VoIP: Do You See What I'm Saying?

Managing VoIP Quality of Experience on Your Network

NetQoS, Inc.

Chapter 4 – Unified Communications

In previous chapters we've discussed how you can manage the VoIP quality of experience on your network. We've provided strategies and advice to help you estimate the levels of quality that users experience in their interactions with the VoIP phone system and continually deliver a high level of quality. But VoIP is just the beginning. It provides a great starting point on the path to Unified Communications. The network requirements for VoIP, call setup performance, and call quality management are important foundational concepts to build on as you expand the capabilities of your network. Deploy VoIP and get it right; then, you are ready to move on to other Unified Communications applications like video, presence, and unified messaging.

Many enterprises are beginning to take a close look at the substance behind the hype surrounding Unified Communications, or "UC." A recent survey showed that nearly 30% of enterprises had a Unified Communications strategy, and more than 31% viewed Unified Communications as one of their top three IT initiatives. [1] Software heavyweights such as Microsoft® and IBM® have entered the UC market, which was previously owned by the IP PBX vendors, such as Cisco® and Avaya®. As the solutions evolve over the next several years, expect business applications and integrated soft phones to play a greater role and to find much broader acceptance.

This chapter explores the topic of network management for Unified Communications. As these new communications applications are added to the network, what are the key network performance considerations, and how can you manage them?

UC adoption is usually a slow, staged process and not a "forklift" upgrade. In this chapter, we'll examine the path toward Unified Communications and consider the following:

- What is Unified Communications?
- What are some of the new Unified Communications applications?
- How will a Unified Communications deployment affect network performance?
- How can you manage Unified Communications applications?

Unified Communications offers the vision of great productivity gains as integrated, multi-modal communications interfaces are built into our most commonly used business applications. But anytime new applications are added to the network, it's always good to take a step back and analyze the potential impact on network performance.

Before we do that, let's take a look at exactly what we mean by the catch-all term *Unified Communications*.

> The term *Unified Communications* is often abbreviated as "UC". We'll use *UC* and *Unified Communications* interchangeably throughout the rest of this chapter.

Defining Unified Communications

Ask the question "What is Unified Communications?", and you are likely to get many different answers. We thought it would be interesting to take a look at the Websites of several Unified Communications vendors and see how they define the term.

Cisco

Cisco is known for robust IP communications solutions. Their version of UC adds applications beyond VoIP. Here's how Cisco defines UC on their Website: "Besides helping organizations integrate their communications more closely with business processes, unified communications ensures that information reaches recipients quickly, through the most appropriate medium, no matter where they may be working or what device they may be using. Unified communications allows businesses to collaborate in real time using advanced applications from an integrated, easy-to-use interface. These applications include: video conferencing, integrated voice and web conferencing, mobile IP soft phones, and voicemail." [2]

Avaya

Avaya has an excellent primer available on their Website titled "Unified Communications for Dummies". That sounds like a good place to find a definition for Unified Communications. Here's what it says:

"Unified Communications is an evolving approach to communications that solves countless issues in the modern, mobile work environment, or, more accurately, wherever you're doing business these days. UC simplifies your communications by logically blending and combining previously separate services and features so that communications by any means with anyone is possible over any of your devices." [3]

Microsoft

Microsoft made a huge, market-altering splash with its entry into the Unified Communications sector. You may have heard the buzz surrounding their stated aim of killing off the IP PBX. Microsoft is leveraging its strength in enterprise messaging to enable other UC applications. Their collateral defines Unified Communications as follows:

"Microsoft unified communications technologies offer customers choices in how their communications and collaboration software is delivered, managed, and maintained [b]y uniting your existing communication systems and tools [w]ith the productivity tools your people use every day[, u]sing integrated servers plus services and client applications [t]o deliver complete communications tools . . . across multiple convenient applications and devices." [4]

Unified Definitions

After sorting through all the marketing messages from the major vendors, how should we define Unified Communications? Let's start by pointing out what UC is *not*.

Unified Communications is NOT:

- **VoIP only**. VoIP-based call processing is a building block for UC, but VoIP alone is not enough to provide UC.
- **Unified Messaging**. The idea of getting all your messages – email, voicemail, fax – in a single interface has been around for some time now. While UM simplifies message access and is generally part of a UC strategy, it is not, by itself, UC.
- **Closed, proprietary systems**. UC depends on interoperability between applications and infrastructure. If you can't communicate with a colleague because she is using a different vendor's communication system, there's not much point in being "unified" in other areas.

- **Rip and replace**. You likely have a communications infrastructure in place already. UC should work side-by-side with your existing infrastructure to enable new applications, not force the replacement of existing infrastructure.
- **About big cost savings**. UC may not save you money. It requires deployment and management of new components and applications. UC vendors often tout user productivity benefits as a cost justification. While we think these benefits can be substantial, it's difficult to put a hard dollar value on soft benefits like these. (Of course, you can always hire an expensive consultant to analyze productivity gains from a UC deployment, but that supports our point about the cost).

Now let's define UC from a *performance-first* perspective:

> Unified Communications is the integration of multiple modes of communication within applications and infrastructure that allows people, teams, and organizations to communicate more effectively. The IP network provides the unifying factor for UC, and network performance is a critical enabler.

There's one common factor when discussing Unified Communications applications: they all make use of a converged IP network. In order to provide benefits from the real-time presence status, point-to-click calling, video conferencing on demand, and other features built into UC applications, the network must be managed and tuned for optimal performance. From the perspective of user experience, UC applications will place greater demands on your network than any other networked applications to date.

Unified Communications solutions provide applications that allow for communications in a variety of different modes. Let's discuss some of these applications and their impact on network performance.

