CHAPTER 4

Call Control Protocols and IPv6 in IP Video Solutions

Revised: March 30, 2012, OL-27011-01

Protocols provide a complete set of specifications and suite of standards for communications between devices, This chapter does not discuss all the information available about the protocols but rather focuses on their most important features and characteristics in the context of handling video communications.

Call Control Protocols in IP Video Solutions

The primary call control protocols used in most IP video solutions today are H.323, Session Initiation Protocol (SIP), and Skinny Client Control Protocol (SCCP).

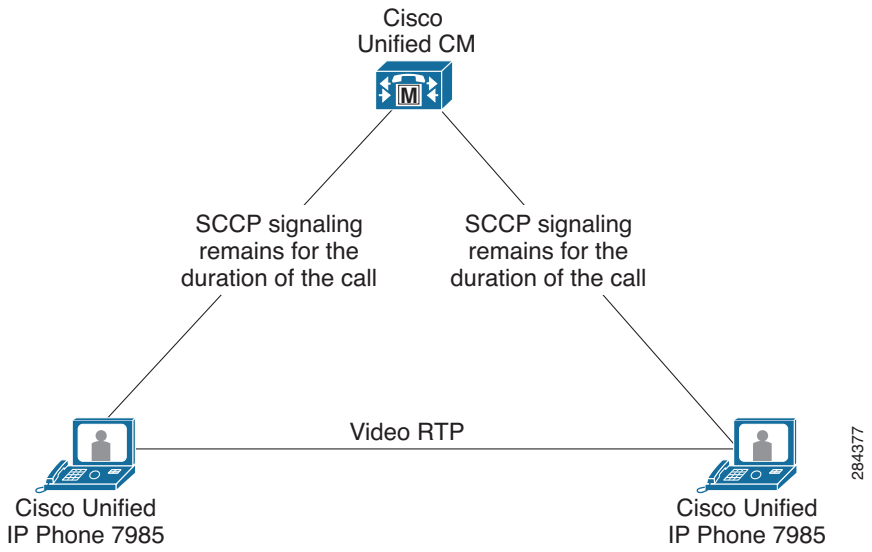
SCCP

Skinny Client Control Protocol (SCCP) was first developed by Cisco for IP Telephony applications. As IP Telephony matured, it integrated video as well and gave rise to Cisco IP Video Telephony. SCCP defines Transmission Control Protocol (TCP) as the transport protocol and a call agent in an architectural relationship with the endpoints (also known as a master/slave relationship). The call agent is the most fundamental difference between SCCP and the rest of the call control protocols discussed in this section. Because SCCP employs a central call agent, it inherently enables very advanced call functions for video endpoints that might not be available in other call control protocols.

Because SCCP defines a master/slave (or client/server) relationship between the call agent and the endpoints, the call agent must always remain available to the endpoint for call features to function. Therefore, SCCP might not be suitable for certain environments where the endpoints are expected to function independently from a call agent component.

Figure 4-1 illustrates the role of SCCP call control signaling in a deployment where Cisco Unified Communications Manager (Unified CM) is the call agent.

Figure 4-1 SCCP Signaling



As stated earlier, the SCCP specification provides support for advance call features in a video environment. Among those features, hold, resume, mute, and conferencing function exactly as they do for regular audio calls. The features that are most distinctive in SCCP endpoints are ad-hoc video conferencing and mute. Although support for ad-hoc video conferencing is not exclusive of SCCP, SCCP and the implementation of reservationless video conferencing in the video endpoints have made it easier for users to engage in ad-hoc video conferences. When the call control server is coupled with a compatible SCCP MCU, SCCP video phones are able to launch a conference by having users press a single key without making a previous conference reservation. This is an important difference from H.323, in which users must dial an always-on conference destination to establish a reservationless meeting.

The SCCP mute feature for video also functions differently than in other protocols. Unlike the mute function in H.323 and SIP, when mute is activated on the SCCP video terminal, both audio and video are muted simultaneously.

