Scaling the Video Conferencing Environment

Benefiting from Scalable Video Coding and Media Routing without a Forklift Upgrade

Ira M. Weinstein
July 2013

Sponsored by:

Polycom
Introduction

The video conferencing industry has experienced a wide range of advancements in recent years. Some of these developments are what Wainhouse Research (WR) would categorize as evolutionary, such as the introduction of HD video resolution and the continuous price/performance improvements across the board. Other advances are more revolutionary, such as the ability to run video conferencing infrastructure on standard PC servers and virtual / cloud servers, and the availability of high-quality video conferencing on mobile devices.

This study focuses on the repercussions of the revolutionary advancements within the video conferencing space and the fact that video conferencing has finally stepped out of the boundaries of the enterprise meeting room and into the hands of the masses. Although relatively recent, the migration of video conferencing from the board room and onto users’ desktops, notebooks, tablets, smart phones, and even into their living rooms has already increased the value of visual collaboration for many information workers and organizations.

The ability to video conference with workers in meeting rooms, at their desks, and on the road has given rise to a new challenge ... the need for scale. Note, however, that this challenge does not directly relate to the video devices or users themselves. The crux of the issue is the need to scale the back-end infrastructure to provide the necessary multipoint video connections to support the requirements of not just a handful of group video rooms, but of tens of thousands of video users – all expecting to meet with each other.

The Drivers for Increased Multipoint Video Calling / Connections

There are a number of factors that are driving increased demand for multipoint video calls including:

Social Acceptance of Video – the recent entrants into the work force have grown up with video. For these workers, video is something they expect to use at home, in the office, and on the road. The introduction of these video-savvy folks into the workplace will yield increased video utilization and an increased need for multipoint video connections.

Increased Deployment Sizes / Expanded Device Support - In the traditional group video conferencing world, the number of systems deployed around an organization was limited; a typical installation might include 10 – 20 group video systems, while a large installation might include 100 – 200 systems.

Today, things are quite different. In addition to a few dozen group video conferencing systems, a typical enterprise might have a few hundred or thousand users on video-enabled PCs, tablets, and smartphones - all expecting to be able to participate in multi-point video calls on any of their devices at any time. The expansion of video conferencing beyond the meeting room has driven increased demand for multipoint video connections.
Increased Use of Video for Business Communications - Organizations today depend on video conferencing for both internal and external communications. Savvy companies are using video to communicate not only with colleagues and peers, but also with clients, prospects, business partners, and members of their supply chain. WR expects the use of Business to Business (B2B) and even Business to Consumer (B2C) video conferencing to increase dramatically over the next few years. This increased usage will drive additional demand for multipoint video connections.

Changing Meeting Paradigm – traditionally, video meeting participants would go to a video-enabled conference room to participate in a video meeting. Depending on the situation, a single group video room might support 5, 10, or even more participants. As a result, a video call involving 10, 20, or more participants might involve only 3 or 4 video systems (and use only 3 or 4 multipoint video connections).

In the new conferencing environment, many attendees will NOT go to the meeting room. Instead, these people will use their personal video systems (PC-based, tablet, smartphone, video phone, etc.). While this will not increase the number of people in the meeting, it can significantly increase the number of video systems / devices that need to connect to the multipoint video call. The table below shows the potential impact of this new meeting paradigm on a 24 person meeting.

<table>
<thead>
<tr>
<th>Typical Video Meeting</th>
<th>Old School (all participants report to video meeting room)</th>
<th>New School (half of the participants use personal video systems)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # of Participants</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td># of VC Rooms</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td># of Participants / VC Room</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td># of Participants across VC Rooms</td>
<td>24</td>
<td>12</td>
</tr>
<tr>
<td># of Participants using Personal VC</td>
<td>0</td>
<td>12</td>
</tr>
<tr>
<td># of Multipoint Video Connections Required</td>
<td>4 (1 per room)</td>
<td>16 (1 per room, and 1 per personal VC user)</td>
</tr>
</tbody>
</table>

Figure 1: Multipoint Video Connections Required Per Meeting

As shown above, the “old school” method of supporting this single video meeting would have required only 4 multipoint video connections, while the “new school” method requires 16 connections – a 4x increase! Imagine the number of simultaneous connections a busy enterprise might use.