Unified Communications Applications

Unified Communications applications are designed to streamline business processes. Communications are a key part of any business, and ineffective or unavailable communications media can directly affect the bottom line. Think about the flow of information through your company. Your business has processes in place to route information to appropriate parties who act on the information and often, in turn, create additional information that must be acted on. These processes are prone to inefficiencies.

The Business Problem

UC applications strive to improve end-user productivity by addressing the business problem of communications inefficiency. Communications inefficiencies are created in a number of ways:

- **Phone Tag.** We've all participated in this little game. You don't know that someone is unavailable or in a meeting, so you call and leave a voicemail message. Then that person calls you back, but doesn't realize that you're now out of the office – so they leave another voicemail. Phone tag results in wasted time. If you knew the current availability information for a contact, you might try a different method of communication based on their status.

- **Number Lookup.** You get an urgent email from a colleague, and you need to quickly give him a call. Unless you have freakish memory recall, it's a good bet that you'll have to look up the colleague's phone number. Hopefully, your contact information is up-to-date and accessible. Wouldn't it be nice to just click on the username in the email message to call that person back immediately?
- **Switching Applications.** You're working with a business application and need to communicate with a colleague about a report you are viewing. Switching out of the business application and launching an email program takes time and loses the context of your working environment. If you can instead communicate from within the application, you save time and maintain the context for the communication.
- **Human Latency.** We all know the effect that network latency can have on application performance. Human latency actually has a comparable impact on your business. Gartner states that "the largest single value of UC is its ability to reduce human latency in business processes."[5] Consider this example: You are working on a project and need immediate input from a supervisor to move to the next step. Unfortunately, she is out of the office, meeting with a client. The time it takes for you to communicate with that supervisor and get a response could be considered human latency. Reducing this lag time can directly improve productivity.
- **Globalization.** The workplace is growing more and more global in nature each and every day. Working on a team that is geographically distributed – with team members in multiple global locations – requires excellent communications. Collaboration over a long distance requires superior-quality communications, such as high-definition videoconferencing, extremely frequent communications using several methods, or both. Applications that can make these communications more effective and more closely integrated with other collaboration tools enable the team to be more productive.

UC applications are geared toward addressing these efficiency issues and providing more effective communications for the enterprise.

UC Solutions

In many cases, UC applications have other common features as well. For example, many of these applications can be accessed via a "communicator" client program that provides a single interface for multi-mode communications. These lightweight applications may run on a PC, a mobile device, or a desktop phone, providing users with a common interface for most of their daily communications. In the next few sections, we'll extend our definition of UC to the components that can usually be found in a solution labeled as "Unified Communications" by a given vendor.

Communicator

Star Trek made the idea of personal communicators famous. Many of us look forward to the day when we can say "Beam me up, Scottie" into our own communicator devices. Today's UC communicators may not be able to physically transport you, but they do provide an integration platform for a number of different communications applications.

The communicator application is software that runs on a computer, phone, or other mobile device. It keeps an organized list of the people with whom you communicate on a frequent (or infrequent) basis: your contact list. Figure 4-1 shows screen shots of Cisco Unified Personal Communicator 1.2.2 and Microsoft Office Communicator 2007.

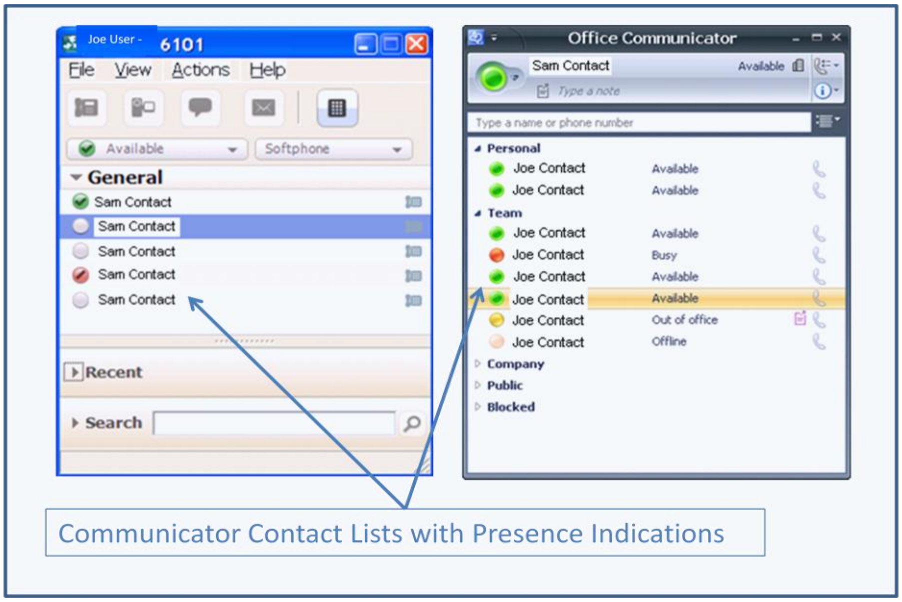

Figure 4-1 – Communicator applications provide contact lists and presence indication.

The contact list provides important information about those you communicate with. You can see their "presence" – their current status and/or availability for communications. From your contact list, you can launch different modes of communication to interact with a given contact. You might need Instant Messaging (IM) to send a message to a co-worker with a quick question. Once the IM session is started, you may need to escalate to a phone call to clarify certain points. Adding video or Web conferencing to an IM session for enhanced collaboration might be a one-click option in a true UC implementation.

Communicator applications often integrate email and the relatively newer IM capabilities that have recently become very popular in the typical enterprise. But they also recognize that audio is still an important means of communicating, which leads us to the next UC application: voice.

Voice (VoIP)

Voice is always a key part of any communications strategy, and UC is no exception. Many enterprises have found that VoIP plays a key role in enabling other UC applications. This is true for a number of reasons, primarily because VoIP places real-time communications requirements on the network with its strict latency and jitter thresholds. By tuning your network to support VoIP well, you are preparing it for other UC applications.

