(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 October 2012 (11.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/138573 A1

(51) International Patent Classification:
H04N 21/2343 (2011.01) *H04L 1/00* (2006.01)
H04N 7/15 (2006.01) *H04L 12/56* (2006.01)
H04L 12/18 (2006.01) *H04N 21/262* (2011.01)
H04L 29/06 (2006.01) *H04N 21/2365* (2011.01)

(21) International Application Number:
PCT/US2012/031606

(22) International Filing Date:
30 March 2012 (30.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
13/066,086 6 April 2011 (06.04.2011) US

(71) Applicant *(for all designated States except US)*: **CISCO TECHNOLOGY, INC.** [US/US]; 170 West Tasman Drive, San Jose, CA 95134-1706 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **PURI, Rohit** [US/US]; 1910 White Oaks Road, Campbell, CA 95008 (US). **CHOU, Jim** [US/US]; 6566 Sloping Meadow Ct., San Jose, CA 95135 (US).

(74) Agent: **KAPLAN, Cindy, S.**; P.O. Box 2448, Saratoga, CA 95070 (US).

(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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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.

(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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(54) Title: VIDEO CONFERENCING WITH MULTIPOINT CONFERENCING UNITS AND MULTIMEDIA TRANSFORMATION UNITS WITH ACTIVE SPEAKER DETECTION

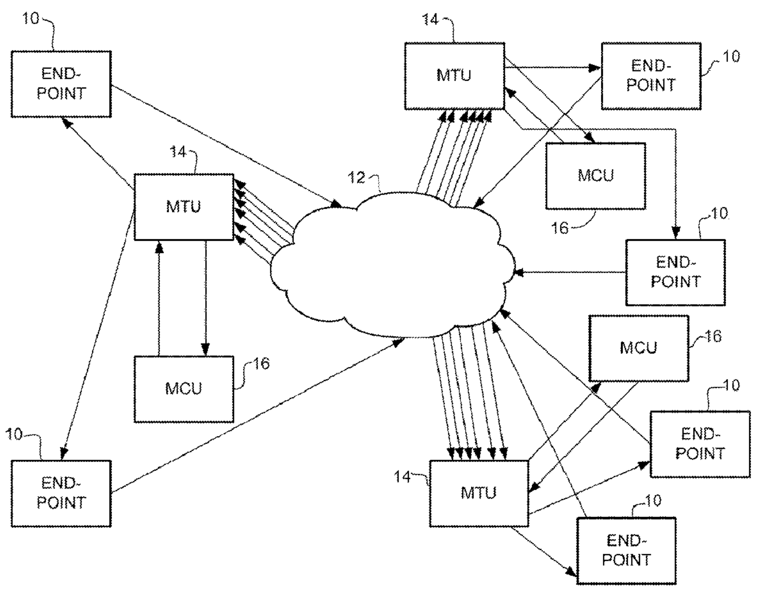

Fig. 1

(57) Abstract: In one embodiment, a method includes receiving at a multimedia transformation unit, media streams from a plurality of endpoints, transmitting audio components of the media streams to a multipoint conferencing unit, receiving an identifier from the multipoint conferencing unit identifying tone of the media streams as an active speaker stream, processing at the multimedia transformation unit, a video component of the active speaker stream, and transmitting the active speaker stream to one or more of the endpoints without transmitting the video component to the multipoint conferencing unit. An apparatus is also disclosed. In other embodiments redundant streams are used to replace a failed stream, also packet offsetting is used to decorrelate channels.

WO 2012/138573 A1

VIDEO CONFERENCING WITH MULTIPOINT CONFERENCING UNITS AND MULTIMEDIA TRANSFORMATION UNITS WITH ACTIVE SPEAKER DETECTION

TECHNICAL FIELD

The present disclosure relates generally to communication networks, and more particularly, to video conferencing.

