Integrated Management Architecture for IP-Based Networks

Ammar Rayes and Karen Sage, Cisco Systems

ABSTRACT

IP telephony will bring about a dramatic change in the way IP services are planned, provisioned, managed, and billed. In order to build and retain a strong customer base for these new services, service providers need to meet, if not exceed, the customer expectations set by today's traditional voice services. Acceptance of IP telephony will depend on the quality and efficiency with which service providers offer, deliver, and manage IP services. Installation, configuration, and activation must be rapid and error-free. Furthermore, customers will want direct control over the reconfiguration of services and real-time visibility into the impact change has on their operating costs. Once the service is activated, customers will want the provider to guarantee service quality as defined by industry standards. Corporate customers in particular will need to be assured that the provider is proactively monitoring performance to avoid problems and providing them visibility into the performance data collected. This article discusses an integrated management support system for IP-based networks illustrating the functions needed to support the unique challenges of managing VoIP services. An example of a service management system is also described.

INTRODUCTION

IP networking is a booming market for telecom and datacom service providers and equipment vendors. Service providers are quickly defining and bringing to market differentiated IP services including voice transport, virtual private networks (VPNs), application/policy prioritization, differentiated services (e.g., gold, silver, bronze), multimedia, and transport LANs. Ensuring profitability from these services requires a comprehensive service management architecture that enables service providers to carefully plan, quickly provision, efficiently operate, and accurately bill these services.

Integrated management architecture leverages common data and human resources across application functions, different services, and heterogeneous network technologies. The requirements for each individual management function and service offering, however, may vary. For example, unlike data-transfer-based applications such as e-mail, voice transport across an IP-based infrastructure demands low latency and jitter. For example, the round-trip time (RTT), which is the time required by a network to travel from the source to the destination and back, including the time to process the message and generate a reply, should to be less than 250 ms. Not only will the performance threshold triggers be different, the servicing of these packets will be done based on a priority scheduling scheme.

User Datagram Protocol (UDP) and Real-Time Transport Protocol (RTP) are typically used to transfer VoIP packets. UDP is a connectionless transport layer protocol in the TCP/IP protocol stack. It is a simple protocol that exchanges datagrams without acknowledgment or guaranteed delivery, requiring that error processing and retransmission function be handled by other protocols. RTP is an IPv6 protocol, which is designed to provide end-to-end network transport functions for applications transmitting real-time data, such as voice, video, or simulation data, over multicast or unicast network services. RTP provides services such as payload type identification, sequence numbering, time stamping, and delivery monitoring of real-time applications.

The particular scheme used and the type of traffic will determine the corrective measures needed to fix a problem. Examples of priority scheduling systems include expedited forwarding as defined in the differentiated services (DiffServ) framework, and Priority Queuing with Class-Based Weighted Fair Queuing. Integrated management architecture helps mediate these differences, limiting exposure to the differences for management operators and end customers.

Figure 1 illustrates the typical structure of an IP network including core, edge, and access subnetworks. The backbone technology, in this particular example, is assumed to be based on multiprotocol label switching (MPLS) such as core and edge label switch routers. MPLS networks integrate IP routing protocols directly over asynchronous transfer mode (ATM) switches, allowing efficient support of services such as IP multicast, IP class of service, and IP VPNs. IP services can also be offered over ATM networks, synchronous optical network/synchronous digital hierarchy (SONET/SDH), or directly over wavelength-division multiplexing (WDM). Customer sites are connected to the edge network, which is typically owned and controlled by a service provider.