A VoIP deployment also provides key infrastructure for UC. The IP PBX or communications server handles VoIP call control, and this function is needed for UC voice applications as well. Unified messaging solutions, which can, for example, send voicemail as an email message, have flourished in VoIP environments and play a key role in UC.

The communicator applications discussed above can usually act as "soft phones," meaning that they are capable of making and receiving phone calls. A click on a contact name shows the available methods for calling that person. The call you make from the communicator application could be PC to PC, PC to desktop phone, PC to PSTN, or PC to mobile.

Video

Ever since the Jetsons started using their visaphone in the early 1960s, the idea of video conferencing has been appealing, especially if you spend a lot of time on conference calls or on airplanes traveling to meetings. The advantages of videoconferencing are well defined: You get to see the body language and read the lips of the people you are talking to; there's an opportunity for visual aids, such as the ubiquitous corporate whiteboard; and you're likely to pay closer attention to a video call than you do to an audio-only call.

However, the usage of video conferencing has varied greatly. In most cases, it has been reserved for executives or isolated in conference rooms. Now, UC applications enable easier video capability by integrating it with other communications applications. In some systems, you can call a colleague and then decide to add video by clicking an icon.

But users are not going to take the trouble to learn how to use video-conferencing equipment, which is often wholly unfamiliar, if the quality is poor, or if all they see is an image the size of a postage stamp. UC applications now offer desktop video capability with high-definition quality. Cisco provides a rich Telepresence solution that creates a specialized environment for users and creates an experience that is close to being there in person.

Presence

Presence has come a long way from the days of AOL Instant Messenger's® announcement that a user has logged on or off. Now presence has many states and can reflect not only availability, but willingness to communicate as well. Users can set their presence state and control multiple levels of access, controlling who can contact them and how they can be contacted. In the Microsoft Office Communications Server (OCS) solution, Outlook and even previously nonintegrated applications like Word can convey user presence information.

Users may be logged in to multiple devices; presence data must therefore be aggregated to determine status accurately. And for a really useful system, presence information is integrated with call routing so that if someone is "present" on his mobile phone, incoming calls for him aren't routed to the telephone on his desktop 500 miles away.

Presence updates need to happen in near real time if they are to provide a true representation of a contact's status. But the inclusion of contact and calendar information can make presence update messages quite large. Some planning is needed to mitigate potential network performance issues associated with potentially voluminous presence updates.

Instant Messaging

Instant messaging (IM) has now become a staple business application and a popular way to communicate. For quick questions or discussions, IM is often the preferred mode of communications, even over email. IM requires less overhead and is easy to use. There's no Inbox to clean out periodically.

IM usually has basic features, like sending and receiving text, changing fonts and colors, basic file transfer, and multi-party chat. With Unified Communications, a core feature is the ability to "escalate" an IM session to audio, video, and conference/collaboration as the communication needs dictate.

The performance characteristics of IM are likely not at the top of every network manager's hot list. However, as users grow more and more accustomed to the quick response that is part of the nature of IM, network performance becomes more important. If messages are delayed or do not make it through the network due to performance issues, users are left with a poor IM experience.

Email and Unified Messaging

Email is a foundational UC application because such a large percentage of daily communication is accomplished via email nowadays. Unified messaging applications provide the capability to access voicemail, email, and fax, all from a single interface. Frequently, email systems are used to store the voicemail data as email attachments that can be played back on a computer.

In UC applications, email clients may be integrated with other UC applications to allow calls or other communications to be routed to the users who sent the email. Presence information can show you the current status of the email sender, giving you valuable information about how best to respond.

Web-Based Conferencing and Collaboration

Collaboration is a big part of any UC solution. We mentioned earlier that globalization was a driver of UC development and adoption. Team members separated by geography need easy–to-use communications applications that enable collaboration. A whiteboard interface, plus the ability to share presentations, notes, and other documents are essential to effective teamwork.

You've probably used Web-based collaboration applications like Cisco's WebEx™, Citrix's GoToMeeting®, or Microsoft's LiveMeeting. These applications allow presentations to be shared with clients in Web sessions that previously required a face-to-face meeting. These applications can provide audio, video, and collaboration via a single Web-based interface.

As you can see, UC applications offer a vast array of integrated, streamlined communications features for users. The converged IP network provides the foundation for all of these applications and features.

The Network Foundation

A good network foundation is required to prepare for current and future UC deployments. Enterprises will follow different timelines as they begin this journey. But no matter how slowly or how rapidly the deployment proceeds, the need for a sound network management strategy increases as Unified Communications applications are rolled out beyond the initial pilot phase. The 2007 Yankee Group Research report titled "A Guide to Managing Enterprise Unified Communications" offers a look at the migration path from VoIP to Unified Communications and various management considerations along the way. [6]

Exhibit 1.

