White Paper

Factors to consider when deploying Microsoft® Office Communications Server (OCS) 2007 Release 2

Introduction

In today's global economy, many companies face the challenge of workforces that are increasingly distributed and yet must work together more effectively than ever before. Unified communications addresses this challenge by simplifying how people connect and interact with one another as well as with applications. With unified communications, presence; real-time communications, including IM, telephony and video and application sharing; and near-real-time communications such as email and voicemail are unified into a single environment that enables workers to collaborate in real-time using the most appropriate channel. Tight integration between enterprise applications and communications allows users to share information within seconds. Whether at an employee's desktop or on the private and public networks that provide connectivity to mobile workers around the world, this unification makes life easier for users and IT departments, and leads to dramatic cost savings, improved productivity and increased revenue.

As unified communications technology takes businesses toward new frontiers in the enterprise market, Nortel is leading the way with the development of unique new technology, products and professional services that preserve and enhance the value of customers' past communications investments. Nortel's collaborative multimedia technology and tools for the voice communications network are but one example of these innovative new products. This technology and these tools have already demonstrated how they preserve and enhance value through integration of Microsoft's unified communications domain to the SIP-based converged voice communications infrastructure using Session Initiation Protocol for Computer Telephony Integration (SIP-CTI).

Deploying Microsoft OCS

Planning, building, operating and maintaining a unified communications model brings many challenges; to ensure success, it is important to identify and understand the business case for these technologies — as well as their risks.

Whatever the overall deployment strategy, there are a few prerequisites that must be met in both the IP domain and the voice network infrastructure if migration from a legacy TDM environment to a unified communications solution is to be successful and cost-effective.

This white paper highlights some of the factors that should be considered when planning to implement a unified communications environment based on the integration of enterprise telephony with Microsoft Office Communications Server 2007 Release 2. Additionally, the paper highlights implementation issues and potential solutions to unified messaging architectures, but does not discuss the relevant advantages and disadvantages of various unified messaging architectures.

Factors to consider when implementing a unified communications environment based on integration of enterprise telephony with Microsoft's Office Communications Server

This paper is divided into three sections: voice network infrastructure, IP network readiness and OCS deployment.

Section one: Voice network infrastructure

For more than a century, telephony has played an increasingly important role in business communications. The advent of mobile networks — which enable users to communicate almost anywhere, regardless of location and situation — has increased the business users' reliance on telephony. This increased reliance has led enterprise telephony to become increasingly complex in terms of architecture and feature functionality, as well as in its integration with business processes. Today, the average moderate-sized enterprise probably owns multiple installations of PBXs that span several generations of technology. OCS 2007 Release 2 is not meant to be a PBX, but rather offers an alternative communications mechanism to that offered by a PBX. In fact, an existing PBX can coexist with OCS 2007 Release 2 and Office Communicator 2007 Release 2 to provide a flexible and powerful combination of traditional telephony and unified communications.

Coexistence scenarios

There are two general scenarios for coexistence between PBX, OCS 2007 Release 2 and Office Communicator 2007 Release 2 — the first of which is the native integration of a PBX with OCS 2007 Release 2 and Office Communicator 2007 Release 2. For this integration to be effective, the PBX must natively support SIP and IP media in a form that is interoperable with Microsoft unified communications. Microsoft announced the availability of this interoperability in March 2007 and Nortel fully supports this coexistence. Indeed, the strategic Innovative Communications Alliance formed by Nortel and Microsoft is critical to the successful integration of OCS 2007 Release 2 with Nortel's enterprise telephony systems.

The second coexistence scenario is to use a Microsoft-certified IP-PSTN media gateway to provide integration between OCS 2007 Release 2 and the PBX.

The enterprise dialing plan in an OCS 2007 Release 2 coexistence scenario

In an OCS 2007 Release 2 coexistence scenario, the enterprise dialing plan will remain on the existing PBXs. However, when OCS 2007 Release 2 is deployed as a departmental solution, OCS 2007 Release 2 stand-alone is distinct from OCS 2007 Release 2 in coexistence in that the enterprise dialing plan is partitioned between OCS users and PBX users. This partition means that users have either a PBX phone or an Office Communicator endpoint, but not both.

Standardizing dialing procedures and configuration

Progression toward VoIP converged networks makes it increasingly important to establish consistency between voice and data network domains. Standardizing private voice network dialing procedures and configuration options is an important part of this evolution to ensure harmony with the more virtual nature of the IP-networking domain that uniformly encompasses the regions of all call servers.

