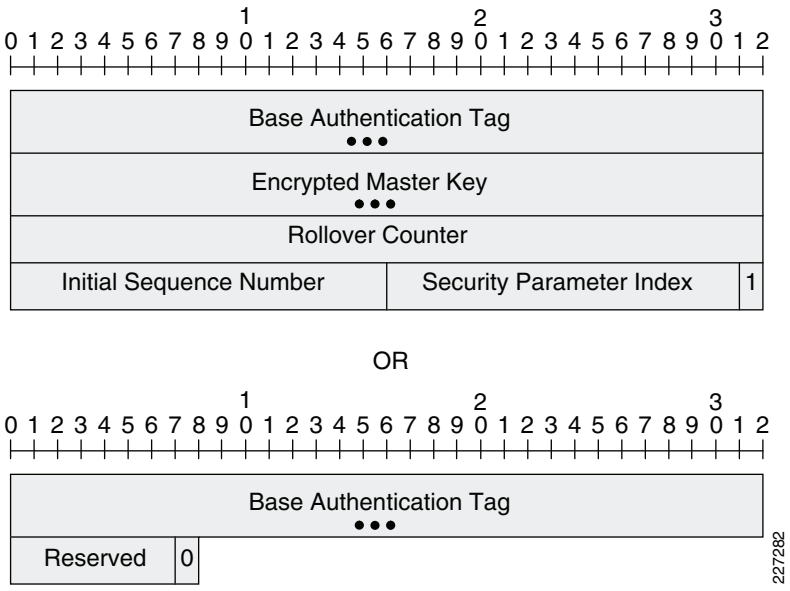
Encrypted Key Transport (EKT) and CTMS Secure Communications

Revised: November 2014, OL-18391-01

Understanding Encrypted Key Transport (EKT) and CTMS Secure Communications

EKT is an extension to SRTP which provides for the transport of SRTP master keys securely within SRTP and SRTCP packets. EKT accomplishes this by sub-dividing the SRTP and/or SRTCP Authentication Tag into one of the two formats shown in Figure 1.

Figure 1 Encrypted Key Transport (EKT) Packet Formats



In the first (long) format, the final bit of the Authentication Tag is set to 1. This indicates the presence of the following fields:

- **Base Authentication Tag**—Configured length field used to hold the message authentication data for the RTP or RTCP header and payload for the particular packet.
- **Encrypted Master Key**—Variable length field which contains the encrypted SRTP master key corresponding to the SSRC within the SRTP or SRTCP packet.
- **Rollover Counter**—32-bit field used to hold the value of the SRTP rollover counter associated with the SSRC contained within the SRTP or SRTCP packet. Since the RTP sequence number is only a 16-bit long field, the rollover counter is necessary for codecs to maintain synchronization and minimize re-keying in long term media streams.
- **Initial Sequence Number**—Indicates the RTP sequence number of the first RTP packet which will be protected by the SRTP master key contained within the Encrypted Master Key field of this SRTP or SRTCP packet.
- **Security Parameter Index**—16-bit field similar to the IPSec SPI. It is used to identify a particular security context (Key Encrypting Key (KEK) used to produce the Encrypted Master Key, KEK cipher, SRTP cipher, SRTP master salt, etc.) corresponding to a particular SRTP flow.

In the second (short) format, the final bit of the Authentication Tag is set to 0. This indicates the presence of the following fields:

