CHAPTER 27

Overview of Cisco Unified Communications Operations and Serviceability

Revised: February 29, 2012; OL-27282-05

Once the network, call routing, call control infrastructure, and applications and services have been put in place for your Cisco Unified Communications System, network and application management components can be added or layered on top of that infrastructure. There are numerous operations and serviceability applications and services that can be deployed in an existing Cisco Unified Communications infrastructure. These applications and services can be classified into four basic areas:

- User and Device Provisioning Services — Provide the ability to centrally provision and configure users and devices for unified communications applications and services.
- Voice Quality Monitoring and Alerting — Provide the ability to monitor on an ongoing basis various call flows occurring within the system to determine whether voice quality is acceptable and to alert administrators when the voice quality is not acceptable.
- Operations and Fault Monitoring — Provides the ability to centrally monitor all application and service operations and to issue alerts to administrators regarding network and application failures.
- Network and Application Probing — Provides the ability to probe and collect network and application traffic information at various locations throughout the deployment and to allow administrators to access and retrieve this information from a central location.

This part of the SRND covers the applications and services mentioned above. It provides an introduction to the various network management applications and services, followed by discussions surrounding architecture, high availability, capacity planning, and design considerations. The discussions focus on design-related aspects of the applications and services rather than product-specific support and configuration information, which is covered in related product documentation.

This part of the SRND includes the following chapters:

- Network Management, page 28-1

  This chapter examines unified communications network and application management services, a common and prevalent set of services within most unified communications deployments, which allows administrators to provision and configure users and devices, monitor network and application operations as well as voice quality, and receive alerts and alarms when issues arise. This chapter also examines the impact of these management applications and services on deployment models and provides design and deployment best practices for network and application management services and applications.

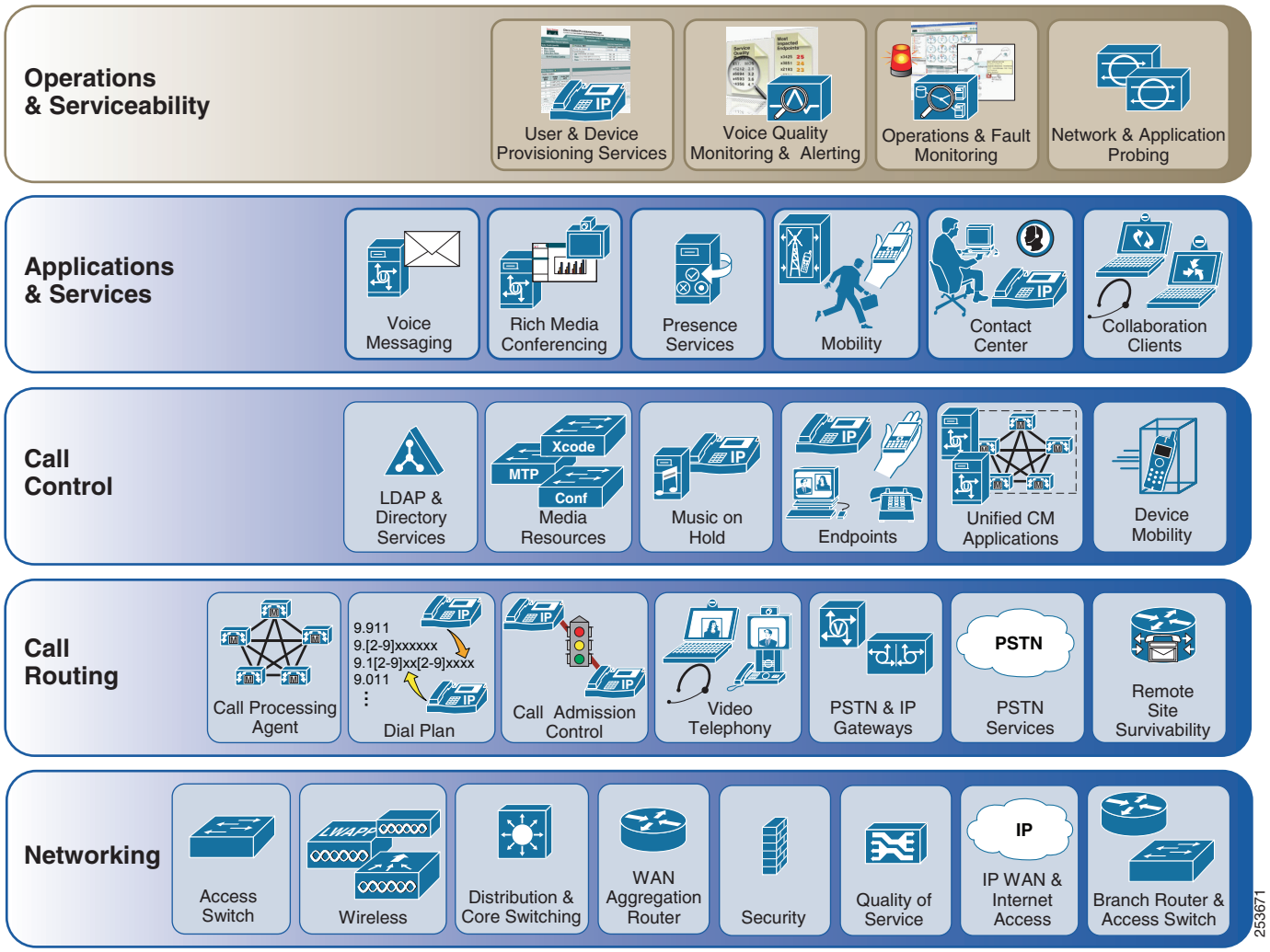
- Unified Communications Design and Deployment Sizing Considerations, page 29-1

  This chapter discusses the sizing of individual Unified Communications components as well as systems consisting of several components communicating with each other. This chapter also discusses the performance impact of the different functions that the various Unified Communications products support, and it explains why "designing by datasheets" is not the preferred way to deploy a complex Unified Communications network. In addition, this chapter provides insights on how to work with the various sizing tools available, mostly notably the Cisco Unified Communications Sizing Tool.