Just as SCCP enables intrinsically advanced call functionality on video endpoints through its phone-like technology and architecture, it also imposes some legacy phone-like behavior for video. Among the legacy behavior is the lack of support for uniform resource identifier (URI) dialing and data sharing. Therefore, when SCCP interoperates with some other protocol in a video deployment, it is important to consider any architectural limitations of SCCP. Table 4-1 lists other features not implemented in SCCP that should be considered when interoperating with H.323 or SIP for video.

Table 4-1 Features Not Implemented in SCCP

Feature not available in SCCP	Result	Workaround (if available)
Dynamic addition of video capability	Cannot promote an audio call to a video call	Ensure that video capabilities are available and broadcast at the beginning of the session
Far-end camera control (FECC) for SCCP endpoints	Cannot adjust remote cameras	Not available
Video codec renegotiation	Call session might be terminated if renegotiation occurs	Not available

The SCCP messages are encoded in hexadecimal, therefore reading them directly from the transmission is challenging. However, this encoding mechanism comes with the advantage that SCCP messages are generally smaller than with other call control protocols. For instance, an SCCP phone averages 256 bps in unencrypted call control traffic while a SIP phone averages 538 bps.

Another benefit of SCCP when used with video is that it allows authentication and encryption of media and signaling through Secure Real-time Transport Protocol (SRTP) and Transport Layer Security (TLS), respectively. When encryption is used, an SCCP video phone averages 415 bps while a SIP phone using encryption averages 619 bps.

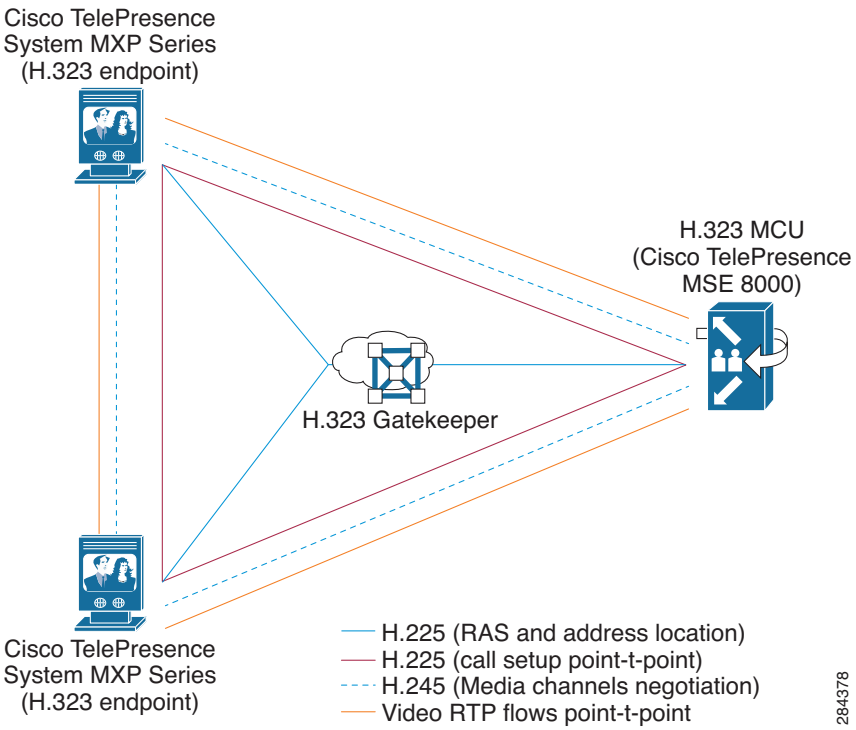
H.323

Unlike SCCP, H.323 is not a single standard or protocol but rather a suite of protocols and recommendations established by the International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T). H.323 is very strict in the definition of its features, expected behavior, and implementation, which puts H.323 in an advantageous position for interoperation between telecommunication vendors and providers. Because H.323 implementation is so well defined, it leaves very little room for misinterpretation of what is expected from the vendors when they interoperate.

