March 2021

COLLABORATION SERVICE TO FACILITATE DYNAMIC CAPTIONING IN REAL-TIME

Kaouther Abrougui

Sandhya Kerehalli

Jeff Apcar

Geetha Anandakrishnan

COLLABORATION SERVICE TO FACILITATE DYNAMIC CAPTIONING IN REAL-TIME

AUTHORS:
Kaouther Abrougui
Sandhya Kerehalli
Jeff Apcar
Geetha Anandakrishnan

ABSTRACT

Network issues can affect the quality of audio and video for online collaboration sessions, which can result in participants missing important parts of speech. Current collaboration tools often provide closed captioning (CC) options, however, captioning is typically not automatically started on bandwidth issues and, further, captioning is often generated in the cloud with no guarantee that the original stream is received intact by meeting participants. Presented herein are techniques to address these issues by providing a collaboration service that facilitates real-time Dynamic Captioning. In the event of network issues during an online meeting, at either the speaker end or the listener end, Dynamic Captioning is activated in accordance with techniques presented herein so that a conversation can still be followed in the event that the video and audio has been badly affected by the network conditions.

DETAILED DESCRIPTION

Have you ever experienced audio dropouts during a video conferencing session? Such a video conferencing session could be a work meeting to share information or to discuss important ideas. In another example, videoconferencing could be provided for live training or an online class, where a consistently high quality flow is important to the clarity of information being communicated with customers, trainees, and/or students.

As a listener, in the event of a disruption, it is frustrating not to be able to decipher speech or lose important parts of a conversation. In many situations, a meeting participant may choose to ask the speaker to repeat information. However, if all the participants start asking (or speaking over themselves), it can entirely disrupt the flow of a meeting. As a speaker there is a good chance that the speaker will not be aware that some of the audience

are having difficulties following up with a conversation due to network issues. Critical points may be missed and/or listeners may get frustrated and lose interest if network issues persist. The result is a poor experience for all participants possibly negating the fulfillment of a meeting.

This proposal provides for the ability to solve the problem of intermittent voice content loss for online collaboration tools in real-time via a dynamic captioning mechanism. Collaboration tools often provide closed captioning (CC) options, however, current captioning options are typically not automatically started on bandwidth issues and, further, captioning for current captioning options often involves captioning that is generated in the cloud with no guarantee that the original stream is received intact by meeting participants.

Presented herein are techniques to address these issues by providing a real-time Dynamic Captioning (DC) collaboration service. In particular, techniques herein provide a solution to current captioning limitations by recording each speaker and generating captioning for each speaker locally (e.g., on a user's device or a local server if the user is in a branch location). This guarantees a higher probability that audio will be recorded and captioned as originated from a local speaker. Next, bandwidth issues can be detected for all participants and transcripts can be published to the impacted participants. Whenever a participant is experiencing network bandwidth issues, the DC feature can be automatically activated and transcripts from a previous number of minutes (e.g., the previous 2 minutes) can be displayed on the affected participant's device to ensure continuity. The captioning can continue until the network issue is resolved.

Figure 1, below, illustrates example details of a system in which the DC feature can be provided.

