



- (51) International Patent Classification:
H04L 7/00 (2006.01) *H04N 7/14* (2006.01)
H04N 7/00 (2011.01) *H04N 7/15* (2006.01)
- (21) International Application Number: PCT/IB2017/050051
- (22) International Filing Date: 06 January 2017 (06.01.2017)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 16306515.4 18 November 2016 (18.11.2016) EP
- (71) Applicant: ALE INTERNATIONAL [FR/FR]; 32 avenue Kleber, 92700 Colombes (FR).
- (71) Applicant (for ZW only): ALE USA INC. [US/US]; 26801 Agoura Road, Calabasas, CA 91301 (US).
- (72) Inventors: FADILI, Moulay; 7 rue des 3 Cèdres, 95000 CERGY (FR). TRINITE, Benoît; 7 rue Henri Sellier, 78360 Montesson (FR). HELBERT, Emmanuel; 71 rue du Maréchal Joffre, 67140 Andlau (FR).
- (74) Agent: THERIAS, Philippe; PHIP Avocat, 75, rue de Tocqueville, c/o Bignon Lebray, 75017 Paris (FR).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(54) Title: VIDEOCONFERENCE SERVER, CLIENT DEVICE, VIDEOCONFERENCE SYSTEM AND METHOD

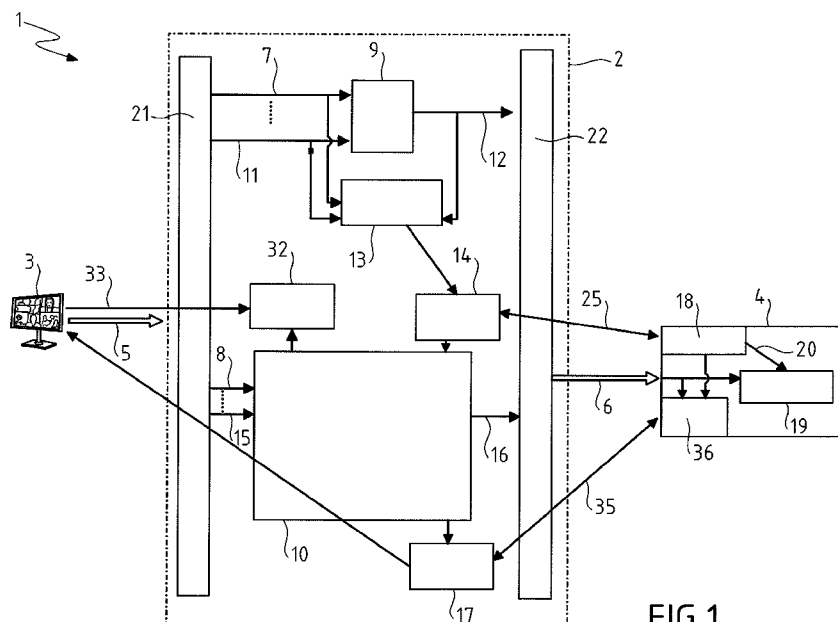


FIG.1

(57) Abstract: The invention comprises a videoconference server (2) comprising : - an audio processing module (9) configured to, o calculating an average audio processing delay of processing audio packets intended to the second client device, in order that the second client device may synchronize a received video packet with the associated received audio packet, - a video router (10) configured to: o duplicate the video packet and temporarily store the duplicated video packet, o receive a retrieving request from the second client device (4) for retrieving a lost video packet, o retrieving the duplicated video packet associated with the identification information in the video buffer (17).



(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

VIDEOCONFERENCE SERVER, CLIENT DEVICE, VIDEOCONFERENCE SYSTEM AND METHOD

Field of the invention

5 The invention relates to synchronization of audio and video signals in a videoconference system. In particular, the invention relates to a videoconference server, a client device and a method for synchronizing audio packets and video packets, especially for lip synchronization.

10 Background

 In Real Time Protocol (RTP) communications, videoconference servers receive audio packets and associated video packets incoming from a client device in order to route the audio and video packets to another client device. A videoconference server processes the received audio packets, for example by compressing a received
15 audio packet and decompressing received compressed packets or by mixing several audio streams received from several client devices. Therefore, a processing delay is added to the transmission delay of the audio packets. In order to synchronize associated video packets to the audio packets, the processing delay is also applied to the video packets by the videoconference server in order to synchronize the video
20 packets to the audio packets prior to sending the audio and video packets to a client device.

 In the case where a video packet is lost in the transmission between an emitting client device and a receiving client device, one of the widely used techniques in real time video conferencing is the packet retransmission mechanism.
25 In this mechanism a recovery request is sent from the receiving client device in order that the emitting client device re-sends the lost packet.

 Therefore, the transmission delay for the lost video packet can exceed the maximum delay of lip synchronization in the receiving client device, which leads to the inefficiency of the retransmission mechanism.

Summary

An idea at the basis of the invention is to improve the recovery of a lost video packet in a video communication session. An idea at the basis of the invention is to add flexibility in the transmission delay of video packets. An idea at the basis of the invention is to carry out the synchronization between an audio packet and a video packet in the client devices rather than in the videoconference server.

In an embodiment, the invention provides a videoconference server comprising

- 10 - a packet input interface configured to receive audio and associated video packets from a first client device,
- a packet output interface configured to send the received audio and associated video packets to a second client device,
- an audio processing module configured to, when the packet input interface receives an audio packet from the first client device,
 - 15 o processing the audio packet,
 - o calculating an average audio processing delay of audio packets intended to the second client device,
 - o associating the average audio processing delay to the second client device,
 - 20 o sending the average audio processing delay to the second client device in order that the second client device may synchronize a received video packet with the associated received audio packet,
- a video router comprising a video buffer configured to duplicate a video packet received by the packet input interface from the first client device,
25 wherein the video packet is associated to the audio packet, and temporarily store the duplicated video packet in relation with an identification information of the video packet,

the video router being further configured to:

- receive a retrieving request from the second client device for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet,
- retrieving the duplicated video packet associated with the identification information in the video buffer,

the packet output interface being further configured to send the retrieved duplicated video packet to the second client device.

Thanks to these features, the retransmission delay in the case where a packet is lost is improved. For example, thanks to these features, the lip synchronization is improved in a video conference.

According to embodiments, such a videoconference server may comprise one or more of the features below.

In embodiments, the video router is further configured to select a subset of the received video packets, and to send to the second client device only video packets of the selected subset of received video packets.

In embodiments, the video buffer is configured to duplicate all the video packets received in the packet input interface. Alternatively, the video buffer is configured to duplicate only video packets of the selected subset of received video packets.

In embodiments, the packet input interface is further configured to receive audio and associated video packets from the second client device, and the packet output interface is further configured to receive the received audio and associated video packets from the second client device to the first client device.

