Network Working Group
Internet Draft
Intended status: Informational
Expires: April 27, 2015

P. Jones (Ed.)
N. Ismail
D. Benham
N. Buckles
Cisco Systems
J. Mattsson
Y. Cheng
Ericsson
R. Barnes
Mozilla

October 27, 2014

Requirements for Private Media in a Switched Conferencing Environment
draft-jones-avtcore-private-media-reqts-00

Abstract

This document specifies the requirements for ensuring the privacy and integrity of real-time media flows between two or more endpoints communicating in a switched conferencing environment. This document also provides a high-level overview of switched conferencing in order to establish a common understanding of the goals and objectives of this work.

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on April 27, 2015.

Table of Contents

1. Introduction

Users of multimedia communication products and services have privacy expectations that are largely satisfied with the use of SRTP [RFC3711] and related technologies when communicating point-to-point over the Internet. When communicating in a conferencing environment with two or more participants, though, it is necessary for an endpoint to share the SRTP master key and salt with the conference

server so that it can authenticate and decrypt received RTP and RTCP packets. The conference server also needs the master key and salt in order to transmit media packets it receives to other participants in

the conference. The need for conferencing servers to have the master key is a security risk for users.

Within a corporate or other isolated environment where conferencing servers are tightly controlled, this security risk can be effectively managed. However, managing this risk is becoming increasing difficult as conferencing resources are being deployed in networks that are less than fully trusted, including virtualized conferencing servers deployed in cloud environments.

There are also public voice and video conferencing service providers in which users must place full trust in order to use those services, as it is necessary for an endpoint to share the SRTP master key with those conferencing servers. This exposes corporations, for example, to a higher risk of being subjected to corporate espionage. While it is not the intent of this draft to suggest that any existing service provider would permit or condone any illicit use of its service, the fact is that security threats can come from external sources and remain undiscovered for long periods of time.

It is possible to ensure communication privacy within the context of a switched conferencing environment with limited changes in the security mechanisms used today. This document discusses this possibility in more detail and presents a set of requirements for meeting this objective.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119] when they appear in ALL CAPS. These words may also appear in this document in lower case as plain English words, absent their normative meanings.

3. Terminology

[Editor's Note: we may want to refine these or add/remove terms]

Adversary - An unauthorized entity that may attempt to compromise the performance of a conference server through various means, including, but not limited to, the transmission of bogus media packets or attempt to gain access to the plaintext of the media.

Switching conference server - A conference server that does not decrypt RTP media flows or perform processing on the media payload, but instead simply forwards the received media from a sender to the other participants in a multimedia conference. A switching conference server may modify some RTP headers.

4. Background

Traditional multimedia conferencing servers would mix, transcode, transrate, and/or recompose media flows from one or more conference participants, sending out a different audio and video flow to each participant. For audio, this might entail mixing some number of input flows that appear to contain audio intended to be heard by the other participants, with each participant receiving a flow that does not contain that participant's own audio. For video, the conference server may elect to send only video showing the current active speaker, a tiled composition of all participants or the most recent active speakers, a video flow with the active speaker presented prominently with other participants presented as thumbnail images, or some other composite arrangement. It is also common for audio or video to be transcoded. A typical traditional conferencing server is depicted in Figure 1.

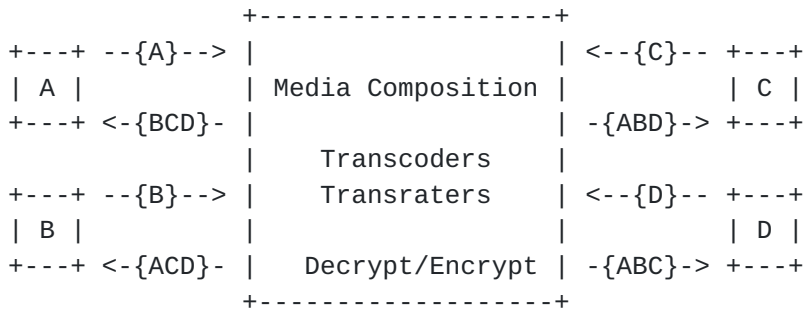

Figure 1 - Traditional Conferencing Server

Traditional conference servers require a significant amount of processing power, which in turn translates into a high cost for conferencing hardware manufacturers. Significantly, too, it is very difficult to deploy these servers in a cloud environment due to the high processing demands, as the specialized hardware found in the traditional voice and video conferencing server does not exist in a cloud environment.