Dialing plan standardization is achieved by maintaining a simple and uniform dialing procedure for the call server and establishing consistent configuration rules across all call server systems. An example of a simple and uniform dialling procedure would be one in which all voice users at all locations on the enterprise private voice network would dial 'nine' as an access code before making a public voice network call, and 'eight' as an access code before making a private voice network call.

Private voice network calls might also require users to dial a three-digit location code after the private network access code to reach any four-digit telephone extension at a remote site. Calls would then route across the enterprise private voice network of call servers to the desired location. Under this dialing plan, each enterprise voice user would have a seven-digit voice number qualifier.

SIP-service domain

The SIP-service domain name is a virtual component of the VoIP domain and serves to integrate the address space and signaling protocols of the voice network and IP domain.

A SIP URI Map is created within the SIP-service domain to support the translation of voice number qualifiers to SIP-equivalent messages used in the negotiation of voice-related transactions. The map must, therefore, be compatible with all SIP endpoints in the service domain. Voice number qualifiers are those generally used in the legacy voice network environment, and consist of numbering plan identifiers (public/private) and types of numbers (TON) within each, such as national, subscriber, UDP and CDP, etc.

Please note that in the OCS 2007 Release 2, the SIP URI Map needs a voice number qualifier to be in E.164 format. The private voice network dialing plan, therefore, requires careful consideration when implementing an OCS 2007 Release 2 coexistence scenario.

Nortel call server/IP PBX

The call server provides the basic telephony services traditionally found in a PBX as well as the new services required to deal with an IP infrastructure. Depending on requirements and network infrastructure, call servers are deployed in either a centralized or distributed architecture, each of which has advantages and disadvantages. Neither architecture can be considered 'best' without first understanding factors such as uptime requirements, network infrastructure and carrier capabilities if a Wide Area Network (WAN) is required.

In a centralized call server approach, uptime is heavily dependent on the capability of the WAN to be responsive and resilient. Without proper provisioning, network outages can potentially affect even local services, including intra-site calling and emergency services capabilities. The key is to have a common dial plan and feature transparency across

the network. This can be achieved with a distributed architecture that takes the burden off the WAN for all capabilities except media path bandwidth.

Ensure that call servers for VoIP or multimedia are put in what is referred to as a Secure Voice Zone. A Secure Voice Zone is a subnet that requires special VPN and/or Firewall access to protect it from security attacks and isolate it from any unrelated traffic.

Unified messaging

When designing a network for a centralized unified messaging platform, much care should be taken with the network architecture and unified messaging platform. Due to bandwidth savings, most VoIP networks today are based on the G.729 codec, which is known to deliver near toll-quality performance. However, the G.729 codec is also based on a single transcoding. Many store-and-forward application servers, such as a unified messaging platform, can introduce an anomaly called multiple transcoding. Multiple transcoding happens when a message is transmitted across the network at G.729 and then transcoded back to G.711 when it hits the application platform and is stored in another compression mode. When the message is picked up, it may be compressed back to G.729 for transmission on playback. Depending on the quality of the network, voice quality can degrade substantially. In this case, it is necessary to design a network that is able to record or playback at G.711 to eliminate multiple transcoding. Alternatively, a message can be stored in the original algorithm or as an RTP stream.

Section two: IP network readiness

Typically, IP networks have been designed to satisfy the requirements of data services, which are principally those services that tolerate delay and packet loss. Often the services involve no fast user interaction and use TCP to adjust bandwidth consumption and retransmit in the event of packet loss. The design of these networks has been based on the "best-effort" service model and has required no significant extension to TCP or IP, as has been demonstrated by the immense growth in the Internet for data services.

With telephony-over-IP, achieving PSTN equivalence is likely to be a required objective. That is, quality should not be perceptibly worse when traffic is carried over the IP network than when it is carried over the TDM PSTN. This objective gives rise to targets, such as for delay and packet loss, that are more demanding than those needed for data services. Moreover, the new services typically use UDP, which does not adjust its bandwidth consumption dynamically in the same way as TCP, so when deployed without proper planning, the services could have poor performance and inflict poor performance on existing services.

To ensure that routers and switches are configured optimally to support telephony and multimedia traffic, it is recommended that a detailed network assessment is undertaken before deployment is initiated.

Capacity

Capacity requirements for services offered over an IP network impose demands on endpoints, such as communication servers, as well as on network links. One major influence on the capacity required for the IP network is the bandwidth needed to support traffic of a given type passing between each source and destination. The bandwidth needed varies between and within traffic types in accordance with performance requirements.

Enterprise networks should be engineered to provide bandwidth that supports projected maximum concurrent voice calls and video sessions.

