submission, OMG Document orbos/96-07-08, July 1996.
[42] Dupuy, F., Nilsson, G., and Inoue, Y., The TINA Consortium: Toward Networking Telecommunications Information Services, *Proc. of International Switching Symposium (ISS)*, Berlin, Germany, 1995.
[43] Rubin, H. and Natarajan, N., A Distributed Software Architecture for Telecommunications Applications, *IEEE Network Magazine*, Vol. 8, No. 1, January/February 1994.

Further Information

All public architectural documents and interface specifications are available at the TINA-C home page (http://www.tinac.com). Other useful information on TINA-C such as FAQ, tutorials from the past TINA conferences, and updates on current activities are also available from the TINA-C home page. Since 1998, TINA-C has been in phase II, and the latest updates and revisions are performed by its working groups. Latest activities of the working groups are also available from the home page. The TINA Book [40] is written by the foremost experts on TINA, and the book stands to be the standard source of reference.

International TINA conferences, which have been held annually since 1990, are excellent sources of the foremost research papers on TINA. For the historical development of TINA and TINA-C, [1] and [42] give good information. For Bellcore Information Networking Architecture (INA) project, a predecessor of TINA, [43] is recommended.

Due to limitation of space, several important topics of TINA had to be omitted from this chapter. For distributed processing environment (DPE), the DPE architecture document [39] is the most reliable source of information. Since a major part of DPE is made compliant with OMG CORBA, the difference between the two distributed architectures is practically minimal. There still remain some differences, however, as influences from ODP are more apparent in TINA DPE. In particular, ODP concepts such as capsule, cluster, stream interface, and multiple interface objects are not yet fully supported in the current version of CORBA [36]. The latest updates on OMG activities are available from its web site (http://www.omg.org). For TINA-related activities at ITU-T, TINA ODL [5] is currently on standardization track at ITU-T SG 10. TINA is also being studied as Long Term Architecture (LTA) at ITU-T SG 11.

Connection management is another important topic in TINA. Network Resource Information Model (NRIM) [11] and Network Resource Architecture (NRA) [37] are the two most reliable sources of information. NRIM describes the information viewpoint, whereas NRA describes the computational viewpoint of network resources.

Many TINA-related researchers have been supported as part of the European ACTS program (http://www.uk.infowin.org/ACTS) and Eurescom (http://www.eurescom.de). Project reports from Eurescom are available (http://www.eurescom.de/public/deliverables/dfp.htm).

For TINA service architecture, the service architecture (SA) document [6] and its companion document, the service component specification (SCS) [7] are the most reliable sources. SA defines basic concepts of TINA service architecture, and SCS describes computational models, interfaces, and event sequences of service components.

3.6 Telecommunications Support Processes

Kornel Terplan

This contribution is based on the SMART TMN Telecom Operations Map by the TeleManagement (TM) Forum (TELE98). The TM Forum is an international, non-profit organization serving the telecommunications industry. Its mission is to help service providers and network operators automate their business processes in cost- and time-effective ways. Specifically, the work of the TM Forum includes:

- Establishing operational guidance on the shape of business processes
- Agreeing on information that needs to flow from one function to another
- Identifying a realistic systems environment to support the interconnection of operational support systems
- Enabling the development of a market and of real products for automating telecom operations processes
- TM Forum makes use of international and regional standards when available, and provides input to standards bodies whenever new technical work is done

The members of the TM Forum include service providers, network operators, and suppliers of equipment and software to the telecommunications industry. With that combination of buyers and suppliers of operational support systems, TM Forum is able to achieve results in a pragmatic way that leads to product offerings as well as to specifications of business processes and support systems. Members meet regularly; several working groups are in action, targeting real challenges in the telecommunications industry. This documentation is extremely valuable to old and new providers who are in the process of re-engineering their support processes. The editor and author thanks the TM Forum for the assistance with this contribution.

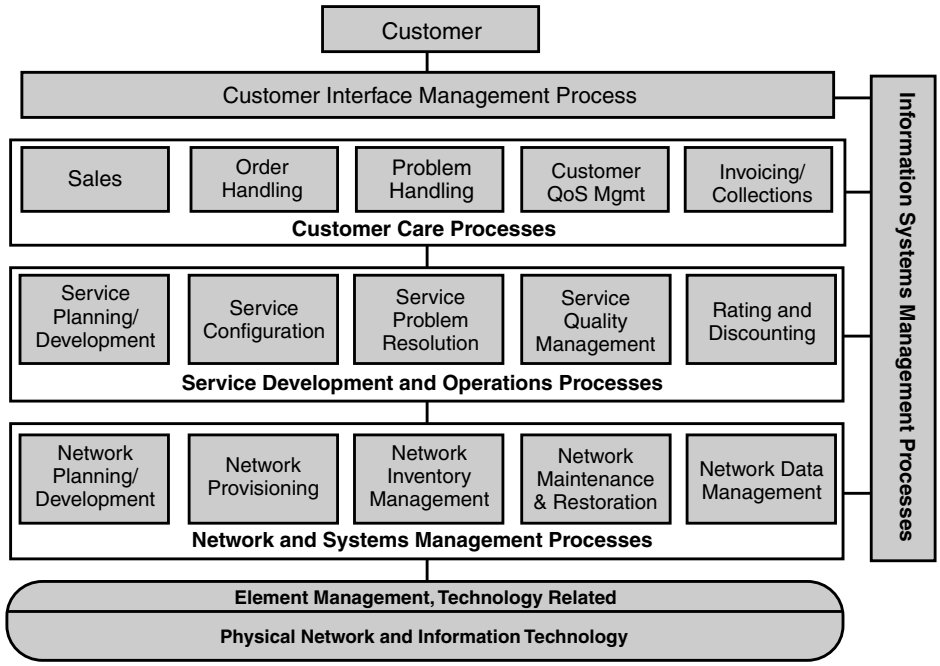

FIGURE 3.6.1 Business model of telecommunications support processes.

A telecommunications service provider efficiently and effectively conducts business by managing its essential business and support processes. These processes can be aggregated to deliver the major requirements common to any service-oriented business, which include:

- Service fulfillment (timely delivery of what the customer has ordered)
- Service assurance (maintaining the service — timely response and resolution of customer and network triggered problems, managing and reporting performance for all aspects of a service)
- Billing for the service (timely and accurate bills, including invoicing, timely adjustment handling and payment collections)

Figure 3.6.1 shows the business process model as recommended by the TM Forum (ADAM96).

3.6.1 High-level Breakdown of Support Processes

Figure 3.6.2 shows a broad breakdown of the business model into the three customer-focused activities identified earlier. The purpose is to show the second dimension in more detail and highlight the predominant processes that need to be involved in the end-to-end tasks of supporting customer services. The network inventory management and network data management processes are split to display their significant role, more than an interface, in both overarching processes; e.g., network inventory management is important to both the fulfillment and assurance processes.

Each of the principal areas are to be described with simplified flow charts. The TM Forum expects to develop robust process flows both at a generic level and in some more specific areas. These process flows are representative, and need customization, but they are an excellent basis for re-engineering processes.

Fulfillment Example

This example of a fulfillment process shows a possible sequence of activities to support a customer inquiry, subsequent order for service, the configuration of the service, and the installation and completion of the request (Figure 3.6.3). Depending on the service provider process, orders can be placed through the sales process and/or directly through the order management process. For a specific service provider, some

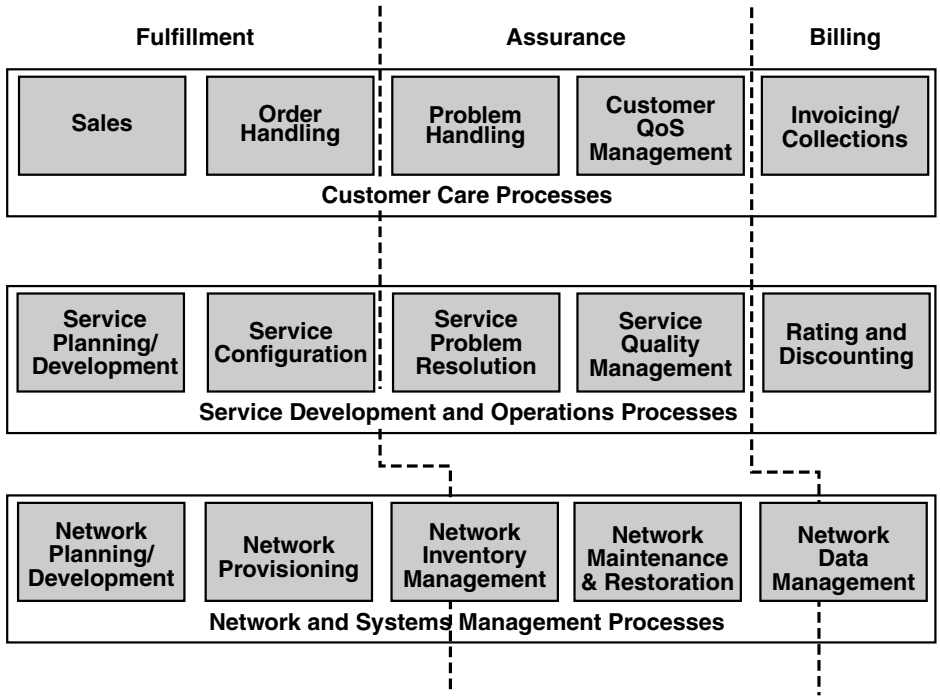

FIGURE 3.6.2 High-level process breakdown.

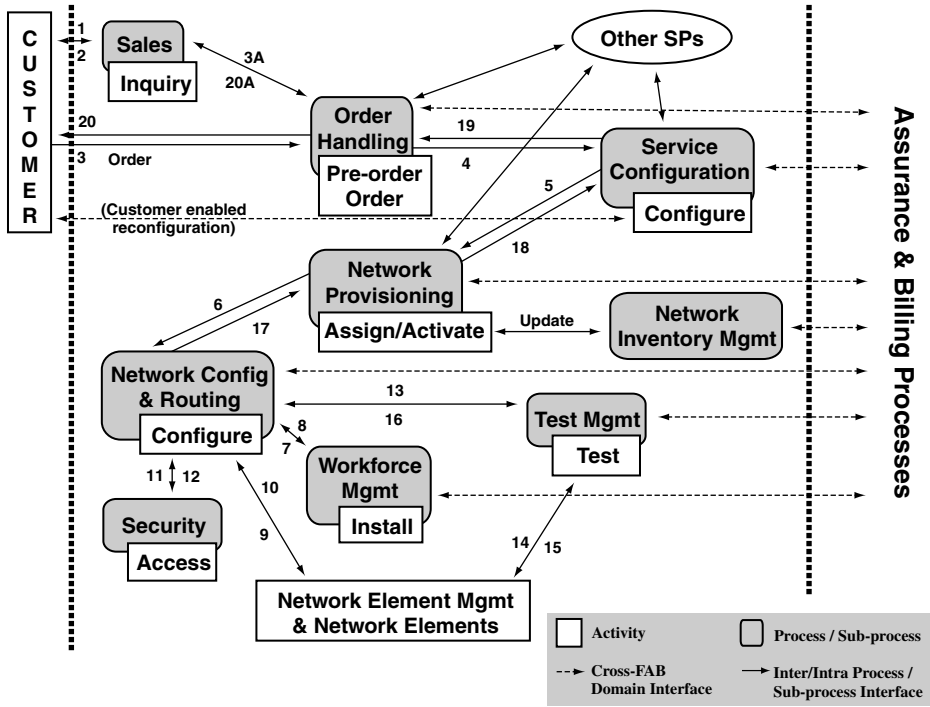

FIGURE 3.6.3 Fulfillment process example.

customers may be supported by a specific sales team that places some or all orders for the customer and tracks them to completion. These dual-trigger process interfaces and follow-ups are shown as 3/3A and 20/20A.

Information exchanges are numbered as follows:

1 Selling
2 Selling
3 Order
3A Order status and completion
4 Service request
5 Assignment request
6 Network configuration request
7 Installation request and completion
8 Installation request and completion
9 Element configuration complete
10 Element configuration complete
11 Network access check and complete
12 Network access check and complete
13 Test request
14 Perform test and test data
15 Perform test and test data
16 Test complete
17 Network configuration complete
18 Assignment complete
19 Service complete
20 Order status and completion
20A Order status and completion

The complete fulfillment flow-through may not actually be required every time for some simple services, which have preassigned service capacity. For example, the flow for an instance of a service set-up could be bypassed at network provisioning, when configured and tested facilities have been preprovisioned. This depends upon a particular provider's operational process and policy. It will also impact the timing of interactions with network inventory management, hence the interface sequence number has been omitted. Interfaces may be required with other service providers or network operators when the service offered to a customer is one of many different kinds of joint service arrangements.

Service Assurance Example

Most service providers are driving their service assurance processes to become proactive, meaning triggered by automation rather than triggered by the customer. This is important for improving service quality, customer perception of service, and for lowering overall costs. Customer care processes have been basically reactive in the past. The extreme pressure on cost, customer demand for more control, and customer demand for more proactive service support are driving a major shift to proactive support through automation. With the advent of internet access, the goal for processes and automation is now interactive support, including giving the customer the ability to see and act on service performance.

Service assurance processes are interlinked to each other. Figure 3.6.4 shows a possible sequence of activities in response to a network-detected problem. The figure shows two ways a potential service-affecting problem could be identified, e.g., by either an "alarm event" or by synthesis of network data through Network Data Management. Neither is exclusive. Network data management logically collects and processes both performance and traffic data as well as usage data. The usage data is used as logical part of the billing process.

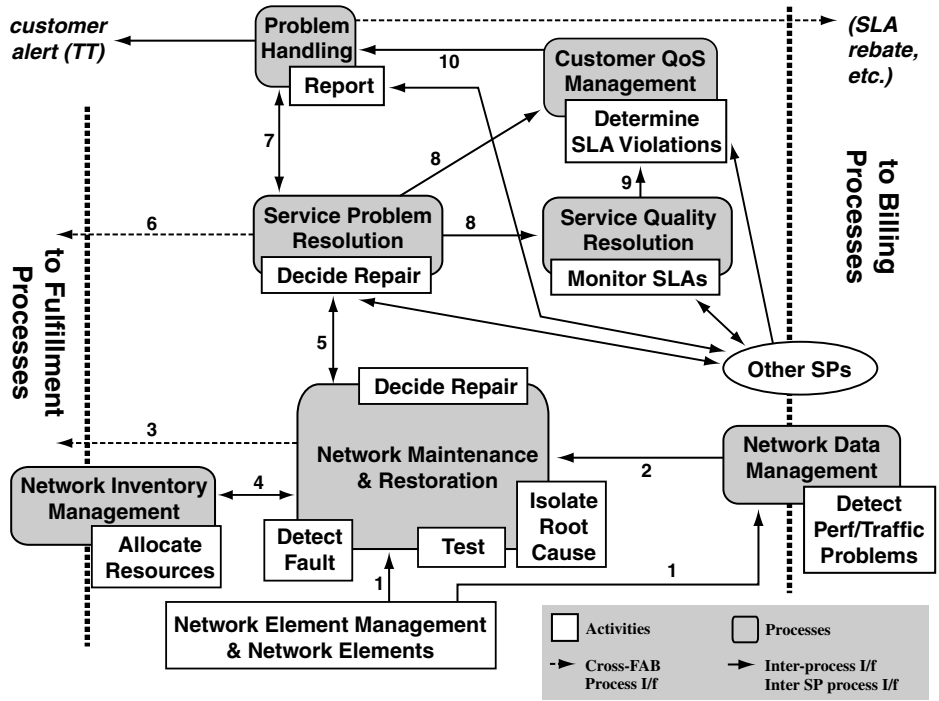

FIGURE 3.6.4 Service assurance process example.

Information exchanges are numbered as follows:

1 Network data, alarm data, events
2 Report degradation
3 Network reconfiguration or change
4 Work order
5 Notify problem/fix
6 Service reconfiguration or change
7 Trouble report or ticket
8 Report problem data
9 SLA impact
10 Service impact

Billing Example

Figure 3.6.5 shows a typical sequence of activities to generate a bill that has flat rate elements (one-time installation, monthly recurring), usage charges, and possible SLA adjustments. Service providers may also choose to apply discounts or rebates for outages and/or SLA breaches to a specific customer's bill, according to service type, promotion, customer relationship, and/or according to its policy or customer contract.

When a service is provided by a combination of different service providers, usage and/or other billing data may be aggregated by the main service provider from input by other secondary service providers, and one bill presented to the customer. This is a trend, but it depends on service provider billing strategy, customer wishes, the actual service arrangement provided, and/or service provider process capability and policy.

Information exchanges are numbered as follows:

1 Network usage data
2 Aggregated usage data

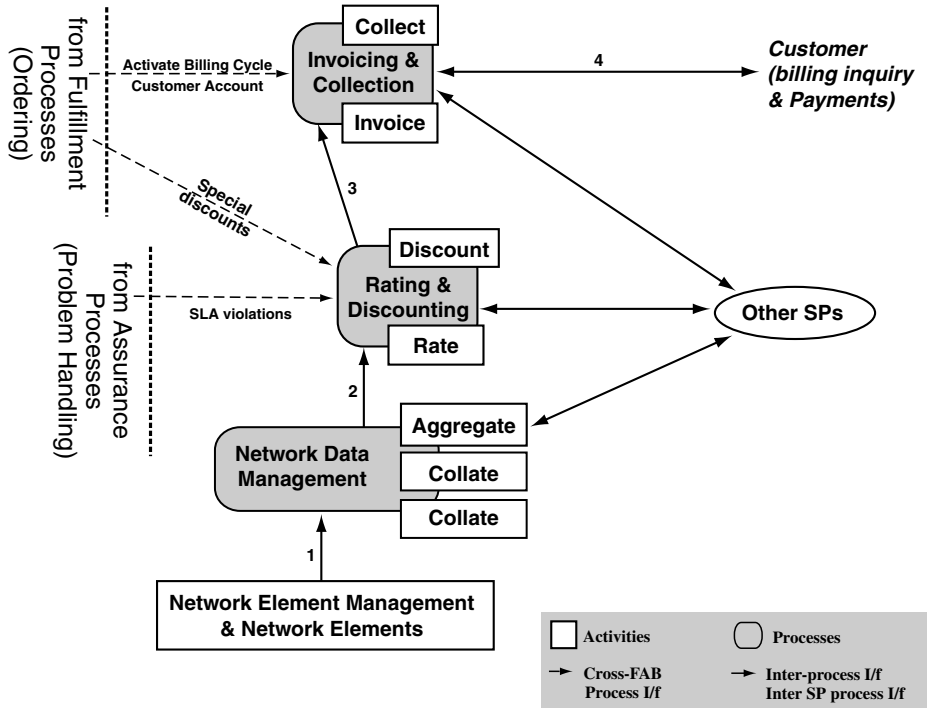

FIGURE 3.6.5 Billing process example.

3 Summarized bill content
4 Generate bills

In the next chapters, each process will be described by its tasks, input, and output connections with internal and external processes and activities.

3.6.2 Customer Care Processes

These processes involve direct interaction with an end customer to provide, maintain, and bill for network services. The end customer is the ultimate buyer of a network service.

3.6.2.1 Customer Interface Management Process

This process may be distinct, or may be performed as part of the individual customer-care processes on an individual service or cross-service basis. These are the processes of directly interacting with customers and translating customer requests and inquiries into appropriate events, such as the creation of an order or trouble ticket or the adjustment of a bill. This process logs customer contacts, directs inquiries to the appropriate party, and tracks the status to completion. In those cases where customers are given direct access to service management systems, this process assures consistency of image across systems, and security to prevent a customer from harming their network or those of other customers. The aim is to provide meaningful and timely customer contact experiences as frequently as the customer requires.

Figure 3.6.6 shows the customer interface management process. Principal functions include:

- Receive and record contacts
- Direct inquiries to appropriate processes
- Monitor and control status of inquiries, and escalate
- Ensure a consistent image and secure use of systems

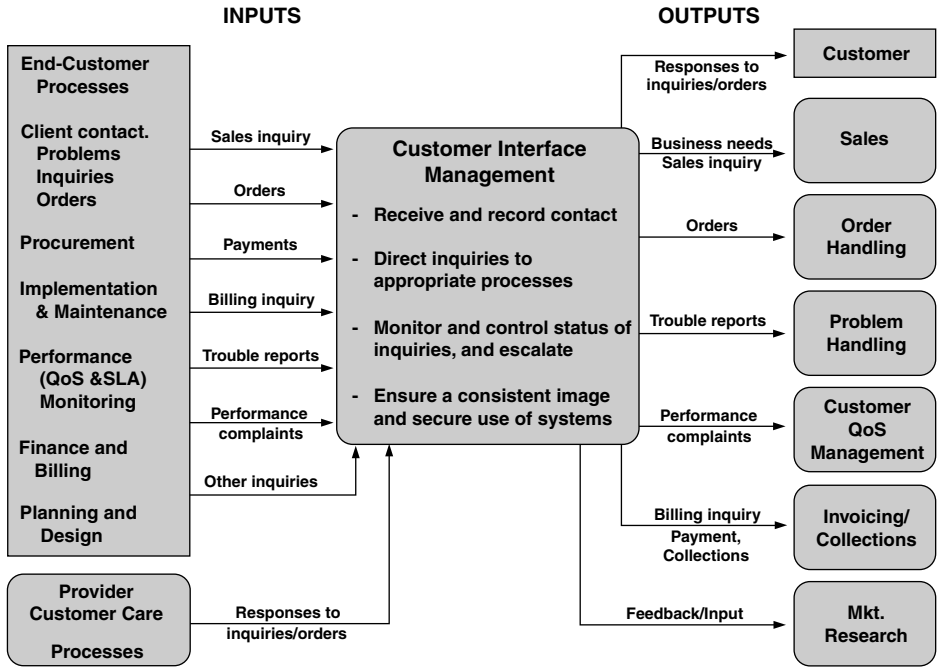

FIGURE 3.6.6 The customer interface management process.

3.6.2.2 Sales Process

This process encompasses learning about the needs of each customer, and educating the customer about the communications services that are available to meet those needs. It includes working to create a match between the customer's expectations and the service provider's ability to deliver. Depending on the service provider process it can be pure selling or can include various levels of support. The sales process may include preorder work and interfaces. The aim is to sell the correct service to suit the customer's need and to set appropriate expectations with the customer. SLA negotiation, Request for Proposal (RFP) management, and negotiation are led from this process.

Figure 3.6.7 shows the sales process. Principal functions include:

- Learn about customer needs
- Educate customer on services
- Match expectations to offerings and products
- Arrange for appropriate options
- Forecast service demand
- Manage SLA and RFP negotiations

3.6.2.3 Ordering Process

The ordering process includes all the functions of accepting a customer's order for service, tracking the progress of the order, and notifying the customer when the order is complete. Orders can include new, change, and disconnect orders for all or part of a customer's service, as well as cancellations and modifications to orders. Preorder activity that can be tracked is included in this process. The development of an order plan may be necessary when service installation is to be phased in, and the need for preliminary feasibility requests and/or pricing estimates may be part of this process when certain services are ordered. The aim is to order the service the customer requested, support changes when necessary and to keep the customer informed with meaningful progress of the order, including its successful completion.

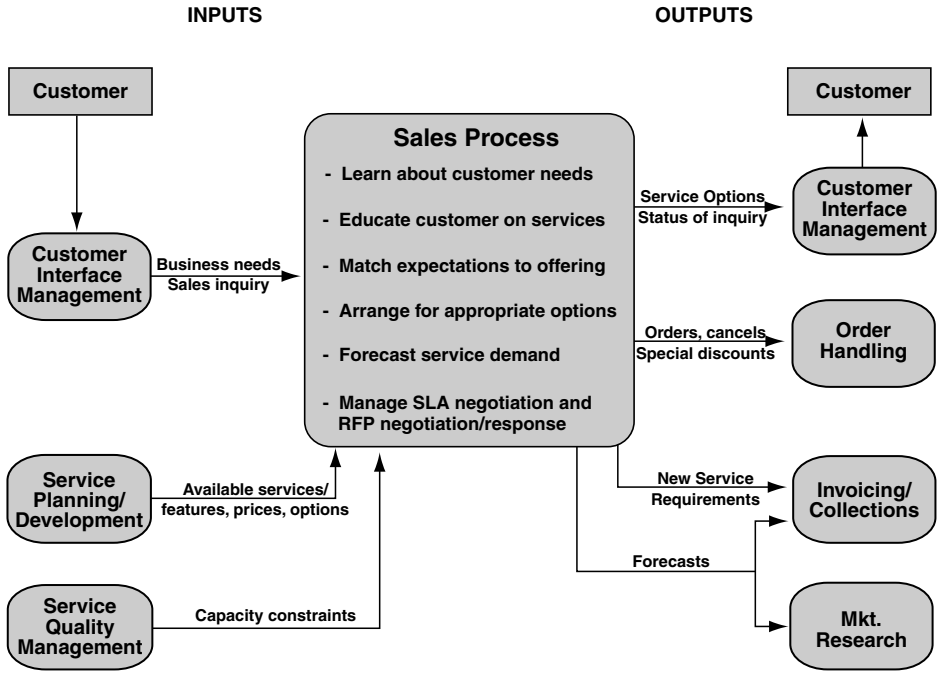

FIGURE 3.6.7 The sales process.

Figure 3.6.8 shows the ordering process. Principal functions of this process include:

- Accept orders
- Determine preorder feasibility
- Prepare price estimates and SLA terms
- Develop order plan
- Perform credit check
- Request customer deposit
- Initiate service installation
- Reserve resources
- Issue orders, and track status
- Complete orders, notify customers
- Initiate billing process

Service orders to other providers can be generated by the ordering process or by the service configuration process, depending on the nature of the service ordered by the customer. They can also be generated from network and systems management processes when part of a network infrastructure.

3.6.2.4 Problem-Handling Process

This process is responsible for receiving service complaints from customers, resolving them to the customer's satisfaction and providing meaningful status on repair or restoration activity. This process is also responsible to be aware of any service-affecting problems, including:

- Notifying customers in the event of a disruption — whether reported by the customer or not
- Resolving the problem to the customer`s satisfaction
- Providing meaningful status on repair or restoration activity

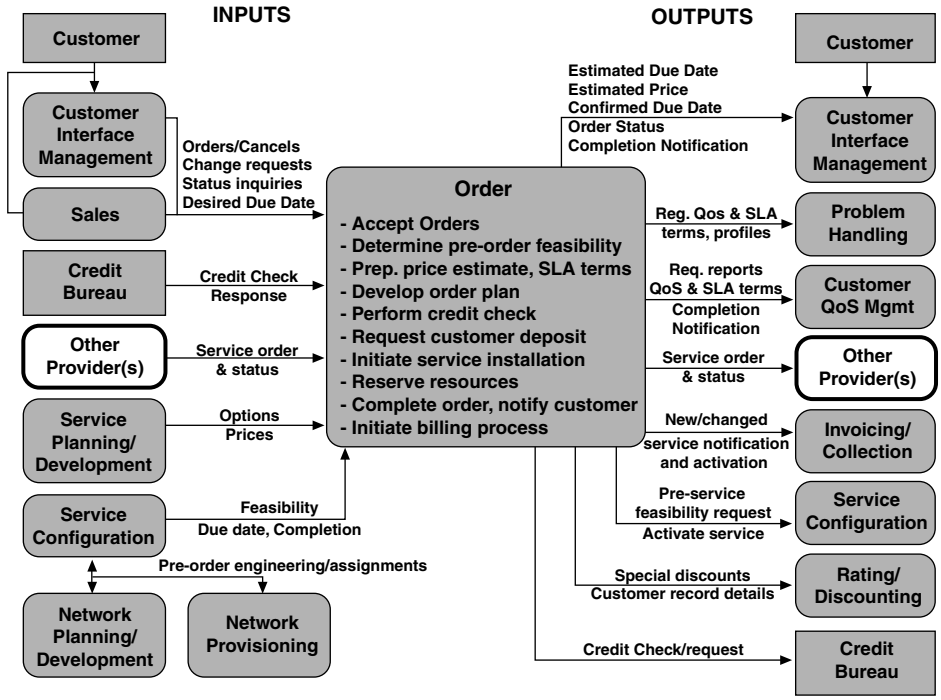

FIGURE 3.6.8 The ordering process.

This proactive management also includes planned maintenance outages. The aim is to have the largest percentage of problems proactively identified and communicated to the customer, to provide meaningful status, and to resolve in the shortest timeframe.

Figure 3.6.9 shows the problem-handling process. Principal tasks include:

Receive trouble notification
Determine cause, resolve, or refer
Track progress of resolution
Initiate action to reconfigure if needed
Generate trouble tickets to suppliers
Confirm trouble cleared, notify customer
Schedule with and notify customer of planned maintenance

When trouble is reported by the customer, a trouble report may be sent to service problem resolution for correction. When a trouble is identified by service problem resolution, then problem handling is notified in order to inform the customer of the problem.

3.6.2.5 Customer Quality of Service (QoS) Management Process

This process encompasses monitoring, managing, and reporting of quality of service (QoS) as defined in service descriptions, service level agreements, and other service-related documents. It includes network performance, but also performance across all service parameters, e.g., orders completed on time. Outputs of this process are standard (predefined) and exception reports including, but not limited to, dashboards, performance of a service against a SLA, reports of any developing capacity problems, reports of customer usage patterns, etc. In addition, this process responds to performance inquiries from the customer. For SLA violations, the process supports notifying problem handling, and for QoS violations, notifying service quality management. The aim is to provide effective monitoring. Monitoring and reporting must provide

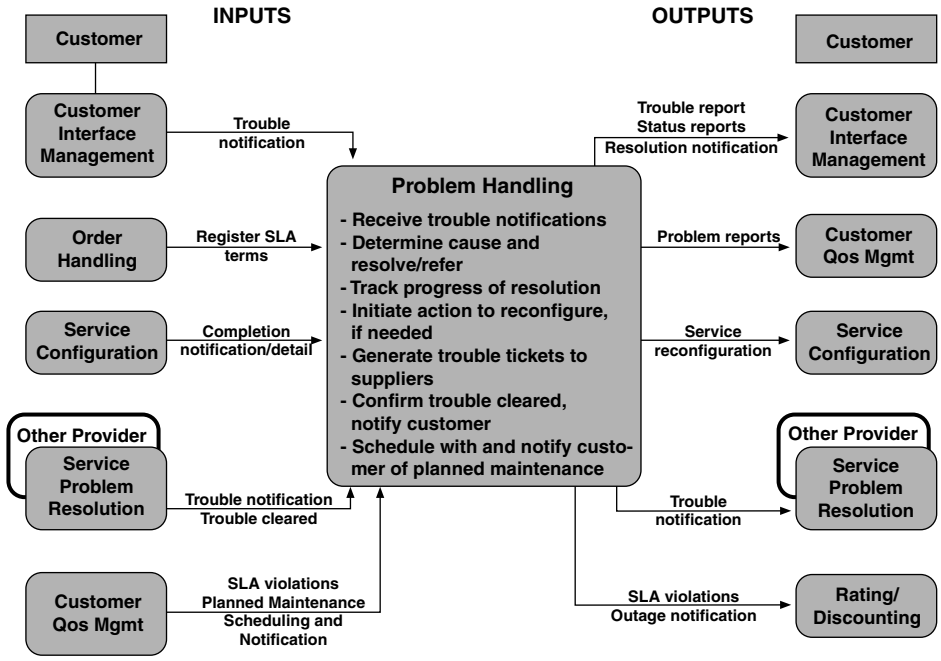

FIGURE 3.6.9 The problem-handling process.

SP management and customers with meaningful and timely performance information across the parameters of the services provided. The aim is also to manage service levels that meet specific SLA and standard service committments.

Figure 3.6.10 shows the customer QoS process. Principal tasks include:

- Schedule customer reports
- Receive performance data
- Establish reports to be generated
- Compile and deliver customer reports
- Manage SLA performance
- Determine and deliver QoS and SLA violation information

3.6.2.6 Invoicing and Collection Process

This process encompasses sending invoices to customers, processing their payments, and performing payment collections. In addition, this process handles customer inquiries about bills, and is responsible for resolving billing problems to the customer's satisfaction. The aim is to provide a correct bill and, if there is a billing problem, resolve it quickly with appropriate status to the customer. An additional aim is to collect monies due the service provider in a professional and customer-supportive manner.

Some providers allow invoicing and collections functions for other providers as a service.

Figure 3.6.11 shows the invoicing and collection process. The principal functions include:

- Create and distribute invoices
- Collect payments
- Handle customer account inquiries
- Manage debt
- Bill on behalf of other providers

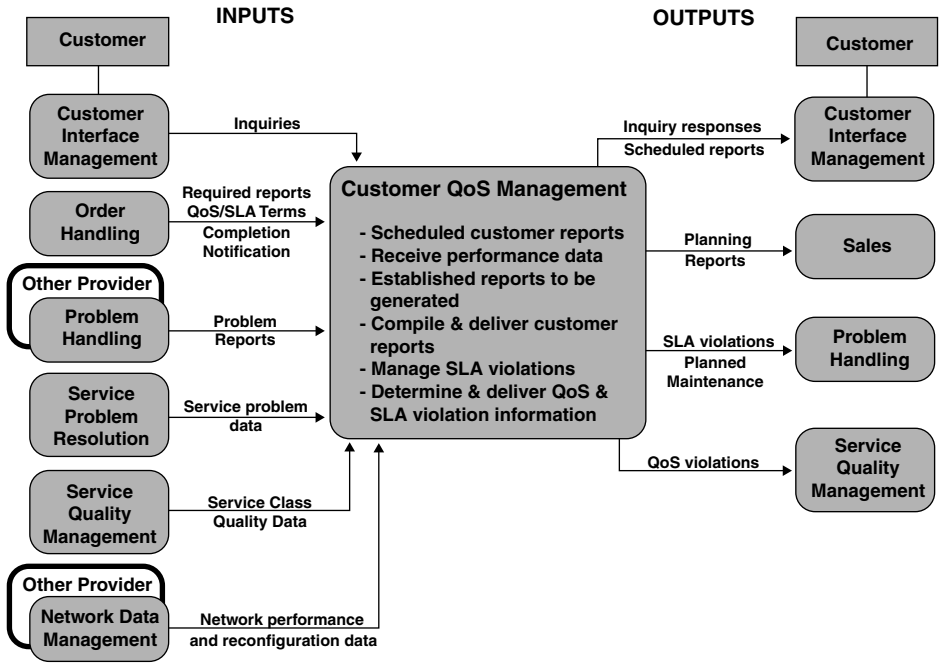

FIGURE 3.6.10 The customer quality of service (QoS) management process.

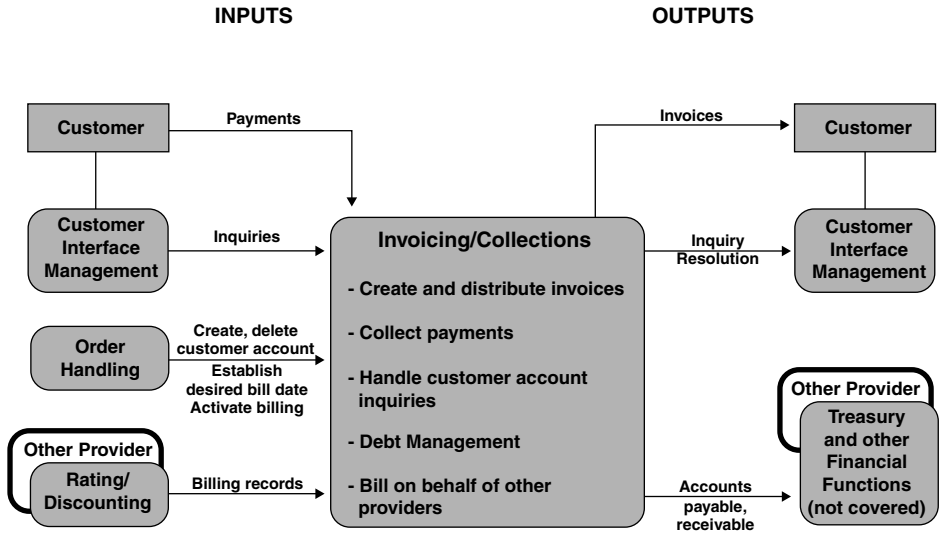

FIGURE 3.6.11 The invoicing and collection process.

3.6.3 Service Development and Operations Processes

These processes are generally one step removed from day-to-day customer interaction. Focus is on service delivery and management as opposed to the management of the underlying network and information technology. Some of these functions are done on a one-time basis, such as designing and delivering a

new service or feature. Other functions involve the application of a service design to specific customers or managing service improvement initiatives, and are closely connected with the day-to-day customer experience.

3.6.3.1 Service Planning and Development Process

This process encompasses the following functional areas:

- Designing technical capability to meet specified market need at desired cost
- Ensuring that the service (product) can be properly installed, monitored, controlled, and billed
- Initiating appropriate process and methods modifications, as well as initiating changes to levels of operations personnel and training required
- Initiating any modifications to the underlying network or information systems to support the requirements
- Performing preservice testing to confirm the technical capability works and that the operational support process and systems function properly
- Ensuring that sufficient capacity is available to meet forecasted sales

Figure 3.6.12 shows the service planning and development process. Principal functions are:

- Develop and implement technical solutions
- Develop and implement procedures
- Define and implement systems changes
- Develop and implement training
- Develop customer documentation
- Plan rollout, test, start service, and project management
- Set product/service pricing

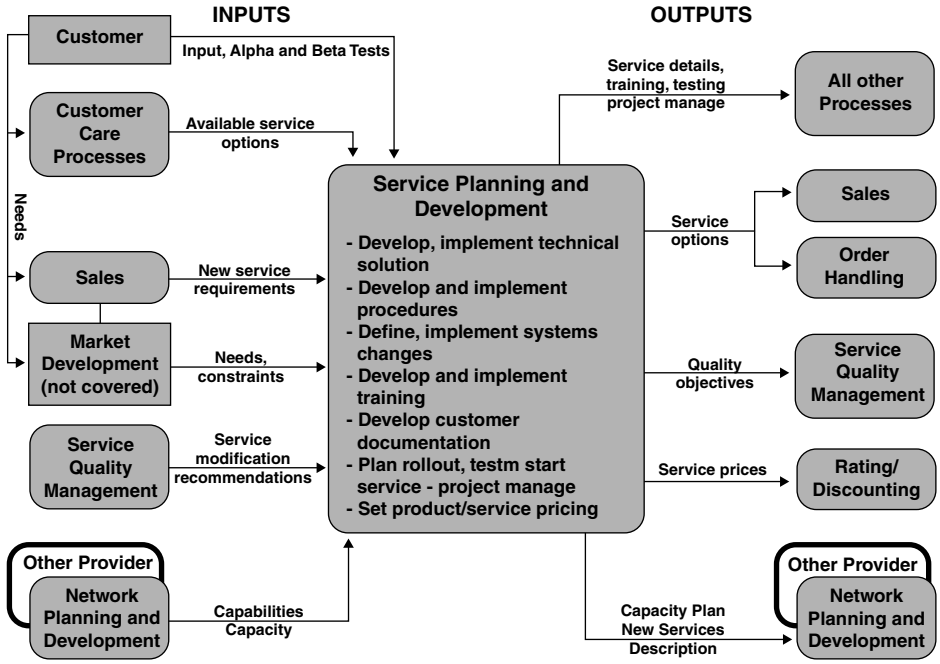

FIGURE 3.6.12 The service planning and development process.

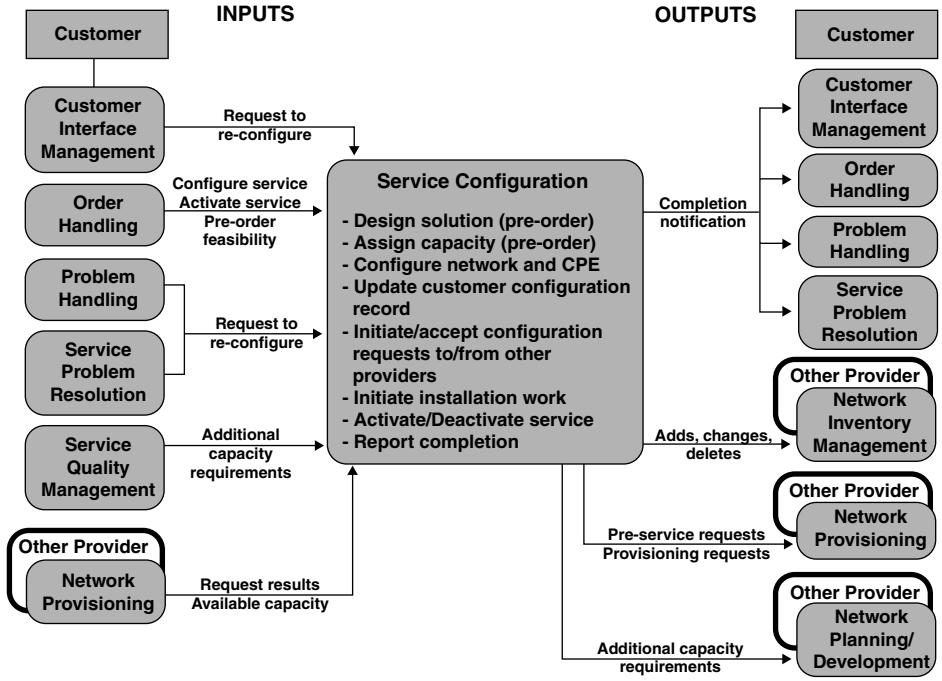

FIGURE 3.6.13 The service configuration process.

3.6.3.2 Service Configuration Process

This process encompasses the installation and/or configuration of services for specific customers, including the installation/configuration of customer premises equipment (CPE). It also supports the reconfiguration of service (either due to customer demand or problem resolution) after the initial service installation. The aim is to correctly provide service configuration within the timeframe required to meet ever-decreasing intervals.

Figure 3.6.13 shows the service configuration process. Principal functions include:

Design solution (preorder)
Assign capacity (preorder)
Configure network and CPE
Update customer configuration record
Initiate/accept configuration requests to/from other providers
Initiate installation work
Activate/deactivate service
Report completion

3.6.3.3 Service Problem Resolution Process

This process encompasses isolating the root cause of service affecting and non-service affecting failures and acting to resolve them. Typically, failures reported to this process impact multiple customers. Actions may include immediate reconfiguration or other corrective actions. Longer-term modifications to the service design or to the network components associated with the service may also be required. The aim is to understand the causes impacting service performance and to implement immediate fixes or initiate quality improvement efforts.

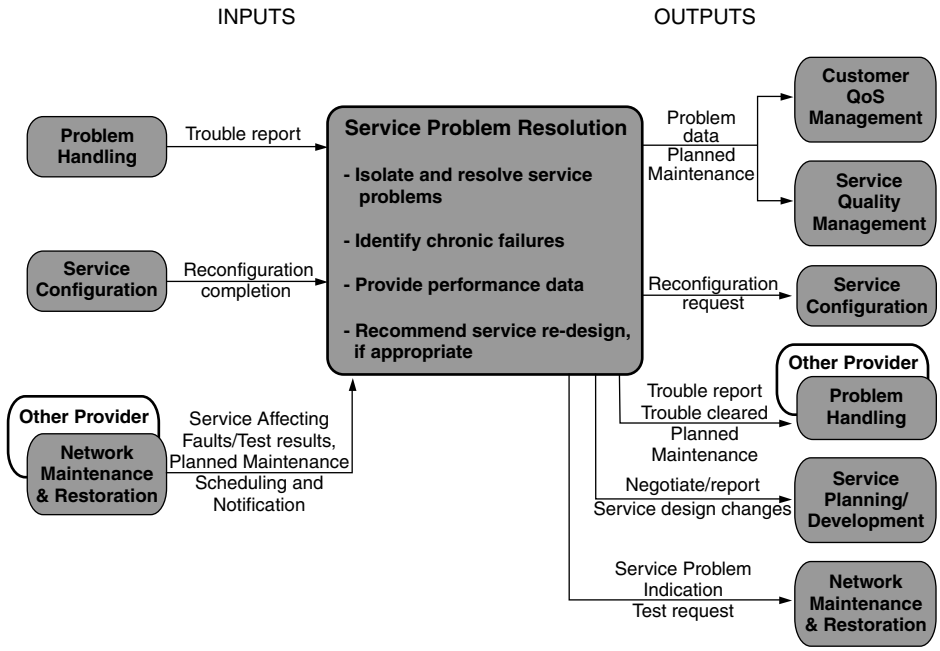

FIGURE 3.6.14 The service problem resolution process.

Figure 3.6.14 shows the service problem resolution process. The principal tasks include:

- Isolate and resolve service problems
- Identify chronic failures
- Provide performance data
- Recommend service redesign, if appropriate

When multiple troubles are reported by customers of a service, a report may be sent from problem handling to service problem resolution for correction. When trouble is identified by service problem resolution, Problem Handling is notified.

3.6.3.4 Service Quality Management Process

This process supports monitoring service or product quality on a service class basis in order to determine whether:

- Service levels are being met consistently
- There are any general problems with the service or product
- The sale and use of the service is tracking to forecasts

This process also encompasses taking appropriate action to keep service levels within agreed targets for each service class and to either keep ahead of demand or alert the sales process to slow sales. The aim is to provide effective service-specific monitoring, to provide management and customers with meaningful and timely performance information across the parameters of the specific service. The aim is also to manage service levels to meet SLA commitments and standard commitments for the specific service.

Figure 3.6.15 shows the service quality management process. The principal tasks include:

- Manage life cycle of service/product portfolios
- Monitor overall delivered quality of a service class
- Monitor available capacity/usage against forecasted sales

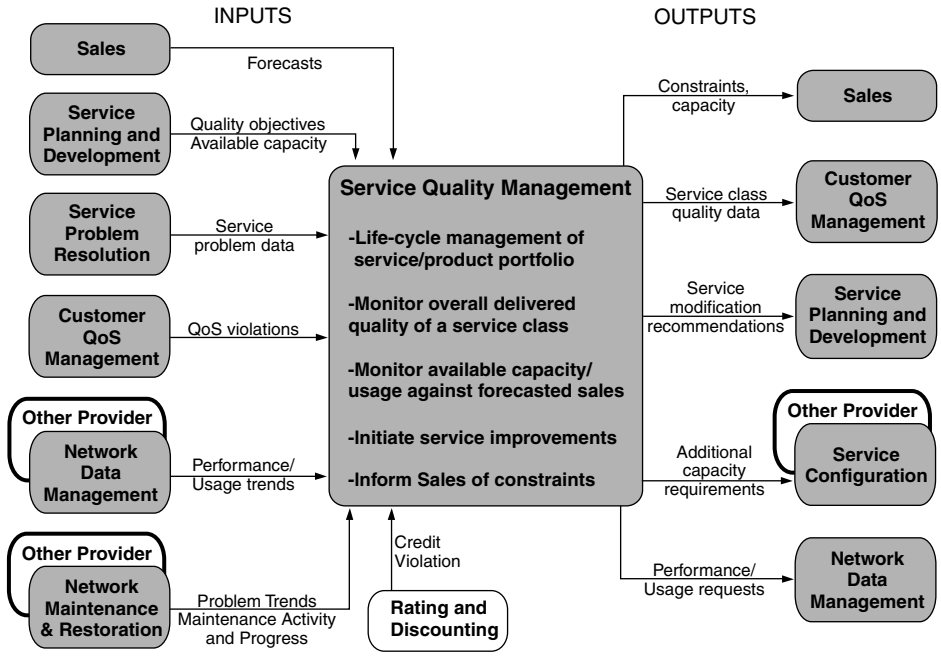

FIGURE 3.6.15 The service quality management process.