H.323 uses a peer-to-peer protocol model that supports user-to-user communication without a centralized call control element. Because of H.323 robustness, it is not uncommon to find call control elements such as gatekeepers peered with endpoints from different vendors. As described earlier, H.323 is an umbrella protocol. An H.323 peer negotiates call setup and call admission control using H.225 and media channels using H.245. While H.225 uses User Datagram Protocol (UDP) and TCP as transport protocols, H.245 uses TCP only. Although this seems inconvenient for firewalls, H.323 is so well established in the telecommunications industry that most firewall vendors can efficiently inspect H.323 packets.

Figure 4-2 illustrates the use of H.323 with a gatekeeper as the call control element.

Figure 4-2 H.323 Signaling



H.323 provides strong support for a wide variety of video conferencing features, the most prominent of which are application sharing and far-end camera control (FECC). H.323 endpoints use H.224 and H.281 for FECC and H.239 for data sharing. FECC and application sharing in H.323 are key architectural differences between H.323 and other call control protocols. For instance, while SIP does not define how application sharing should be implemented, H.323 defines it clearly through Annex Q and its implementation of H.281 and H.224. With FECC in H.323, the camera control instructions are embedded into H.281 and later encapsulated in H.224, and RTP therefore provides a robust approach for transmission of the FECC instructions in the existing network infrastructure.

Application sharing is also very well defined in H.323, which uses H.239 to support it. H.239 defines how the management and addition of extra video channels must be implemented, and then the application video is sent over the additional video channel. Moreover, using a token system, H.239 ensures that only one participant at a time utilizes the application sharing functionality in the meeting.

Some features differ greatly between H.323 and other protocols. For example, some H.323 endpoints implement ad-hoc conferencing, but H.323 does not specify a central call control element in its architecture to execute conference resource tracking and establish conferences. Therefore, the ad-hoc conferencing behavior in most H.323 endpoints requires users to dial an always-on conference bridge.

Another example of protocol differences is that H.323 defines media encryption through H.235, but the definition of signaling encryption is not in the scope of H.323. Therefore, H.323 implementers commonly use either TLS or Internet Protocol Security (IPsec) when they need to secure the call signaling. This could potentially cause interoperability problems between endpoints from different vendors that use different approaches for securing the call signaling.

Although H.323 is highly evolved in its specification, H.323 features support only video and voice; they do not extend to instant messaging, presence, or other services. The lack of support in H.323 for new services should be thoughtfully considered when designing an IP Video network that might later integrate more communication methods besides voice and video alone.