In embodiments, the video router is further configured to detect that a video packet initially sent by the first client device is not received by the packet input interface,

the video router being further configured to send a retrieving request to the first client device, and to receive from the first client device the initially sent video packet that has not been initially received.

In embodiments, the videoconference server further comprises:

- a client device configuration detection module configured to, prior the initiation of a communication session with the second client device, send a configuration request to the second client device,

5 - receiving a configuration response from the second client device, the configuration response comprising an indication whether the second client device is adapted to locally synchronize video packets with associated audio packets or not adapted,

10 the videoconference server being further configured to, when the configuration response comprises an indication that the second client device is not adapted to locally synchronize video packets with associated audio packets,

- delaying the received video packet by the calculated average audio processing delay prior to sending the received video packet to the second client device.

15 In embodiments, the videoconference server is further configured to, when the configuration response comprises an indication that the second client device is not adapted to locally synchronize video packets with associated audio packets,

- inhibiting the sending of the calculated average audio processing delay to the second client device.

20 In embodiments, the audio packets are identified by audio packet identifiers and the associated video packets are identified by video packet identifiers, the video buffer being configured to detect that a video packet initially sent by the first client device is not received by the packet input interface by controlling the video packet identifier of a received video packet and by comparing the video
25 packet identifier to an expected video packet identifier, wherein the expected video

packet identifier is associated with the audio packet identifier of a received audio packet to which the expected video packet is associated.

In embodiments, the audio and video packets are each identified by a packet number or sequence number, the packet number indicating the position of
5 the packet in a sequence of emitted packets, wherein the retrieving request comprises the packet number or sequence number of the video packet.

In embodiments, processing the audio packet comprises mixing the audio packet with other received audio packets.

In embodiments, the audio processing module is configured to periodically
10 recalculate the average audio processing delay and to send the recalculated average audio processing delay to the second client device in order that the second client device may update the average audio processing delay used to synchronize a received video packet with the associated received audio packet.

In embodiments, the video buffer is a circular buffer.

15 In embodiments, the packets are Real Time Protocol (RTP) packets.

The invention also provides a videoconference endpoint comprising

- an emitting interface configured to emit audio and associated video packets for a videoconference server,
- a receiving interface configured to receive audio and associated
20 video packets from the videoconference server,
- a video buffer configured to temporarily store the received video packets,
- a synchronization module configured to receive an average audio processing delay from the videoconference server, and to delay the received video

packets by the average audio processing delay in order to synchronize each video packet with a received audio packet associated to the video packet,

- a packet loss detector configured to detect that a video packet initially sent by the videoconference server is not received by the receiving interface,

5 the packet loss detector being further configured to send a retrieving request to the videoconference server for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet,

the receiving interface being further configured to receive a duplication of the lost video packet in response to sending the retrieving request.

10 In embodiments, the packet loss detector is further configured to inhibit the delaying of the duplication of the video packet by the average audio processing delay.

In embodiments, the videoconference endpoint further comprises an audio speaker configured to render a voice signal carrier by the audio packets and a video
15 display configured to display a video signal carrier by the received video packets, the video display being configured to display a portion of the video signal carried by a video packet associated with an audio packet at the same time that the audio speaker renders a portion of the voice signal carried by the associated audio packet. For example, the videoconference endpoint is a client device, such as a personal
20 computer device, a digital screen device, a smartphone device etc.

The invention also provides a videoconference system comprising:

- a videoconference server comprising:
 - o a packet input interface configured to receive audio and associated video packets from a first client device,
 - 25 o a packet output interface configured to send the received audio and associated video packets to a second client device,

- 5
- an audio processing module configured to, when the packet input interface receives an audio packet from a first client device,
 - processing the audio packet,
 - calculating an average audio processing delay of audio packets intended to the second client device,
 - associating the average audio processing delay to the second client device,
 - sending the average audio processing delay to the second client device in order that the second client device may
- 10
- a video router comprising a video buffer configured to, duplicate a video packet received by the packet input interface from a first client device, wherein the video packet is associated to the audio packet, and temporarily store the duplicated video packet in relation with an
- 15
- identification information of the video packet,

the video router being further configured to:

- receive a retrieving request from the second client device for retrieving a lost video packet, the retrieving request
- 20
- comprising an identification information of the video packet,
 - retrieving the duplicated video packet associated with the identification information in the video buffer,

the packet output interface being further configured to send the retrieved duplicated video packet to the second client device,

25 the videoconference system further comprising,

- the second client device comprising

- an emitting interface configured to emit audio and associated video packets for the videoconference server,
- a receiving interface configured to receive audio and associated video packets from the videoconference server,
- 5 ○ a video buffer configured to temporarily store the received video packets,
- a synchronization module configured to receive an average audio processing delay from the videoconference server, and to delay the received video packets by the average audio processing delay in
- 10 order to synchronize each video packet with a received audio packet associated to the video packet,
- a packet loss detector configured to detect that a video packet initially sent by the videoconference server is not received by the receiving interface,

the packet loss detector being further configured to send a retrieving request to the videoconference server for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet,

the receiving interface being further configured to receive a duplication of the lost video packet in response to sending the retrieving request.

In an embodiment, the video router is further configured to

- 20 ▪ forwarding the retrieving request to the first client in case the video packet is not available in the video buffer,
- dropping the retrieving request in case the lost video packet had already been detected by the video router, this detection having triggered the sending of a retrieving request by the
- 25 video router towards the first client,

In an embodiment, the packet loss detector is further configured to inhibit the delaying of the duplication of the video packet by the average processing audio delay.

The invention also provides a method for synchronizing audio and video packets from a first client device to a second client device, wherein the first client device and the second client device are involved in a video communication session, the method comprising:

- receiving, in a videoconference server, audio packets and associated video packets from a first client device to be forwarded to a second client device,
- 10 - processing, in the videoconference server, the received audio packets and calculating an average audio processing delay of audio packets intended to the second client device,
- sending the processed audio packets to the second client device,
- duplicating and temporarily storing, in the videoconference server, at least
15 some of the received video packets associated with the received audio packets,
- sending by the videoconference server the received video packets to the second client device,
- receiving, in the second client device, the processed audio packets and the
20 video packets,
- sending, by the videoconference server, the calculated average audio processing delay to the second client device,
- delaying, by the second client device, the received video packets by the average audio processing delay.

25 In embodiments, the method further comprises retrieving a video packet that has not been received by the second client device, wherein retrieving comprises:

- detecting, by the second client device, that a video packet sent by the videoconference server is not received,
- sending, by the second client device, a retrieving request for retrieving the video packet that has not been received, the retrieving request comprising
5 an identification information of the video packet that has not been received,
- receiving, by the videoconference server, the retrieving request from the second client device,
- retrieving, by the videoconference server, the duplicated video packet associated with the identification information,
- 10 - sending the retrieved duplicated video packet to the second client device,
- receiving, by the second client device the retrieved duplicated video packet.