- Initiate service improvements
- Inform sales on constraints

3.6.3.5 Rating and Discounting Process

The rating and discounting process encompasses the following functional areas:

- Applying the correct rating rules to usage data on a customer-by-customer basis, as required
- Applying any discounts agreed to as part of the ordering process
- Applying promotional discounts and charges
- Applying outage credits
- Applying rebates due because SLA were not met
- Resolving unidentified usage

The aim is to correctly rate usage and to correctly apply discounts, promotions, and credits. Figure 3.6.16 shows the rating and discounting process. Principal functions are:

- Apply service rates to usage
- Apply negotiated discounts
- Apply rebates

3.6.4 Network and Systems Management Processes

These processes are responsible for ensuring that the network infrastructure supports the end-to-end delivery of the required services. Network Management is a key integration layer between the element management layer and the service management layer. Its basic function is to assemble information from element management systems, and then integrate, correlate, and in many cases, summarize that data to pass on the relevant information to service management systems.

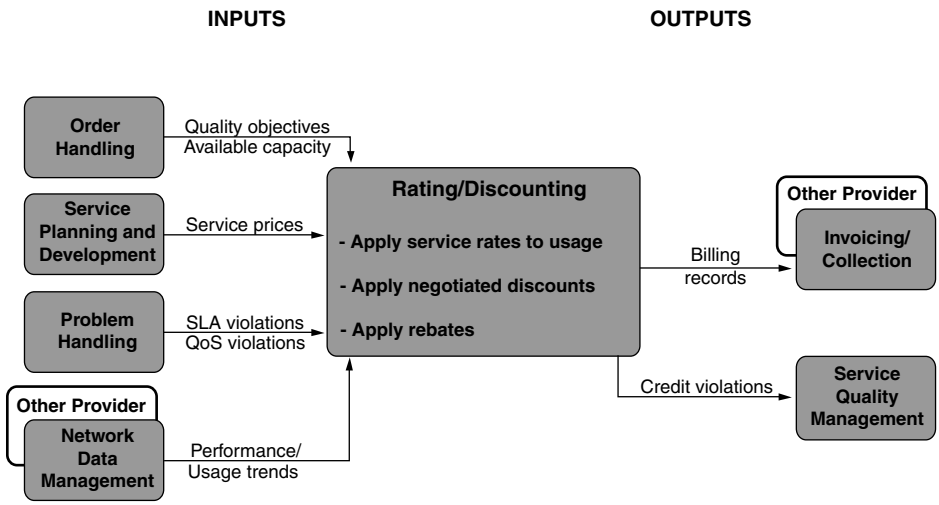

FIGURE 3.6.16 The rating and discounting process.

Network Management is more than just a mediator between the element and service management layers. It has its own responsibilities, e.g., network provisioning and network fault management. The key issue is that management responsibility will be placed at a level where adequate information is present, instead of shifting all responsibilities to service management. At this point, systems management is not explicitly addressed, but would logically need to be integrated into the architecture at this level.

The network and systems management processes manage the complete service provider network and subnetwork. Much of the interface is through element management.

3.6.4.1 Network Planning and Development Process

This process encompasses development and acceptance of strategy, description of standard network configurations for operational use, definition of rules for network planning, installation, and maintenance.

This process also deals with designing the network capability to meet a specified service need at the desired cost and for ensuring that the network can be properly installed, monitored, controlled, and billed. The process is also responsible for ensuring that enough network capacity will be available to meet the forecasted demand and support cases of unforecasted demand. Based on the required network capacity, orders are issued to suppliers or other network operators and site preparation and installation orders are issued to the network inventory management or a third-party network constructor. A design of the logical network configuration is provided to network provisioning.

Figure 3.6.17 shows the network planning and development process. Principal functions are:

- Develop and implement procedures
- Set up framework agreements
- Develop new methods and architectures
- Plan required network capacity
- Plan the mutation of network capacity
- Issues orders to suppliers and other network operators
- Plan the logical network configuration

3.6.4.2 Network Provisioning Process

This process encompasses the configuration of the network, to ensure that network capacity is ready for provisioning of services. It carries out network provisioning as required, to fulfill specific service requests and configuration changes to address network problems. The process must assign and administer identifiers

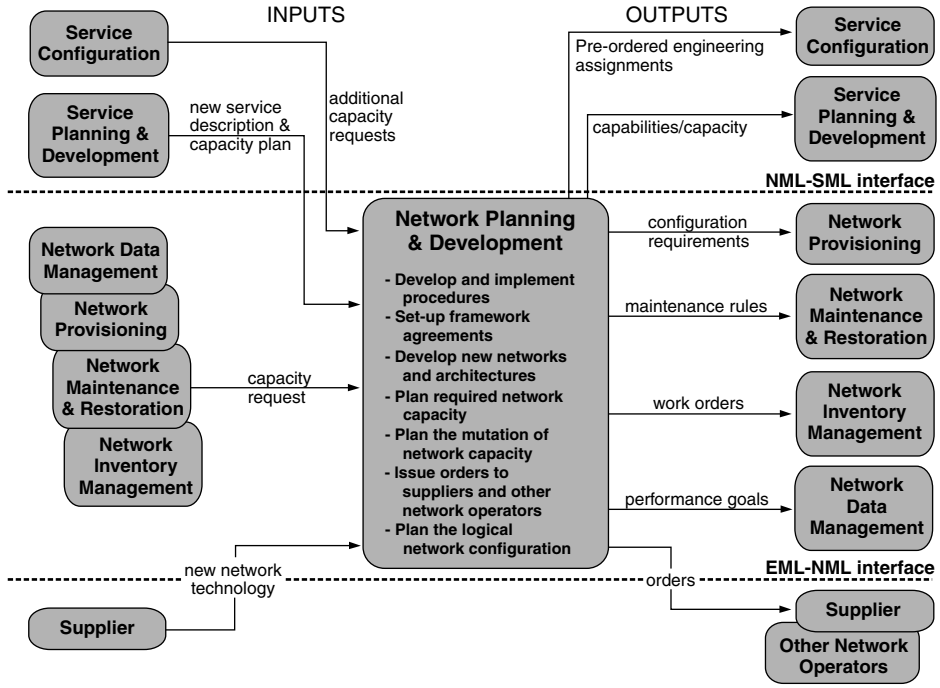

FIGURE 3.6.17 The network planning and development process.

for provisioned resources and make them available to other processes. Note that the routine provisioning of specific instances of a customer service — in particular, simple services such as plain old telephone service (POTS) — may not normally involve network provisioning, but may be handled directly by service provisioning from a preconfigured set.

Figure 3.6.18 shows the network provisioning process. Principal functions of this process are:

- (Re)configuration of the network; installation of initial configuration and reconfiguration due to capacity problems
- Administration of the logical network, so that it is ready for service
- Connection management
- Network testing

3.6.4.3 Network Inventory Management Process

This process encompasses anything to do with physical equipment and the administration of this equipment. The process is involved in the installation and acceptance of equipment, with the physical configuration of the network, but also with handling of spare parts and the repair process. Software upgrades are also a responsibility of this process.

Figure 3.6.19 shows the network inventory management process. The principal tasks of this process are:

- Install and administer the physical network
- Perform work in the network
- Manage the repair activities
- Align inventory with network
- Manage spare parts
- Manage faulty parts

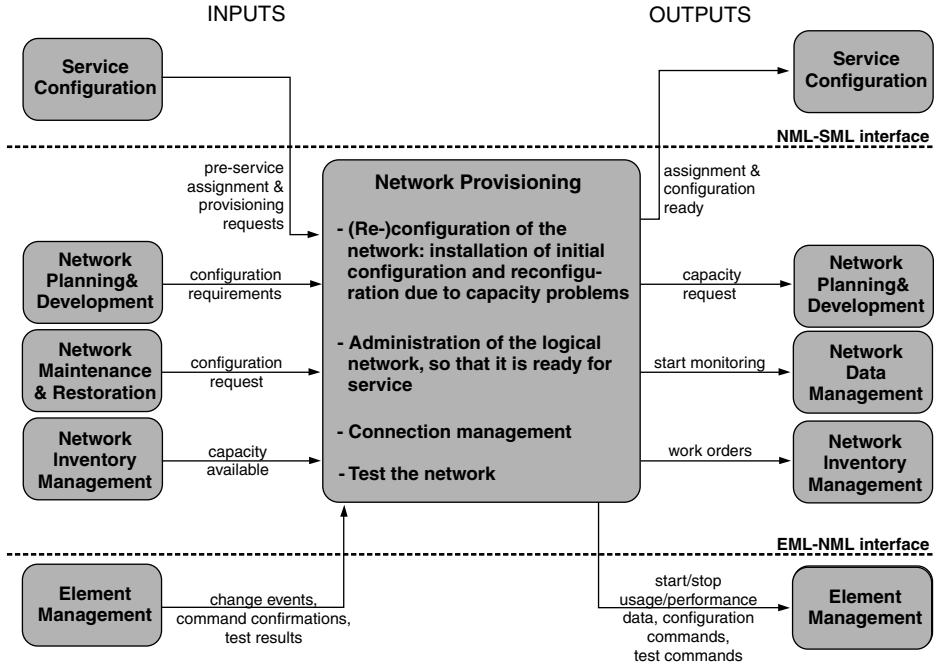

FIGURE 3.6.18 The network provisioning process.

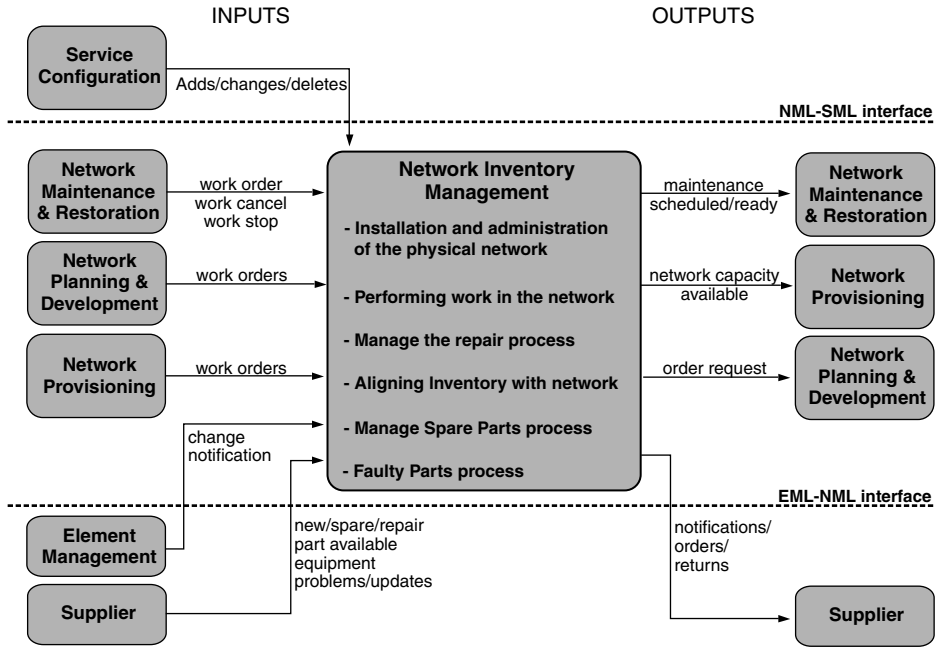

FIGURE 3.6.19 The network inventory management process.

3.6.4.4 Network Maintenance and Restoration Process

This process encompasses maintaining the operational quality of the network, in accordance with required network performance goals. Network maintenance activities can be preventative — such as scheduled routine maintenance — or corrective. Corrective maintenance can be in response to faults or to indications

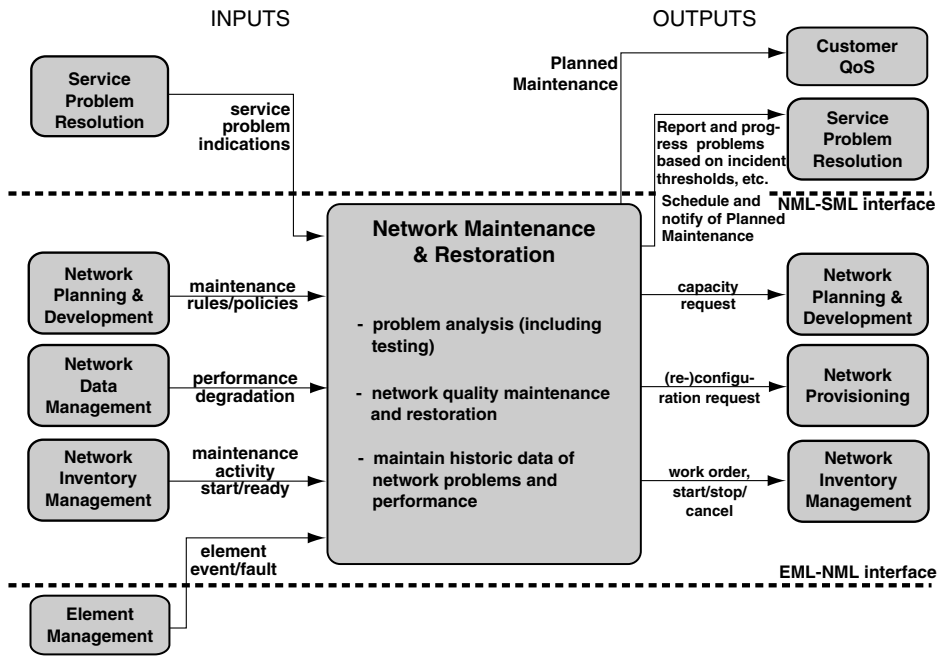

FIGURE 3.6.20 The network maintenance and restoration process.

that problems may be developing (proactive). This process responds to problems, initiates tests, does analysis to determine the cause and impact of problems, and notifies Service Management of possible effects on quality. The process issues requests for corrective actions.

Figure 3.6.20 shows the network maintenance and restoration process. Principal tasks include:

- Problem analysis, including testing
- Network quality maintenance and restoration
- Historic data maintenance of network problems and performance

3.6.4.5 Network Data Management Process

This process encompasses the collection of usage data and events for the purpose of network performance and traffic analysis. This data may also be an input to billing (rating and discounting) processes at the service management layer, depending on the service and its architecture.

The process must provide sufficient and relevant information to verify compliance/noncompliance to SLAs. The SLAs themselves are not known at the Network Management Layer. The process must provide sufficient usage information for rating and billing.

This process must ensure that the network performance goals are tracked, and that notification is provided when they are not met (threshold exceeded, performance degradation). This also includes thresholds and specific requirements for billing. This includes information on capacity, utilization, traffic, and usage collection. In some cases, changes in traffic conditions may trigger changes to the network for the purpose of traffic control. Reduced levels of network capacity can result in requests to network planning for more resources.

Figure 3.6.21 shows the network data management process. Principal tasks include:

- Collect, correlate, and format usage
- Determine performance of capacity and utilization
- Provide notification on performance
- Initiate traffic metrics collection function

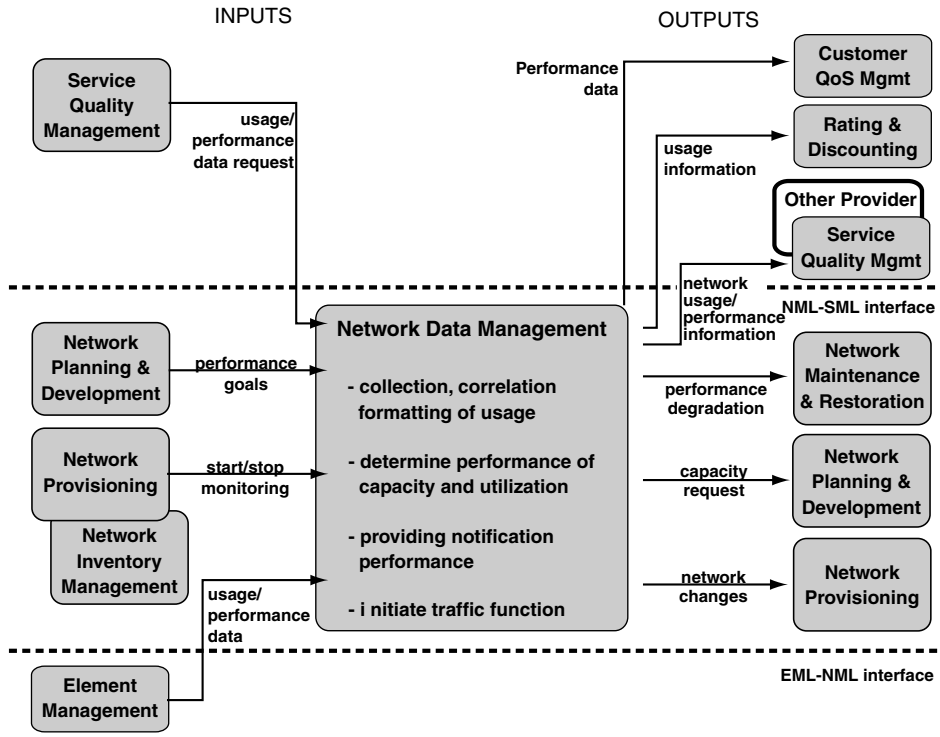

FIGURE 3.6.21 The network data management process.

3.6.5 Summary and Trends

The telecommunication industry has embraced the TMN model as a way to think logically about how the business of a service provider is managed. The model itself is simple, although its implementation is complex. The sheer number of standards now available that address the various interfaces between management systems sometimes makes it difficult to see and appreciate the big picture. These ITU standards are mainly concentrated in the Element Management and Network Management Layers. They have been developed from the bottom up, making it difficult to apply the standards as part of a business case. It is also difficult to have a customer-centric focus.

Smart TMN™ is a holistic, business-driven approach to implementing TMN. It provides investment direction as well as development specifications needed to produce management systems. It starts with the layered TMN model, but goes much further, to address concrete business problems in a pragmatic way. Specifically, Smart TMN™ brings three key elements to bear on the TMN layers:

- A business process-driven approach: Smart TMN™ starts from the premise that service providers and network operators need to automate their business processes, which means information needs to flow from end-to-end across many different systems. Only when a process is understood and the linkages are clear is it possible to apply standards in a way that delivers business value. Unless that is known, a great deal of money can be spent implementing standards that simply don't contribute to the overall business objective.
- Technology-independent agreements: Business agreements about what information will flow between processes must be kept independent of the protocols used to implement those agreements. Technology will continue to change, becoming cheaper and easier to use, and delivering more power. Smart TMN™ applies the right technology for the job instead of forcing a single technology to serve every need. Further, the Smart TMN™ approach documents all agreements in technology-neutral form, so that the same agreement can be implemented in multiple technologies as they continue to evolve.

- Products, not just paper: A main premise of Smart TMN™ is that paper standards are not sufficient to solve business problems. Products, not paper, are the end goal, provided documentation is produced to support the replications of industry agreements across multiple vendors' products.

This contribution has put emphasis on a telecom operations map which is a common model for telecommunications operations processes and a guide for re-engineering such processes.

References

ADAM96 Adams, E., Willets, K.: *The Lean Telecommunications Provider,* McGraw-Hill, New York, 1996.

TELE98 Telemanagement Forum: Smart TMN — Telecom Operations Map, TM Forum, October 1998, Morristown, NJ.

3.7 Management Frameworks and Applications

Kornel Terplan

Management frameworks consist of an application platform and management applications. The application platform consists of management services, computing hardware, and operating system. Management frameworks show unique features that differentiate them from management systems, particularly from proprietary management solutions. This contribution will introduce these differentiating features first. This step is followed by showing typical framework examples for both telecommunications service providers and enterprise users. NetExpert from Objective Systems Integrators, TeMIP from Compaq, TNP from Network Programs, TNG from Computer Associates, OpenView from Hewlett Packard, and FrontLine Manager from Manage.Com are briefly discussed. Management applications can be categorized as device-dependent — provided by vendors of networking equipment, and device-independent — usually provided by independent software vendors (ISV). For the integration between the application platforms and applications, APIs are provided by the framework suppliers assuming that these APIs are going to be supported and used by application providers.

3.7.1 Evolving Management Frameworks

The enrichment of application platforms by generic and specific management applications makes them more poweful. Figure 3.7.1 illustrates this evolutionary process. When application platforms are introduced to the market, they usually provide support for a few management applications in addition to core services that are part of the framework. If the framework is well accepted by users, the number of vertical management applications grows. Step by step, they will be partially or fully integrated into the framework.

3.7.2 Features and Attributes of Management Frameworks

Actually, there is no scientific definition of a management framework. In order to decide whether certain products qualify as a framework, an elaborated list of attributes is going to be addressed first. When products are able to support the basic framework attributes, they are qualified as frameworks. The advance attributes may then serve as differentiators between frameworks. Figure 3.7.2 shows the basic architecture of management frameworks. It consists of a runtime environment, development runtime tools, and application programming interfaces (APIs). The runtime environment is subdivided into management applications, management services, and the basic infrastructure (GHET97). Management services can be further subdivided into basic and advanced services, differentiating management frameworks from each other. Between the runtime environment, development tools, and the implementation, there are APIs to interconnect pieces with each other.

3.7.2.1 Basic Infrastructure

Basic infrastructure concentrates on the hardware and software features of the management frameworks. The most important attributes are listed below.

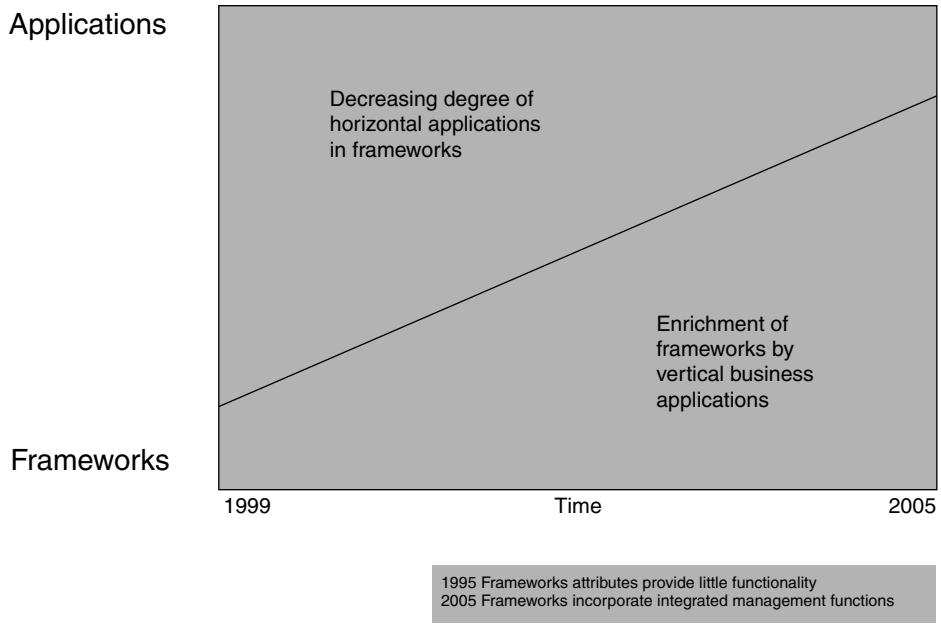

FIGURE 3.7.1 Evolving management frameworks.

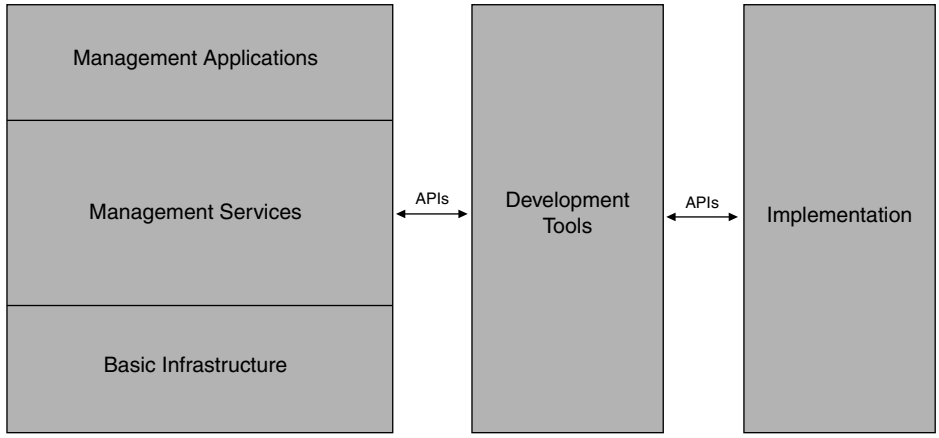

FIGURE 3.7.2 Architecture of management frameworks.

3.7.2.1.1 Hardware Platform

Hardware platform of the product involves a wide variety of items, including Intel 386/486, Pentium, HP 9000, RS/6000, Sun Sparc, Tandem, Alpha, System/88, and eventually others. Backup support should be addressed here as well.

3.7.2.1.2 Operating Systems

The industry expects a certain streamlining or, in other words, a shakeout of operating systems supporting the management platform. At present, the operating systems to be considered include: AIX, DOS/Windows, OS/2, SunOS, Ultrix, Sinix, Unix, Windows, Windows NT, and eventually others. Future solutions will concentrate around Unix versions and Windows NT.

3.7.2.1.3 Directory Services

Management frameworks deal with a great variety of entities and a great number of resources. The service of allocation of human-readable names to each managed resource (object) is the goal of directory services. Directory services are based on commonly agreed standards that model the naming paradigm, provide naming notations, allocate identifiable names to managed resources, translate names into physical addresses of resources, and ultimately provide location transparency of the resource in the system. All these considerations are valid for network and systems management frameworks. Framework management services and management applications use naming information from the directory services in order to perform their functions in relationship with managed resources and other management frameworks. The principal directory service requirements are (GHET97):

- Global information directory service and universal access to directory information
- Separation between the names of managed objects and the underlying physical networks
- Translation capabilities between various directory systems
- Translation between logical addresses and network addresses or routing addresses
- Storage of directory information and access to directory information, including metadata
- APIs in order to easily incorporate directory services into applications

Considering directory service capabilities, the following components are used as examples (GHET97):

- Directory services users: People, management applications, electronic messaging, routers/servers, other management framework services
- Resources requiring naming: People, organizations, computers, processes, files, mail boxes, network devices, printers, object class abstractions, object instances, management applications, management services, management agents
- Directory system types: Centralized, distributed, standard, proprietary; interpersonal communications directory (human), intersystem communications directory (computers and software systems)
- Directory service generic operations: Query (read, list, search, compare), modification (add/remove entry, modify entry, modify naming space, quit), binding/unbinding (security authentication)

Major directory systems include:

- OSI Directory Services (X.500)
- Internet Domain Name System (DNS)
- OSF Cell Directory Service (CDS)
- DEC Naming Service (DECdns)
- Novell Netware Directory Service (NDS)
- AppleTalk Name Binding Protocol (NBP)
- Windows NT Directory service
- Lightweight Directory Access Protocol (LDAP)
- Netscape Directory Server
- Banyan StreetTalk

3.7.2.1.4 Time Services

In distributed systems and network environments, where processes and applications are running on different machines, it might happen that time differences occur between systems. The time difference becomes critical when correct time stamps determine sequencing of events, job scheduling, measurements timing, and reporting intervals. A consistent use of time services is imperative when dealing with global networks, which span multiple time zones. Time service general requirements can be summarized as follows (GHET97):

- Use of an absolute, universal, coordinated time reference source
- Consistent synchronization services across hardware and software components
- Translation of universal time to local time for networks spanning multiple time zones
- Automatic resynchronization of manager and agent platforms after service interruptions
- Ability to operate in a heterogeneous computer and network environment
- Ability to keep the service running in case of major network instabilities
- Ability to provide both clock corrections and time source synchronization

In most cases, Internet Network Time Protocol (NTP) is used. Its primary reference time source is the absolute Universal Time Clock (UTC) or sources directly derived from UTC.

3.7.2.1.5 Software Distribution

In complex environments, management systems are usually distributed. They consist of servers, clients, and the communications paths between them. In order to keep them in synch, software versions or releases running on servers and clients must be compatible with each other. Manual software distribution is too slow and not reliable enough. Electronic software distribution offers two popular alternatives: push and pull. Distributing software by push offers easier scheduling, better automation, and does not require the physical presence of administrators. But, receiver servers and clients should be prepared for the distribution. Distributing by pull offers better control by administrators, and changes during distribution at a price of low automation. Scheduling is flexible and depends solely on human decisions. At this time, pull is priviliged in web environments.

3.7.2.1.6 Security Services

Open distributed network environments consist of an increasing number of interconnected computing resources, networks, and users. The networks are no longer closed networks but mixtures of private and public networks. The networks include heterogeneous components, which has bearing on security services as well. Security of a network depends on the security of adjacent networks or other trusted partners. Frequent changes, like adding new resources and new users, bring additional concerns regarding security. Security can be seen as the security management functional feature built into certain management applications, namely security management applications. Since management frameworks control resources, security becomes an issue, as other framework services have to operate securely in order to make the whole system secure. Security is often embedded in framework services such as communications, database management, and object manipulation sevices, which perform management oprtaions.

Basic security requirements include (GHET97):

- Support for basic security features such as authentication, access control, and data integrity
- Ability to protect the system against potential intrusions
- Security features should be included in the whole life cycle of software development
- Distinctions in user security access profiles according to their role in the network
- Ability to group resources and users and apply common security policies to them
- Need to test security features and services against possible violations
- Protection of passwords and encryption keys by storing them in protected, encrypted files
- Mechanism to provide automatic clearing of disabled user accounts, user I.D.s, and passwords
- Capacity to communicate security data in a secured fashion

In typical management environments, management frameworks are protected by using a combination of the following security services:

- Identification of users: Identification is a unique representation of a user, computer, application, or remote system in order to provide accountability and to record action of the identified entity.
- Access control: Allows the requesting party to actually access the system and networking resources if the party was authorized to do so. It is supported by login functions where passwords are built in.

- Authentication: The verification of the entity prior to accessing the system and networking resources. In this case, the entity should prove its identity by using various techniques, such as personal attributes, digital signatures, and others.
- Data privacy: An encryption mechanism using trusted third-party secret keys. Through a combination of private and public keys, the encrypted information can be verified for integrity and accessed for processing.
- Data integrity: A security feature which allows verification of cryptographic data checksums. The correctness of this verification is a proof that the data was not tampered with or corrupted through network transmission.
- Security auditing: Allows the generation of audit logs. These logs should be encrypted and protected against unauthorized access attempts.

3.7.2.2 Management Services

Management services address more specific items toward management applications. The most important features are listed below.

3.7.2.2.1 Communication Services

Network architectures: The targeted networks to be managed are very different. Many products are expected to manage legacy networks and more open networks at the same time. The most widely used protocols supporting network architectures include DECnet, IPX/SPX, OSI, SNA, DSA, APPN, TCP/IP, Guardian, and eventually others. Capabilities of managing SNA and TCP/IP are the highest priority.

Network management protocols: The products are expected to support at least SNMP. It is an additional advantage when they can do more. SNMP support may include the capabilities of working with proxy agents that are capable of converting non-SNMP into SNMP. Protocols to be supported include CMIP, CMOT, LNMP, NMVT, RMON, SNMPv1, SNMPv2, and eventually DMI to manage desktops.

The management platform provides SNMP support in several ways. First and foremost is the ability to poll SNMP devices and receive SNMP Traps, as described previously. However, in order to configure polls on MIB variables of various devices, one must first know what those variables are. Management platforms provide MIB "browsers" for this purpose. A MIB browser queries user-selected SNMP network devices and displays their MIB values. In addition, most platforms can display line or bar graphs of those MIB values, provided they are in numeric form (counters, etc.).

MIB browsers are crude tools at best, displaying raw and often cryptic, low-level device information. For this reason, platforms also provide MIB application builders that allow users to quickly create applications for displaying information on MIB objects in a more meaningful way. MIB applications may include graphing real-time information on selected network nodes. However, even MIB applications builders are limited in supporting the high-level analysis more openly provided by third-party applications.

MIB compilers allow users to bring in third-party, device-specific MIBs (also called "private" or "extended" MIBs) and register them with the management platform. While most platforms ship with a number of third-party MIBs, they do not include all possible MIBs from all vendors. A MIB compiler is necessary for adding support for third parties whose MIBs are not shipped as part of the standard platform.

Some MIB compilers are more robust than others. Some MIB compilers will fail or abort processing if there is an error in the MIB being compiled. Unfortunately, errors in third-party MIBs are not rare. Therefore, it is desirable to have a MIB compiler that can flag errors and recover, rather than stop dead.

3.7.2.2.2 Core Management Services

The management framework is expected to offer core services and interfaces to other applications. The basic management applications to be provided are discovery/mapping, alarm management, and platform protection.

Discovery and Mapping

Device discovery/network mapping discovery refers to the network management system's ability to automatically learn the identity and type of devices currently active on the network. At minimum, a

management platform should be capable of discovery-active IP devices by retrieving data from a router's IP tables and Address Resolution Protocol (ARP) tables.

However, even this capability does not guarantee that all IP devices on a given network will be detected. For example, relying solely on routing tables is inadequate in purely bridged networks where there are no routers. Thus, a more comprehensive discovery facility should also include other mechanisms such as broadcast messages (PING and others) that can reach out to any IP device and retrieve its address and other identifying information.

On the other hand, discovery mechanisms that rely completely on broadcasting (e.g., PING) will incur a tremendous amount of overhead in finding devices out on the network. Ideally, a management platform should support a combination of ARP data retrieval and broadcasting.

Furthermore, a complete network discovery facility should be capable of detecting legacy system nodes, such as DECnet and SNA. Currently, most platforms rely on third-party applications or traffic monitoring applications to supply discovery data on non-TCP/IP devices.

Another desirable feature is the ability to run automatic or scheduled "dynamic discovery" processes after the initial discovery, to discern any changes made to the network after the initial discovery took place. In large networks especially, overhead and consumed bandwidth for running a dynamic discovery process continually in background mode may be too great; therefore, the ability to schedule discovery at off-peak hours is important.

It is also important for the user to have the ability to set limits on the initial network discovery. Many corporate networks are now linked to the Internet, and without predefined limits, a discovery application may cross corporate boundaries and begin discovering everything on the global Internet. Some management platforms allow users to run discovery on a segment-by-segment basis. This can help the discovery process from getting out of hand too fast.

Many management platforms are capable of automatically producing a topological map from the data collected during device discovery. However, these automatically generated maps rarely result in a useful graphical representation. When there are hundreds of devices, the resulting map can look very cluttered enough to be of little use.

Even when the discovery process operates on a limited or segment-by-segment basis, there is eventually going to come a time when the operator must edit the automatically generated network map to create a visual picture that is easier for human beings to relate to. Therefore, the ability to group objects on the map, and move them around in groups or perform other types of collective actions, can be a real time-saving feature.

Alarm Capabilities

Management platforms act as a clearinghouse for critical status messages obtained from various devices and applications across the network. Messages arrive in the form of SNMP traps, alerts, or event reports when polling results indicate that thresholds have been exceeded.

The management platform supports setting of thresholds on any SNMP MIB variable. Typically, management platforms poll for device status by sending SNMP requests to devices with SNMP agents, or Internet Control Message Protocol (ICMP) echo requests ("pings") to any TCP/IP device.

The process of setting thresholds may be supported by third-party applications or by the management platform. Some platforms allow operators to configure polls on classes of devices; most require operators to configure a poll for each device individually.

Most platforms support some degree of alarm filtering. Rudimentary filtering allows operators to assign classifications to individual alarms, such as informational, warning, or critical, triggered when thresholds are exceeded. Once classifications are assigned, the user can specify that only critical alarms are displayed on the screen, while all other alarms are logged, for example.

More sophisticated alarm facilities support conditional alarms. An example of a conditional threshold may be "errors on incoming packets from device B > 800 for more than 5 times in 25 minutes." Conditional alarms can account for periodic spikes in traffic or daily busy periods, for example.

Finally, the platform should support the ability to automatically trigger scripts when specific alarms are received.

3.7.2.2.3 User Interface Services

GUI's basic job is to provide color-coded display of management information, multiple windows into different core or management applications, and an iconic or menu-driven user interface. By providing a standardized interface between the user and the underlying tools, the GUI simplifies what a user needs to learn and provides a standard tool for application developers.

Most management operations are available from a menu bar; others from contex menus. Point-and-click operations are standard features, as is context-sensitive help. Most platforms allow some degree of customization of maps and icons.

While most platform GUIs are the same, there can be a few subtle differences. Some GUIs have larger icons than others. While this makes it easier to read information on the icon and distinguish status changes more quickly, a screen can quickly become cluttered with just a few large icons. Icon size is strictly a matter of user preference. The most widely used GUIs are Motif, OpenLook, OS/2 Presentation Manager, and Windows.

3.7.2.2.4 Database Services

The database is the focal point for key data created and used by the management applications. They include MIB data, inventories, trouble tickets, configuration files, and performance data.

Most platforms maintain event logs in flat-file ASCII format for performance reasons. However, this format limits the network manager's ability to search for information and manipulate the data. Therefore, links to relational database management systems (RDBMSs) are now important aspects of the framework architecture.

A RDBMS is essential for manipulating raw data and turning it into useful information. Users can obtain information from a RDBMS by writing requests, or queries, in Structured Query Language (SQL), a universally standard language for relational database communication.

While most management platforms also supply report writer facilities, these tools are generally not top-notch. However, most higher quality third-party reporting applications can extract data from a RDBMS using SQL.

3.7.2.2.5 Object Manipulation Services

Object-oriented and object-based technologies are helpful in relation to user interfaces, protocols, and databases. The use of object request brokers (ORB) and CORBA provides a glue needed to accomplish interoperability between heterogeneous systems. These services provide support for information exchange between objects as abstractions of physical and logical resources ranging from network devices computing systems resources to applications and management services. It includes operations on MIBs, object support services providing location transparency for objects exchanging requests and responses, persistent storage for MIBs, and support for object-oriented applications development.

3.7.2.2.6 Network Modeling Services

Network modeling is an artificial intelligence capability that can assist in automated fault isolation and diagnosis as well as performance and configuration management. Modeling allows a management system to infer status of one object from the status of other objects.

Network modeling is facilitated by object-oriented programming techniques and languages such as C++. The goal of modeling is to simplify the representation of complex networks, creating a layer of abstraction that shields management applications from underlying details.

The building block of this technology is the model,which describes a network element such as a router. A model consists of data (attributes) describing the element as well as its relationships with other elements. Abstract elements such as organizations and protocols can also be modeled, as can nonintelligent devices such as cables. A model may use information from other models to determine its own state; modeling can reduce the complexity of management data and highlight the most important information. In this

way, fault isolation and diagnosis can be automated. In addition, models can be used to depict traffic patterns, trends, topologies, or distributions to assist in performance and configuration management.

3.7.2.3 Application Programming Interfaces (APIs) and Development Toolkits

API and developer's toolkit platform vendors encourage third-party applications by providing published APIs, toolkits that include libraries of software routines, and documentation to assist applications developers. Another aspect to this effort is the "partners programs" — the marketing angle of encouraging third-party applications development.

An API shields applications developers from the details of the management platform's underlying data implementation and functional architecture. Management platform vendors generally include in their developer's kits several coded examples of how APIs can be used, as well as the APIs themselves.

In most cases, when an application takes advantage of platform APIs, it must be recompiled with the platform code, resulting in a tightly integrated end product. Many ISVs and other third-party developers lack resources necessary to pursue this level of integration. Or, perhaps a more accurate way of stating this is that ISVs aren't convinced that putting out the extra effort to fully integrate their applications with all leading management platforms will result in a proportionally larger revenue stream. ISVs and other third-party developers face a choice: tightly integrate their products with one management platform vendor, or loosely integrate them with all leading platform providers. Most third parties have chosen the latter route, as they are unwilling to turn off prospective customers who may have chosen a different platform vendor as their strategic management provider.

As a result, at least 80% of the third-party applications available today are only loosely integrated with the underlying management platform — at the menu bar — and completely ignore APIs and other environment libraries. This is expected to change as the market matures, and as platform vendors begin to offer high-level APIs which make porting applications from one management platform to another into an almost trivial exercise.

In summary, published APIs and libraries make it possible for lSVs and other third parties to write applications that take advantage of other basic services provided by the management platform. To date, few third parties have taken full advantage of platform APIs, although this is expected to change over the next several years.

3.7.2.4 Management Operations Support Services

Any management framework consists of framework services and management applications. The services are implemented as a set of related processes, databases, and file sets. The basic thrust of management implies collection and processing of management-related information. The coordination of all the framework processes, including those which are part of the development environment, is done through additional framework components commonly called management operations support services. These services are also responsible for application integration with framework services, and for multiple national language systems support.

Management frameworks are basically a set of interconnected software programs which run on one or more computing platforms. Management operations support services provide supervision, coordination, maintenance, and management of processes, applications, and databases which are part of the management framework. The requirements of management operations support services are the following (GHET97):

- Facilitating interactions between framework services
- Allowing overall coordination and supervision of background processes
- Supporting integration between management services
- Allowing configuration and customization of framework services and associated processes
- Supporting registration of management applications which run on management platforms
- Providing easy integration of management applications with framework services
- Supporting multiple national language systems
- Facilitating incorporation of management information models into frameworks

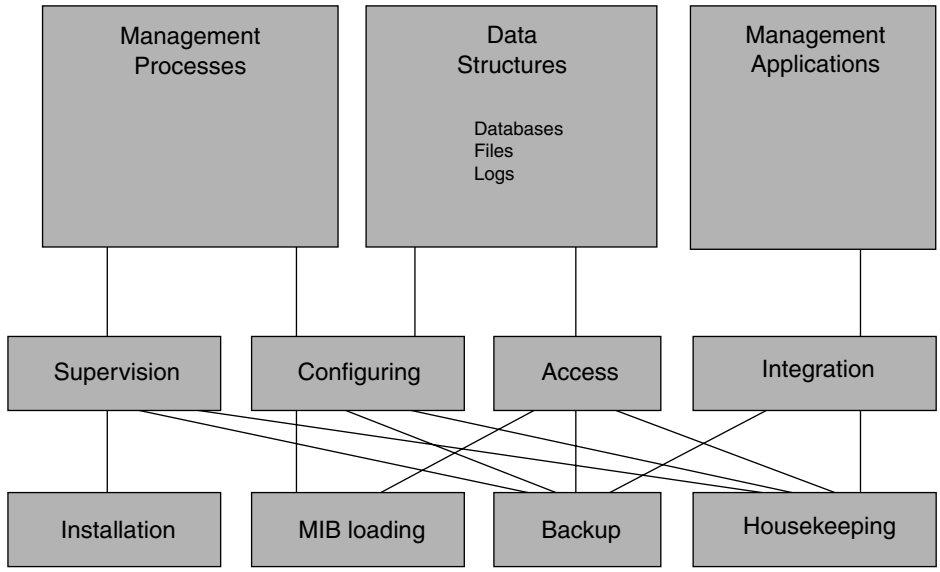

FIGURE 3.7.3 Overview of management operations support services.

- Supporting installation of framework services and management applications
- Supporting MIB loading, backup, and clean-up facilities
- Supporting distribution of management frameworks services and associated databases

This list of requirements indicates that management operations support services play a critical role in monitoring, administration, and management of the management framework itself.

The structure of management operations support services is characterized by a layered architecture. The upper layer consists of management processes, data structures, and management applications (GHET97). The middle layer presents important support functions, such as supervision and synchronization of management processes, configuring processes and databases, access to databases and files, and integration between framework services and management applications. The lowest layer consists of tools, supporting installation, MIB-loading, backups, and other usual housekeeping functions. Figure 3.7.3 shows these layers.

3.7.3 Management Framework Examples for Telecommunications Providers

These examples show very powerful and scalable frameworks with a number of capabilities for both wireline and wireless services. In all cases, third-party management applications can be integrated into the frameworks.

3.7.3.1 TeMIP from Compaq

At the highest level, TeMIP consists of a management information repository (MIR) for storage of data stuctures, functions, and management information, an executive kernel responsible for supporting all the interactions beween components, and a set of interfaces to all the management modules belonging to the framework. Figure 3.7.4 shows the TeMIP architecture.

Three types of management modules interface the kernel:

- Access modules which provide access to various agents attached to real management entities such as physical network elements or systems logic resources
- Presentation modules provide the user interfaces
- Functional modules provide the actual management services such as event management, object manipulation, and management operation support services

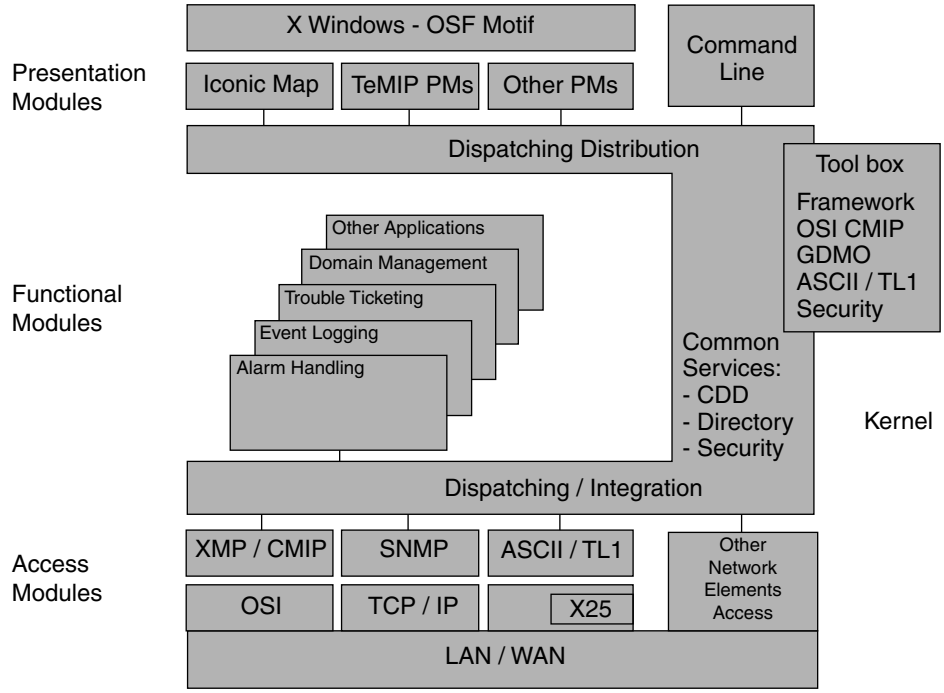

FIGURE 3.7.4 TeMIP architecture from Compaq.