Migration Path Toward Unified IP Communications

Source: Yankee Group, 2007

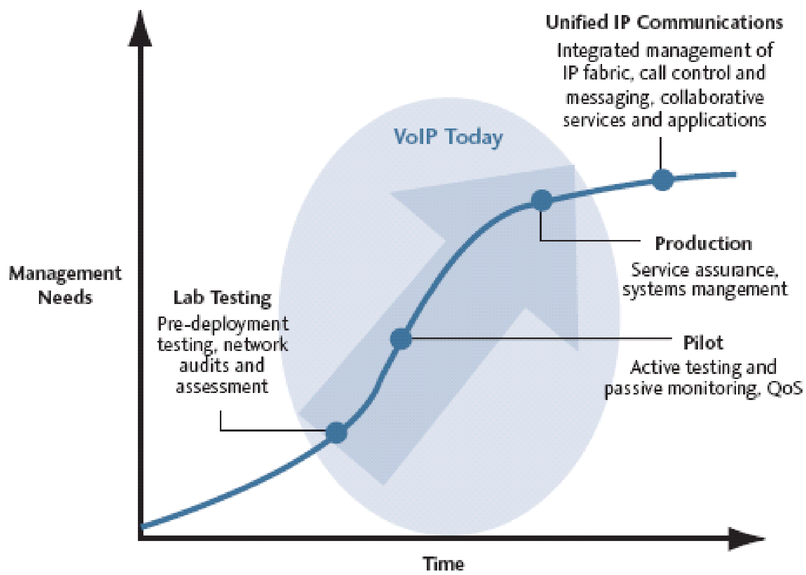

Figure 4-2 – Management needs increase on the path to Unified Communications

It's interesting to note that network management plays a key role in every stage of the overall UC deployment cycle. It begins with Lab Testing and continues through Pilot and Production phases – all in preparation for deployment of Unified Communications applications. At each step along the way, it's also important to look at key metrics for your existing applications to ensure that the changes haven't harmed their performance. Network performance is always an important component of end-user satisfaction with application delivery, but as we've pointed out in previous sections discussing VoIP performance, the network directly impacts the user experience with UC applications. According to Zeus Kerravala, Senior Vice President of Enterprise and Enabling Technologies at Yankee Group Research, "When UC applications encounter performance issues, they don't just slow down, they become unusable. The complexity of presence, voice, and video on your IP network demands that network management tools present a unified view of the performance of these applications."

With that advice in mind, let's take a closer look at the impact that UC applications can have on network performance.

Unified Communications' Impact on Network Performance

Any UC solution comprises lots of moving parts. Plenty of new applications, new endpoints, and new infrastructure must be added to your network. As Unified Communications applications are purchased, installed, and configured, prepare yourself for the inevitable changes by finding preliminary answers to several important questions:

- How will the new UC applications affect the performance of my existing networked applications?
- How will the new UC applications themselves perform?
- If performance is sub-par, will the user experience be good enough to make deployment worth the effort and expense?

To answer these questions, it's worth considering the kinds of requirements that typical UC applications can place on the network in terms of bandwidth, packet loss, jitter, and latency. Understanding how the new applications work is important as well. The Session Initiation Protocol (SIP) is the key network protocol that enables communication for almost all UC applications. In the following sections, we'll provide a quick refresher on SIP.

SIP – The Enabler

SIP was defined by the IETF in RFC 3261 and in many other follow-on RFCs. SIP can operate over both TCP and UDP, making use of port 5060 for non-secure communications and port 5061 for secure communications. SIP can use Transport Layer Security (TLS) to provide secure, encrypted communications. Encryption is important for SIP because it is a text-based, ASCII protocol, meaning that if you capture unencrypted SIP packets, you can easily read their contents. SIP uses an addressing format called a Uniform Resource Identifier (URI). The URI format resembles an email address, but is prefixed by "sip:". For example:

sip:user@netqos.com

or

sip:user2@10.10.10.1

The URI is used to set up a SIP session with another user. Whenever you make a call, start up a conference, or IM another user, SIP looks up and contacts the other user(s) based on their URI. The SIP addressing format means that instead of memorizing someone's phone number, now you can contact them via the more user-friendly URI. Figure 4-3 shows a snippet from a SIP INVITE packet. This INVITE was sent in order to establish a voice call from a user with extension 6101 to a user with extension 6110.

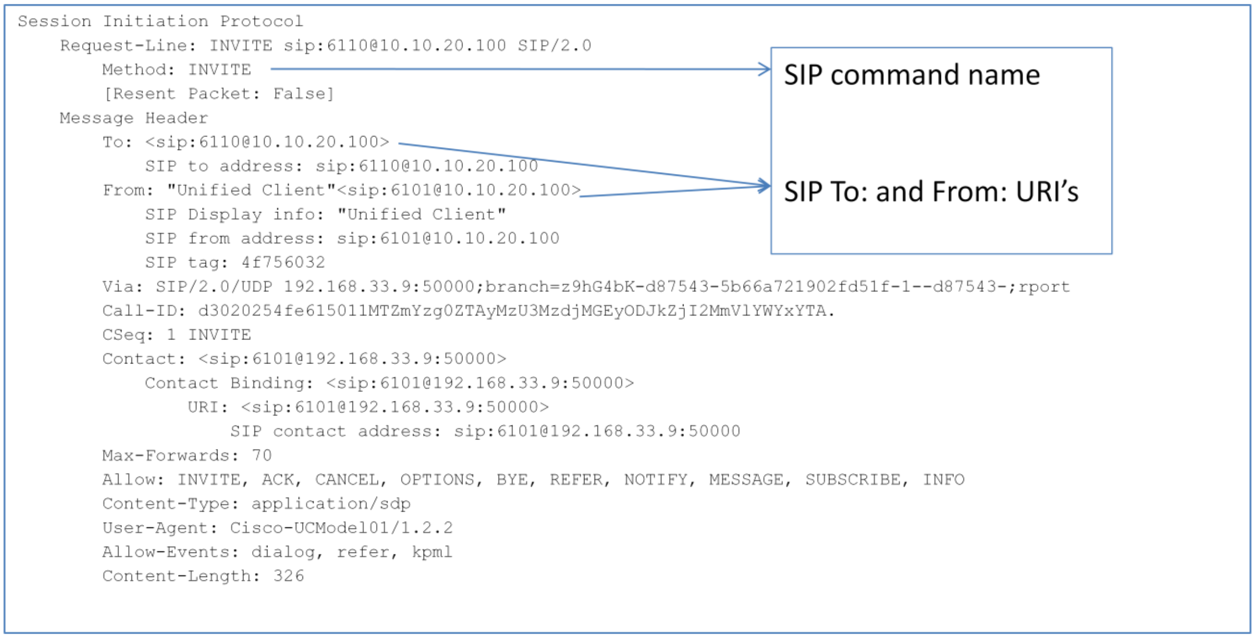

Figure 4-3 – SIP messages are text-based and contain URI addressing information.