In embodiments, the method further comprises inhibiting, by the second client device, the delaying of the retrieved duplicated video packet by the average audio processing delay.

15 In embodiments, the method further comprises retrieving a video packet that has not been received by the videoconference server, wherein retrieving comprises:

- detecting, by the videoconference server, that a video packet sent by the first client device is not received by the videoconference server,
- sending, by the videoconference server, a retrieving request to the first client
20 device, wherein the retrieving request comprises an identification information of the video packet that has not been received,
- receiving, by the videoconference server, the video packet that has not been received.

25 In embodiments, the method further comprises inhibiting, by the second client device, the delaying of the retrieved video packet by the average audio processing delay.

Brief description of the drawings

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiments described hereinafter, by way of example, with reference to the drawings.

- 5 - Figure 1 schematically illustrates a videoconference system comprising a videoconference server, a first client device and a second client device, each involved in a videocommunication session;
- Figure 2 schematically illustrates the transmission of a video packet between the first client device, the videoconference server and the second client device of
10 Figure 1, wherein the video packet is lost between the videoconference server and the second client device;
- Figure 3 schematically illustrates the transmission of a video packet between the first client device, the videoconference server and the second client device of the
15 Figure 1, wherein the video packet is lost between the first client device and the videoconference server;
- Figure 4 schematically illustrates the second client device of Figure 1.

Detailed description of the embodiments

Figure 1 illustrates a videoconference system 1 comprising a
20 videoconference server 2, a first client device 3 and a second client device 4.

For the sake of illustration, the videoconference system 1 represented on Figure 1 only comprises two client devices 3 and 4, but the videoconference system 1 may as well comprise any number of client devices involved in the videoconference session without departing from the scope of the invention.

25 A first user uses the first client device 3 and a second user uses the second client device 4. The client devices may be in two different remote places or in the same place.

Each client device involved in the videoconference session comprises a microphone configured to capture the voice of the user of the client device and an
30 audio speaker configured to render the voice of users of the remote client devices.

At least two client devices involved in the videoconference session further comprise a camera configured to capture a video of the user of the client device and a video display configured to display a video from the user(s) of the remote client device(s).

Each client device exchanges audio and video streams of captured voice
5 and video with the other client device through the videoconference server 2 in order that the users of the client devices may see and hear the other users through their personal client device, even if the users are in different places.

For example, the video and voice of the user of the client device 3 is played
10 by the client device 4 and the video and voice of the user of the client device 4 is played by the client device 3.

The exchanges of audio and video streams of audio and video packets will be now described in more details.

The client devices, for example the first and the second client devices 3 and 4 pictured, are involved in a videoconference session through the videoconference
15 server 2. In the videoconference session, the first client device 3 sends audio and video streams to the second client device 4 and the second client device 4 sends audio and video streams to the first client device 3. The audio and video streams comprise audio packets and associated video packets.

The audio packets and the video packets are Real Time Protocol (RTP)
20 packets. The audio packets and the video packets comprise a packet header comprising control information and a packet body comprising the audio or video data content. The packet header namely comprises a sequence number, which indicates the sequential position of a packet in a stream of emitted packets. The packet header also comprises a timestamp which indicates the timing of emitting
25 the packet. The packet body comprises encoded audio or video data which were captured during a time slot before emitting the packet. For example, video data of a packet comprises a sequence of images encoded by Red Green Blue (RGB) pixels and spatial coordinates, each image being captured per sub- time slot such that the concatenation of each sub-time slot is the time slot. For example, video data of a
30 packet comprises a sequence of transition of images, wherein each transition is

encoded by differences between two successive images, each captured during a sub-time slot.

The videoconference server 2 comprises a packet input interface 21 configured to receive audio and video packets from the client devices 3 and 4, and a
5 packet output interface 22 configured to transmit audio and video packets to the client devices 3 and 4.

For the sake of illustration, the only streams of audio and video packets that are represented on Figure 1 are an audio and video stream 5 emitted by the first client device 3 and an audio and video stream 6 transmitted by the videoconference
10 server 2 to the second client device 4.

The audio and video stream 5 of audio and video packets received in the packet input interface 21 of the videoconference server 2 is split in an audio stream 7 and a video stream 8.

The audio stream 7 is routed to an audio mixer 9 and the video stream is
15 routed to a video packet router 10.

The audio mixer 9 and the video packet router 10 and their functionalities will now be successively described.

The audio mixer 9 is a processing module dedicated to the processing of received audio streams of audio packets.

20 The audio mixer 9 is configured to mix the audio stream 7 with other audio streams. For the sake of illustration, another audio stream 11 is illustrated. The audio stream 11 is the audio part of an audio and video stream received from the second client device 4 (not represented).

The pictured resulting audio stream 12 is a mix of the audio streams 7 and
25 11. In other words, the resulting audio stream 12 is a stream of mixed audio packets from the audio packets of the audio streams 7 and 11. More precisely, an audio packet of the audio stream 7 is mixed with an audio packet of the audio stream 11 whose timestamp corresponds to the timestamp of the audio packet the audio stream 11 to generate an audio packet of the audio stream 12. The generated audio
30 packet has a header comprising the timestamp of the mixed audio packets. Two

corresponding timestamps are either identical or separated by a time value lower than a threshold.

The resulting audio stream 12 is transmitted by the audio mixer 9 to the output packet interface 22 in order to be routed to the second client user 4.

5 The audio mixer 9 is further configured to perform compression and decompression processing on the audio packets in order to mix the audio packets.

The videoconference server 2 further comprise an audio delay calculating module 13 configured to calculate an average audio processing delay introduced by the audio mixer 9 for the audio packets from the audio stream 7 and the audio
10 packets from the audio stream 11. The average audio processing delay depends namely on di-jittering algorithm of audio, compression type, usage of Noise Reduction, and the audio mixing algorithm. The average audio processing delay is computed for each client device, namely for the client device 4. The average audio processing delay is sent to a client detection module 14.

15 The video packet router 10 and its functionalities will now be described.

The video packet router 10 is a processing module dedicated to the processing of received video streams of video packets.

For the sake of illustration, another video stream 15 is illustrated. The video stream 15 is the video part of an audio and video stream received from the second
20 client device 4. For the sake of simplicity, the audio stream is not represented.

The video packet router 10 switches the video packets of the video stream 8 and 15.

In the pictured example, when the video packets are switched, the video packet router 10 selects the video stream corresponding to an active audio stream
25 (i.e. the audio stream incoming from a client device whose user is currently speaking). After selecting the video stream between the video streams incoming from the different client devices, the video packet router 10 transmits the selected video stream to the client devices.

Selecting the video stream corresponding to the active audio stream is
30 equivalent to selecting a subset of video packets received by the input interface.