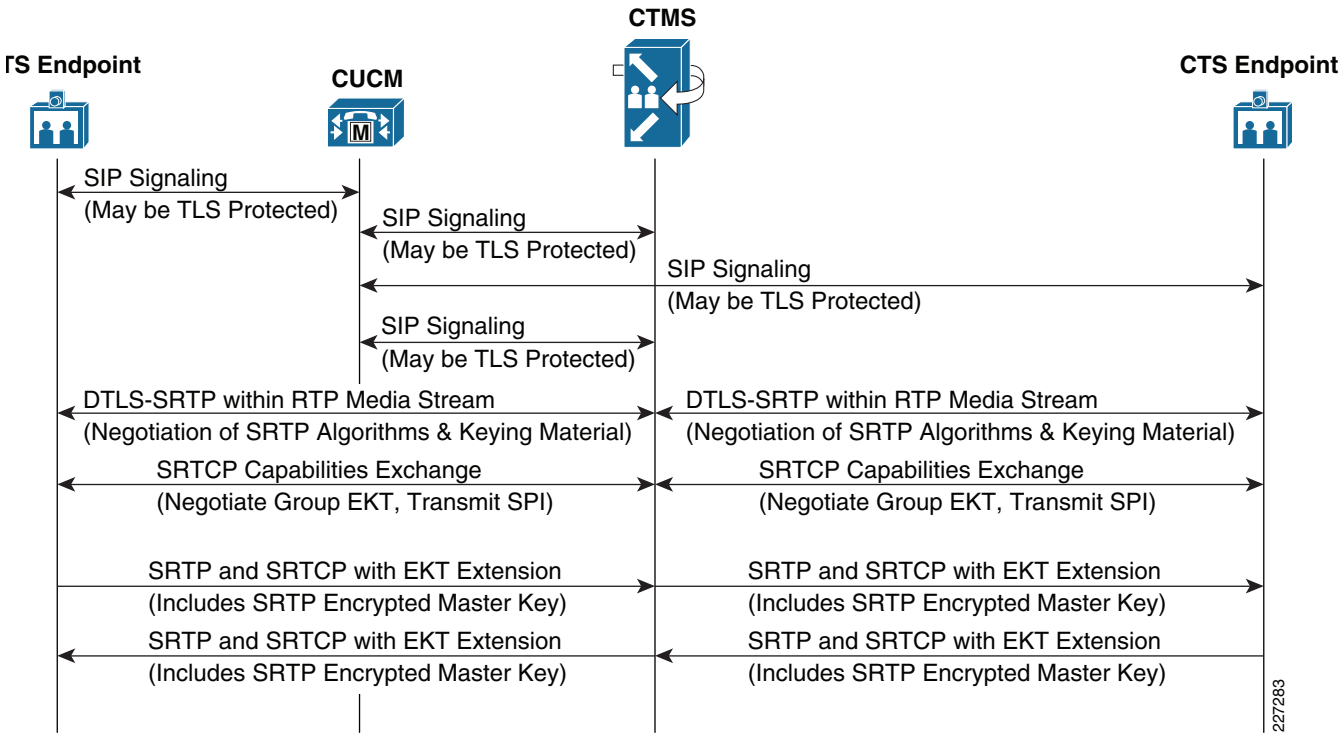
- **Base Authentication Tag**—Configurable length field used to hold the message authentication data for the RTP or RTCP header and payload for the particular packet.
- **Reserved**—7-bit field reserved for future use.

With EKT, the SRTP master key is distributed directly between CTS endpoints (CTS 3200s,CTS 3000s, CTS 1300s, CTS 1000s, and CTS 500s) in a multipoint call by encrypting it with a Key Encrypting Key (KEK), and sending it within SRTP and SRTCP packets which contain the long format of the EKT extension shown in Figure 1 above. This is done for each of the voice and video media streams. The following section provides a high-level overview of the key exchange process within a multipoint Cisco TelePresence call using DTLS-SRTP and EKT.

Cisco TelePresence Multipoint Call Operation

Figure 2 shows the process that occurs for establishing a secure multipoint Cisco TelePresence meeting. Figure 3 shows two CTS 1000 systems in a multipoint meeting via a CTMS. As with previous figures, only one set signaling and media has been shown for simplicity.

