September 2020

MEDIA OPTIMIZATIONS FOR CLOUD ENDPOINTS USING A LOCAL GATEWAY (LGW)

Naveen Rajasekaran

Krishna Kumar Rai

Prasadaraju Yarakaraju

MEDIA OPTIMIZATIONS FOR CLOUD ENDPOINTS USING A LOCAL GATEWAY (LGW)

AUTHORS:
Naveen Rajasekaran
Krishna Kumar Rai
Prasadaraju Yarakaraju

ABSTRACT

Presented herein are techniques that provide efficient media anchoring in hosted cloud collaboration deployments. A media path optimization may be achieved by utilizing a local gateway (LGW) that intelligently detects media loops and avoids sending such media to the cloud. The media path optimization may provide a bandwidth savings and may improve user experience by providing a better call quality.

DETAILED DESCRIPTION

Figure 1, below, illustrates a typical deployment that may be utilized to facilitate cloud calling. Figure 1 illustrates a single site with a local gateway (LGW) and a co-located public switched telephone network (PSTN) gateway (GW) and Session Border Controller (SBC).

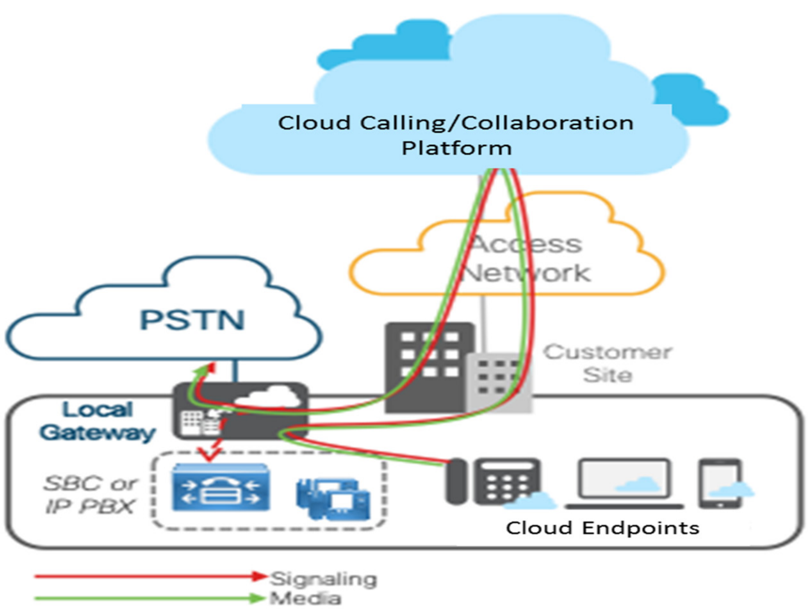

Figure 1: Single Site with a LGW and Co-located PSTN GW/SBC

The deployment as shown in Figure 1 may provide various features, such as enabling a bring your own PSTN (BYoPSTN) option for videoconference calling, providing connectivity to a customer-subscribed PSTN service, and/or providing connectivity to an on-premise Internet Protocol (IP) private branch X-change (PBX) or dedicated SBC/PSTN GW.

As illustrated in Figure 1, the local gateway may be used as a fusion element between the cloud calling/collaboration platform and on-premise calling such that the local gateway may act as a bridge for calls between cloud registered endpoints and on-premise registered endpoints or the PSTN. A primary function of the local gateway is to act as a border controller that terminates and re-originates the media/signaling between cloud and on-premise/PSTN calls and also provides interworking functionalities.

Currently, for a call originating from a cloud registered endpoint to the PSTN as shown in Figure 1, media (e.g., Real-time Transport Protocol (RTP) media) that originates on-premise is communicated all the way to the cloud SBC through the local gateway and then comes back to the cloud registered endpoint. This is a not an optimized path. Apart from unnecessary bandwidth consumption, such a non-optimized communication path could also cause media quality issues due to latency/jitter as the media has to traverse through internet.

Provided herein is an optimization technique that can be realized via a local gateway (LGW) in order to facilitate media path optimizations without putting constraints on other system elements, such as endpoints, PBX elements, the cloud SBC, PSTN elements, etc. For the optimization technique, the LGW intelligently detects media loops and avoids sending media to cloud while still anchoring media internally.

Different approaches may be utilized to facilitate detection of associated calls that may be causing a media loop in order to apply media path optimizations on the LGW. For example, various detection approaches may include Session Initiation Protocol (SIP) Session Identifier (Session-ID) based detection, detection using calling and called numbers and/or Uniform Resource Identifiers (URIs), or detection using initial media/RTP packet headers.