These management modules are a set of cooperative processes rather than independent ones. Compaq/Digital has been adding access and functional modules over the last couple of years, such as SNMP, OSI CMIP, ASCII/TL1, and TMN support. A more detailed view of the framework can be seen in Figure 3.7.5 (GHET97).

An important emphasis is placed on the TeMIP distributed framework which allows any of the constituent modules to run on physically distributed systems. Each of these systems is considered a peer director. Among directors, some play the role of servers, others play the role of clients. Direct communications and management information exchange is provided only between director servers. The implementation of the TeMIP architecture can range from a standalone centralized management system to hierarchical or a cooperative network of manager topologies.

The TeMIP GUI is based on OSF/Motif and XWindows Systems and provides a common view of all the managed resources. The Icon Map PM provides presentation and language localization to the alarm handling, event logging, and trouble ticket FMs. The icon map provides map windows, a navigation box, graph windows, and a toolbox for customization. Forms and command line interfaces are also available.

The platform provides many generic functional modules. The alarm handling and the event logging FMs are based on ISO standards. A log panel window allows the user to customize the logging environment. The trouble ticket FM is based on the recommendations of the Telemanagement Forum. The performance analyzer FM provides normalized and statistical data for TCP/IP hosts, RMON probes, and DECnet nodes. It collects information about DECnet/OSI end systems, data links, intermediate systems, routing ports and routing circuits, circuits, nodes, and protocols. The statistics collected include throughput rates, counts, averages, overhead percents, and utilization metrics.

The information manager is the platform's object request broker and is similar but not compatible with CORBA from OMG. It receives requests from clients along with their arguments. Then, acting as a client, the information manager connects through a RPC binding to the appropriate server. Location transparency is achieved through Distributed Name Services, which provides a global directory service.

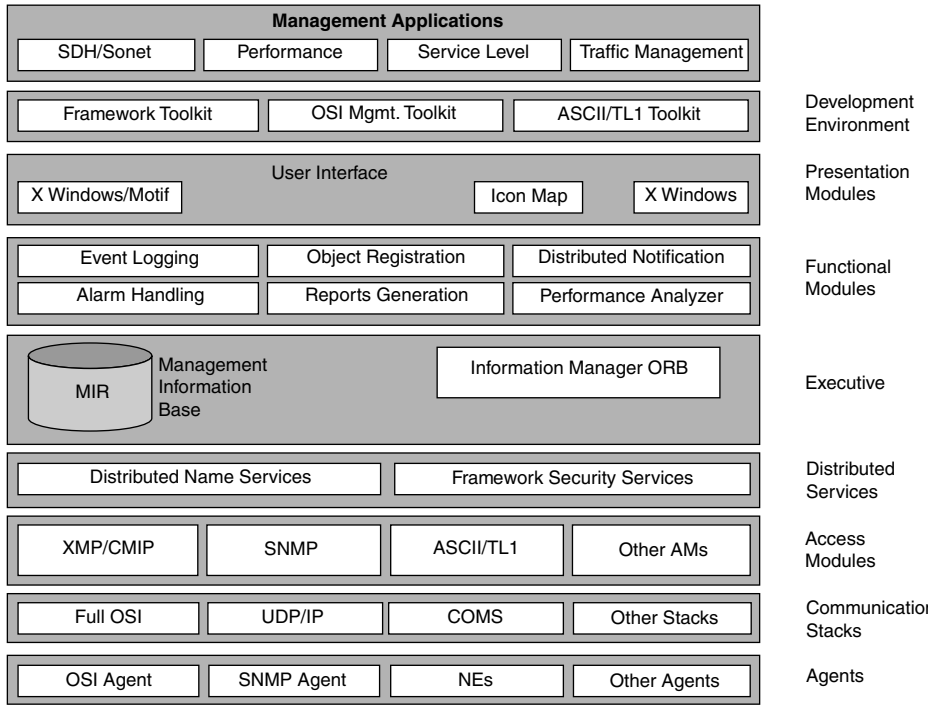

FIGURE 3.7.5 TeMIP framework in detail.

Security services consist of access control (access control filters, user profiles, access control management), logging of operator commands (storage of prefiltered commands entered by users), and a security development toolkit.

The TeMIP framework provides access to managed resources through access modules. All of the relevant network protocols are supported by the framework. The SNMP AM supports the MIB II management information base. In addition, a MIB compiler is provided to check the Concise MIB syntax and to support loading of the MIB into MIR. The SNMP AM allows get and set operations on the agents and can test reachability of an object at the IP level by using the ICMP ping protocol.

The TeMIP applications map is shown in Figure 3.7.6. This map includes three major groupings for external management applications:

- Network management
- Telecommunications management
- Unix systems management

Strengths and weaknesses of the TeMIP framework are summarized in Table 3.7.1.

3.7.3.2 NetExpert from Objective Systems Integrators

The NetExpert framework consists of a series of coordinated modules that fall into three general groups:

- External network element and non-NetExpert subsystem gateways
- Object persistence and behavior servers
- User/operator workstations and web interfaces

NetExpert is a robust, scalable, and distributable archietcture that supports a high degree of configuration flexibility while maintaining individual component independence. Easy to use, modify, and initiate, it is quick to roll out and integrate with existing platforms and management applications.

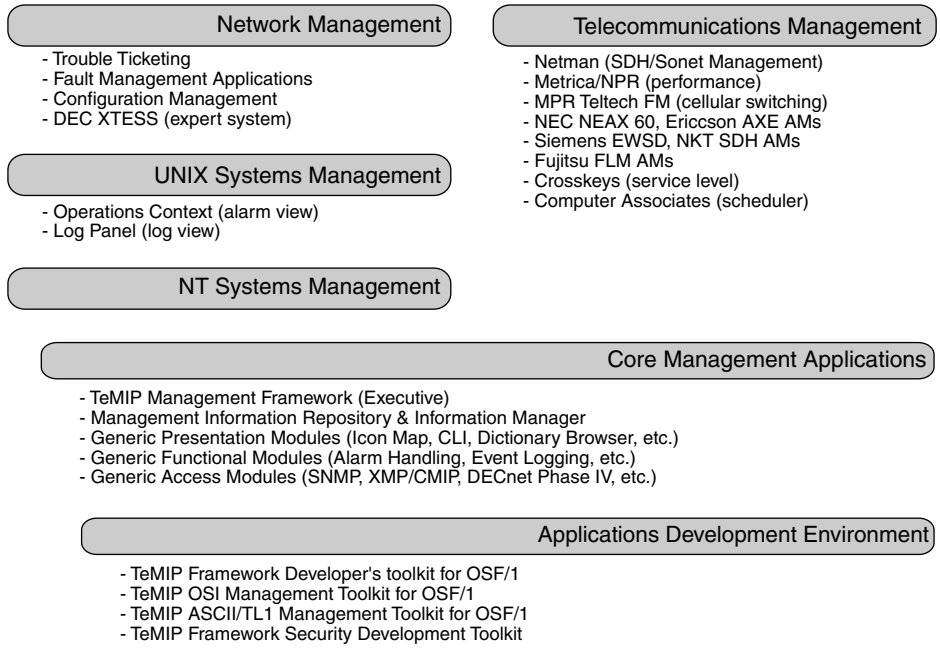

FIGURE 3.7.6 Applications map of TeMIP.

TABLE 3.7.1 Strengths and Weaknesses of TeMIP

Strengths
Modular and functionally distributed architecture
Direct communication capability between TeMIP director servers
Framework functional modules based on OSI management standards
Incorporation of an object request broker
Home-grown management framework design
Distributed notification mechanism
Distributed name and security services
Policy-based management domains selection capability
Partnership with telecommunication service providers and manufacturers
Self-management capabilities
Weaknesses
Very complex architecture and development environment
The Information Manager is not CORBA compliant
Complex documentation
Long learning curve
Small market share
Proprietary internal database and database API
Limited set of systems management application availability
Limited choices of hardware and operation systems platforms
No scaled-down TeMIP management platform alternative

Figure 3.7.7 provides a high-level description of how NetExpert receives "from" and "send" messages to external network elements and operations support systems.

The main attributes of principal subsystems are the following:

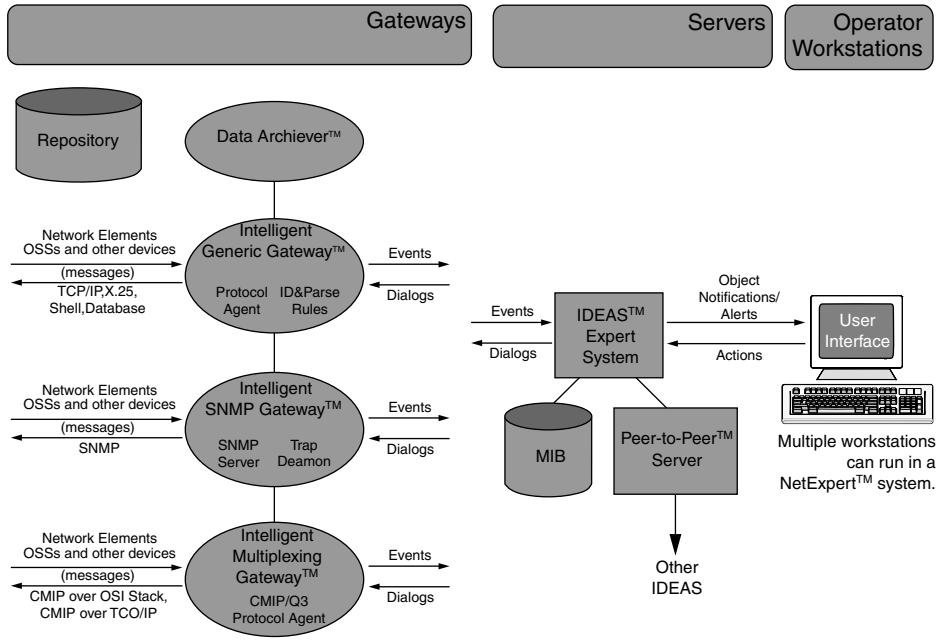

FIGURE 3.7.7 NetExpert framework operational overview.

Gateways:

- Receive raw data from network elements
- Identify important messages, parse relevant data into attributes, and package them into events
- Perform analysis
- Forward events to the IDEAS Expert System Server
- Generate and send dialogs and polls to devices and receive responses
- Forward messages to DataArchiver

Servers:

- Receive events from gateways
- Perform analysis and execute rules
- Generate alerts and send to operator workstations
- Initiate dialogs and polls and send to gateways
- Modify MIB values
- Forward notification to peer systems

Operator workstations include the following modules:

- Gateway control
- Visual agent client windows
- Alert display
- Command and response system
- Managed objects configuration system
- Trouble ticket
- Report maker
- Data browser
- Interface to pagers

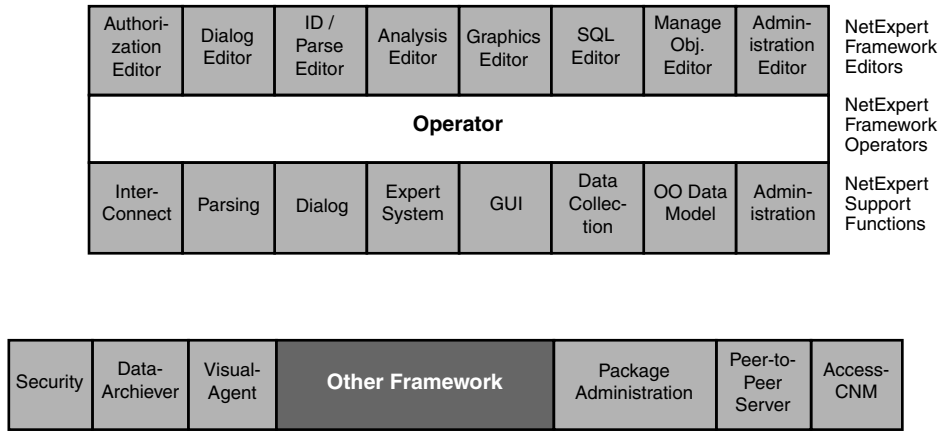

FIGURE 3.7.8 NetExpert functional framework.

The NetExpert framework is a set of modules covering the basic functions that a distributed frameworks needs, including gateways to the system, a way to send messages and events, the intelligence to act on those events, and a consistent operator interface. A customer can distribute these modules across a network, gaining the foundation required to monitor continuous and large volumes of events and traffic. The framework is controlled by rules that replace complex programming languages and enable network analysts to model desired system behaviors. Rules are written with the product's implementation tools. Existing rule sets, called application components, eliminate the complex traditional development process, which entails writing requirements and building a complete solution from scratch. Rule writing is estimated as 10–15 times more productive than traditional development methods.

The functional framework consists of editors, operators, support functions, and other framework enablers. These are shown in Figure 3.7.8.

NetExpert's modifiable application packages provide a comprehensive subset of functions. These can be further tailored to individual customer requirements. This is how the framework accomodates configuration-specific solutions and the demands of the customer's business model. Because they are object oriented, these rule packages can deliver a large number of services; manage any number of tangible elements, such as switches or routers; and model intangible elements, such as knowledge of subject matter experts. Rules make it possible for the same NetExpert framework to manage diverse networks, such as digital cellular, traditional telephony, high-speed data, or hybrid fiber/coax.

Application rules ride on top of the NetExpert framework. They are categorized as point, domain, or corporate level application packages. The differences between each depend on the business focus they are designed to address. Point applications define the native messages required by a network element during, for example, the provisioning process. Domain applications group higher level commands into those associated with, for example, all switch or transport network devices constituting a service provider's network. Corporate applications perform, manage, and control functions associated with the domain- and point-level applications. The layering of corporate, domain, and point applications is illustrated in Figure 3.7.9.

The framework running in concert with NetExpert applications enables users to generate revenue by quickly delivering new services. However, getting to market first is not enough. With NetExpert, users protect past investments, increase the life span of aging equipment, incorporate new elements, and integrate disparate management systems and software rule set packages. Another advantage OSI users have is their ability to deploy network management systems and OSSs in formerly uncharted niches, and integrate these with existing infrastructures. Users are closer than before to automating their business models because OSI delivers the tools they need to translate key processes across systems that support entire telecommunications operations.

Table 3.7.2 lists strengths and weaknesses of NetExpert.

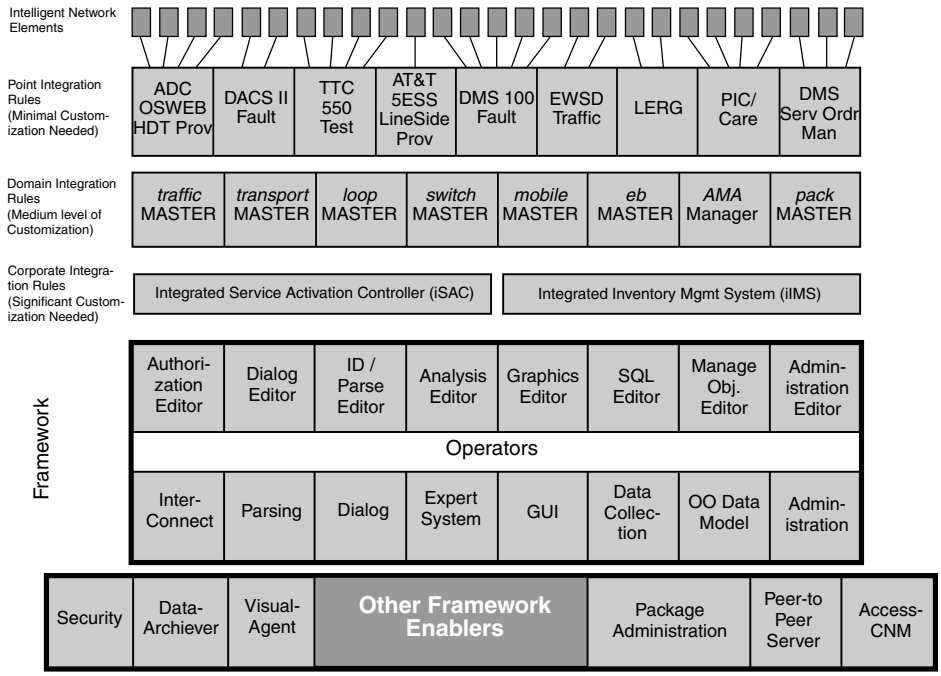

FIGURE 3.7.9 Rules-based applications of NetExpert.

TABLE 3.7.2 Strengths and Weaknesses of NetExpert

Strengths
Support of accelerated design procedures
Flexible integration of new network elements and OSSs
Cross-vendor functions and domain correlation
Common maintenance and operations procedures
Integrated problem resolution capability
Substantial reductions in software costs for development and operations
Support of multiple hardware platforms
Support of CORBA for interprocess communications
Heavy use of Java to support presentation services
Support of Web-based front ends
Large market share with wireless operators
Weaknesses
Portability of rules sets is limited to non-Unix environments
Extensive training is required
Learning curve for subject matter experts is relatively long
Lack of third-party support for integrating management applications
Rule sets are heavily fault management-oriented
No scaled-down alternative for smaller operators
No presence in enterprise environments

3.7.4 Management Framework Examples for Enterprise Users

These examples represent flexible and scalable frameworks with a number of integration capabilities. Some management applications provided by ISVs are the same or at least similar to those provided for the telecommunications industry.

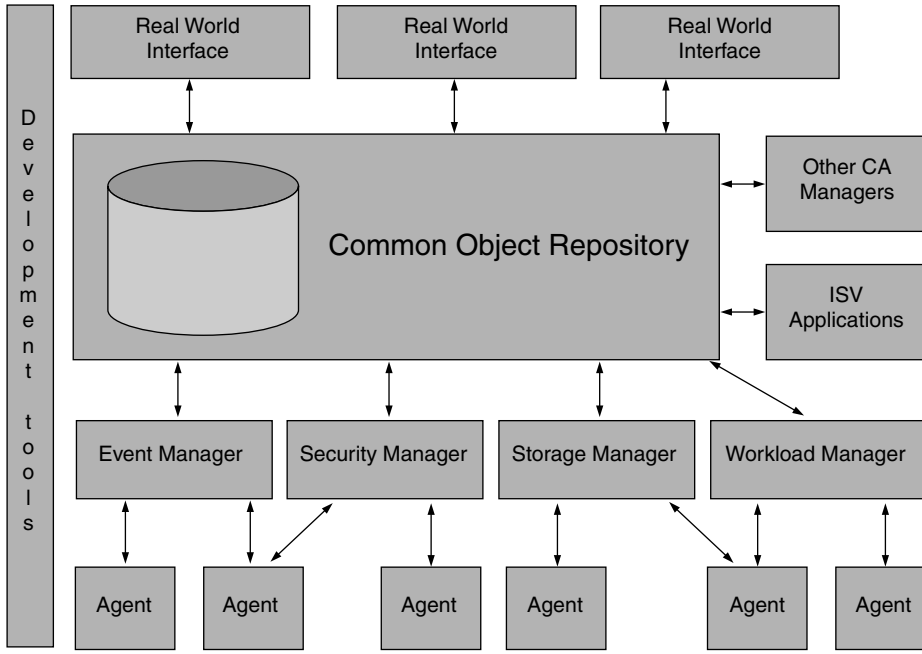

FIGURE 3.7.10 Simplified view of the TNG architecture.

3.7.4.1 TNG from Computer Associates

TNG features, as an absolute novelty for a management platform, a 3-D, animated, graphical user interface called the Real World Interface. The core platform is bundled under the Common Object Repository component which hides the object manipulation and object storage processes. Objects representing the abstraction of actual managed resources and objects created by the platform services are stored in the platform object database. Query and search capabilities allow core management functions and applications to access the management information. Figure 3.7.10 shows the simplified view of the TNG architecture.

The availability of a Java browser, based on either a 2-D or VRLM 3-D interface, provides an alternative graphic environment. This Web interface, in particular, delivers on the framework's promise of managing everything from anywhere.

TNG includes many functional modules, which provide event management, security management, storage management, workload and performance management, as well as backup and recovery functions. Management of distributed resources is based on a manager–agent infrastructure which relies on a mix of proprietary and standard agents. TNG allows scalable, multilevel, hierarchical build-up of manager–agent structures as required by managing large enterprise networks.

The following application packages are running on top of the TNG platform:

- Software delivery
- Advanced help desk
- Open Storage Manager
- Single sign-on
- Internet Commerce Enabler

The company is acquiring products, but more than in the past, pays a lot of attention to integration. Table 3.7.3 summarizes strengths and weaknesses of the TNG architecture.

TABLE 3.7.3 Strengths and Weaknesses of Unicenter TNG

Strengths
Unicenter TNG applications are running on multiple platforms
Extensive experience with mainframe-based management applications
Integration of systems and network management
Interoperability capabilities between various CA-products
Use of advanced 3-D techniques
Multiple alliances
Use of neural technology to deal with large data volumes
Support of Web technology
Promotion of a developer partner program
Filling functionality gaps by acquiring the right best-of-breed products
Weaknesses
Limited experience in network management
Proprietary agent implementations and information exchange
Customer support could be better
No telecom industry-specific management applications are available
Limited application development tools
Support for open, standard systems and APIs is not readily seen in real products
Quality of documentations is not always good
No low entry, PC-based framework version available

3.7.4.2 OpenView from Hewlett-Packard (HP)

The OpenView family provides an integrated network and systems management solution. It consists of a number of products from HP and also from Solution Partners. The most important components of the OpenView family are:

- Network Node Manager (NNM) — meets the requirements for a powerful SNMP core solution.
- IT/Operations — an advanced integrated operations and problem management solution for networks and systems.
- IT/Administration — an integrated solution for change management. It also includes inventory, asset, software, and user management.
- PerfView, NetMetrix, and MeasureWare — performance management solutions for networks and systems; may be considered as the foundation for service level management.
- OmniBack and OmniStorage — typical systems management solutions for powerful backup and storage management.

Figure 3.7.11 shows OpenView with its principal components.

HP is targeting OpenView Network Node Manager at managing the Internet/Intranet infrastructure rather than Web servers and services. HP promotes a three-tier OpenView strategy for managing and leveraging the Internet:

- Manage the corporate Intranet infrastructure, including network infrastructure, servers and Internet applications, and security
- Manage the infrastructure of Internet service providers
- Leverage Internet technologies in OpenView solutions.

The enhancements in NNM have significantly increased the product's scalability, making OpenView-based management of corporate Internet/Intranet infrastructures possible. Management of Web servers and applications is largely provided by the generic server and application management capabilities of IT/Operations. HP is targeting management of Internet service providers infrastructure through its HP OpenView DM offering, and HP OpenView Event Correlation Services. Internet service providers may

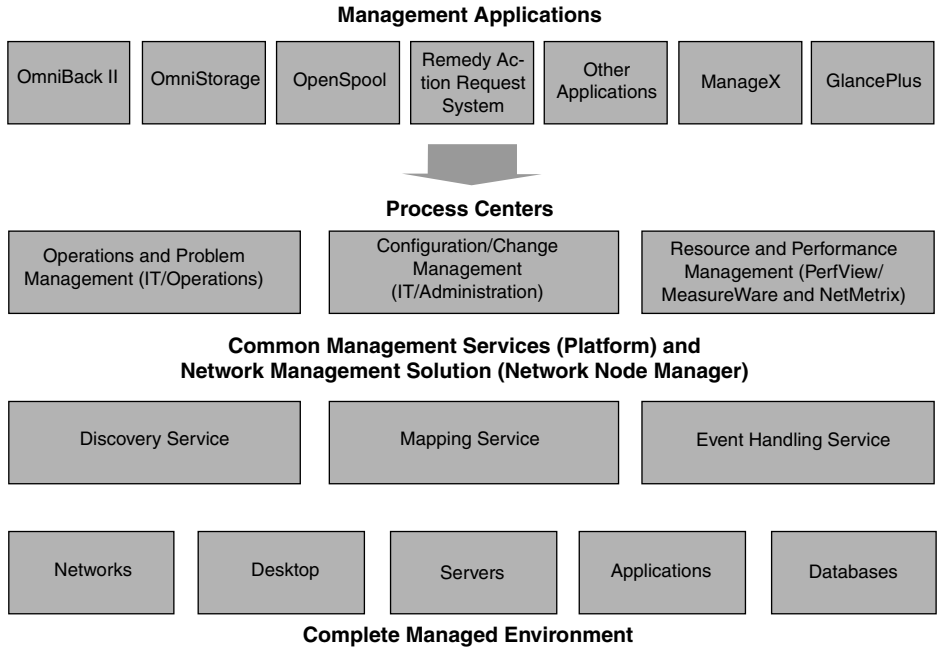

FIGURE 3.7.11 Principal components of OpenView.

include carriers, cable companies, value-added networks, and others. Finally, HP is exploring and prototyping Web technology extentions to OpenView products, including:

- Web access to OpenView event repositories for problem management support
- Web access to the OpenView map
- Internet as a software transport vehicle

Using IT/Operations for Internet Management

IT/Operations is capable of managing processes and applications running on any computer for which HP provides an IT/Operations agent. Supported systems include HP-UX, Solaris, AIX, SCO, and Windows NT, among others. IT/Operations agents are capable of intercepting SNMP traps, Unix logfile messages, and events generated when IT/Operations agents detect threshold crossings. Using these attributes, Netscape Commerce Servers can also be managed. They can run under HP/UX and support secure electronic commerce and communications on the Internet and TCP/IP Intranets. The server permits corporations to publish HTML-formatted documents (Web pages) and deliver them using HTTP. To ensure data security, the Netscape Commerce Server provides server authentication, data encryption, and user authorization. Communications support also includes the Common Gateway Interface (CGI) and the Secure Socket Layer (SSL) protocol.

To support manageability, the Netscape Commerce Server records several kinds of errors, all of which can be collected by an IT/Operations agent reading the logfile of the server. These errors include:

- Unauthorized — occurs when users attempt to access protected server documents without proper permission
- Forbidden — occurs when the server lacks file system permissions needed to execute a read or to follow symbolic links
- Not found — occurs when the server cannot find a document or has been instructed to deny a document's existence
- Server error — occurs when the server has been misconfigured or affected by a core dump, out of memory, or other catastrophic error

HP provides an IT/Operations template for handling these errors. Users can derive proper responses, including forwarding events to the appropriate IT/Operations or database operators, or triggering a script for deleting hypertext links to documents that no longer exist.

Each error type described above can be associated with error codes.

IT/Operations agents are capable of collecting these error messages and forwarding user-specified events to the IT/Operations console for operator attention and problem resolution. For example, in the case of server error, possible causes of the problem may include the following:

- CGI is not enabled on the Web server, preventing electronic commerce application from running permissions that have not been specified properly
- CGI script is not specifying a shell or other program to be run
- Syntax error in the script

Syntax errors are typically resolved by tweaking application scripts, which may be written in CGI, Practical Extraction and Report Language (PERL), or Tool Command Language (TCL). Many Web server applications for electronic commerce are written in C language and implemented with CGI, PERL, or TCL scripts.

Monitoring Web Page Availability with IT/Operations

IT/Operations can be deployed to monitor Web page status as well as Web Server status. Specific functions supported include:

- Monitoring Web access logfiles and error logfiles
- Monitoring the HTTP domain
- Viewing server access statistics
- Integrating the native Netscape administration and configuration tools into IT/Operations
- Starting up and shutting down the Web server and administrative interface
- Modifying access configuration

HP has developed a script that can be used by the IT/Operations agent monitor to check the availability of the Web server system, the HTTP port, and the Web page. The script uses the Korn shell, one of four major Unix command and script interpreters in use today. This script, designed primarily for Netscape Commerce Server, can theoretically be modified and extended to monitor other Web servers as well.

In order to meet the needs of today's IT organizations, a new Java-based user interface has been added to IT/Operations. The features and benefits of such an interface can be summarized as follows:

Ease of use:

- The Java-based interface combines the familiar concepts of IT/Operations with Windows-like concepts (similar to ExplorerView), to minimize training time and reduce the operator's learning curve.
- Most functions available in the IT/Operations Motif Operator user interface are supported in the Java version. The characteristics of Java also add functionality that is not available in the standard user interface, such as sorting and shifting columns in the IT/Operations' message browsers.
- It is available on the Windows NT operating system. The Java user interface allows the management of large heterogenenous environments from PCs running the Windows NT operating system.
- Through a special application-bank entry, local Windows NT applications can be tightly integrated with the Java user interface, resulting in a more powerful, integrated Windows NT operator workstation.

Scalability and distribution:

- The number of operators that can concurrently access IT/Operations from a Java user interface is greatly increased.
- This addresses the needs of customers with large environments. The size of environments is continually increasing.
- The Java user interface minimizes network traffic, enabling it to work over low-bandwidth lines.
- It is not common to have a LAN connection, for example, at home or in a remote office, yet management is available whether at home or on the road.
- The Java user interface runs on any machine with a Java-compliant browser or as a stand-alone application on HP-UX or the Windows NT operating system.

Lower total cost:

- Previously, the operators had to have a Unix workstation or special tools on NT PCs to obtain the same functionality.
- No additional IT/Operations software or hardware needs to be installed and maintained on the client systems, other than a Web browser.

The Java-based interface is designed to take up minimal resources on the client.
Table 3.7.4 summarizes the strengths and weaknesses of the OpenView architecture.

TABLE 3.7.4 Strengths and Weaknesses of OpenView

Strengths
Front runner in the implementation of the framework concept
Modular design with the SNMP and distributed manager framework
Good coverage in HP-developed management applications and products
Extensive coverage in Unix-based applications developed by ISVs
Leader in management framework source code licenses and OEM partnerships
HO provides a testing and certification program for partner applications
Distributed management consoles and GUI services
Partnerships to serve the telecommunications market
Manager-to-manager capabilities for integrated IT/Operations and IT/Administration
Support of application management standardization
Strong performance management applications
Web-based management front ends
Weaknesses
No built-in middleware object request broker
The communication between HP managers contains proprietary elements
Too many processes, too many API calls, no built-in security features
CMIP OSI stack support is environment dependent
Telecom industry support in management applications is still limited
Contains proprietary components and extentions
The application development tools are not yet mature
Delay in delivery of a common management repository
Insufficient customer support for platform implementation and tuning
Scalability is limited for large enterprise networks and systems
Delays in supporting the NT platform

3.7.4.3 FrontLine Manager from Manage.Com

FrontLine Manager unifies the management of rapidly prolferating intranet computing resources across network devices, systems, services, and applications. Designed for systems and network administrators, help desk agents and others who staff the frontline, responding to and solving user support calls. FrontLine Manager uses Web technologies to simplify day-to-day operational tasks and thus lower the cost of intranet management. It begins managing out-of-the-box by discovering and identifying resources, while creating a unified management view of the entire intranet environment. It goes beyond passive monitoring to identifying and diagnosing problems proactively. Embedded software intelligence determines the ideal operating state of each resource and notifies support staff when healthy operating conditions are exceeded. Rapid installation and ease of use are combined with low administrative overhead to maximize the productivity. It is a typical first-tier support tool. All management functionality and the unified management view are accessed via a standard Web browser. The components of FrontLine Manager are:

- FrontLine Manager Server: Each server manages a typical LAN or LANs supporting up to 255 network devices and systems, along with a base set of intranet services and applications. The FrontLine Manager server incorporates a Web server and a scalable object database. As a result, a distributed group of servers can manage up to 1,000,000 nodes.
- Web browser: All functionality and the unified management view are accessed via a standard Web browser. The browser interface simplifies the presentation of complex management information, while giving frontline managers the freedom to manage securely from anywhere, locally or remotely. An active window displays the most recent information for each managed resource.
- Managed Agents: Each managed resource has an associated SNMP or Java agent. Agents transmit management data to the server and can also conduct management tasks. A base set of SNMP and Java agents developed by Manage.Com are included with FrontLine Manager. Third-party SNMP agents already installed on network devices and systems can also be used.

FrontLine Manager is prebuilt with key management features needed to manage the majority of intranet comounting environments. No time is wasted on complex installation, customization, or integration. Quickstart installation automatically discovers all resources and begins monitoring them so that productive management can begin immediately. It begins by proactively discovering and classifying resources during installation, a process that completes within a few hours. It then associates an ideal operating state with each resource and monitors accordingly. If abnormalities are discovered, FrontLine Manager immediately begins to diagnose and isolate the causes. For maximum efficiency, it helps to identify and resolve problems before users report them. The intelligence to identify the healthy operating state of specific resources is built into the product. As a result, it is able to take samples of the intranet continually and determine its overall health. It also launches automatic analysis to diagnose and segment operating problems, often before they are reported by users.

Figure 3.7.12 shows the principal management functions.

This Web-based solution differentiates itself from individual device- or application-dependent products, because it integrates the management of network devices, systems, services, and applications.

Table 3.7.5 summarizes the strengths and weaknesses of FrontLine Manager.

3.7.5 Management Applications

Application platforms are powerless without management applications. They are provided by equipment vendors or by ISVs, and serve various purposes.

3.7.5.1 Device Dependent Applications

Equipment vendors develop and deploy management applications in order to promote sales of their equipment. Today, it's not possible anymore to sell networking gear without element management systems — in other words, without management applications. These applications are offered and sold at reasonable prices. Equipment vendors don't make much revenue with these element management systems

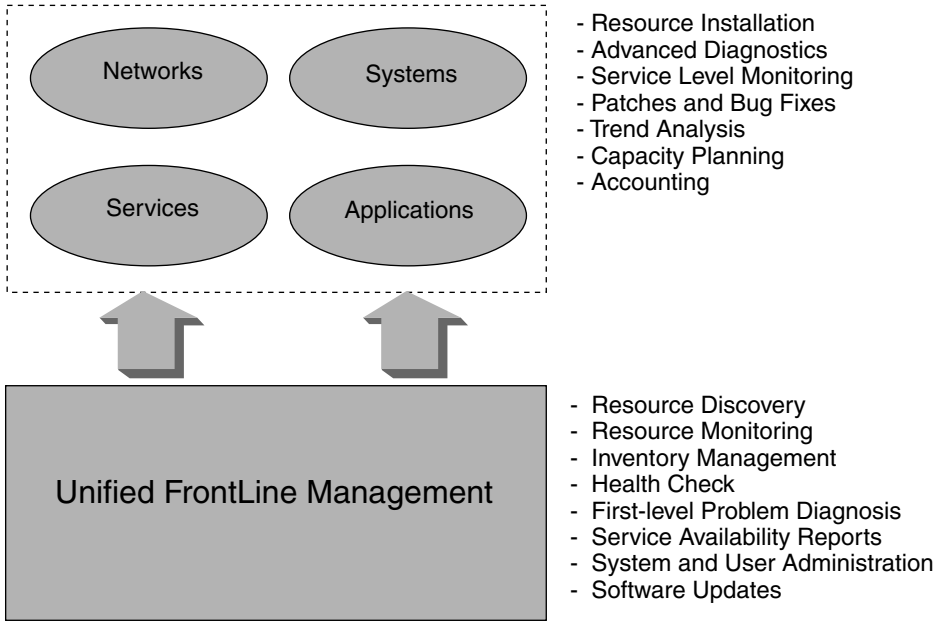

FIGURE 3.7.12 Architecture of FrontLine Manager.

TABLE 3.7.5 Strengths and Weaknesses of FrontLine Manager

Strengths
Unified management
FrontLine management
Intelligent management
Changes are dynamically executable
Distributed architecture
Scalable
Flexible deployment
Extensible and customizable
Use of Internet technologies
Weaknesses
Targeted for small and medium-sized businesses
Not yet widely used
Support of third-party management applications is limited
No compliance to Object Request Broker standards

because they must support multiple frameworks. Web-based management will bring relief by offering an unified interface to management applications. This interface is expected to be supported by all framework vendors.

3.7.5.2 Device-independent Applications

They are designed, developed, and deployed to work in different environments. Usually, they address the following management areas:

- Trouble ticketing
- Performance analysis and reporting

- Security management
- Modeling

Also, these management applications can be integrated into frameworks using Web-based technology. The big benefit is that management applications can be loosely coupled with the framework and with each other.

3.7.6 Summary

Management frameworks are the key for successfully managing communication infrastructures. The frameworks of the future will show very strong core components, and a rich set of management applications. Management applications will be provided by independent software vendors and will address key management process areas of telecommunications services suppliers and of enterprise users. Integration depth is different; the telecommunications are most likely deeper than in the enterprise environment. Some of the management applications are the same for both areas.

References

BALL94 Ball, L. L.: *Network Management with Smart Systems,* McGraw-Hill Series on Computer Communications, New York, 1994.

DORF93 Dorf, C.R.: *Handbook — Electrical Engineering,* CRC Press, Boca Raton, 1993.

GARE95 Gareis, R. and Heywood, P.: Tomorrow's Networks Today, *Data Communications,* September 1995, p. 55-65, McGraw-Hill, New York, 1995.

GHET97 Ghetie, I. G.: Networks and Systems Management — Platforms, Analysis and Evaluations, Kluwer Academic, Norwell, USA, 1997.

NMF95 *Network Management Forum: Discovering OMNIPoint 1 and OMNIPoint 2 — A Common Approach to the Integrated Management of Networked Information Systems,* Prentice-Hall, Englewood Cliffs, USA, 1995.

STAL96 Stalling, W.: *SNMP, SNMP2 and RMON — The practical guide to network management standards,* Addison-Wesley Publishing Company, Reading, MA, 1996.

TERP92 Terplan, K.: *Communication Networks Management,* Second Edition, Prentice-Hall, Englewood Cliffs, USA, 1992.

TOWL95 Towle, T. T.: TMN as Applied to the GSM Network, *IEEE Communications Magazine,* March 1995, p. 68-73.

YAMA95 Yamagishi, K. and co.: An Implementation of a TMN-Based SDH Management System in Japan, *IEEE Communications Magazine,* March 1995, p. 80-88.

3.8 Customer Network Management

Kornel Terplan

3.8.1 Definitions

Customer network management lets corporate users of communication services view and alter their segments of a provider's network. Once such a standardized and open interface is in use, both the service provider and corporate users benefit.

Service providers offer these advantages:

- Keep network loads to a minimum, despite the inexact nature of traffic prediction
- Provide customers with safe access to pertinent OSS and network data, from port assignments to billing and account details
- Isolate individual customer domains without revealing details of the carrier network configuration

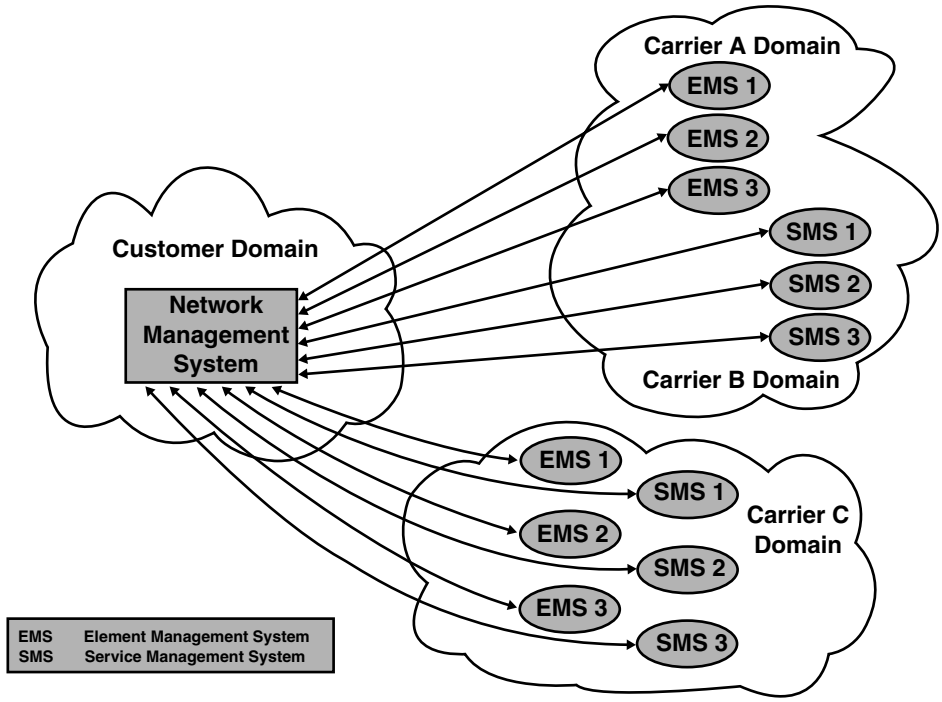

FIGURE 3.8.1 Interfacing multiple services offered by multiple providers.

- Accomplish even the most complex mapping by gathering values from across the network or among OSSs
- Establish customer network domains with full assurance that customers can make only authorized changes

The advantages of corporate users are:

- Alter data network configurations without the delays of paperwork or telephone calls
- Produce any level of report, from performance on a single switch to comprehensive management overviews of account histories
- Manage faults dynamically, reducing the need for carrier intervention
- Streamline troubleshooting with easily generated reports and automatic fixes, even in multicarrier environments
- Integrate to the carrier network whether or not current end-user management systems are robust

The way to a standardized and open interface is long. Today, there are many interfaces and manual information exchange is typical. The customer has to support multiple interfaces that are usually different for each of the providers. Figure 3.8.1 shows this typical case.

This solution can be characterized as follows:

- Support of many proprietary element management systems; most of them are legacy-type systems. They address PBX management, multiplexer management, modem management, management of packet switching nodes, frame relay management, ATM management, wireless management, etc.
- Support of many proprietary service management systems; they are evolving without any core management functions. They address service provisioning, bandwidth management, service assurance, etc.
- Lack of well-understood management protocols
- No easy way of exchanging management information because database and MIB structures are very different.

3.8.2 Concerns of Customers

Customer network management (CNM) has been a long time coming for several reasons. First, it is difficult to measure how much CNM benefits the bottom line. CNM is generally on the cost-saving side which is very hard to sell to management. Selling is easier when the CNM user is actually a value-added provider, which makes CNM a critical component of the value-added service.

Second, when it comes to CNM, many network managers simply do not know what services to require from the supplier. Third, some network managers have serious security concerns about letting an outsider get a detailed look at mission-critical data. Others are afraid that CNM is an attempt by the carrier to lock the customer into a long-term relationship. And some others worry that CNM is the first step toward outsourcing.

Fourth, CNM is very complex to implement. Enabling customers to perform both read and write operations on the internal operations support systems of the telecommunications providers places a considerable stress on those OSSs. Most OSSs are not designed for extra transaction handling and security imposed by CNM. Further, integrating a CNM interface with the network management system of the customer is a difficult task.

Prior to the decision making about implementing CNM functions and features, corporate users should complete a diligency phase consisting of the following tasks (HOLL95):

- Which services come with CNM?
- Are different services integrated in some way?
- What software, hardware, and management platforms do CNM applications run on (Unix, Windows, Sun, HP, IBM, etc.)?
- Can they be easily ported to the company's current network management platform?
- What facilities are furnished to help integrate CNM functions into the existing corporate management infrastucture (CPE, management applications, accounting systems, databases, documentation systems, workflow solutions, etc.)?
- What is the end-user interface to the CNM applications (Windows, Openlook, Motif, etc.)?
- Have provisions been made for training users and technical staff on the CNM system?
- How is corporate data protected against unauthorized access and use?
- What is the cost of the CNM system on a component-by-component basis, including access charges, transport of CNM data to customers, initial installation, integration, and ongoing support?
- How is the CNM system supported?
- What services are offered to help integrate CNM with other management systems?
- What are the procedures in case of significant changes of the OSS?
- What are the impacts on the CNM interfaces and gateways?

After completing this diligency phase, the corporate network manager is well informed about what management functions, databases, and applications can be integrated into the corporate network management systems.

3.8.3 Basic Structures and Core Components

Corporate networks must be able to perform various management tasks. In particular, the following tasks should be supported (HOLL95):

- Fault management, including fault detection, analysis and reporting, tracking and resolution
- Perfomance and quality of service management
- Configuration management, including inventory management, service control, service ordering, and tracking

- Security management, including the protection of the network and its management from both outside and within
- Accounting management, including invoicing, maintaining user and usage profiles, scenario analysis, trend reporting, and exceptional reporting

Most of these tasks must be duplicated for equipment and services of the providers. This redundancy can lead to serious inconsistencies between the provider, corporation, and reality due to the lack of synchronization between inventory files and databases. Moreover, without near real-time information about the provider's network, it is difficult to establish and maintain coherent, end-to-end views of the network, its services, and its performance.

Corporations that buy services from multiple providers find that their problems multiply as the number of interfaces to the service provider rises: operational, fault reporting, inventory, service modification, accounting, and so on. However, even when these interfaces were unified into a one-stop-shopping concept at the provider end, integration with the corporation's internal management systems remains a problem. There are a number of issues to be resolved:

Accounting management: If a customer wants to receive billing information from the provider in near real time (end of shift or end of day) to update an accounting system, some form of electronic interface between customer and provider is needed. Alternatives like e-mail or sending a tape via courier are not the best solutions.

Bandwidth management: Without integrated CNM and enterprise network management, customers who want to change the bandwidth of a service or add more channels to voice and data have to contact the provider through its interface. After confimation, which may take long, customers can start to reconfigure their routers and other network devices. Ideally, using a CNM system, it would require a single application that would accept the request for additional bandwidth. A component of this application would wait for notification that the change has been made and then initiate reconfiguration of the customer network.

Quality of service (QoS): Many customers use their own network management systems to verify that the provider is meeting contracted QoS committments. Doing this properly involves a significant amount of resources. The provider on the other end is probably collecting the same data for the same purpose. It would be best if both parties were working from the same view of the service.

Fault management: Customers will likely perform initial detection and diagnosis of fault using their own network management and monitoring systems. Without CNM, they then must relay this data via phone or fax to the provider and track the progress of fault rectification using the same medium. Assuming the high level of sophistication on both ends, this is not the most efficient way to solve problems.

Table 3.8.1 (HOLL95) summarizes the core CNM components and high priority tasks.

In order to avoid redundancy and inconsistencies, state-of-the-art CNM solutions request a very tight connection between the management architectures and products of the provider and the corporation.

There are a number of ways in which a CNM system could integrate or fail to integrate with customer systems. The first alternative is no integration at all. The provider's CNM system could continue as an independent stand-alone system that provides a convenient point of access to services such as PBX management. Beyond that, the provider could supply customers with a standard interface that encapsulates a particular combination of protocols, information models, and behaviors such as a CNM agent and MIB. This will be the integration point for the management applications at the customer promises. But this still will cause problems if different providers define different interfaces with various information/object models for similar services. At the next level, integration could be achieved via a common graphical interface at the user interface level. The provider would supply a Windows or Motif CNM application that runs alongside the customer's management application on its net management platform. In some cases, the provider would furnish applications as part of the CNM system that uses a private provider CNM agent MIB on the provider side. This is really an extension of the previous approach: the provider offers more of the application functionality to the customer.