To enable the traditional conferencing server to perform its job, the server establishes an SRTP session with each of the conference participants so that it can get the keys required to decrypt and encrypt media flows from and to each participant. This means that the conference server is necessarily a fully trusted entity in the communication path. Anytime these servers are deployed in a network that is not tightly controlled, it increases the risk that an attacker might gain access to cryptographic key material, thus allowing the attacker to be able to see and listen to ongoing conferences. In some instances, depending on how the hardware is designed and how keys and certificates are managed, it might be

possible for an attacker to see and listen to previously recorded conferences or future conferences.

The Secure Real-time Transport Protocol (SRTP) [RFC3711] is a profile of RTP, which can provide confidentiality, message authentication, and replay protection to the RTP traffic and to the RTP Control Protocol (RTCP). Encryption of header extension in SRTP [RFC6904] provides a mechanism extending the mechanisms of [RFC3711], to selectively encrypt RTP header extensions in SRTP. [RFC3711] and [RFC6904] solves end-to-end use cases between two endpoints, and does not consider use cases where a sender delivers media to a receiver via a cloud-based conferencing service.

5. Motivation for Private Media in Switched Conferencing

5.1. Switched Conferencing in Cloud Services

There is a trend in the industry for enterprises to use cloud services to host multi-party conferences and meet-me services, either exclusively or to meet peak loads on-demand. At the same time, there is huge shift toward using light-weight, cost-effective switching conference servers in cloud services that do not necessarily need to mix audio or composite/transcode video. Also fueling the use of such light-weight conference servers is the desire to fully exploit virtualized computing resources and dynamic scalability potential available in cloud computing environments.

The increased use of cloud services has exposed a problem. There are two different trust domains from a media perspective: endpoints and other devices in a trusted domain, and conference servers controlled by the cloud service in an untrusted domain. Other examples of conference devices spread across trusted and untrusted domains are likely, but the cloud service trend is triggering the urgency to address the need to allow for lightweight media conference while enabling media privacy at the same time.

With a switching conference server, each participant transmits media to the server as it would with a traditional conferencing server. However, the switching conference server merely forwards media to the other participants in the conference (where the other participant may be associated with a cascaded conference server or an endpoint on the same server), leaving composition to the receiving endpoint. Since some endpoints may have a limited amount of bandwidth, each endpoint might negotiate with the switching conference server to receive only a subset of the available media flows. Each transmitting endpoint might also send multiple media flows of varying frame sizes and/or frame rates (e.g., simulcast or scalability layers), so that the server can select the streams most appropriate for each receiver's bandwidth and capabilities. This allows, for example, an endpoint to receive and display higher quality video for the active speaker and thumbnails for other participants. It is also worth noting that, for

switched media to work successfully, each endpoint in the conference must support the media formats transmitted by all other entities in

the conference. More modern endpoints support multiple codecs and formats, making this commercially practical.

Figure 2 depicts an example of a switching conference server wherein each participant is receiving the media flows transmitted by each of the other participants in the conference.

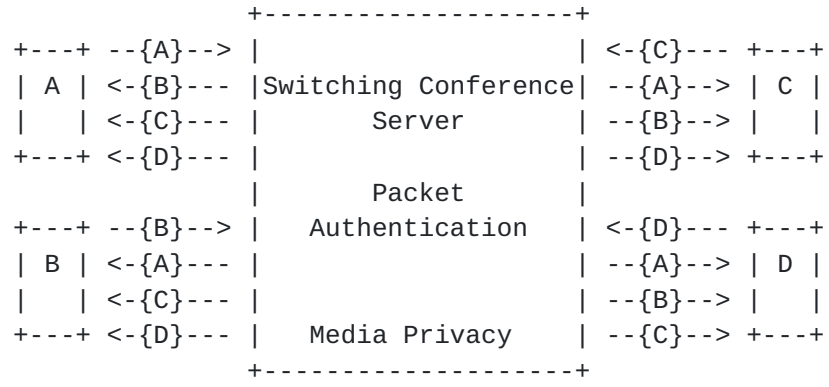

Figure 2 - Switching Conference Server

Note - The use of multiple arrows directed toward each endpoint is not intended to suggest the use of separate RTP sessions.

By using methods such as those described in [RFC6464], it is possible for the switching conference server to transmit the appropriate audio and video flows to conference participants without having knowledge of the contents of the encrypted media. The examples that follow help to illustrate this point.

In the Figure 3 below, endpoints A, B and D receive the video streams from endpoint C, the currently active speaker, which is receiving video from endpoint A, the previous active speaker. Later when endpoint B becomes the active speaker (Figure 4), endpoints A, C and D will start to receive video from B, while endpoint B continues to receive video from endpoint C. Finally in Figure 5, endpoint A becomes the active speaker.