For the SIP Session-ID based detection approach, consider that the SIP Session-ID header is carried end-to-end by SIP entities unlike SIP Call-ID that gets changed by back-to-back user agents (B2BUAs) and/or SBCs. For detecting a media loop utilizing this approach, the LGW can use the Session-ID to identify associated calls that are traversing through the cloud and coming back to the LGW. This approach requires SIP entities, specifically, the LGW and the cloud SBC, to support the Session-ID header as described in Internet Engineering Task Force (IETF) Request For Comments (RFC) 7329.

For detecting a media loop using calling and called numbers/URIs, the LGW can match the calling and called numbers/URIs to identify the associated calls that are traversing through the cloud and coming back to the LGW. This approach may not work if number translations are applied to calling or called numbers/URIs in the cloud.

For detection using initial media/RTP packet headers, consider that since the cloud SBC may be relaying the media packets without changing RTP headers fields (e.g., Synchronization Source (SSRC), sequence numbers, and timestamps) the LGW can use these fields to detect redundant and/or non-optimized media paths.

For example, the LGW can globally store the SSRCs, sequence numbers, and timestamps of the first packet sent to the cloud SBC for every call, with the SSRC being a key to perform a look-up for packets received from the cloud SBC. Upon receiving a first packet from the cloud SBC for any call, detection of a media loop can be performed by determining whether the packet satisfies the following conditions:

1. The SSRC matches to one of the stored SSRCs;
2. An RTP sequence number difference between the stored data and that of the received packet is within a window of 'X' packets; and
3. An RTP timestamp difference between the stored data and that of the received packet is within a window of time 'T' milliseconds.

For the conditions above, the parameters 'X' and 'T' can be defined based on implementation (e.g., to small values). The windows for the conditions may be utilized instead of an exact match of sequence numbers and timestamps between the stored data and the received packet data because an initial few number of packets could be dropped in the network due to multiple reasons. Additionally, it may not be possible to rely on an SSRC match alone because SSRC collision is possible. Thus, using RTP sequence numbers and timestamps may reduce the probability of collision.

Once associated calls are detected/identified, the LGW may perform the following optimizations:

1. Trigger a hold locally towards the cloud SBC for both calls so that the cloud calling/collaboration platform does not expect media; and
2. Internally bridge the media for the two calls:
   a. Internet Protocol (IP) phone to the cloud calling/collaboration platform, and
   b. Cloud calling/collaboration platform to the PSTN.

Consider various example call flows shown below via Figures 2A and 2B in which Figure 2A illustrates a call flow involving a cloud endpoint calling a PSTN using a LGW without the optimization techniques of this proposal and Figure 2B illustrates a call flow with an optimized media flow using a LGW in accordance with the optimization techniques of this proposal.