The acronym FCAPS is often used in the literature to refer to integrated management of various types of networks. It refers to the open systems interconnection (OSI) five functional

0163-6804/00/$10.00 © 2000 IEEE

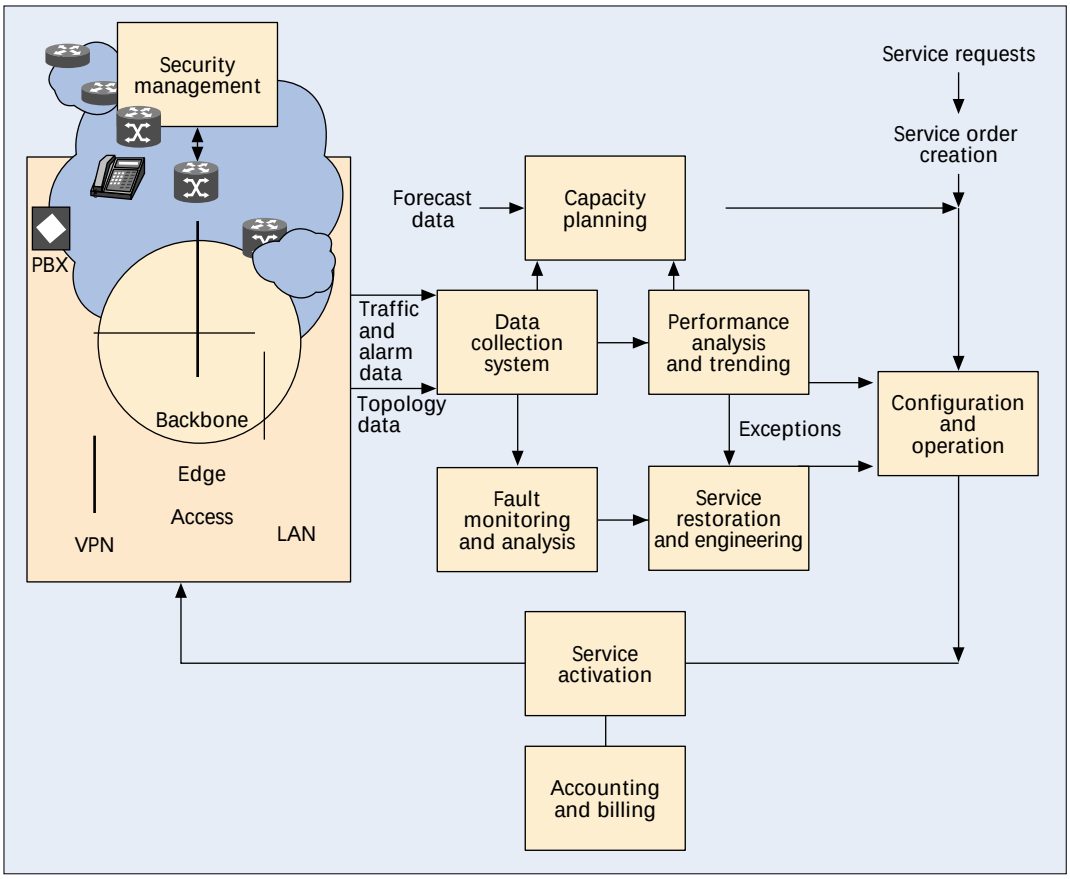

Figure 1. *FCAPS extensions for IP networking.*

Integrated management architecture leverages common data and human resources across application functions, different services, and heterogeneous network technologies. However, the requirements for each individual management function and service offering may vary.

areas: fault management, configuration management, accounting management, performance management, and security management [1]. This article discusses an extended framework of the FCAPS functions for IP-based networks. The extended functions, as shown in Fig. 1, include service restoration, traffic engineering, data collection, service activation, and network planning.

Capacity planning provides a long-term view of network demands and requirements. It computes network element growth rates and generates a long-term capacity expansion plan. The network administrator who wants to do hypothetical (what-if) studies typically carries out the capacity planning function to determine the required capacity as well as optimal equipment locations (or homing arrangements) based on forecast or expected demands.

Once the network is in place, the data collection function collects and forwards data on a regular basis to the appropriate module. For instance, alarms and related fault statistics data are forwarded to the fault module to provide comprehensive diagnosis capabilities, including alarm coronations and pinpointing. Traffic statistics data is typically forwarded to the performance module for data analysis and detection of potential performance exceptions (based on the collected and trending data). Traffic statistics as well as network topology data may also be forwarded to the planning module to estimate the base exogenous traffic [2]. Estimated growth factors can then be used to estimate the future forecasting traffic loads. Finally, traffic statistics can also be used as the basis to observe traffic trends and estimate the load for use in network engineering. It should be noted that the data collection functionality has been implemented in several element management systems, especially for ATM and IP networks.