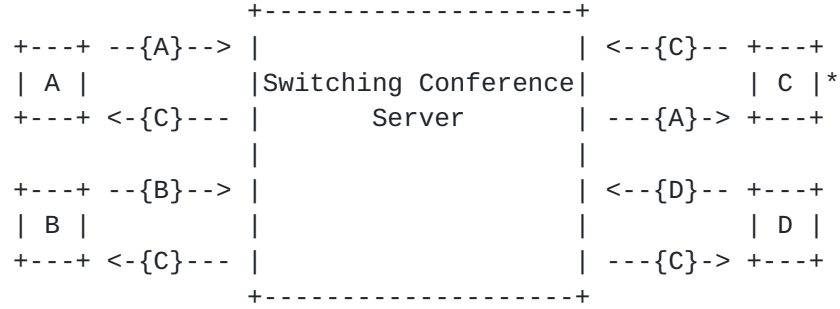

Figure 3 - Endpoint "C" is the Active Speaker

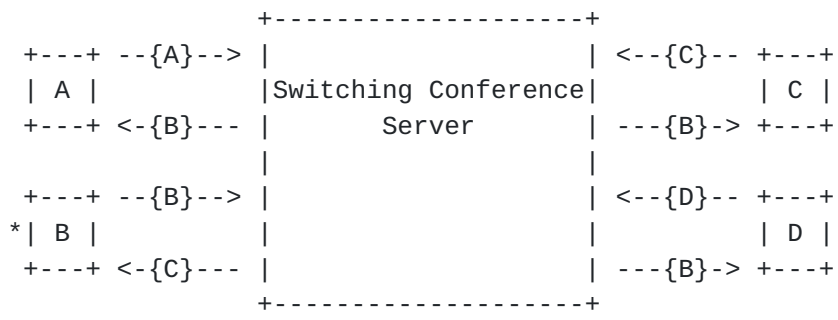

Figure 4 - Endpoint "B" is the Active Speaker

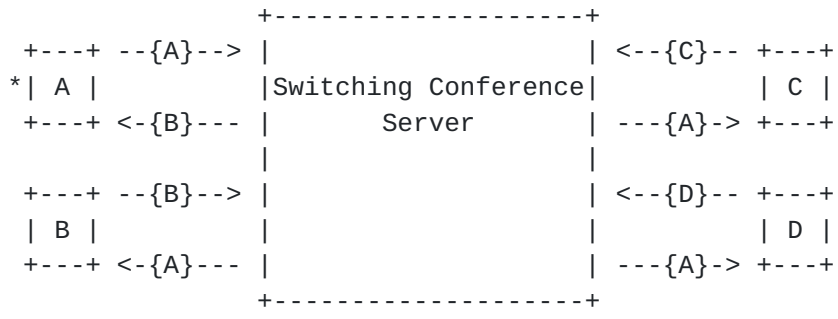

Figure 5 - Endpoint "A" is the Active Speaker

Switched conferencing can also enable conferences to scale to include many more simultaneous participants than would be possible with a traditional conferencing server. Like traditional conferencing servers, switching conference servers can also be cascaded or interconnected in a meshed topology to increase the size of the conference without putting undue burden on any particular server.

5.2. Private Media Security through Switching

A traditional conferencing server, or MCU, establishes an SRTP session with each participating endpoint separately, and needs to decrypt packets containing media presented to other endpoints. By using a switching conference server, it is possible to keep the media encryption keys private to the endpoints such that the conference server does not have access to the keys used for media encryption. The switching conference server just forwards media received to each of the other participants in the conference.

This provides for a significantly improved security model, as one can, for example, utilize conferencing resources in the cloud that do not necessarily have to be trusted. That said, there may be situations where the switching conference server needs to modify the RTP packet received from an endpoint, such as by adding or removing an RTP header extension, modifying the payload type value, etc. It

would be the responsibility of the switching conference server to ensure that media of the expected type and containing the correct information is received by a recipient.

Thus, there is a need to utilize an end-to-end encryption and authentication key (or pair of keys) and a hop-by-hop encryption and authentication key (or pair of keys). The purpose for the hop-by-hop encryption key is to optionally encrypt RTP header extensions. The current SRTP specification and related specifications do not define use of a dual-key approach presently. However, such an approach is possible and would result in ensuring the privacy of media while also enabling the more scalable switched conferencing model.