Many SIP messages contain eXtensible Markup Language (XML) content. A generic standard for structuring and communicating data, XML is often used by Web-based applications to transfer data between a client and a server. SIP uses XML content to transfer information like presence status between endpoints. An embedded, XML-like protocol called Key Press Markup Language (KPML) is also used by SIP to transfer messages related to "keys" that a user presses. The call server subscribes to key press events, and each time the user presses a key, the digit information is sent by means of a SIP NOTIFY message to the call server.

SIP is not just a voice call setup protocol. It was designed to be more than that; in many current solutions, it enables real-time communications for audio, video, IM, and presence. Table 4-1 shows how SIP is used for the main applications typically present in a UC solution.

UC Application	SIP Usage
Voice	Set up and take down voice calls
Video	Set up and take down video sessions
Instant Messaging	Establish IM session between users and send/receive user text messages.
Presence	Used to allow endpoints to subscribe to presence status and receive presence change notifications.

Table 4-1 – SIP is a key enabler protocol for UC applications.

Because SIP is practically ubiquitous when it comes to UC applications, it is important to understand its performance characteristics. For example, one of the advantages of SIP is that it's a text-based protocol, easy to parse and read. But this advantage can also cause network performance issues: being ASCII-based means that many required SIP messages are quite verbose, consuming more bandwidth. We've observed from packet analysis of a standard phone call that the call signaling information from a SIP phone can consume up to four times more bandwidth than a phone using Cisco-proprietary SCCP. Call signaling is not a large bandwidth consumer in general, but when you deploy thousands of endpoints, the extra bytes can add up.

Another set of performance issues is inherent in the networking architectures that carry SIP data. As more and more UC applications that use SIP are deployed in the enterprise, organizations will need to communicate with other organizations outside their domain. In the past, the PSTN network functioned like glue to tie everything together. Islands of VoIP existed within enterprises, but PSTN connectivity was still required to call other users outside the enterprise. More recently, service providers have begun offering SIP trunks that can connect the SIP islands together and allow SIP-based communications between different domains.

A *SIP trunk* is basically a network connection to transfer SIP packets and data traffic up to an allocated amount of bandwidth. A service provider provides the network connection using a device called a *session border controller (SBC)* or *border element*. The session border controller acts as a proxy for SIP connections between the enterprise and the service provider network. Anytime boundaries are introduced between networks, and anytime proxies are involved, the potential exists for performance issues. Calls may travel exclusively on the IP network to other SIP-connected entities, or they may be routed out to the PSTN. As the migration from PSTN-connected VoIP islands to interconnected SIP networks progresses, new performance issues can arise. Calls may traverse other IP networks that you do not control, and end-to-end QoS may be difficult to achieve.

Anytime you have a hand-off between networks, the potential for performance problems arises. A best practice is to monitor your SIP trunk connections to see how many sessions are traversing them at a given time, and to measure the performance and quality of those SIP sessions. A good metric to watch is bandwidth utilization on the trunk – are you at capacity, approaching capacity, or not even close? QoS metrics are always important for SIP-controlled sessions. Be sure to monitor packet loss, jitter, and latency.

We've touched on some performance considerations for SIP. Now let's look at the performance of the UC applications.

UC Performance Considerations

Deploying UC applications can raise a whole host of new network performance issues above and beyond those associated with the SIP protocol and system architecture. One feature that makes UC attractive is the real-time nature of the communications – it fits the model of how humans like to communicate. But in order for unified communications to fit the "real-time" model, you need a network that's ready to provide optimal real-time performance.

Each of the typical UC applications we discussed earlier entails some specific ramifications for network performance.

Communicator

The communicator application, which is typically very lightweight, is a key enabler for Unified Communications because it ties the contact information and presence status information together. In most implementations, it opens in a very small, nonintrusive window, and can take only a few seconds to install on a client system. However, its small portion of desktop real estate belies its rather substantial network footprint. The first thing to consider is what happens when the communicator application starts.

Initialization

How many contacts do you have in your contact list? 10? 20? 50? The cool thing about the "unified" contacts in the communicator application is you don't just see the person's name and phone number. You can see all kinds of information about each person in your enterprise, such as their physical address, their department and title, their calendar, even a photo of the person. All of this information must be downloaded to your client when the application first starts up at login so that you have the latest information. To get an idea of the potential problem, we looked at the data transfers to download contact information for one of the leading communicator applications.

Contacts

The amount of bandwidth required for initial download is directly proportional to the number of contacts defined. We took a reasonable number of contacts—10—and started up one of the communicator applications. The amount of data transferred was approximately 75,000 bytes. Now imagine 100 users logging in and starting the day at 8:00 AM. Perhaps they are all at a remote site, with a slow-speed WAN link separating them from the server at company headquarters. Without traffic shaping mechanisms, the WAN link could be completely saturated while the UC applications download contact and presence information.

Large corporate directories can provide additional challenges, as most communicator applications allow you to search for a user in the directory. This information must either be queried on demand or cached locally by the application. Network flows associated with either of these actions could be significant.