Performance management is the process of converting raw IP traffic measurements into meaningful performance measures. It can be divided into real-time (or near-real-time/short-term) and long-term management. Real-time performance management typically includes snapshots of the behavior of bottleneck network elements (e.g., backbone link elements that affect the operation of the entire network) as well as mission-critical applications. It is intended for network operators to make certain that network capacity is used efficiently by the mission-critical applications most important to the business (e.g., the most profitable services, network control missions, critical backbone links).

The real-time performance management process is a mechanism to guarantee that enough bandwidth is reserved for time-sensitive IP voice traffic while other applications sharing the same link get their fair share without interfacing with the mission-critical traffic. Another example of real-time performance management is constant monitoring of high-priority customer services (e.g., gold) as well as customers who have been complaining about the performance of their services.

Long-term performance management, on the other hand, supports studies that monitor the ability of the existing IP networks to meet ser-

SLA reports are intended to correlate fault and performance data, and then provide end users and network operators the freedom to establish quality and grade of service objectives specific to their applications.

vice objectives. The purpose of this type of study is to identify situations where corrective planning is necessary. This is needed when objectives are not being satisfied and, where possible, to provide early warning of potential service degradation so that a corrective plan can be formulated before service is affected.

Typically, raw traffic measurements are collected, validated by data collection systems (or element management systems), and then stored in batch mode in a database. One of the most critical steps in developing comprehensive management methods is to define the required traffic measurements that are the basis for performance, fault, service-level agreement (SLA), and traffic engineering algorithms. Examples of IP performance raw traffic measurements include:

- Number of packets received per interface
- Number of packets transmitted per interface
- Number of packets dropped due to mild congestion[1] per interface
- Number of packets dropped due to severe congestion per interface
- Number of packets dropped due to protocol errors
- Amount of time a network element is in a mild congestion state
- Amount of time a network element is in a severe congestion state
- Number of times a network element enters a mild congestion state
- Number of times a network element enters a severe congestion state

The performance management process then converts the validated raw measurements into meaningful network element loads (utilization, packet loss ratio, delay, jitter, etc.). Next, it calculates statistics to characterize the load for traffic engineering purposes (e.g., average peak values, average busy season). The process then computes network element performance measures (route delay, end-to-end packet loss, average and peak packet loss, etc.) based on the characteristic engineering loads. Finally, the performance management process compares the calculated performance results for the short and long terms with the service objectives to identify service or performance exceptions.

The fault management process is similar to the real-time performance process except that it uses the collected alarms and fault statistics to detect and correct problems by pinpointing and correlating faults through the system. It simplifies the service provider's ability to monitor customer services by providing the status of the subscribed services. The ability to monitor a service inherently includes all the network elements that constitute it.

SLA reports are intended to correlate fault and performance data, and then provide end users and network operators the freedom to establish quality and grade of service objectives specific to their applications. SLA reports are intended to be tailored to a specific customer or organization. Examples of IP SLA metrics include:

- Service availability: the percentage of time each polled element was active and running
- Network latency: elapsed time between receipt of the last bit in a frame at network ingress to delivery of the first bit in the same frame at network egress
- Jitter: variation in network latency
- Response time: measures how quickly the network moves information
- Loss ratio: percentage of sent frames discarded or not received
- Mean time to repair: average down time (from when an outage is detected until it is reported fixed)
- Mean time between failures: average down time between consecutive failures
- Throughput: total traffic volume (usually in bits per second)
- Network uptime: percent of time the network is operating without a "hard" failure, usually better than 99.9+ percent

Traffic engineering is perhaps the most challenging function of the management process for IP networks. It represents the action that the network (or network administrator) should consider in order to relieve a potential servicing problem before the service is affected. This may include rehoming, rerouting, load balancing, and congestion control. Traffic engineering is also an essential input for capacity expansion,[2] network dimensioning, and network planning.

