Unified Communications System Integration within VMDC Architecture

The adaptation of an enterprise-wide IP infrastructure has enabled the migration of a variety of services to the IP network. The Cisco Unified Communications System (UCS) is an example of such a service, enabling data, voice, and video to be transmitted over a single network infrastructure using standards-based Internet. The evolution of virtualized compute systems are enabling greater scalability, added flexibility and ease of management of the various components associated with the Cisco Unified Communications System components. The integration of Cisco Unified Communications System family of products and its various components within existing data center architectures is imperative in achieving a consistent, seamless, secure and highly-available telephony service over enterprise network.

Document Goal

This document describes the manner which Cisco's Unified Communications System can be integrated within the Virtual Multiservices Data Center (VMDC) framework. The Virtualized Multiservices Data Center (VMDC) architecture, focuses on a shared data center infrastructure supporting multiple tenants and secure separation of cloud resources. The flexibility of this architecture allows for seamless integration of services such as UC within its framework.

The rapid adaptation of virtualized cloud infrastructure is accelerating the virtualization of many legacy applications. Cisco's Unified Communications System is such an application. Many components of Cisco Unified Communications family of products offering have been virtualized. The virtualization of these components provides a great opportunity for cloud architects to integrate IP telephony services within the VMDC architecture and take advantage of the numerous VMDC benefits. This document describes the following:

- **Unified Communications System Components—**The functional overview of Unified Communications system and traffic flows between the various functional components.
- **Virtualization Unified Communication Components—**The architectural description of virtualized Unified Communications components.
- **Overview of VMDC Architectural Framework—**The VMDC architectural components, overall framework, and functional capabilities.
- **Unified Communication Deployment Models—**The description of UC deployment models.
- **Overlay of Unified Communication on VMDC—**The mapping of Unified Communication components and its integration within the VMDC architecture.
- **Network Architecture—**Network considerations, such as firewalls, traffic engineering, network virtualization, QOS, load sharing, and redundancy.

- **Management**—System management, logging, system visibility, and secure connectivity of mobile and stationary viewing-stations.
- **Security Considerations**—The secure access, visibility, threat assessment and isolation of the end-to-end IP Telephony Components.

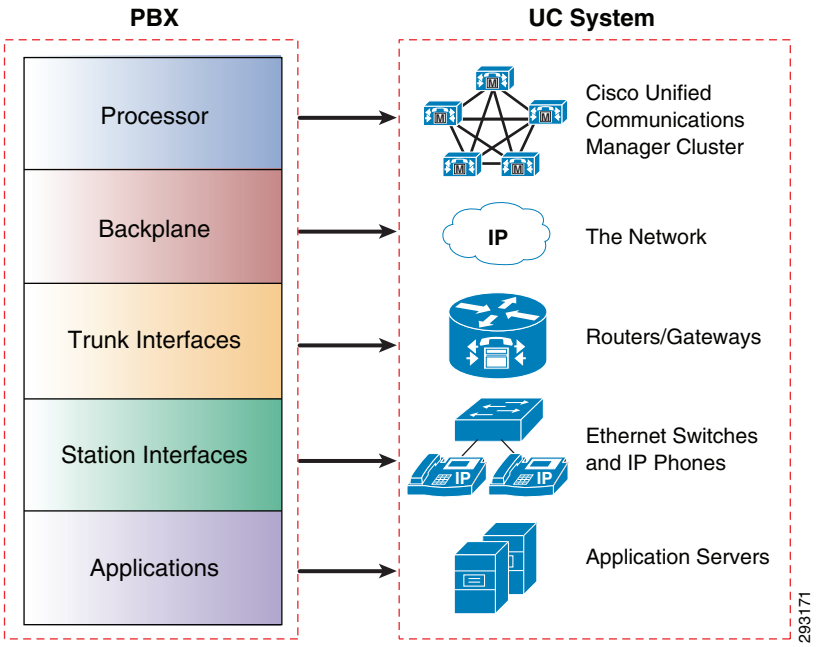
Audience

The target audience for this document includes sales engineers, field consultants, professional services, IT managers, Cisco channel partner engineering staff, and customers who have requirements for a cloud-ready data center and wish to integrate various Cisco Applications such as Unified Communications.

UC Components and Functionality

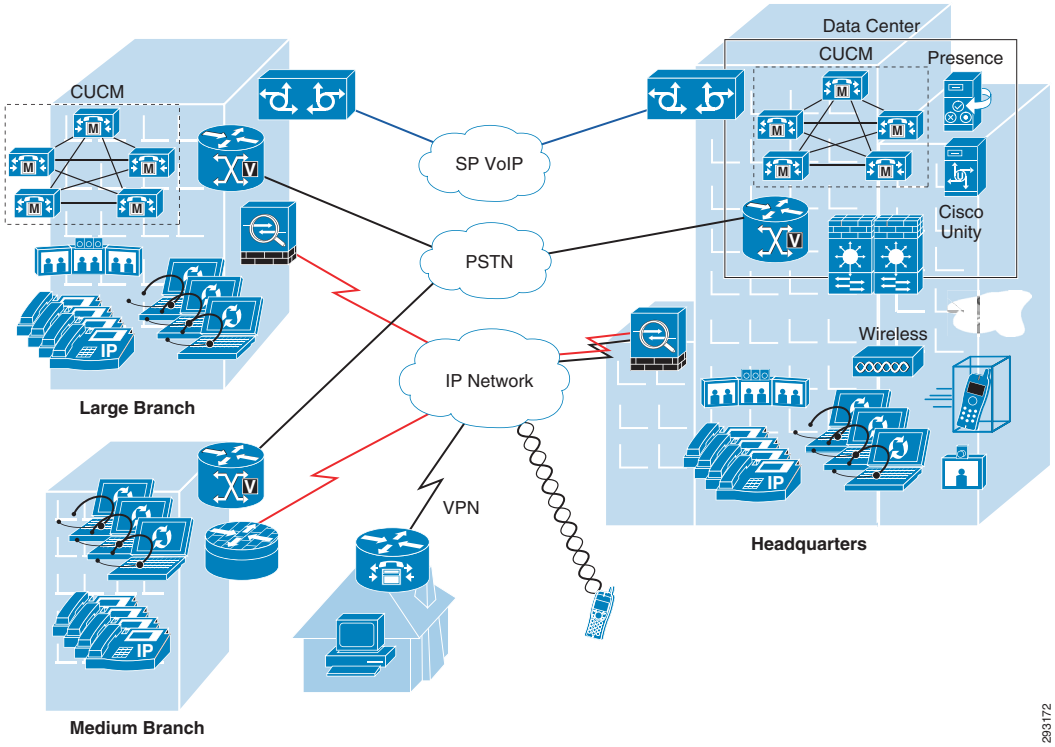
The Cisco Unified Communications System provides IP-based Telephony services. This IP-Based Telephony system replaces traditional PBX systems and consist of various functional components such as Cisco Unified Communication's Manager, Routers and gateways, switches, end point devices and application servers. Figure 1 shows the mapping of these functional components to the traditional PBX legacy systems.

Figure 1 Unified Communications System Components Mapping to PBX



A typical enterprise consists of branches and mobile workers and teleworkers and the main Campus. An IP network, PSTN network or an SP VoIP infrastructure provides connectivity between the various branches, mobile devices, and teleworker to the main campus as shown in Figure 2.

Figure 2 Unified Communication Deployment in a Typical Enterprise



The Cisco Unified Communications System incorporates and integrates the following major communications technologies:

- **IP Telephony**

  IP telephony refers to technology that transmits voice communications over a network using IP standards. Cisco Unified Communications includes a wide array of hardware and software products such as call processing agents, IP phones (both wired and wireless), voice messaging systems, video devices, and many special applications.

- **Customer Contact Center**

  Cisco Unified Contact Center products are a combination of strategy and architecture that promote efficient and effective customer communications across a globally capable network by enabling organizations to draw from a broader range of resources to service customers. They include access to a large pool of agents and multiple channels of communication as well as customer self-help tools.

- **Video Telephony**

  The Cisco Unified Video Advantage products enable real-time video communications and collaboration using the same IP network and call processing agent as Cisco Unified Communications. With Cisco Unified Video Advantage, making a video call is as easy as dialing a phone number.

- **Rich-Media Conferencing**

  Cisco Unified MeetingPlace, Cisco Unified Videoconferencing, and Cisco WebEx Software as a Service enhance the virtual meeting environment with a integrated set of IP-based tools for voice, video, and web conferencing.

- **Mobility**

  Cisco wireless and mobility solutions enable users to increase productivity and responsiveness by enabling access to network resources and applications securely, regardless of location or client device.

- **TelePresence**

  Cisco TelePresence delivers real-time, face-to-face interactions between people and places in their work and personal lives using advanced visual, audio, and collaboration technologies. These technologies transmit life-size, high-definition images and spatial discrete audio that make users feel like they are in the same room even when they are half a world away.