BACKGROUND

Video conferencing provides real-time, face-to-face interactions between people from distributed geographic locations and has become increasingly popular as a way to reduce travel expenses and enhance collaboration. Conventional video conferencing systems often use a centralized multipoint conferencing unit (MCU) to couple communication equipment at various conference sites. The MCU combines video, audio, and data streams from multiple endpoints into a single, multiple location, interactive session. Video conferencing systems developed by different suppliers often do not interoperate very well due to signaling and format mismatches between different endpoints.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 illustrates an example of a network in which embodiments described herein may be implemented.

Fig. 2 depicts an example of a network device useful in implementing embodiments described herein.

Fig. 3 is a block diagram illustrating details of a multimedia transformation unit (MTU), in accordance with one embodiment.

Fig. 4 is a flowchart illustrating a process for using redundant streams at the MTU to provide robustness, in accordance with one embodiment.

Fig. 5 is a block diagram illustrating an example for processing redundant streams at the MTU, in accordance with one embodiment.

Fig. 6 is a block diagram illustrating optimization at a multipoint conferencing unit (MCU), in accordance with one embodiment.

Fig. 7 is a flowchart illustrating a process for MTU optimization, in accordance with one embodiment.

Corresponding reference characters indicate corresponding parts throughout the several views of the drawings.

DESCRIPTION OF EXAMPLE EMBODIMENTS

Overview

In one embodiment, a method generally comprises receiving at a multimedia transformation unit, media streams from a plurality of endpoints, transmitting audio components of the media streams to a multipoint conferencing unit, receiving an identifier from the multipoint conferencing unit identifying one of the media streams as an active speaker stream, processing at the multimedia transformation unit, a video component of the active speaker stream, and transmitting the active speaker stream to one or more of the endpoints without transmitting the video component to the multipoint conferencing unit.

In another embodiment, an apparatus generally comprises a processor configured to transmit media streams received from a plurality of endpoints to a multipoint conferencing unit, receive a plurality of redundant streams from the multipoint conferencing unit, decode two or more substreams generated from the redundant streams, identify a failure of one of the substreams, and utilize another one of the substreams in place of the failed substream. The apparatus further comprises memory for storing a redundancy parameter corresponding to a number of substreams decoded at the apparatus.

In yet another embodiment, an apparatus generally comprises an interface for receiving media streams from a multimedia transformation unit, the media streams transmitted from a plurality of endpoints, and a processor for interconnecting the endpoints in a conference session. The processor is configured for offsetting delivery of packets to the multimedia transformation unit to decorrelate channels at the multimedia transformation unit.

Example Embodiments

The following description is presented to enable one of ordinary skill in the art to make and use the embodiments. Descriptions of specific embodiments and applications are provided only as examples, and various modifications will be readily apparent to those skilled in the art. The general principles described herein may be applied to other applications without departing from the scope of the embodiments. Thus, the embodiments are not to be limited to those shown, but are to be accorded the widest scope consistent with the principles and features described herein. For purpose of clarity, details relating to technical material that is known in the technical fields related to the embodiments have not been described in detail.

The embodiments described herein provide an efficient implementation for joint deployment of a multipoint conferencing unit (multipoint control unit) (MCU) and multimedia transformation unit (MTU) to provide video conferencing. As described in detail below, one or more of the embodiments provide robustness through the use of redundant streams at the MTU or controlled delivery of packets at the MCU. One or more of the embodiments provide an optimization to increase capacity at the MTU or minimize bandwidth between the MCU and MTU.

The embodiments operate in the context of a data communications network including multiple network devices (nodes). Some of the devices in the network may be switches, routers, gateways, servers, call managers, service points, media sources, media receivers, MTUs, MCUs, or other network devices.

Referring now to the figures, and first to Fig. 1, an example of a network in which embodiments described herein may be implemented is shown. The network includes a plurality of endpoints (e.g., media sources/receivers) 10 in communication with one or more conferencing systems (MTU 14 and MCU 16) via network 12. The network 12 may include one or more networks (e.g., radio access network, public switched network, local area network, metropolitan area network, wide area network, enterprise network, Internet, intranet, or any other network). The nodes 10, 14, 16 are connected via communication links (wired or wireless). Media flow paths between the endpoints 10 may include any number or type of intermediate nodes (e.g., routers, switches, gateways, or other network devices), which facilitate passage of data between the endpoints.