The development of appropriate models for traffic engineering depends primarily on clear understanding of quality and grade of service requirements, and the statistical characteristics of the traffic. While there are more than a hundred years of experience in traffic engineering circuit-switched networks, engineering IP-based networks is new. Traffic engineering functionality has been added through the use of tunneling mechanisms or forced route algorithms.

Several traffic models and network dimensioning methods for packet networks have been proposed in the literature [3]. In general, the models can be divided into two categories: those that exhibit long-range dependency (e.g., the fractional Brownian motion model, on/off model with heavy-tailed distributions for the on/off duration, and M/Pareto/∞ models); and Markovian models that exhibit only short-range dependence (e.g., on/off models with exponential on/off distributions, Markov-modulated Poisson process, and Gaussian auto-regressive models, which typically have exponentially decaying correlation functions). The on/off model has been proposed to model voice-over-IP (VoIP) calls with alternating active periods (talk spurts) and silent periods. The parameters of the on/off models can be estimated from actual traffic traces or by using typical default values.

Finally, traffic engineering methods depend on the function of the network element. For instance, traffic techniques for IP edge routers include packet classification, admission control, and configuration management, whereas congestion management and congestion avoidance are typical considerations of backbone routers or switches.

Configuration management deals with the physical and geographical interconnections of various IP network elements such as routers, switches, multiplexers, and links. It includes the

[1] *Mild and severe congestion states are typically defined by the network administrator for each service. For instance, the thresholds for gold IP services should be lower than bronze and silver services so that network operators have more time to react to any potential problems before gold services are affected.*

[2] *Capacity expansion typically is an optimization process that involves a set of algorithms which determine the required network resources (capacity) to meet a specific set of performance objectives.*