The following table demonstrates typical bandwidths for voice traffic on an enterprise network.

Type of traffic	Codec	Bandwidth* (kbps)
Voice	RT Audio	45
Voice	G.711	80

** Bandwidth includes voice payload plus RTP, UDP and IP packet overhead.*

The table below demonstrates typical bandwidths for video traffic using the DivX codec.

Type of client	Resolution	Frames/second	Quality	Bandwidth (kbps)
Home	160 x 120	8	Low	40-80
Office	320 x 240	10	Medium	150-300
Conference	352 x 288	15	High	400-800

Quality of service

As already stated, IP networks are inherently "best-effort" networks that treat all packets in the same way. A “best-effort” network has no specified parameters, does not guarantee how fast data is transmitted, and provides no assurances that data will even be delivered. A means of providing guarantees is obviously required.

The purpose of quality of service (QoS) mechanisms is to guarantee that the network treats certain packets in a specified way. QoS mechanisms refer to packet tagging mechanisms and network architecture decisions on the TCP/IP network that expedite packet forwarding and delivery. QoS is especially important for low-speed links, where the usual amount of bandwidth available is only several hundred kbps; data traffic could easily use all of the available bandwidth on such a link, causing voice-quality problems. QoS mechanisms can be used to guarantee that network bandwidth is available for voice traffic. To achieve good voice quality, end-to-end QoS is needed for IP telephony applications. End-to-end QoS can be achieved by ensuring that different parts of the network apply consistent treatment to the telephony packets.

QoS versus bandwidth

One approach to network engineering states that QoS is not needed and that increasing bandwidth provides enough QoS for all applications. This approach also states that implementing QoS is complicated and that adding bandwidth is easy. However, due to the nature of IP network traffic, even very large amounts of bandwidth might not be enough to prevent congestion during a burst of traffic at a particular instance in time. If all networks had infinite bandwidth available to ensure that network congestion never occurred, QoS technology would not be needed. However, while having adequate bandwidth provisioned on a network is very important, over-provisioning might not be very realistic and, therefore, QoS mechanisms are necessary.

Performance

For the purposes of this white paper, the term “performance” means the extent to which services that are operational meet the sensory expectations of users. Examples of where this performance is important include the audibility of voice, legibility of fax and timeliness of tone generation. The quality experienced by users of a service (QoE) can be predicted by characterizing the network using certain metrics. Quality is then regarded as acceptable if the values of these metrics are no worse than certain targets. Typically targets apply equally to all sources and destinations of traffic served by the network, at least when the sources and destinations are not too far apart. However, because different traffic types, such as media, signaling and management, affect perceived quality to different extents, targets can be different for different traffic types.

Voice quality is affected by distortion and delay. As TDM switches introduce little delay, the majority of delay in a TDM PSTN is due to the propagation time of the transport medium, which is proportional to distance for a given medium. Propagation time contributes equally to delay for TDM and IP networks, and can be discounted when estimating the impact on quality of replacing a TDM connection with an IP one on part of, or the entire, route taken by a call.

A network with low delay might be able to tolerate the introduction of additional delay without significant degradation in the voice quality, providing the total one-way delay remains less than 150 ms to 200 ms. Degradation to voice quality is very small below 150 ms and remains small until about 175 ms to 200 ms, when it starts to become much more severe. Delay and distortion in voice calls are affected by various factors that are specific to IP networks, including:

- Media gateway processing
- Packet delay variation or “jitter”
- Packet mis-sequencing
- Packet loss

Voice-quality metrics

Various metrics have been devised to quantify the overall perceived voice quality of a system component or an entire system. This paper discusses three common metrics:

- The subjective measure called Mean Opinion Score (MOS)
- An objective MOS estimator called PESQ (Perceptual Evaluation of Speech Quality, pronounced “pesk”)
- A computed metric called Transmission Rating (R), which is calculated from objective measurements of fifteen contributing parameters using an ITU standard tool called the E-Model

While voice quality metrics are evaluated for individual calls, there is no single value to describe a network. Networks carry many calls, both simple and complex, and quality is determined by the access types, transport technology, number of nodes the call passes through, distance, packet-transport link speeds, and many other factors that differ from one connection to another.

Mean opinion score

Mean opinion score (MOS) started as a subjective measure, but is now more often used to refer to an objective approximation of subjective MOS. Although all MOS metrics appear very similar (values between one and five with one or two decimal places) and are intended to quantify quality-of-experience (QoE) performance, the various metrics are not directly comparable to one another. This lack of comparability between metrics can result in confusion, since the particular metric used is almost never reported when MOS values are cited.