The term 'conference' as used herein may include any communication session between a plurality of users transmitted using any audio or video means, including signals, data, or messages transmitted through voice or video devices. The term 'media' as used herein refers to video, audio, data, or any combination thereof (e.g., multimedia). The media may be encrypted, compressed, or encoded according to any format. The media content may be transmitted as streaming media or media files, for example.

The endpoints 10 are configured to originate or terminate communications over the network 12. The endpoints 10 may be any device or combination of devices configured for receiving, transmitting, or receiving and transmitting media flows. For example, the endpoint 10 may be a personal computer, media center device, mobile device (e.g., phone, personal digital assistant), or any other device capable of engaging in audio, video, or data exchanges within the network 12. The endpoints 10 may include, for example, one or more processor, memory, network interface, microphone, camera, speaker, display, keyboard, whiteboard, and video conferencing interface.

In the example shown in Fig. 1, the network includes an MTU 14 and MCU 16 for each conferencing system. One or more media streams or components of the streams received at the MTU 14 are processed and then forwarded to the MCU 16. The MCU 16 uses the streams to create a representation for the multi-location interactive session. The MTU 14 receives this representation from the MCU 16 and performs processing on the representation with the objective of tailoring it according to the capabilities and limitations of the associated endpoints 10. This representation is distributed to various endpoints 10 involved in the interactive session.

The MTU 14 enables different endpoints 10 to interoperate with one another by performing protocol and codec conversions on input streams from source endpoints 10 to create output streams that are well matched to destination endpoints. The MTU 14 shields any peculiarities or limitations of various endpoints 10 involved in the video conference session by performing all of the necessary processing in the network. The processing may include, for example, conversion from one codec format to another codec format, multimedia re-sampling, spatial resolution scaling for video streams, etc. The term 'multimedia transformation unit' as used herein refers to any network device configured for processing media streams received from two or more endpoints.

The MCU 16 is configured to manage a multi-party conference by connecting multiple endpoints 10 into the same conference. The MCU 16 collects audio and video signals transmitted by conference participants through their endpoints 10 and distributes the signals to other participants of the conference. The MCU 16 combines video, audio, and data streams into a single multi-location, interactive session. The term 'multipoint conferencing unit' as used herein refers to any network device configured for interconnecting a plurality of endpoints in a conference session.

It is to be understood that the network shown in Fig. 1 and described herein is only an example and that the embodiments described herein may be implemented in networks having different network topologies and network devices, without departing from the scope of the embodiments.

An example of a network device 20 (e.g., MTU 14, MCU 16) that may be used to implement embodiments described herein is shown in Fig. 2. In one embodiment, network device 20 is a programmable machine that may be implemented in hardware, software, or any combination thereof. The device 20 includes one or more processors 22, memory 24, network and MTU/MCU interfaces 26, and one or more media processing components 28 (e.g., encoder, decoder, etc. as shown in Figs. 3 and 5). Memory 24 may be a volatile memory or non-volatile storage, which stores various applications, modules, and data for execution and use by the processor 22.

Logic may be encoded in one or more tangible media for execution by the processor 22. For example, the processor 22 may execute codes stored in a computer-readable medium such as memory 24. The computer-readable medium may be, for example, electronic (e.g., RAM (random access memory), ROM (read-only memory), EPROM (erasable programmable read-only memory)), magnetic, optical (e.g., CD, DVD), electromagnetic, semiconductor technology, or any other suitable medium.

The network and MTU/MCU interfaces 26 may comprise one or more wireless or wired interfaces (linecards, ports) for receiving signals or data or transmitting signals or data to other devices. The interfaces 26 may include, for example, an Ethernet interface for connection to a computer or network.