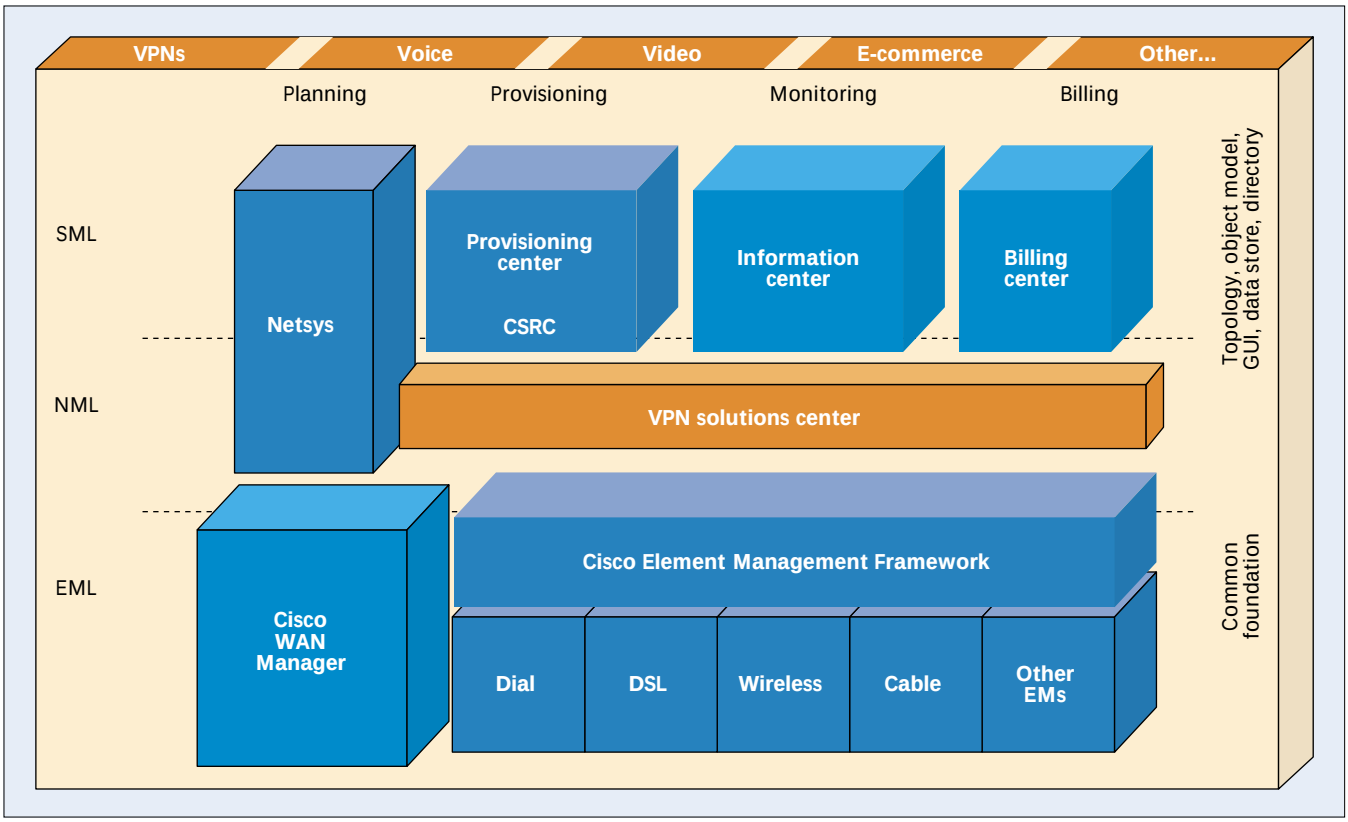

Figure 2. *An example: Cisco Service Management Systems.*

procedure for initializing, operating, setting, and modifying the set of parameters that control the day-to-day operation of the networks. Configuration management also deals with service provisioning, user profile management, and collection of operational data, which is the basis for recognizing changes in the state of the network.

The main functions of configuration management are creation, deletion, and modification of network elements and network resources. This includes the action of setting up an IP network or extending an already existing network, setting various parameters, defining threshold values, allocating names to managed IP objects, and taking out existing network elements.

Security management includes authentication, authorization, and other essential secure communications issues. Authentication establishes the identity of both the sender and the receiver of information. Integrity checking of confidential information is often done if the identity of the sending or receiving party is not properly established [5]. Authorization establishes what a user is allowed to do once the user is identified. Authorization usually follows any authentication procedures.

Issues related to authentication and authorization include the robustness of the methods used in verifying an entity's identity, the establishment of trusted domains to define authorization boundaries, and the requirement of namespace uniqueness.

Billing and accounting management deals with the generation and processing functions of end-user usage information [6]. This includes measuring the subscribers (and possibly the network resources for auditing purposes) and managing call detail information generated during the associated call processing. Of growing importance in IP networks are the records created in the application servers. Such records are the source of content and services delivered by the network. Billing data collection and mediation systems between the IP architecture and the extant billing platforms may aggregate usage-related raw data and produce usage detail records. The access usage detail data can then be transferred to a billing system to render invoices to the subscribers that use IP services. Fraud detection and subscriber-related profile information, such as authorization to charge, is also a function of accounting and billing management.

Several billing approaches have been considered for IP networks ranging from flat rate, such as the voice-world call detailed record (CDR) approach, to full IP-usage-based. A known working IP usage approach is to integrate the IP rated records with existing telephony customer care and billing systems. Such arrangements exist and are fully functional in both the United States and Europe.

It should be mentioned that IP, as a connectionless protocol, does not have an elapsed time or distance-sensitive component. This is not to say that time of day is irrelevant for discounting and promotional considerations, when it clearly is. However, when a customer clicks on a URL in France after a search, the cost does not go up. Additionally, long distance tariffs are being bypassed with the Internet-based backbone.

The Provisioning Center is a service activation system that provisions layers 2 and 3 of the OSI model. It provides an integrated provisioning solution for network service providers who offer IP-based as well as other services, including Internet access, IP-VPN, ATM, frame relay and data link switching.

Many believe that voice will eventually be offered free, in addition to other value-added IP-based services.