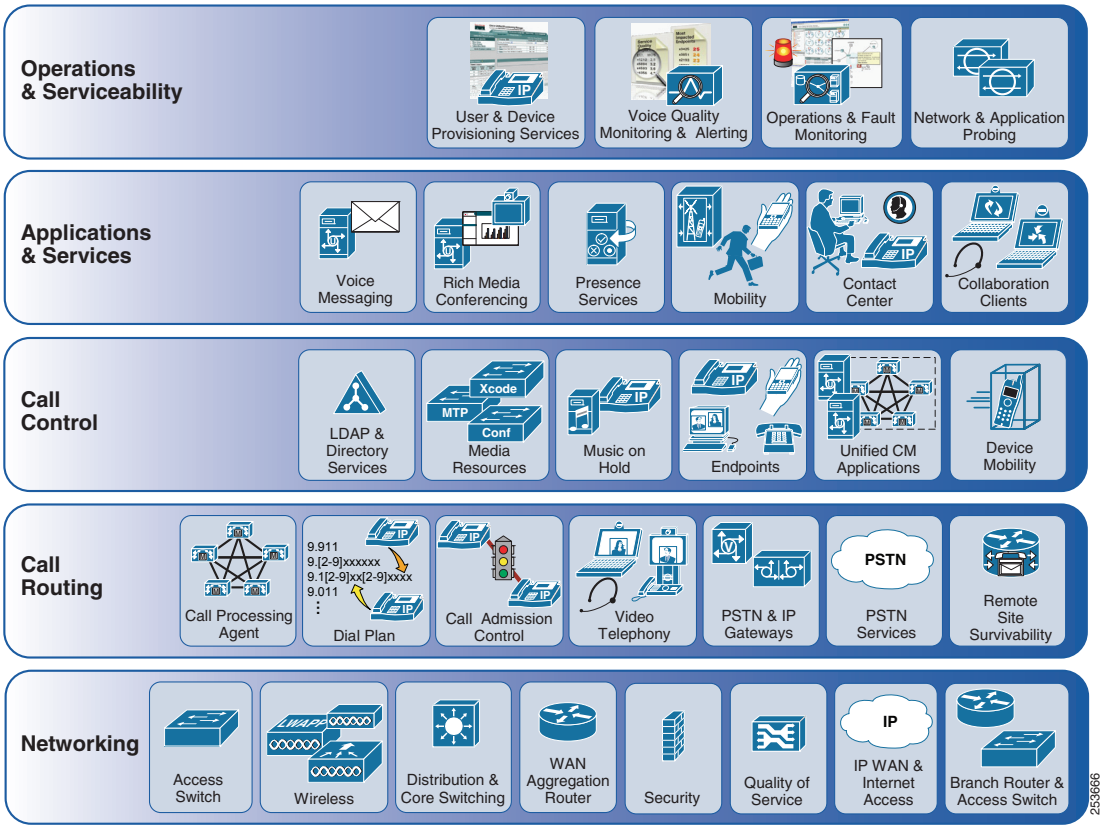
- **Applications**

  Cisco provides numerous embedded applications and also works with leading-edge companies to provide the broadest selection of innovative third-party unified communications applications and products focused on critical business needs such messaging, customer care, and workforce optimization.

Cisco Unified Communications Architecture

Figure 3 shows the layered architecture of Cisco Unified Communications System.

Figure 3 Cisco UC System Layered Architecture



The various layers of the Cisco Unified Communications System perform the following major tasks and roles:

- **Networking**

  This layer forms the foundation for the Unified Communications network. It includes components Such as Quality of Service (Qos) and Network Security.

- **Call Routing**

  This layer handles the processing and routing of calls throughout the system. It includes components such as call processing and routing, dial plan configuration, call admission control and video telephony services.

- **Call Control**

  This layer enables users to initiate and manage calls. It includes components such as LDAP integration, such as unified communication end point feature sets, and device mobility roaming features.

- **Applications and Services**

  This layer contains numerous applications and services that can be deployed on top of an existing Cisco Unified Communications infrastructure to add enhanced user features to the system. It includes components such as voice messaging, mobility services, contact center applications and collaboration client services.

- **Operations and Serviceability**

  This layer contains system-level services for monitoring and managing the Unified Communications network and applications. It includes components such as user and device provisioning, voice quality monitoring, and network and application probing capabilities.

For more information regarding Cisco Unified Communications System, refer to documentation available at the following locations:

- Cisco Unified Communications and Collaboration Solutions Design Guidance
- Cisco Unified Communications System Documentation

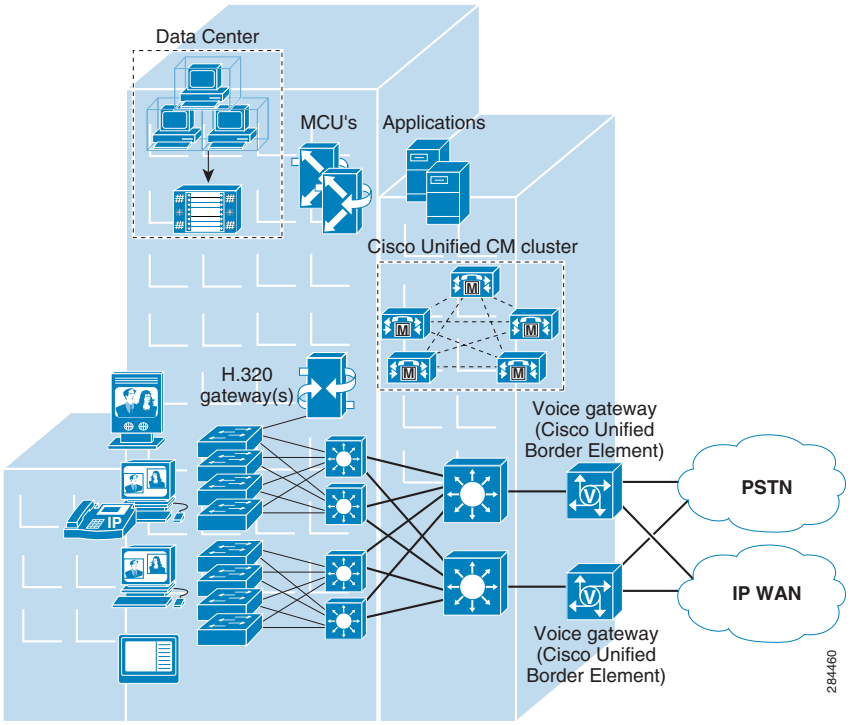
Unified Communication Deployment Models

The Unified Communication uses a variety of deployment models based on the various characteristics of the enterprise network. Reliability requirements, size of the organization, and required services are such considerations that can affect the deployment model deployed within a specific customer enterprise environment. The following is a summary of some of the UC deployment models.

Single-Site Deployment Model

In this call processing deployment model, the Unified Communications services and the endpoints are co-located in the campus, and the QoS-enabled network between the service nodes, the endpoints, and applications is considered highly available, offering virtually unlimited bandwidth with less than 15 ms of latency end-to-end (Figure 4). Likewise, the quality and availability of power are very high, and services are hosted in an appropriate data center environment. Communications between the endpoints traverses a LAN or a MAN, and communications outside the enterprise goes over an external network such as the PSTN. An enterprise would typically deploy the campus model over a single building or over a group of buildings connected by a LAN or MAN.

Figure 4 UC Single-Site Deployment Model

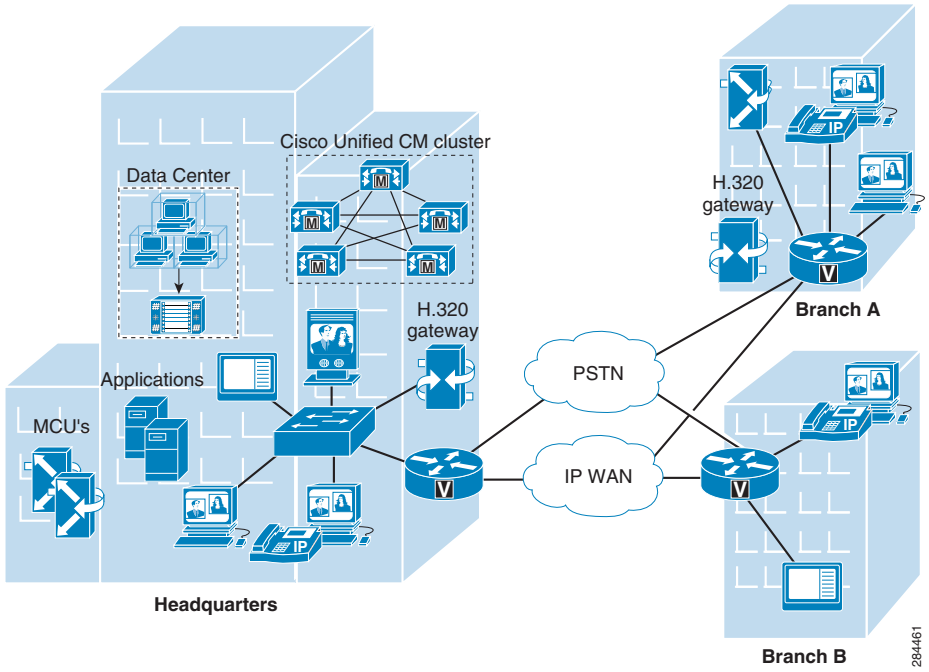


Multi-Site Deployment Model with Centralized Call Processing

In this call processing deployment model, endpoints are remotely located from the call processing service, across a QoS-enabled Wide Area Network. Due to the limited quantity of bandwidth available across the WAN, a call admission control mechanism is required to manage the number of calls admitted on any given WAN link, to keep the load within the limits of the available bandwidth. On-net communication between the endpoints traverses either a LAN/MAN (when endpoints are located in the same site) or a WAN (when endpoints are located in different sites).

The IP WAN also carries call control signaling between the central site and the remote sites. Figure 5 shows a typical centralized call processing deployment, with a Unified CM cluster as the call processing agent at the central site and a QoS-enabled IP WAN to connect all the sites. In this deployment model, other Unified Communications services such as voice messaging, presence and mobility are often hosted at the central site as well to reduce the overall costs of administration and maintenance. In situations where the availability of the WAN is unreliable or when WAN bandwidth costs are high, it is possible to consider decentralizing some Unified Communications services such as voice messaging (voicemail) so that the service's availability is not impacted by WAN outages.

