Invention Title:	A mechanism to enable A/V conferences to be hosted by multiple Service Providers
Invention Summary:	Most IP-based audio/video conferencing solutions today are built around a centralized media server. This architecture enables the conference to scale to very large numbers of conferees by limiting the number of media streams that must be sent/received by each client device (as compared to a mesh conferencing solution, where each conferee must send and receive media to/from every other conferee). Typically, the media server is centralized and provided by a single service provider (e.g., GoToMeeting). This invention provides a way to distribute the media server resources within a service provider (SP) network, or across multiple service providers. This will improved scalability, and for the multi-SP case, will enable each SP to tailor the service to meet the special needs of their users.
Invention Description:	Today's audio/video conferencing solution is shown in Figure-1, where all users across all peering SPs access the conference bridge of the host SP directly. The peering SPs 2, 3, and 4 provide IP transport of media only, and do not provide any value-added services. Figure-2 shows an incremental improvement, where the peer SPs 2, 3, and 4 insert a media bridge device in the media stream to/from each client device. This would enable the SP to provide additional value-added services to their customers that might not be supported by the host SP-1; e.g., real-time speech-to-text translation for the hearing impaired, real-time language translation, audio/video recording, transcript generation, etc. Figure-3 shows the case where each SP hosts its own Selective Forwarding Unit (SFU). The SFU in the conference host SP-1 is designated "master", while the SFUs in the peer SPs 2,3,4 are designated "slave". In this example, each SFU is configured to selectively forward the audio/video streams of only the loudest talker. Each slave SFU makes its own decision regarding which user is the loudest talker, and which user has video focus (usually the loudest talker has video focus), and sends these media streams to the master SFU. The master SFU-1 makes an overall determination of loudest talker and video focus, and transmits these media streams to its own users, and to the slave SFUs. The slave SFUs in turn transmit the media streams received from the master SFU to their respective users. The slave SFUs do not relay any media received from their directly-connected users to their directly-connected users; all media from the peer SFUs is hair-pinned through the master SFU. There is a single entity - the master SFU-1 - making the overall decision on loudest talker and focus, so all conferees have the same/consistent view of the conference. The configuration shown in Figure-3 has the advantage over the configurations shown in Figures 1 and 2 in that there are fewer media streams exchanged among the SP peers (in this case the media streams associated with one user, instead of all users). This greatly improves scalability. This configuration would also enable the peer SPs to provide special services to their users; the services mentioned above for Figure-2, plus additional "conferencing" services such as the ability to have a private voice conversation among two/multi users during the overall conference (say a private client/lawyer consultation, or medical experts privately discussing a patient's diagnosis during a health-care session). The configuration in Figure-4 is the same as Figure-3 except that the master SFU is configured to transmit the media of loudest to talkers to the slave SFUs. In this example the loudest talker has video focus; i.e., each slave SFU receives two video streams - one for user-3 and one for user-5 - and they provide video focus to the loudest talker user-3.

	The configuration in Figure-5 shows a case where video focus is given to the presenter, instead of the loudest talker. In this case, since focus can't be determined implicitly by the loudest talker, the master SFU must signal the video focus to the slave SFUs. This is done by adding an RTP header field to the video media packets indicating which stream has focus. This mechanism to explicitly signal video focus in the media stream has the advantage over the implicit loudest-talker mechanism in that it guarantees all SFUs are presenting the same video focus (e.g., for the case where there are multiple talkers talking more-or-less at the same volume, or there's a talker in a noisy environment. Figure-6 shows the case where each SFU sends/receives media from all other SFUs; i.e., a mesh configuration. Each SFU relays media received from its own user to its own user - there is no media hair-pinning through the master SFU. This would slightly improve media latency over the Figure-3 configuration. To avoid media loops, the SFUs do not forward the media streams received from other SFUs; an SFU only forwards the media for its directly connected users to other SFUs (selecting the subset of media streams based on local configuration). To provide a consistent view of conference focus across all conferees, the master SFU must signal focus to the slave SFUs via an out-of-band mechanism.
Invention Commercial Value/Customers:	Improves scalability of audio/video conferencing solutions. Enables Service Providers participating in an audio/video conference to tailor the services and user experience to meed the special needs of their users.
Invention Differences:	Today's audio/video conferences are hosted by a single Service Provider.