Another important requirement for IP billing and accounting functions is an interface to the SLA profiles, and the resulting performance and fault reports. This includes the generation of automatic credits for customers that their SLA agreements were validated.

EXAMPLES

Figure 2 shows an example of service management systems for IP-based networks. It consists of applications that provide end-to-end services across a set of network technologies and service-defined domains delivered on top of element management systems. In this section we highlight the IP-related functions of service management systems.

NETWORK PLANNING

Netsys is a network planning system that automatically generates a comprehensive topology and paths across the network. This information can then be used to interactively plan for network expansion and migration, and the introduction of new services. It can also be used to detect and correct configuration errors, and allows offline changes to be made to the network in order to determine how these changes will affect the network before the actual changes are made.

NETWORK PROVISIONING

The provisioning process includes service activation and subscriber and access registration. The Provisioning Center is a service activation system that provisions layers 2 and 3 of the OSI model. It provides an integrated provisioning solution for network service providers who offer IP-based as well as other services, including Internet access, IP-VPN, ATM, frame relay, and data link switching. The Provisioning Center supports a generic southbound interface to various element management systems as well as northbound flowthrough to enable integration with existing management systems: order management, billing, capacity planning, and others. It supports service customization so that service providers can quickly define and deploy new services. The Provisioning Center provisions services across multiple network elements and displays the results for the network operators.

The Subscriber Registration Center (CSRC) is a suite of products used to configure and manage broadband modems, and to enable and administer subscriber self-registration. CSRC is a directory-enabled solution that consists of a User Registrar for subscriber provisioning and administration, a Modem Registrar for cable modem management, Network Registrar™ technology for domain name system (DNS) and Dynamic Host Configuration Protocol (DHCP) services, and Access Registrar™ technology for broadband telco return.

Access Registrar technology provides RADIUS services for deployment of high-speed data and roaming services. It returns RADIUS configuration parameters to network access server (NAS) clients based on per-subscriber policies.

NETWORK MONITORING

As service providers endeavor to provide VoIP and other New World services, service-level management (SLM) tools that enable performance guarantees are an essential purchase factor. SLM tools enable service providers to demonstrate to customers that contracted SLAs are being satisfied. Together with the Performance SLM and Info Center, the Service Operations and Assurance module offers tools that effectively and efficiently monitor performance and faults in the network to improve service quality to users.

The Performance SLM solutions consist of two main components: statistics available in embedded intelligent agents/management information bases (MIBs), and element management systems, collection points, and end-to-end performance reporting applications. The types of statistical data available fall into three main categories: element-level statistics from routers, switches, and access devices; traffic-flow-based statistics from NetFlow collectors and remote monitoring probes; and response time statistics from the Ping MIB and service-level assurance agent. Leveraging the reporting technology of third-party vendors makes it possible for service providers to guarantee performance on IP, ATM, and frame relay networks.

Info Center provides an integrated family of products that provide multitechnology SLM and assurance capabilities for service providers. Info Center consolidates fault and alarm information from multiple sources and technologies, filters and correlates the information in a real-time database, and distributes it to custom views of the network. It is a service-level monitoring and diagnostics tool that provides network fault and performance monitoring, trouble isolation, and real-time SLM for large networks. It is designed to help operators focus on important network events, offering a combination of alarm reduction rules, filtering, customizable alarm viewing, and partitioning. Info Center works in conjunction with network element management systems such as WAN Manager to provide fault and alarm management across LANs and WANs.

THE BILLING AND ACCOUNTING CENTER

The service management systems offer a suite of data collector products designed to aggregate raw data and produce mediated usage detail records. NetFlow Collector, for example, supports methods to aggregate large volumes of raw usage data. The billing center will access usage details through open interfaces designed to enable rapid integration and third-party development. Key functional areas targeted by the billing center include billing, subscriber management/customer care, and trouble ticketing.