In the pictured example, a resulting video stream 16 corresponds to the result of switching a video packet from the video stream 8 or a video packet of the other video stream 15.

The video packet router 10 also comprises a video buffer 17 configured to
5 duplicate the resulting video stream 16 and to temporarily store a duplication of the video packets of the video stream 16.

Once duplicated, the resulting video stream 16 is transmitted by the video packet router 10 to the output packet interface 22 in order to be routed to the second client user 4.

10 The audio and video stream 6 transmitted by the videoconference server 2 to the second client device 4 comprises both the resulting audio stream 12 and the resulting video stream 16.

Alternatively, in a non-pictured example, the video packet router 10 is configured to route the video packets. The video packet router 10 routes both the
15 video stream 15 and the video stream 8 to a third client device (not represented). The mixing of the video streams 15 and 8 is then performed by the third client device. The video packet router 10 may also comprise a video buffer 17. The video buffer 17 is configured to duplicate both the video stream 15 and the video stream 8 and to temporarily store a duplication of the video stream 15 and a duplication of
20 the video stream 8. The video packet router 10 may also comprise one video buffer per connected client device, for example a first video buffer dedicated to the first client device, a second video buffer dedicated to the second client device and a third video buffer dedicated to the third client device. In this example, the first video buffer is configured to duplicate and store all the video packets of the video stream
25 8 and the second video buffer is configured to duplicate and store all the video packets of the video stream 15.

The processing of received audio and video stream by the client device 4 will now be described.

The client device 4 comprises an audio packet reader (not represented) for
30 reconstituting the voices from reading the body of the audio packets of the audio

stream 12, and an audio speaker (not represented) for rendering the reconstituted voices. The audio packets are read in real-time as the audio packets are received.

The client device 4 further comprises a video delay module 18 configured to receive a request 25 comprising the average audio processing delay from the client detection module 14 of the videoconference server 2.

The client device 4 further comprises a video buffer 19 configured to temporarily store the video packets of the video stream 16.

The client device 4 further comprises a video packet reader (not represented) for reconstituting the video from the reading of the video packets of the video stream 16, and a video display (not represented) for displaying the reconstituted video.

The video buffer 19 receives instructions 20 from the video delay module 18 in order to introduce the average audio processing delay in the video packets stored in the video buffer 19 in order to synchronize the displaying of each video packet with the rendering of the voices of each audio packet of corresponding timestamp.

The synchronization of the video packets with the audio packets is thus performed in the client device 4, and not in the videoconference server 2.

Recovery of lost video packets

It may happen that a video packet of a video stream is lost at different stage of the travel between the first client device 3 and the second client device 4.

In order to recover a lost packet in a fast and efficient way, the videoconference server 2 and the client device 4 are equipped with loss detection and recovery functions.

The recovery of a lost video packet between the first client device 3 and the videoconference server 2 will now be described.

The videoconference server 2 comprises a packet loss detector 32 connected to the video packet router 10 in order to check if the sequence number of each received video packet is the successive sequence number expected, or not.

If the packet loss detector 32 detects that a sequence number is missing, then the packet loss detector 32 sends a request 33 to the first client device 3, the

request comprising the missing sequence number or the previous sequence number and an instruction to re-send the missing video packet.

The client device 3 receives the request 33 and re-send the missing video packet to the videoconference server 2.

5 The recovery of a lost video packet between the videoconference server 2 and the second client device 4 will now be described.

The client device 4 comprises a packet loss detector 36 configured to check if the sequence number of each received video packet is the successive sequence number expected, or not.

10 If the packet loss detector 36 detects that a sequence number is missing, then the packet loss detector 36 sends a request 35 to the video buffer 17 of the videoconference server 2, the request comprising the missing sequence number or the previous sequence number and an instruction to re-send the missing video packet.

15 The video buffer 17 receives the request 35 and re-send the missing video packet to the second client device 4 if the missing video packet is available.

Optionally, when a packet loss occurs and is detected, the videoconference server 2 sends an inhibition request to the client device 4, wherein the inhibition request comprises an instruction to inhibit the introduction, by the video delay
20 module 18, of the average audio processing delay in the recovered missing video packet.

Thanks to these features, the recovery delay is limited, because the videoconference server 2 detects a loss of a packet in an early manner. Moreover, the recovery delay is also limited as the lost packet is already stored in the video
25 buffer 17.

The advantages of the early recovery of lost video packets on lip synchronization will now be described with reference to Figures 2 and 3.

Figure 2 schematically illustrates the transmission of a video packet between the first client device 3, the videoconference server 2 and the second client

device 4 of the Figure 1. As represented by the flash 34, the video packet is lost between the videoconference server 2 and the second client device 4.

The client device 4 detects the loss in a detection time T_0 and sends a NACK request to the videoconference server 2 for recovery of the lost video packet.

5 The NACK request between the client device 4 and the videoconference server 2 has a latency D_{2S} . The videoconference server 2 retrieves the lost video packet in the video buffer 17 in a delay of recovery DR . Then, the videoconference server 2 sends the retrieved lost video packet PKT . The transmission of the video packet PKT has a latency DS_2 .

10 The total recovery time 38 is the addition of the detection time T_0 , latency D_{2S} , delay of recovery DR and latency DS_2 .

The advantage of storing temporarily the video packets in the video buffer 17 of the videoconference server 2 is to avoid adding supplementary latency in the total recovery time as no retransmission is required from the first client device 3.

15 Figure 3 schematically illustrates the transmission of a video packet between the first client device 3, the videoconference server 2 and the second client device 4 of the Figure 1. As represented by the flash 37, the video packet is lost between the first client device 3 and the videoconference server 2.

The videoconference server 2 detects the loss in a detection time T_0 and
20 sends a NACK request to the first client device 3 for recovery of the lost video packet. The NACK request between the videoconference server 2 and the first client device 3 has a latency DS_1 . The first client device 3 retrieves the lost video packet in a delay of recovery DPR . Then, the first client device 3 sends the retrieved lost video packet PKT . The transmission of the video packet PKT has a latency DS_1 . The
25 videoconference server 2 forwards the retrieved video packet to the second client device 4 in a delay of packet forwarding DPF . The videoconference server 2 transmits the video packet PKT to the second client device 4 in a latency DS_2 .

The total recovery time 39 is the addition of the detection time T_0 , latency DS_1 , delay of recovery DPR and latency D_{1S} .

The advantage of detecting that a video packet has been lost at an early stage, i.e. by the videoconference server 2, is to avoid adding supplementary latency in the total recovery time as no detection is needed from the second client device 4.

The client device 4 is represented more in detail with reference to Figure 4.