TABLE 3.8.1 Core Components of Customer Network Management (CNM)

Fault Management	Configuration Management	Accounting Management	Performance Management	Security Management
Reporting, tracking, and resolution of faults	View inventory of telco-provided customer premises equipment and services	Expenditure tracking on services in near real time	Monitoring of service quality (throughput, delay, and availability)	Access authentication and authorization
Interface to customer trouble ticket or workflow systems	Order new services	Interface to customer accounting system	Ability to generate reports and verify against service contract	Separation of customer data
Fault domain identification	Reconfigure services and network	Extract of histories and usage profiles by customer cost center; cost comparison of rival telco services (ISDN, leased line, xDSL, etc.)	Performance comparison of rival telco services	Separation of telco and customer data

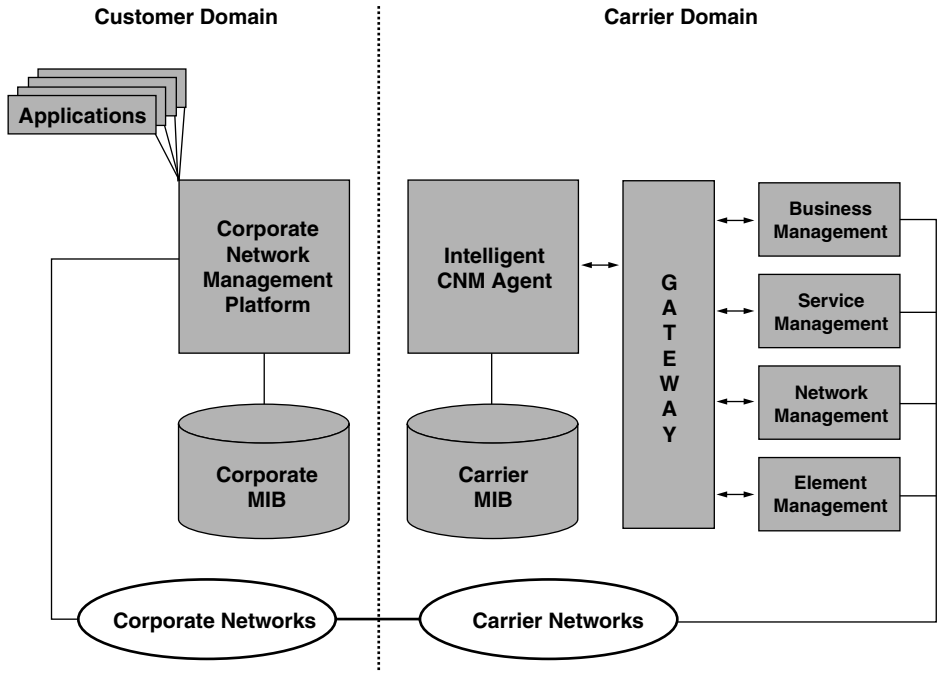

FIGURE 3.8.2 Customer network management.

In order to enhance efficiency and simplification at the same time, the network management platform of the corporation should be connected to a very intelligent "agent" on behalf of the provider. This agent unifies and coordinates the work of multiple managers who are responsible for business applications, along with service, network, and element management. It is also responsible to synchronize data files, databases, and MIBs.

Figure 3.8.2 shows an integrated structure. There are two connections between the systems: one at the physical level and one at the network management level.

This high level of integration is expected to be reached in multiple phases. Telecommunications providers are on the move to select, customize, and deploy powerful management frameworks that will play the role of the intelligent agent.

3.8.4 AccessCNM from Objective Systems Integrators

The functions possible with CNM technology must address the individual customer's need for network management data. They also must meet the provider's need to maintain absolute and integrated control and visibility of their entire network, all OSSs that support it, and each domain contained in it.

Each of these perspectives must be considered if the overall service offering developed is to provide acceptable and competitive features to customers while it answers the nonnegotiable operations and business concerns of service providers.

This seemingly self-conflicting goal defines the challenge that Objective Systems Integrators (OSI) meets with AccessCNM. The company brings to the CNM market an expert awareness from both perspectives and the resources to develop, support, and enhance a comprehensive solution.

3.8.4.1 Basic Functions of AccessCNM

CNM services allow a customer access to segments of a public network that is shared by many users and services. The individual customer's need to know must be balanced against the privacy concerns of the entire customer base and the need for network security and integrity. Therefore, individual access must be controlled. AccessCNM accomplishes this by segmenting the view of network elements and securing them against unauthorized use through three basic functions:

- User authentication, which enforces established access authority
- Flow control to avoid overloading network elements
- View mapping to translate how objects are represented between customer and carrier network views

3.8.4.2 AccessCNM Architecture

The customer access network primarily transports data from the customer site to AccessCNM. It also performs rudimentary flow control and serves in the authentication process, which is a feature of the underlying network service. Access is part of any customer configuration with or without CNM. However, OSI can consult in the design of the access portion to optimize the authentication process and traffic regulation for use with AccessCNM.

The Access Regulation Module (ARM) is the gateway to AccessCNM and acts as firewall. Essentially an IP relay engine, the ARM regulates the load from the server to a module that translates and filters SNMP messages. ARM functions include applying customer-specific flow control; discarding non-SNMP, malformed or errored packets, and those from nonregistered users; and hiding the rest of the carrier network. Traffic from the reverse direction is also regulated.

The deployment of AccessCNM is shown in Figure 3.8.3.

The AccessCNM Core is the system's request processing engine. This is the only module that interprets SNMP messages after the initial phase of the ARM, thus freeing the rest of the system to handle generic database and flow control functions.

The AccessCNM Core performs mapping in three stages; first, the customer view presented in the packets is transformed into the view appropriate to the carrier network. Second, the request is transmitted to one of many potential internal handlers. Third, the handler processes the customer request and, if required, formulates a response.

Handler functions may be customized and their actions extended depending on the desired extent of the resulting CNM features. Some handlers may use SNMP to contact various network elements. Others may obtain database information either locally or from various OSSs. AccessCNM Core is shown in Figure 3.8.4.

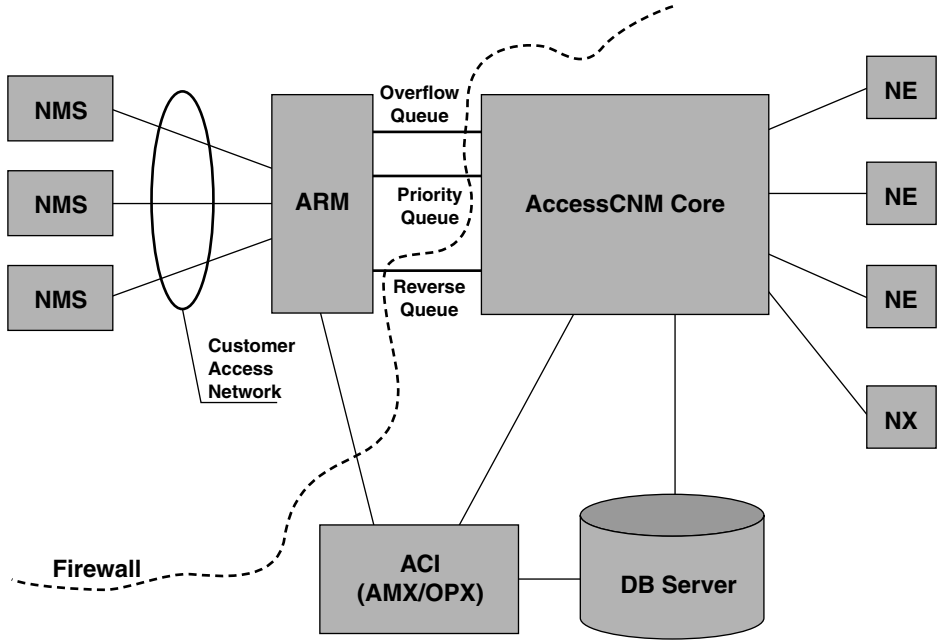

FIGURE 3.8.3 AccessCNM architecture.

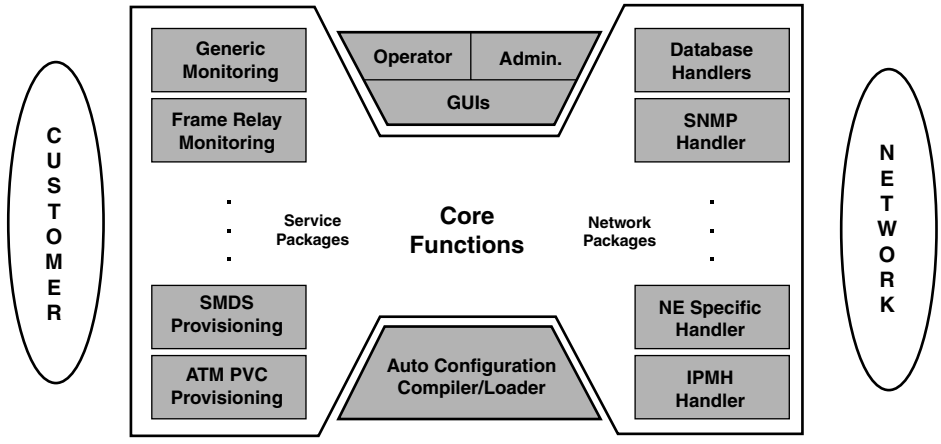

FIGURE 3.8.4 Core of AccessCNM.

Still other handlers and NetExpert interfaces may trigger more complex processes such as provisioning additional services or allocating higher bandwidth, updating customer account files, and confirming that actions have been completed. As the technology evolves, many of these functions will be available off the shelf as feature-specific CNM packages. The AccessCNM Core also is an engine containing the logic and data required to manage customer information. It provides the basis for the MIB traversals, identification of view translator and handler functions, as well as the tables of parameters used in handler processes.

Because AccessCNM associates the actual network view to the segmented and limited customer views, knowing how its processes are executed is important to network security. Information from AccessCNM Core will not flow to the customer unless a corresponding mapping entry is found in the ARM. The AccessCNM Core also includes caching features that keep database lookups to a minimum by storing certain previously accessed data in memory.

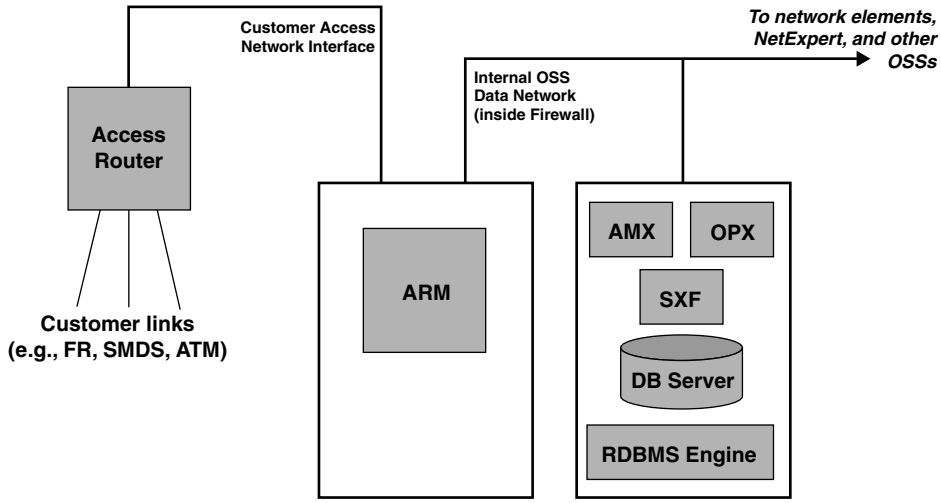

FIGURE 3.8.5 Structure of AccessCNM workstations.

3.8.4.3 Provider User Interfaces

AccessCNM includes operator interfaces for routine modifications, such as provisioning, and maintenance interfaces for privileged commands used for maintenance, monitoring, initialization, and reporting. Interface design may be customized and later revised to match organizational structures and operator skill levels. Initally, these interfaces will allow these basic functions:

- Assigning interfaces to customers
- Modifying flow control parameters
- Generating configuration reports

An important AccessCNM feature is the ability to perform additions and deletions of user data and handler functions dynamically, without system restart. AccessCNM offers an open architecture with editors and compilers (Figure 3.8.5).

3.8.5 Summary

The rising dependence of commercial businesses on internetworking and data communications presents an excellent chance for service providers to gain the loyalty and partnership of the largest spenders in their markets. Even in markets where competition has not yet occurred, providers are finding that customer relations and the efficient implementation of requested services directly impact profit margins and the provider's general image. If competition is expected soon, the edge may be retained by providing CNM solutions today and expanding the features offered as the market opens.

Offering comprehensive CNM services complements the most basic reason for being in business: getting and keeping customers so that revenue generation is assured. End users, demanding the power to view and alter their segments of the carrier network, gladly pay for the added value and, increasingly, are comparing CNM functions when deciding on service providers. The right CNM technology gives providers confidence that their networks and OSSs are secure and that customers have access only to their own domains. Risks are minimal and profit goes up with AccessCNM because its functions can be enhanced on demand.

Empowering the customer by implementing a comprehensive CNM solution brings a hidden bonus to providers: end-user self-government decreases operator intervention and lessens trouble calls, thereby reducing the cost of doing business. The characteristics of AccessCNM — standards-based, open, and object-oriented — are the very features required of OSSs as entities. They also are the traits of systems

most likely to return investment quickly because such systems are ultimately less costly. AccessCNM is flexible enough to integrate with existing OSSs and is ready for anticipated expansion.

References

HOLL95 Holliman, G., Cook, N.: Get ready for real Customer Network Management, *Data Communications,* McGraw-Hill, September 1995, p. 67-72.

TERP98 Terplan, K.: *Telecom Operations Management Solutions with NetExpert,* CRC Press, Boca Raton, 1998.

3.9 Aspects of Managing Outsourcing Solutions: Aiming for Success

Carel Marsman

3.9.1 Introduction

The aspects of management of outsourcing solutions and the aim for success — this chapter will handle this broad topic from the customer focus perspective. What are the related processes, what matters from this perspective, and which factors can make a difference? People, as in most service delivery organizations, are key. They can be the competitive edge, if organized and facilitated in an adequate way.

This chapter will discuss how an outsourced solution could be managed in order to deliver world-class services and support, with an emphasis on *could be managed* because it's merely the view of the author based on his experience in the field. Various related topics will be addressed such as the partnership approach of the solutions, service level performance indicators and reporting, integration with vendors, management of and integration with customer operations, decisions to expand or exit, and other related concepts.

These topics will be discussed through the following themes:

- Customer problems and needs in the marketplace
- Strategic outsourcing alliance
- Managing the strategic supplier relationship
- Business processes: What's to be managed?

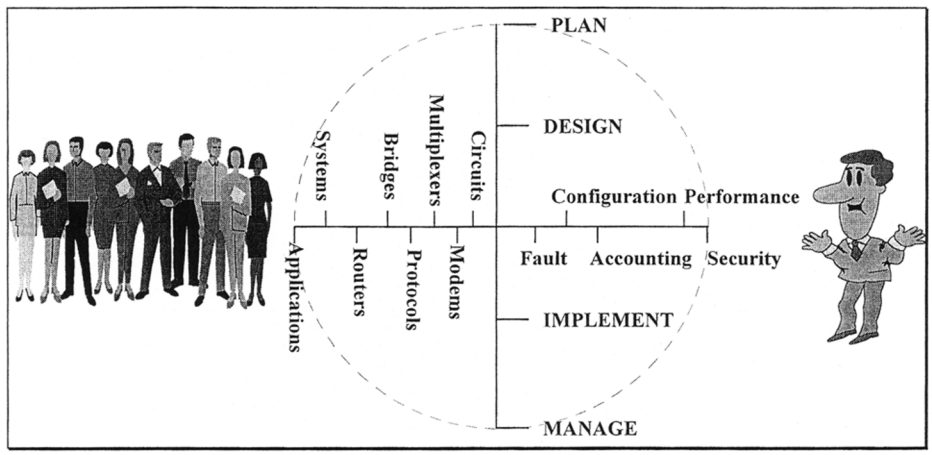

FIGURE 3.9.1

- The partnership approach of service management
- Organizing for success

The aim is not to cover all topics and concepts in the area of management of outsourcing solutions, but to give a comprehensive view of the most important aspects in this exciting and demanding "game." At the end of the day, a customer is looking for a partner to rely upon and to be trusted in handling the company's telecommunications infrastructure. Such a partner offers the right people, processes, and tools for the job, along with the business knowledge, experience, and relationships in the industry — in other words, the core business of the IT outsourcing vendor.

This chapter is written from the customer focus perspective and, for that reason, it emphasizes more the processes, partnership, and people side of the outsourcing business, and not so much the tools and technique side. However, in the references and bibliography section, you will find authors who have contributed in more detail on the technical issues of outsourcing.

3.9.2 Outsourcing — the Evolution

Outsourcing is the product of a long-existing tradition of the make-vs.-buy decision concept. In order to understand the evolution of this product, an overview of the characteristics along the years is given below, including the categories as we still know them:

Decade	Computer Services 1970–1980	Facilities Management 1980–1990	Outsourcing 1990–
Technique	Batch programs on large mainframes	More complex, downsizing of hardware	Highly complex networks, downsizing, telecommunications
Equipment	High purchase costs	Costs decline	Costs decline more
Applications	Costs relatively low to equipment	Costs incline	High costs
Know-how	Low costs	Costs incline	High costs
Customer	Small	Small	Small and large
Presence	National	National/international	International/national
Product	Large amounts of "bread and butter" programs	Standard offers for managing data centers	Managing (parts) of the IT environment based on specific customer requirements
Economic motivation	Shared usage of expensive main frame	Flat costs, higher service level	Saving costs, increased flexibility, better results

In the following sections and examples you will get a better picture of outsourcing. A typical recent end-to-end managed outsourced solution is presented in Figure 3.9.2 in order to create an image of what could be part of the scope of the service delivery.

3.9.2.1 Setting the Stage — Customer Problems in the Marketplace

Before we start setting the stage, we should have a common definition of what outsourcing really is. A definition — not the only one — is the following: *Outsourcing is the transfer of part or all of an organization's existing data processing hardware, software, communications network, and systems personnel to third party* (Due, 1992).

Much of the recent activity in the outsourcing market has concerned vendors developing more desktop services outsourcing capability. These services have been developed to address the challenges that organizations are facing in managing new technologies in the form of distributed LANs. However, changing technology and business requirements are also beginning to have an even more major impact on the ability of organizations to manage their WANs.

In this context, in order to understand what matters from the perspective of the customer, it is important to understand the scope of the customer's problems. A customer's need for integrating and managing multinational infrastructures is hampered by:

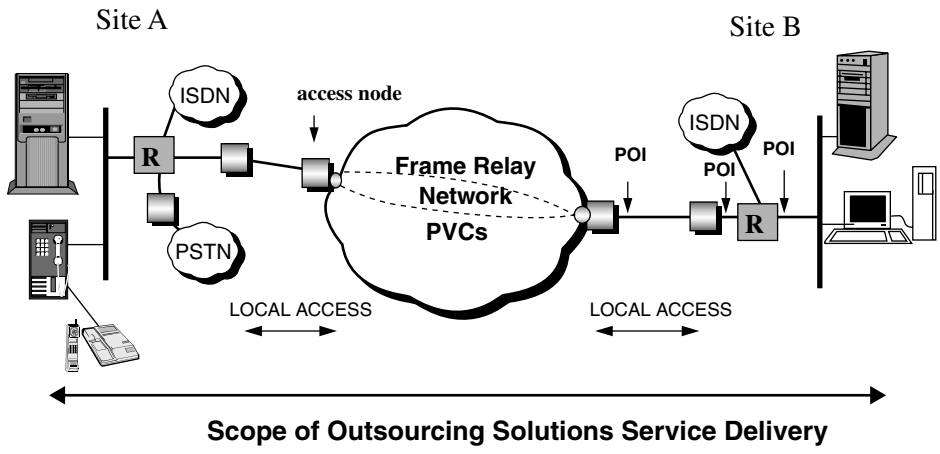

FIGURE 3.9.2 Typical end-to-end managed outsource solution.

- Incompatible and inflexible networks and local infrastructures (LAN)
- Multiplicity of suppliers and products, e.g., multiple telecommunications carriers and vendors
- Difficulty of managing systems and equipment conforming to different standards
- Shortages of affordable skilled resources to manage existing networks and infrastructures
- Lack of global networking infrastructure
- No adequate capabilities to combine voice, data, and image
- Regional barriers — time zones, cultures, languages, lead times, holiday schedules, work shifts — resulting in poor service processes

3.9.2.2 Drivers to Outsourcing — Customer Needs

Altogether, the potential characteristics of outsourcing vendors of perceived importance to customers can be listed in short, i.e., the most common drivers, as follows:

1. Cost savings
2. Focus on core business
3. Avoid headcount increase
4. Availability of new services (voice/data)
5. Staffing/skill problems
6. Headcount reduction
7. Better local support required
8. Better cost control and allocation required
9. Desire for uniform international services
10. One-stop shopping/maintenance
11. Inflexibility of private network
12. Bill consolidation
13. New application requirements
14. Better reliability/resilience required
15. Poor utilization of private network

16. Other companies are doing it
17. Poor network management
18. Poor performance of private network

3.9.2.3 The Importance of the Strategic Outsourcing Alliance

Given these challenges — fast-changing business requirements, technologies, and scarce resources — it is important to understand the concept of the strategic alliance.

Strategic alliances are combinations between firms, designed to support or shape the competitive strategy of one or more of the allies, all for the benefit of the customers served via the alliance. As stated before, companies are not just looking for vendors but also for partnerships. An example:

> Lasher, Ives, and Järvenpää (1991) describe the example of strategic alliances. The objective of the strategic alliance between the United Services Automobile Association (USAA) and IBM was to build a large-scale image processing system for USAA in order to introduce the paperless office. IBM invested heavily in the solution. The requirements of USAA were comparable to those of IBM — a general image solution for all of the application areas of USAA. The advantage for IBM obviously was to integrate this solution in their standard offering. USAA is operating in a niche of the market aiming for financial services to military personnel. The company, in comparison to others, is not worried by IBM's interest to market the product in the insurance industry.
>
> In the end, the product supported both USAA and IBM in order to achieve important organizational targets and reach a sustainable competitive edge within their own market areas.

3.9.3 Managing the Strategic Relationship — Supplier Management

Given the concept of such strategic relationships: what is important in managing this relationship from the customer's perspective?

In other words — yes, the customer is looking for a partnership but in the same respect wants, exactly for that reason, to get the best "deal" around — one that creates a win–win situation for both parties, as in the case of USAA and IBM. You could say that in outsourcing deals, the offering company has to be some sort of "super-supplier." It has to bring the right cards to the table on strategic, tactical, and operational levels in order to offer the ultimate synergy.

Bearing that thought in mind, let's take a look at what a super-supplier should bring to the table, looking through the customer's eyes. Knowing what they are looking for and putting yourself — as a potential partner — in the position of the customer will give the potential outsourcing partner a head start on the competition.

In an average outsourcing deal, the customer will look for partners performing strong on:

- New technology development
- Taking total cost of ownership
- Globalization and localization issues

For the customer, managing their supplier successfully means looking at what the potential partner's views are on quality management aspects and the integrated business management system. Organizations like Xerox, Honda, and Ricoh[4] are known to be leading practitioners in implementing quality management principles for managing their suppliers.

The right partner should be organized and have:

- Shared values that govern decisions, behavior, and relationships
- Leadership and management systems that provide clear direction and develop, empower, and recognize people
- Organizational structures that provide efficient and effective roles and responsibilities

[4]Masaaki, Imai, KAIZEN The Key to Japan's Competitive Success, 1986.

- Processes that create value for their customers
- Information Systems that provide the facts upon which decisions and actions are based
- Their procurement, engineering, product, and quality assurance teams and customers working together early on as one team, in planning and implementing these elements to achieve mutual goals

The underlying principles such a super-supplier should be committed to are:

- Supply and technology strategies driven by business objectives
- The knowledge that allies can achieve more than adversaries
- A willingness to work together to achieve mutual business advantage from the relationship — being open and honest with each other
- Clear objectives, goals, and requirements are always established
- Performance is measured and reviewed on a regular basis
- Strive for continuous improvement to deliver maximum value to the customer and their end customer/end user of the product or service

This means that customers will be looking for the tangible and intangible — after all, it's a people business — proof of all of these high-quality organizational building blocks.

Knowing that the customer is looking for all the above qualities makes it clear that the strategic alliance is very much about building relationships. The more complex the relationship, the more its effectiveness depends on those same factors that are important to personal relationships: trust, mutual respect and dependence, and a shared vision. The joint intent to achieve the spirit, in addition to the letter of a contract, characterizes the most effective relationships.

So, an important step in establishing this relationship is to first understand the characteristics of the customer organization. What are the behavior drivers, the unwritten performance standards, the unspoken generally accepted truths? The customer will be looking for relationships with partners — super-suppliers — that have similar or complementary characteristics. These characteristics will consist of values, beliefs, and norms. Clear, consistent, and *well-communicated* values, beliefs, and norms support both sides of the strategic relationship.

When the partnership is established, the real effort is to implement all of these mindsets on strategic, tactical, and operational levels. Really sharing these values and business senses, living them in the day-to-day business, is what the partnership should be about.

Last but not least, as in any healthy relationship, it should be possible to expand or exit the relationship. The customer could prefer to test the relationship on an operational level through, for instance, a pilot program. Most regular supplier relationships are on a tactical level. Trusting each other to start a long-term strategic relationship should in all cases also be based on clear performance indicators and service levels combined with well-defined contractual exit points. If the relationship starts on an operational level, expansion to a strategic relationship is a logical consequence of the customer's requirements and the partner's capabilities.

3.9.4 Business Processes and Outsourcing — What's to be Managed?

3.9.4.1 Introduction

We now have an impression of the challenges, requirements, and strategic needs of the outsourcing customer.

For a customer, outsourcing (part of) their IT infrastructure means being able to focus on their core business processes and the remaining supporting processes such as finance and human resource management.

This section will provide an overview of the entities within, and the scope of, outsourced solutions. A model in which the to-be-managed processes (core processes for the outsourcing partner!), functions, and objects are displayed serves to build the thesis of this article — how an outsourced solution should be managed in order to deliver world-class service and support.

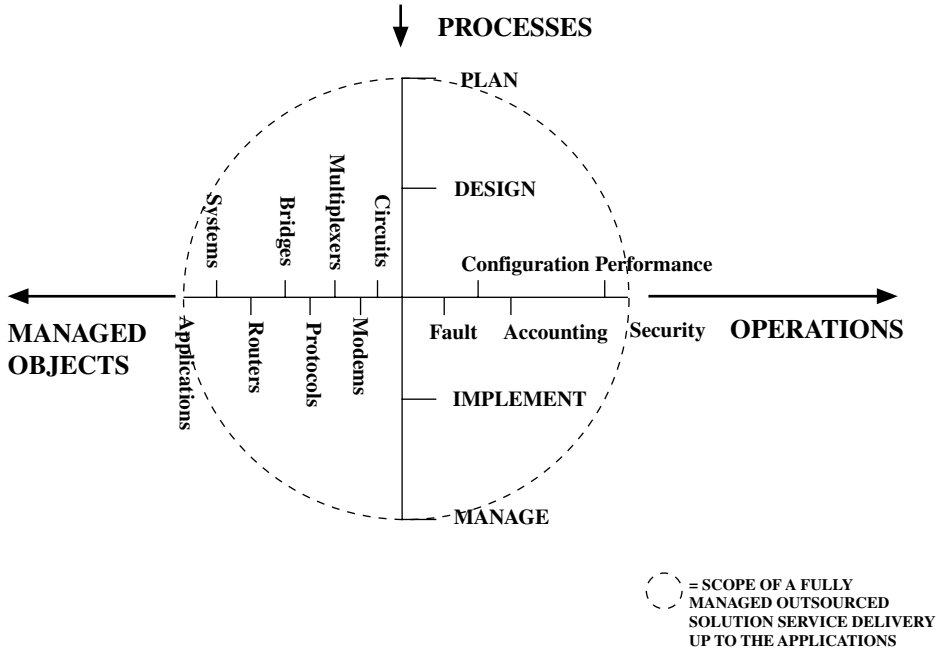

FIGURE 3.9.3 The relationship needed for high-quality services and solution management.

The planning, design, implementation, and management processes will be discussed, as well as the day-to-day management functions. An additional model will offer comprehensive insight into the management functions and related information streams. Furthermore, the various dimensions and definitions of what can be outsourced will be addressed in relation with the to-be-managed objects.

3.9.4.2 The Life Cycle of Processes, Functions, and Managed Objects

In order to be able to present a clear picture to the customer and create a mutual understanding of what the important processes, functions, and elements are, the next model (Figure 3.9.3) depicts the relationship between the different elements needed for high-quality services and solution management.

This model will serve as the basis for explaining how outsourced solutions should be managed in order to create world-class service.

For that reason, the essence of the model will be discussed. It should be noted that any other model could work as well. It's not about this specific model, but more the thought that a picture paints a thousand words. That's exactly the strength of using clear models and pictures in this complex business, and one of the primary added values of a high-quality service provider. In other words: making the provided services and the level of services crisp and clear is really for the benefit of both the customer and the outsourcing partner. After all, as in most relationships, trust is very much about expectations being in sync. This model can be used to aim the combination of processes, daily operations, and managed objects toward the customer requirements.

Horizontal Axis Left: Outsourced Solutions Elements

Every outsourced solution (or part thereof) consists of a number of elements that, working together, establish the service provision to the end user. This combination of hardware and software needs to be managed.

Vertical Axis: Process Elements

Once the decision has been made to implement the solution or to implement changes to an existing solution, a number of processes have to be followed:

- *Analyses.* It is important to do a thorough analysis of the specific requirements that the business processes and the end-user wish to demand from the network. These requirements need to be documented during the analysis phase. This has sometimes already been partly done by the customer and will be complemented by the outsourcing partner in collaboration with the customer.
- *Solution planning/design.* Based on the technical and functional requirements obtained, the solution or the changes on the network can be planned and designed. This is accomplished by outsourcing partner specialists who will regularly update the customer. The functional requirements should be documented in a service level agreement (SLA). The SLA will function as a benchmark and a means for controlling and checking the quality and the performance level of the solution and the solution management service delivery.
- *Implementation.* After the solution design has been approved and accepted by the customer, the implementation will take place. For this, the outsourcing partner should use project management and quality control processes that ensure a correct and timely implementation. After the solution has been implemented, a technical and functional acceptance test should be performed. These tests will be coproduced by the customer and the outsourcing partner.
- *Management.* Once the solution has been approved and accepted by the customer, the management phase will commence. The management service elements are mentioned on the horizontal axis and are described below.

Horizontal Axis Right: Service Elements for Outsourcing Management

The outsourced solution elements need to be maintained and managed on a daily basis, the so-called logical and physical solution management. Logical solution management entails the management of the network protocols and the software for the solution components such as bridges and routers. The logical solution management often can be done remotely. Physical solution management is the management of, for instance, cabling and hardware components. The physical outsourcing management must often be done locally — for example, some of the desktop services.

Outsourced solution management services should be able to perform both the physical and logical solution management for the customer.

To enable this, the following outsourced solution management service elements should be delivered on a daily basis:

- *Fault Management:* This comprises the proactive monitoring of the solution in order to detect possible causes for future problems and prevent them from happening. Next to that, it also should offer help desk services to the customer's user representatives in order to be able to receive user calls for help and have specialists solve the problems. Concerning logical solution management, the outsourcing partner staff will diagnose and solve the problems. In case the problems are caused by the physical solution, the outsourcing partner staff should be able to dispatch a specialist to the faulty component in order to fix the problem.
- *Configuration Management:* The outsourcing partner should be able to perform both physical as well as logical solution configuration management. This means that logical adjustments should be made in case of changes to the information stream requirements, and physical adjustments to the solution should be made when user units change location. Also, the administration of the solution configuration will be performed and reports should be regularly submitted to the customer. The outsourcing partner should control the administration of, for instance, the customer network addresses on the WAN within the points of interconnection (POI). The customer could make the choice to be in control of the addressing on the LANs and the network number registration, depending on the scope of the agreement of the outsourced solution.
- *Performance Management:* The outsourcing partner should constantly monitor and evaluate the solution on the consistent compliance with the performance requirements — indicators — as documented in the functional design. Proactive measurements should be undertaken if the performance level tends

to degrade. The customer representative should regularly receive reports concerning the solution performance.

- *Accounting Management:* Depending on the customer's specific accounting management requirements and the available technology, the outsourcing partner should categorize, quantify, and report the usage of the network.
- *Security Management:* Depending on the customer's specific security management requirements and the available technology, the outsourcing partner should perform the security management of the solution. This could include, for example, access management, password security, and encryption.

3.9.5 The Partnership Approach of Service Management

3.9.5.1 Introduction

In this section, you will learn what service management is, what it offers to the customer, how this should be performed, and the required interaction with the operations in order to deliver a seamless service that surpasses the agreed-upon SLAs.

The "how" will cover the aspects of partnering or relationships with other suppliers/vendors being a truly single point-of-contact for the customer; management around the clock, especially the challenge of the time zones, integration with customer operations; managing customer perception; and building the service level agreement and reporting based on the SLA (service performance indicators, etc.).

Between these points, some customer examples (no names will be mentioned) will be highlighted, along with market research findings.

3.9.5.2 Service Management

The outsourcing partner is a service provider company. What does that mean?

To begin with, services are a mix of tangible and intangible aspects. Tangible aspects are the service levels, the reports, etc. Intangible aspects are the perceived qualities of the service received by the customer, and the way the solutions provider handles the customer. As you will see, there is a strong connection with the essentials of building a strategic partnership.

When we look at a quality service organization, it should distinguish itself through a market-oriented approach. The enterprise of the outsourcing partner should have the following fundamental features related to the marketing concept:

- **Attitude of mind**: the customer is the basic reason for the existence of the enterprise
- **Organizing the enterprise**: all organizational design should stem from the customer and ensure that the customer is "created," won over, and kept by the enterprise
- **Range of activities**: activities necessary to ensure the serving of customer needs emerge as a matter of course (creation, production, delivery of services)
- **Techniques and tools**: enabling the organization to operate as efficiently and effectively as possible in the customer's interest (motivation research, linear programming, discounted cash flows, etc.).

Knowing that services are partly tangible and partly intangible, substantial attention should be paid to the marketing aspects.

Furthermore, the four standard Ps (product, price, place, and promotion) of the marketing mix are expanded with three more Ps (people, physical evidence, and process) in the marketing mix for Services.[5] The three additional Ps should be incorporated in the marketing mix as input for the design of a service organization. Because of their importance to the design, the three extra Ps will be discussed individually. By the way, this is my opportunity to introduce maybe the most important — most forgotten — P of a service organization: pleasure (or fun)! The amount of energy this "P" generates should never be underrated by the management of any service providing company! A happy employee radiates pleasure toward the customer.

[5]Cowel, D. The Marketing of Services, 1984.

3.9.5.3 People and Service Organizations

Although the building, equipment, and financial assets are also resources required by organizations, employees — the human resources — are particularly important. People provide the creative spark in any organization. They design the service, control quality, market the service, allocate financial resources, etc. With respect to personnel, close attention should be paid to training, discretion, commitment, incentives, appearance, and interpersonal behavior. With respect to other customers, close attention should be paid to their attitudes, behavior, degree of involvement, and customer/customer contact.

People are the most important assets of a service providing company!

3.9.5.4 Physical Evidence and Service Organizations

In the physical design, it is important to pay attention to the internal environment in terms of:

- Environment: furnishing, color, layout, noise level
- Facilitating goods: presentation equipment, audio-visuals
- Tangible clues: packaging, manuals

3.9.5.5 Process and Tools in Service Organizations

The behavior of people in service organizations is critical. So too is the process — the what and how — of the service delivery. Cheerful, attentive, and concerned staff can help alleviate the customer's problems of having to queue for service or soften the blow of the breakdown of technology. They cannot, however, compensate entirely for such problems. How the overall system operates — the policies and procedures adopted, the degree of mechanization (or high tech) used in the service provision, the amount of discretion employees have, the flow of information and service, etc. — these are operational management concerns that need attention when designing service organizations.

More details of how to successfully organize the service organization will be presented in the Section titled 3.9.6.

3.9.5.6 What Does Service Management Mean in Practice?

Service management in practice means making sure that the agreed service levels are reached and, if possible, surpassed. Consequently, this also means service management is about the art of managing the customer's expectations. No matter how great the quality of service, there always will be dips in the service level. Implementation projects will always run into unforeseen, unexpected, or new hurdles.

The real art lies in really knowing your customer's requirements, the service delivery and operations processes, your supplier requirements, and — last but not least — your customer processes (Figure 3.9.4). A solutions provider should be very aware of this principle. The strongest example is perhaps the role of the (inter) national carriers in global solutions.[6] If a national carrier — a supplier of a major link in the end-to-end service provisioning — closes down the help desk on the weekend, one can tell the impact on the service level if a leased line goes down on Friday at 5:30 p.m.

Knowing this and having the right tools and procedures in place will enable the outsourcing solution partner to proactively inform the customer of any interruptions to the service. Always try to place yourself in the position of the customer—know that they have to face *their* end users.

So knowing the customer requirements — your output, and looking at what that means for the service delivery processes; your input, and its requirements toward the suppliers, e.g., carriers — means having measurements in place, providing and receiving feedback, and communicating. Improving these customer focused business processes[7] is the key to creating a sustainable competitive advantage as a solution provider.

A short and simple example on informing the customer of any interruptions in the service, anytime, anywhere:

---

[6] *Datacommunications International,* May 1997, Rating the World's Best Carriers.

[7] In the References and Bibliography section, a couple of authors are mentioned on this and related topics.

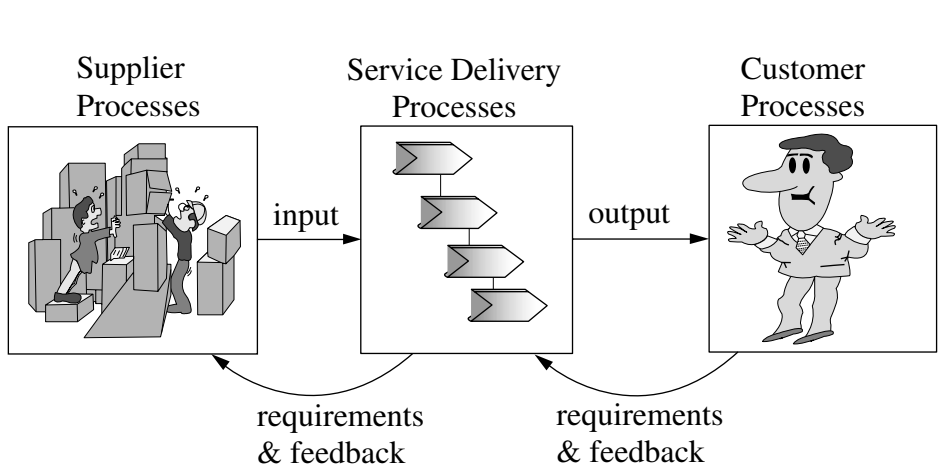

FIGURE 3.9.4 Essential areas of awareness for solutions providers.

The service manager of an outsourcing company was instantly informed by the operations manager of a severe problem with the international backbone of a customer. Because of the proactive Management of the managed objectives, the service manager could contact his counterpart within the customer organization. It appeared his counterpart could not be reached that day for family reasons. The service manager, in close cooperation with the customer organization, was able to quickly track down the MIS manager — the counterpart of the service manager — and get him out of an important meeting with one of his end users, an organization. The service manager could now briefly inform the MIS manager of the problem, the corrective actions taken, and the problem analysis so far. The MIS manager was informed from first-hand knowledge and could get the message to his customer — the end user — instead of being confronted with the problem by his customer.

This example may seem like an open door; however, in talking to customers one will find that even the simple and obvious things are not always common practice in the service they receive from their outsourcing partners.

An extra dimension in international service delivery is, of course, the difference in cultures. Minor problems will probably not even be reported from, for instance, the Nordics, but the same problem could sound huge when reported by one of the countries in the south of Europe. The nature of people and customs should always be taken into account. This means taking the time to get to know what these cultural aspects are, and being open minded.

3.9.5.7 Service Reporting

As stated before, the services, in opposition to products, are part tangible and intangible. Therefore, service reporting takes an important place in managing outsourcing solutions. Clear and correct information on the service levels and other relevant information will help both the customer and the partner to control the managed solution, and plan ahead.

After all, the process of analysis, plan and design, implementation, and managing the solution is a never-ending cycle because of changes in business planning and focus, closing/opening of new subsidiaries, short life cycles of products, emerging technologies, and so on.

For instance, monthly service reporting is input for this never-ending cycle and enables the combined customer and outsourcing team to investigate trends, bottlenecks, and other interesting topics emerging from the reporting.

Service reporting should roughly have the following objectives.

- Provide means for customer control of the solution process elements.
- Ensure that the level of solution services remains in line with the customer's business objectives.
- Provide information to authorized customer personnel with an interest in the solution infrastructure.

The contents of the report should be built along the previously discussed elements of the management model. An example of the added value of customer reporting follows:

A customer (a managed outsourced solution much like the typical 1990s managed solution picture mentioned earlier) was confronted on a regular basis with complaints from his end users about the network's performance. After closer examination and further investigation of the customer and outsourcing service provider jointly, as partners, the discovery was made that the degradation in service was the result of heavy file transfers. These file transfers were not part of the regular business processes of the customer's decentralized international user organization. However, the detailed WAN traffic reports showed these bursts in traffic from the central host to the requesting end user.

The customer was able to quickly trace the root cause of the problem that was influencing the service for all the end users — the customer.

By jointly analyzing and discussing the service reports, the outsourcing service manager and the service manager of the customer were able to fix the problem and advise the specific end user's organization to locally request the information. As it turned out, the end user was unaware of some new local data retrieval capabilities.

3.9.5.8 Service Level Performance Indicators and Reporting

Reports should reflect the performance against the agreed service levels. Performance indicators will be agreed upon in SLAs, bearing in mind the process between customer solutions providers/suppliers. A couple examples of indicators important to the customer from a Fault Management perspective (knowing that they also have their customers — end users — to respond to) are given below. Most operational SLAs will have indicators as well on change, accounting, performance, and security management.

- Number of faults closed in the month
- Total number of faults over entire managed solution[8]
- Faults whose outage exceeded the customer-agreed threshold
- Chronic faults (x or more faults per month caused by the same problem)
- Mean time to repair (MTTR) for faults by, for instance, link and total solution
- All equipment failures whose outage exceeded the customer-specified threshold
- Per vendor number of faults (committed MTTR and actual MTTR)
- etc.

Again, the real art is to extract the needed *data* from the managed objects and convert this data into useful *information* to present to the customer. In other words, add value as a solution provider, and not just simply deliver the agreed service levels. Be proactive, look at and discuss trends, indicate opportunities and threats — act as a Partner.

As indicated in the introduction, distributed LANs — and, linked to that, distributed network management — is a major topic and challenge. Quality reporting and indicators on LAN traffic — resulting from adequate solution management — is, for a solution provider, a competetive edge.[9]

---

[8]Up to the agreed point of interconnection, e.g., the boundary of the management responsibility.

[9]Read, for instance, *Network and Distributed Systems Management* by Morris Sloman and *Lan Traffic Management* by Peter Phaal.

3.9.6 Organizing for Success — A People Business

3.9.6.1 Introduction

In this section, the most important aspects of organizing a team in order to deliver world-class services and support will be presented.

What does it mean to create and maintain a team of top specialists, and how do you manage them? What makes or breaks a team? How can you create a learning team? The relevance? The people managing the solution are the competitive edge of the IT outsourcing vendor. They are the human capital of the company and can make the difference in winning or losing a deal or an existing customer.

3.9.6.2 Structuring the Service Organization

Organizing people means some sort of structure should be put in place enabling these people to work together in order to accomplish the company targets. An outsourcing service providing company will mainly have highly educated and dedicated professionals on the payroll. Professionals are people that know their job and — most of the time — are not too fond of too many rules and procedures. However, processes, procedures, and rules are necessary in order to deliver seamless high-quality (inter) national services. The structure of the service organization should take all of these aspects into consideration. Next, a short overview of some of the essentials.

In essence, the *structure* of the service organization should be built around the service delivery processes — the management processes (FCAPS) — in the model. The related subprocesses, activities, and functions should be derived from the major processes. The organization should be able to adequately support the processes. This should be the starting point.

The *systems* — tools and techniques — play an important role in enabling the organization to effectively and efficiently service several customers. After all, the economies of scale aspect is one of the major reasons for an outsourcing company to be able to be cost effective. This goes for the management systems but also for the internal information systems. These last ones should be aimed at the crucial data (customers, orders, design rules, etc.) and the functionality should be fast adaptable.

The classical *line and staff* — advising functions — should have a balanced mix, thus creating "think overhead" sense in the line functions. In other words, at all times it should be avoided that the line functions are caught in turbulence of just managing the day-to-day business. Time and resources should be allocated to think things over, analyze trends, and implement and monitor quality and security policies — thus enabling the organization to learn. Consequently, the modern service organization should strive for horizontal and vertical job specialization together with handing over more responsibilities and authorizations to the work floor. By cutting down on management levels — less coordination and confusion — and by adding value and importance to the tasks performed by the people working close to the customer, the Organization improves its learning curve and self-steering ability of the teams, thereby creating a more *flexible* Service Enterprise.

The management *style* is a very important part of the success of the service organization. Having *shared values* in place and having the management actively practicing these — walking the talk — is rather essential. These values should be the common bond between the employees of a company. Having a solid bond in place enables people to have a basis they can rely upon and which enables them to go that extra mile and be flexible service-minded persons.

The *skills* of the people are probably the most important, because these skills will enable them to participate in multifunctional teams. Organizing skill pools and focusing on key skills will enable the service organization to run an efficient organization. From the people's perspective, they will be able to learn a great deal and have challenging projects to participate in. In this turbulent and dynamic environment, it is not possible and desirable to centrally manage the working force. The outsourcing company employees skill-set should be a mix of soft skills — communication, presentation, teamwork, and hard-skills — management platform technology, specific customer solution technology skills, etc.

The *overall aim* for the structure of the service organization should be to create a learning organization. As stated before, mostly the people, the employees working with the customer every day, will be the

competitive edge. Coaching, or enabling, them to learn and grow, be committed, and have fun and pride in their job, will always create value for and to the customer.

3.9.6.3 Teamwork

Last but not least, a couple of words on teamwork. A lot of the success of outsourcing companies managing outsourcing solutions is based on teamwork. This starts during the engagement/sales process, continues during the solution implementation process and, you could actually say, begins during the management process of the solution. After all, it is ten times as expensive to win back a customer that walks away than it is to keep the customer in the first place.

To have professionals work as a team requires special skills from the "coach." The ability to create a team with skill sets that complement each other, stress the team targets, and still reward/penalize the individual for (non) performance are not given to all managers (as it's also a challenge for professionals to be team players!).