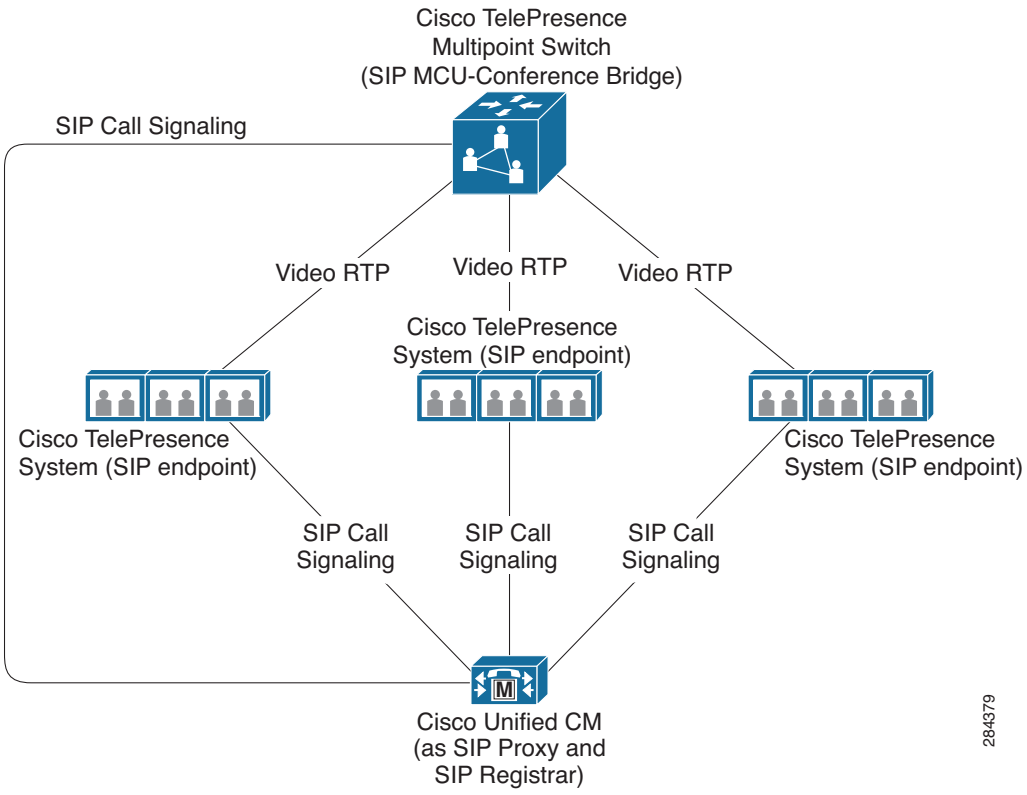
In addition, H.323 messages are encoded in binary, making it fairly challenging to interpret them without an appropriate dissector and potentially resulting in little-endian and big-endian errors when implementing the protocol messages. Although the H.323 messages are smaller than SIP messages, the difference in bandwidth can be considered negligible.

SIP

Session Initiation Protocol (SIP) is a peer-to-peer protocol. In the simplest implementation, SIP endpoints do not need a call control entity to contact each other, assuming they know their location. However, SIP also defines a client/server relationship so that the endpoints can make use of services, resources, and dialable destinations that are unknown to the endpoints. SIP is defined by the Internet Engineering Task Force (IETF) and is conglomeration of Requests for Comments (RFCs). Although the SIP core rules are defined by RFC 3261, the SIP standard includes more than a dozen RFCs.

In most enterprise deployments that use SIP, it is deployed with a call control element (client/server model) to provide a feature-rich experience, control over the dialable domains, and centralization of call control. SIP elements consist of two basic categories: user agent client (UAC) and user agent server (UAS). The element requesting connection to another element is the UAS, while the element receiving the request is the UAC. During a session, the same end-party can be a UAC for one transaction and a UAS for another, and the role is limited to a single transaction.

Figure 4-3 SIP Signaling



In many regards, SIP is better categorized as a communications session signaling protocol than a telecommunications signaling protocol because SIP enables more than just the establishment of voice and video communications. SIP can enable instant messaging, presence, and so forth, whereas SCCP and H.323 are purely telecommunications protocols. Part of the strength of the SIP protocol specification to support a myriad of services comes from the fact that UAS and UAC elements must ignore what they do not understand or support. On occasion, however, this strength becomes one of SIP's disadvantages because it complicates interoperation between vendors. Furthermore, SIP is less detailed in its specification than SCCP or H.323, making vendor interoperation somewhat challenging at times. For example, in SIP there is more than one way to implement some features. If different vendors implement the same feature in different ways, they would be incompatible.

It is also important to note that some features defined in other call signaling protocols are either not defined in SIP or function differently than in the other protocols. For instance, before RFC 4353, there was no standard to define how ad-hoc conferencing should be implemented, and SIP implementers took different approaches to fill the void. In Cisco IP Video Telephony, ad-hoc conferencing was implemented by creating a proprietary approach using XML.

Another example of a gray area in SIP is application sharing. Some implementers use the 'm' (media-type) attribute to specify when application sharing media will be sent and when an additional video channel will be set up. However, SIP does not clearly define how these features should be implemented, which makes application sharing between SIP vendors challenging.