Architecture

As with other network and application technology systems, operations and serviceability applications and services must be layered on top of the underlying network, system, and application infrastructures in order to be able to monitor and control these infrastructures. Figure 27-1 shows the logical location of unified communications operations and serviceability in the overall Cisco Unified Communications System architecture.

Figure 27-1 Cisco Unified Communications Operations and Serviceability Architecture



Unified communications operations and serviceability services such as user and device provisioning, voice quality monitoring and altering, operations and fault monitoring, and network and application probing, all rely on the underlying network infrastructure for network connectivity for various operations and serviceability applications and probes. While there is no direct reliance on the unified communications call routing, call control infrastructure, or unified communications applications and services, these infrastructures and applications are what the various operational and management services actually manage and configure. For example, user and device provisioning services as well as various monitoring and alerting services leverage the network infrastructure for connectivity to various

unified communications applications and service nodes in order to configure and monitor various components and operations. These same services also communicate directly with, and in some cases change configurations on or receive alerts from, components such as call processing agents, PSTN and IP gateways, media resources, endpoints, and various unified communications applications for voice messaging, rich media conferencing, and collaboration clients. In addition to relying on these infrastructure layers and basic unified communications services and applications, services pertaining to operations and serviceability are also often dependent upon each other for full functionality.

High Availability

As with network, call routing, and call control infrastructures and critical unified communications applications and services, unified communications operations and serviceability services should be made highly available to ensure that required provisioning, monitoring, and altering will continue even if failures occur in the network or applications. It is important to understand the various types of failures that can occur as well as the design considerations around those failures. In some cases, the failure of a single operations and management application or server can impact multiple services because the unified communications operations and serviceability components are dependent on other components or services. For example, while the various application service components of a network management deployment might be functioning properly, the loss of network connectivity to, or a failure of, a network probe would effectively eliminate the ability to monitor network health or voice quality unless redundant network probes had been deployed along with alternate paths of connectivity.

For operations and serviceability functions such as user and device provisioning, high availability considerations include temporary loss of functionality due to network connectivity or application server failures resulting in the inability of administrators to provision users and devices or to make changes to these user account or device configurations. In addition, failover considerations for these types of operations include scenarios in which portions of the functionality can be handled by a redundant operation or management application that allows administrators to continue to facilitate some configuration changes in the event of certain failures.

High availability considerations are also a concern for operations and serviceability applications that provide services such as voice quality monitoring or application and operations fault monitoring. Interrupted network connectivity or server or application failures will typically result in a reduced ability to monitor and/or alert, and in some cases complete loss of such functionality. For voice quality monitoring, this can mean that voice quality measurements for some call flows or devices will be unavailable. For operations and fault monitoring services, high availability considerations include the potential for loss of operational change tracking data or fault alerts and indications.

Capacity Planning

Network, call routing, and call control infrastructures as well as unified communications applications and services must be designed and deployed with an understanding of the capacity and scalability of the individual components and the overall system. Similarly, deployments of operations and serviceability components and services must also be designed with attention to capacity and scalability considerations. When deploying various operations and serviceability applications and components, not only is it important to consider the scalability of these applications themselves, but you must also consider the scalability of the underlying infrastructures. Certainly the network infrastructure must have available bandwidth and be capable of handling the additional traffic load these operations will create. Likewise, the call routing and control infrastructure must be capable of handling required inputs and outputs as facilitated by the various operations and serviceability components in use. For example, with operational applications and services such as voice quality monitoring and alerting and operations and fault

monitoring, there are capacity implications for each of these individual applications or services in terms of the number of devices and call flows that can be monitored at a given time, but just as important is the scalability of the underlying infrastructure and monitored applications to handle the added network traffic and connections required for monitoring and alerting. While the monitoring and alerting application or service itself may be able to support the monitoring of many network devices and call flows, the underlying network or devices might not have available capacity to handle the probing connections or the alarm messaging load generated by these monitoring and alerting services.

For operation applications or services that provide user or device provisioning capabilities, capacity planning considerations include things such as ensuring that the provisioning application can handle the requested load and also that user or device provisioning operations not only do not exceed the number of support devices or users for a particular underlying unified communications application or service, but also that provisioning or configuration change transactions do not exceed either the capacity of the underlying network or the rate at which a particular application can handle transactions. In most cases additional capacity can be added by increasing the number of operational provisioning application servers or by increasing the size or number of underlying unified communications applications or service instances, assuming the underlying network and call routing and control infrastructures are capable of handling this additional load.

For a complete discussion of system sizing, capacity planning, and deployment considerations related to sizing, refer to the chapter on Unified Communications Design and Deployment Sizing Considerations, page 29-1.