Presence

When your communicator starts up, it has to register with the presence server in order to receive presence notifications. In a SIP environment, presence information is disseminated by means of SUBSCRIBE and NOTIFY messages. Your client must SUBSCRIBE to the contacts that you are interested in (the ones you have added to your list), and the presence server will NOTIFY you when their presence status changes. Figure 4-4 shows how presence updates work with Microsoft OCS. [7]

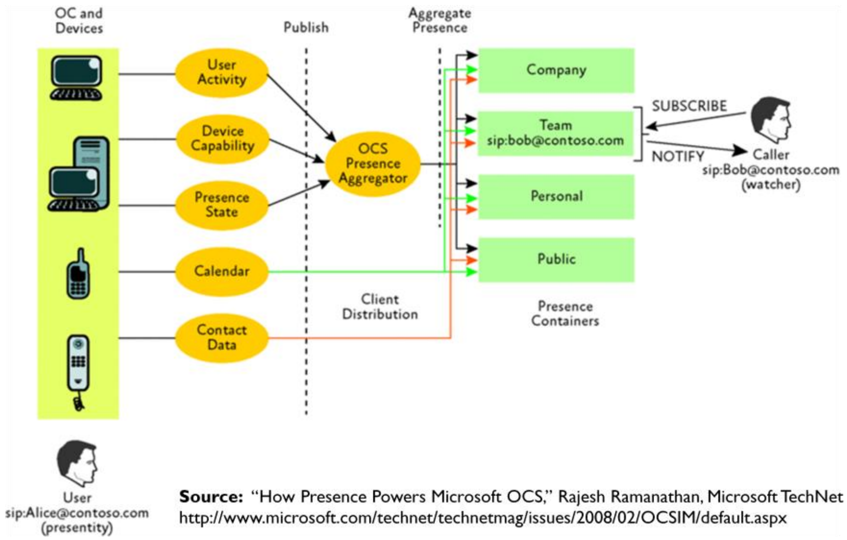

Figure 4-4 – Clients subscribe for presence updates, which may be aggregated.

The more contacts you have in your contact list, the more presence updates the server will need to send, and the more presence updates that you will need to receive. When the status of a contact changes, the server must relay that change to each user who has subscribed for that contact. It can become quite a large mesh of bidirectional updates as more users are added. Presence updates are also pretty frequent. They need to be sent and received in near real time; otherwise the information you are using to make decisions about whom to contact and by which method won't be up-to-date.

Voice

VoIP technology is an underlying building block for UC applications. We've covered VoIP network performance extensively in Chapters 1-3 of this ebook. We'd encourage you to review these if you are considering rolling out, are actually rolling out, or have already deployed VoIP. The same network performance issues that are present with VoIP phones are also applicable to UC applications.

The Quality of Experience (QoE) for voice calls is very important from a management perspective, along with the ability to map that QoE to the underlying Quality of Service (QoS) that the network provides.

Video

Bandwidth, bandwidth, and more bandwidth. UC desktop video conferencing now enables point-to-point video streams between the endpoints. Communicator applications allow users to easily participate in video conferences from their own computer. No longer is the video conferencing solution confined to a conference room with dedicated equipment, a dedicated expert on staff to help users activate it, or even dedicated network connections. Is your network ready for point-to-point video? Do you at least know how much bandwidth is already being used on the same links, and who is using it?

A single video stream takes anywhere from 300 to 400 kbps of the available bandwidth in each direction. Add in the audio component, and you've got over 800 kbps for combined video and audio streams. Voice packets are usually relatively small, while video packets are usually crammed with as much information as they can contain. Proper QoS mechanisms (see the "QoS" section later in this chapter for more details) are required so that the video doesn't crowd out the voice (and everything else).

Like voice, video is carried using RTP and is sensitive to packet loss, jitter, and latency. Packet loss can create missing blocks in the video image, or if severe enough, can cause the image to "freeze" for a few seconds. Too much jitter can have the same impact as packet loss as the packets arrive too early or too late to be contained in the jitter buffer. Latency can result in lag time, making quick interaction very difficult.

Ensuring user Quality of Experience (QoE) is immensely challenging for video applications. For one thing, it's hard to measure your success in delivering high-quality video. There's not a widely accepted video quality standard that's equivalent to the MOS for audio. "Video quality" is an even more subjective category than call quality because at least the MOS has been mapped to a set of metrics with defined value ranges. Video formats can vary, from low-resolution all the way to high-definition (HD) quality. Determining the user-perceived video quality is a big challenge, especially as more users become accustomed to watching HD programming on their televisions – the lower-resolution video just doesn't look as good anymore.

However, despite the lack of a generally accepted quality standard for video over IP, network performance metrics like packet loss, jitter, and latency can give you a good indication of the quality of service being provided by the underlying network. In addition, more video-specific metrics are usually measured by the video endpoints themselves. These include frame rate, frame loss, and frozen video and can provide some insight into the user experience with the video application.

Most UC solutions enable a couple of different types of video calls, with different requirements that can affect network performance. Let's consider them.

Point-to-Point

Adding desktop video presents an entirely new challenge to the network. Most networks are tuned for client-to-server business communications. VoIP has changed that somewhat as calls can traverse many different paths to get to the end-user. Now, with the addition of video calls, a lot more bandwidth may be required in places you didn't expect. What if Jim in Raleigh makes a video call to Steve in Austin? Is bandwidth sufficient at each hop along the path between Raleigh and Austin? You now have to consider capacity issues, such as: I can make X calls over this link now. If I add video, then I can make Y audio only calls, or Z calls with audio and video, or some number of both. Traffic shaping or rate-limiting techniques may be required to ensure optimal quality for all users.

And don't forget your remote workers who may connect to the network via a VPN connection. If they are participating in point-to-point video sessions, they may be consuming bandwidth in additional places where you were not expecting to see a bandwidth shortage. Depending on the VPN that is accessed, these users may or may not be near other users to whom they are talking via desktop video.

Unless your network links are tremendously oversubscribed, you will need more bandwidth tomorrow for every place where you can make a VoIP call today when you enable UC desktop video.