The assumptions with this model are that the endpoints are trusted entities, as they clearly have access to the media keys for encryption. Some call processing functions for the administrative domain, such as SIP [RFC3261] proxy servers or B2BUAs, are trusted in exactly the same way they are with the traditional conferencing model, meaning they must be trusted to keep signaling secure as certificate information (e.g., fingerprints) might be conveyed via signaling. The switching conference server is not fully trusted and is not given visibility into the actual contents of the SRTP payload. However, the switching conference server in the untrusted domain is at least trusted to perform its core duties of forwarding media and processing signaling; it simply isn't trusted with the media encryption keys.

The assumption is that no changes are made to SRTCP, i.e. SRTCP is protected hop-by-hop with a single security context.

This dual-key model does necessitate a change in the way that keys are managed. However, the topic of key management is outside the scope of this requirements document. However, high-level assumptions like if the end-to-end contexts use a group key as SRTP master key or if individual SRTP master keys (that may be derived/negotiated from another group key) is likely to influence the solution derived from this document.

6. Goals and Non-Goals

6.1. Goals

6.1.1. Ensure End-To-End Confidentiality

The content of the communication and all media needs to be confidential within the group of entities explicitly invited into the conference. An external monitoring adversary should not be able to deduce the human-to-human communication that actually occurred from capturing the media packets.

At the same time, it is necessary to allow switching media servers to manipulate certain RTP header fields like the payload type value.

6.1.2. Ensure End-To-End Source Authentication of Media

In a conference system with multiple participants it is vital that the multimedia content presented to any of the human participants is from the stated participant, and not an adversary that attempts to inject misleading content. Nor should an adversary be able to fool the system into becoming a trusted party in the conference. Only explicitly invited parties shall be able to contribute content.

6.1.3. Provide a More Efficient Service than "Full-Mesh"

A multi-party conference that has the goals of confidentiality and source authentication can be established as a "full mesh" (i.e., each participating endpoint directly addresses each of the other participants). However, this has a significant issue with the amount of consumed resources in both the uplink and the downlink from each participant.

A switched conferencing model would yield the efficiencies desired.

6.1.4. Support Cloud-Based Conferencing

To achieve cost-effective and scalable conferencing, it must be possible to run the conference node instances in a cloud-based virtualized environment.

From a security standpoint, this is a significant issue since the virtualized server instance and the underlying hardware and software upon which it runs might not be secure from an adversary.

6.1.5. Limiting a User's Access to Content

Since an invited user will be provided with the content protection keys, the user can decrypt content from time periods before and after the user joined the conference. However, this is not always desirable. It should be possible to re-key the content protection keys every time a user joins or leaves the conference so each particular set of conference participants uses a unique key.

This also changes the trust level required on the conference roster handling at any point and how to keep that accurate and secured.

It should be noted that timely completion of the re-keying operations become an obstacle in system design and operation. Thus, it is a goal to allow for this possibility when it is deemed essential, but it should not be a requirement on a system to re-key each time the participant list changes.

6.1.6. Compatibility with the WebRTC Security Architecture

It is a goal of this work to ensure compatibility with the WebRTC security architecture as described in [I.D-rtcweb-security-arch]. As an example, local resources that are considered a part of the trusted computing base (TCB), such as keying material derived using DTLS-SRTP, will remain within the TCB and not exposed to untrusted entities.

The browser is reliant on an external calling service to convey signaling information that may open the door for a man-in-the-middle attack, such as the conveyance of certificate fingerprints over the interface between the browser and the calling service. However, as described in [I.D-rtcweb-security-arch], the browser may utilize additional services, such as a trusted identify provider, to mitigate such risks.

6.2. Non-Goals

6.2.1. Securing the Endpoints

The security of a communication session requires that the endpoints are not compromised and that the users are trustworthy. If not, credentials and decrypted content may be shared with third parties. However, this is hard to prevent through system design. Thus, it should be assumed that the endpoint is secure and the user is trustworthy; how to achieve this is out of scope this document.

6.2.2. Concealing that Communication Occurs

A non-goal is to attempt to prevent a pervasive monitoring adversary from knowing that the communication session has occurred. The reason for excluding this as a goal is that it is extremely difficult to achieve, as a pervasive monitoring adversary can be expected to be able to have knowledge of all IP flows that enter or exit local ISPs, across links that straddle nation borders or internet exchange points. To hide the fact communication occurred, the flows required to achieve the communication session need to be highly difficult to correlate between different legs of the communication.

At this stage this is deemed too difficult to attempt and will need to be a subject for further study. Existing attempts include The Onion Router (TOR), against which it has been claimed to be possible to monitor, at least partially, by an adversary with sufficient reach.

Also of consideration is that trying to conceal the fact that communication occurred actually makes it more difficult for network administrators to effectively manage and troubleshoot issues with

conference calls.