Figure 5 UC Multi-Site Deployment Model with Centralized Call Processing



Multi-Site Deployment with Distributed Call Processing

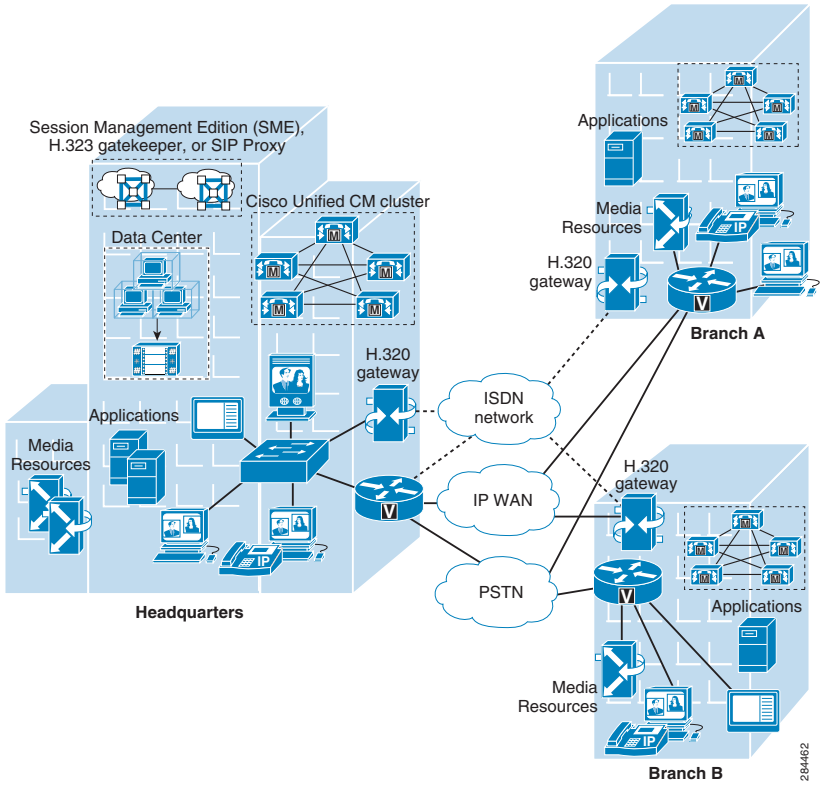
The model for a multi-site deployment with distributed call processing consists of multiple independent sites, each with its own call processing agent cluster connected to an IP WAN that carries voice traffic between the distributed sites. Figure 6 shows a typical distributed call processing deployment.

Each site in the distributed call processing agents can be one of the following:

- A single site with its own call processing agent such as Cisco Unified Call manager, Cisco Unified Communication Manager Express or other IP PBX appliances.
- A centralized call processing site and all of its associated remote sites
- A legacy PBX with Voice over IP (VoIP) gateway

An IP WAN interconnects all the distributed call processing sites. Typically, the PSTN serves as a backup connection between the sites in case the IP WAN connection fails or does not have any more available bandwidth. A site connected only through the PSTN is a standalone site and is not covered by the distributed call processing model. Cisco Unified Communications Manager Session Management Edition clusters, H.323 gatekeepers, or Session Initiation Protocol (SIP) proxy servers can be used to provide intercluster call routing and dial plan aggregation in multi-site distributed call processing deployments.

Figure 6 UC Multi-Site Deployment with Distributed Call Processing

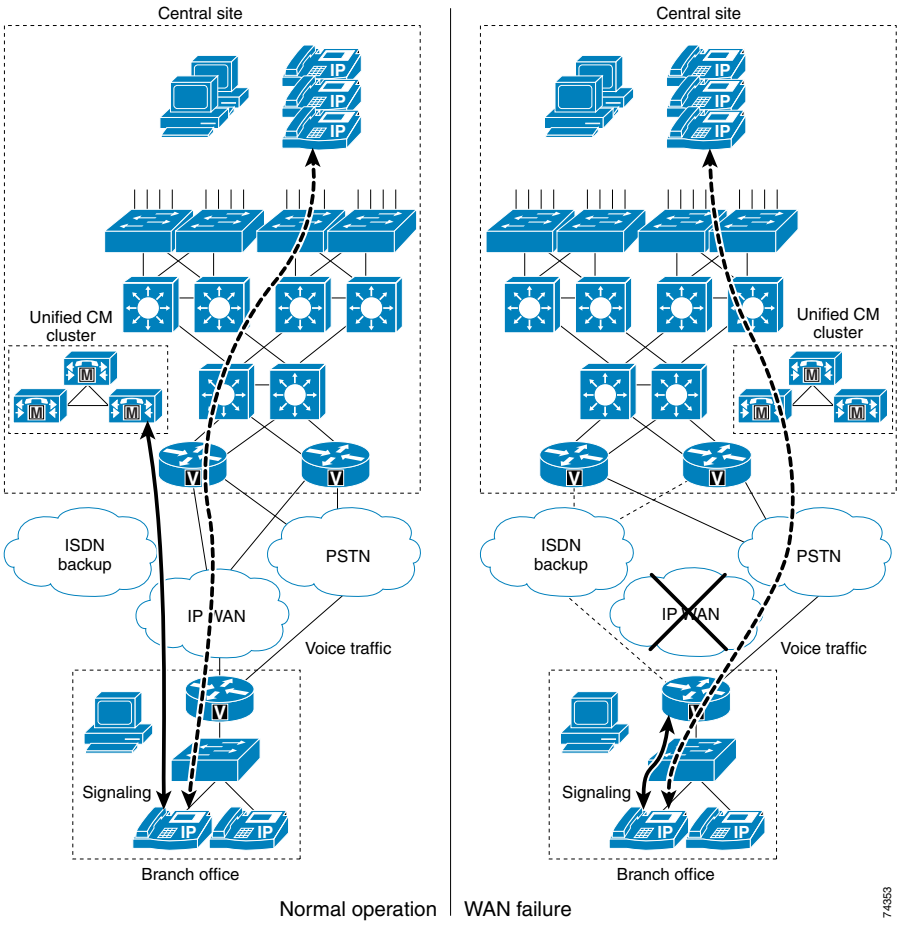


Remote Site Survivability

When deploying Cisco Unified Communications across a WAN with the centralized call processing model, one can take additional steps to ensure that data and voice services at the remote sites are highly available. The choice of one of these strategies may depend on several factors, such as specific business or application requirements, the priorities associated with highly available data and voice services, and cost considerations.

Under normal operations shown in the left part of Figure 7, the branch office connects to the central site via an IP WAN, which carries data traffic, voice traffic, and call signaling. The IP phones at the branch office exchange call signaling information with the Unified CM cluster at the central site and place their calls across the IP WAN. The branch router or gateway forwards both types of traffic (call signaling and voice) transparently and has no knowledge of the IP phones. If the WAN link to the branch office fails, or if some other event causes loss of connectivity to the Unified CM cluster, the branch IP phones re-register with the branch router in SRST mode. The branch router, SRST, or Unified CME running in SRST mode, queries the IP phones for their configuration and uses this information to build its own configuration automatically. The branch IP phones can then make and receive calls either within the branch's network or through the PSTN. The phone displays the message “Unified CM fallback mode,” and some advanced Unified CM features are unavailable and are grayed out on the phone display. When WAN connectivity to the central site is reestablished, the branch IP phones automatically re-register with the Unified CM cluster and resume normal operation.

Figure 7 Normal Centralized Call Processing Operation



Virtualization of Unified Communication Components

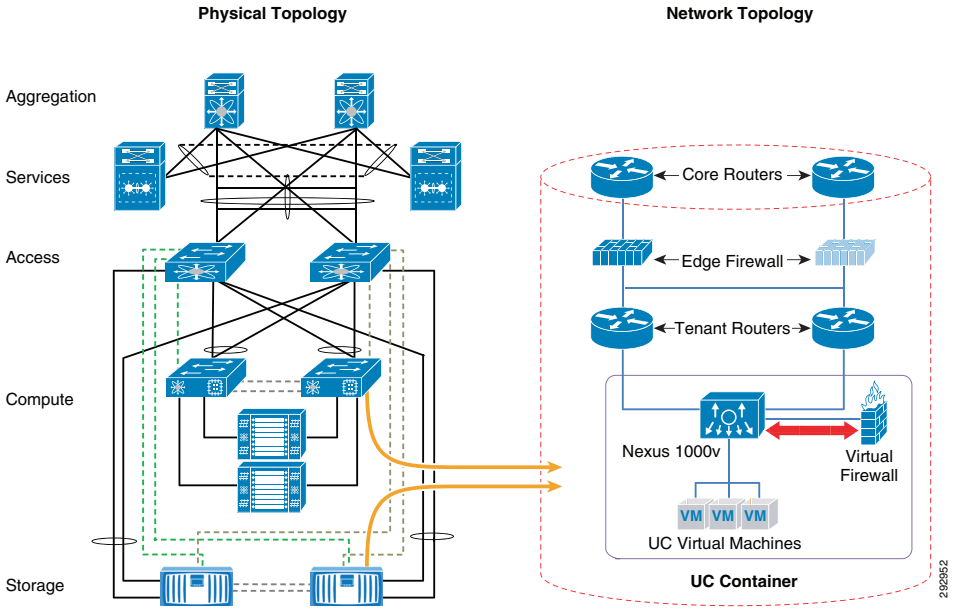
Until recently the implementation of an IP Telephony solution required the deployment of a standalone server to host the Cisco Unified Communication software components such as Cisco Unified Communication Manager (CUCM), Cisco Unity Presence (CUP)**.** The virtualization of these components running on the Cisco Unified Computing System- has significantly improved performance, ease of use, scalability and ease of management. With the deployment CUCM and CUP and other software components as a virtual machine, one can take advantage of all the benefits of a virtualized cloud environment. Some of the advantages include:

- Modular UCS platform building blocks support great scalability, versus dedicated server solutions, allowing physical security personnel to deliver thousands of end devices that can deliver network IP telephony solutions for a variety of use cases.
- Highly secure separation capabilities for multi-tenant environments increase logical security and access control, by utilizing many virtual appliances such as Cisco's Virtual Secure Gateway.
- A virtual infrastructure allows infrastructure architects to easily integrate Cisco's IP Telephony components solution within an existing virtualized infrastructure, and hence providing IP Telephony as a service on the network within a cloud infrastructure.

- A virtualized infrastructure allows deployment of efficient, high-density storage platforms such as industry leading storage options from EMC and NetApp. By using these platforms one takes advantage of storage virtualization capabilities which include, sophisticated data disaster recovery features, remote storage capabilities, logical separation of storage spaces within a multi-tenant environment, and a centralized storage solution for all data center applications.

Figure 8 shows Cisco's Unified Communication components deployment within a virtualized environment.

Figure 8 Cisco Unified Communication Components Deployment in a Virtualized Environment



These are positioned behind the Nexus 1000V virtual switch and protected by the Cisco's VSG virtual firewall. The physical edge firewalls, provide additional inter-tenant security, as well as traffic security and policy enforcement for devices located outside the tenant container. The physical topology shows storage, and UCS fabric connectivity through access switches. The services node provides additional services, such as edge firewall functionality, load balancing and intrusion prevention capabilities.