5 The client device 4 further comprises a data repository 22 connected to a processing unit 23 and to a communication interface 40 for receiving the audio and video stream 6. The video delay module 18 is also connected to the communication interface 40 in order to receive the request 25 through the communication interface 40. The client device 4 also comprises a video display 26, for example a liquid crystal
10 display (LCD) screen, a camera 27, for example a webcam, an audio speaker 28, for example an audio headset, and a microphone 29. The client device 4 also comprises an analog-to-digital and/or digital-to-analog converter 24 connected to the video display 26, camera 27, audio speaker 28 and a microphone 29. The processing unit 23 is configured to generate packets from the video and audio streams received
15 from the converter 24 and to generate audio and video streams for the video display 26 and the audio headset 28 from the audio packets of the audio and video stream 6 and video packets of the video buffer 19.

As represented, the video display 26 displays a video of users 30, 31 participating to the videoconference involving the client device 3 and another client
20 device. The mixed video packets of the received audio and video packets comprises a compositing of the captured videos of each user 30, 31 located in different geographical area.

Client detection functionality

Another functionality of the client detection module 14 will now be
25 described in relation to Figure 1.

The client detection module 14 of the videoconference server 2 is further configured to send a capable request to the client device 4, the capable request comprising instructions for detecting if the client device 4 is equipped with a video delay module 18.

If the client device 4 sends back an answer that the client device 4 is equipped with the client detection module 14, then the client detection module 14 sends periodically the request 25 in order to transmit an update of average audio processing delay in order that the client device 4 locally introduces the average
5 audio processing delay to the received video packets for lip synchronization.

If the client device 4 sends back an answer that the client device 4 is not equipped with the client detection module 14, then the average audio processing delay is introduced in the video packets by the videoconference server 2.

The invention is not limited to the described embodiments. The appended
10 claims are to be construed as embodying all modification and alternative constructions that may be occurred to one skilled in the art, which fairly fall within the basic teaching here, set forth. The use of the verb "to comprise" or "to include" and its conjugations does not exclude the presence of elements or steps other than those stated in a claim. Furthermore, the use of the article "a" or "an" preceding an
15 element or step does not exclude the presence of a plurality of such elements or steps. The invention may be implemented by means of hardware as well as software. The same item of hardware may represent several "means".

The client devices and the videoconference server described hereinabove may be implemented through the use of dedicated hardware as well as hardware
20 capable of executing software in association with appropriate software. When provided by a processor, the corresponding functions may be provided by a single dedicated processor, by a single shared processor, or by a plurality of individual processors, some of which may be shared. Moreover, explicit use of the term "processor" or "controller" should not be construed to refer exclusively to hardware
25 capable of executing software, and may implicitly include, without limitation, central processing unit (CPU), digital signal processor (DSP) hardware, network processor, application specific integrated circuit (ASIC), field programmable gate array (FPGA), read-only memory (ROM) for storing software, random access memory (RAM), and non-volatile storage. Other hardware, conventional and/or custom, may also be
30 included. A client device described hereinabove may be implemented in a unitary

manner or in a distributed manner. The videoconference server described hereinabove may be implemented in a unitary manner or in a distributed manner.

The videoconference server and the first client device may be collocated. The videoconference server and the first client device may be integrated in a single
5 device.

In the claims, any reference signs placed between parentheses shall not be construed as limiting the scope of the claims.

CLAIMS

1. A videoconference server (2) comprising
 - a packet input interface (21) configured to receive audio and associated video packets from a first client device (3),
 - 5 - a packet output interface (22) configured to send the received audio and associated video packets to a second client device (4),
 - an audio processing module (9) configured to, when the packet input interface (21) receives an audio packet from the first client device (3),
 - o processing the audio packet,
 - 10 o calculating an average audio processing delay of audio packets intended to the second client device,
 - o associating the average audio processing delay to the second client device,
 - o sending the average audio processing delay to the second client
 - 15 device in order that the second client device may synchronize a received video packet with the associated received audio packet,
 - a video router (10) comprising a video buffer (17) configured to duplicate a video packet received by the packet input interface (21) from the first client device (3), wherein the video packet is associated to the audio packet, and
 - 20 temporarily stores the duplicated video packet in relation with an identification information of the video packet,

the video router (10) being further configured to:

 - o receive a retrieving request from the second client device (4) for
 - retrieving a lost video packet, the retrieving request comprising an
 - 25 identification information of the video packet,
 - o retrieving the duplicated video packet associated with the identification information in the video buffer (17),

the packet output interface (22) being further configured to send the retrieved duplicated video packet to the second client device (4).

2. The videoconference server according to claim 1, wherein the video router (10) is further configured to detect that a video packet initially sent by the first client device (3) is not received by the packet input interface (21), the video router (10) being further configured to send a retrieving request to the first client device (3), and to receive from the first client device (3) the initially sent video packet that has not been initially received.
 3. The videoconference server according to claim 1, wherein the videoconference server (2) further comprises:
 - a client device configuration detection module (14) configured to, prior the initiation of a communication session with the second client device (4), send a configuration request to the second client device (4),
 - receiving a configuration response from the second client device (4), the configuration response comprising an indication whether the second client device (4) is adapted to locally synchronize video packets with associated audio packets or not adapted,
- the videoconference server (2) being further configured to, when the configuration response comprises an indication that the second client device (4) is not adapted to locally synchronize video packets with associated audio packets,
- delaying the received video packet by the calculated average audio processing delay prior to sending the received video packet to the second client device (4).
4. The videoconference server according to claim 1, wherein the audio packets are identified by audio packet identifiers and the associated video packets are identified by video packet identifiers, the video buffer being configured to detect that a video packet initially sent by the first client device is not

- received by the packet input interface by controlling the video packet identifier of a received video packet and by comparing the video packet identifier to an expected video packet identifier, wherein the expected video packet identifier is associated with the audio packet identifier of a received audio packet to which the expected video packet is associated.
- 5
5. The videoconference server according to claim 1, wherein the audio and video packets are each identified by a packet number, the packet number indicating the position of the packet in a sequence of emitted packets, wherein the retrieving request comprises the packet number of the video
- 10
- packet.
6. The videoconference server according to claim 1, wherein processing the audio packet comprises mixing the audio packet with other received audio packets.
7. The videoconference server according to claim 1, wherein the audio
- 15
- processing module (9) is configured to periodically recalculate the average audio processing delay and to send the recalculated average audio processing delay to the second client device (4) in order that the second client device (4) may update the average audio processing delay used to synchronize a received video packet with the associated received audio
- 20
- packet.
8. The videoconference server according to claim 1, wherein the video buffer (17) is a circular buffer.
9. The videoconference server according to claim 1, wherein the packets are Real Time Protocol (RTP) packets.
- 25
10. A videoconference endpoint comprising
- an emitting interface (40) configured to emit audio and associated video packets for a videoconference server (2),
 - a receiving interface (40) configured to receive audio and associated video packets from the videoconference server (2),

- a video buffer (19) configured to temporarily store the received video packets,
- a synchronization module (18) configured to receive an average audio processing delay from the videoconference server (2), and to delay the received video packets by the average audio processing delay in order to synchronize each video packet with a received audio packet associated to the video packet,
- a packet loss detector configured to detect that a video packet initially sent by the videoconference server (2) is not received by the receiving interface (40),

the packet loss detector being further configured to send a retrieving request (35) to the videoconference server (2) for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet,

the receiving interface (40) being further configured to receive a duplication of the lost video packet in response to sending the retrieving request (35).