Subjective MOS is a direct measure of user perception of voice quality, and is thus a direct measure of QoE. Subjective MOS is the average of ratings assigned by subjects to a specific test case using methods described in ITU-T P.800 and P.830. Subjective MOS can be obtained from listening tests, where people rate the quality of recorded samples, or conversation tests, where people rate the quality of experimental connections. Quality ratings are judged against a five-point scale where five is equivalent to excellent, four is good, three is fair, two is poor and one is bad. MOS is computed by averaging all the ratings given to each test case, and falls somewhere between one and five. The higher the MOS, the better the perceived quality.

Perceptual evaluation of speech quality (PESQ)

Subjective studies take significant time and effort to carry out. MOS estimators such as PESQ can provide a quick, repeatable estimate of signal distortion. However, the score does not reflect the conversational voice quality, since listening level, delay and echo are excluded from the computation. Separate measures of these characteristics must be considered along with a PESQ score to estimate accurately the overall performance of a channel. PESQ is defined in ITU-T recommendation P.862.

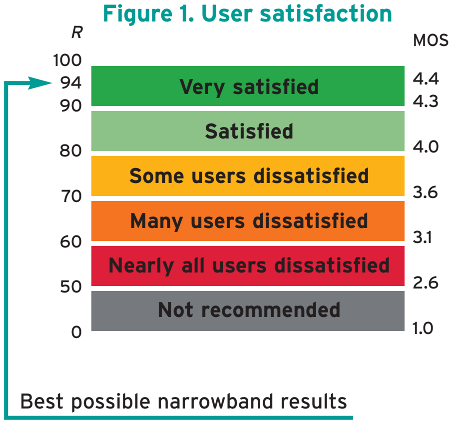
PESQ is an intrusive test, meaning that a tester must commandeer a channel and put a test signal through it. One or more speech samples are put through a device or channel, and the output (test signal) is compared to the input (reference signal). The more similar the two waveforms are, the less distortion there is, and the better the assigned score. The raw PESQ score is usually converted into a MOS estimate using one of several available conversion rules. An ITU-T standard conversion is defined in P.862.1.

Transmission rating (R)

Transmission rating *(R)* is an objective metric that indicates the overall quality of narrow-band conversational voice. *R* is the main output variable of the ITU E-Model (Rec. G.107). Fifteen parameters are used to compute *R*, including listening level, noise, distortion, codec(s) used, packet loss, delay and echo. *R* accounts for all factors that contribute to conversational voice quality, which makes it the only value needed to completely describe quality.

The table below shows scales of voice quality for *R* and MOS metrics. Ideally, voice quality should be above the *R*=80/MOS=4.0 mark; falling below this level will likely result in some users becoming dissatisfied with voice quality.

Figure 1. User satisfaction



Section three: OCS deployment

As with any deployment of applications within an enterprise, planning is crucial to a successful deployment of OCS 2007 Release 2. Key factors to reflect upon during planning include what features to deploy, what type of availability is required, what is the geographic distribution of users, what is the location user density, what external user access is required and selection of appropriate topology. To protect existing network equipment investment and avoid compromising security or carrier-grade reliability, planning and decision-making should occur in phases.

Phase one: Determine key planning considerations

In the initial planning phase, several key decision factors will affect how OCS 2007 Release 2 is deployed in any particular organization. Evaluate your organization's specific requirements and decide what OCS features will be enabled.

Phase two: Decide what OCS features will be deployed

OCS 2007 Release 2 Standard Edition server and Enterprise Edition pools deliver IM, presence, enterprise voice, audio/video conferencing and web conferencing for users within an organization. OCS also provides several supplementary features that can be offered by deploying additional server roles, depending on the functionality that an organization wants to support. For example, if an organization wants to provide internal users with the ability to communicate with external users, then an Edge server will be required. If an organization wishes to archive instant messages, an Archiving server will be needed.

Phase three: The importance of high availability

If an organization requires its OCS 2007 Release 2 topology to provide high availability, then one or more Enterprise pools should be deployed in the internal topology. If high availability is not a consideration, and simplicity and economy are more important, Standard Edition Server may be an appropriate choice.

If an organization plans to enable external access in a highly available topology, multiple Edge servers connected to a hardware load balance must be deployed in the perimeter network. Conversely, if an organization does not need high availability in the perimeter network, a single Edge server can be deployed.

If an organization must meet compliance requirements to archive instant messages and/or capture call-detail records (CDRs), the Archiving server and Monitoring server respectively can be deployed with a topology that delivers high availability.