Table 1 shows the various UC applications that can be implemented in a virtual environment.

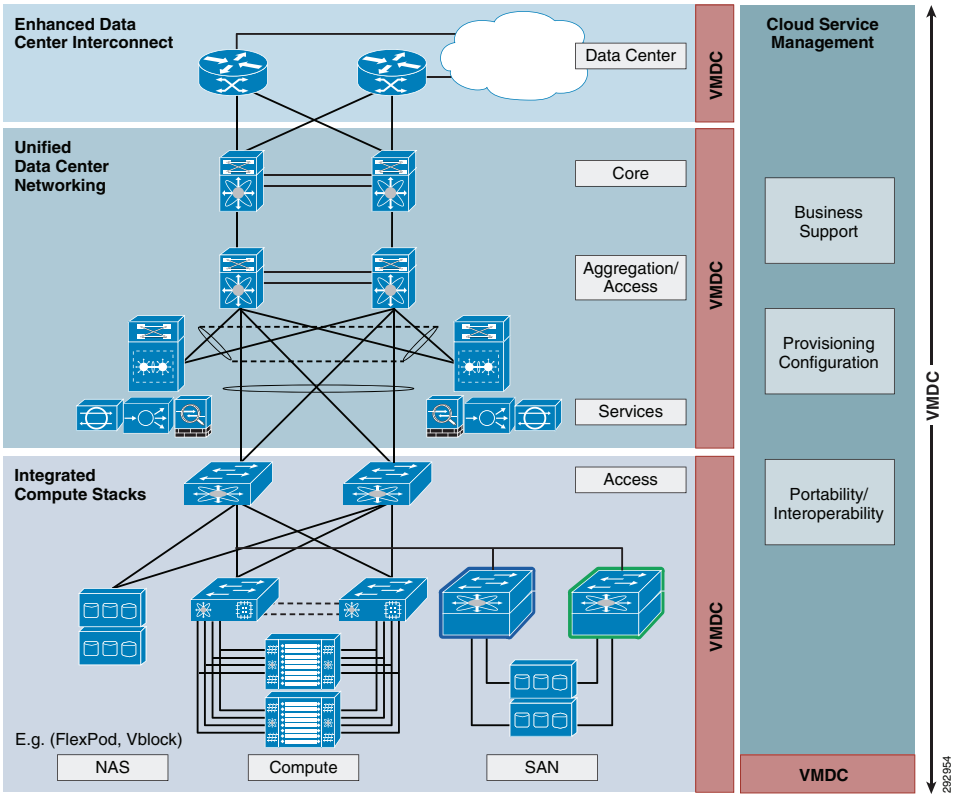
Table 1 UC Applications Implemented in a Virtual Environment

Component	Hardware	Role
CUCM	Virtualized	Call Control
CUCxn	Virtualized	Voicemail/Mailbox
CUP	Virtualized	Presence/IM Control
CER	Virtualized	Cisco Emergency Responder
CUEAC Server	Virtualized	Auto Attendant

VMDC Architectural Overview

The Cisco Virtualized Multiservices Data Center solution provides design and implementation guidance for enterprises deploying private cloud services and service providers building virtual private and public cloud services. The Cisco VMDC solution integrates various Cisco and third-party products that are part of the cloud computing ecosystem. Cisco's VMDC system defines an end-to-end architecture, which an organization may reference for the migration or build out of virtualized, multi-tenant data centers for new cloud-based service models such as Infrastructure as a Service (IaaS). Figure 9 shows the basic architectural framework for VMDC.

Figure 9 Basic VMDC Architectural Framework



VMDC System Overview

VMDC is an end to end system that integrates compute, network and storage components with an architectural frame work, The various design paradigms and functional layer are defined within this framework.

Hierarchical Network Layers

The data center within the VMDC reference architecture is based on the classic multi-layer hierarchical network model. Hierarchical model benefits include scalability, resilience, performance, maintainability, and manageability and its design represents a structured approach to building the infrastructure, allowing for relatively easy expansion in modular increments. Redundant nodes and links

at each level insure no single point of failure, while link aggregation can be engineered for optimal bandwidth and performance through the aggregation and core layers. In general, this hierarchical model uses three layers:

- **Core Layer**—Characterized by a high degree of redundancy and bandwidth capacity and thus optimized for availability and performance.
- **Aggregation Layer**—Characterized by a high degree of high-bandwidth port density capacity and thus optimized for traffic distribution and link fan-out capabilities to access layer switches. Functionally, the nodes in the aggregation layer typically serve as the Layer 2/Layer 3 boundary.
- **Access Layer**—Serves to connect hosts to the infrastructure, providing network access, typically at Layer 2 (L2) (i.e., LANs or VLANs).

VMDC Functional Layers

The VMDC architecture can also be functionally classified into the following categories.

- Network
- Services
- Compute
- Storage
- Management

The **Network** layer includes the WAN/PE router, which forms the data center perimeter to the Enterprise wide area or provider IP backbone, and to the public Internet. These perimeter nodes may be dedicated to Layer 3 routing functions, or may be multiservice in nature, providing Layer 2 interconnects between data centers as well as Layer 3 services. The VMDC topologies support two variants of the three-layer hierarchical model: a collapsed core/aggregation version, and a collapsed aggregation/access version. These allow for fine-tuning of port capacity and bandwidth to the level of aggregation or access density required to accommodate current and anticipated scale requirements.

The **Services** layer comprises network and security services such as firewalling, server load balancing, SSL offload, intrusion prevention, network analysis, and gateway functions. Within the VMDC reference architecture, the Data Center Services Node (DSN) provides firewalling and server load balancing services, in a service module form factor; alternatively, these are available in appliance form-factors. This layer also serves as the termination point for remote access IPSec or SSL VPNs; within the VMDC architecture, the Cisco physical appliances connected to the DSN fulfills this function, securing remote tenant access to cloud resources.

The **Compute** layer includes several sub-systems. The first is a virtual access switching layer, which allows for extension of the Layer 2 network across multiple physical compute systems. This virtual access switching layer is of key importance in that it also logically extends the Layer 2 network to individual virtual machines within physical servers. The feature-rich Cisco Nexus 1000V generally fulfills this role within the architecture. A second sub-system is that of virtual services. These may include security, load balancing, and optimization services. Services implemented at this layer of the infrastructure will complement more centralized service application, with unique applicability directly to a specific tenant or workgroup and their applications. Specific application based services validated within the VMDC architecture currently include the Cisco Virtual Security Gateway (VSG), providing a security policy enforcement point within the tenant virtual data center. The third sub-system within the Compute layer is the computing resource that includes the Cisco Unified Compute System consisting of physical servers, hypervisor software providing compute virtualization abilities, and the virtual machines thus enabled.

The **Storage** layer provides storage resources. Data stores reside in SAN (block-based) or NAS (file-based) storage systems. SAN switching nodes use an additional level of resiliency, interconnecting multiple SAN storage arrays to the compute resources, via redundant FC or Ethernet links.

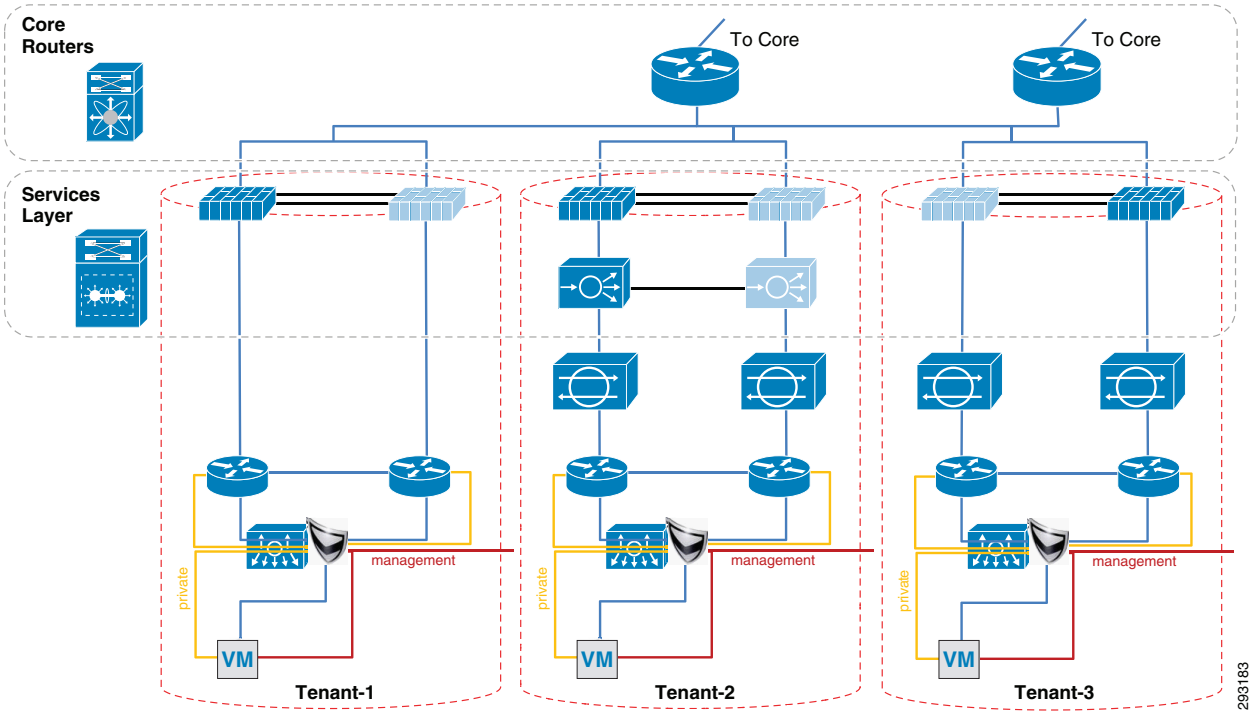
The **Management** layer consists of the "back-end" hardware and software resources required to manage the multi-services infrastructure. Such infrastructure include Active Directory, logging collection applications, and various device management software applications.