Figure 2 Multipoint Cisco TelePresence Key Exchange



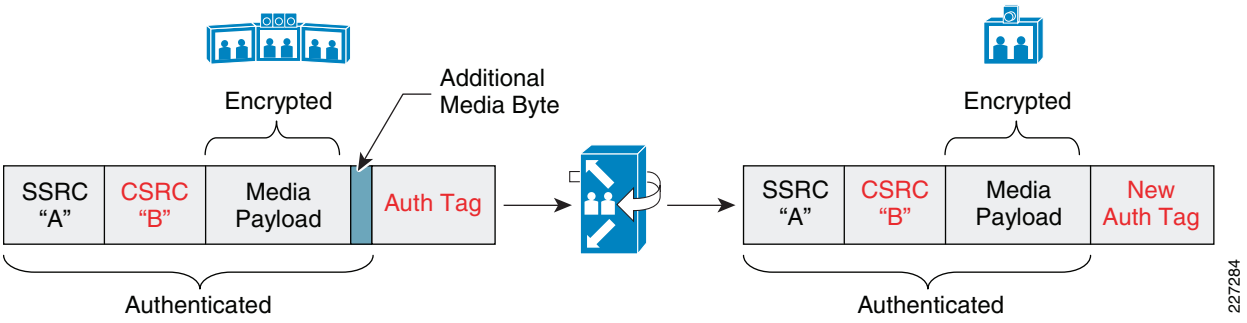
SIP signaling occurs first between the CTS endpoints and the Cisco Unified CM, as well as between the CTMS and the Cisco Unified CM. SIP establishes the RTP media streams and the RTCP control streams between the CTS endpoints and the CTMS. Immediately after the RTP media streams are set up, a DTLS-SRTP session is established over each media stream between the CTS endpoints and the CTMS. This is used to negotiate the SRTP encryption and authentication algorithms, as well as keying material. Following this, the CTS endpoints and the CTMS exchange SRTCP packets (using the SRTP keys established within DTLS) in order discover their capabilities.

During this step the CTS endpoints discover that they are communicating with a CTMS, and that EKT is required for SRTP master key exchange between the CTS endpoints. EKT is necessary for SRTP master key exchange because the CTMS does not actually de-encrypt the SRTP media packets and re-encrypt them. Therefore, each CTS endpoint needs the SRTP master keys used to encrypt the media flows from each of the other CTS endpoints in order to decrypt the media. Also during this step, a group EKT parameter set is established. The group EKT parameter set includes the Key Encrypting Key (KEK), which will be used to encrypt the SRTP master keys sent within SRTCP and SRTP packets which include the long format of the EKT extension shown in Figure 1 above.

SRTP encrypted master keys are sent by the individual CTS endpoints to the CTMS via SRTCP Source Description (SDES) packets which contain the long format of the EKT extension. The CTMS will forward these packets to other CTS endpoints within the multipoint call. Additionally, for video streams, the first SRTP packet of each video frame contains the EKT extension with the SRTP encrypted master key. All other packets of the frame contain the shorter, second format of the EKT extension shown in Figure 1 above. For voice, every third RTP audio packet contains the EKT extension with the SRTP encrypted master key. The other audio RTP packets contain the shorter EKT extension. By forwarding these voice and audio SRTP EKT packets to the other CTS endpoints, the CTMS further guarantees that each endpoint acquires the SRTP encryption key for each transmitter.

Figure 3 shows a high-level example of the video media switching performed by the CTMS during a secure multipoint Cisco TelePresence meeting. The SRTP packet has been intentionally simplified for this example.

Figure 3 Media Switching in a Secure Multipoint Cisco TelePresence Meeting



Cisco TelePresence endpoints use the Contributing Source Identifier (CSRC) field to indicate the source camera or microphone position of the media stream, as well as the destination display or speaker of the media stream, as opposed to the Synchronization Source Identifier (SSRC) which was previously used. The CTMS may modify the CSRC value in cases such as when the video from CTS 3000 *right* camera needs to be displayed on the *right* display of a CTS 3000 and on the *center* display of a CTS 1000 within the same multipoint meeting. However, the SSRC value of the media packets is not modified, as they are switched through the CTMS.

CTS endpoints also append an additional unencrypted byte to the payload of audio and video media packets before sending the packets to the CTMS. For audio packets, the additional byte is used to determine the level of audio energy within the packet. This is used by the CTMS to determine which CTS endpoints should be transmitting video streams. For video packets, the additional byte is used to determine if the video packet is the beginning of a new reference frame (IDR). For both the audio and video packets, the CTMS strips off the additional byte before forwarding it to the CTS endpoints.

Therefore, the CTMS does need to recompute the authentication tag as it switches the media, as shown in Figure 3 above. However, the CTMS does not decrypt and re-encrypt the actual media streams. This helps to ensure high performance and low delay of the CTMS even with encryption enabled on the multipoint call, as well as preserving the existing hardware investment of the CTMS.

For more information about EKT and Cisco TelePresence secure communications and signaling, see the *Design Zone for Video: Cisco TelePresence Secure Communications and Signaling Guide* on Cisco.com.