NETWORK AND ELEMENT MANAGEMENT

The Voice Manager is a Web-based solution to manage applications for configuring, monitoring, and diagnosing (on the element management layer) VoIP applications across multiple routers.. It detects voice support and provides call detail information including call quality (e.g., end-to-

end call data record with time stamps and call details such as source, destination, and duration of call quality), call history, quality of voice exception reports, active calls reports, and network delay reports.

The WAN Manager is a network and element management system that addresses operations, maintenance, and management of WAN multiservice networks. Core features include topology management via real-time topology displays, connection management, performance management (real-time statistics data collection and reporting), and element management (configuration and monitoring of network elements). In addition, the WAN Manager provides standards-based interfaces to facilitate automated, flow-through, programmatic interactions with external systems.

The Element Management Framework (CEMF) is the new homogeneous element management layer of the service management system. It allows service providers to install the mix of element managers they need to support their dynamic businesses. It also enables rapid development and deployment of new element managers, permitting service providers to more rapidly introduce and manage new services.

The CEMF provides common interfaces and element management services to applications on the network and service management levels. Solution integrators can also build third-party element management systems to provide element management support in mixed network environments.

CONCLUSION

This article describes an integrated and extended management framework for IP-based networks, including fault management, configuration management, accounting management, performance management, and security management as well as service restoration, traffic engineering, data collection, service activation, and network planning. The article discusses an example of a service management system to plan, provision, operate, and bill for a wide range of IP differentiated services.

REFERENCES

[1] H. Hegering, S. Abeck, and B. Neumair, *Integrated Management of Networked Systems*, Morgan Kaufmann, 1998.

[2] M. Guizani and A. Rayes, *Designing ATM Switching Networks*, McGraw-Hill, 1998.

[3] N. L. S. Fonsenca and M. Zukerman, "ATM Dimensioning and Traffic Management and Modeling Tutorial," *IEEE GLOBECOM*, Phoenix, AZ, Nov. 1997.

[4] W. E. Lenals *et al.*, "On the Self-Similar Nature of Ethernet Traffic," *IEEE/ACM Trans. Networking*, vol. 2, no. 1, Feb. 1994, pp. 1–5.

[5] M. Kaeo, *Designing Network Security*, Macmillan Technical.

[6] S. Aidarous and T. Plevyak, *Telecommunications Network Management*, IEEE Press, 1998.

BIOGRAPHIES

AMMAR RAYES (rayes@cisco.com) is a solutions manager and principal solutions architect at Cisco Systems. He is currently working on network management/operation support systems for IP wireless, fixed broadband wireless, IP, ATM, VoIP, and WDM networks. Prior to joining Cisco Systems, he was a director at Telcordia Technologies (formally Bellcore). He has authored/co-authored over 30 books, patents, and papers on advances in numerous communications-related technologies including a recent book on ATM switching and network design published by McGraw Hill. He received his B.S. and M.S. degrees in electrical engineering from the University of Illinois at Urbana in 1986 and 1988, respectively. He received his Doctor of Science degree in electrical engineering from Washington University, St. Louis, Missouri, in 1994, where he received the Outstanding Graduate Student Award in Telecommunications. He served as a board member of IEEE Computer Society Press and is currently a member of the International Advisory Committee of the IEEE IP-Oriented Operations and Management Workshop (IPOM). He is also the industry Liaison Chairman of the International Conference on Parallel and Distributed Computing and Systems.

KAREN M. SAGE is currently a senior manager for architecture and technical marketing of service provider network management systems at Cisco Systems. Prior to this position she was an engineering manager of NETSYS Performance Management Tools, also at Cisco Systems. She came to work for Cisco through Cisco's acquisition of NETSYS Technologies, Inc. where she was principal architect of the NETSYS Enterprise Performance tools. Her pioneering work in the communications field began with Master's and doctoral research at the University of Virginia. Her seminal work in the areas of performance analysis and design of communication networks continues to be the foundation for many products on the market today. She is also the published author of several press articles and technical papers. In addition, she is a reviewer for several journal and book publishers, and an invited speaker at universities around the world.

The Element Management Framework allows service providers to install the mix of element managers they need to support their dynamic businesses, and enables rapid development and deployment of new element managers.