Multi-Tenancy Architecture

Virtualization of compute and storage resources enables sharing across an organizational entity. In contrast, virtualized multi-tenancy, a concept at the heart of the VMDC reference architecture, refers to the logical isolation of shared virtual compute, storage, and network resources. In essence, this is “bounded” or compartmentalized sharing. A tenant is a user community with some level of shared affinity. For example, within an enterprise, a tenant may be a business unit, department, or workgroup. Depending upon business requirements or regulatory policies, a tenant “compartment” may stretch across physical boundaries, organizational boundaries, and even between corporations.

A tenant container may reside wholly within their private cloud or may extend from the tenant's enterprise to the provider's facilities within a public cloud. The VMDC architecture addresses all of these tenancy use cases through a combination of secured data path isolation and a tiered security model which leverages classical security best practices and updates them for the virtualized multi-tenant environment. Figure 10 shows the implementation of multi-tenancy within the VMDC architecture.

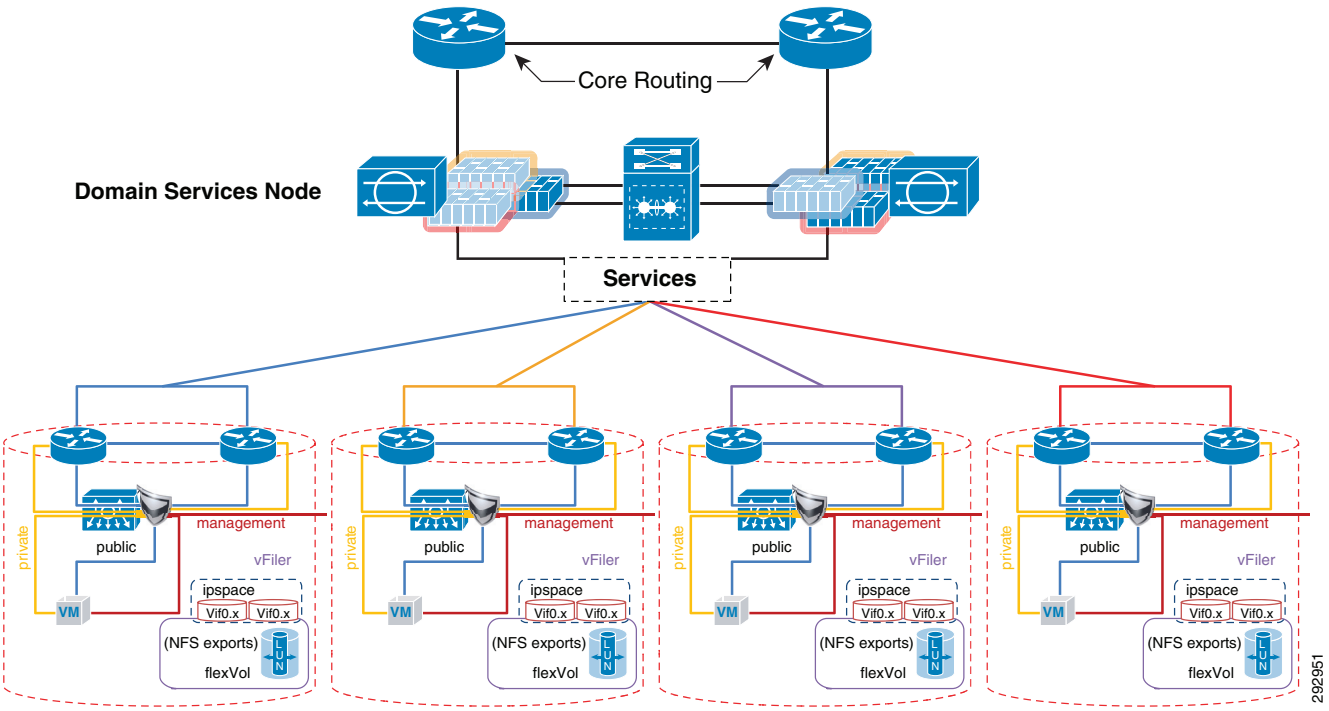
Figure 10 VMDC Multitenant Implementation



Services Overlay within VMDC

VMDC framework facilitates seamless overlay and integration of various services securely and reliably by providing Infrastructure as a Service (IaaS). Such services include physical security, collaboration and IP Telephony. The container model within VMDC's architecture can be leveraged to overlay various services. Figure 11 shows the integration of services within VMDC.

Figure 11 VMDC Services Integration



Services Mapped to Separate Virtual Containers

As it can be seen a number of services can be overlayed on VMDC by placing the various virtualized software components associated with each service within its own separate container. In case of Unified Communication, all of its components can be placed in its own separate container where it can be securely and reliably deployed. Each services-container can access the various common infrastructure service and the various tenants through the access layer or edge layer firewall. Firewall, load balancing and intrusion prevention services can be enabled for each services container as necessary.

IP Telephony within VMDC Framework

The virtualization of UC components in the data center allows the integration UC software applications, IP networking, network-based storage and virtualization into a single highly available system. This level of integration provides simplified server connectivity into the network, dynamic application repositioning between physical hosts, and pooled disk storage capacity. In addition the security framework within VMDC provides increased visibility, threat mitigation, simplified policy enforcement and secure isolation. The VMDC framework allows seamless deployment of the different UC deployment models.

An IP-based Telephony system has certain network characteristics and requires certain network services which effects how such a service is supported within the VMDC framework. The deployment of UC applications requires the availability of these services, described below. Some of these services may be implemented within the UC applications themselves.

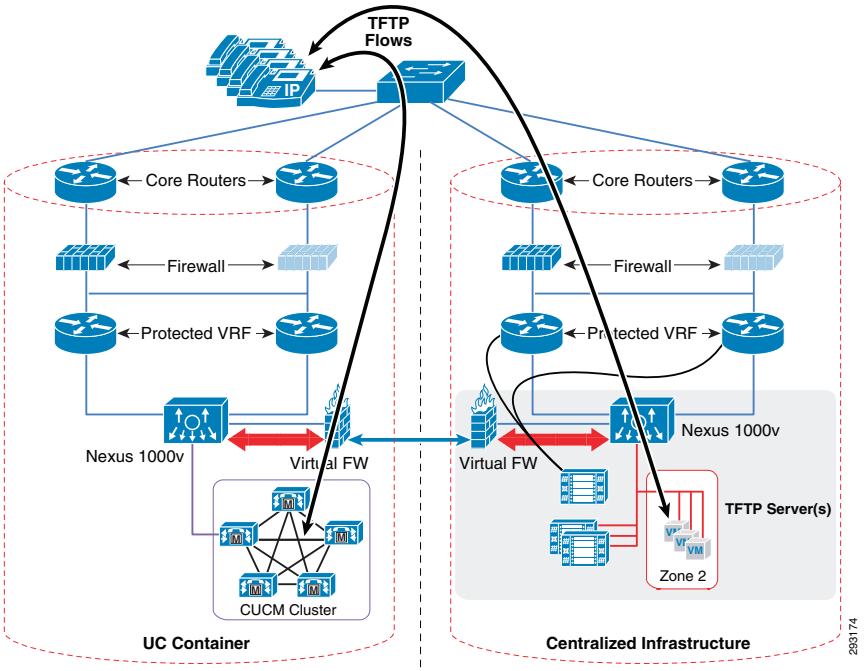
Network Services

The following network services within Cisco Unified Communications system are defined:

Trivial File Transfer Protocol (TFTP)

Within a Cisco Unified CM system, endpoints such as IP phones rely on a TFTP-based process (Figure 12) to acquire configuration files, software images, and other endpoint-specific information. The Cisco TFTP service is a file serving system that can run on one or more Unified CM servers. It builds configuration files and serves firmware files, ringer files, device configuration files, and so forth, to endpoints. Each time an endpoint requests a file, there is a new TFTP transfer session. For centralized call processing deployments, the time to complete each of these transfers will affect the time it takes for an endpoint to start and become operational as well as the time it takes for an endpoint to upgrade during a scheduled maintenance. The time to complete each file transfer via TFTP is predictable as a function of the file size, the percentage of TFTP packets that must be retransmitted (which is effected by the available bandwidth), and the network latency or round-trip time. Therefore the delay introduced by appliances within the network and the available bandwidth effect the operational and the perceived efficiency of the IP telephony service.

Figure 12 TFTP Traffic Flow

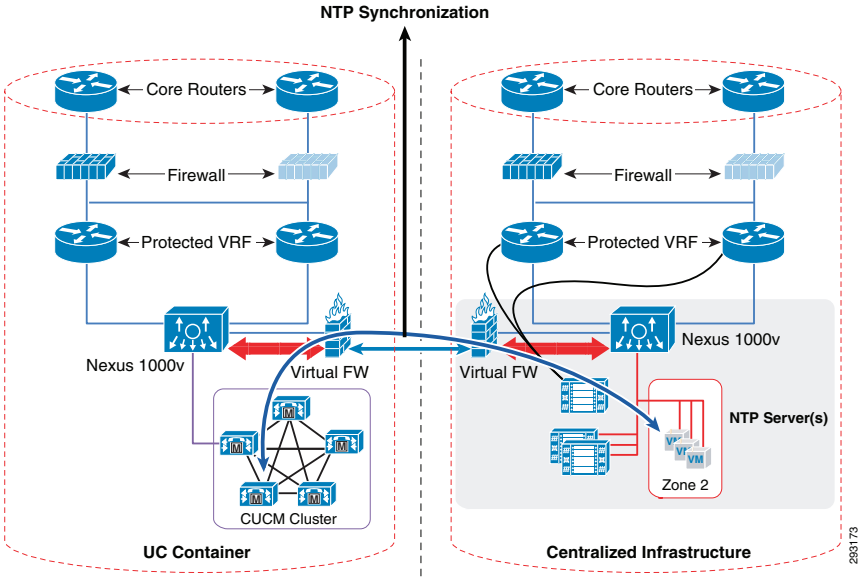


Network Time Protocol (NTP)

NTP allows network devices to synchronize their clocks to a network time server or network-capable clock. NTP is critical for ensuring that all devices in a network have the same time. When troubleshooting or managing a telephony network, it is crucial to synchronize the time stamps within all error and security logs, traces, and system reports on devices throughout the network. This synchronization enables administrators to recreate network activities and behaviors based on a common timeline. Billing records and call detail records (CDRs) also require accurate synchronized time.