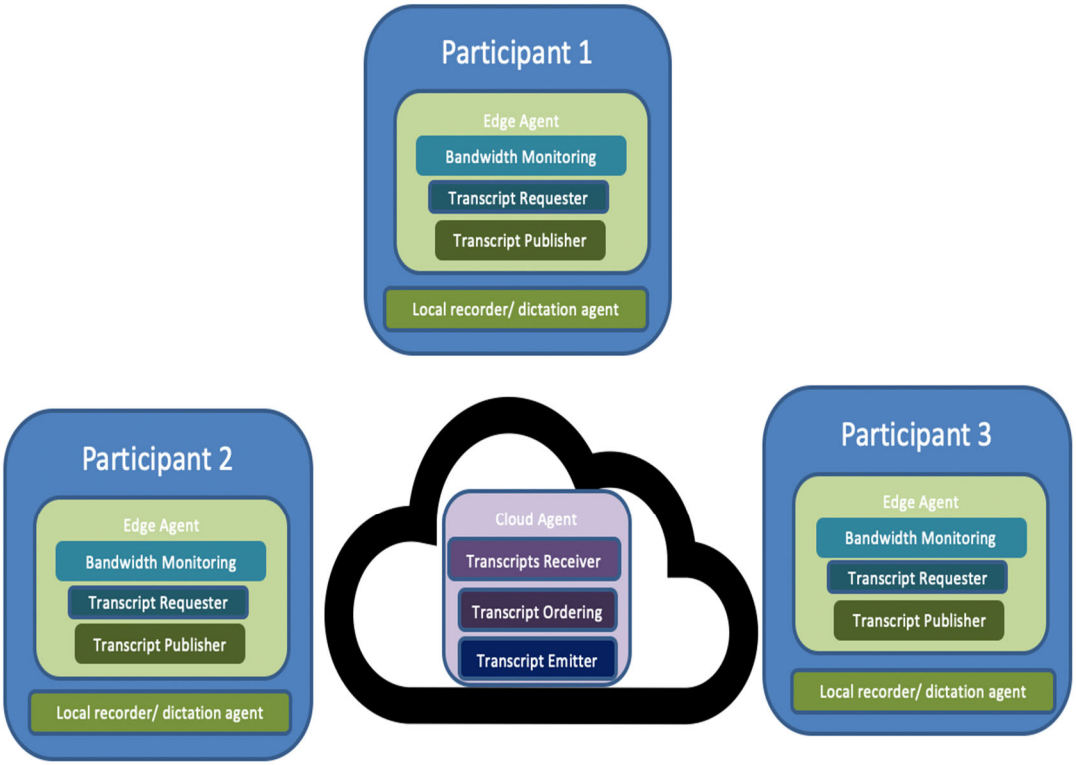

Figure 1: Example Dynamic Captioning System

For the system as illustrated in Figure 1, the participants can include all members participating a video conference. Each participant has two different roles: speaker and listener. An edge agent for each participant is the mediator between a participant and a cloud agent.

An edge agent can have different roles. For example, an edge agent can initiate recording using a local participant device and can generate transcripts from the recording. An edge agent can also monitor network bandwidth issues using known mechanisms such as the Real-Time Transport Control (RTCP) feedback protocol. An edge agent can also send transcripts to the cloud agent and can push commands to the cloud agent in order to push transcripts to all other participants if the current participant is a speaker and experiences bandwidth issues. An edge agent can also pull transcripts from the cloud agent if the current participant is a listener and experiences bandwidth issues. Thus, an edge agent may facilitate displaying transcripts dynamically on impacted participant devices. If

the speaker is impacted by bandwidth issues, then all listeners will receive a Dynamic Captioning transcript in real-time.

The cloud agent communicates between other cloud agents and also with edge agents. The cloud agent also provides global time synchronization between edge agents, collects all transcripts locally generated at an edge agent, and orders transcripts based on their timestamps. The cloud agent can be integrated with CC functionality of a collaboration tool in order to send all the transcripts from other participants to a current participant but with enhanced CC. The cloud agent can also push transcripts to participants if a push command is received from an edge agent with an impacted participant and can send transcripts on demand to edge agents with impacted participants.

Consider an example workflow, involving the system of Figure 1, as follows. At startup, the cloud agent synchronizes the clock of each edge agent to ensure that received transcripts can be reordered. Each edge agent locally records the speech of each participant, generates time stamped transcripts, and transmits the time stamped transcripts to the cloud agent. The cloud agent orders the transcripts based on the timestamps. Each edge agent also monitors the bandwidth of each corresponding participant.

For potential listener issues, if the edge agent at a listener participant detects bandwidth issues, the edge agent can request transcripts from the cloud agent for a previous period of time from which the bandwidth issues were detected. For example, the cloud agent can request transcripts for the previous two minute period from which the issues were detected. The time period can be configured to include any number of minutes, seconds, etc. The cloud agent can send the transcripts to the requesting edge agents and keep streaming any new transcripts until a stop command is received from the edge agent. The edge agent turns on transcripts dynamically on the participant's screen. Once the bandwidth issue is resolved for the participant, transcripts are turned off.

Figures 2A and 2B illustrate example details for an example listener issue scenario. For example, Figure 2A illustrates a 'Before' diagram in which Listener 1 is experiencing some corruption of video, possible caused by oversubscribing on the download capacity. In contrast, Figure 2B illustrates an 'After' diagram in which Dynamic Captioning is automatically activated on Lister 1's screen.