Conferencing

Most UC communicator applications allow video "calls" to be placed that involve more than two participants. In the case of a multi-party conference call, a conference server is usually involved. Users establish connections to the conference server, which multiplexes the audio and video and sends dedicated streams to each of the participants. In this scenario, the network between each client and the conference server must be able to support the audio and video streams. If you have a large number of clients in the same remote location participating in the conference, you could face performance issues as separate unicast streams are sent to each client.

However, this type of conferencing solution is not your only option. Many video streaming applications provide a multicast option or other features to reduce the network bandwidth required for video conferencing. You should consider the likely usage scenarios when you evaluate different UC video conferencing solutions.

QoS

As with VoIP, QoS mechanisms are required to enable UC applications to run with acceptable quality. The real-time nature of UC applications and their strict latency and jitter requirements dictate the usage of QoS. Due to differences in the protocols and bandwidth consumption, you will need multiple classes to represent SIP, audio, and video flows. Cisco's Class Based QoS (CBQoS) is an ideal mechanism for UC applications.

CBQoS allows a network administrator to define application traffic classes that should receive different handling in the network. Policies can be applied to the traffic classes to control bandwidth and provide low-latency queuing, with the ability to create separate queues for different traffic classes if needed. In a UC environment, you could define separate voice, video, and SIP traffic classes. Each of these classes would be prioritized appropriately by intermediate network devices that carry the UC traffic.

UC applications mark the traffic flows with different bit settings in the DiffServ codepoint. Network devices can use the different bit settings to map the traffic to the appropriate class. Different queues can be used for each class so that the voice, video, or SIP traffic does not become stuck behind other data traffic.

UC is all about the user experience. A poorly performing network can make that experience a bad one. UC applications introduce a number of network performance issues. Earlier we mentioned the need for a unified approach to managing the performance of UC applications. Let's discuss such an approach.

Managing Unified Communications

Managing the performance of Unified Communications applications presents a challenge to the traditional organizational hierarchies at many enterprises, where separate teams of trained staff have responsibility for different communications components. In the past, a typical enterprise had a group to manage their PBX, another group to manage the network, a different group to manage servers, and possibly even another group to manage specialized infrastructure like video conferencing. The transition to VoIP has initiated the convergence of not only the networks, but also the management groups. Many telecom and data management groups are becoming a single entity. Now with the addition of UC, we see the further convergence and blurring of traditional boundaries.

Successful delivery of UC applications will require a performance-first – proactive quantifying of network and application performance – mentality, applied to the management of the applications and infrastructure. It's not enough to know whether the server is running, or the router is up. UC will cross all the boundaries of application, voice, server, and network management and demands a unified approach to management.

Managing Unified Communications starts with understanding the components that create a better QoE for your users. You need tools to provide the visibility into the QoE metrics and the ability to map those metrics to underlying network quality of service. Perhaps most importantly, you have to do away with the finger-pointing traditionally associated with the separate IT and Telecom teams of the past. Keeping users happy as you roll out not only new applications, but new ways of communicating with coworkers, both distant and close by, requires a cooperative and comprehensive mindset.

Let's consider a unified approach to UC management that prioritizes performance.

UC Management Methodology

Since UC applications span a broad spectrum in terms of network performance, to manage them all, it's useful to look at a number of different performance data sources and pull them into a single UC Performance "Dashboard"-type report. A performance-first UC management strategy should include performance characteristics for all of the UC applications, always bearing in mind that they have the potential to drag each other down. Each application will have one or more performance metrics associated with it that is best suited for the approach outlined here. As the performance information from the data-gathering tools is synthesized in a single report, it should present a comprehensive management view of the UC environment. Figure 4-5 demonstrates some of the items that should be part of UC performance management.