Time synchronization is especially critical on CUCM servers (Figure 13). In addition to ensuring that CDR records are accurate and that log files are synchronized, having an accurate time source is necessary for any future features to be enabled within the cluster and for communications with any external entity. Unified CM automatically synchronizes the NTP time of all subscribers in the cluster to the publisher. During installation, each subscriber is automatically configured to point to an NTP server running on the publisher. The publisher considers itself to be a master server and provides time for the cluster based on its internal hardware clock unless it is configured to synchronize from an external server. Cisco highly recommends configuring the publisher to point to a Stratum-1, Stratum-2, or Stratum-3 NTP server to ensure that the cluster time is synchronized with an external time source.

Figure 13 NTP Synchronization



Domain Name System (DNS)

DNS enables the mapping of host names and network services to IP addresses within a network or networks. DNS server(s) deployed within a network provide a database that maps network services to hostnames and, in turn, hostnames to IP addresses. Devices on the network can query the DNS server and receive IP addresses for other devices in the network, thereby facilitating communication between network devices.

However complete reliance on a single network service such as DNS can introduce an element of risk when a critical Unified Communications system is deployed. If the DNS server becomes unavailable and a network device is relying on that server to provide a hostname-to-IP-address mapping, communication can and will fail. For this reason, in networks requiring high availability, Cisco recommends that you do not rely on DNS name resolution for any communications between Unified CM and the Unified Communications endpoints. For standard deployments, Cisco recommends that you configure Unified CM(s), gateways, and endpoint devices to use IP addresses rather than hostnames.

There are some situations in which configuring and using DNS might be unavoidable. For example, if Network Address Translation (NAT) is required for communications between the IP phones and Unified CM in the IP Communications network, DNS is required to ensure proper mapping of NAT translated addresses to network host devices. Likewise, some IP telephony disaster recovery network configurations rely on DNS to ensure proper failover of the network during failure scenarios by mapping hostnames to secondary backup site IP addresses. If either of these two situations exists and DNS must be configured, DNS servers in a geographically redundant fashion should be deployed so that a single DNS server failure will not prevent network communications between IP telephony devices. When DNS is used, Cisco recommends defining each Unified CM cluster as a member of a valid sub-domain within the larger organizational DNS domain, defining the DNS domain on each server, and defining the primary and secondary DNS server addresses on each server. VMDC framework with its separate container configuration provides the capability to implement DNS in situations described above.

Dynamic Host Configuration Protocol (DHCP)

DHCP is used by UC endpoints on the network to obtain initial configuration information, including IP address, subnet mask, default gateway, and TFTP server address. DHCP eases the administrative burden of manually configuring each host with an IP address and other configuration information. DHCP also provides automatic reconfiguration of network configuration when devices are moved between subnets. The configuration information is provided by a DHCP server located in the network, which responds to DHCP requests from DHCP-capable clients. Because IP telephony devices are configured to use and rely on a DHCP server for IP configuration information, you must deploy DHCP servers in a redundant fashion. VMDC framework provides three options for providing DHCP functionality:

- **Centralized DHCP Server**

  Typically, for a single-site campus IP telephony deployment, the DHCP server should be installed at a central location within the campus. As mentioned previously, redundant DHCP servers should be deployed. If the IP telephony deployment also incorporates remote branch telephony sites, as in a centralized multi-site Unified CM deployment, a centralized server can be used to provide DHCP service to devices in the remote sites.

- **Centralized DHCP Server and Remote Site Cisco IOS DHCP Server**

  When configuring DHCP for use in a centralized multi-site Unified CM deployment, you can use a centralized DHCP server to provide DHCP service to centrally located devices. Remote devices could receive DHCP service from a locally installed server or from the Cisco IOS router at the remote site. This type of deployment ensures that DHCP services are available to remote telephony devices even during WAN failures.

- **Unified CM DHCP Server**

  Typically DHCP servers are dedicated machine(s) in most network infrastructures, and they run in conjunction with the DNS and/or the Windows Internet Naming Service (WINS) services used by that network. In some instances, given a small Unified CM deployment with no more than 1000 devices registering to the cluster, you may run the DHCP server on a Unified CM server to support those devices.

Network Isolation and Virtualization

Before the phone has its IP address, the phone determines which VLAN it should be in by means of the Cisco Discovery Protocol (CDP) negotiation that takes place between the phone and the switch. This negotiation allows the phone to send packets with 802.1q tags to the switch in a "voice VLAN" so that the voice data and all other data coming from the PC behind the phone are separated from each other at Layer 2. Voice VLANs are not required for the phones to operate, but they provide additional separation from other data on the network. Voice VLANs can be assigned automatically from the switch to the phone, thus allowing for Layer 2 and Layer 3 separations between voice data and all other data on a network. A voice VLAN also allows for a different IP addressing scheme because the separate VLAN can have a separate IP scope at the Dynamic Host Configuration Protocol (DHCP) server. Using separate VLANS for voice achieves a a degree of separation and network Isolation.

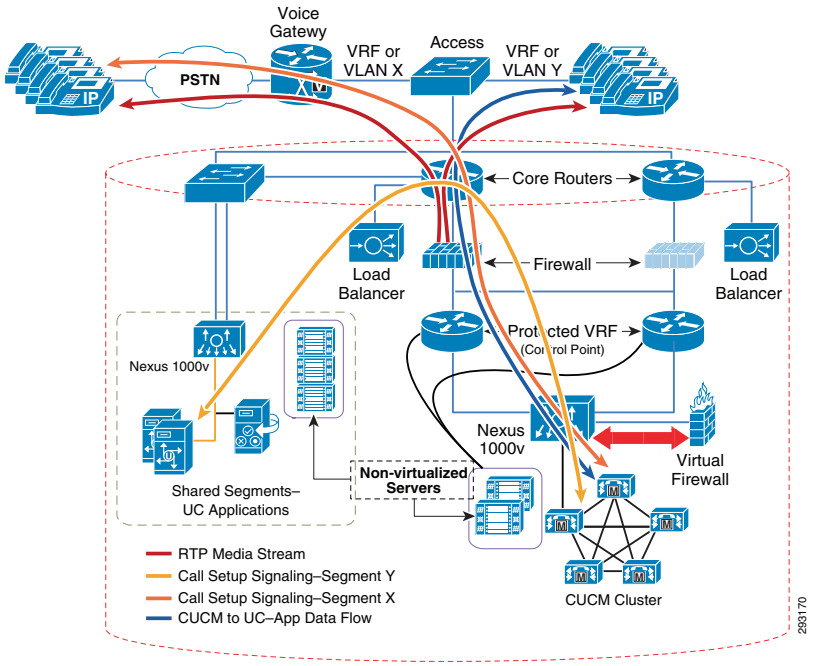
In addition to using VLANs as a means for network isolation, one can use network virtualization to achieve a greater degree of isolation. When a network is based on virtualization technology, there is a logical separation of traffic at Layer 3, and separate routing tables exist for each virtual network. Due to the lack of routing information, devices in different virtual networks cannot communicate with one another. Regardless of how the virtual networks are arranged – whether by department, location, type of traffic (data or voice), or some other basis – the core issue is the same: endpoints in different Virtual Private Network Routing and Forwarding tables (VRFs) do not have the capability to communicate to one another.

Using VRF virtualization capabilities incorporates a data center router with the capability to route packets to any VRF. The following base requirements apply to this scenario:

- Campus routers send packets for other campus VRFs toward the core router via default routing, so all router hops must route by default to the fusion router. The data center shared VRF has route information about each campus VRF. All VRFs other than the shared VRF have no direct connectivity.
- A Unified CM cluster is located in a shared VRFs in the data center, and communication within that shared VRF is totally unhindered.
- The shared VRF(s) is located in the data center

Figure 14 shows a solution that uses a shared VRF located in the data center to provide connectivity between a software-based phone located in one VRF and a hardware phone located in another VRF. The end point devices can be a separate VRFs or separate VLANs. Network Virtualization requires that fire-walling of the data center be implemented for the demarcation between the data center and the campus networks, and the following discussion shows how this can be implemented.

Figure 14 Network Virtualization



QOS Considerations

Until recently, quality of service was not an issue in the enterprise campus due to the asynchronous nature of data traffic and the ability of network devices to tolerate buffer overflow and packet loss. However, with new applications such as voice and video, which are sensitive to packet loss and delay, buffers and not bandwidth are the key QoS issue in the enterprise campus. Due to the delay-sensitive nature of voice traffic, ant IP-Telephony solution requires end-to-end QoS implementation This oversubscription, coupled with individual traffic volumes and the cumulative effects of multiple independent traffic sources, can result in the egress interface buffers becoming full instantaneously, thus causing additional packets to drop when they attempt to enter the egress buffer. The fact that campus

switches use hardware-based buffers, which compared to the interface speed are much smaller than those found on WAN interfaces in routers, merely increases the potential for even short-lived traffic bursts to cause buffer overflow and dropped packets.

The following types of QoS tools are needed from end to end on the network to manage traffic and ensure voice quality:

- **Traffic Classification**

  Classification involves the marking of packets with a specific priority denoting a requirement for class of service (CoS) from the network. The point at which these packet markings are trusted or not trusted is considered the trust boundary. Trust is typically extended to voice devices (phones) and not to data devices (PCs).

- **Queuing or Scheduling**

  Interface queuing or scheduling involves assigning packets to one of several queues based on classification for expedited treatment throughout the network.

- **Bandwidth Provisioning**

  Provisioning involves accurately calculating the required bandwidth for all applications plus element overhead.

In the VMDC framework the QoS framework is defined in the Service Assurance section of the VMDC Design Guide.

Table 2 shows VMDC classification used in VMDC framework. Voice traffic can be classified as VoIP traffic and thus marked with a COS of 5 and a DSCP value of CS5. The control plane traffic for CUCM and UC applications, which includes all other traffic from endpoints, clients and servers, should be marked with CS3 or CS2 and COS 3 and 2, respectively.