All four of these drivers (social acceptance of video, increased deployment sizes, increased use of video for business communications, and the new meeting paradigm) will serve to increase the multipoint video calling and connection requirements of the typical video-enabled organization.

The section that follows provides insight into the options available to enterprises needing to support wide-scale multipoint video calling.
The Traditional Video Bridging Approach

Traditionally, video conferencing used advanced video coding (AVC) to create a compressed, single stream version of the incoming video signal. In the AVC world, conducting a video conference with more than two parties or locations (a.k.a. multi-way, multi-party, or multi-point video call) required the use of a centralized device called a video bridge.

To participate in a multipoint video meeting, each AVC system connects to the video bridge the same way it would connect to any other video system. While hosting an AVC multipoint video call, the video bridge must perform the following steps in real time and for the entire duration of the session:

1. Receiving – the MCU receives video, audio, and in some cases the content streams from each participating system.
2. Decoding – the MCU decodes (decompresses) the incoming video and audio streams.
3. Compositing – the MCU combines / mixes the decompressed streams to create the appropriate video and audio streams for each participating location.
4. Encoding – the MCU encodes (compresses) the new streams created during compositing.
5. Sending - the MCU transmits the appropriate newly created, compressed streams (video, audio, and possibly content) to each participating system.

The process of decoding, compositing, and encoding the video streams (steps 2, 3, and 4 above) is called transcoding. Because transcoding is extremely processor-intensive, the majority of traditional (AVC) video bridges in use today are hardware-based appliances with a large number of digital signal processors (DSPs) that are optimized for audio and video compression.
Options for Deployment

1) Customer Premise (CPE) Deployment
The most obvious way an organization can increase its AVC video bridging capacity is to buy more video bridges or expand the capacity of its existing deployment.

For example, a multi-national organization might choose to do the following:
- Add capacity (via software licenses or additional hardware) to its New-York video bridge
- Install new video bridges in its London and Singapore offices

Thanks to centralized management systems and load balancing technologies, this organization can manage these bridges via a single user interface and intelligently route its multipoint call traffic to the most appropriate bridge (based on location, capacity, etc.).

Authors Note - while technically possible, support for hosting a single transcoded multipoint video call across multiple video bridges (a.k.a. cascading) remains problematic with issues ranging from unacceptably long delays, small images of the participants on the other bridge(s), and the ability to view only the currently speaking site. Therefore, instead of dividing meeting participants across several bridges, a better approach would be to host the video meeting on a bridge with adequate capacity to host all participants.

Furthermore, depending on the situation, the cost per additional transcoded multipoint connection can range from $1,000 (for a software license expansion for an existing software-based video bridge) to $2,000 - $10,000 or more (for the purchase of a hardware video bridge). Given the widespread availability of low cost or FREE video endpoints today, the number of video-enabled devices can be huge. Hence the cost to meet the resultant multipoint demand is likely to tax an organization’s budget.

2) External Video Bridging Provider
Instead of buying video bridging hardware, an organization could choose to leverage the resources of an external service provider. High quality video bridging services are available from scores of network service providers, video bridging providers, managed service providers, and video resellers.

The primary benefit of this approach is the ability to outsource the cost, burden, and risk associated with video bridging to a third-party for whom video bridging is a core competency and for whom costs can be distributed over multiple clients.

The primary negative of the services approach is the need to pay for multipoint bridging on a usage-based model (per-minute, per-month, or per-user/system, etc.). For organizations with limited usage, this OPEX-based model may be far less expensive than the CAPEX model. However, as the number of video-enabled users increases, the usage volume and monthly fees will also increase. And once video becomes pervasive, the monthly fees could easily exceed the total cost of owning a video bridge(s).
Furthermore, in order to use an external provider, the enterprise endpoints must be able to connect to the provider’s video bridges. Depending on the situation, this may require any (or all) of the following:

- Modifications to the enterprise firewall rules
- The deployment of additional hardware or software solutions
- The purchase of dedicated links between the customer’s WAN and the service provider
- The deployment of additional Internet or WAN bandwidth to host the video traffic

In addition, some organizations may have security concerns associated with hosting their meetings on shared, hosted infrastructure.

**Summary**

Traditional (AVC) video bridges were designed to prioritize user experience and flexibility above scalability and cost effectiveness. As a result, in some situations traditional video bridges may be too expensive to support large scale multipoint calling.