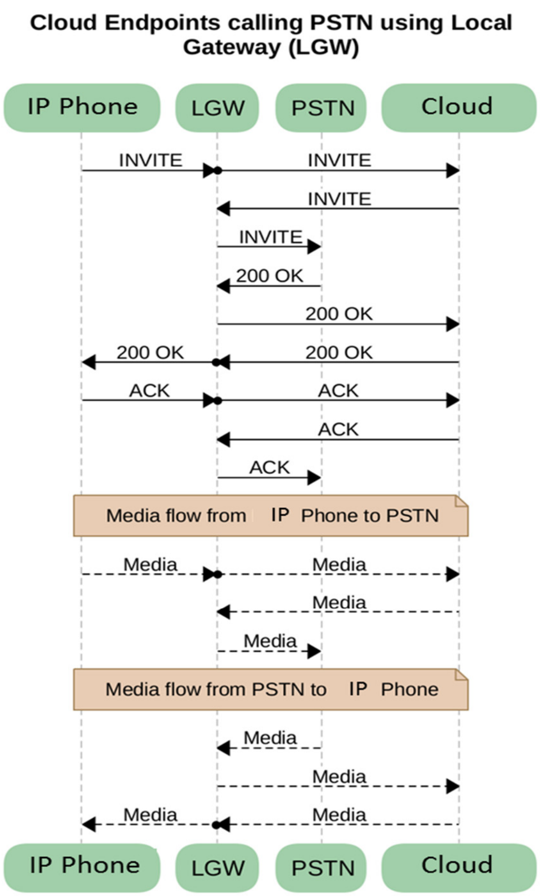

Figure 2A: Without Optimization

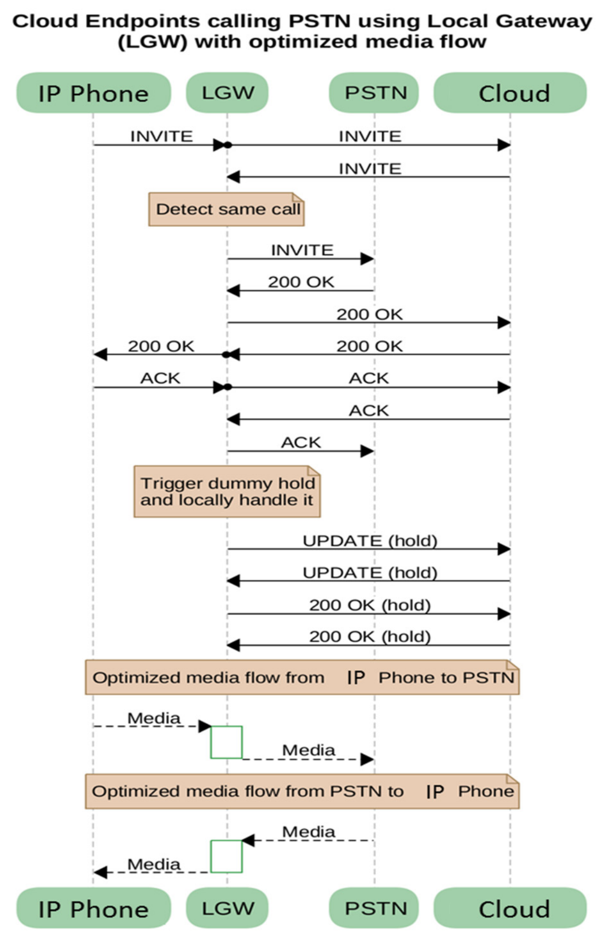

Figure 2B: With Optimization

Additionally, consider the system of Figure 1 in which the optimization can be provided via the LGW to facilitate an optimized media flow, as shown below in Figure 3.

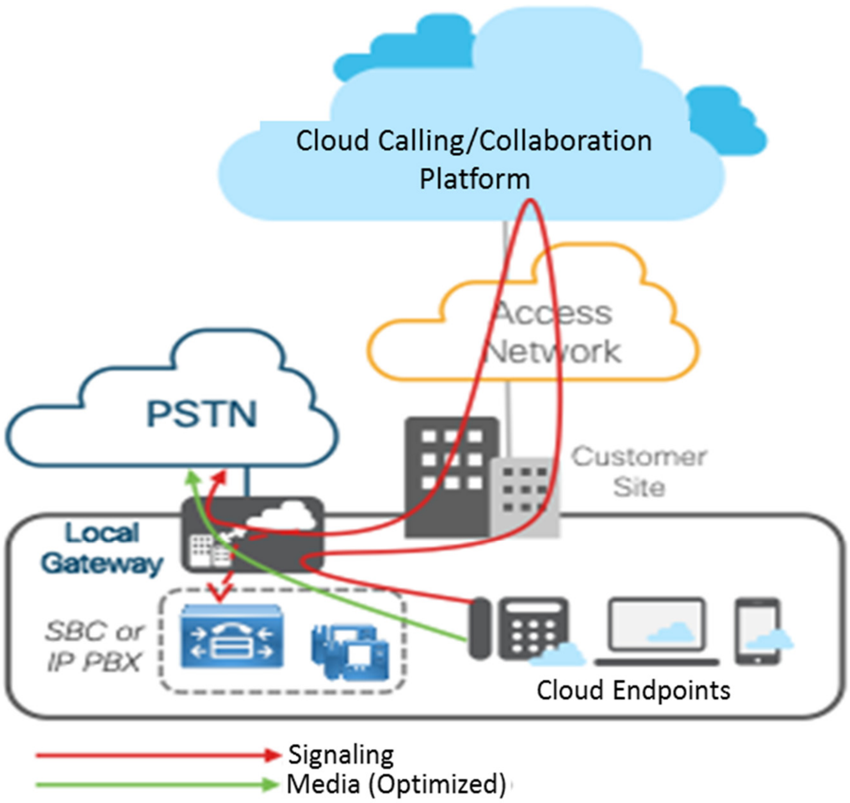

Figure 3: Single Site with LGW Providing an Optimized Media Flow

In summary, techniques herein may provide efficient media anchoring in hosted cloud collaboration deployments. A media path optimization may be achieved by utilizing a LGW that intelligently detects media loops and avoids sending such media to the cloud. The media path optimization may provide a bandwidth savings and may improve user experience by providing a better call quality.