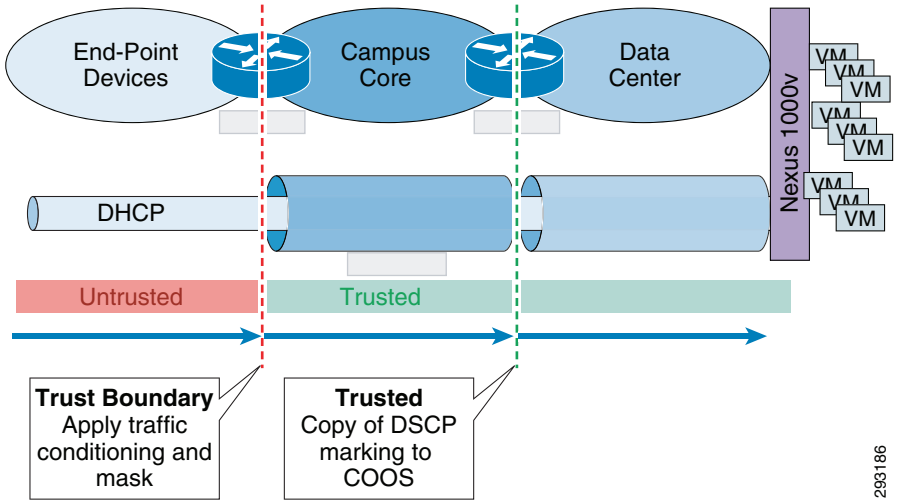
Table 2 Traffic Classes and Bandwidth Reservation (Eight-Class Reference)

Traffic Class	EXP/CoS	BW Reserved (Remaining After Priority)	Actions
Utility Compute Data: Bronze-Standard	0	15% (17%)	WRED
Out & Webex Collaboration Data (Interactive)*	In (CoS 2) Out (CoS 1)	60% (70%)	WRED, Out of Contract dropped before in contract
Utility Compute Data: Gold-Business Critical (In/Out of Contract)			WRED, Out of Contract dropped before in contract
Storage—FCoE & VoIP Call Control	3	3% (4%)	
Video Streaming (Future)*	4	x% (x%)	WRED, egress policing per tenant
VoIP Bearer & Video Conference	5	15%	Priority, egress policed per tenant
Network Control	6	4% (5%)	
Network Mgmt & Service Control	7	3% (4%)	

As shown in Figure 15, it is a best practice to mark traffic at the source-end system or as close to the traffic source as possible to simplify network design. If the end system is not capable of marking, or cannot be trusted, ingress marking may be used. If the endpoint devices are capable of marking their own

voice traffic, so ingress marking is not necessarily needed unless the access switch administrator does not want to trust traffic marking coming from end stations. For all other components of the IP telephony solution, ingress marking must be used. Most of these servers are virtual machines in a VMDC deployment, which means ingress marking would need to be configured on the Nexus 1000V used in the access layer.

Figure 15 Traffic Source-End System



Traffic generated by the all the components of the IP Telephony solution can take full advantage of the queuing and scheduling mechanisms implemented in the VMDC design. VoIP traffic will benefit from the use of class-based weighted fair queuing/low latency queuing (CBWFQ/LLQ) on the Nexus 1000V and other access switches at the southern edge of the DC QoS domain, and priority queuing (PQ)/CBWFQ on the core router at the northern DC WAN edge. These implementations, along with marking the VoIP traffic, will bound delay and jitter for the voice traffic.

Security Considerations

One important consideration in the implementation of IP telephony within a VMDC solution architecture and implementation is security. Securing the various components in a Cisco Unified Communications System is necessary for protecting the integrity and confidentiality of voice calls.

VMDC provides a comprehensive security framework that can be used by the network architect to secure the end-to-end Unified Communications System. General principles of the VMDC security framework are as follows.

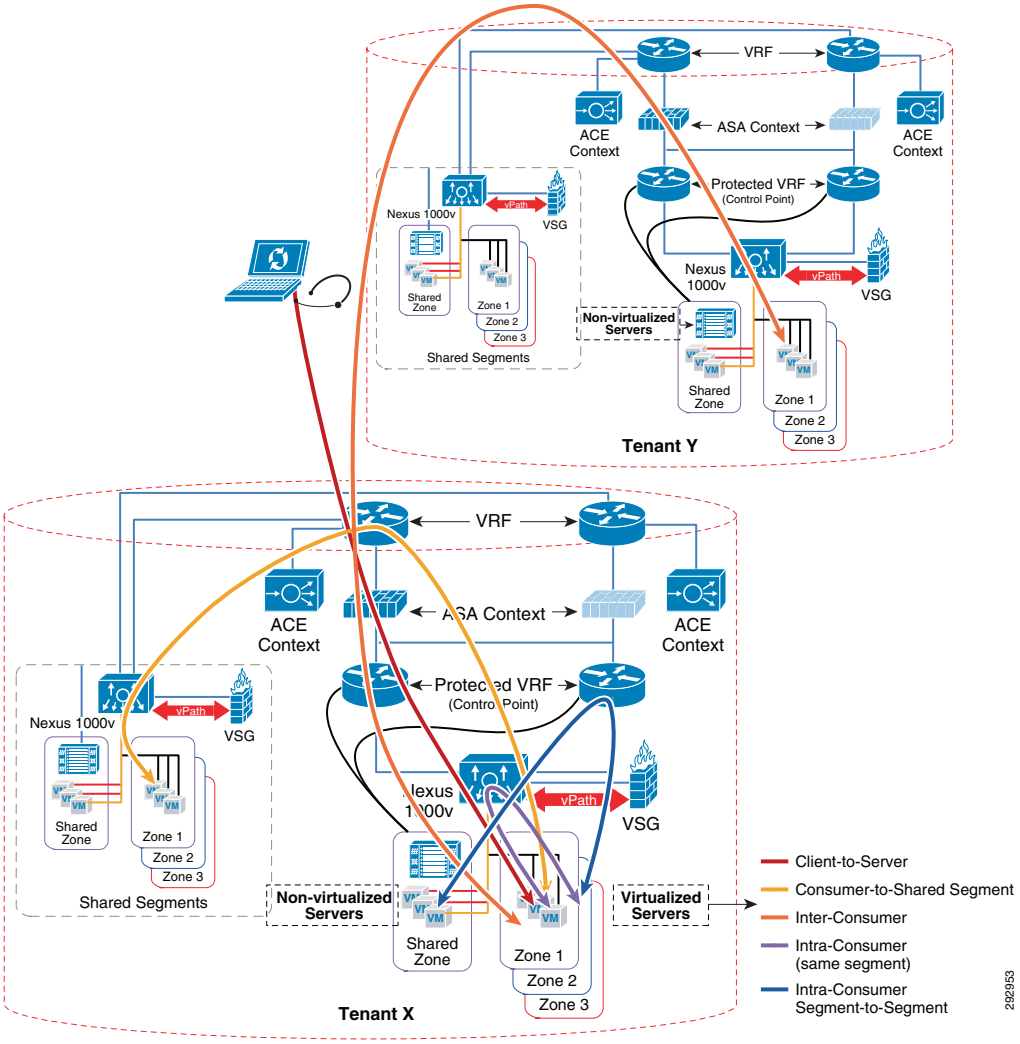
- **Secure Separation**—The partition that prevents one tenant from having access to another's environment and prevents a tenant from having access to the administrative functions of the cloud infrastructure.
- **Isolation**—Isolation within the VMDC framework is defined as the logical separation of network, compute, and storage resources. Depending on design goals, it can be achieved by using firewalls, access lists, VLANs, virtualization, storage, and physical separation. A combination of these provides appropriate levels of security enforcement to server applications and services within various tenants or services. Each tenant container has its own VRF, a set of distinct VLANs, access to a separate set of compute resources and its own firewall instance. The storage space can also be segmented and mapped to each tenant.

- **Policy Enforcement and Access Control**—Within the VMDC multi-tenant environment, access control and policy enforcement describes device and appliance capabilities within each layer of the architecture leveraged to create complex policies, and secure access control that enhances secure separation of all resources and services offered.
- **Visibility**—Total visibility implies that all resources within the network are used to facilitate threat detection and mitigation capabilities available at each layer of the network, to monitor traffic flows and gather alarm, data, and event information, to dynamically visualize attack paths, and suggest with optional enforcement response actions.
- **Resiliency**—Resiliency implies that end-points, infrastructure, and applications within the VMDC multi-tenant environment are protected and can withstand attacks that cause service disruption, data enclosure, and unauthorized access. Proper infrastructure hardening, providing application redundancy, and implementing firewalls are some steps needed to achieve the desired level of resiliency.

Traffic Flows

There are diverse traffic flows within the VMDC network. Understanding these various scenarios is significant when implementing firewall policies. Figure 16 shows different traffic patterns.

Figure 16 QoS Implementation



Virtual Firewall

Cisco's Virtual Security Gateway (VSG) firewall can be used to securely separate the UC components, to enforce firewall rules on data flows from the devices, and to enforce remote access by the management station. Possible ways that VSG's capabilities can be used to provide more security are:

- IP end points at different locations can be separated into different zones, where different security policies can be enforced. End points at onsite-campus locations may require different security policies than devices in the branch. Also, onsite end-point devices may be subdivided further into zones where separate security policies can be applied to them.
- Traffic to CUCM and application servers can be restricted to allow flows based on the ports and IP addresses of devices. Traffic to UC related components can be restricted if it resides on insecure hosts. There are some regulatory requirements that restrict sensitive virtual machines (such as CUCM) from co-residing on the same hypervisor with out-of-scope insecure virtual machines. VSG's hypervisor-based rules can be used to enforce such a requirement.

Physical Firewall

In addition to the virtual firewall-the VMDC architecture incorporates a physical firewall where each tenant is mapped to a separate firewall context. The use of a physical firewall at the edge provides increased security and additional flexibility. The physical firewall can be used to enforce policies specifying inter-tenant traffic flows, management viewing-stations, access policies and policies defining remote device connectivity to Unified Communication ecosystem. Since Unified Communication traffic flow is latency-sensitive, its may be preferable to route some of the UC-based traffic flows through a physical firewall.