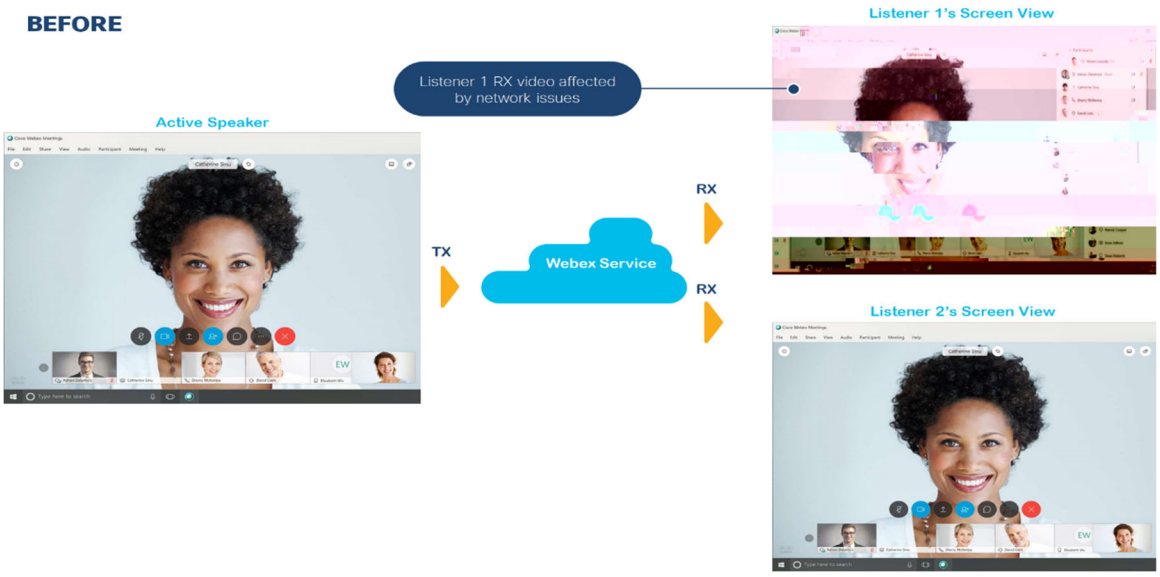

Figure 2A: Listener Issue Scenario before Dynamic Captioning is Activated

Figure 2B: Listener Issue Scenario after Dynamic Captioning is Activated

For potential speaker issues, if the edge agent at the speaker participant detects a bandwidth issue then it attempts recovery by performing various operations, including turning off the video at speaker to save bandwidth and send only CC, providing suggestions regarding what needs to be turned off in the speaker's system to improve user experience, and notifying the cloud agent. The impacted speaker edge agent streams the transcripts to the cloud agent, and the cloud agent streams the transcripts to the other edge agents (listeners). The edge agents of the impacted participants will be notified and Dynamic Captioning will be activated on all the listener's participant devices. If the impacted speaker agent detects that the bandwidth issue is resolved, then it notifies the cloud agent to abort the streaming of transcripts to other participants.

Figures 3A and 3B illustrate example details for an example speaker issue scenario. For example, Figure 3A illustrates a 'Before' diagram in which the speaker is experiencing significant bandwidth issues, which manifests itself on all listener screens. Figure 3B illustrates an 'After' diagram in which the edge agent detects this issue and sends a command to the cloud agent to automatically activate Dynamic Captioning on all listener screens. The speaker participant is also advised that their video has been shut off and Dynamic Captioning is active.