Therefore, teams should have clear objectives — both as a team and on an individual basis. Complementing skills sets, thinkers and doers should be committed, and have a sense of urgency. Research has proven that successful teams should have a strong desire to perform — encouraged, for instance, by the company's competitor.

3.9.7 Conclusions

In short, the winners of the outsourcing game are the companies — as far as managing outsourcing solutions — that have a thorough and well-defined plan for their people, processes, and tools. Furthermore, managing the solution is not only having a good plan — it is about the smart planning around the customer of all these three!

Finally, I believe that the outsourcing trend will continue to grow. Reading, for instance, through Tapscott's *The Digital Economy* and Aidarous and Plevyak's *Telecommunications Network Management into the 21st Century,* I see enough indications that lead to the conclusion that companies will seek strategic relationships with suppliers — outsourcing solution providers — that will provide them with the competitive edge of new technologies without having to worry about how to manage these wonders of mankind.

References and Bibliography

Aidarou, S., Plevyak, T., *Telecommunications Network Management into the 21st Century,* 1994.

Brelin H.K., Davenport K.S., Jennings, L.P., Murphy, P.F., *Focused Quality Managing for Results,* 1994.

Cowel, D., *The Marketing of Services,* 1984.

Datacommunications International, May 1997, Rating the World's Best Carriers.

Due, R.T., The real costs of Outsourcing, *Information Systems Management,* winter 1992.

Harrington, H. James, *Business Process Improvement,* 1991.

Hoogeveen, D., *Outsourcing,* 1994.

Katzenbach, J.R., Smith, D.K., *The Wisdom of Teams Creating the High-Performance Organization,* 1993.

Khandpur, N.K., Laub, L., *Delivering World-Class Technical Support,* 1997.

Lasher, D.R., Ives, B., and Järvenpää, S.L., USAA — IBM partnerships in information technology: Managing the image project, *MIS Quarterly,* December 1991.

Masaaki, Imai, KAIZEN The Key to Japan's Competitive Success, 1986.

Minoli, D., *Analyzing Outsourcing Reengineering Information and Communication Systems,* 1995.

Phaal, Peter, *Lan Traffic Management,* 1994.

Sloman, Morris (edited by) *Network and Distributed Systems Management,* 1994.

Tapscott, D., *The Digital Economy: Promise and Peril in the Age of Networked Intelligence,* 1995.

3.10 Support Systems for Telecommunication Providers

Kornel Terplan

The telecommunications industry shows both evolutionary and revolutionary signs. Evolution is seen with incumbent carriers; revolutionary attributes are visible with new entrants. The technology itself shows a mixture of wireline and wireless services, supporting all major telecommunication forms, such as voice, data, and video.

3.10.1 Status, Definitions, and Markets of Operations, Business and Marketing Support Systems

OSSs (operations support systems), BSSs (business support systems), and MSSs (marketing support systems) represent a very complex but increasingly significant segment of the communication industry. All three types of support systems together will be called 3SS. OSS, BSS, and MSS software enables the support, administration, and management from day-to-day operations to traffic trending, capacity planning, and forecasting of communication network services providers. Customer care, billing, provisioning, order processing, and networks operational management are all functions implemented via OSSs, BSSs, and MSSs. Until recently, there was little opportunity for direct investments in this important telecommunications segment. Increasingly, however, both incumbent and new service providers have come to view these systems as critical service differentiating assets. As a result, there is a growing number of public and private companies expected to benefit from the strategic importance of these support systems.

Financial estimates are on the basis of 15% CAGR (compound annual growth rate) over the next few years, approximately until 2002–2005. Unlike the average annual growth rate (AAGR) method, CAGR takes into account the changes from year to year, not only in revenues but also in revenue growth rate. CAGR is the rate at which the amount in the final year represents the future value of the amount in the first year after a specific interval. This CAGR percentage is an average over all market segments, such as customer care and billing, provisioning and order processing, and network operations management.

Industry issues of OSSs, BSSs, and MSSs are:

- Upgrade cycles in support systems: As a result of global deregulation, carrier competition is driving the demand for new, more efficient back-office solutions. In addition to reducing operating expenses, advanced 3SSs improve time to market and often facilitate the introduction of new, revenue-producing solutions.
- Product-based vendor driven solutions: Carriers increasingly demand solutions, rather than raw technology and development kits for custom-developed 3SS solutions. The advent of technology standards encourages the use of best-of-breed vendor solutions.
- Emergence of complex, multiplatform environments: reliability and scalability of large centralized systems remain excellent. Service providers incorporate a multiplatform strategy augmenting existing investments in legacy solutions with newer technologies targeted at profitable customer market sectors.
- Emphasis on telecom systems integration: complex multiplatform, multivendor telecom networks require substantial systems integration for interoperability. With multiple client–server and legacy 3SSs in place, integration capabilities of vendors are in high demand.
- 3SS growth is tied to share-shift among telecom end-markets and carriers: the strongest near-term growth has been achieved by vendors targeting the fast growing telecom end markets, emerging LECs, and wireless carriers.
- Developing 3SS markets: 3SS growth is dominated by new carrier adoptions and incumbent upgrades. Developing markets, such as data solutions, local number portability, and carrier interconnection are likely to justify the next wave of 3SS spending.

- Convergence and telecom consolidation: this accelerates the use of advanced 3SSs. Consolidation of carriers across multiple end markets creates advantages for 3SSs targeting multiple end markets. It increases the complexity of telecom networks and demands for 3SS integration.
- Outsourcing: ongoing structural changes in the telecom industry will place new requirements on 3SSs. In order to concentrate on customer management, some back-office functions may be outsourced to service bureaus. These service bureaus might use 3SSs from the same vendors, but they use them in a shared fashion among multiple service providers.

3.10.2 Market Drivers for 3SSs

The market is changing very rapidly. 3SSs should be positioned well, and should meet telco expectations in a timely fashion. Principal market drivers are addressed in this segment.

3.10.2.1 Growth of the Global Telecommunications Market

Explosive telecom expansion driven by internal growth and acquisition is forcing telecommunications providers to assess the productivity of their current support systems. Growth and acquisition mean that the number of subscribers grow for existing services; new services are provisioned on existing infrastructures and completely new services on new infrastructures are deployed or acquired. Several 3SS vendors have striven to capitalize on this opportunity with solutions that reduce complexity. These 3SS vendors do not usually replace existing systems, but add functionality to accomodate new services, such as:

- Internet, intranet, and extranet
- Special data services on top of voice networks
- Wireless services
- Cable and video services
- Voice services on top of IP

Adding functionality and interoperational features with each other opens new business opportunities for 3SS vendors.

3.10.2.2 Increasing Network Complexity

As a result of customer expectations, the time-to-market of new services is extremely short. Incumbent and new telecommunications services providers do not have the time to build new, but to combine existing and new infrastructures, such as copper, fiber, and wireless. They are deploying emerging services on the basis of a mixture of infrastructures as an overlay. Emerging services use emerged and emerging technologies, such as:

- Emerged technologies (voice networks, ISDN, circuit switching, packet switching, message switching, frame relay, Fast Ethernet, Fast Token Ring, and FDDI/CDDI)
- Emerging technologies (ATM, mobile and wireless, SMDS, Sonet/SDH, cable, xDSL and B-ISDN)

Each of these technologies has its own support system solutions. The only elements in public switched telephone networks (PSTNs) that should be managed are the switches themselves. On average, the ratio of managed elements to subscriber lines is around 1:10,000. The advent of distributed, software-based switching and transmission created a large number of additional managed elements, about one for each 500 subscriber lines. Moreover, multiple elements per subscriber in digital loop carrier systems, digital cellular networks, or hybrid fiber/coax systems may cause an explosion in terms of managed elements. As a result, the size of configuration databases and event messages generated by more intelligent network elements have grown exponentially over the last 20 years.

Growth in the number of network elements has been accompanied by an increase in the complexity of items to be managed. Sonet/SDH, ATM, and digital wireless are highly complex, with a high degree of interdependence among network elements. This in turn makes service activation and fault isolation

a challange, especially as the number of service providers increases. As networks shift from lower-speed, dedicated-rate, and inflexible services to mobile, fully configurable, bandwidth-on-demand and high-speed services, 3SSs must adapt to this new situation.

When services are offered in combination, 3SSs should be modified, re-engineered, and connected to each other. This opens new business opportunities for 3SS vendors.

The introduction of standards for support systems is accelerating the demand for third-party 3SSs. Legacy systems are primarily proprietary systems not integrated across functional areas. Service providers depend upon custom development by internal development staff and outside integrators to connect various support systems. The introduction of technology standards such as telecommuication management network (TMN), Distributed Communication Object Model (DCOM), Common Object Request Broker Architecture (CORBA), Telecommunications Information Networking Architecture (TINA), and Web-based Enterprise Management (WBEM) have begun to gain critical support by new 3SS vendors.

TMN is a special network in its own right that is implemented to help manage the telecommunication network of the service provider. As such, it interfaces to one or more individual networks at several points in order to exchange information. It is logically separate from the networks it manages, and may be physically separate as well. However, TMN may use parts of the telecommunication networks for its own communications.

TMN is an extension of the OSI standardization process. It attempts to standardize some of the functionality and many of the interfaces of the managed networks. When fully implemented, the result will be a higher level of integration. TMN is usually described by three architectures:

- The functional architecture describes the appropriate distribution of functionaliy within TMN, appropriate in the sense of allowing for the creation of function blocks from which a TMN of any complexity can be implemented. The definition of function blocks and reference points between them leads to the requirements for the TMN-recommended interface specifications.
- The information architecture, based on an object-oriented approach, gives the rationale for the application of OSI systems management principles to the TMN principles. The OSI systems management principles are mapped onto the TMN principles and, where necessary, are expanded to fit the TMN environment.
- The physical architecture describes interfaces that can actually be implemented together with examples of physical components that make up the TMN.

TMN distributes management responsibilities into several layers, such as business management layer (BML), service management layer (SML), network management layer (NML), element management layer (EML), and into the actual network elements layer (NEL).

DCOM is the heart of Microsoft's ActiveOSS product suite. Basically, DCOM is an integration infrastructure designed to facilitate communication between software components operating on the same host or with DCOM on multiple networked hosts. It was originally developed to create interoperability between components. It is the most widely deployed component object model. Active OSS acts as a centralized management and translation point for an OSS network. Conceptually, applications ride on top of the framework, but communicate through it. DCOM abstracts various application interfaces into objects, basically mapping the functions of the application into a common model that can be stored in a database. The common model allows the various applications to communicate in a uniform manner within the framework or across multiple networked frameworks.

By abstracting interfaces into software objects, applications theoretically can be upgraded and/or changed without affecting surrounding systems because integration is based upon independent software components that communicate, not applications that are heavily modified to fit together one-to-one. In this sense, upgrading an application means mapping a new interface into the framework, or modifying an existing one. The frameworks need to work with the interface, but do not need to affect details of the application. The framework is intended to create uniformity among application services without any

modifications to source code. Application services are built into and managed by the framework. The overall architecture also incorporates Smart TMN business process model and related work by TINA.

CORBA is a generic communication framework to connect various network management applications. The object request broker is the coordinator between distributed objects. The broker receives messages, inquiries, and results from objects, and routes them to the right destination. If the objects are in a heterogeneous environment, multiple brokers are required. They will talk to each other in the future by a new protocol based on TCP/IP. There is no information model available; no operations are predefined for objects. But an object does exist containing all the necessary interfaces to the object request broker. For the description, the Interface Definition Language (IDL) is being used. There are no detailed MIBs for objects because OMA is not management specific.

The functional model consists of the Object Services Architecture. It delivers the framework for defining objects, services, and functions. Examples for services are instanciation, naming, storing objects' attributes and the distribution/receipt of events and notification. CORBA services and facilities represent more generic services; they are expected to occur in multiple applications or they are used in specific applications. The driving force beyond designing common facilities for systems management is the X/Open Systems Management Working Group. The Managed Set Service, defined by this group, encourages grouping of objects in accordance to their management needs, with the result of easier administration. In the future, more services are expected to be defined; the next is an Event Management Service that expands the present Object Event Service by a flexible mechanism of event filtering.

Telecommunications Information Networking Architecture (TINA) is based on the concept that call processing in networks, and its control and management are separated from each other. TINA is actually a concept-integrator from IN, TMN, and Open Distributed Processing (ODP) from ISO and CORBA from OMG. The core is OSI-based network management, expanded by the layered structure of TMN. The emphasis with TINA is not on the management of network elements, but on the network and services layers. TINA is going to be standardized by a consortium consisting of telecommunications suppliers, as well as computer and software vendors.

WBEM is a joint initiative of many manufacturers, led by Compaq, Microsoft, and Cisco. The initial announcement called for defining the following specifications:

- HyperMedia Management Schema (HMMS): an extensible data description for representing the managed environment that was to be further defined by the Desktop Management Task Force (DMTF).
- HyperMedia Object Manager (HMOM): data model consolidating management data from different sources; a C++ reference implementation and specification, defined by Microsoft and Compaq, to be placed in the public domain.
- HyperMedia Management Protocol (HMMP): a communication protocol embodying HMMS, running over HTTP and with interfaces to SNMP and DMI.
- Common Information Model (CIM): basis of the information exchange between various management applications.

WBEM is helpful to unify and simplify network management.

The implementation of standard gateways enables interaction between newer client/server solutions with existing legacy systems and eases interoperability among all 3SS systems. In particular, TMN may help to streamline 3SS processes and to position support systems.

3.10.2.3 Deregulation and Privatization

Telecommunications service competition began in the 1980s in the U.S., led by MCI with 3SSs playing a key role. The AT&T divesture in 1984 marked a major breakthrough. The second significant milestone was the Telecom Act of 1996. As telecom deregulation continues, with RBOCs actively pursuing the long distance market and long distance carriers moving into local services, major 3SS re-engineering efforts are expected.

Under the pressure of the European Commission (EC), Europe is in the process of deregulation and privatization. It is a much slower process than in the U.S., because multiple countries are involved, each with their own agenda. Interoperability of 3SSs is more difficult than in the U.S.; but at the same time, it offers opportunities for 3SS vendors.

It is assumed that Asia/Pacific, South America, Eastern Europe, and Africa will follow these deregulation and privatization trends.

Competition is everywhere; long distance, local exchange, ISP, cable, and wireless. In many cases, 3SSs are the differentiators. The best 3SS opportunities are seen with CLECs. 3SS requirements vary substantially from carrier to carrier. As a result, CLEC-3SS-strategies are ranging from internal development to outsourcing to systems integrators and to third-party software/service providers. CLECs could be small or mid-size, they may own facilities, or are facilityless. In all cases, they must interoperate with ILECs by opening the 3SS to permit access by CLECs in various phases of provisioning and order processing, and service activation. Key issues are:

- Local number portability (LNP): it allows customers to retain their telephone numbers even if they change service providers. It is not only the telephone number that is important, customers also typically want to retain access to advanced features they have come to expect from an intelligent network.
- Extranets connecting 3SSs of ILECs and CLECs: ILECs are required to provide access to information on five classes of 3SSs: preordering, ordering, provisioning, repair, and maintenance.
- Directory services: real-time service processing requires additional customer-related data. The expanded directory role includes end-user authorization and authentication. It also includes the real-time allocation of network resources according to a user's class of service and other policy-based variables. Directory Enabled Networks (DENs) promise to increase momentum for directory services by bringing physical infrastructure under the directory umbrella and tackling the standardization of directory information.

Incumbent service providers have turned to advanced 3SSs to differentiate their long distance or local exchange services from each other. After a substantial investment in custom systems over the last few years, many incumbents have begun to focus on upgrading select 3SS systems with best-of-breed technologies. Many of them try to augment older systems to add more flexibility while sustaining traditional levels of performance and reliability. This creates additional complexity and requires that new management solutions designed for advanced equipment also work with older technologies.

As a result, umbrella-types of 3SSs are in demand, opening new opportunities for 3SS vendors with integration capabilities. To remain competitive, incumbent carriers need to deliver an increasingly larger number of new products and services. This has created a mixture of equipment, software, and services within many carriers.

Innovation and re-engineering on behalf of the incumbent carriers show:

- Better customer care: based on call detail record (CDRs) and other resource utilization-related data, unsophisticated customer analysis can be accomplished. It includes discovering trends in customer behavior, traffic patterns, reasons for frauds, and also service-related items.
- Convergent billing: the customer may expect to receive one bill for all services, such as voice, data, video, and Internet. The minimal requirement is to receive multiple bills with electronic staples.
- Rapid provisioning new services: based on additional 3SSs, provisioning can be expedited by better interfaces and more accurate data.
- Service differentiation: still using the same infrastructures, new services can be created and deployed. By carefully defining the value-added nature, it may be considered by customers as differentiators.
- Offering new services: incumbent service providers are expected to react rapidly to new communication needs, including offering Internet access for reasonable money, the deployment of xDSL, VPNS and VoIP.

In each of these cases, either the deployment of new 3SSs or the customization of existing 3SSs are required. In both cases, additional market opportunities open for 3SS vendors.

3.10.2.4 Communication Convergence

Advanced technology, coupled with deregulation, is driving communications convergence. Customers prefer to get all types of services, such as long distance and local voice, data/Internet, cable/video, and wireless access from the same service provider. Voice is expected to support both local and long distance, requiring to play a LEC and IEX role at the same time. Data is gaining importance for both local and long distance, and does usually include Internet access. Data is supposed to reach voice volumes within 5 years, requesting the total rebuilding of circuit switching technology. Cable is expected to accommodate voice and data in addition to video. Wireless does include all kinds of mobile services and satellites supporting voice, video, and data.

Deregulation is meant to encourage competition through the proliferation of new entrants. Looking to gain share, carriers are entering each other's market, blurring traditional lines between services, geographic coverage, and communication platforms. Aggressive new carriers have moved rapidly to establish nationwide service networks, consolidating local, long distance, Internet, wireless, and cable services under one umbrella. Incumbent carriers are trailing this way of convergence. The U.S. shows an excellent example of this convergence, the "big eight" convergence carriers

	Local	Long Distance	Data/Internet	Wireless	Cable/Video
AT&T	Teleport TCI	AT&T	AT&T WorldNet	AWS/McCaw	TCI
Bell Atlantic	Bell Atlantic Nynex		BA New Media	Bell Atlantic	
Bell South	Bell South		BS New Media	Bell South Mobile	Wireless Cable
GTE	GTE	GTE	BBN		
SBC	SBC Pactel Ameritech			Ameritec Wireless	Ameritech in-region cable TV
Sprint	Sprint ION	Sprint	Parenet/Earthlink	Sprint PCS	
Qwest/US West	US West		Internal Enterprise Networking Division (Cisco)		
WorldCom	Brooks MFS	MCI Worldcom	MCI CompuServe/ ANS UUNet		

cover most end markets. But they still leave room for hundreds of point products, mostly best-of-breed telco products and services. Communication convergence necessitates the deployment of next generation 3SSs. Relying upon advanced technologies, client/server or Web-based 3SSs enable convergence carriers to offer their customers higher total value through new, innovative products and services, superior customer service, and customized pricing and billing. At the same time, 3SSs guarantee profitability by increasing effectivity of processes by automation of all routine processes and by supervising quality of services metrics.

3.10.2.5 Customer Orientation

Competition is driving telco service providers to emphasize customer management. Driven by global competition, carriers are likely to focus on improving the total value of their services — quality, support, and price — as means to retain customers. Many of these improvements will come from advanced 3SSs. Besides improving the customer interface — e.g., offering Web-access — granular data available with new 3SSs can be utilized to retain key customers and reduce the number of customer churn. Over a

longer range, further differentiation is expected. High-margin customers may receive special treatment, average customers just average services — similar to other industries.

Customer network management (CNM) incorporates a class of 3SSs that enable end users to securely view, troubleshoot, reconfigure, and generate reports on their subscribed telecommunication services. CNM provides strategic links to the customer and allows service providers to further differentiate their offerings. 3SS vendors are expected to offer the following:

- Performance: extraction of the information from the network without slowing overall network operations
- Customization: packaging information so that customers can receive an appropriate level of detail, in a way they can understand
- Security: delivery of the information to the customer in a cost-effective and secure manner, so that customers see only relevant information about their portion of the network

It is expected that Web technology will primarily be used to deliver this service. CNM represents a modest source of incremental growth for 3SS suppliers.

Certain 3SS services can also be outsourced. The customers may not be aware where the 3SS services come from. Today's outsourced solutions are service bureaus. They may outsource all or part of the carrier's support systems. In the latter case, the vendor relies upon remote access to the carrier's existing solution to deliver incremental functionality. For most emerging carriers, the benefits of outsourcing outweigh the negatives.

3.10.3 Strategic Benefits of Advanced 3SSs

Once deployed, advanced 3SSs offer the following strategic benefits:

- Improved operating efficiencies in data, inventory, and network management: it is expected that the management of various objects, such as equipment, applications, databases, etc. is more integrated, and requesting less human resources to manage.
- Reduced support and maintenance costs associated with legacy systems: due to more automation and interconnection, the support and maintenance expenses are decreasing.
- Shorter product development cycles: products and services can be created, tested, and deployed faster due to advanced technology used in 3SSs.
- Speedier deployment of new services and pricing schemes: processes are connected to each other. Rapid service provisioning in combination with pricing guarantee rapid deployment.
- Flexibility to modifying pricing and marketing schemes: due to interconnected processes, changes can be deployed very quickly. Even modeling and simulating resource utilization scenarios is easy to implement.
- New synergistic products and convergent services: products' bonding is very helpful to support convergent services. This bonding integrates OSSs, BSSs, and MSSs.
- Strategic marketing to target and acquire profitable business customers: due to rich information on customers and their traffic generation patterns, marketing strategies can be customized.
- Superior customer management to establish customer loyalty: The significant improvement of customer care will help to avoid customer churn and to sell value-added communication services to loyal customers.

The three principal 3SS process segments are:

- Customer care and billing
- Provisioning and order processing
- Network operational management

Figure 3.10.1 shows the high-level flow between these three process segments. It is important to observe that the corporate database or repository or data warehouse is shared between the principal process groups.

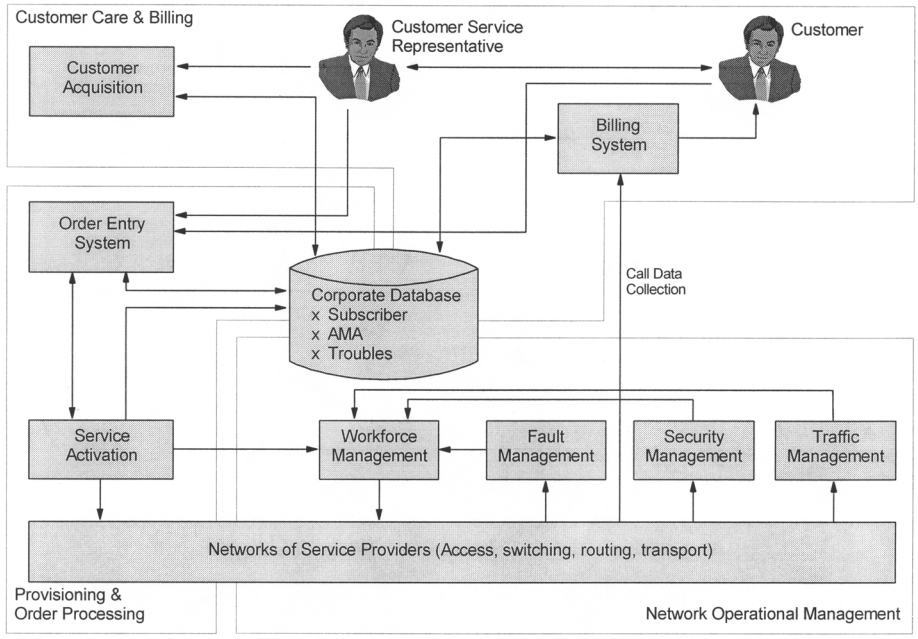

FIGURE 3.10.1 Principal 3SS process segments.

Table 3.10.1 shows the principal processes and functions for each of these segments. The present estimated market shares by segments are:

Customer care and billing	42%
Provisioning and order processing	38%
Network operational management	20%

The ratio of this market share will not change substantially during the next 3–5 years. Support tools are available for each segment offered by approximately 250 to 300 vendors, but it is rare that one product or one vendor can offer solutions for each market segment.

3.10.4 Providers of Operational, Business, and Marketing Support Systems

There are approximately 250 to 300 companies that are successfully competing in the 3SS arena. The categorization of providers is the following:

Software framework vendors (e.g., Telcordia, OSI, TCSI, Micromuse)
Consulting companies (e.g., Andersen Consulting, American Management Systems)
Computer vendors (e.g., IBM, Compaq/DEC, HP, Bull)
Telco equipment vendors (e.g., Ericsson, Siemens, Nortel, Nokia)
Application vendors (e.g., Versant, Vertel, Saville, Kenan, Metrica, Concord, NetScout)
Outsourcers (e.g., EDS, Perot Systems)

The present market shares can be characterized as follows:

- Telcordia and Lucent are equally strong and take up to 20% of the total 3SS market.
- Another 18 companies that are well known in the 3SS branch take approximately 10%

TABLE 3.10.1 Principal 3SS Processes and Functions

3SS Processes and Functions	Definition
	Customer Care and Billing
Data analysis and mining	Process of analyzing call data details collected from switches and transmits
Mediation	An intermediate step for pre-processing and analyzing CDRs; fraudulent calls can be removed, data input from different switches in multiple formats can be converted into a format appropriate for bill processing. Also, pricing schemes can be inserted here, rather than by the call rating module. Call data can be selected and then transmitted to individual billing platforms, such as for voice, data, wireless, Internet, etc. Mediation is increasingly used for convergent and real-time billing.
Call rating and discounting	Prices call data according to current plan; it does include threshold plans currently popular among wireless carriers. Also, discounts are considered with this function.
Bill compilation and processing	Aggregates the rated call detail records and adds data for multiple services, handles advanced charges and payments.
Bill presentment	Customizes bill formats on a customer or service provider basis, may consolidate multiple statements; delivers bills via mail, online, e-mail, tape, or Internet.
Revenue assurance	Factoring, finding of receivables, credit checks, remittance processing, and customer deposit management.
Collection and credit analysis	Collecting outstanding debt, usually by the help of third-party collectors.
Customer care	Evaluates historical customer requirements, traffic patterns, expectations; reports and solves technical and billing problems.
Customer analysis and acquisition	Billing platforms tend to maintain the most complex picture of telecom customers in terms of resource usage, habits, and traffic patterns. Using these data intelligently, customer churn can be avoided and new services can be sold to customers.
Data warehousing	Call detail records, and additional data sources can be transmitted into warehouses. Data mining and other applications help to determine customers and end-product profitability.
	Provisioning and Order Processing
Local number portability	Allows customers to retain their telephone numbers with multiple service providers. Also, access to value-added services can be retained.
Inventory management	Allows maintaining first of all technical inventory data about equipment and circuits for the geographical reach of the service providers. Both CAD/CAM and GIS solutions may be implemented. Connections to the data warehouse are obvious.
Service creation	Process of creating and testing new or advanced services on the basis of the existing infrastructure of the service providers.
Service activation, provisioning, and assignment	Process of allocating equipment, assigning numbers, and activating circuits or ports at switches and activating customer services.
Service order processing	Based upon customer requests submitted to customer service representatives, creation and activation of services to customers.
Handling service change requests	Based upon customer service change requests submitted to customer service representatives, changing and reactivation of services to customers.
Service assurance	Allows continuous supervision of service level agreements on service indicators, such as availability, throughput, call congestion, packet losses, CDR losses, and others. For violations, the billing module is informed to initiate discounts or reimbursements.
Capacity management	Process of periodic surveillance of capacity in equipment and circuits. If capacity thresholds are exceeded, capacity extensions are initiated automatically.
	Network Operational Management
Call data collection	Collects call detail records (CDRs) from switches and transmits them to a billing database or mediation device. State-of-the-art solutions use complete automation.
Reactive fault management	Process of determining, diagnosing, and resolving faults, detected and reported by customers or by fault monitoring devices.
Proactive fault management	In order to detect problems early, allows the continuous supervision of fault indicators, the identification of causes for chronic troubles, and the evaluation of vendor performance.

TABLE 3.10.1 (continued) Principal 3SS Processes and Functions

3SS Processes and Functions	Definition
Preventive fault management	Allows evaluation of usage statistics, the causes of performance threshold violations and the impact of additional payload on equipment and circuits.
Performance monitoring	In order to further support preventive fault management, equipment and facilities (circuits) are monitored continuously. In addition, performance metrics are maintained in a repository, which can be part of the data warehouse.
Error repair and maintenance	Allows repair of chronic faults and deployment of preventive maintenance techniques to equipment and to facilities.
Installation and inspection	Allows as part of the provisioning process the physical deployment of equipment and facilities on the basis of provisioning and service order change requests of customers.
Security management	Process of identifying security risks in equipment and facilities, deploying security procedures and tools, creating and evaluating security logs, and protecting operations, business, and marketing support systems.
Workforce management	Allows the central, policy-based, dispatch of workforce to monitor, test, maintain, inspect, and install equipment and facilities.
Testing	Process of testing equipment and facilities prior to deployment or as a part of the error repair process.
Design and planning	Allows, as a result of capacity bottlenecks, initiation of design processes that may include the deployment of new technology to equipment and facilities.
Traffic management	Process of observing typical traffic patterns by customers, customer groups, geographical areas, equipment, and facilities types. As a result, parameters and controls can be changed in equipment and facilities.
Network systems administration	May be considered as part of the maintenance process, limited, however, to version control, backup, archiving, and distribution of software to equipment.

- The remaining 70% is distributed between literally hundreds of companies that are eager to emerge as dominant suppliers of 3SS solutions.

The requirements for being a winner are tough. The profile may look like the following:

- Solutions are scalable: given the large and growing number of network devices, services, and subscribers, 3SSs must grow with the service provider. While there may be a low-end market for small-scale "telco-in-the-box" solutions, it is not expected that solutions that do not scale well will capture significant market segments. Prepackaged functionality will help to reduce the demand of customization required to match 3SSs to a particular service provider's business objectives.
- Domain knowledge (best-of-breed): the implementation and deployment of 3SSs require a sound knowledge of service provider operational procedures. This domain knowledge is not always available, but without it, successful work in the domain is not possible.
- Integration capabilities (best-of-suite): it is absolutely necessary to connect existing point products using electronic bonding or extranets. Standards are emerging to facilitate this work. First implementation results are seen with CORBA and DCOM.
- Supporting multiple products and services: the ability to manage traditional, enhanced, wireless, data, and video products and services in a unified convergent manner is widely viewed as critical to the success of advanced service providers.
- Willingness for partnerships: it is not possible to exhibit open multivendor support without effective partnerships. It is highly likely that leading 3SS vendors will establish partnerships with other industry stakeholders, such as equipment vendors, system integrators, customers, and other 3SS vendors.
- Strong references: perhaps the best selling argument for a 3SS is its existing customer list. Service providers, recognizing the high cost associated with maintaining and enhancing a 3SS platform, view a strong customer base as a way to share development costs for basic 3SS functionality and reduce risk.

3.10.5 Positioning and Evaluating Products

3SS products can be positioned, evaluated, and compared to each other using the following three dimensions:

- Compliance to TMN layers, such as BML, SML, NML, EML, and NEL
- Support of principal 3SS processes and functions (see Table 3.10.1 for details)
- Support of various end markets, such as long distance, local, data/Internet, wireless, and cable

In order to help position, evaluate, and compare support tools, the following tables are recommended:

TABLE 3.10.2 Principal 3SS Process Groups and TMN Layers

Process Areas	Customer Care and Billing	Provisioning and Order Management	Network Operational Management
TMN Layers			
BML	x		
SML	x	x	
NML		x	x
EML			x
NEL			x

The allocation between principal 3SS process groups and TMN layers shows a very clear trend (Table 3.10.2):

- Customer care and billing supports upper TMN layers
- Provisioning and order processing supports middle TMN layers
- Network operational management supports lower TMN layers

This allocation remains relatively stable even when the TMN standard changes over time.

TABLE 3.10.3 Principal 3SS Processes and Functions and TMN Layers

TMN Layers	BML	SML	NML	EML	NEL
OSS Process Areas					
Billing and Customer Care					
Data analysis and mining	x	x	x	x	x
Mediation	x	x	x		
Call rating and discounting	x				
Bill compilation and processing	x				
Bill presentment	x				
Collections and credit analysis	x				
Revenue assurance	x				
Customer care	x	x			
Customer analysis and acquisition	x				
Data warehousing	x				
Provisioning and Order Management					
Local Number Portability		x			
Inventory management		x	x	x	x
Service creation		x			
Service activation, provisioning and assignment		x	x	x	x
Service order processing		x	x	x	
Handling service change requests		x	x	x	
Service assurance (SLA)		x			
Capacity management			x		x

TABLE 3.10.3 (continued) Principal 3SS Processes and Functions and TMN Layers

TMN Layers	BML	SML	NML	EML	NEL
Network Operational Management					
Call data collection		x	x	x	x
Reactive fault management			x	x	x
Proactive fault management			x	x	x
Preventive fault management			x	x	x
Performance monitoring		x	x	x	x
Error repair and maintenance			x	x	x
Security management	x	x	x	x	x
Installation and inspection			x	x	x
Workforce management			x	x	x
Testing		x	x	x	x
Design and planning			x		x
Traffic management		x	x		x
Network systems administration			x	x	

Breaking down the principal 3SS process groups (Table 3.10.3) does not change the allocation to TMN layers significantly. But for certain processes, exceptions can be observed. In other words, the allocation is not unique. The result is that TMN needs additional work to clarify responsibilities of layers and their functions more accurately.

TABLE 3.10.4 Support of 3SS Processes and Functions by Individual Support Tools

Companies	Functions	Products
	Customer Care and Billing:	
Acc*Comm	Billing data collection	TIBS, NetPlus, DCMS, DCMS/NEDS, ANMS TREX*COM NetPlus Pro*Vision
Alltel Information Services	Billing	Virtuoso II
Amdocs	Billing applications	Ensemble
American Management System	Customer care Billing	Mobile 2000, Spectrum 2000 Tieline UP, Tapestry
Andersen Consulting	Billing applications Customer care	IABS (Integrated Access Billing System), Flexcab
Axiom	Billing data collection Fraud management	Sterling Billing Data Collection Sterling Real-Time Fraud Management
Beechwood	Carrier-to-carrier OSS interconnection New carrier systems integration Post-merger IT integration IP network OSS implementation	
Billing Concepts	Service bureau billing	Modular Business Applications (MBA)
Cable Data	Billing	Intelecable
CBIS	Billing software and services	Wireline: Precedent 2000 Wireless: Advantage
Clarify	Customer care applications	
Commsoft	Billing	CommVergence
Corsair	Fraud detection systems	
CSG Systems	Large number of functions, services, and products	CCS with ACSR
Daleen Technologies	Billing	BillPlex
EDS	Service bureau billing	BSM (Billing Services Management), IXPlus, CMIS, Empower

TABLE 3.10.4 (continued) Support of 3SS Processes and Functions by Individual Support Tools

Companies	Functions	Products
Ericsson	Customer service and billing	TIMS, BIP (Billing Information Processor) BMP (Billing Mediation Platform), Progressor
IBM	Customer care and billing	TFS/ICMS CARTS (Centralized AMA Records Transfer System)
InfoDirections	Billing	CostGuard
Infozech	Billing	eBill
Intertech Management Group	Billing applications	Network Strategies
Kenan Systems	Billing applications	Arbor/BP, Acumate ES, Arbor Strategist, EC/Arbor
LHS Group	Billing applications	Business Support and Control System (BSCS)
Lightbridge	Customer care applications Fraud detection systems	Telesto
Lucent Technologies	Billing data collection	BILLDATS
Metapath	Billing data collection	
Objective Systems Integrators	Billing data collection	NetExpert, AMA Gateway
Portal Software	Internet billing applications Customer care applications	Infranet
Saville Systems	Billing applications	Convergent Billing Platform (CBP)
Sema Group	Customer care and billing	CABS 2000 Mobile
USCS	Cable billing	Cable Data
	International billing	IBS (Intern. Billing System)
	Provisioning and Order Processing	
Amdocs	Order management	
Applied Digital Access	Service activation, provisioning	Provisioner
American Management Systems	Service order management	Tieline SOMS
Architel	Service activation	FAMIS, ASAP, OMS
Atlantec	xDSL service activation	
Beechwood	Workflow for provisioning	Flow-through
Bellcore/SAIC	Provisioning	TransportEMS
	Switch administration	NetMemory
	Workforce management	FORCE
	Service order management	Delivery
	Legacy 3SS maintenance	
Call Technologies	Provisioning and enhanced services	Call Profiler, Call Activate, Call Verify, Call Notify, Call Courier, Call Builder, Call Codence, Call Care, Call Plus
CBIS	Configuration management	Switch Manager
CommTech	xDSL and Centrex service activation	BECAS Facility Management
Crosskeys	Service management	Resolve
DSET	Service order administration center	LSOA
	Local service management system	LSMS
EDS	Workforce management	FMS (Force Management (System)
	Inventory and service provisioning	EMAC (Enhanced Mechanized and Control System)
Evolving Systems	Workflow management Event management	
FirsTTel	Service activation for phone systems	
Hewlett-Packard	Service management	OpenView ITA (Admin)
Illuminet		
Lucent Technologies	Service provisioning	CONNECTVU
	Switch administration and management	NetMinder
	Service activation	ACTIVIEW
MetaSolv	Service activation, repository, and provisioning	TBS (Telecom Business Solution)
Network Programs	Rapid service deployment	TNP (The New Platform)

TABLE 3.10.4 (continued) Support of 3SS Processes and Functions by Individual Support Tools

Companies	Functions	Products
Nortel	Provisioning	DSS-II
Objective Systems Integrators	Service activation	NetExpert, iSOP, iSAC
Quintessent Communications	3SS interconnections for provisioning	QConnect
SmallWorld	Inventory management	GIS-based 3SS
	Network Operational Management	
ADC Metrica	Wireless performance management	NPR (Network Performance Reporting)
ADC NewNet	SS7 management	AcceeMANAGER
Applied Digital Access	Testing	T3AS Test
	Fault detection and isolation	T3AS Monitoring
Ascend	Managing POPs	Navis Access, Navis Core
Axiom	Traffic management	Manifest
Bellcore/SAIC	Operations management	NMA
	Testing	OcuSpan
	SS7 management	NetPilot
	Legacy 3SS maintenance	
Cisco	Service management	Cisco Service Management System
Clear Communications	Data collection and reporting	ClearView Probable Cause, Early Warning, ReportCard, CircuitView, Legacy Gateway
CommTech	Centrex monitoring	Macstar
Compaq/Digital	SS7 management	DECss7
Concord Communications	Data service level monitoring	Network Health
CrossKeys	Network management	Exchange Performance Management, Exchange Traffic Management
	Testing	Open/Test
DeskTalk	Data service level monitoring	TREND
DSET	NPAC-SMS simulator	Simulator
Ericsson	Network management	TMOS Network Traffic Manager, XM (Exchange Manager)
	SS7 management	TMOS SS7
Hewlett-Packard	Network management	OpenView ITO (Operations) OpenView MeasureWare
Illuminet	SS7 monitoring	
INET	SS7 monitoring	
ISR Global Telecom	Cable, SDH/Sonet management	ObjectEngine, Mask
Lucent Technologies	Testing	SARTS
	Fault detection, isolation, and reporting	ITM, NMF, NOCI
	ChoiceNet	Dynamic filter management
Micromuse	Alarm correlation and analysis	Netcool suite
NetScout	Data network performance monitoring	RMON Probe
Nortel	Network management	S/DMS, NetWORKS, DFMS (Digital Facility Management System)
Objective Systems Integrators	Fault detection, isolation, and resolution	NetExpert
	Alarm correlation and analysis	NetExpert IDEAS
OpenCom Systems	Integration of element managers	TMS2000 Global EMS Product
Remedy Corporation	Trouble ticket management	AR System
Team Telecom	Fault management applications	
Technically Elite	Data network performance monitoring	MeterWorks
Visual Networks	Data network service level monitoring	

This table (Table 3.10.4) groups 3SS products around three principal process groups, such as

- Customer care and billing
- Provisioning and order processing
- Network operational management

Due to acquisitions and changes in product portfolios, such a table needs frequent updates. In addition to these groups, an additional list (Table 3.10.5) is provided for vendors of frameworks and platforms that offer enabling technologies and integration services.

TABLE 3.10.5 Framework and Platform Vendors, and System Integrators

Companies	Functions	Products
ADC SoftXchange	Interconnection platform	DataXchange
Alcatel	Systems integration	1320NM, TSM, CBCU
AllTel	Remote outsourcing	
Amdocs		
American Management Systems	Consulting and integration	
Andersen Consulting	Consulting and integration	
Beechwood	Carrier-to-carrier integration	
	New carrier system integration	
	Post merger IT consolidation	
Bellcore/SAIC	Consulting, integration, and custom design	
Bull	Management platform	Open/Master
CapGemini	Systems integration	OSS, TTM, BSCS, MARS
Compaq/Digital	Systems integration	TeMIP management platform
DMR Consulting	Consulting and integration	
DSET	Development tools	TMN Agent and TMN Manager
	Electronic bonding	EC-Lite, PIC/CARE
	Intelligent networks	LNP
EDS	System integration	
Euristix	Tools and custom development	Raceman EMSX
	Q-Adapter	
Evolving Systems	Platform and management	

TABLE 3.10.6 Support of End Markets by Individual Support Tools

Company	Product	Local	Long distance	Data/Internet	Wireless	Cable/Video
			Customer Care and Billing			
Acc*Comm	TIBS, NetPlus, DCMS, ANMS, TREXCOM			x		
Amdocs		x	x		x	
AMS						
Andersen						
Axiom						
Beechwood						
Billing Concepts		x				
CBIS	Wireline Wireless	x	x		x	x
Clarify						
Corsair						
CSG Systems				x		x
Daleen		x	x			
Illuminet		x		x		
Intertech		x		x		
EDS						
Ericsson						

TABLE 3.10.6 (continued) Support of End Markets by Individual Support Tools

Company	Product	Local	Long distance	Data/Internet	Wireless	Cable/Video
IBM						
Kenan	Arbor/BP, Acumate ES, EC/Arbor	x	x	x	x	x
LHS Group					x	
Lightbridge						
Lucent						
Metapath						
Objective Systems						
Portal Software	Internet Bill			x		
Saville Systems	Convergent Billing Platform	x	x	x	x	
USCS	Cable Data, IBS					x

Another dimension is offered in this table by grouping 3SSs in accordance with end markets (Table 3.10.6) they support. This table provides selected examples for customer care and billing only.

TABLE 3.10.7 Support of TMN Layers by Individual Support Tools

Company	Product	BML	SML	NML	EML	NEL
	Customer Care and Billing					
Acc*Comm	TIBS, NetPlus, DCMS, Dcms/Neds, ANMS, Trex*Com	L	L	H	H	H
Amdocs		H	H			
Axiom		L		H	H	H
AMS	Mobile 2000 Spectrum 2000, Tieline UP	H	M			
Andersen Consulting	IABS	M	M			
Beechwood						
Billing Concepts		H				
CBIS		H	H			
Clarify		M	M			
Corsair						
CSG Systems						
Daleen		H	H			
Intertech						
EDS	BSM	H	M			
Ericsson	TIMS, BIP	H	H	M	M	M
IBM	TFS/ICMS CARTS	H	H			
Kenan	Arbor/BP, Acumate ES, Arbor Strat EC/Arbor	H	H			
LHS Group		H	H			
Lightbridge		H	M			
Lucent	BILLDATS	H	H			
Metapath				H	H	
Objective Systems	NetExpert AMA, Gateway	M	M	H	H	
Portal Software		H	H			
Saville Systems	Convergent Billing Platform	H	H			
USCS						
	Provisioning and Order Processing					
Amdocs						
Applied Digital Access	Provisioner			H	H	H

TABLE 3.10.7 (continued) Support of TMN Layers by Individual Support Tools

Company	Product	BML	SML	NML	EML	NEL
AMS	Tieline SOMS		H	M		
Architel	FAMIS, ASAP, OMS		H	H	M	
Atlantec		M	H	H		
Beechwood	Flow Through	H	H			
Bellcore/SAIC	MediaVantage, NCON, Transport, EMS, FORCE, Delivery, NetMemory		H	H	H	
Call Technologies			H	M	M	H
CBIS	Switch Manager			M	M	
CommTech			L	H	H	H
Crosskeys	Resolve		H	H	H	
DSET			L	M	H	H
Evolving Systems		M	H			
FirsTel			H	M	H	
Hewlett-Packard	OpenView ITA		M	M	H	
Illuminet						
Lucent	ACTIVIEW, OneVision, CONNECTIVU, NetMinder		H	H	H	M
MetaSolv	TBS	M	H	M		
Network Programs	The New Platform					
Nortel	DSS-II			M	M	M
Objective Systems	NetExpert, iSOP, iSAC		H	H	M	M
Quientessent		H	H			
SmallWorld			M	H		
	Network Operational Management					
ADC Metrica	NPR			H	H	H
ADC NewNet	SS7		M	M	M	M
Applied Digital Access	T3AS			H	H	H
Ascend	NavisAccess, NavisCore		M	H	M	
Bellcore/SAIC	NetPilot, NMA, OcuSpan	M	H	H	H	
Cisco	Service Management System		H	M	M	M
Clear Comm	ClearView Probable Cause, Early Warning, ReportCars, CircuitView, Legacy gateway	M	H	H	L	
CommTech			L	H	H	H
Concord	Network Health	M	H	M		
Compaq/Digital	DECss7			H		
Crosskeys	Exchange Performance and Traffic Management Open/Test		H	H	H	
DeskTalk		M	H	M		
Ericsson	TMOS Traffic and Exchange Manager, TMOS SS7		M	H	M	M
Hewlett-Packard	OpenView ITO and Measure Ware		M	H	M	M
Illuminet			H	H		
INET				H	H	L
ISR Global	ObjectEngine Mask			M	H	
Lucent	ITM, NOCI, NMF, NetMinder, SARTS, ChoiceNet		H	H	H	H
Micromuse	NetCool		M	H		
NetScout	RMON Probe	H	H	H	H	
Nortel	S/DMS, NetWORKS, DFMS		M	M	M	M
Objective Systems	NetExpert, NX Ideas		H	H	H	H
Remedy	AR Systems		H	H	M	
Team Telecom				H	M	
Technically Elite	MeterWorks		M	M	M	H
TCSI	SolutionCore		L	H	H	M
Visual Networks		M	H	M	M	M
Vertel			L	M	H	H

Using the same 3SSs as in Table 3.10.4, the support of TMN layers is evaluated (Table 3.10.7). As can be observed, there are many empty cells, indicating gaps in support.