6.2.3. Individual Media Source Authentication

Although the participants in the conference are authenticated, it is not a goal to provide source authentication of the media at the individual user level, instead being satisfied with being able to authenticate media as coming from an invited conference participant or not.

There exist solutions that can provide individual media source authentication (e.g., TESLA). However, they impact the performance or security properties they provide. Thus, further study is required to determine impact and resulting security properties if desired to have individual source authentication.

6.2.4. Support for Multicast in Switched Conferencing

Multicast traffic is, by design, transmitted to every participant in a conference. The focus of this document is only on centralized unicast conferencing that utilizes a switched conferencing architecture.

7. Requirements

The following are the security solution requirements for switched conferencing that enable end-to-end media privacy between all conference participants.

Note that while some switching media servers might be fully trusted entities, the intent of this solution and purpose for these private media (PM) requirements is to address those servers that are not fully trusted.

PM-01: Switching conference server MUST be able to switch the media between participants in a conference without having access to the media encryption keys.

PM-02: Solution MUST maintain all current SRTP security goals, namely the ability to provide for confidentiality, provide replay protection, and ensure message integrity.

PM-03: Solution MUST extend replay attacks protection to cover each hop in the media path. It MUST be possible to detect if a packet received by either an endpoint or a switching conference server was previously received by that entity or if the packet is not intended for that entity.

PM-04: Keys used for end-to-end encryption and authentication of RTP payloads and other information deemed unsuitable for access by the switching conference server MUST NOT be generated by

or accessible to any component that is not in the fully trusted domain.

PM-05: The switching conference server MUST be capable of making changes to the RTP header and, optionally, the RTP header extensions.

PM-06: The switching conference server, or any entity that is not fully trusted, MUST NOT be involved in the authentication of identities for the purpose of media key distribution.

PM-07: The switching conference server MUST be able to switch an already active SRTP stream to a new receiver, while guaranteeing the timely synchronization between the SRTP context of the transmitter and its current and new receivers.

PM-08: It MUST be possible for the switching conference server to determine if a received media packet was transmitted by a valid conference participant.

PM-09: It MUST be possible for a conference to be optionally re-keyed as desired, such as each time a participant joins or leaves the conference.

PM-10: To decrypt packets, the receiving endpoint needs to be able to know the SSRC and RTP sequence number used by the sending endpoint. These values need to be integrity protected end-to-end, either explicitly by inclusion in an end-to-end MAC or implicitly like the MKI field in [RFC3711].

8. IANA Considerations

There are no IANA considerations for this document.

9. Security Considerations

[TBD]

10. References

10.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3711] Baugher, M., McGrew, D., Naslund, M., Carrara, E., and K. Norrman, "The Secure Real-time Transport Protocol (SRTP)", RFC 3711, March 2004.

[RFC6464] Lennox, J., Ivov, E., and E. Marocco, "A Real-time Transport Protocol (RTP) Header Extension for Client-to-Mixer Audio Level Indication", RFC 6464, December 2011.

[I.D-rtcweb-security-arch]
E. Rescorla, "WebRTC Security Architecture", Work in Progress, July 2014.

[RFC6904] J. Lennox, "Encryption of Header Extensions in the Secure Real-time Transport Protocol (SRTP)", RFC 6904, December 2013.

10.2. Informative References

[RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.

11. Acknowledgments

The authors would like to thank Marcello Caramma, Matthew Miller, Christian Oien, Magnus Westerlund, Cullen Jennings, Christer Holmberg, and Bo Burman for their invaluable input.

12. Contributors

[TBD]

Authors' Addresses

Paul E. Jones
Cisco Systems, Inc.
7025 Kit Creek Rd.
Research Triangle Park, NC 27709
USA

Phone: +1 919 476 2048
Email: paulej@packetizer.com

Nermeen Ismail
Cisco Systems, Inc.
170 W Tasman Dr.
San Jose
USA

Email: nermeen@cisco.com

David Benham
Cisco Systems, Inc.
170 W Tasman Dr.
San Jose
USA

Email: dbenham@cisco.com

Nathan Buckles
Cisco Systems, Inc.
170 W Tasman Dr.
San Jose
USA

Email: nbuckles@cisco.com

John Mattsson
Ericsson AB
SE-164 80 Stockholm
Sweden

Phone: +46 10 71 43 501
Email: john.mattsson@ericsson.com

Yi Cheng

Ericsson
SE-164 80 Stockholm

Sweden

Phone: +46 10 71 17 589
Email: yi.cheng@ericsson.com

Richard Barnes
Mozilla
331 E Evelyn Ave.
Mountain View
USA

Email: rlb@ipv.sx