

No Jitter

Affordable Ubiquitous Telepresence

By Bob Emmerson | Apr 14, 2010

Telepresence is an application to die for, but the cost of telepresence rooms and of the network limits usage to large corporations. But increasingly, vendors are coming out with low-cost, high-quality solutions for the SMB price bracket. One such example is from Vidyo, which has introduced a telepresence room system that supports two video streams at 60 frames/second and the price is just under \$7K. To that you have to add the cost of one or two monitors. A "traditional" telepresence solution would have more monitors plus a fancy environment and the price would start at around \$100K.

In the newer iteration of telepresence, one monitor will display up to eight participants at a time; adding a second screen enables data to be displayed. This is judged to be sufficient for small and medium-sized businesses.

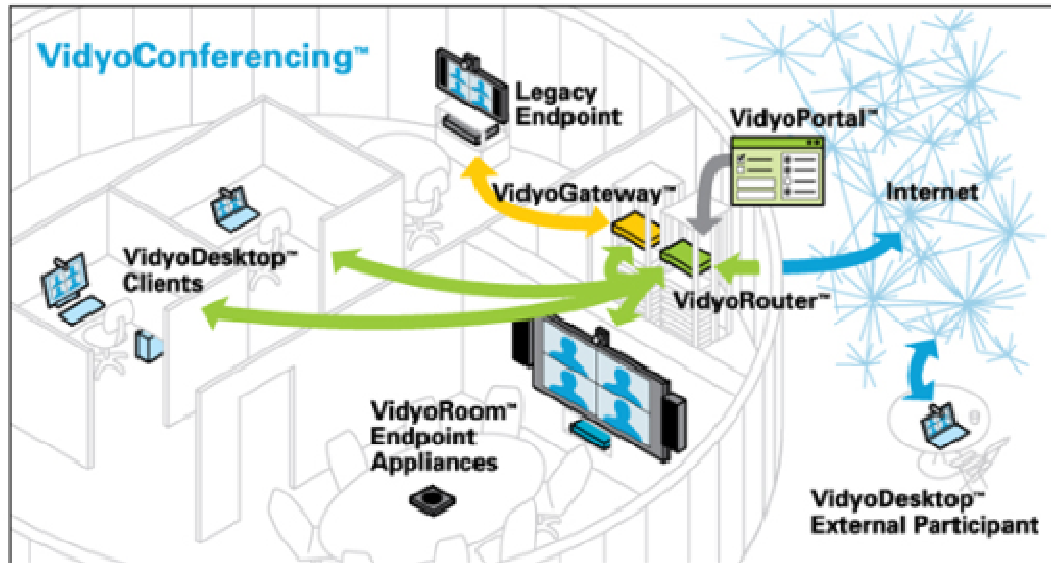
Vidyo in particular has gained attention for its unique enabling technology. For openers, unlike most telepresence systems, there's no MCU (multipoint conferencing unit). MCUs take encoded video from each participant, decode the signals, aggregate them into a composite stream, re-encode the result, and send it to the receiving participants where it is decoded.

Transcoding signals in this way is a slow process, one that introduces a delay of around 200ms. The network will typically introduce an additional 70 ms and the endpoints another 200 ms. The total delay can therefore exceed 400 ms, which is noticeable and annoying.

Vidyo's solution employs a router that routes encoded packets at multiple frame rates and resolutions to each endpoint. Because there is no transcoding, the added latency is under 10 ms. The transcoding encode/decode cycle is not required because the technology provides multiple endpoints with video streams that have been scaled to match that endpoint's available bandwidth, processing power, and resolution capability. This is done dynamically, throughout the call, and it has been realized using advanced software algorithms that run on Intel multi-core processors. The technology therefore enables ubiquitous, personal telepresence, i.e. desktop PCs and Macs that have a webcam can participate in conference calls. Video-enabled phones are set to follow.

As well as reducing latency to around 10 ms, Vidyo's high-speed implementation of video routing involves layering the media so that different endpoints receive individual treatments based on available bandwidth, processing power and resolution.

The solution for desktop end points is software-based and it is employed and managed via a Web-based portal (**see figure below**). This is basically an environment that administrators use to manage the system. However, Vidyo says that usage is very simple and this allows regular conference participants to initiate meetings via the Web from a standard browser.



The VidyoRouter architecture eliminates transcoding, routes packets to each endpoint individually, and enables transmission over best-effort networks like the Internet. The VidyoGateway provides interoperability with Polycom, Tandberg or other MCU-based systems. Teleworkers and road warriors can also participate in conferences.

SVC and the Net

Regular telepresence solutions need high quality networks and managed services, and this represents around two-thirds of the total cost. Vidyo claims its solution delivers equivalent or even superior quality over a best-effort network such as the Internet. This allows companies to build videoconferencing capabilities on top of their existing infrastructure.

Scalable video coding (SVC) is the technology that provides the flexibility in the video stream for the VidyoRouter to intelligently route only the packets needed by the decoding endpoint in order to create an optimal experience (for more in-depth descriptions of SVC, see blogs by John Bartlett [here](#) and [here](#)). In turn this allows the Internet to be employed. The following explanation comes from Dr. Thomas Wiegand, who is an authority on video compression and one of the chairmen of the Joint Video Committee responsible for the H.264 standard for video compression. Dr. Wiegand is also a member of Vidyo's advisory board:

An efficient system architecture for video conferencing over general-purpose IP networks has to look very similar to the rest of the Internet--that is, with little processing being required *inside* (my italicization) the network and instead being handled at the edge of the network, and with the network itself being a best-effort network. That means the video encoder and decoder at the endpoints should do almost all the processing, with the media routers in the network left to do only lightweight packet operations with practically no delay.

Error Resilience

The base layer of SVC is Advanced Video Coding (AVC), which is used in traditional video coding. However, AVC is sensitive to transmission errors and they will normally show up on screen.

The scalable video coding standard, SVC, adds enhancement layers that contain additional information used to achieve higher resolutions and frame rates for endpoints that can make use of this information. It can also be employed to compensate for packet loss without

noticeably impacting the user's experience. In addition, SVC technology features spatial scalability. This allows the router to drop some resolution details temporarily when packet loss takes place, while temporal scalability enables frame rate to be dynamically adjusted as changes take place on the network. Therefore the base layer is fully protected and the picture never breaks up.

Implementation and Usage

One or more VidyoRouters sit on the various LANs: the topology is flexible. They can be geographically dispersed, in which case the local end points default to the router that is closest to them.

Floating port licenses enable reuse of HD conferencing ports. They can "float" automatically to whichever router needs them. They are software-based ports running on Intel x86 hardware platforms, which will allow the solution to benefit from the same performance growth curve as the PC industry.

Only one screen is needed per end station: it can be a notebook PC or a monitor mounted on the wall of a conference room. There can be up to 100 concurrent participants in a single call (an unlikely scenario). A maximum of eight participants (the previous eight speakers) will be displayed at a time. A second screen can be employed to display data, e.g. a PowerPoint presentation. In this case it would become a collaborative videoconference.