SIP is text-based and uses the ISO 10646 character set encoded in 8-bit Unicode Transformation Format (UTF-8). A SIP phone averages 538 bps for call control traffic in unencrypted mode, while an SCCP phone averages 256 bps. SIP can use either TCP or UDP. SIP implementations typically use port 5060 but SIP can also be implemented on a different port.

Call Control Protocol Selection in IP Video Solutions

Selecting the right protocol for the design of the IP video solution is crucial to its success. The wrong choice of protocol could result in scalability issues and/or the inability of users to execute expected features.

When selecting the call control protocol for an IP video solution or a call leg section, consider the following factors:

- What call features are currently needed by the users and what features are planned for the future? (For example, data sharing, encryption, and so forth)
- Which transport protocol (TCP or UDP) will be used? Some call control protocols are better suited to a particular transport protocol.
- Are network characteristics such as Network Address Translation Traversal (NAT-T) or deep inspection (security) needed? Sometimes video endpoints might need to be behind a firewall, NAT support might be required, or payload encryption might form part of the requirements.
- What interoperability will be needed (for example, with third-party H.323), and what type of endpoints and MCUs will be used? A particular call leg might include devices that do not support the protocol selected for the overall design. For example, interoperability might be required with the IP PBX where the audio deployment resides, and that IP PBX might use H.323.
- Will business-to-business (B2B) communications be needed? If so, will a B2B vendor be used or will there be a direct connection to a third-party company? If a B2B vendor is used, what call control protocol has the B2B vendor implemented?
- What are the application sharing requirements? For example, will H.239 be required?

The more information you can gather about the call control protocol usage and roadmap, the better the decision making process will be. Any additional information that is specific and relevant to you solution should be included in the criteria for protocol selection.

After gathering all the information needed for protocol selection, you can begin the selection process. When selecting the protocol, pay particular attention to the following areas:

- Scalability — Based on the information gathered, how much growth is expected in the IP Video solution deployment and how will that impact the protocol selected and the call control elements in the deployment.
- Use cases — Based on the call flows that are meaningful for the success of the deployment and the requirements gathered, real-world scenarios should be developed and careful studied to determine how the protocols affect them. For example, if users are expected to share applications through the video endpoints without access to laptops, that would reduce the protocol options to H.323 and SIP only.
- Customer requirements — Occasionally there might be requirements that do not fall into a clear use case or scalability area. Depending on the importance of the requirement, it can be assigned a certain weight for purposes of the protocol selection process.

IPv6 in IP Video Solutions

IP version 4 (IPv4) has by now exhausted all the public IP address assignments. There still is room in the private address ranges for a large enterprise to be able to increase its operations. Nevertheless, mobile devices are increasing exponentially the number of IP devices connecting in the enterprise, and as that trend continues, the implementation of IP version 6 (IPv6) eventually will be necessary to increase the number of available IP address assignments.

The risks of ignoring IPv6 and not planning ahead could range from inability to connect new devices to diminished business-to-business capabilities. Those risks, however, are in the mid-to-long-term future, given the fact that there are currently enough addresses for private usage.

Cisco already supports IPv6 in certain devices, such as the Cisco TelePresence C Series and Video Communication Server (VCS), while at the same time working on integration of IPv6 into the rest of its IP video portfolio. However, not many IP video equipment manufacturers support IPv6 at this time.

The best approach is to be prepared, know your network, and track the IP address assignments to understand when migration to IPv6 will be necessary for your network. When deploying a new IP video solution, ensure that the manufacturers and products you select have a clear roadmap for IPv6, and understand how much work the migration will take so that you can plan accordingly.

If your IP video solution requires internetworking of IPv4 and IPv6 devices, the Cisco VCS currently offers address translation between IPv4 and IPv6. For further information, refer to the documentation at

http://www.cisco.com/en/US/products/ps11337/prod_maintenance_guides_list.html