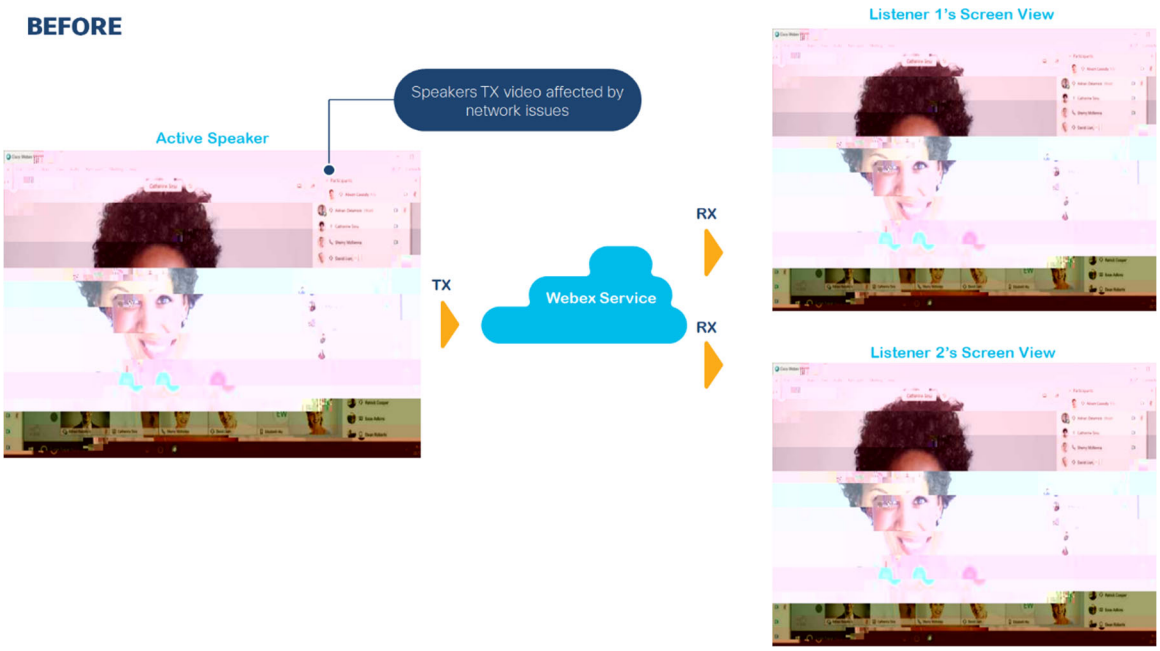

Figure 3A: Speaker Issue Scenario before Dynamic Captioning is Activated

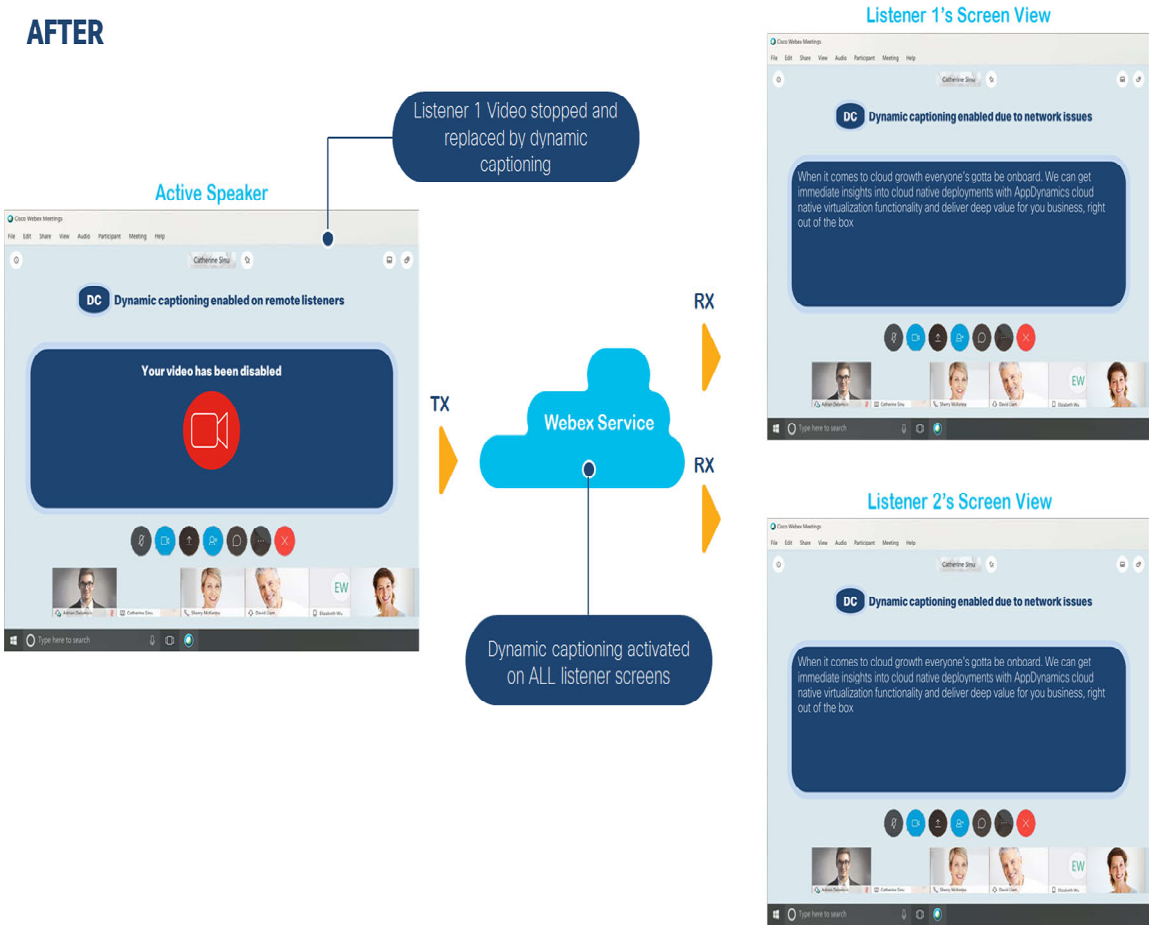

Figure 3A: Speaker Issue Scenario after Dynamic Captioning is Activated

For bandwidth monitoring, there are existing tools, such as RTCP feedback protocol that can be used to facilitate monitoring. Accordingly, techniques herein can be integrated with one or more known bandwidth monitoring solutions in order to facilitate bandwidth monitoring. Figure 3, below, illustrates an example flow associated with bandwidth monitoring that can be utilized to trigger the Dynamic Captioning techniques described herein.

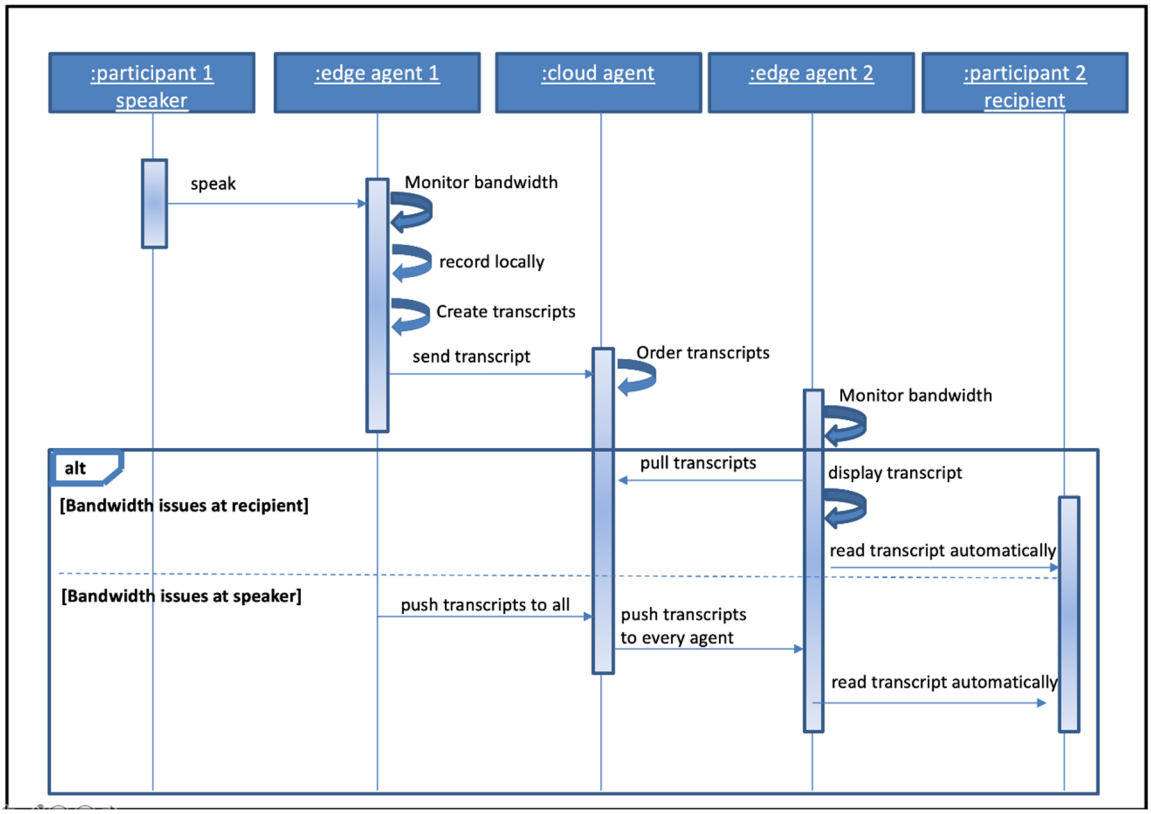

Figure 3: Example Bandwidth Monitoring to Trigger Dynamic Captioning

In summary, techniques are presented herein that provide real-time Dynamic Captioning features for a collaboration system. In the event of network issues during an online meeting, at either the speaker end or the listener end, Dynamic Captioning is activated in accordance with techniques presented herein so that a conversation can still be followed in the event that the video and audio has been badly affected by the network conditions. The captioning can be performed at an edge (local) device and then transmitted to the cloud for distribution in the event of a network condition. Thus, a correct copy of the original speech in text form is available as it is recorded and processed locally.