Securing Unified Communication Components within a Data Center

Securing the various components in a Cisco Unified Communications System is necessary for protecting the integrity and confidentiality of voice calls. One of the more difficult issues with a security policy that includes IP Telephony is combining the security policies that usually exist for both the data network and the traditional voice network. It is vital to ensure that all aspects of the integration of the voice data onto the network are secured at the correct level for your security policy or corporate environment.

Securing the end-to-end Unified Communications System implies the hardening and securing the IP phone endpoints, the network from the phone to the access switch, to the distribution layer, into the core, and then into the data center. VMDC provides the security framework to harden and secure UC application. Within the data center, the security policies should define what security is needed for the IP Telephony applications servers.

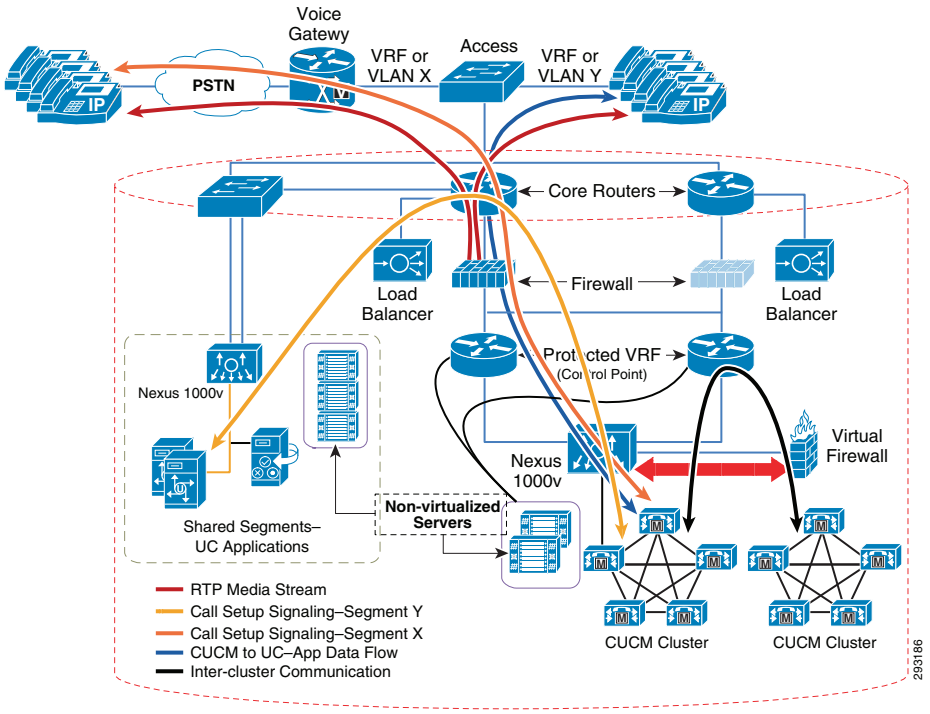
Because the Cisco Unified Communications servers are based on IP, the security that you would put on any other time-sensitive data within a data center could be applied to those servers as well. If clustering over the WAN is being used between data centers, any additional security that is applied both within and between those data centers has to fit within the maximum round-trip time that is allowed between nodes in a cluster. The following summarizes the mapping of VMDC security capabilities to the Unified Communication ‘s security components.

Firewall Deployment

Firewalls can be used to protect the voice servers and the voice gateways from devices that are not allowed to communicate with IP Telephony devices. Because of the dynamic nature of the ports used by IP Telephony, having a firewall does help to control opening up a large range of ports needed for IP Telephony communications. Given the complexities that firewalls introduce into a network design, one must take care in placing and configuring the firewalls and the devices around the firewalls to allow the traffic that is considered correct to pass while blocking the traffic that needs to be blocked. IP Telephony networks have unique data flows. The phones use a client/server model for signaling for call setup, and Unified CM controls the phones through that signaling. The data flows for the IP Telephony RTP streams are more like a peer-to-peer network, and the phones or gateways talk directly to each other via the RTP streams. Normally this peer-to-peer traffic flows occur outside of the data center. Inspecting this peer to peer traffic can increase the load of the firewall. Performance includes the amount of latency, which can be increased by a firewall if the firewall is under high load or even under attack. The general rule in an IP Telephony deployment is to keep the CPU usage of the firewalls to less than 60% for normal usage. However if the signaling flows do not go through the firewall the RTP streams could be blocked because the firewall will not know which ports need to be opened to allow the RTP streams for a conversation. A firewall placed in a correctly designed network can inspect the signaling traffic flows while minimizing the load on the firewalls; within the VMDC framework way that firewalls could be deployed within the data center, with Unified CMs behind them.

In this example, the Unified CMs are in a centralized deployment, single cluster with all the phones outside the firewalls. To keep the CPU utilization within the firewall to a minimum, RTP streams should avoid the firewalls. The gateways can be placed outside the firewalls, in the shared segment within the VMDC container. The voice applications servers are placed within the shared segment and firewall policies are used to control access to and from the Unified CMs and to the users in the network. This configuration will limit the amount of RTP streams through the firewall that inspects the signaling traffic, which will minimize the impact to the firewalls when the new voice applications are added to the existing network.

Figure 17 VMDC Traffic Flows



Within the data center, the security policy should define what security is needed for the IP Telephony applications servers. Because the Cisco Unified Communications servers are based on IP, the security that you would put on any other time-sensitive data within a data center could be applied to those servers as well. If clustering over the WAN is being used between data centers, any additional security that is applied both within and between those data centers has to fit within the maximum round-trip time that is allowed between nodes in a cluster. In a multi-site or redundant data center implementation that uses clustering over the WAN, if your current security policy for application servers requires securing the traffic between servers across data center firewalls one may use encryption for this traffic between the infrastructure security systems already deployed.

Securing Unified Communication Manager Servers

One characteristic of the CUCM servers associated with Unified Communications System is that they only operate using one interface. This interface is used for signal handling between end-point IP phones, management connectivity, data flows to the common enterprise-wide infrastructure services and inter-cluster communication. It is therefore imperative to secure these servers, while at the same time providing seamless traffic flows that need to be passed to and from these servers. VMDC provides the

flexibility to use a combination of physical firewall or a virtual firewall to secure these types of traffic. For example one may use a physical firewall to secure inter-cluster communications, since this type of traffic is time-sensitive. Securing different types of traffic is explained below.

- IP Telephone to Unified Communication Manager Data Flow.

  The only data flow between End-point IP Telephone devices and Cisco Unified Communication manager are signaling flows for call processing. In most cases all actual media traffic between end points occur outside the data center.

- Phones are controlled via SCCP or SIP from Unified CM.
- In the case of Cisco Unity Express deployment, Cisco Unity Express is controlled via JTAPI (CTI-QBE) from Unified CM.
- The Message Waiting Indicator (MWI) on the phone is affected by Cisco Unity Express communicating a change of mailbox content to Unified CM via CTI-QBE, and by Unified CM in turn sending a MWI message to the phone to change the state of the lamp.
- IP phones access to services, PIN authentications etc is achieved through HTTP or HTTPS signaling to and from the Unified CM.
- The voice gateway communicates via H.323, SIP, or MGCP to Unified CM.

Firewall policies can be used to secure and harden traffic flows shown above to and from UCUM.

Securing Unified Communication Application Servers

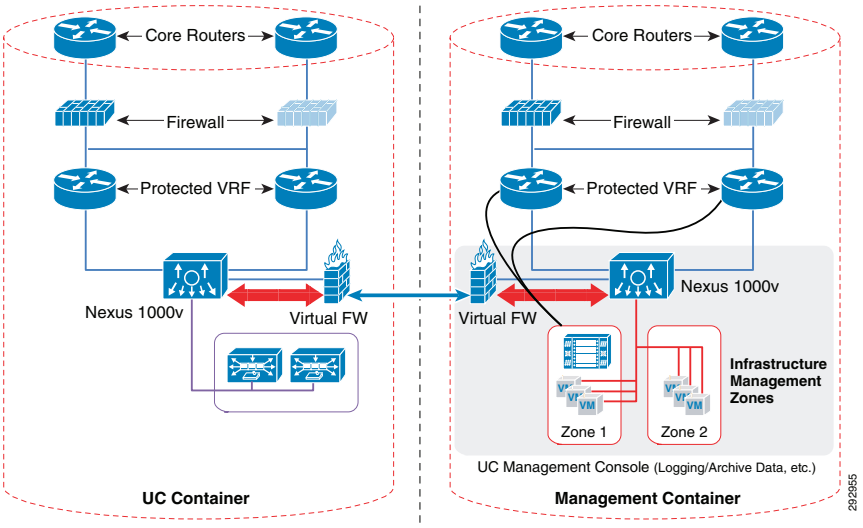
UC application servers such as voice mail, can be mapped and placed within the shared segment within VMDC. As explained below by placing these application servers in the shared segment one can reduce load on the firewalls that do the inspection. Firewalls outside the data center maybe used to secure these applications servers.

UC Management

The CUCM n be managed by opening a browser and using the management URL, and the management address of the UC virtual machine. The CUCM supports one interface for both data and management traffic. The use of one interface necessitates the use of a firewall to securely manage traffic to and from that interface The management stations managing the UC can be located at a different VMDC network container and the firewall policies can provide management access to the CUCM cluster securely.

Figure 18 shows how to use the management infrastructure within VMDC to securely manage the CUCM cluster using a virtual firewall.

Figure 18 Traffic Flows within a Unified Communication Deployed Network



Conclusion

This document describes the framework to integrate Cisco Unified Communications System solution components within the Cisco's Virtualized Multiservices Data Center (VMDC) architecture. By utilizing the VMDC framework, infrastructure architects can easily overlay UC components as an add-on service to an existing architectural framework. This would provide the certainty that the UC solution operates within a validated and consistent architecture that at the same time can support multiple services and customer use cases.

With virtualization of the UC Server and its associated components, the infrastructure architect can take advantage of the many virtual network services that are supported within VMDC. Virtual appliances, such as a virtual firewall can provide flexible policy enforcement, secure access and increased protection of the UC components than traditional stand-alone implementations.