Using an external video bridging provider is a solid option for many organizations – especially smaller companies or those with limited video usage. However, this approach does not fully address the problem of scaling the video bridging environment ... it simply outsources the burden to another entity and changes the associated expenses from a CAPEX to an OPEX model.

**Option 2 – The Media Routing Approach**

Another option for enterprises seeking to host a high volume of multipoint video calls and connections is to use media routing. Essentially, media routing involves the routing (meaning switching without processing) of video streams between the participating systems without the need for expensive, processor-intensive transcoding.

There are two basic forms of media routing in use in the video conferencing space today:

**Method #1 – Stream Routing (a.k.a. simulcasting or multi-encoding)**

Simulcasting involves the simultaneous creation of multiple complete and independent video streams from the original signal. For example, a video conferencing system might encode a camera signal into an H.264 180p / 30 video stream, an H.264 360p / 30 video stream, and an H.264 720p video stream.

Each of these video streams is then sent to a central device, called a media router, which sends each participating system a copy of the appropriate stream from each of the other participating systems.
The primary advantages of the simulcast approach are as follows:

1) Efficient, low cost infrastructure – this approach avoids the need for an expensive transcoding video bridge to support multipoint video calls.

2) Strong interoperability – each of the streams created by each system is 100% complete and can be received / decoded by any compatible system.

3) Future friendly – this approach allows the use of almost any video codec including commonly used codecs for video conferencing (e.g. H.264, H.263), alternative codecs (e.g. VP8), proprietary codecs (e.g. Microsoft RTV), and codecs that will be released in the future (e.g. H.265).

4) Support for hardware encoders – this approach allows the use of hardware-based encoders and decoders available from leading chip manufacturers (e.g. Texas Instruments). This is especially useful for systems leveraging x86-based architectures with relatively limited processing power.

The primary disadvantages of the simulcast approach include:

1) Processor power – this approach requires the video systems to have sufficient processor power to encode several streams simultaneously.

2) Increased upload bandwidth required – this approach requires each video system to send several complete streams to the media router, which may utilize as much as 30% more upload bandwidth than a traditional, single-stream approach.

Method #2 – Layer Routing (a.k.a. scalable video coding or SVC)
Layer routing requires the use of a codec (e.g. H.264 SVC which is Annex G of the H.264 / MPEG-4 AVC video compression standard) that can encode the source video stream into multiple sub-streams called layers. What makes the resulting video stream “scalable” is that the resulting stream can be decoded / viewed even if some of the layers have been removed.

The first and only required layer in an SVC video stream is the base layer, which in itself is a complete, video stream. The additional (and optional) layers in an SVC stream are called enhancement layers, each providing increased video resolution, increased frame rate, and/or increased video quality. The base layer can then be combined with one or more enhancement layers to create video streams with higher resolutions, frame rates, and quality levels.

Once created, the various layers are then sent to a central device, called a media router, which subsequently sends each participating system a copy of the appropriate layers from each of the other participating systems.
The primary advantages of the layer routing / SVC approach are as follows:

1) Efficient, low cost infrastructure – this approach avoids the need for an expensive transcoding video bridge to support multipoint video calls.

2) Exceptional network resiliency – the ability to view / decode the video stream even if some of the layers are missing packets or corrupted allows the layer routing method to operate well over lossy networks (e.g. the public Internet).

3) Real-time adaptation – the layers (number, type, and level) created by each system, as well as the layers provided to each system by the media router, can be adjusted in real time to adapt to changing network conditions and the capabilities of the participating systems.

4) Flexibility – the ability to combine the base layer with one or more enhancement layers allows the media router to provide each participating system with a combination of layers that match its requirements and capabilities.

The primary disadvantages of the layer routing / SVC approach are as follows:

1) Need for a scalable codec – unlike the simulcast model which can utilize almost any video codec, the layer routing / SVC approach requires a scalable codec. As of this writing, WR is aware of only a handful of scalable codecs including H.264 SVC and proprietary flavors of VP8.

2) Limited interoperability – the need to use of a scalable codec means that this approach does not support native interoperability with legacy, single-stream (AVC) video systems.