The overall conclusions of 3SS evaluations are:

- Even the most powerful companies in the 3SS business (Bellcore and Lucent) and their products cannot support all principal 3SS process groups.
- There are many best-of-breed, and very few best-of-suite products.
- Support for higher TMN layers (BML and SML) is not sufficient.
- Serious segmentation of product's applicability for various end markets exists.
- Enabling technologies for integration and interconnection are absolutely necessary.

3.10.6 Future of Telecom 3SSs

In order to position 3SSs and their vendors, future trends of support systems should be estimated. With other words, the dynamic of principal market drivers should be analyzed in depth.

In order to match the rich service offerings of new entrants, ILECs have implemented multiple upgrade strategies, including modifications by internal staffs, custom development by external system integrators, and integration of third-party products. Most likely, they won't completely replace their existing 3SSs. Several incumbent carriers are incorporating best-of-breed solutions with their legacy systems. This trend opens great opportunities for point 3SSs and for professional services.

Deregulation of the LEC market has stimulated and still stimulates significant demand for CLEC-3SSs. Most of them start from scratch and invite all types of 3SSs vendors with point and integrated products. Larger CLECs with custom-designed in-house solutions are enhancing these to accommodate new services and technologies; they show some similarities with incumbent providers. Replacement of these 3SSs is not expected soon. But, in particular, back-office operational efficiency is expected to improve. Also, network operational management solutions are in demand.

CLECs may want to outsource 3SS services. They have started to evaluate the benefits of outsourcing their back-office services entirely. Outsourcing would eliminate the need for the carrier to invest scarce research and development dollars in 3SSs, allocating spending to their networks and/or customer management systems. Essentially, it allows CLECs to focus on their core business.

Less well-known CLECs either purchase or license point products from third parties, or take advantage of service bureaus. It is highly unlikely that these CLECs are interested with in-house development and maintenance. 3SS-vendors can sell to these CLECs directly or to service bureaus that may share their products between multiple CLECs.

The international market is not easy for 3SS suppliers. Systems integrators are in good position with PTTs and new entrants; they usually subcontract for point solutions.

Carrier interconnections open excellent opportunities for 3SS vendors. The unbundling of local exchange elements for resale requires that resellers of local exchange services provide electronic links to incumbent carriers for ordering, service activation, troubleshooting, and billing. Present 3SSs do not have these interconnect features. There is a significant opportunity for incremental 3SS sales by emerged and also by new vendors. Specialized vendors for LNP will play a significant role during the next 10 years. The best-of-breed solutions are expected to offer provider portability, location portability, and also service portability.

Telecom industry consolidation creates new 3SS requirements, but the need is situation specific. It is difficult to estimate the timeframes of re-engineering or consolidating 3SSs of the consolidated telecommunication providers. Consolidated carriers are likely to work to fully integrate multiple 3SS platforms (customer care and billing, provisioning and order processing, network operational management), to create synergies in products and markets, and to reduce costs. 3SS vendors with system integration capabilities are in demand. 3SSs will mirror trends in the telecommunication industry; full-service 3SS vendors could emerge to serve convergence carriers.

Table 3.10.8 summarizes the present goals and the 2002 targets from the perspective of incumbent carriers, emerging carriers, customers, equipment vendors, 3SS vendors, and system integrators.

TABLE 3.10.8 Goals and Targets

Stakeholder	Goals	Targets for 2002
Incumbent service providers (ILECs, PTTs, global carriers)	Rapid introduction of new services Address Year 2000 issues Customer retention Multi-vendor management Cost reduction	Less internal development More use of systems integrators More 3SS packaged software Pervasive 3SS interconnection
Emerging service providers (CLECs, ISPs, and wireless)	Build network capacity Customer acquisition Improve service quality Offer new differentiated services	Minimal internal development Fully automated processes More 3SS packaged software Less service bureaus Strong 3SS interconnections
Customers (end users)	Increase service reliability Lower transport costs Faster service provider responsiveness Customer network management	Self-provisioning Custom quality of service reporting (QoS) Flexible billing formats
Equipment vendors	Sell more equipment	Outsource element management systems to 3SS vendors Use of open interfaces
3SS vendors	Sell more software Sell more professional services	Full-line of 3SS offerings Target ILEC legacy 3SS replacement Acquire other 3SS vendors Compete with system integrators
System integrators	Sell consulting, custom programming, and integration services	Acquire 3SS vendors Conduct many projects

Acronyms

3SS	Operations, business, and marketing support systems
ASR	Access service request
ATM	Asynchronous transfer mode
BIP	Billing information processor
BMP	Billing mediation platform
BSM	Billing services management
BML	Business management layer
BSS	Business support system
CAGR	Compound annual growth rate
CARTS	Centralized AMA records transfer system
CBIS	Cincinnati Bell Information Systems, Inc.
CDR	Call detail record
CLEC	Competitive local exchange carrier
CNM	Customer network management
CBP	Convergent billing platform
DEN	Directory enabled networks
DFMS	Digital facility management system
DSL	Digital subscriber line
EML	Element management layer
EMS	Element management system
IABS	Integrated access billing system
IEX	Interconnect exchange carrier
ILEC	Incumbent local exchange carrier

ISDN	Integrated services digital networks
LEC	Local exchange carrier
LNP	Local number portability
LSR	Local service request
MSS	Marketing support system
NEL	Network element layer
NML	Network management layer
NPR	Network performance reporting
OSS	Operations support system
PSTN	Public switched telephone network
RBOC	Regional Bell Operating Company
SML	Service management layer
SDH	Synchronous data hierarchy
TBS	Telecom business solution
TNP	The new platform
TSM	Transmission status monitor
VoIP	Voice over IP
VPN	Virtual private network

References

ARL197 Arlitt, M., Williamson, C.: Internet Web Servers: Workload Characterization and Performance Implications, *IEEE/ACM Transactions on Networking,* Vol. 5, No. 5, October 1997.

BULO98 Bulow, D.: Dynamic Compute services, *Datacom,* 7/1998, p. 58–61, Bergheim, Germany.

CASE97 Case, J.: Finding the Right Job, www.nwfusion.com, April 21, 1997.

DETE95 DeTeBerkom: Intelligent Agents: Concepts, Architectures, and Applications, Part 2: Impact of IA Concepts on the telecommunications Environment, June 1995.

FORB97 Forbath, T.: Web-based Management: A recipe for success, *Network World,* May 5, 1997.

GARE97 Gareiss, R.: Casting the Web Over ATM, *Data Communications,* June 1997, p. 35–36.

GHET97 Ghetie, I.G.: *Networks and Systems Management — Platforms, Analysis and Evaluation,* Kluwer Academic Publishers, Boston, 1997.

HERM97 Herman, J.: Web-Based Net Management Is Coming, *Data Communications,* October 1997, p. 139–141.

HEYW97 Heywood, P.: An Impartial Interpreter of Service-Level Agreements, *Data Communications,* November 1997, p. 32–34.

HUNT96 Huntington-Lee, J., Terplan, K., Gibson, J.: *HP OpenView, McGraw-Hill Series on Computer Communications,* New York, 1996.

JAND96 Jander, M.: Distributed Net Management — In Search of Solutions, *Data Communications,* February 1996, p. 101–112.

LARS96 Larsen, A.K.: Mastering Distributed Domains via the Web, *Data Communications,* May 21, 1996, p. 36–38.

LEMA95 Lemay, L.L.: *Web publishing with HTML,* SAMS Publishing, Indianapolis, 1995.

MEGA97 Megandanz, T., Rohermel, K., Krause, S.: *Intelligent Agents: An Emerging Technology for Next Generation Telecommunications,* Research Paper with GMD Fokus, Berlin, Germany, 1997.

NAIR96 Nair, R., Hunt, D., Malis, A.: Robust Flow Control for Legacy Applications over Integrated Services ATM Networks, *Proceedings of Global Information Infrastructure, Evolution Internetworking Issues, Nara, Japan,* IOS Press, Amsterdam, 1996, p. 312–321.

POWE97A Powell, T.: The Power of the DOM, *InternetWeek,* p. 61–74, September 29, 1997.

POWE97B Powell, T.: An XML Primer, *InternetWeek,* p. 47–49, November 24, 1997.

REAR98 Reardon, M.: Need Management That Fast? *Data Communications,* 1998, p. 30–31.

ROBE98 Roberts, E.: Load balancing: On a different track, *Data Communications,* May 1998, p. 119–126.
RUBI98 Rubinson, T., Terplan, K.: CRC Press, Boca Raton, 1998.
SANT97 Santalesa, R.: Weaving the Web Fantastic — Review of authoring tools, *InternetWeek,* November 17, 1997, p. 73–87.
SPER98 Spero, S.: Analysis of HTTP Performance Problems, www.w3.org/Protocols/HTTP-NG/http-prob.html.
STEV94 Stevens, W.R.: *TCP/IP Illustrated,* Addison-Wesley, 1994.
TATE97 Tate, D.: Picking Through Piles of Web Pages, *LanTimes,* February 17, 1997, p. 30.
TERP94 Terplan, K.: *Benchmarking for effective network management,* McGraw-Hill, New York, 1994.
TERP96 Terplan, K.: *Effective Management of Local Area Networks, McGraw-Hill Series on Computer Communications,* New York, 1996.
TERP98 Terplan, K.: *Telecom Operations Management Solutions with NetExpert,* CRC Press, Boca Raton, 1998.
THAL98 Thaler, D.; Ravishankar, C.: Using Name-Based Mappings to Increase Hit Rates, *IEEE/ACM Transactions on Networking,* Vol. 6, No. 6, February 1998.

3.11 Performance Management of Intranets

Kornel Terplan

Abstract

After outlining generic and specific challanges of managing intranets, this presentation focuses on emerging new measurement and management tools, such as log file analyzers, traffic monitors, Web server managers, load distributors, and traffic shapers. The presentation ends with a discussion of integration opportunities of these new tools with existing management platforms and applications.

3.11.1 Introduction — Internet, Intranets, and Extranets

Intranet management means deploying and coordinating resources in order to design, plan, administer, analyze, operate, and expand intranets to meet service-level objectives at all times, at reasonable cost, and with optimal capacity of resources. Intranet management can utilize all the experiences collected over the last 25 years with managing data networks. Existing management concepts are still valid. Critical success factors are applicable as well. In managing intranets, those critical success factors include:

- Management processes that can be grouped around fault, configuration, performance, security, and accounting management
- Management tools that are responsible to support management processes and are usually assigned to human resources
- Human resources of the management team, with their skills and network management experiences

Intranet management instrumentation shows similarities with the management of other networks. The architecture is shown in Figure 3.11.1. The management framework is the center and is in charge of consolidating, processing, displaying, and distributing information to authorized persons. The framework is expected to be equipped with Web capabilities meeting the expectations of the majority of users. It means that views and reports are converted into HTML pages and are accessible from universal browsers. Management applications are a mix of well-known ones, such as trouble ticketing, asset management, and change management; and brand-new ones, dealing with log file analysis, load balancing, packet shaping, content authoring, and Web-server management.

The remaining part of this chapter addresses specific challenges of intranet management toward management processes.

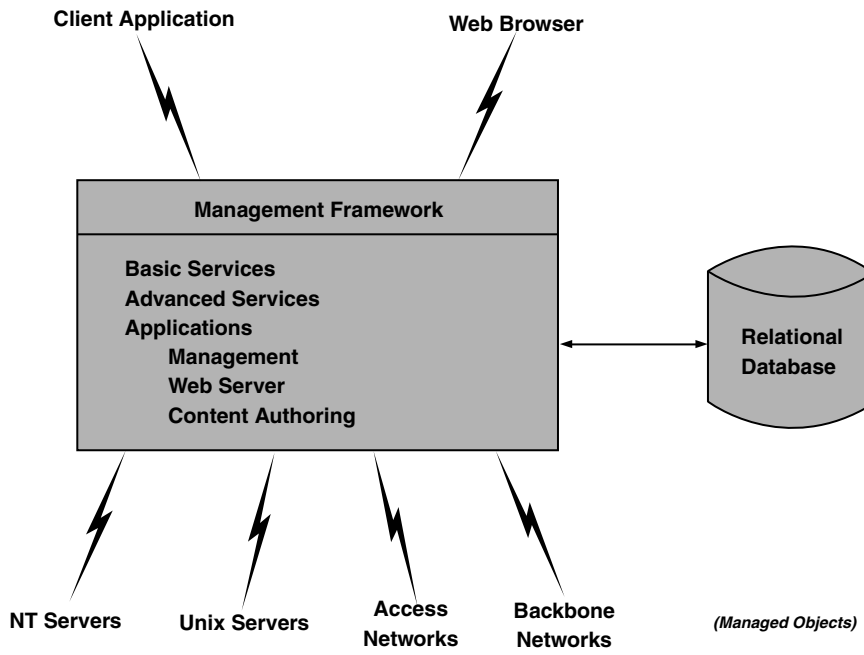

FIGURE 3.11.1 Intranet management framework.

The Internet is an existing network used by millions of people every single day. At the same time, the Internet is a generic term for a bundle of technologies available under the Internet umbrella. The Internet shows a number of similarities with the global phone system. Whoever is a subscriber can be reached by dialing the right country code, area code, and the actual phone number. In the case of the Internet, visitors type in the right universal resource locator (URL) to access the necessary information. Even the billing process shows similarities; the longer the talk or surfing, the higher the bill.

The ownership is not so clear with the Internet as with public phone systems. There are multiple owners of the Internet physical backbone, but they are hidden from users. Administration and management are getting more important as the number of subscribers is growing very fast. Just one administration issue — address management — causes a lot of headaches. Country institutions are coordinated by an independent U.S.-based company. Basically, the Internet can support multiple communication forms, such as voice, data, and video. The predominant use is still data.

This standardization is a threat to proprietary networking architectures, such as SNA from IBM. In order to support both, gateways are being deployed to interconnect both types with each other. It is very tempting to consider the Internet as the central switching point of corporate networking. But performance and security considerations drive corporate network managers to use privately owned Internet-like networking segments, called intranets. Intranet examples are shown in Figure 3.11.2; A, B, C, and D are communicating parties that use the intranet(s) offered by their own company.

Intranets are company-internal networks that are using Internet technology. In particular, Web technology is used for information distribution (e.g., company documentation will be unified this way, internal hiring procedures made visible, etc.) and Web protocols are used for internal information transfer. The backbone of intranets is based in IP on Layer 3. If interconnection is required to other networks, e.g., SNA or to other companies, then firewalls are deployed to protect the company-owned intranet. Firewalls are actually filters; certain packets without the necessary authorization code cannot pass the firewall.

If partnerships are the targets, networking equipment of partnering companies can be connected to each other. In such a case, the connected intranets are called extranets. Doing so, requirements toward firewalls are much lighter. Typical application cases are: car manufacturers and their suppliers of parts;

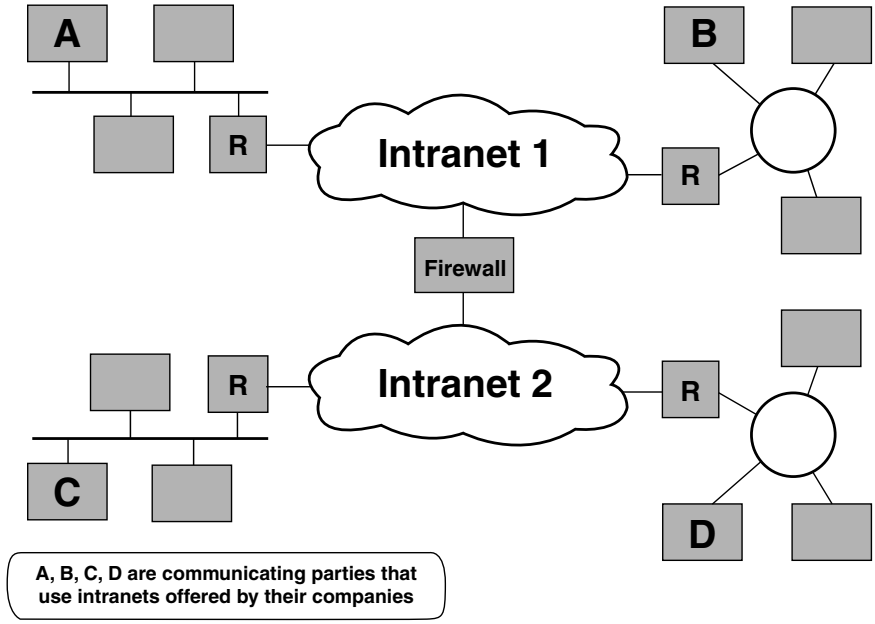

FIGURE 3.11.2 Use of intranets.

airlines in alliance; airlines and travel agencies; telcos with each other to complement local, long distance, and international services; and service providers and customers.

The Internet can still be utilized as part of intranets and extranets. Virtual private networks (VPN) are offering this by just securing channels that are part of Internet, to be used by communicating parties in intra- and extranets. There are a couple of technical solutions that are based either on Layer 2 or Layer 3 technologies.

3.11.2 Generic Intranet Management Challenges

This segment investigates how management functions can be reimplemented in intranets. Challanges will be highlighted in each functional area as well.

3.11.2.1 Performance Management Challenges

Feasible network architectures for intra- and extranets are shown in Figures 3.11.4 and 3.11.5. The components of the intranets are similar to other types of networks. Principal components include:

- Web servers that maintain home pages
- Web browsers that directly support users to view, download, and upload information to/from Web servers
- Backbone offering broader bandwidth for high data volumes
- Access network offering narrower bandwidth for lower data volumes
- Networking components including routers, switches, traffic shapers, and firewalls
- Communication protocols such as IP for the backbone, and higher layer protocols such as HTTP, SNMP, and FTP to support management applications

Figure 3.11.3 shows a typical arrangement in a simplified form. From the management perspective, all these components are managed objects. One additional managed object type must be considered; this object type is the application running in Web servers.

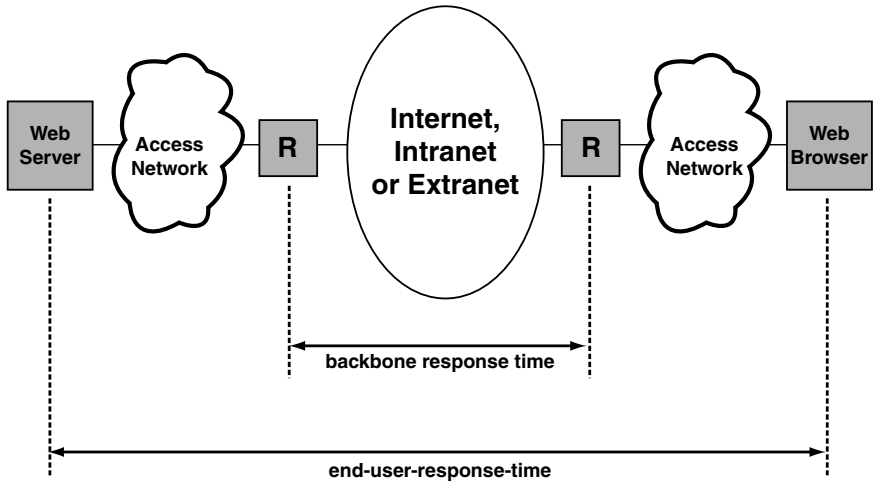

FIGURE 3.11.3 Principal structure of systems and networking components.

By early 1999, approximately 37 million people were accessing the Internet every single day. Altogether, approximately 830 million Web pages were being accessed every day.

Due to these special patterns, performance metrics are extremely important. From the technical viewpoint, everything can be measured. From the practical point of view, however, a few indicators are of prime interest. In particular, two of them are considered in every enterprise: response time and resource utilization.

For the response time, not only the resource-level, but also the user-level response time should be measured. Now, there are several types of tools to choose from: some of them measure throughput rates, some simulate network traffic and tally the results, some gauge performance by running within the applications themselves, and some rely on a combination of those techniques. Altogether, there are four approaches:

- Monitors or packet analyzers
- Synthetic workload tools
- Application agents
- Application Response Measurement (ARM) MIBs

End-user-level response time is helpful for service level agreements. Performance optimization needs more details about the contributors, such as the networks, systems, and applications. When segments of the response time are known, resource optimization by proper capacity planning is possible.

The utilization of resources has a direct impact on the response time. The payload is always an issue with resource utilization. Operating systems put load on the servers; control characters of protocols mean additional bytes to be transferred. Both represent overhead, but they cannot be avoided completely. The same is true with monitors and the transfer of monitored data for further processing. But overhead can be controlled, and then productive operations are not impacted. Further details on performance-related metrics in intranets are shown in other chapters.

In summary, tuning and optimizing intranets may be very different than traditional networks. User behavior, application performance, unusual traffic patterns, asynchronous resource demand, and additional protocols cause unique challanges to performance management of intranets.

3.11.2.2 Security Management Challenges

Due to opening networks, connecting partners, and using a public domain, such as the Internet, security risks increase considerably. VPNs are a possible answer to combine existing infrastucture with acceptable

protection. Security expectations may be different in various industries, but the generic security management procedures are identical or at least very similar. Security management enables intranet managers to protect sensitive information by:

- Limiting access to Web servers and network devices by users both inside and outside of enterprises
- Notifying the security manager of attempted or actual violations of security

Security management of intranets consists of:

- Identifying the sensitive information to be protected
- Finding the vulnerable access points to sensitive information
- Securing these access points
- Maintaining the secure access points

Identifying sensitive information means the classification of information. Most organizations have well-defined policies regarding what information qualifies as sensitive; often it includes financial, accounting, engineering, and employee information. But, in addition, there are environments that can have sensitive information unique to them. The main purpose of intranets is to improve the internal documentation and communication within enterprises. Web servers are the focal point of information maintenance. Evidently, not everything is for everyone. Depending on the individual responsibilities, access rights to information sources can be relatively easily structured and implemented. In summary, sensitive information are the home pages with particular content residing on Web servers.

Once the webmaster and network managers know what information is sensitive and where it is located, it must be found out how users can access it. This often time-consuming process will usually require that webmasters and network managers examine each piece of hardware and software offering a service to users. In this respect, intranets are not different from any other complex networks. Generic sensitive access points are (Figure 3.11.4):

- End-user devices, such as browsers
- Access and backbone networks
- Web servers maintaining sensitive information

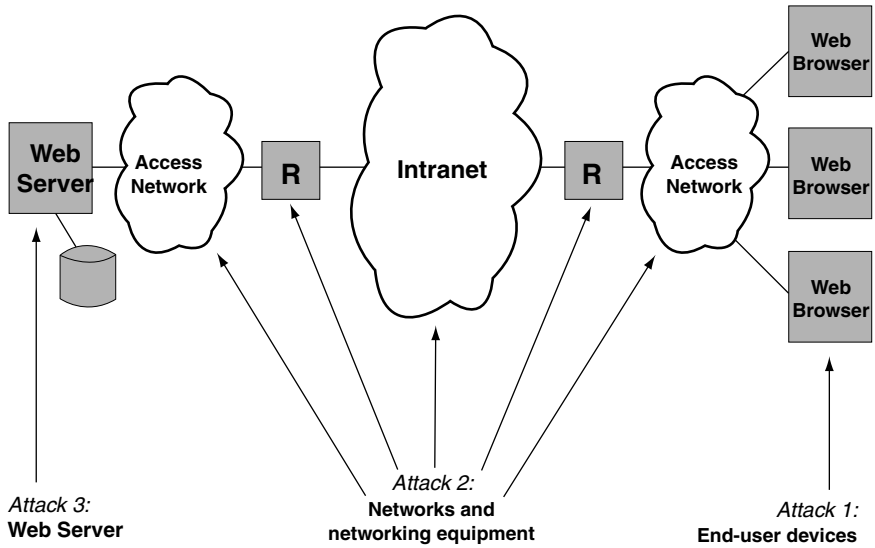

FIGURE 3.11.4 Access points with security risks.

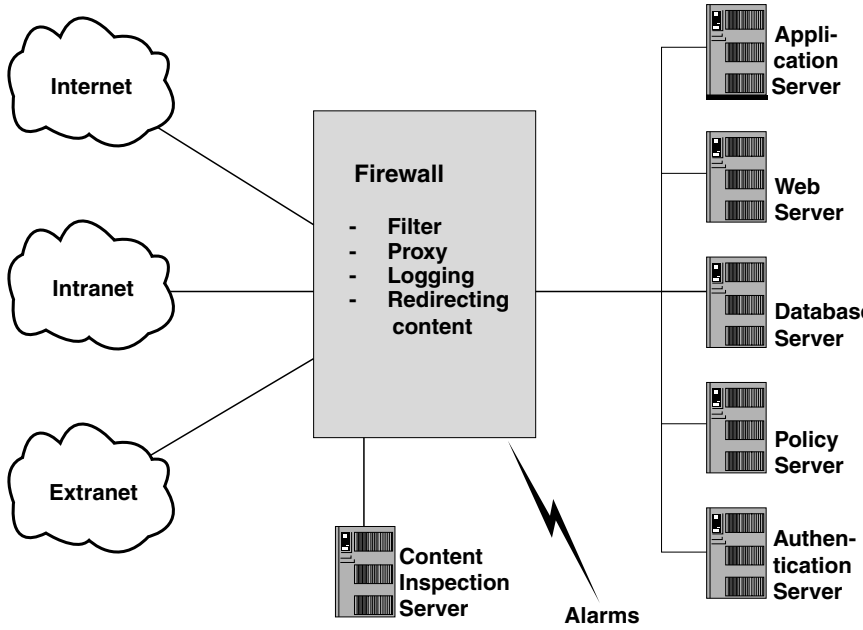

FIGURE 3.11.5 Firewall architecture.

The next step in security management is to apply the necessary security techniques. The sensitive access points dictate how the protection should be deployed using a combination of policies, procedures, and tools. In this respect, the following levels of security techniques must be considered:

- End-user devices, such as universal browsers (use of chip cards or chip keys)
- Access and backbone networks (use of encryption, authentication, and firewalls)
- Web servers (use of server protection, operating systems protection, special tools, and virus protection)

The last step in effectively securing access points in intranets is maintenance. The key to maintenance is locating potential or actual security breaches. It requires an ongoing effort of stress testing intranets, assigning tasks to outside professional security companies, reviewing case studies of security violations, and evaluating new security management techniques and tools.

Firewalls play a significant role in security management of intranets. A firewall (Figure 3.11.5) is a device that controls the flow of communication between internal and external networks, such as the Internet. A firewall serves several functions. First, it acts as a filter for inbound Internet traffic to the servers of enterprises. As a filter, the firewall prevents unnecessary network packets from reaching Web and application servers. Second, it provides proxy outbound connections to the Internet, maintaining authentication of the internal Internet users. Third, the firewall also logs traffic, providing an audit trail for usage reports and various planning purposes.

Firewalls are not without risks. Many companies assume that once they have installed a firewall, they have reduced all their network security risks. Typically, firewalls are difficult to penetrate, but when they are broken, the internal network is practically open to the intruder.

Furthermore, a firewall does not address internal network compromise. Approximately 70% of all network security breaches occur from within the corporation, that is, by persons already past a firewall. A modem dial-up established by the company or by an engineer for remote access is one easy way past a firewall. Also, misconfigured firewalls may cause problems. Firewalls are highly susceptible to human error. In a dynamically changing environment, system managers routinely reconfigure firewalls without

regard to security implications. Access control lists on a firewall can be numerous and confusing. Intranet managers should be sure that firewalls have been set up correctly and that they are performing well.

For intranets, a network-based intrusion detection system is required to protect the perimeter network from hacker attack. Network-based intrusion detection systems may be deployed as probes or agents running on servers. Probes are the most effective method at providing network-based intrusion detection. This probe minimizes the impact to existing systems because it is a passive listener reporting back to a centralized console without interruption. Intrusion detection will perform the following functions at the network device level:

- Inspection of data streams as they pass through the network, and identification/action on the signatures of unauthorized activity
- Activation of an alarm immediately upon detection of the event
- Notification of the appropriate security personnel, and triggering of an automated response to several issues to be considered.

In addition to intrusion detection, a TCP proxy aggregator may be considered. This will tighten security through the firewall by limiting the exposed ports. It also provides an offload for session/connection management and a more robust technical implementation in terms of port permutations supported.

Tunneling and encryption are used to deploy networks needing to appear point-to-point, but in fact consisting of various routes to an endpoint, providing data integrity and confidentiality. Usually, tunneling protocols, such as Layer 2 Tunneling Protocol (L2TP), Point-to-Point Tunelling Protocol (PPTP) and Internet Protocol Security (IPSec) and encryption standards such as DES, MD5, Triple DES, and others are used.

Mobile-code programs, such as Java and ActiveX, pose an increasing security threat. Content inspection software should:

- Provide full control over Java, ActiveX, and other mobile code activity in the corporation
- Prevent undetected, costly mobile code attacks, such as industrial espionage and data modification
- Enable safe Internet/intranet/extranet surfing while taking full advantage of Java and ActiveX technologies

A content inspection server will accept mobile contents redirected from a firewall in order to scan for attack signatures. If the scan detects a vulnerability, the contents will be blocked and the client prevented from downloading the mobile code. This denial will alert an appropriate administrator and notify the requesting client. If the scan does not detect any vulnerability, the mobile code is redirected to the firewall for routing to the client.

In summary, security management challenges increase in intranets due to many access points in the network. New techniques and new tools are required in combination.

3.11.2.3 Accounting Management Challenges

As far as the components of intranets are concerned, there are no differences to other types of networks and systems. But there are fundamental differences in terms of traffic patterns that may impact the right accounting strategies. Accounting management involves collecting data on resource usage in order to establish metrics, check thresholds, and finally bill users. Billing is a management decision, but usage measurements are a must in intranets. Principal steps of accounting are:

- Gathering data about the utilization of Web servers, the access and backbone networks
- Setting usage quotas as part of service level agreements with users
- Billing users for their use of resources

In order to gather data on usage, proper instrumentation is necessary. Standalone monitors, log file analyzers, and built-in accounting agents are most commonly used. Accounting management requires continuous measurements, but the amount of collected data is usually not critical in terms of overhead.

Service level agreements may include an expected level of resource utilization by single users or user groups. Either time duration or byte volumes may be agreed upon. Exceeding the agreed data volumes quota, the service and/or the price may change. The agreements and their continuous surveillance help to plan for the right amount of capacity.

Billing for intranet services is a new area, not yet well understood. Users are often billed based on one of the following:

- One-time installation fee and monthly fees
- Fees based on the amount of resources used

The first case is very straightforward. The user is billed for the installation of the intranet access and then a standard fee for each month of use. Using this method, accounting management is not necessary for billing. Although this is the easiest system to implement, it becomes difficult to justify why users with very different traffic patterns and volumes are billed for the same amount.

The second case is more difficult, and requires more engineering. Again, there are more alternatives, such as

- Billing is based on the total number of visits
- Billing is based on the total number of packets sent or received
- Billing is based on the total number of bytes sent or received

The accounting and billing cases are more complicated when multiple suppliers are present in intranets. If so, they must use a clearing house to gather usage data, allocate them to each other, and then generate convergent bills to the users. It is expected that the user receives just one bill for the intranet service.

In summary, the accounting management process can be fundamentally different in intranets in comparison to WANs and LANs of private enterprise networks. In particular, usage-based data collection and convergent billing are the real challenges to accounting management.

3.11.2.4 Configuration Management Challenges

Configuration management is the process of identifying systems and network components and of using that data to maintain the setup of all intranet resources.

Configuration management consists of the following steps:

- Identification of the current intranet environment
- Modifying that environment by moves, adds, and changes
- Maintaining an up-to-date inventory of the components, and generating various reports

Identification of the Current Intranet Environment

This process can be done manually by engineers or automatically by intranet management systems. Intranets don't require special treatment. This discovery and mapping step is identical with other networks and systems.

SNMP-oriented platforms offer configuration and topology services in two different ways; the discovery function identifies all managed objects with valid IP addresses on the LAN or across LANs; the mapping function goes one step further and displays the actual topology of the LAN or across LANs. Both functions can be successfully used for intranets. Managed objects without IP addresses are not discovered. The discovery and mapping processes need time and may impact production. Careful selection of the periodicity is required. Many companies deploy intranet visualization tools, instead of or in addition to discovery and mapping. Usually, they are very user friendly and easy to use, but they are independent from the actual network. Without synchronization of the tool's database with the actual network, these visualization tools are useless. But a combination of the discovery feature of the management platform with a visualization application can be very successful.

Modifying the Configuration Environment by Moves, Adds, and Changes

The intranet environment shows an over-average moves, adds, and changes (MAC) rate. Moves, adds, and changes are due to a user's move, restructuring buildings and infrastructures, deployment of new

applications and the usual equipment changes. In order to offer service to mobile users, the change rate is not even predictable. Modification would probably be manual if the data collection method were manual, and automatic, if the data collection method were automatic. It requires stable and well-implementable procedures. Intranets become a very important part of the IT-infrastructure, requiring high availability and good performance. The MAC window is narrowing with the requirements that MACs must be prepared very carefully. The requester is expected to fill in forms, detailing the nature of changes, their impacts on other managed objects, fall-back procedures, desired dates, its priority, and human resources requirements. Also, the MAC process should be carefully monitored. When problems occur, fallback procedures are expected to be triggered. After successfully completing the MAC process, all related files and databases must be updated accordingly.

Maintenance of the Configuration

Asset and inventory management is one of the critical success factors of intranet management. Usually, relational databases are used to store and maintain technical and financial data on systems and network components. Access is usually via SQL; reporting is supported by standard or additional third-party reporting tools.

Asset management is expected to work together with other management tools that are implemented in other management areas. In particular, the following links are obvious in managing intranets:

- Trouble ticketing and asset management
- Performance tuning and asset management
- Security violation traces and asset management
- Accounting details and asset management

In summary, managing the configurations of intranets does not introduce additional challenges to configuration management.

3.11.2.5 Fault Management Challenges

Fault management is the process of detecting, locating, isolating, diagnosing and correcting problems occurring in intranets. Fault management consists of the following steps (Figure 3.11.6):

- Detecting and locating problems: Intranet components generate a number of messages, events, and alarms. Meaningful filtering, combined with user input helps to detect abnormal operations. Management platforms and their management applications are usually able to determine the location of faults. This phase indicates that something is wrong.
- Determining the cause of the problem: Based upon information generated by element managers or correlation results provided by management platforms, the cause of the problem is being determined. This phase indicates what is wrong.
- Diagnosing the root cause of the problem: In-depth measurements, tests, and further correlating messages, events, and alarms will help to determine the root cause of problems. This phase indicates why the problem happened.
- Correcting the problem: Using various hardware and software techniques, managed objects are being repaired or replaced, and operations can return to normal. This phase indicates that the problem has been resolved.

In summary, managing faults in intranets does not introduce additional challenges to fault management.

3.11.3 Specific Challenges to Intranet Performance Management

The emergence of intranets is dramatically altering the way information is accessed within and outside the enterprise. Components of intranets, such as servers, networks, and browsers are known, and are individually well manageable. But their integrated management, as an intranet, generates several challenges to IT managers. Content, server, networks, and browser management are all critical success factors.

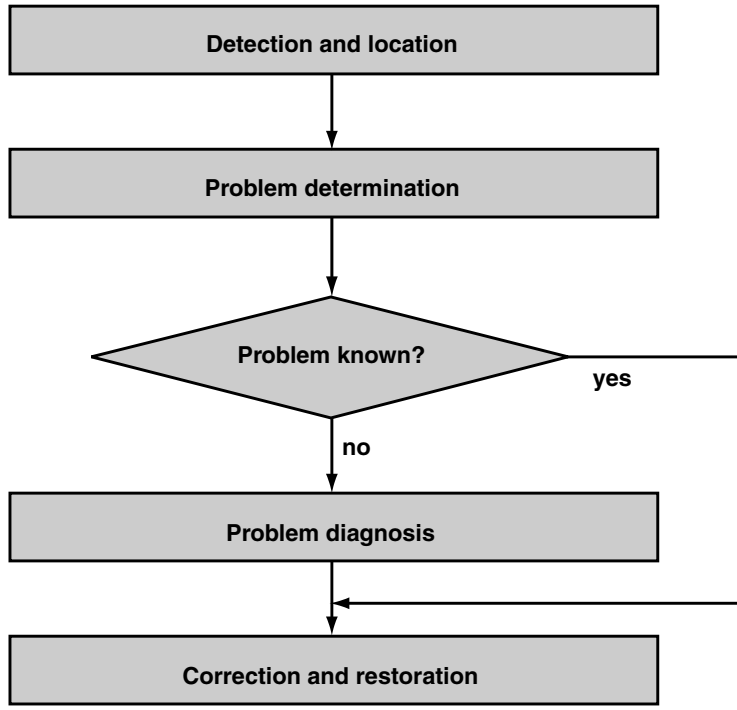

FIGURE 3.11.6 Fault management functions.

Not giving enough attention to any of them will cause IT managers to fail their intranets. Figure 3.11.7 shows the components of intranets.

The emergence of Web computing is dramatically altering the way information is accessed. The heavy use and popularity of the World Wide Web (WWW) is the most dramatic evidence. Looking at the enterprise, there is evidence that the Web browser has become the window of choice into corporate documentation and information. There are several important implications of this trend:

- All information can be viewed as Web content, accessible directly through a Web browser, a browser plug-in, or a dynamic piece of code, (e.g., Java) which is downloaded automatically to the client. This content can be as varied as a static Web page, a CGI script front-ending an existing database application, or new media such as streaming audio or video.
- The information access model has changed from one in which client-specific configuration is required in order to access information to one in which access is always available unless policies are explicitly defined to prevent it.
- Flash crowds, where certain content in the intranet generates significant unexpected traffic, are frequent observations, making traditional network design techniques based on measuring peak and average loads obsolete.
- Information accessed on or through Web servers comprise the bulk of traffic on the intranet (around 80%). Therefore, effective management of Web resources, bandwidth, and traffic is critical if acceptable quality of service (QoS) is required for Web-based computing.

3.11.3.1 Content Management

All information can be viewed as content. Structuring and arranging the content will finally decide about success and failure. Depending on the content for targeted visitors, page layouts may differ considerably. Not only the content of single pages, but also their links to each other have a great impact on visitor satisfaction. Individual visitors expect:

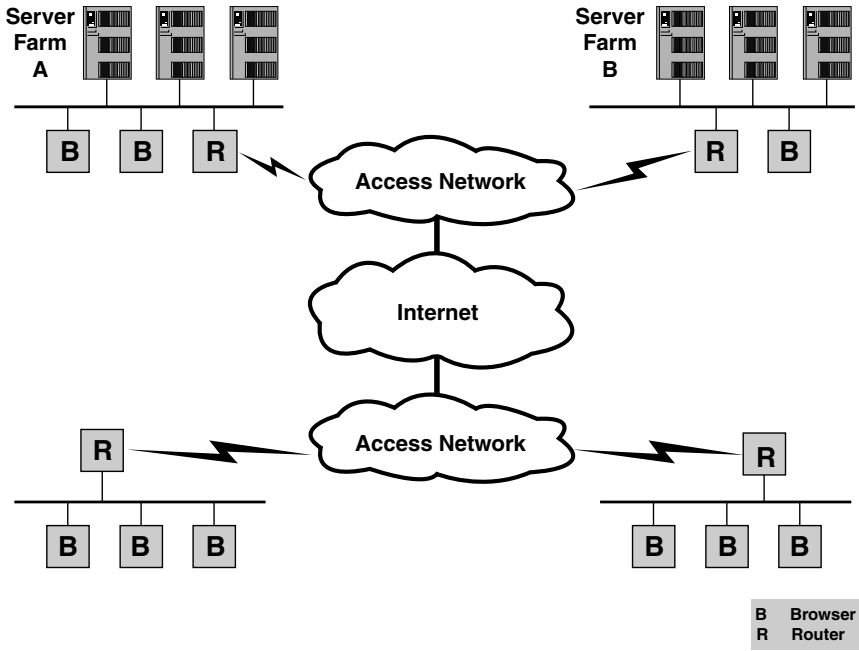

FIGURE 3.11.7 Components of intranets.

- Easy to read layout combining text and graphics
- Easy navigation between pages
- Easy return to the home page
- Rapid painting of pages
- Efficient links to interactive services
- Up-to-date status of pages
- Visualization of the site structure
- Site-wide change management of pages
- Easy ways of selecting pages to print or download

Goals and interests of companies offering information in home pages include:

- Rationalize information distribution to internal customers
- Fully meet content expectations of external visitors
- Manage intranet resources effectively
- Meet performance expectations of external visitors
- Meet business goals by using intranet technologies
- Provide the opportunity of deploying extranets to link business partners
- Meet high security standards
- Monitor visitor's behavior in order to make rapid changes to increase user satisfaction

Improvements in content management will have a great positive impact on overall performance. While Web server performance improvements are part of the performance optimization solution, they must be accompanied by improvements in network and content management technology to have a true impact on WWW scaling and performance. Specifically, developments in the following three areas are critically important:

- Content distribution and replication — By pushing content closer to the access points where users are located, backbone bandwidth requirements can be reduced and response time to the user can be improved. Content can be proactively replicated in the network under operator control or dynamically replicated by network elements. Caching servers are examples of network elements that can facilitate the dynamic replication of content. Other devices and models are likely to emerge over time.
- Content request distribution — When multiple instances of content exist in a network, the network elements must cooperate to direct a content request to the "best fit" server at any moment. This requires an increasing level of "content intelligence" in the network elements themselves.
- Content-driven Web farm resource measurement — A server or cache in a server farm ultimately services a specific content request. Local server, switching, and uplink bandwidth are precious resources which need to be carefully managed to provide appropriate service levels for Web traffic.

3.11.3.2 Web Server Management

Web traffic poses a significant number of challenges to existing Internet and intranet infrastructures. Most Web sessions are short-lived. As such, they have fewer TCP packets compared to batch mode operations such as file transfer. In addition, HTTP traffic tends to spike and fall radically. This creates instant demand for hot content that in turn causes network and server congestions. Web site traffic is highly mobile in that a unique event on a particular Web site could trigger a significantly high hit rate within a very short period of time. This would be typical in cases with periodic management report distribution and major systems and network outages.

Web traffic behavior is significantly different from today's client/server paradigm. It has the following unique characteristics:

- The amount of data sent from a server is significantly larger (5:1) than the amount of data sent from a client to a server. This suggests that optimization of server to client traffic has more significant impact to the intranet and that client request redirection to the best-fit server could have significant performance advantages for Web traffic flows.
- The median transfer size for Web server documents is small (e.g., 5 KB). This implies that Web flows are mostly short-lived flows. They are more likely to create instantaneous congestion due to their bursty nature. This suggests a resource management model must deal appropriately with short-lived flows. Even though HTTP supports persistent connections, due to interoperability issues with existing network caches, it is unclear how widespread deployment will be, or how soon.
- The top 10% of Web server files are accessed 90% of the time and are accountable for 90% of the bytes transferred. This suggests that Web server selection, caching, and content replication schemes that focus on this top 10% will yield the greatest gain.
- A significant percentage (e.g., 15–40%) of the files and bytes accessed are accessed only once. That is, some small number of large files often consumes a disproportionate amount of total server and network bandwidth. In addition, servers suffer performance degradation when subject to significant job size variation. This is due primarily to memory fragmentation, which occurs when buffering variable size data in fixed length blocks. Furthermore, subjecting servers to workloads consisting of both hot and one-time requests will result in lower performance due to frequent cache invalidation of the hot objects. Therefore, a server selection strategy that takes into account content, job size, and server cache coherency can significantly improve network and server resource allocation and performance. In addition, requests for large files may be good candidates for redirection to a server that has a shorter round-trip time to the client.
- Hosts on many networks access Web servers, but 10% of the networks are responsible for over 75% of this usage. This suggests that resource management strategies that focus on specific client populations may yield positive results in some cases.

Real-time traffic is becoming an increasingly significant proportion of Web traffic. Web site resource management strategies must take into account an increasing demand for support of real-time applications

such as voice, distance learning, and streaming media. To deal with both legacy and Web traffic as well as real-time Web traffic, these strategies will need to include admission control as well as bandwidth and buffer allocation components.

The hardware of Web servers is practically the same seen with other servers. The software is divided in most cases between Unix and NT; industry analysts expect a clear shift toward NT for price reasons in the future. Web server sizing should follow generic guidelines, and also criteria specified by analyzing Web traffic patterns. If resource demand is higher than server capacity, multiple servers can be put together into server farms. This solution may satisfy the resource demand criteria, but requires careful attention of allocation and flow control.

3.11.3.2.1 Content Smart Quality of Service (QoS) and Resource Management

In a typical Web site, the top 10% of Web server files are accessed 90% of the time and are accountable for 90% of the bytes transferred. Therefore, techniques that optimize performance for these files will have the most significant impact on total Web site performance. This requires that the network itself be aware of which content is hot and which servers can provide it. Since content can be hot one instant and cold the next, content-smart switches must learn about hot content by tracking content access history as it processes content requests and responses.

To effectively manage Web site servers, network, and bandwidth resources, something must also be known about the content size and quality of service requirements. These content attributes can be gleaned through the processing of active flows, through proactively probing servers, or through administrative definitions. In addition, it is important to track server performance relative to specific pieces of content. All of this information can be maintained in a content database that provides an analogous function to a routing table in a router or switch. Content-smart switches make a content routing decision based on the information contained in the database, connecting a client to a best fit server in either a local or remote server farm. This enables the emergence of a business model based on replicating content in distributed data centers, with overflow content delivery capacity and backup in the case of a partial communications failure. Additionally, overflow content capacity intelligence minimizes the need to build out to handle flash crowds for highly requested content.

3.11.3.2.2 Content Smart Flow Admission Control

Two factors often contribute to congestion in a server farm. One is that servers are not up to the task of handling the amount of incoming traffic. The other is that the link bandwidth from servers to the Internet is overwhelmed by the combination of inbound and outbound traffic; this is complicated by the fact that the amount of outbound traffic from servers is on average about 5 times that of the inbound. As a result, a TCP/HTTP connection could be made successfully only to find out that the server could not be allocated the necessary bandwidth to deliver the requested content. To make matters worse, some server implementations come to a grinding halt when presented with an excessive number of TCP/HTTP connections — sometimes requiring a hard reboot.

3.11.3.3 Load Distribution and Balancing

In order to satisfy the high performance expectations of site visitors, bandwidth in backbone and in access networks should be managed effectively. Usually, servers are consolidated into server farms that are using the infrastructure of LANs. It is very unlikely that the LAN causes any bottlenecks. Larger enterprises may use multiple server farms deployed at various locations. In order to optimize content allocations, traffic and page references should be monitored and evaluated. At different locations in the network, hardware and software are expected to be installed that intelligently analyze the requests and direct the traffic to the right destination. The right destination could be threefold:

- Server farm destination with the requested content
- Server farm destination with the lightest load
- Server farm destination with the closest location to the visitor

There cannot be any compromise on item 1, but there could be a trade-off between 2 and 3, depending on the networking traffic.