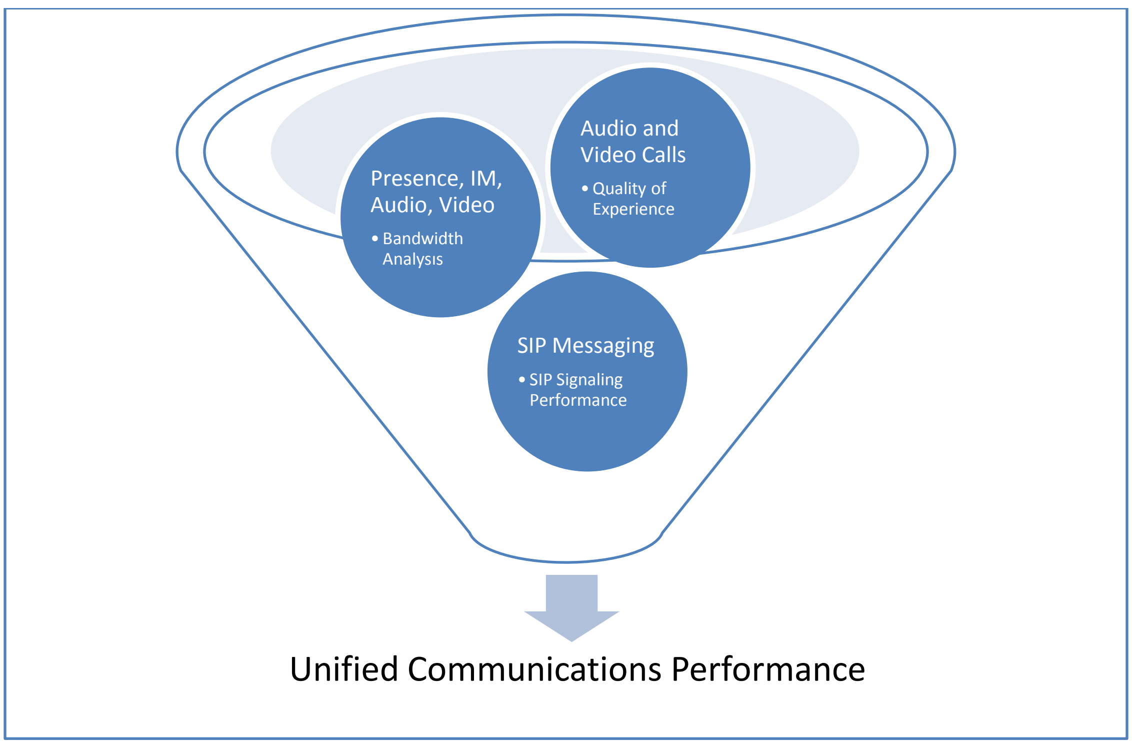

Figure 4-5 – A comprehensive approach to UC performance management

UC performance management begins with the Quality of Experience. What is the QoE for the user audio and video calls? What is the underlying QoS that supports that QoE? These are key questions that should be answered by UC performance monitoring tools.

For voice, the QoE is best measured by the call setup and call quality performance. For call setup, look at metrics like delay to dial tone, post-dial delay, and call setup failures. For call quality, begin with MOS, but also understand the relevant network metrics like packet loss, jitter, and latency. We've discussed call setup and call quality management extensively in Chapters 2 and 3 of this ebook.

For video, quality standards are not as well defined as they are for voice. It's hard to use software to quantify something like "lip-sync delay," for example, even though a viewer most certainly notices the effect when the image and audio are poorly synchronized. In the absence of a de-facto quality standard for video, a more user-friendly approach to video quality monitoring relies heavily on metrics that are measured by the video components themselves, such as video frame loss and frozen video. You can then correlate spikes in these values with the underlying network metrics for packet loss, jitter, and latency. Keeping all of these metrics is as important to video quality as it is for VoIP call quality.

Moving from user experience to bandwidth analysis of UC applications is a logical progression. We identified some of the performance concerns around bandwidth for UC applications like video and presence. The data collection tools you use to manage UC performance need to provide enough visibility into traffic composition on network links so that you gain the necessary visibility to understand:

- how much bandwidth is being used by each of the UC applications, and
- who is using that bandwidth

We mentioned earlier the potential that desktop video has for consuming bandwidth in places you don't expect. If a specific user is making video calls all day long to other users and saturating your WAN link, you need to know about it right away. You also need microscopic visibility into every network link where QoS is being applied. A QoS misconfiguration at any point in a communication flow between two users can make VoIP and video nearly unusable as the tiny VoIP packets get queued behind the huge video packets, or as video traffic is queued with other data traffic.

Finally, we touched on the fact that SIP was the underlying enabler for all UC applications. It only makes sense that we should keep an eye on the performance of SIP signaling and message flows. Within the SIP messaging performance, it's important to understand what kind of Network Round Trip Time (NRTT) is typical between client UC applications and your UC servers. What is the server response time, and is it degrading over time?

A network performance product like NetQoS Performance Center can help answer these questions and report on underlying metrics that affect UC application performance.

In Figure 4-6, we've created a Unified Communications Dashboard using the NetQoS Performance Center. The dashboard shows at a quick glance the UC Quality of Experience, a UC Bandwidth Analysis, and a view of SIP Messaging Performance.

Figure 4-6 – A Unified Communications dashboard provides comprehensive view of UC performance

The different views on the "dashboard" report provide insight into UC performance. The Call Quality Breakdown view provides a breakdown of audio calls and the user experience based on the MOS for those calls. The Call Performance by Location view shows ratings for audio and video performance metrics for calls in specific network locations. Selecting a "location" allows you to drill in and see the metric details, including key call-quality and call-setup metrics, measured on specific subnets. The Performance by Application view provides ratings for performance metrics associated with SIP messaging. Selecting the "SIP Messaging" applications allows you to drill in to see the details of salient metrics such as transaction time and network round trip time.

Bandwidth analysis is provided by the Top Enterprise Protocols by Volume and Top Enterprise Hosts by Volume data views. With these views, we can see how much bandwidth is consumed by voice, video, and SIP. In addition, we can see who is consuming the bandwidth so that we can make performance-enhancing adjustments. For example, if video conferencing is using up a lot more bandwidth on two or three key links, we might need to move the conferencing server to another location.

Chapter Summary

Unified Communications is another in a long line of new applications that offer compelling features, all of which are accompanied by network performance ramifications. We always recommend that before you deploy a new networked application, first, understand its network performance requirements, and second, understand the potential impact on your existing applications.

As applications like UC have increased user interaction, understanding the user's quality of experience and being able to map that quality to performance levels for management and troubleshooting are crucial. Use monitoring tools to gain the visibility that you need to provide an excellent unified communications experience for your users.

References

1. Kelly, Brent. "Do You See What I See in UC?" *Business Communications Review*, November 2007. Posted on: http://www.nojitter.com/showArticle.jhtml?articleID=205901063
2. Cisco Website. http://www.cisco.com/en/US/products/sw/voicesw/products_category_technologies_overview.html
3. Gregory, Peter. *Unified Communications for Dummies, Avaya Custom Edition*. Wiley Publishing, 2007. https://www1.avaya.com/forms/premcontent/us/index.html?cid=AVPC97D&urlcode=1336&revisit=0
4. Microsoft Website. http://www.microsoft.com/uc/what.mspx
5. Gartner. "Magic Quadrant for Unified Communications," August 20, 2007.
6. Yankee Group Research, Inc. "A Guide to Managing Enterprise Unified Communications," January 2007.
7. Ramanathan, Rajesh. "How Presence Powers Microsoft OCS". Microsoft TechNet. http://www.microsoft.com/technet/technetmag/issues/2008/02/OCSIM/default.aspx.