3) Processor power – this approach requires the video systems to have sufficient processor power to encode the incoming signal into numerous layers.

4) Increased download bandwidth – the need to send multiple layers to each participating system results in the use of 15 – 30% more bandwidth than a traditional, single-stream approach.

Summary
While there are pros and cons associated with each method, both simulcast and SVC support large scale, cost-effective multipoint video calling without the need for an expensive, transcoding video bridge.

Options for Deployment
Organizations seeking to leverage media routing (simulcast and/or SVC) have three basic options:

Option #1 - Forklift Upgrade
The enterprise can opt to decommission its existing, single-stream (AVC) video conferencing systems and video infrastructure devices (meaning primarily video bridges) and replace them with media routing capable solutions. The organization can either buy the required infrastructure (a CPE deployment), use a hosted service, or use a combination of locally installed (CPE) and hosted services.
The primary advantage of this approach is the simplicity of managing a homogeneous, all media routing environment. The primary disadvantage of this approach is the need to discard expensive and fully functional video systems and infrastructure. This alone makes this approach a non-starter for most organizations.

Another concern with this approach is that the resulting deployment may not natively interoperate with legacy AVC video systems that may be in use by clients and partners.

Option 2 – Parallel Environments
The enterprise may choose to keep its existing AVC video environment intact and deploy a media routing video environment in parallel. The advantage of this approach is that it allows an enterprise to continue benefiting from its prior video conferencing investments, while simultaneously supporting the low cost, highly scalable media routing architecture.

There are several disadvantages associated with this approach including:

- The need to use media gateways to connect the legacy AVC environment with the SVC environment. While functional, media gateways tend to add to the complexity of an environment, often requiring specialized dialing strings and advanced configurations.
- A compromised experience during calls between the legacy AVC environment and the simulcast environment due to the inability of the legacy systems to provide multiple streams.
- The need to purchase, maintain and support two distinctly separate environments. Once again, this is far from ideal.

Option 3 – A Hybrid Deployment
In this context, a hybrid deployment is one that natively supports both traditional (AVC) and media routing (simulcast and/or SVC) video conferencing solutions without the need for gateways. Depending on the existing deployment, this can be implemented in two ways:

- Forklift upgrade – replacing the existing video conferencing infrastructure with dual-capability devices. The obvious negative here is the up-front cost associated with the new infrastructure.
- Capability upgrade – adding media routing capabilities onto the existing infrastructure. The key benefits of this approach include:
  - Cost reduction compared to a forklift upgrade or a parallel media-routing architecture
  - Expedited deployment since no additional hardware is required
  - Decreased support burden since the hybrid environment can be managed using a single set of management tools
Solution Spotlight – Polycom’s Hybrid Architecture

The sponsor of this study, Polycom, has implemented a hybrid architecture that supports both traditional AVC (single-stream) and media routing (both SVC and simulcasting) video conferencing.

Video Conferencing Systems
Polycom offers a number of dual-capability (AVC and media-routing-capable) video conferencing endpoints including the RealPresence Group 300, 500 and 700 systems, the RealPresence Desktop software client, and the RealPresence Mobile application for iOS and Android devices.

When used in traditional environments, these video systems provide a single AVC stream to a centralized video bridge / MCU.

When used in a media routing environment, these systems create several (three by default) simulcast streams at different resolutions. Each of these streams is then “scaled” to include several (three by default) layers, each of which increases the number of frames per second. These three streams, each supporting three different frame rates, are then sent to the media router.

By using both scalable video coding (SVC) and simulcasting within its media routing architecture, Polycom provides its users with a compelling set of benefits including low cost, highly scalable multipoint video calling and enhanced interoperability.

Video Conferencing Infrastructure
Polycom’s hardware and software video bridges (listed as a ‘universal bridge’ in the diagram below) also support both media transcoding (for single-stream AVC systems) and media-routing.¹

---

¹ These capabilities are available out-of-the-box within the company’s software MCUs and as a software upgrade on its hardware MCUs.

---

Figure 3: Mixed AVC / SVC Meeting on a Universal Video Bridge
As a result, Polycom-powered organizations can conduct meetings involving traditional (AVC) and media routing systems without having to deploy any additional hardware or servers (e.g., gateways), or change their workflow. Users simply dial the call and the architecture does the rest – regardless of whether the users are using AVC or media routing endpoints.