The emergence of Web computing and Web traffic over the Internet or intranets has created some unique new problems. It is estimated that over 80% of Internet traffic is related to Web-based HTTP traffic. Even applications such as FTP and RealAudio, which run over TCP and UDP, respectively, typically use HTTP to set up the transfer. Since HTTP is an application protocol that runs over TCP, LAN switches and routers, which run Layers 2, 3, and 4, have very limited ability to influence Web traffic behavior. This burden is left to Web servers, which take on the function of TCP/HTTP connection management and, in some cases, the responsibility to distribute HTTP requests to servers within a server farm. This creates inevitable scaling problems as Web sites grow.

The current Internet can be described by using a model where local bandwidth is plentiful in the premise LAN located at the edge of the Internet. However, the uplink from LAN or remote user to the Internet is often severely bandwidth constrained by orders of magnitude. Although congestion can occur anywhere in the Internet path between a client and a server, the most frequent culprits are the WAN connection between the client and the Internet and the WAN connection between the Web farm and the Internet. Actions taken to ensure that this bandwidth is not overcommitted will help improve end-to-end performance.

Instantaneous bandwidth mismatches can occur for a network device that functions as the demarcation point between the public Internet and the Web farm. Examples are:

- The incoming link of the traffic is a faster media type (e.g., fast Ethernet) and the outgoing link is a slower type (e.g., T1 or T3).
- The instantaneous fan-in, i.e., the number of flows being sent at the same time to the same output port, can vary dynamically from one instant to the next.
- A number of traffic sources (e.g., outbound server traffic) may be sharing the bandwidth of a 45 Mbps T3 pipe in a bursty manner over a very high-speed switching fabric (e.g., 10 Gbps). This creates a need to regulate flow admission into a slower pipe from multiple higher speed traffic sources.

Information about the use of Web pages, their users, the frequency of access, resource utilization, and traffic volumes can also be collected in the network or at the interfaces of the network. In many cases, the borders between tools and techniques in the server and networking segments are not clear. Tools are different from each other; the differentiators are data collection technologies, performance metrics used, and reports offered.

In the Internet and intranet area, effective bandwidth management is a critical success factor. The role of network planners is going to be redefined. Real-time and near-real-time bandwidth allocation definitions are needed. Network managers agree that load balancers are needed.

There is little progress in standardizing on load distribution performance metrics. But the following few metrics can be successfully used:

- Number of referrals to server farms
- Number of lost requests due to load situations
- Number of requests with unacceptable response time
- Number of broken connections due to network problems

3.11.3.3.1 Content Smart Link Management

This technique can ensure that more flows are not admitted than can be handled through the switch or on the uplinks on average. It is still critical, however, to deal appropriately with traffic bursts and temporary congestion on these links to ensure that Web flows get the appropriate quality of service. Priority queuing provides a way to prioritize requests based on their type precedence. Fair queuing and weighted queuing methods improve over the priority queuing scheme by addressing the low priority

traffic starvation problem with a scheme that separates traffic into well-identified flows so that each receives a "fair" or "weighted fair" share of transmission bandwidth.

Class based queuing (CBQ) was developed by the Network Research Group at Lawrence Berkeley Laboratory, as an improvement upon these existing bandwidth management techniques. It proposes a model which traffic is categorized in hierarchical classes. Flows inherit their flow characteristics from their parent flow class tree and can have local characteristics of their own. Flows are identified based on the IP address and the inner attributes within the IP header and payload. CBQ provides more granular control of transmission bandwidth and distributes it to member flow classes in accordance with their allocation policies. The model itself is independent of the scheduling techniques that run underneath it, therefore implementation details will vary based on the target architecture.

Content smart link management borrows concepts from CBQ. However, where CBQ is a model which operates on a packet-by-packet basis based on Layer 3 and 4 classification techniques, content smart link management classifies flows at admission time based upon the content requested, its attributes, and configured policies. These policies support the enterprise and service provider service models described in an earlier section of this chapter. This facilitates the classification of flows in a two-level hierarchy which includes owners (or customers) and content. Actual scheduling of flows is managed by a hardware-based flow scheduler which supports guaranteed bandwidth flows, prioritized/weighted flows, and best effort flows. Hardware-based scheduling is critical in order to scale the Web farm.

3.11.3.3.2 Content Smart Load Balancing

Simple load balancing techniques such as round robin, weighted round robin, and least connections are inadequate for Web traffic. For example, Web traffic load balancers must support "sticky" connections, which allow a particular server to be selected regardless of server load due to content locality or transaction integrity. Because of the disproportionate ratio of hot content files to total content (1:10), it is highly desirable to support a content replication model that does not require that content be equally and fully mirrored among servers in a server farm. This means a load balancing technique must be intelligent enough to recognize if content is available on a particular server before making the selection decision.

Content smart load balancing takes into account several factors that have a significant impact on the overall performance and cost of a Web server farm:

- Server cache hit rate — By directing requests for hot content to a server that has recently serviced that content, this technique ensures that cache hit rate, reducing disk access latency for the most frequently accessed content. Since a significant percentage (15–40%) of the files are accessed only once and 90% of the files are accessed only once or not accessed at all, it is important to keep those infrequently accessed files from thrashing a server cache. That is, an infrequently accessed file should be invalidated in server cache promptly to increase the chances that a more frequently accessed file can remain in cache.
- Burst distribution — Short-lived, bursty flows can best be handled by distributing them among eligible servers so long as the servers have been performing below a defined threshold for a period of time.
- Web flow duration — Most Web flows are short lived. However, a relatively small number of infrequent, long-lived flows have a far significant impact on overall bandwidth and server resource consumption. For that reason, long-lived flows should be separated from short-lived flows from a load balancing perspective and short-lived flows of similar QoS requirements should be aggregated to increase TCP flow intensity and reduce per flow resource allocation overheads.
- Content biased server performance measurement — Current server loading can best be measured by examining the request/response time interval of a server as it handles requests. This measurement is most accurate when connection between the switch and the server is direct. In addition, server performance is not uniform across all content. For example, computer intensive applications may perform better on one server than another. Other servers may perform better for other types of content. Server performance information needs to be qualified by content.

In the Internet and intranet area, effective bandwidth management is a critical success factor. The role of network planners is going to be redefined. Real-time and near-real-time bandwidth allocation definitions are needed. Network managers agree that load balancers are needed. The decisions are if:

- Hardware or software-based load balancers are better
- Embedded or standalone solutions should be preferenced
- Use of the combination of both

In the first case, considering high traffic volumes, hardware solutions should be preferred. Software solutions in critical load situations may slow down processes, and risk performance. At this time, there are no accurate guidelines for tolerable workload, but a range up to 5% seems to be reasonable.

Switches, routers, and firewalls are almost everywhere in Internet access networks and in intranets. To embed traffic control and sharing functions would save extra components, but would — as stated earlier — generate additional load and may impair the principal functions. The embedded solution may also include the use of RMON capabilities for real-time load profiling. The standalone solution is sensitive against single point of failure, but would offer an overhead-free traffic and load management. The following attributes may play an important role when evaluating alternatives:

Use of Load Balancing Switches

Benefits:

- Load balancing is performed in a device that is needed anyway in the network
- Centralized management
- Good opportunity to control and guarantee QoS

Disadvantages:

- Performance may be impacted by management functions
- Single point of failure for both switch and management functions

Use of Load-balancing Firewall

Benefits:

- Load balancing is performed in a device that is needed anyway in most networks
- Centralized management
- Includes special functions and services, such as traffic management and application-based load balancing

Disadvantages:

- Switches are still needed
- Single point of failure for both firewall and management functions
- Performance depends on hardware and operating system configuration

Use of Load-balancing Traffic Shapers (Figure 3.11.8)

Benefits:

- Load balancing is performed by a device most likely present in the networks anyway
- Centralized management
- Offers traffic shaping and balancing for Internet or intranet access in addition to server access

Disadvantages:

- In most cases, switches and firewalls are needed in addition to these devices
- Single point of failure for both traffic shaping and load balancing
- Little experience yet with performance and scalability

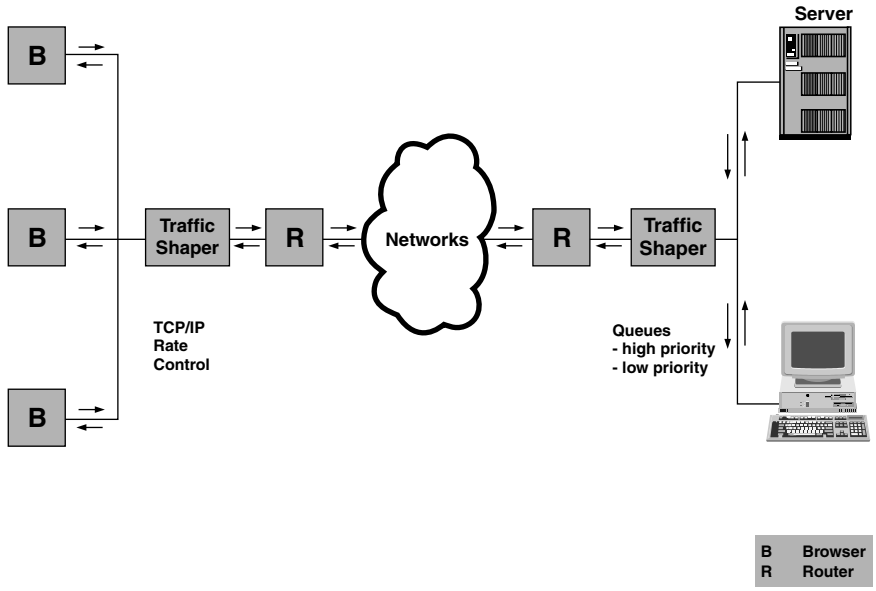

FIGURE 3.11.8 Load balancing packet shapers.

3.11.3.4 Technologies of Access Networks

There are many alternatives of designing and deploying access networks. The basis technology chosen has a significant impact on overall performance. Table 3.11.1 evaluates the most important technological choices according to criteria, such as suitability, maturity, scalability, distance limitations, and costs.

The right choice of access network technology must be seen in connection with content smart control of the bandwidth provided in the access networks.

TABLE 3.11.1 Comparison of Technologies for Access Networks

Criteria	T	ISDN	Frame	ATM	Cable	xDSL
Suitability	Medium	Good	Good	Excellent	Excellent	Good
Maturity	High	High	High	Medium	Low	Low
Scalability	Good	Medium	Mefium	Excellent	Medium	Good
Distance limitations	None	None	None	None	Some	High
Costs	High	Low	Medium	High	Low	Low

3.11.4 Content Management

Authoring tools present a standalone environment in which to build pages. While this requires learning a new program specifically for HTML/XML creation, these tools allow users to make the most of HTML/XML, using features that traditional word processors do not support.

Currently, there are two distinct kinds of tools Web authors can use to bring their words to the Web. Tag-based tools automate HTML/XML syntax, allowing users to see and tweak tags without having to enter their syntax manually. In contrast, WYSIWYG tools hide HTML/XML from the user, generating it in the background instead. If these tools do not support a specific feature of HTML/XML, that feature must be added manually after the document's underlying code is visible, usually in a text editor. Some products use dialog boxes or palettes to accept information before displaying it as HTML/XML code in the body of the document. Since these tools generate HTML/XML for users, they minimize the learning curve for new Web authors and can produce syntactically perfect HTML/XML. Many of the publicly available tools have both standard and professional features, the latter being available only in the registered or commercial version.

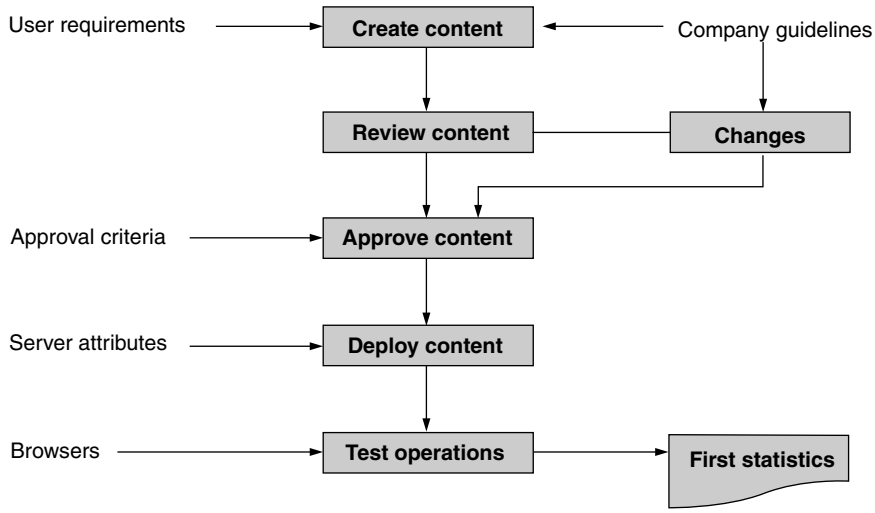

FIGURE 3.11.9 Process of content authoring and management.

3.11.4.1 Design of Home Pages — Content Authoring and Deployment

Most users are challenged by the task of information creation, management, and dissemilation. These activities are time consuming and difficult to control. The Internet and intranets alone cannot solve information management problems unless specific intranet solutions are implemented that directly address the need for document management. The new discipline, called content authoring and deploying, includes the following tasks:

- Creating content
- Reviewing content
- Approving content
- Changing content
- Deploying content

Figure 3.11.9 shows the process of creating, reviewing, changing, enhancing, approving, and deploying home pages.

The prerequisites to successfully execute these tasks are:

- Users must be able to easily add and update content on a periodic basis.
- Users must be able to protect their page contents from changes by other users.
- A content approval process should be defined and in place. This process should emcompass ways to manage and control document revisions, especially changes to shared documents.

As policies and procedures relating to content management are formulated, it is important to designate responsibilities to specific individuals to ensure that they are properly implemented and followed. An internal style guide should be developed that provides page layout, design elements, and HTML/XML code guidelines. Usually, case tools are also involved. The style guide will help the users to maintain a consistent look and feel throughout the Web pages. Sometimes television-like techniques are helpful in this respect. The style guide should contain information on where to obtain standard icons, buttons, video, and graphics, as well as guidelines on page dimensions and how to link the pages to each other. As part of the style guide, it is helpful to create Web page templates. These templates consist of HTML/XML files, and are used to provide a starting point for anyone interested in developing Web pages or content for the intranet. Although it is very easy to create a working Web page and to publish for mass viewing, the real challenges are:

- To maintain the page
- To size the Web server
- To configure the access network

3.11.4.1.1 Site Design Considerations

Content authoring includes a number of tasks. The most important tasks are:

Determine the right page layout, including:

- How to structure a Web site
- How to lay out a Web page
- Ideas for improving usability
- Technical hints to increase display speed
- Collection of examples of well-designed sites that can be used as models
- Consideration of new Web technologies for site design

One of the principal factors in the design of a good Web page is knowledge and understanding of the motivations and goals of the target user as well as the technical platform on which they operate. Given the varying levels of user knowledge and the infinite number of ways a Web page can be constructed, this understanding is essential to creating a usable, effective Web site. Before beginning the design, a user and task-centered analysis should be completed to gain knowledge about the target users and their goals. Important questions are:

- To whom will the page be available?
- What are the business drivers for the site — e.g., to provide information, to collect data, to market products?
- Who are the users — e.g., professional "knowledge workers" or casual intranet users?
- How will a typical user access the page — e.g., fast connection or dial-up?
- What browser will they use?
- What are the most frequent tasks that users perform?

Answers to these questions will provide the necessary background information for the navigational structure of the site. During site design, designers should keep in mind that if the user cannot quickly find what they are looking for and are not engaged by the layout and information contained within the site, they are likely to move on.

Site Registration

The purpose of site registration is to establish content ownership and to facilitate navigation. Through the site registration process, sites are added to the intranet directory and become accessible via the intranet-wide search facility. The Web site is defined as a collection of related Web pages. Typically, a Web site is an administrative unit.

Site Navigation

There are two points to consider when constructing the navigation layout for a Web site — namely, the structure of the information and how access to that information will be provided. First, the layout of the site is usually the most difficult part of the site design process, particularly if a lot of information will be accessible from the site. Adequate time must be put into designing the structure of the information to allow easy access for all users. Second, navigation tools must be clear and easy to use as well as functional within all types of browsers that will be used by the target audience. Navigational design must consider all of the same factors as many other GUI interfaces. Since movement within a Web site is typically non-linear, navigational menus should be planned to allow users quick access to any part of the site.

Content Organization by Menus

A user's ability to move through a Web site and find the information or functions they are searching for plays an important role in determining a site's success. Menus and sub-menus are powerful tools in the design of

a Web site. In the same way that menus are used in traditional Windows-based design, HTML menus can be used to subdivide and group relevant content to allow the user to be guided to their topic of interest gradually. The use of more than four levels of menus forces a user to work too hard to find the information they are looking for. Using too few levels may be equally difficult to navigate, in particular, when the information volumes grow. Three to four levels should generally provide appropriate depth and guidance for the user. However, because of the varying content of sites, this is a flexible guideline. It is important to know that the menu structure for the site should be continually evaluated and improved as the site grows.

Interaction Models

There are many ways to organize information contained within a Web site. The term "interaction model" refers to the structure that is implemented to allow the user access to the various pages within a site. The type of model best suited to a particular page will depend on the content and complexity of the information that the page presents. There are a number of interaction models in use. These models may be used independently or in combination throughout a site. These models are:

- Table of contents — This approach is taken from printed books. Users can easily find the headings they are looking for, and then hyperlink directly to that page. This type of access is useful for sites that provide textual or encyclopedic information.
- Image maps — They are graphics that use an embedded linkage map that relates hot spots on the graphic to URLs within the Web site. In this way, the user can view the graphic and point and click to move to different locations on the site.
- Graphic menus — They provide the same visual approach to site navigation as image maps, without incurring the disadvantages of employing one single large graphic, mapped with links. They employ smaller, simpler graphics, strategically placed to provide visual impact.
- Search — Web site searches provide a useful means of allowing a user to access information contained on a particular Web site. Some form of search facility is usually a requirement for larger sites.
- Indexing — It provides functionality similar to book indices. It allows a user to rapidly locate information pertaining to a specific keyword or topic. It may be used in combination with search.

3.11.4.1.2 Page Design Considerations

The actual layout of a Web page is highly dependent on the type of information that is being presented. This segment provides some fundamentals of good page design.

Header — The Header provides a user with access to commonly used functions within the company-wide intranet and clearly differentiates intranet content from Internet content. The standard header provides links for navigation to common functions via the following graphics:

- Company logo — Links to the company's home page
- Directory — Links to the company's intranet directory Web site
- Services — Links to the company's intranet service page
- Search — Links to the company's search Web site
- Help — Links to the company's intranet help Web site

Preimaged mapped versions of the company's header are available on the intranet development and support site.

Footer — The footer gives the user important information about the page and provides consistency within the company's intranet. The standard footer usually contains the following:

- A standard horizontal rule as a separator
- Copyright statement
- Statement regarding content ownership with an optional e-mail link to the designated page maintainer; not supposed to be a name of an individual
- Date of the last revision

Page size — Page size must be designed with the actual usable space of the browser window in mind. Typically, this would be the lowest amount of useable space for the standard browser configuration in a 640 × 480 video monitor resolution. When designing a Web page, designers want to limit horizontal scrolling as much as possible. Keeping the width of Web sites less than 600 pixels (using tables) makes it much easier for users to navigate information. In some cases, horizontal scrolling is normal and acceptable.

The acceptable size for an intranet page is 100,000 bytes or less. This limit includes all of the images that are embedded on this page. This size will keep performance within acceptable limits for both LAN, WAN, and dial-up users with 28.8 Kbps modems.

Home page — The layout and design of the "home page" of any Web site is extremely important. Besides being the first thing a user sees upon entering a site, it defines the organizational structure and tone for the entire site. Some essential elements for every home page include:

- Visually appealing design
- Overview of site content
- Links to contents of site
- Company/organization identifying information

Page layout — HTML does not provide graphic designers the flexibility they are accustomed to in existing page layout and editing programs (e.g., MS Word, Adobe PageMaker). However, this does not mean that complex and functional applications cannot be created using HTML. Rather, one must realize that, when used inconsistently, the graphic and typographic controls of HTML can result in inconsistent designs. To avoid the haphazard look of documents, designers should take care in how graphics are placed and organized. A consistent style will also allow for a consistent conversion from non-HTML documents. It is better to use simple icons and images, instead of complex ones. Navigation should be kept in a consistent place.

Text style — Text needs to be short and to the point. Text should be organized in sections of a paragraph. When browsing, visitors tend to scan rathen than read. They are usually searching for information and appreciate when sections are arranged in logical order. Similar ideas or facts should be presented in a consistent way, with the same components presented in the same way in the same order. Consistency is a very important consideration in Web design.

Graphics — Graphics images should be used where appropriate to help the user navigate and find information more quickly. Graphics also provide a "look" to the site that will help the user to identify where they are. Graphics should not be overused for internal publishing applications. Whereas external marketing Web sites often are graphically intense to catch attention, use of graphics in internal Web sites should be based on ease of navigation and usage. The type, sizing, and location of graphics throughout a site should be presented in a consistent manner, items of similar importance should have the same size and type of graphic. If a larger-than-normal graphic is used, the user is likely to assume that there is some additional significance. Often, the visibility and intended use of the site will dictate the level of graphics required for the site. Graphic images should be designed for a 256 color envisonment. A common mistake that professional graphic designers make is designing with higher resolutions and greater color depth than the deployment environment. The color scheme that was designed in 16-bit color may look bad in 256 color or even worse in 16 color environments. Design should follow the requirements of the target environment. Most images are between 10 and 30 K. The exception would be image maps on navigational pages or photographic images, which should be around 50 K. One of the drawbacks with using images on a network is the time it takes to download very large files. Images must be kept as small as possible and fit within the size of the browser's viewable space. For image formats, file formats are the best to be used. GIF and JPEG are both compressed formats. GIF format is better for smaller graphic or line art images.

Local navigation elements — Each Web site should include a sitemap, showing a detailed layout of this site with links to all possible sections and documents. Each page within a Web site should include a link to the sitemap page. Users may link to a Web site or Web page from a number of different places (navigation page, search results page, hyperlinks, etc.). The sitemap page gives the user a quick and easy way to locate the information they need. On long pages, the user may want to quickly go to the top of

the page to view the table of contents or other introductory information. The "top of the page" icon helps users more quickly navigate to the top of the current page.

Links — While many Web sites incorporate graphics to support navigation, text links still play an important role in ensuring the usability of a site. Working with text in HTML is easy. In general, because it is easy to create links and change font types, there are several mistakes commonly made. Several guidelines that aid in ensuring a site's readability and usability are listed below:

- Design for scanner, not for readers
- Explain the page's benefit above the "fold"
- Bold typeface will draw attention to a particular section
- Avoid typing in all caps — it is more difficult to read
- Links must be underlined in addition to being colored to assist users who may be colorblind or using black and white monitors
- Avoid blinking text because it is difficult to read and annoying to users

A typical Web page provides both informational text and links to more specific information. Most people are looking for visual clues to whether a page is useful or interesting enough to be worth reading. If they don't find what they want quickly, they will move to another site. One of the difficulties in using text for navigational purposes is the wording of the links. Proper wording of the text allows the user to jump to a new topic or continue reading without losing their place. All links to default pages should be set with a trailing "l." This eliminates the problem of DNS names turning into IP addresses. By default, the Web browser converts any hyperlink that does not include a Web page (such as a link to a home page) to the default page for the server. However, depending on the browser, this may convert the DNS name into the physical IP address of the hosting server. If the DNS name is converted to an IP address and the user adds the page to their favorites, the URL will be stored with the IP address. If the IP address of the site changes, the bookmark will no longer work. To eliminate this problem, simply include a trailing "/" on any link that does not include a page file name. Abstracts and summaries are very helpful for large pages or large graphics. Whenever possible, users should have the opportunity of linking to further information if desired. Very large files or files which are not in a useable browser format (e.g., ZIP files, BMP files, etc.) should have a link which allows the user to download the file to their local PC.

Other graphic elements — Separators are graphic or possible textual elements that are used to break up or visually divide the contents of a single Web page. Separators can be as simple as a horizontal line to a shadowed line graphic or an actual image file. Their use helps to visually discern varying subject matter on the page. While separators can be effective, it is important to remember that separators should not distract the user from page content; rather, their purpose is to divide the information into logical groupings. HTML provides tags for standard information-gathering controls like radio buttons, drop-down menus, and exit boxes. In general, guidelines created for traditional GUI-based development apply to Web page design. Important remarks are:

- In most countries, the eye moves from left to right when reading, so text literals should be left-aligned.
- Exit boxes should be similarly sized and also left-aligned.
- Tabs should move the user downword through the page.
- Controls should be evenly spaced and aligned when possible.
- A default button should be provided.
- Mixed case text should always be used.

Bullets are used in HTML in the same manner they are used in traditional word processing to define a list of items. While textual bullets are fine for use on Web pages, there are also many available graphic bullets that will add just a touch of color to an ordinary Web page.

Background and text colors — The use of appealing backgrounds and text colors can add an artistic look to Web sites, but the way colors are used also affects the usability of the site. Designers must be

wary of improperly using color, as colors may have different meanings to different people and some users may be unable to distinguish some colors clearly.

Some user interface guidelines that are applicable to Web sites include:

- Color is second only to movement in attracting attention
- Three colors are sufficient for a color scheme
- Specific colors should be used carefully
- Shades of red attract attention, while the retina responds to yellow the fastest
- Blue is more difficult to focus on, making unsaturated blue a good choice for backgrounds
- Gaudy, unpleasant colors, and combinations of red/green, blue/yellow, green/blue, and red/blue should be avoided
- If backgrounds are going to be used, they should be either a light-colored pattern or a solid color

Printing — When the nature of a site is documentation, users must have the ability to print individual Web pages or an entire site's content. This can easily be accomplished by adding a link to a printable form of the entire document. Documents may also be provided in multiple formats such as Microsoft Office formats to accomodate the maximum number of users.

3.11.4.2 Issues with Content Authoring

The recommendations for Web page design can be summarized as follows:

- Use of standards for the layout of pages
- Standardize links
- Use one or more interaction models, such as table of contents, image maps, graphic menus, search, indexing
- Use of navigation assistance
- Segment long documents into small ones
- If a site includes a significant amount of pages or data, a local search page should be provided to search the content only
- Design pages for rapid and slow searches alike
- Use text pages for users with narrow bandwidth
- Test HTML pages and links before practical use
- Test content on different browsers
- Use recommended templates from the webmaster to create new pages
- Use abstracts or summaries for larger text pages or large images and give the users the option to link to detailed information if desired
- Provide a link to download a concatenated file of a series of Web pages so that a user can print an entire document rather than printing multiple web pages
- Use backgrounds carefully; make sure that users can easily read the text of a page if a background is used
- GIFs should be used for small graphics where there are a limited number of colors and JPEGs should be used for photographic images
- If users link to another site, the site owner of the link must be informed; this will enable the site owner to notify everybody involved if the link changes

But the users should consider the following facts:

- Too much graphic and animation content slows down operations
- Big pictures slow down loading the pages
- Copyright of graphics should be granted

- Proofreading is always necessary
- Browser compatibility must always be checked
- Avoid one-way-streets in HTML-documents

3.11.4.3 Content Authoring Tools

There are many content authoring tools available. The most important ones are listed in Table 3.11.2. Some of them are combined with analysis tools. FrontPage from Microsoft is becoming part of Site Server, a complex product addressing site development, deployment, search and usage analysis.

TABLE 3.11.2 Content Authoring Tools

Vendor	Product
Adobe	PageMill
Allaire	Homesite
FileMaker	HomePage
Golive	CyberStudio
Micromedia	Dreamweaver
Microsoft	FrontPage
NetObjects	Fusion
Softquad	Hotmetal
Symantec	VisualPage

3.11.5 Log File Analysis

Web site activity reporting involves the analysis of:

- Basic traffic statistics (hits, page views, visits)
- Navigation patterns (referrers, next-click, entrance and exit pages)
- Content requested (top pages, directories, images, downloaded files)
- Visitor information (domains, browsers, platforms)
- Fulfillment of the Web site's objective (purchases, downloads, subscriptions)

Clearly, this last characteristic is the reason that Web site activity analysis has become an enterprise-critical priority for organizations investing massive amounts of time and money in their Web presence. How well the Web site is performing relative to its objective is what justifies continued investment. The easiest way to quantify the return on Investment (ROI) is with meaningful Web activity reports.

Reporting is also essential for making decisions about content. Web site activity reports, by providing statistics about the most popular pages or files, give an organization quantifiable measurements as to what type of content appeals to its audience. Without reliable, comprehensive reports, a Web site's content is designed based on an educated guess by the design team or editorial staff.

Similarly, Web site activity analysis reports also tell an organization about their visitors. Where are they coming from, how do they get to the Web site, and what type of browser or platform are they using? When a corporation decides to deploy a Web site, it usually has an idea about who its audience will be. Does the actual audience resemble the predicted one? How does it change over time? What type of content improves visitor retention or session depth?

3.11.5.1 Usage Analysis

Web server monitors and management tools concentrate on how the Web server is utilized and how performance goals can be met. In addition to these tools, other tools are required that are able to continue the analysis by using log files filled by special features of the server operating system. This segment is devoted to log file analyzer tools that are able to give the necessary data for in-depth usage analysis.

Usage analysis is a means of understanding what is happening on an Internet or intranet server such as a Web server. Usage analysis tools piece together data fragments to create a coherent picture of server activity.

Usage analysis can answer the following questions:

- How many individual users visited the site on a particular day?
- What day of the week is the site busiest?
- How many visitors are from a certain country?
- How long do visitors remain on the site?
- How many errors do visitors encounter?
- Where do visitors enter and leave the site?
- How long did it take most visitors to view the home page?
- Which links on other sites send the most visitors to this site?
- Which search engines send the most visitors to this site?

Reports can span any length of time, making it possible to see trends. They can also display any degree of granularity, allowing users to see both broad-ranging reports and detailed reports. Usage analysis is most frequently thought of in terms of Web servers. The reports created by usage analysis tools can be used throughout organizations to help people make informed decisions. Examples are:

- Web developers use these tools to gauge the effects of site design changes. Using this information, they can make further refinements to the design of the site to maximize its effectiveness.
- Marketers use these tools to analyze the effectiveness of marketing programs and online ads.
- Site administrators can spot Web pages that are causing errors, determine future server hardware needs, and track FTP and proxy server activity.
- Salespersons can gather information about prospects including their geographic location, how many pages they viewed, and how they found the site in the first place.
- Executives use the intelligence gathered with log analyzers as a resource when making a broad range of decisions.

Each time a visitor accesses a resource on a Web server — whether it is an image, a HTML file, or a script — the activity is usually recorded as a line in a text file associated with the Web server. This text file is known as the Web server log file. A single line of a typical Web server log file can be interpreted as follows.

Record of the server log file entry:

foo.bar.com --(31/Oct/1998:23:31:44+ 500) "GET home.html HTTP/1.0" 200 1031 http://www.yahoo.com/ "Mozilla/3.0 (Win32;U)"

Interpretation by elements:

Element	Interpretation
foo.bar.com	Hostname of the visitor's computer
31/Oct/1998:23:31:44	Date and time
GET	Method used to request the resource
home.html	Name of the requested resource
HTTP/1.0	Protocol used to request the resource
200	Status code "200" means that the request was successful
1031	Number of bytes transferred to satisfy the request
http://www.logfile.ana.html	Web page that referred the visitor to this page
Mozilla/3.0	Visitor's Web browser and version
Win32	Visitor's operating system

Most Web servers write out log files in the combined log format. It differs from an older common log format in that it contains browser and referral information. Referral information is important to determine what sites are sending the most traffic to the target address and what sites might have out-of-date

links pointing to specific user sites. Referral information is also critical for gauging the effectiveness of online ads. Other information that can be included into a log file includes:

Cookie	A persistent identification code assigned to a user which allows the user to be tracked across several visits
Session identifier	Tracks each visitor for the length of the visit only
Amount of time the request took to fulfill	Enables server performance reporting

Basically, there are two types of usage analyzer tools: software-based and on-the-wire-collectors. On the high end of usage, analysis tools are packet sniffers which offer on-the-wire reporting by installing an agent against the kernel of the operating system of the Web server. They run as root in the kernel of the operating system on the Web server. Furthermore, they require that a network runs in promiscous mode in order to expose network traffic to the agent. Usually, there are very few reports packet sniffers can create and log file analyzers cannot. Log file analyzers can create reports on the usage of secure/encrypted communications, while packet sniffers cannot. Packet sniffers are more expensive, offer less reports, and offer just a few report distribution capabilities.

3.11.5.2 Issues of Log File Analysis

When selecting products, there are a number of criteria that must be carefully evaluated. The market is big, addressed by a relatively low number of products. These criteria are also important when the webmaster wants to position log file analysis within their IT administration or when they want to deploy this functionality within their organization.

Architecture of a product answers the question whether the product can support a distributed architecture or not. Distribution means that collecting, processing, reporting, and distributing data can be supported in various processors and at different locations. Figure 3.11.10 shows these functions with a distributed solution.

In Figure 3.11.10, Web servers A, B, and C can be from very different types, such as Netscape Navigator and Microsoft Explorer.

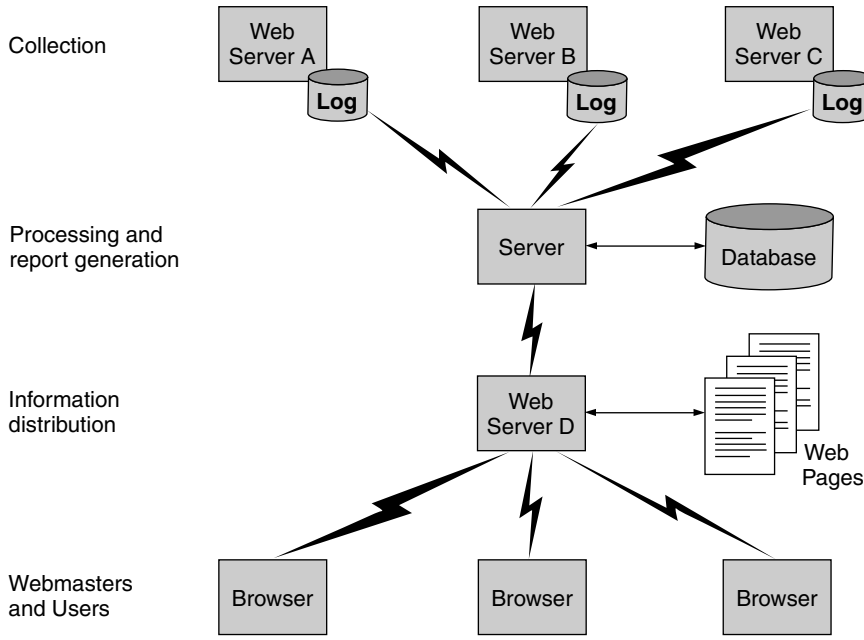

FIGURE 3.11.10 Generic product architecture for log file analysis.

Of course, it is expected that many different Web server types are supported. Also, the hardware and operating system may be a differentiator for products. It is assumed that the Web server hardware has decreasing impact on log file analysis. The role of operating systems is more significant; the product should know exactly how log files are initiated and maintained. No problems are expected with leading Web server solutions, based on Unix and NT.

The data capturing technique is absolutely essential with log file analysis. The first question is where the logs are located. Figure 3.11.10 indicates that they are located in the Web servers. But, more accurate information is required here:

- What memory area is used
- What auxiliary storage area is used
- What is the size of those areas
- What types of log files are supported

If log files are not processed in real time or near real time, it is important to know where they are stored until they are downloaded for processing. Log file analysis is dealing with very large data volumes, and these volumes depend on the visitor's traffic.

Usually, log files are downloaded for processing. It is important to know how downloads are organized and how rapidly they are executed. As indicated in Figure 3.11.10, WANs are involved with sometimes limited bandwidth. The bandwidth is usually shared with other applications, with the result of potential traffic congestion. Bandwidth-on-demand solutions are rare with log file analysis. When transmission is arranged for low traffic periods, the actuality of log file analysis results may suffer. In such cases, local storage requirements increase, and processing, report generation, and information distribution are delayed by several hours or even by days.

Two solutions may help. The first solution is using intelligent profiling at the source of data collection. Redundant data are removed from logs during collection. Data volumes decrease and local storage requirements decrease as well, but processing requirements in Web servers increase considerably. The second solution may use data compression or data compaction with the same results and impacts as with the first solution.

Overhead is a very critical issue with large data volumes. Data capturing is expected to introduce little overhead, when logs are stored away immediately. If local processing is taking place, overhead must be very carefully quantified; if resource demand is high, overall Web server performance may be impacted. Data transmission overhead can be heavy, when everything is transmitted to the site where processing is taking place. WAN bandwidth is still very expensive to be dedicated just to log file analysis. If bandwidth is shared with other applications, priorities must be set higher for business applications than for transmitting raw log file data.

In the case of server farms, a local mediation device could help. The mediation device is connected via LAN; bandwidth is not so critical in LANs in comparison to WANs. Processing and report generation remain at a special server that is consolidating all data from mediation devices.

It is absolutely necessary that all data are captured that are necessary to conduct a detailed Web site analysis of visitors or groups of visitors:

- Who is the visitor?
- What is the purpose of the visit?
- Where is the visitor coming from?
- When has the visit taken place?
- What key words have brought the visitor to the site?
- What search machines helped to access the site?
- How long was the visit?

Data losses cannot be completely avoided. Logging functions of Web servers, storage devices, or components of the transmission may fail; in such cases, there will be gaps in the sequence of events. Backup capabilities may be investigated, but IT budgets won't usually allow too much to spend for backing up large volumes of log file data. In the worst case, certain time windows are missing in reporting and in statistics. Those gaps may be filled with extrapolated data.

Also, the management capabilities are very important. One of the functions here includes automatic log cycling. In order not to lose data, multiple logs are expected to be used. When one of the logs is full, the other log seamlessly takes over. Another function is the translation of domain name service (DNS). Its speed is absolutely important for real-time information distribution. In order to generate more meaningful reports, it is required that results of log file analyzers are correlated with other data sources. These other data may be maintained in other databases. In order to correlate, *ad–hoc* database links should be established and maintained. Management of logs of any log file analyzer can be taken over by the operating system of Web servers. The basic services are supported today; additional services may follow. In the case of server farms or of many individual Web servers, the coordination of log transfers and processing is no trivial task. Event scheduler may help in this respect.

Cookie support is important to speed up work initiated by visitors. It is a logical connection between Web sites and browsers; a persistent identification code is assigned to a user, which allows the user to be tracked across several visits.

Due to considerable data volumes, databases should be under consideration to maintain raw and/or processed log file data. Database managers would then offer a number of built-in features to maintain log files. Clustering visitors may be deployed from various perspectives, such as geography, common applications, common interests on home pages, and data and time of visits. Automatic log cycling can also be supported here by the database managers. Open database connectivity (ODBC) support helps to exchange data between different databases and to correlate data from various databases. Besides log files, other data sources can be maintained in the same data warehouse. Besides routine log files analysis with concrete targeted reports, special analysis may also occasionally be conducted. This special analysis, called data mining, can discover traffic patterns and user/visitor behavior. Both are important to sizing systems and networking resources.

One of the most important questions is how log file analysis performs when data volumes increase. Volume increase can be caused by offering more pages on more Web servers, more visitors, longer visits, and extensive use of page links. In any case, collection and processing capabilities must be estimated prior to deciding for precedures and products.

In order to reduce processing and transmission load of log files, redundant data should be filtered as near as possible to the data capturing locations. Filters can help avoid storing redundant data. Filters can also be very useful in the report generation process. Again, unnecessary data must not be processed for reports. Powerful filters help to streamline reporting.

Not everything can be automated with log file analysis. The user interface is still one of the most important selection criteria for products. Graphical user interfaces are likely, but simple products are still working with textual interfaces. When log file analyzers are integrated with management platforms, this request is automatically met by management platforms.

Reporting is the tool to distribute the results of log file analysis. Predefined reports and report elements as well as templates help to speed up the report design and generation process. Periodic reports can be automatically generated and distributed for both single Web servers and Web server farms. In the cases of many Web servers, report generation must be carefully synchronized and scheduled. Flexible formatting helps to customize reports to special user needs.

Output alternatives of reports are many. The most frequently used solutions include Word, Excel, HTML, and ASCII. Also, the distribution of reports offers multiple choices:

- Reports may be stored on Web servers to be accessed by authorized users who are equipped with universal browsers
- Reports can be uploaded into special servers or even pushed to selected users

- Reports may be distributed as attachments to e-mail messages
- Reports can also be generated at remote sites; this alternative may save bandwidth when preprocessed data instead of completely formatted reports are sent to certain remote locations

Documentation may have various forms. For immediate answers, an integrated on-line manual would be very helpful. Paper-based manuals are still useful for detailed answers and analysis. This role, however, will be taken over by Web-based documentation systems. In critical cases, a hot line can help with operational problems.

Log file analysis is actually another management application. If management platforms are used, this application can be integrated into the management platfrom. There are many ways to integrate; most likely a command line interface (CLI) will be deployed.

3.11.5.3 Drawbacks of Pure Log File Analyzers

Log file analysis can give a good entry-level summary about the activities in and around Web servers. But this technology shows major problems that are analyzed as follows.

The first major problem is traffic volumes. As traffic levels quickly reached exponential growth rates, nightly log file downloads quickly became afternoon-and-evening, and then even hourly downloads, since server disk drives would fill with log file data so quickly. Compounding this problem was the fact that higher-traffic sites needed to load-balance across several servers and physical machines, so that log file downloads needed to be done not only many times a day, but also across several machines each time. The quick fix to this problem was typically an automated script that would download log files on a preset schedule. However, this failed to account for unexpected spikes in traffic and also clogged internal networks with huge log files being transmitted across the network several times a day.

The second major problem is data processing speed. Even if there were an easy way to continuously transfer log file data to a consolidated area, there was still the problem of how to process the gigabytes of log files into database tables in an efficient, continuous, and robust manner. Batch processing of log file data requested a considerable amount of time. In addition, the human resources demand for log file collection, processing support, and report compilation has exceeded the expectations.

The third major problem involved incomplete data. Beside log files, there are significant alternate sources of site activity data which contain more information than even the longest, most complex custom log file format can provide. A log-file only approach cannot guarantee a complete picture for Web activities. A good example of missing data is certain network-level data that the Web server and the server's log file never get to see. For instance, a visitor requests a page that turns out to be too slow to download, and decides to hit the browser STP button, BACK button, or otherwise terminate the request in mid-download. In this case, the network layer will log that action, but it will not notify the Web server about it. Similarly, there is much data that is seen by Web servers, but never written to the log file. Therefore, any measurement approach based solely on log files would occasionally miss critical information about user activity on the Web site.

The fourth major problem with the log file approach is flexibility. As sites become more sophisticated, one of the first obvious enhancements is to add dynamically generated content. Regardless of the type of content management system used, dynamic content typically results in URLs that are very difficult, if not impossible, for a human reader to decipher. Since log files are just transaction records, dump reporting systems simply pass the nonsensical URLs through to the end-user report as the page that was requested, resulting in an unintelligeble report with meaningless page names and URLs. The ideal solution would be to interpose some intelligent classification system between the raw activity data and end-user report. In practice, however, the reality of gigabytes of raw log files often leave an in-house analysis team with few human resources to add even more complexity to an already slow log-based process. The inflexibility of log files to handle the tracking of new technologies has been observed not only with dynamic content but also with personalization applications, applet-based multimedia technologies, and a host of other new capabilities which the log file approach was never designed to handle.

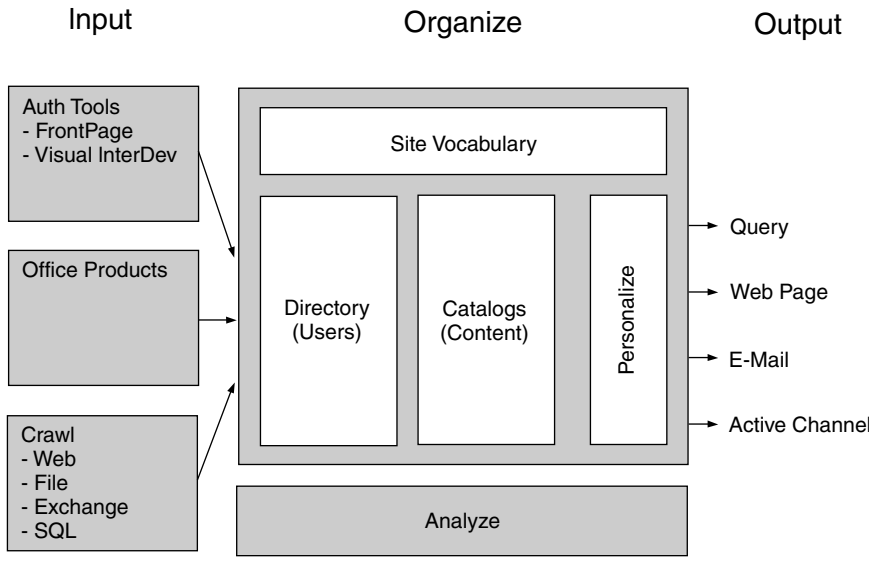

FIGURE 3.11.11 Key components of SiteServer.

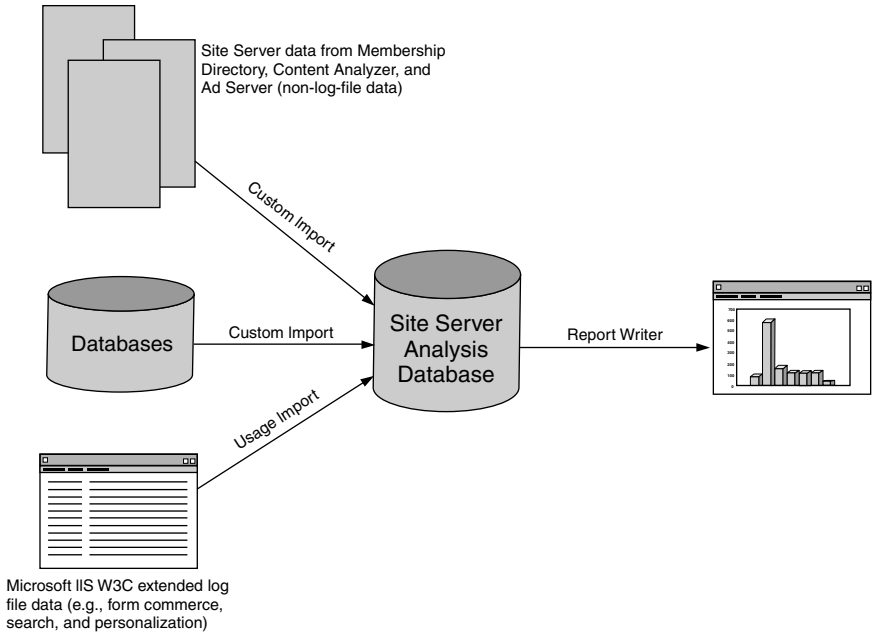

FIGURE 3.11.12 Reporting feature with SiteServer.