Phase four: Geographic distribution

The geographic distribution of a company affects the design of its system topology. If an organization is geographically dispersed across WANs, it might want to consider placing a Standard Edition server or an Enterprise pool at each local site. The use of audio and video features requires greater bandwidth, and a better user experience is achieved with a local server or pool.

Phase five: Number of users at each location

Closely related to the geographic distribution of a company is the number of users at each location. If there are more than 100 users at a remote site that is connected by a WAN to a central site or data center, an organization should consider placing a local server at that site. Similarly, if an organization plans to support external access to web and audio/video conferences, and expects a high amount of use at the site, then it should consider deploying a local Consolidated Edge server because of the higher bandwidth requirements for this type of traffic.

Phase six: Support for external user access

External user access is not only a feature of OCS, but also has important planning implications. If an organization chooses to support external user access to its OCS topology, it needs to also plan for this support in the perimeter network topology. For any type of external user access, a Consolidated Edge server is required. If the deployment is small, a single server may be sufficient. Alternatively, a load balancer with multiple servers can be used to support a larger user base and achieve high availability.

Phase seven: Select topology

A variety of OCS topologies — from a very small or pilot deployment to a large global deployment — have been designed to provide IM and conferencing functionality and other OCS features. This paper features three general topologies, which are provided as sample deployments:

- Small to medium deployment scenarios present topologies appropriate for small or pilot deployments where high availability is not required. These topologies build upon a Standard Edition server to support internal IM and conferencing for a user base of less than 2,000.
- Centralized enterprise deployment scenarios present topologies that deliver high availability and support for a user population of more than 2,000 at a single physical site.
- Global deployment scenarios present topologies that span multiple sites and provide high availability and scalability.

Conclusion

The evolution to unified communications delivers significant productivity benefits: employees have greater flexibility over when, where and how they work, and are able to respond more quickly to the demands of a dynamic business environment. Teams can also use multi-modal collaboration capabilities to accelerate decision-making, project completion and development of new business opportunities. Taken together, these benefits result in significant improvements to business performance.

Nortel leverages partnerships with industry leaders, including Microsoft, to provide unique unified communications technology that delivers an open application ecosystem and network simplification. Under the Nortel and Microsoft Innovative Communications Alliance (ICA), seamless delivery of collaborative multimedia technology and tools to the voice communications network provides effective integration of Microsoft's unified communications domain to the SIP-based converged voice communications infrastructure using Session Initiation Protocol for Computer Telephony Integration (SIP-CTI). The strategic ICA is critical to successfully integrating OCS 2007 Release 2 with Nortel's enterprise telephony systems.

As with any deployment of applications within an enterprise, planning is key to implementing a unified communications environment based on the integration of enterprise telephony with OCS 2007 Release 2. To ensure success, engage targeted user communities early on in the process, and build a positive business case by leveraging existing investments and quantifying specific savings. Ensure the underlying data network requirements — including network capacity; routing efficiency; prioritized queues for QoS, security and redundancy (OCS 2007 Release 2) — are understood, and decide what features will be deployed. Standardize configuration and dialing procedures, and select the appropriate topology. Indeed, as within the real-estate market, a common phrase when planning an OCS topology is "location, location, location!"

In the United States:
Nortel
35 Davis Drive
Research Triangle Park, NC 27709 USA

In Canada:
Nortel
195 The West Mall
Toronto, Ontario M9C 5K1 Canada

In Caribbean and Latin America:
Nortel
1500 Concorde Terrace
Sunrise, FL 33323 USA

In Europe:
Nortel
Maidenhead Office Park, Westacott Way
Maidenhead Berkshire SL6 3QH UK
Email: euroinfo@nortel.com

In Asia:
Nortel
United Square
101 Thomson Road
Singapore 307591
Phone: (65) 6287 2877

Nortel is a recognized leader in delivering communications capabilities that make the promise of Business Made Simple a reality for our customers. Our next-generation technologies, for both service provider and enterprise networks, support multimedia and business-critical applications. Nortel's technologies are designed to help eliminate today's barriers to efficiency, speed and performance by simplifying networks and connecting people to the information they need, when they need it. Nortel does business in more than 150 countries around the world. For more information, visit Nortel on the Web at **www.nortel.com**. For the latest Nortel news, visit **www.nortel.com/news**.

For more information, contact your Nortel representative, or call 1-800-4 NORTEL or 1-800-466-7835 from anywhere in North America.

NN123226-060909

NORTEL

BUSINESS MADE SIMPLE