11. The videoconference endpoint according to claim 10, further comprising an audio speaker (28) configured to render a voice signal carrier by the audio packets and a video display (26) configured to display a video signal carrier by the received video packets, the video display being configured to display a portion of the video signal carried by a video packet associated with an audio packet at the same time that the audio speaker renders a portion of the voice signal carried by the associated audio packet.

12. A videoconference system (1) comprising:

- a videoconference server (2) comprising:
 - o a packet input interface (21) configured to receive audio and associated video packets from a first client device (3),
 - o a packet output interface (22) configured to send the received audio and associated video packets to a second client device (4),

- an audio processing module (9) configured to, when the packet input interface (21) receives an audio packet from a first client device (3),
 - processing the audio packet,
 - calculating an average audio processing delay of audio packets intended to the second client device (4),
 - associating the average audio processing delay to the second client device (4),
 - sending the average audio processing delay to the second client device (4) in order that the second client device (4) may synchronize a received video packet with the associated received audio packet,
- a video router (10) comprising a video buffer (17) configured to, duplicate a video packet received by the packet input interface (21) from a first client device (3), wherein the video packet is associated to the audio packet, and temporarily store the duplicated video packet in relation with an identification information of the video packet,

the video router (10) being further configured to:

- receive a retrieving request from the second client device (4) for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet,
- retrieving the duplicated video packet associated with the identification information in the video buffer (17),

the packet output interface (22) being further configured to send the retrieved duplicated video packet to the second client device (4),

the videoconference system further comprising,

- the second client device (4) comprising

- an emitting interface (40) configured to emit audio and associated video packets for the videoconference server (2),
- a receiving interface (40) configured to receive audio and associated video packets from the videoconference server (2),
- 5 ○ a video buffer (19) configured to temporarily store the received video packets,
- a synchronization module (18) configured to receive an average audio processing delay from the videoconference server (2), and to delay the received video packets by the average audio processing delay in
- 10 order to synchronize each video packet with a received audio packet associated to the video packet,
- a packet loss detector configured to detect that a video packet initially sent by the videoconference server (2) is not received by the receiving interface (40),
- 15 the packet loss detector being further configured to send a retrieving request (35) to the videoconference server (2) for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet,
- the receiving interface (40) being further configured to receive a duplication of the lost video packet in response to sending the retrieving request (35),
- 20 13. A method for synchronizing audio and video packets from a first client device to a second client device, wherein the first client device and the second client device are involved in a video communication session, the method comprising:
- receiving, in a videoconference server (2), audio packets and associated video
- 25 packets from a first client device (3) to be forwarded to a second client device (4),

- processing, in the videoconference server (2), the received audio packets and calculating an average audio processing delay of audio packets intended to the second client device (4),
 - sending the processed audio packets to the second client device (4),
 - 5 - duplicating and temporarily storing, in the videoconference server (2), at least some of the received video packets associated with the received audio packets,
 - sending by the videoconference server (2) the received video packets to the second client device (4),
 - 10 - receiving, in the second client device (4), the processed audio packets and the video packets,
 - sending, by the videoconference server (2), the calculated average audio processing delay to the second client device (4),
 - delaying, by the second client device (4), the received video packets by the average audio processing delay.
 - 15
14. The method according to claim 13, further comprising retrieving a video packet that has not been received by the second client device (4), wherein retrieving comprises:
- detecting, by the second client device (4), that a video packet sent by the videoconference server (2) is not received,
 - 20 - sending, by the second client device (4), a retrieving request for retrieving the video packet that has not been received, the retrieving request comprising an identification information of the video packet that has not been received,
 - receiving, by the videoconference server (2), the retrieving request from the second client device (4),
 - 25 - retrieving, by the videoconference server (2), the duplicated video packet associated with the identification information,
 - sending the retrieved duplicated video packet to the second client device (4),
 - receiving, by the second client device (4) the retrieved duplicated video packet.
 - 30

15. The method according to claim 13, further comprising retrieving a video packet that has not been received by the videoconference server (2), wherein retrieving comprises:

- 5 - detecting, by the videoconference server (2), that a video packet sent by the first client device (3) is not received by the videoconference server (2),
- sending, by the videoconference server (2), a retrieving request to the first client device (3), wherein the retrieving request comprises an identification information of the video packet that has not been received,
- 10 - receiving, by the videoconference server (2), the video packet that has not been received,