Figure 3.11.11 shows the principal components of SiteServer from Microsoft. Figure 3.11.12 displays the reporting process with SiteServer.

Finally, Table 3.11.3 summarizes a general Web server statistic based on log file analysis.

In summary, though log files were a convenient approach to measurement in the early days of using the Web, they rapidly highlighted problems of:

TABLE 3.11.3 General Web Server Statistics

Date and time this report was generated	Friday, December 4, 1998; 07:46:17 a.m.
Time frame	11/01/98 01:03:38–11/30/98 22:05:41
Number of hits for home page	1031
Total number of successful hits	5729
Total number of user sessions	1119
User sessions from the U.S.	0% (not broken down)
User sessions from outside the U.S.	0% (not broken down)
Origin unknown user sessions	100%
Average hits per day	190
Average user sessions per day	37
Average user session length	67:11:49

- Labor intensity
- Slow data processing speeds and turnaround times measured in weeks
- Incomplete data, missing server- and network-level data
- Ineffective tracking of new feature enhancements such as dynamic content, personalization, and applet-based multimedia

In response to these problems, hybrid products have been developed and deployed.

3.11.5.4 Log File Analysis Tools

There are numerous log file analysis tools. Their depth and functionality are very different. Some of them are complex in their nature, and offer more than just log file analysis. The most widely used tools are listed in Table 3.11.4.

TABLE 3.11.4 Log File Analysis Tools

Vendor	Product
met.Genesis	net.Analysis
WebManage	NetIntellect
WebTrend Corporation	WebTrends
Marketware	Hit List
Andromedia	ARIA
Microsoft	SiteServer

3.11.6 Wire Monitors

Log files are not the only source of information for analyzing Web sites. There are other tools that are residing “on-the-wire” or LANs and collecting information on performance and traffic metrics. The information depth and the overhead are significant indicators that may differentiate between log file analyzers and these products. In certain environments, the most effective results can be achieved only when both types of tools are deployed in combination.

Over the past several years, companies have adopted distributed multi-tier network infrastructures, and moved business operations from traditional client/server applications to distributed Web-based applications. However, as more and more users come to depend on Web servers and TCP-based services, IT organizations are discovering that their current infrastructures are unable to offer the performance and availability expected by users; nor do they provide the management and monitoring capabilities required by IT organizations themselves.

3.11.6.1 Changes in Networking Infrastructures

Over the past several years, large corporations have begun re-engineering their enterprise networks and establishing distributed, multi-tier infrastructures. These multi-tier infrastructures typically include three levels:

- At the wide area network (WAN) level to enable communication across multiple points of presence (POPs)
- At the Web level, to support server farms providing a wide range of TCP-based services, including HTTP, FTP, SMTP, and Telnet
- At the application level, to support farms of application servers that offload computation from Web servers to increase overall site performance

IT organizations are deploying newly distributed, Web-based applications to take advantage of this new enterprise infrastructure. In place of fat software clients and centralized application servers, corporations are deploying Web browsers on every desktop, Web servers in departments and divisions, and application servers residing at multiple locations.

The new Web-centric model offers several advantages over the client/server model it replaces. IT departments can deploy Web browsers quickly and affordably to every desktop platform. Basic Web skills can be learned quickly, and are popular with users. If an application need requires modification to reflect changing business practices, IT departments need only modify the application itself, not the complex clients that used to work with the application. Most importantly, distributed, Web-based infrastructures move content and applications closer to users and provide improved reliability and availability. Employees can leverage this new infrastructure to improve internal business practices, communication with partners and suppliers, and services for customers.

While distributed, multi-tier infrastuctures offer considerable advantages over earlier network architectures, they still do not offer the performance and availability expected by end users; nor do they provide the management and monitoring capabilities expected by IT organizations. Multi-tier architectures are physically well connected, but not logically well connected. Standard network equipment enables traffic to flow, but not necessarily to the server best suited to respond. IT departments deploying these networks need traffic management solutions that intelligently direct TCP traffic to optimal resources at each tier of the enterprise infrastructure. An optimal traffic management solution requires communication between tiers. For example, there is little point in a DNS server directing traffic to a local server, if that server is down or overloaded, while another server is available with processing cycles to spare. To perform its job optimally, the DNS server needs availability and load information from the servers to which it directs requests.

The multi-tier model itself, when implemented with the standard software products available today, does not monitor services for system failures or spikes. Nor does it provide other capabilities that IT departments require to manage busy, distributed networks effectively. Specifically, it provides no:

- Policies for scheduling TCP traffic based on specific events centralized
- Remote management reporting integration with standard network management tools

IT organizations need integrated software systems that can be layered on top of the existing infrastructure to provide intelligent scheduling of requests and information.

3.11.6.2 Issues of Data Collection

The targeted metrics are the same as with log file analyzers, but the source of data is different. When selecting products, there are a number of criteria, such as information depth, overhead, and reporting capabilities, that must be carefully evaluated. The market potentials are good, addressed by a few vendors. These criteria are also important when webmasters want to position traffic measurements within their IT administration or when they want to deploy this functionality within their organization.

Architecture of a product answers the question whether or not it can support a distributed architecture. Distribution may mean that collecting, processing, reporting, and distributing data can be supported in various processors and at different locations. Figure 3.11.13 shows these functions with a distributed solution.

The monitors are passively measuring the traffic in the network segments. They are actually microcomputers with ever increasing intelligency. Their operating systems are either proprietary or based on

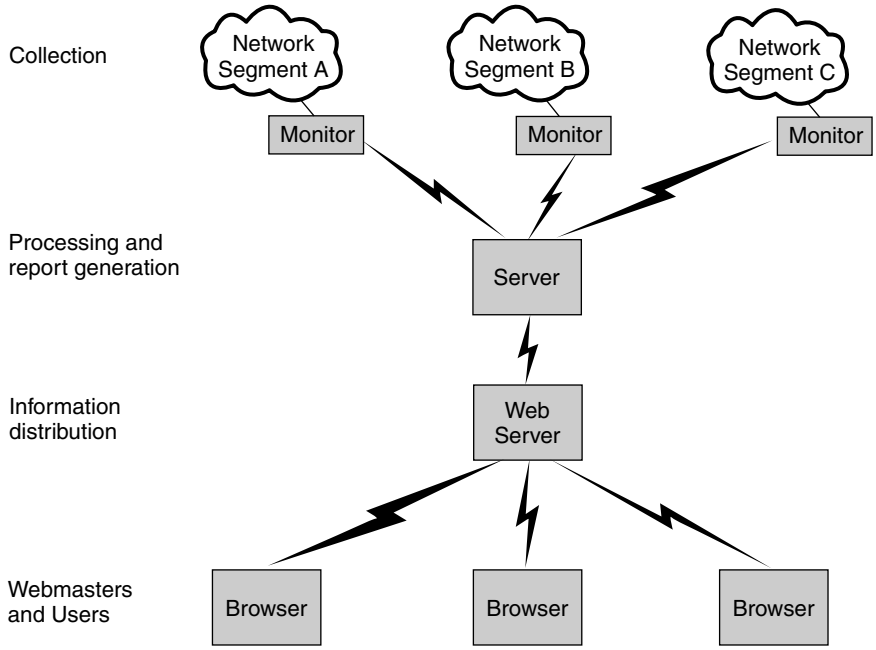

FIGURE 3.11.13 Generic product architecture for processing traffic measurements data.

Unix or more likely on NT. Usually, they are programmed to interpret many protocols. TCP/IP and UDP/IP, and HTTP are high on the priority list of vendors.

The data capturing technique is essential with traffic measurement tools. The measurement probes are attached to the digital interface of the communication channels. They can reside directly on the network (standalone probes) or co-located with networking equipment. In this case, the probe is used as a plug-in. Even software probes can be used and implemented into networking components or into end-user devices. The hardware or software probes usually include event scheduling. It means determining polling cycles and time periods when downloading of measurement data is intended. Transmission should be scheduled for low-traffic periods. Probes are expected to deal with large data volumes. These volumes depend — to a large degree — on visitor's traffic in networking segments. Probes have limited storage capabilities; implementation examples show capabilities up to 24 hours. When this limit is exceeded, measurement data are overwritten by new data. Usually, measurement data are downloaded for further processing. It is important to know how downloads are organized and how rapidly they can be executed. As indicated in Figure 3.11.13, wide area networks are involved that may show bandwidth limitations. The bandwidth is usually shared with other applications with the result of potential traffic congestions. Bandwidth-on-demand-solutions are rare with measurement probes. When transmission is arranged for low traffic periods, the actuality of measurement results may suffer. In such cases, local storage requirements increase, and processing, report generation, and information distribution are delayed by several hours or even by days.

Two solutions may help. The first is using intelligent filtering during and shortly after data collection. Redundant data are removed from captured packets during collection. Data volumes decrease, local storage requirements decrease as well, but processing requirements of the probes increase. The second solution may use data compression or data compaction with the same results and impacts as can be observed with the first solution.

Overhead is a very critical issue with large data volumes. Data capturing is expected not to introduce any overhead in case of hardware-based probes. Overhead is minimal with software-based probes. It is assumed that measurement data are stored away immediately after collection. If local processing is taking

place, overhead must be critically quantified. If resource demand is high, probes must be upgraded properly. Data transmission overhead can be heavy, when everything is transmitted to the site where processing takes place. Dedicated bandwidth would be too expensive for measurement and management purposes only. If bandwidth is shared with other applications, priorities must be set higher for business applications than for transmitting measurement data.

It is absolutely necessary that all data are captured that are necessary to conduct a detailed Web site analysis of visitors or groups of visitors.

- Who is the visitor?
- What is the purpose of the visit?
- Where is the visitor coming from?
- When has the visit taken place?
- What key words have brought the visitor to the site?
- What search machines helped to access the site?
- How long was the visit?

Data losses cannot be completely avoided. Probes, monitors, networking devices, user workstations, or transmission equipment may fail; in such cases, there will be gaps in the sequence of events. Backup capabilities may be investigated, but IT budgets won't usually allow too much to spend for backing up large volumes of log file data. In the worst case, certain time windows are missing in reporting and in statistics. Those gaps may be filled with extrapolated data.

Due to considerable data volumes, databases should be under consideration to maintain raw and/or processed data. Database managers would then offer a number of built-in features to maintain data. Clustering visitors may be deployed from various perspectives, such as geography, common applications, common interests on home pages, data, and time of visits. Automatic log cycling can also be supported here by the database managers. Open database connectivity (ODBC) support helps to exchange data between different databases and to correlate data from various databases. Besides measurement data, other data sources can also be maintained in the same data warehouse. Besides routine log file analysis with concrete targeted reports, special analysis may also occasionally be conducted. This special analysis, called data mining, can discover traffic patterns and user/visitor behavior. Both are important in sizing systems and networking resources.

One of the most important issues is how measurement data analysis performs when data volumes increase. Volume increase can be caused by offering more pages on more Web servers, more visitors, longer visits, and extensive use of page links. In any case, collection and processing capabilities must be estimated and quantified prior to deciding procedures and products.

In order to reduce processing and transmission load of measurement data, redundant data should be filtered out as near as possible to the data capturing locations. Filters can help to avoid storing redundant data. Filters can also be very useful in the report generation process. Again, unnecessary data must not be processed for reports. Powerful filters help to streamline reporting.

Not everything can be automated with analyzing measurement data. The user interface is still one of the most important selection criteria for products. Graphical user interfaces are likely, but simple products are still working with textual interfaces. When measurement data are integrated with management platforms, this request is automatically met by management platforms.

Reporting is the tool to distribute the results of log file analysis. Predefined reports, report elements, and templates help to speed up the report design and generation process. Periodic reports can be automatically generated and distributed for both single Web servers and Web server farms. In the cases of many Web servers, report generation must be carefully synchronized and scheduled. Flexible formatting helps to customize reports to special user needs.

Output alternatives of reports are many. The most frequently used solutions include Word, Excel, HTML, and ASCII. Also, the distribution of reports offers multiple choices:

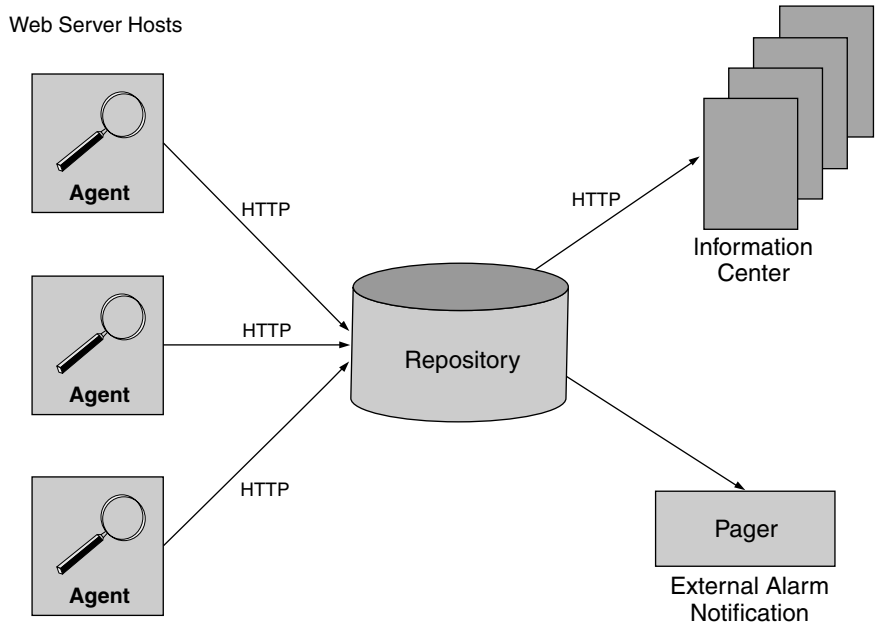

FIGURE 3.11.14 Architecture of WebSniffer.

- Reports may be stored on Web servers to be accessed by authorized users who are equipped with universal browsers
- Reports can be uploaded into special servers or even pushed to selected users
- Reports may be distributed as attachments to e-mail messages
- Reports can also be generated at remote sites; this alternative may save bandwidth when preprocessed data instead of completely formatted reports are sent to certain remote locations

Figure 3.11.14 shows the architecture of WebSniffer, one of the well-known wire monitors. Figure 3.11.15 shows distributed monitoring capabilities with Net.Medic Pro, known for for its rich reporting selection.

Documentation may have various forms. For immediate answers, an integrated on-line manual would be very helpful. Paper-based manuals are still useful for detailed answers and analysis. This role, however, will be taken over by Web-based documentation systems. In critical cases, a hot line can help with operational problems.

Measurement data analysis is actually another management application. If management platforms are used, this application can be integrated into the management platform. There are many ways to integrate; most likely a command line interface (CLI) will be deployed.

3.11.6.3 Traffic Monitoring Tools

There are just a few tools supporting traffic monitoring. Table 3.11.5 displays the list of these tools. The best results are expected in such cases, where these tools are used in combination with load balancers and traffic shapers.

3.11.7 Web Server Management

The content of Web pages is maintained on Web servers. Usually, they are processors running under Unix or NT. They must be flexible and scalable enough to cope with significant workload fluctuations.

Server management is composed of several functions:

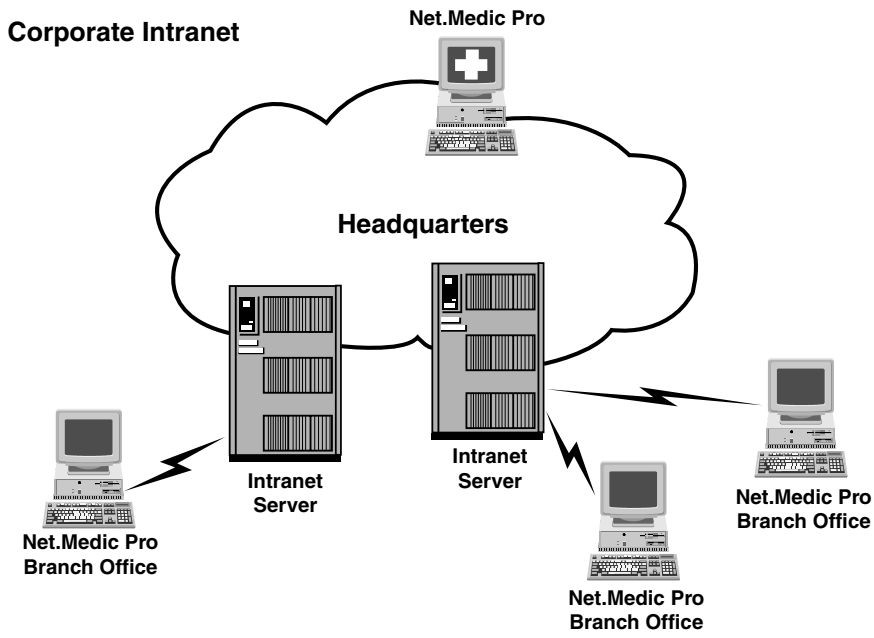

FIGURE 3.11.15 Distributed monitoring with Net.Medic Pro.

TABLE 3.11.5 Traffic Monitoring Tools

Vendor	Product
Network Associates	WebSniffer
Resonate	IntelliFlow
Sane Solutions	NetTracker
Telemate Software	Telemate.Net
Visual Networks	OnRamp
Vital Signs	Net.Medic

- Server monitoring — This function is the base component of server management. This requires someone or something to keep a constant watch on the status of the managed servers. This is a painfully tedious task for human beings, much of which, fortunately, can be automated by management platforms, such as Unicenter TNG or HP OpenView. Monitoring is essential for detecting problems as soon as they occur, and for gathering data for use in performance management.
- Workload management — This function consists of scheduling and tracking the jobs that run across one or more servers in a heterogeneous environment. Workload management takes into account calendar requirements such as time of day, day of week, or holidays. It also considers dependencies between workloads, such as Job A must be finished before Job B should be started, as well as what to do in the case of a failure.
- Server performance management — While monitoring focuses on server availability, the purpose of server performance management is to ensure that servers are working efficiently. The keys to this function are data collection and trend analysis.
- Server capacity planning — While performance management focuses on current effectiveness, capacity planning ensures that servers will work effectively in the future. The keys to this function are historical analysis and forecasting.

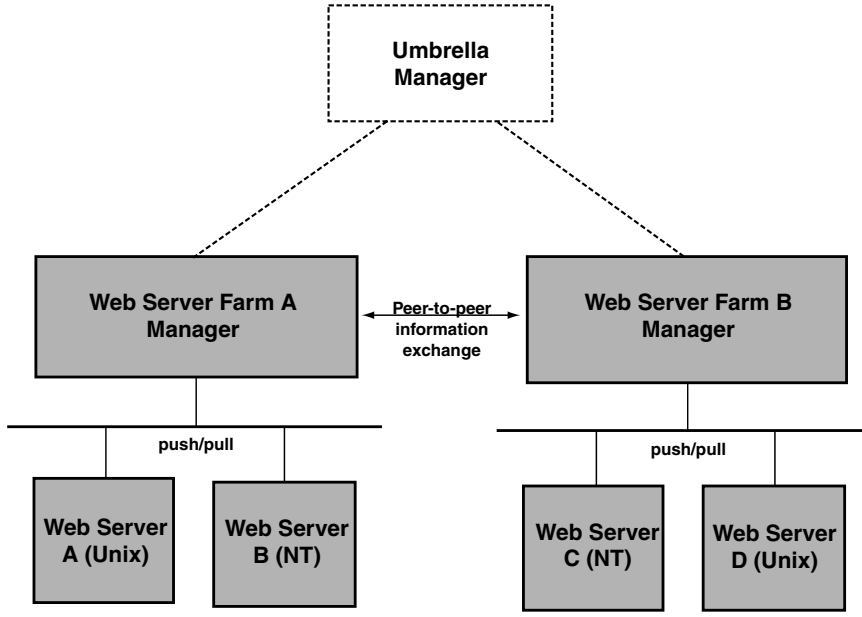

FIGURE 3.11.16 Decentralized Web server management.

The management architecture for Web servers may be central or decentral, or a combination of both. A centralized solution assumes that all Web servers can be managed from one location. When the number of Web servers to be managed exceeds a certain number, this solution could become critical in terms of networking overhead. It is assumed that with the exception of collecting raw data, all processing functions are executed in the manager.

With a decentralized solution, domain managers take over the responsibility of managing a certain number of Web servers (Figure 3.11.16). Each domain is actually a centralized solution on its own. Domain managers may communicate with each other or can even be connected to an umbrella manager. Network overhead can be well controlled and kept to a minimum. Domain managers usually just exchange consolidated data with each other. The result is that the communication overhead can be kept to a minimum.

Practical arrangements usually work with a combination of these two alternatives. If umbrella management is the choice, this manager can also manage other components such as switches, routers, and other components, and can correlate data with server management.

It is important to know when different architectural alternatives are under consideration which operating systems of Web servers can be managed. Web servers are usually deployed on Unix or NT platforms.

In terms of hardware and software of the manager, there are multiple choices. The software is Unix or NT; the hardware is constantly losing importance, because both leading operating systems are working with a number of hardware platforms.

Data capturing techniques are critical for both overhead and performance of the management architecture. Measurement probes or agents are located inside the operating system; they run with relatively high priority. These agents can supervise both hardware and software components of Web servers. Raw data are expected to be stored away immediately. Processing can be done here in the Web server or in the manager. The targeted metrics to be collected include:

- What is the CPU utilization by applications?
- What are physical and logical I/O-rates?
- Can the list of active applications be generated?
- What is the average queue length for CPU?
- What is the average queue length for I/O devices?

- How high is the CPU/I/O overlap?
- Are process wait times measured and displayed?
- How high is the disk utilization?
- How high is the memory utilization?
- Are swap rates measured?
- What are resources that processes are blocked on?
- What reporting is used?
- Can user by application be identified?

Raw data or preprocessed data are stored at the Web servers with the intention of being uploaded for further processing by the manager. Upload may be controlled in two different ways:

- Upload is triggered by events, such as filling percentage of storage spaces, or time, or when critical data are captured, or
- Upload is controlled by polling cycles initiated by the manager

Both alternatives have pros and cons; the selection depends on the actual configuration, data volumes, and the communication protocols in use. Web server management can utilize SNMP for transmitting data, assuming Web server metrics are stored and maintained in MIBs. Another alternative is the use of DMI-like standards for storing and transmissions. The recent alternative is the use of embedded Wbem-agents that are supporting the common information model (CIM) for storing and exchanging data. In this case, HTTP is the protocol of choice.

As for overhead, concerns are similar to those experienced with traffic monitors. Data capturing is expected to introduce little overhead when data are stored away immediately. If local processing is taking place, overhead must be very carefully quantified; if resource demand is high, overall Web server performance may be impacted. WAN bandwidth is still very expensive to be dedicated just to transmitting management data. If bandwidth is shared with other applications, priorities must be set higher for business applications than for transmitting raw log file data.

Here as well, data losses cannot be completely avoided. Data capturing functions in Web servers, storage devices, or components of the transmission may fail; in such cases, there will be gaps in the sequence of events. To protect as much data as possible, we re-emphasize the importance of database use to properly maintain Web server measurement data, and effective processes to filter redundant data and facilitate timely report generation/documentation.

One of the most important questions is how Web server management performs when the number of managed Web servers are maxed out, and as a result of this, data volumes increase. All resources, such as processors, storage devices, I/O-devices within the Web servers, and networking components may become the bottleneck.

Figure 3.11.17 displays information sources for NT management.

Managing Unix and NT servers represents just another management application. If management platforms or umbrella managers are used, these applications can be integrated into the platform. There are many ways to integrate; most likely a Command Line Interface (CLI) solution will be deployed. Integration may even be supported by a management intranet. Every participant is equipped with a universal browser and communicates with management applications residing in managed objects and being equipped with lean Web servers.

The majority of Web-server-implementation is based on Unix or NT. Some of them are on NetWare, but their market share is not significant.

Table 3.11.6 displays examples for Web server management tools. The tools are grouped by operating systems they support, such as Unix and NT.

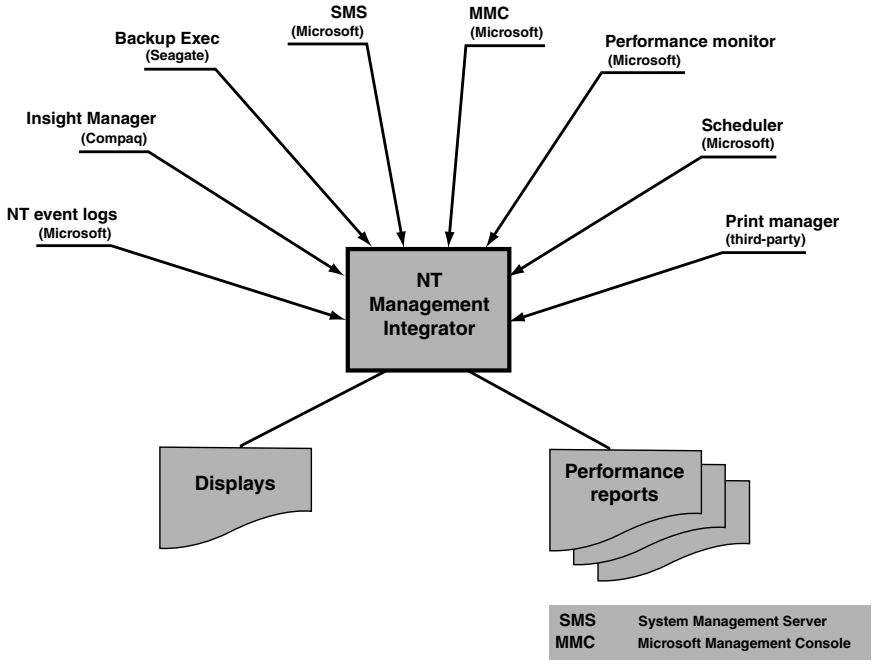

FIGURE 3.11.17 Information sources for NT management.

TABLE 3.11.6 Web Server Management Tools

Vendor	Product
	Unix Management
Computer Associates	Unicenter TNG
BMC	Patrol Knowledge Module for Unix
Hewlett-Packard	PerfView
Hewlett-Packard	GlancePlus
Hewlett-Packard	PerfView RX
	NT Management
BMC	Patrol Knowledge Module for NT
Computer Associates	Unicenter TNG
Heroix	Robomon
Hewlett-Packard	ManageX
NetIQ	AppManager
Seagate	ManageExec

3.11.8 Load Balancing

In order to help IT managers track IP performance and to optimize bandwidth usage across WANs, several new vendors offer hardware- and software-based load balancing products. Load balancers typically reside at the edges of corporate networks and decide about traffic priorities. They apply a policy that defines different traffic types and determine what happens to each. A very simple policy may call for priorities for a specific sender. Other criteria may be TCP port numbers, URLs, and also domain name service (DNS). Traffic shaping may be supported by queueing or via TCP rate control. There are products available for both categories.

Optimization is accomplished by controlling enterprise traffic flows at the boundary between the LAN and the WAN. Because these products give priority to traffic according to application type or even individual users, they will let IT managers take the first steps toward policy-based QoS in their networks. These products are a logical evolution from the passive probes that gave users a certain level of visibility for fault operations monitoring but no actual control over traffic. These products go further, and can manipulate traffic. IT managers expect that this new class of traffic-shaping tools will ease the congestion for bandwidth without forcing purchase of more and larger physical transmission lines.

This segment of the book introduces a couple of innovative solutions provided by start-ups and known flow-control companies.

3.11.8.1 The Needs for Bandwidth, Service Quality, and Granularity

Bandwidth management is rapidly becoming a must for internet service providers (ISPs) as well as corporations running their own global intranets. The reasons for bandwidth management are the following.

The Move to Internet/Intranet-based Business

Corporate networks are rapidly evolving from a classic client/server paradigm toward an intranet-based model, based on information-sharing and Web navigation. Analysts predicted that by the year 2000 there would be over 3 million private intranet sites, compared to approximately 650,000 Internet sites. The result is the demand for significantly more bandwidth. Adding more channels and more bandwidth to each channel will not guarantee availability and performance, where it is needed most. An intranet-based model implies the following factors:

- Changing patterns or network use and unpredictable demands for bandwidth. Global users access the network 24 hours a day, 7 days a week. As information appears and disappears on Web sites, access patterns change and saturation moves around the network.
- Demand for increased amounts of bandwidth. People may stay on the link for extended periods of time and download large amounts of data.
- Demand for guaranteed QoS in terms of bandwidth and minimum delay. Emerging Internet applications are both bandwidth intensive and time sensitive, often requiring support for voice, video, and multimedia applications across the network infrastructure.
- Lack of control by IT staff. Workgroups and departments generally create their Web sites without IT approvals, generating increased traffic without necessarily having the infrastructure to handle it. This often results in excessive traffic at the fringes of the network where Web sites are situated, generating traffic precisely where there is least provision.
- A change in user attitude. Users expect instant access to information without delays or restrictions, especially if that information is critical to their work.

The Need for Guaranteed Bandwidth

Current networking technology has two major limitations:

- The bandwidth available on a link at any given moment cannot be predicted in terms of quantity or quality. Bandwidth management is needed to allow applications which require a specific quality of service, in terms of bandwidth and delay (such as desktop video conferencing), to reserve the bandwidth quality of service they need.
- It is difficult to control which applications or users get a share of the available bandwidth. In some circumstances, an application or a user can take control of all the available bandwidth, preventing other applications or users from accessing the network. To solve this problem, the user can either add extra capacity at additional costs, resulting in an overprovisioned network that still does not guarantee equal access, or the user can introduce bandwidth allocation.

The Need for Service Level Agreements

Virtual private networks (VPN) are a popular value-added Internet service that corporations are increasingly moving toward. Enterprise customers seeking a VPN provider are more likely to sign with an ISP that can offer a contractual service level agreement — one that guarantees quality of service.

While service level agreements (SLA) cannot guarantee end-to-end service across the public Internet, they can be implemented for transport over a single-vendor network or for Internet server hosting. In these areas, a SLA is an important differentiator for an ISP.

Generally, the customer subscribes to a particular class of service, and signs a SLA accordingly. Packet throughput is monitored as part of the agreement. Value-added services were expected to grow at almost 175% in the U.S. up to the year 2000. ISPs that want to get a piece of this additional business clearly need to implement bandwidth management in order to meet SLAs which guarantee QoS toward customers. Only efficient bandwidth management can enable them to tune network behavior so that customers receive the quality of service they are charged for.

The new paradigm is a service-driven network. This is a responsive, reliable, modular infrastructure, based on the latest generation of management technology and built on dynamic, flexible management services. To respond to today's business needs, ISPs and large enterprises must deploy the service-driven network. It delivers innovative services, such as unified roaming, push browsers, multicast, on-line shopping, etc. to customers faster and at a lower cost than ever before.

The Need for Granularity

Bandwidth allocation based simply on filtering by protocol is not sufficient to meet bandwidth management needs. One of the key issues in this area is the extensive and increasing use of HTML/HTTP systems for OLTP. Within the next few years, the volume of HTTP-based OLTP traffic is expected to exceed the volume of traditional OLTP traffic. A fine level of granularity is needed for bandwidth management to take into account more than just the protocol when assesing the relative importance of network traffic. Bandwidth management must base allocation not only on protocol type, but also on the application and users involved.

3.11.8.2 Issues of Deploying Load Balancing Products

Load balancing helps to utilize resources more effectively. At the same time, the end-user response time may be stabilized and improved. This is an emerging area with a number of innovative products that work hardware- or software-based. Even there, a few implement load balancing functions in both hardware and software. The hardware solution is faster; the software offers more flexibilities if changes are required.

The functionality of a load balancer can be deployed in a standalone device or embedded into existing networking components, such as routers, switches, and firewalls. The standalone solution offers broad functionality without impacting any other routing, switching, or firewall functions. But it will add components into the network that must be managed. It may add another vendor that may be managed as well. The embedded solution is just the opposite; easier management at a price of conflicting functions with its host.

Load balancers are only successful when policy profiles can be implemented and used. Policy profiles are most likely based on supporting various transmission priorities. Priorities may be set by applications, by users, or a combination of both. The technology of solution may differ from case to case, and product to product, but most frquently the TCP flow is intercepted.

Load balancers are expected to support a number of services, such as quality control, resource management, flow control, link management, and actual load balancing. Advanced products support all these services in dependency of page content. It requires more work to gather the necessary information about content, but it offers better services for high-priority content.

Functions in a narrower sense include traffic shaping, load balancing, monitoring, and baselining. Baselining means to find the optimal operational conditions for a certain environment. It may be expressed by a few parameters, such as resource utilization, availability, and response time. Load balancers should monitor these few metrics, and act on them. Traffic shaping and load balancing help restore normal conditions by splitting traffic, redirecting traffic to replicated servers, delaying payload transport, etc.

One of the most important questions is how load balancing performs when data volumes increase. Volume increase can be caused by offering more pages on more Web servers, more visitors, longer visits,

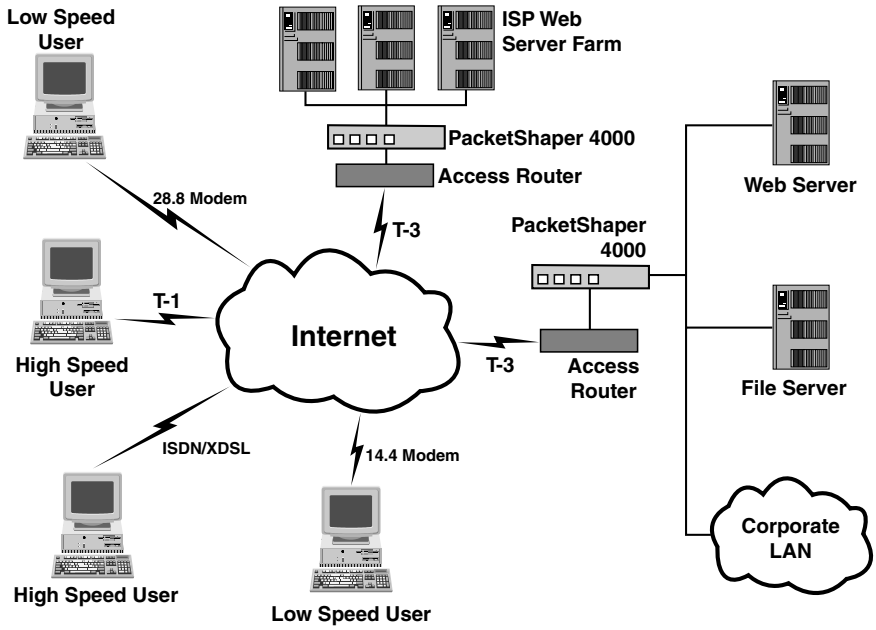

FIGURE 3.11.18 PacketShaper from Packater in operation.

extensive use of page links, etc. In any case, collection and processing capabilities must be estimated prior to deciding for precedures and products.

Load balancing products can be managed by SNMP- or Wbem-agents. They are handled by managers as with any other kind of managed object. As before, various approaches may be taken for documentation and assistance to generate documentation.

Figure 3.11.18 shows PacketShaper in operation.

Managing load balancers out of a management platform offers integration at the management applications level. Baselining and monitoring may even be supported by other applications. In case of using management intranets, universal browsers may be used to view, extract, process, and distribute management information. The only prerequisite is that Wbem agents have been implemented and that CIM is supported for information exchange.

3.11.8.3 Load Balancing Tools

Table 3.11.7 lists the presently available tools supporting load balancing and traffic shaping.

TABLE 3.11.7 Load Balancing Tools

Vendor	Product
Allot Communications	AC200 and AC300
CheckPoint Software Technology	Floodgate-1
Internet Devices	Fort Knox Policy Router
NetGuard	Guidepost
NetReality	WiseMan
Netscreen Technologies	Netscreen 10 and Netscreen 100
Packeter	PacketShaper
RND	Web Server Director
Structured Internetworks	IPath 10 and IPath 100
Sun Microsystems	Bandwidth Allocator
Ukiah Software	Trafficware
Xedia	Access Point

3.11.9 Look-through Measurements

Web application requirements have gone from zero to mission-critical within a very short period of time. The available tools have not kept up with this speed. In a business environment where "Connections failed" means the same thing as "Closed for business," IS/IT professionals are left to struggle with the challenges of building a highly avalable, high-performance server infrastructure.

Many problems interact with each other:

- The majority of Web sites, both Internet and intranet, use single Unix or NT servers. Like mainframe solutions of the past, these centralized servers have become single points of failure. Even minor system upgrades become major service problems for demanding users.
- As the demands of interactivity grow, the cost of WAN bandwidth becomes a major factor. System configurations that force all user access out across the WAN for each request stretch out retrieval times, and raise users' frustration levels.
- The increasing complexity of Web applications add even more overhead; electronic commerce and multi-tier content architectures that build pages on the fly out of applications and databases make high reliability an even more important — and costlier — goal.

The severe problem in addition to all of these is that the Web technology base is narrow. In other words, solutions that can be applied to these problems are expensive and not very effective. Adding WAN bandwidth and a larger server are just the first steps in a never-ending circle. Adding mirrored, distributed servers increases server costs significantly as well as the complexities and costs of content distribution. Hiring more webmasters and Web administrators to reboot downed web applications and servers is not the ultimate solution. And, in a world of increasingly dynamic content and transactions, how effective will server caches and load balancing tools really be?

3.11.9.1 Response Time Measurements

Response time is one of the key metrics in all SLAs. Its definition varies, but most users consider the duration between sending the inquiry until receiving the full answer as response time. There are two alternatives:

- Time up to the first character of the response on the screen of the user
- Time up to the last character of the response on the screen of the user

The second definition is better suited for the working cycle of users. The difference between RT2 and RT1 depends on many factors, such as the throughput of the backbone and access networks, servers in these networks, number of hops, and the hardware/software capabilities of the client's workstation or browser. Present measurement technology offers the following alternatives:

- Monitors and packet analyzers: They filter and interpret packets and draw inferences about application response times based on these results. These monitors are passively listening to the network traffic and calculate the time it takes specific packets to get from source to destination. They can read the content of packages, revealing eventual application errors and inefficiency. But they cannot measure response time end to end.
- Synthetic workload tools: They issue live traffic to get a consistent measurement of response time on a particular connection in the intranet or for a given application. These tools are installed on servers, desktops, or both. They typically send TCP messages or SQL queries to servers and measure the time of the reply. Results from multiple sources are correlated to give a more detailed view about intranet response times. They are very accurate to the end-to-end response time.
- Application agents: They work within or alongside applications, using software that monitors keystrokes and commands to track down how long a specific transaction takes. They can run at both the client and server. They clock specific portions of the application at the server or at the

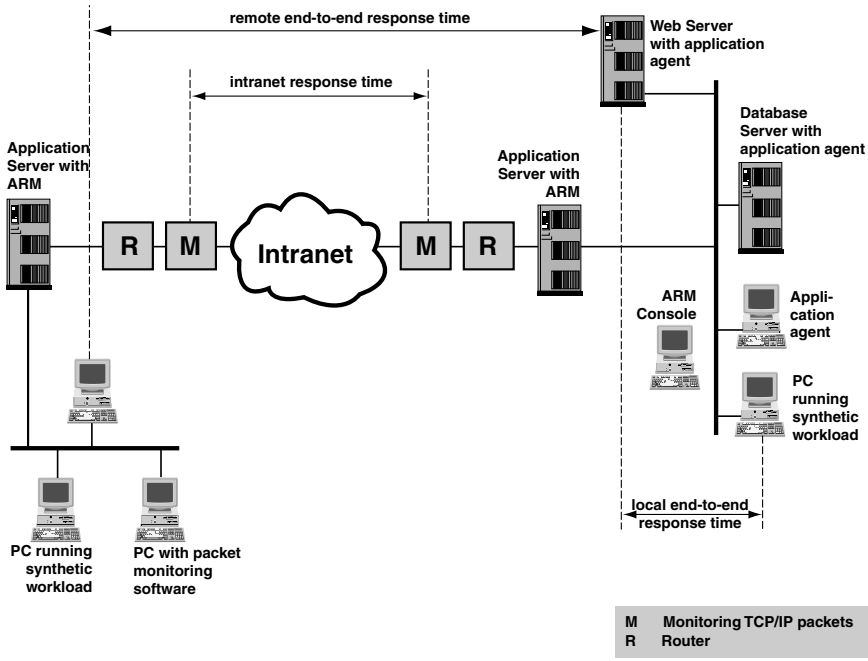

FIGURE 3.11.19 Positioning response time measurement tools.

workstation. The use of agents needs customization and the correlation of many measurements in order to give users a performance estimate about their intranet.

- Use of ARM MIBs: ARM defines APIs that allow programmers to write agents into an application so that network managers and webmasters can monitor it for a range of performance metrics, including response time. It is a complete offer to application management. But it requires rewriting of existing code that many companies are unwilling to do.

Figure 3.11.19 shows the locations of these tools and agents.
When evaluating products, many components must be factored in. These factors are:

- Customization needs
- Maintenance requirements
- Deployment of code
- Overhead of transmitting measurement data
- Load increase due to synthetic workload
- Reporting capabilities
- Capabilities to solve complex performance problems
- Capabilities to conduct root-cause analysis
- Combination with modeling tools
- Price of the tools

3.11.9.2 Highlighting Bottlenecks

End-to-end service level monitoring is getting extremely popular with Web-based applications. Monitoring is targeting availability and response time measurements. Element-centric management platforms "look down" and manage elements. Response time monitoring tools "look through" the infrastructure from one end to the other.

Applications-related measurements can also be done with RMON probes. The way to do this, according to NetScout Systems Inc., is to track an application on its entire path across enterprise. To support that approach, the remote monitoring vendor is able to collect and report traffic statistics IT managers use to measure how quickly an application makes its round-trip run.

NetScout is leading off its application flow management strategy with a new multiport Fast Ethernet probe, a RMON2 agent for NT servers and Web-based reporting software. Applications can be observed and measured as they run using AppScout, a browser-based solution. AppScout monitors SAP R/3, Microsoft Exchange, Lotus Notes, and TCP/IP applications.

Typical look-through products work on the principle of Java applets in combination with C++ scripts. The code is distributed to various selected end points on the network. These agents generate synthetic transactions against targeted applications, such as databases or intranet Web pages. Response time for these scripted transactions — including the response times over each individual "hop" along the route — are logged on a management server, which assembles and organizes the collected data. The data is then available to users through a client-side Java interface.

The new type of network instrumentation closely mimics the end users' actual experience since it measures the end-to-end responsiveness of an application from one or more outlying LAN nodes to the application server and back again. By doing so, it delivers a metric that accurately reflects application and service performance levels on the network. Trying to gauge the end-to-end performance level of an application over the network by monitoring each distinct element along the service delivery path has not proven successful. Element-specific monitoring is still essential for troubleshooting and maintenance, but network managers have to start looking at some new kinds of instrumentation if they want to view the environment from the end-user's point of view.

3.11.9.3 Tools for Look-through Measurements

Table 3.11.8 lists all the tools that may be considered for end-to-end response time measurements.

TABLE 3.11.8 Tools for Look-through Measurements

Vendor	Product
Avesta	Trinity Measurement Software
Freshwater Software	SiteScope
International Network Services	Enterprise Pro
Jyra Research	Service Management Architecture
NextPoint Networks	NextPoint S3
NetScout Systems	Application Management
Proactive Networks	ProntoWatch
Response Networks	VeriServ

3.11.10 Trends of Intranet Performance Management

Intranet management is an emerging area to webmasters and Web administrators. It combines existing processes for fault, performance, configuration, security, and accounting management with new management tools. Performance and security management are the two most challanging areas. Usage pattens, traffic peaks, unbalanced input/output streams from/to Web servers, server overload, and unstable performance mean challenges to webmasters and network capacity planners. Partitioning networking segments properly, selecting and implementing firewalls, stress testing firewalls and the use of the right authentication techniques mean challenges to security officers of all corporations operating intranets.

New intranet-related management tools are content authoring and auditing instruments, log file analyzers, traffic monitors, load balancers, and application monitors. They can be used individually, or in combination with each other. It is expected that they will soon be integrated into systems and network management platforms.

References

ALDR99 Aldrich, S.: Freshwater's Web Application Management, *Patricia Seybold Group e-Bulletin,* January 21, 1999.

BOBR98 Bobrock, C.: Web developers follow old scripts, *Interactive Week,* November 2, 1998, p. 29.

BOCK98 Bock, G.E.: Microsoft Site Server — Organizing and Sharing the Contents of a Corporate Intranet, *Workgroup Computing Report, Patricia Seybold,* August 1998.

BRUN99 Bruno, L.: *IP Balancing Act: Sharing the Load Across Servers, Data Communications,* February 1999, p. 29.

GIBB98 Gibbs, M.: Pinning down network problems, *Network World,* March 2, 1998, p. 43.

HERM98 Herman, J., Forbath, T.: Using Internet Technology to Integrate Management Tools and Information, http://www.cisco,com/warp/public/734/partner/cmc/bmi_wi.htm.

HUNT96 Huntington-Lee, J., Terplan, K., Gibson, J.: *HP OpenView — A Manager's Guide,* McGraw-Hill, New York, 1996.

JAND98 Jander, M.: Clock watchers, *Data Communications,* September 1998, p. 75–80.

JAND99 Jander, M.: Network Management, *Data Communications,* January 1999, p. 75.

KAPO98 Kapoor, A., Ryan, J.: Reassessing networks for an IP architecture, *Telecommunications,* October 1998, p. 48.

LARS97 Larsen, A.K.: All Eyes on IP Traffic, *Data Communications,* March 1997.

LEIN93 Leinwand, A., Fang, K.: *Network Management — A Practical Perspective,* Addison-Wesley Publishing Company, New York, 1993.

POWE97B Powell, T.: An XML Primer, *InternetWeek,* p. 47–49, November 24, 1997.

REAR98 Reardon, M.: Traffic Shapers: IP in Cruise Control, *Data Communications,* September, 1998, p. 67.

RUBI98 Rubinson, T., Terplan, K.: Network Design — Management and Technical Perspectives, CRC Press, Boca Raton, 1998.

SANT98 Santalesa, R.: Weaving The Web Fantastic — Authoring Tools, *InternetWeek,* November 17, 1997.

SCHU97 Schultz, K.: Two Tools for Monitoring Your Web Site, *InternetWeek,* October 27, 1997, p. 60–61.

STUR98 Sturm, R.: *Working with Unicenter TNG,* QUE Publishing, Indianapolis, 1998.

TAYL96 Taylor, K.: Internet Access: Getting the Whole Picture, *Data Communications,* March 1996, p. 50–52.

TERP96 Terplan, K.: *Effective Management of Local Area Networks,* Second Edition, McGraw-Hill, New York, 1996.

TERP98a Terplan, K.: Web-based systems and network management, Xephon Briefing, London, October 14, 1998.

TERP98b Terplan, K.: *Telecom Operations Management Solutions with NetExpert,* CRC Press, Boca Raton, 1998.

TERP99 Terplan, K.: *Web-based Systems and Network Management,* CRC Press, Boca Raton, 1999.