The power of this multi-architecture support is further enhanced by the following:

1) Dynamic resource allocation which allows any Polycom video bridge to operate as a dedicated traditional AVC video bridge, a dedicated media router, or any combination in between. These capabilities allow organizations with Polycom video bridges to gain immediate access to media routing, while simultaneously enabling a gradual migration from traditional AVC video calling to media-routing --- all without the need to purchase and deploy additional hardware or servers.

2) The use of scalable audio coding (SAC) which extends the benefits of scalable coding to the audio streams. Key benefits include:

   a. Reduced burden on the infrastructure (the audio signals are routed, not processed)
   b. Exceptional network resiliency / protection from packet loss – in some cases up to 70%!

Given the importance of audio quality during video sessions, these benefits are significant.

**Conclusion**

Interest in video conferencing, for high-impact communications with both internal and external (partners, customers, prospects, etc.), continues to grow. This, in turn, has created an increased need for multipoint video calling.

Traditional multipoint video conferencing architectures were designed to provide an exceptionally high quality experience. Despite their strong performance, these traditional architectures are not well suited for large scale deployments involving thousands of users and systems.

A next generation architecture, dubbed media routing, resolves many of these issues and allows organizations to cost-effectively conduct wide scale multipoint video calling. For this reason, WR expects the video conferencing industry to migrate from traditional methodologies to media routing over the next few years.
Organizations wishing to embrace media routing have two basic options:

1. Implement a forklift upgrade; an option that is not only expensive, but also disruptive to the organization.
2. Run parallel environments connected by media / signaling gateways; an option that forces the organization to support and manage two environments, and is likely to complicate the workflow for the end-users.

The sponsor of this study, Polycom, offers a third option; a hybrid approach that:

1) Leverages the commonly used media routing methodologies; scalable video coding (SVC) and simulcasting.
2) Adds media routing capabilities to the existing AVC video conferencing environment.

This hybrid approach allows organizations to migrate gradually from traditional AVC video conferencing to media routing (SVC and simulcasting) in order to support a large scale multipoint video conferencing environment without the need to install new infrastructure, change end-user workflow, or shoulder an additional support burden.
About Wainhouse Research

Wainhouse Research, www.wainhouse.com, is an independent market research firm that focuses on critical issues in the Unified Communications and rich media conferencing fields, including applications like distance education and e-Learning. The company conducts multi-client and custom research studies, consults with end users on key implementation issues, publishes white papers and market statistics, and delivers public and private seminars as well as speaker presentations at industry group meetings. Wainhouse Research publishes a variety of reports that cover all aspects of rich media conferencing, and the free newsletter, The Wainhouse Research Bulletin.

About the Author

Ira M. Weinstein is a Senior Analyst and Partner at Wainhouse Research and a 20-year veteran of the conferencing, collaboration, and audio-visual industries. His prior experience includes senior positions with conferencing and AV vendors, distributors, and resellers. In addition, Ira ran the global conferencing department for a Fortune 50 investment bank. As the lead analyst of WR’s visual collaboration team, Ira’s focus includes video conferencing endpoints (mobile, desktop, group, and telepresence / immersive) and infrastructure, streaming / webcasting, and the visual communication elements within unified communications. Ira has published hundreds of articles, documents, and reports on AV and collaboration, is a frequent speaker at industry events, and regularly consults with end-users, vendors, service providers, and investment firms seeking to understand the collaboration space. Ira has been an industry analyst and consultant since 2001 and can be reached at iweinstein@wainhouse.com.

About Polycom

(Copy provided by Polycom)

Polycom is the global leader in open, standards-based unified communications and collaboration (UC&C) solutions for voice and video collaboration, trusted by more than 415,000 customers around the world. Polycom solutions are powered by the Polycom®RealPresence®Platform, comprehensive software infrastructure and rich APIs that interoperate with the broadest set of communication, business, mobile, and cloud applications and devices to deliver secure face-to-face video collaboration in any environment. Polycom and its ecosystem of over 7,000 partners provide truly unified communications solutions that deliver the best user experience, highest multi-vendor interoperability, and lowest TCO. Visit www.polycom.com or connect with us on Twitter, Facebook, and LinkedIn to learn how we’re pushing the greatness of human collaboration forward.