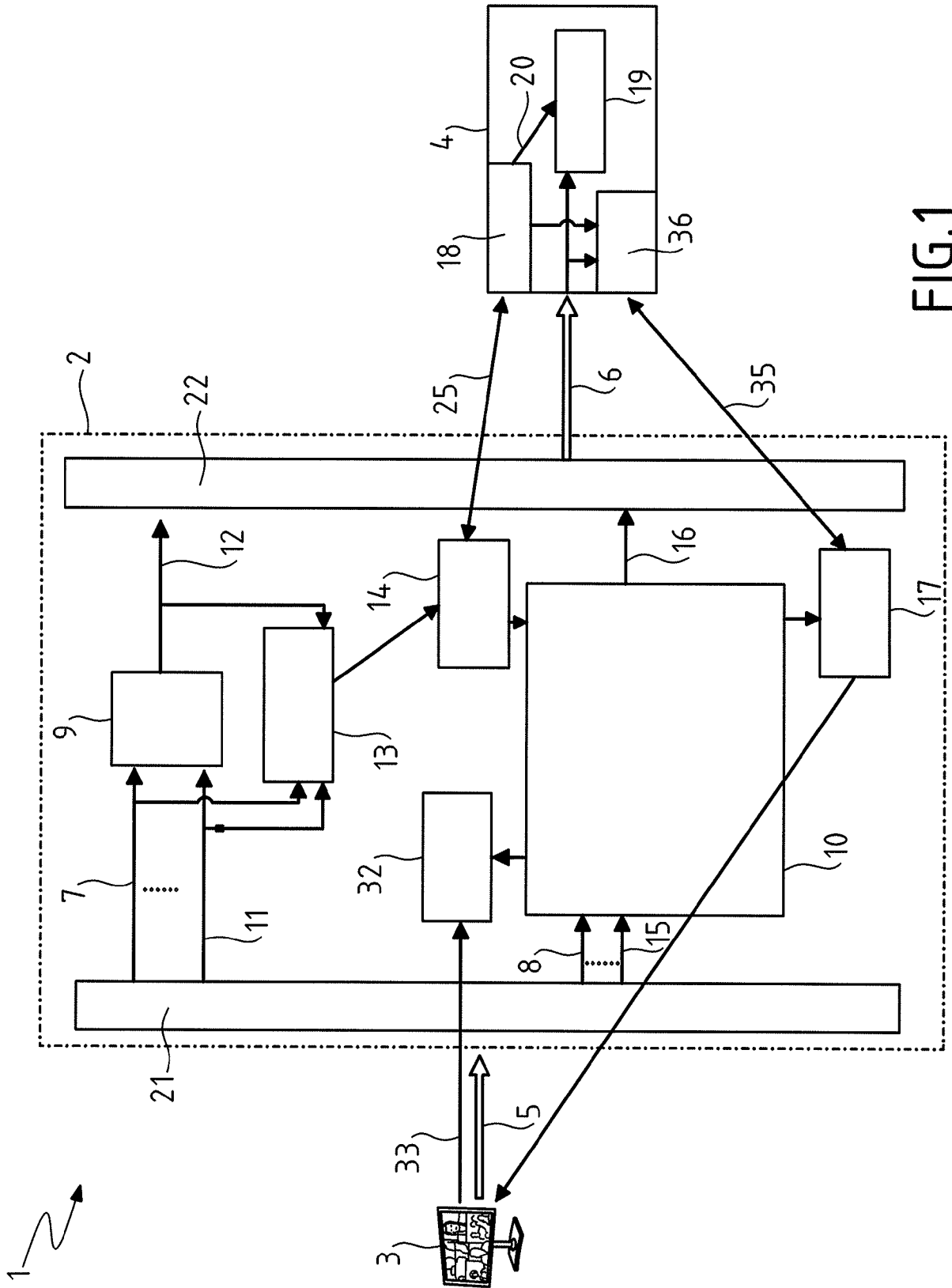


FIG. 1

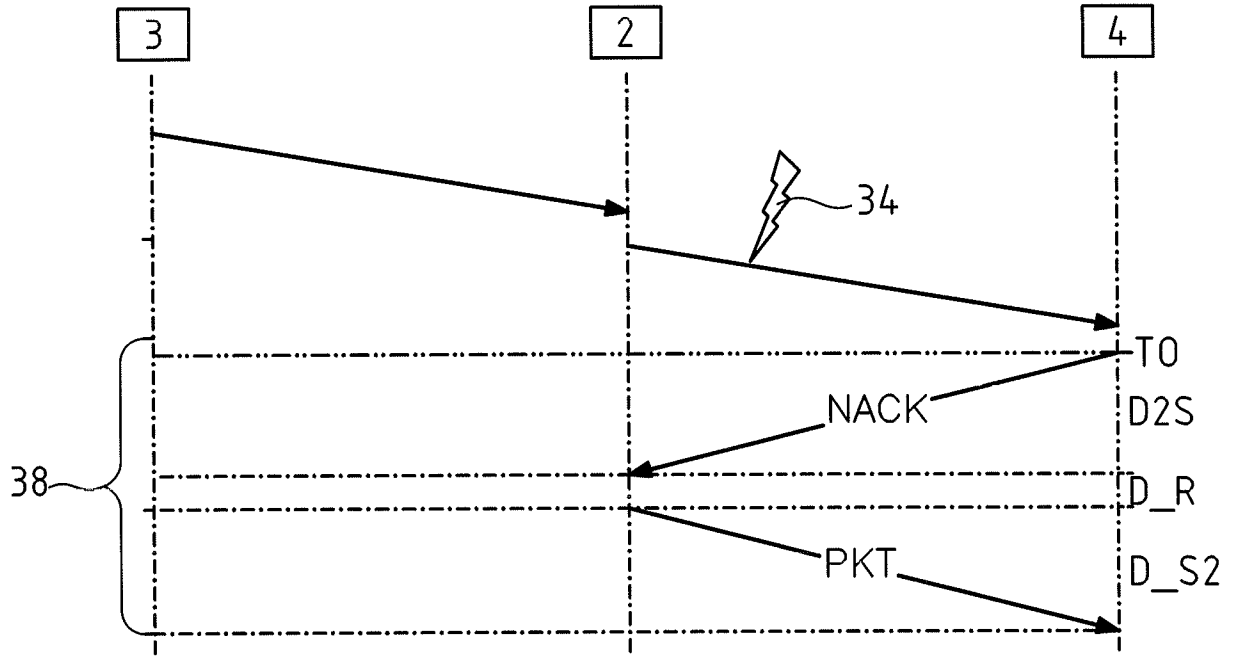


FIG.2

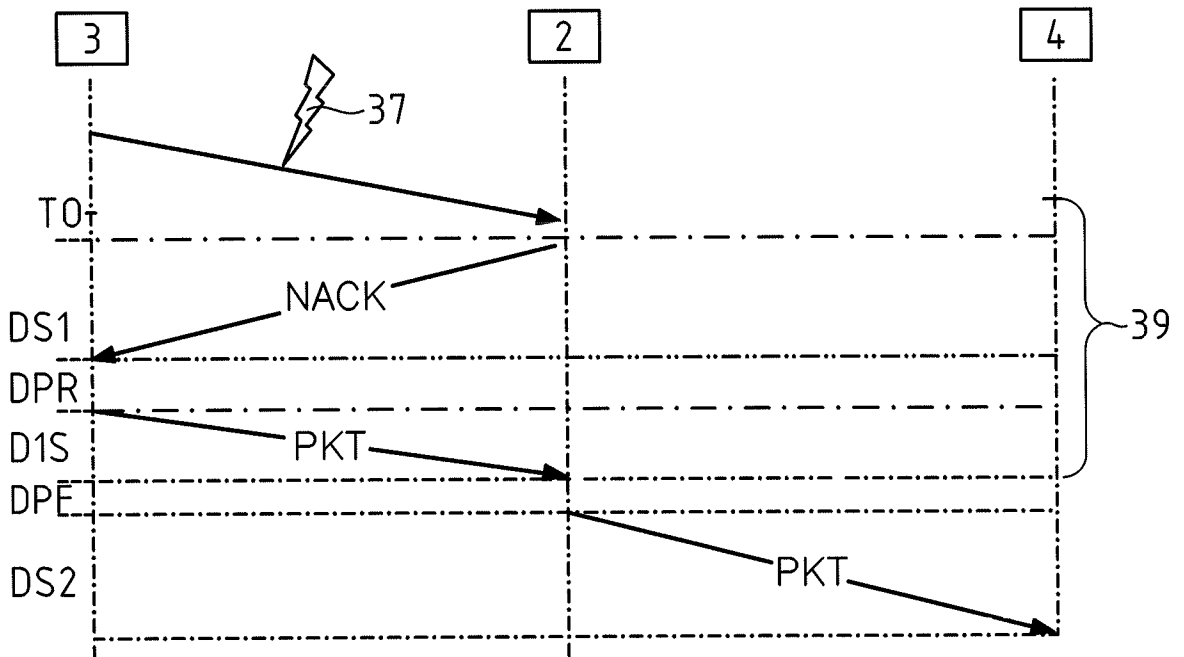


FIG.3

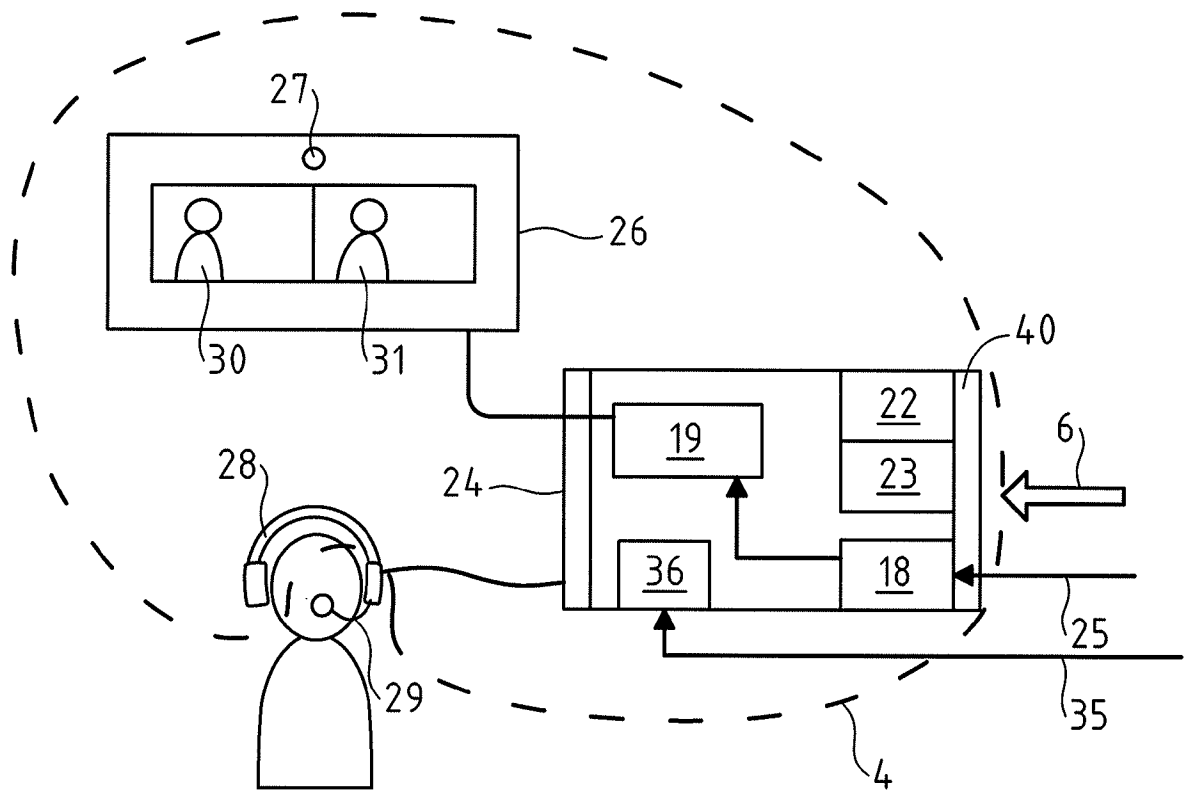


FIG.4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2017/050051

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - H04L 7/00; H04N 7/00; H04N 7/14; H04N 7/15 (2017.01)

CPC - H04L 7/00; H04L 65/00; H04N 7/00; H04N 7/14; H04N 7/141; H04N 7/147; H04N 7/15 (2017.02)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 348/14.010; 348/14.080; 370/259; 370/260; 709/204 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X ---	US 2012/0140018 A1 (PIKIN et al) 07 June 2012 (07.06.2012) entire document	1-4, 7, 9, 13-15
Y		5, 6, 8
Y	US 2006/0007943 A1 (FELLMAN) 12 January 2006 (12.01.2006) entire document	5, 6
Y	US 2016/0283860 A1 (MICROSOFT TECHNOLOGY LICENSING, LLC) 29 September 2016 (29.09.2016) entire document	8
A	US 2008/0034104 A1 (KARITI et al) 07 February 2008 (07.02.2008) entire document	1-9, 13-15
A	US 2014/0168354 A1 (RABBIT, INC) 19 June 2014 (19.06.2014) entire document	1-9, 13-15
A	US 2015/0029856 A1 (ROZENBERG) 29 January 2015 (29.01.2015) entire document	1-9, 13-15
A	US 2016/0021149 A1 (SWYME IP BV) 21 January 2016 (21.01.2016) entire document	1-9, 13-15

 Further documents are listed in the continuation of Box C. See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

05 May 2017

Date of mailing of the international search report

26 MAY 2017

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2017/050051

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
See Extra Sheet(s)

1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-9, 13-15

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/IB2017/050051

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claims 1-9,13-15, drawn to a videoconference server.

Group II, claims 10-12, drawn to a videoconference endpoint.

The inventions listed as Groups I-II do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical feature of the Group I invention: a video router (10) comprising a video buffer (17) configured to duplicate a video packet received by the packet input interface (21) from the first client device (3), wherein the video packet is associated to the audio packet, and temporarily stores the duplicated video packet in relation with an identification information of the video packet, the video router (10) being further configured to: receive a retrieving request from the second client device (4) for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet as claimed therein is not present in the invention of Group II. The special technical feature of the Group II invention: a packet loss detector configured to detect that a video packet initially sent by the videoconference server (2) is not received by the receiving interface (40), the packet loss detector being further configured to send a retrieving request (35) to the videoconference server (2) for retrieving a lost video packet, the retrieving request comprising an identification information of the video packet, the receiving interface (40) being further configured to receive a duplication of the lost video packet in response to sending the retrieving request as claimed therein is not present in the invention of Group I.

Groups I and II lack unity of invention because even though the inventions of these groups require the technical feature of an emitting interface configured to emit audio and associated video packets for a videoconference server; sending the average audio processing delay to the second client device in order that the second client device may synchronize a received video packet with the associated received audio packet, this technical feature is not a special technical feature as it does not make a contribution over the prior art.

Specifically, US 2008/0034104 A1 (KARITI et al) 07 February 2008 (07.02.2008) teaches an emitting interface configured to emit audio and associated video packets for a videoconference server (Paras. 5 and 65-67 and Fig. 1); sending the average audio processing delay to the second client device in order that the second client device may synchronize a received video packet with the associated received audio packet (Paras. 81-83).

Since none of the special technical features of the Group I or II inventions are found in more than one of the inventions, unity of invention is lacking.