The network device 20 may further include any suitable combination of hardware, software, algorithms, processors, devices, components, or elements operable to facilitate the conferencing capabilities described herein. It is to be understood that the network device 20 shown in Fig. 2 and described above is only one example and that different configurations of network devices may be used.

Fig. 3 illustrates an example of processing performed on a media stream by the MTU 14, in accordance with one embodiment. A top leg (as viewed in Fig. 3) illustrates a path taken by an input stream at the MTU 14, and a bottom leg illustrates a path taken by an output stream at the MTU. For simplification, only one input stream and one output stream are shown, however, any number of input streams or output streams may be processed at the MTU 14. The media stream (e.g., audio component, video component, audio and video components) received from the endpoint 10 is processed by a decrypter 30, decoder 32, pixel processor 34, and encoder 36 before being transmitted to the MCU 16. The stream received from the MCU 16 is processed by decoder 32, pixel processor 34, encoder 36, encrypter 38, and jitter adapter 40. It is to be understood that the processing components shown in Fig. 3 are only an example and that additional, fewer, or different processing components may be utilized at the MTU 14, without departing from the scope of the embodiments. Also, in one or more embodiments only a portion of the media stream may pass through one or more of the components. For example, one or more video components may follow path 42 and bypass the encoder 36 in the top leg of the MTU and decoder 32 and pixel processor 34 in the bottom leg of the MTU (as shown in Fig. 3), while the audio components pass through all of the processing components, as described below.

In one example, various outbound streams from disparate endpoints 10 with heterogeneous capabilities are combined into a single stream, with their audio components mixed and their video components composited together by the conferencing system. The top leg of the MTU 14 (Fig. 3) prepares the input streams for the MCU 16 by equalizing the heterogeneity (e.g., spatial and temporal resolution) of the various endpoints 10. The bottom leg of the MTU 14 then adapts the stream for the limitations and capabilities of the various endpoints 10. If the MCU 16 is incognizant of the presence of the MTU 14 between itself and the endpoints 10, it would transmit the composite stream individually to all of the endpoints 10, with each of the identical bit-streams being intercepted by the MTU 14 and tailored for the target endpoints 10. As described below, the embodiments described herein allow for interaction between the MTU 14 and MCU 16 to provide an efficient joint deployment of the MTU and MCU.

The MTU 14 serves as a proxy for the associated endpoints 10, terminating a logical connection between MCU 16 and an endpoint 10, and replacing it with two physical connections, one between the MCU 16 and MTU 14, and the second from the MTU to the corresponding endpoint. As a result, the transmission from the MCU 16 to the MTU 14 is highly redundant (i.e., many of the streams from the MCU are identical or close to identical). The following describes embodiments that use this redundancy to provide robustness in the conferencing system. The redundancy may be used as a means for providing N-to-1 redundancy, where N is equal to the number of redundant streams received from the MCU 16. In the following examples, a redundancy parameter (k) is used to identify the number of substreams (e.g., streams, substreams, sets of frames) processed at the MTU 14.

Fig. 4 is a flowchart illustrating a process for utilizing the redundancy at the MTU 14 to provide system robustness, in accordance with one embodiment. The MCU 16 transmits each of the redundant streams to the MTU 14 and the redundancy is used inside of the MTU to provide high availability. At step 40, the MTU 14 receives N redundant streams from the MCU 16. The MTU decodes k (e.g., two or more) substreams generated from the redundant streams, where k<=N (step 42). If one of the k substreams fails due to hardware or software unavailability (step 44), one of the other k streams is used in place of the failed stream (step 46).

In a first example, the substreams (k) comprise a subset of the N redundant streams received from the MCU 16. This provides k-to-1 redundancy at the expense of k-1 extra resources.

In a second example, the redundancy is used to provide forward error correction. For example, this may correspond to an (N, 1)-repetition code, where N-1 erasures may be corrected for. In this case, as long as 1 of the N streams is received, reliable packet delivery to the endpoint 10 can take place.

In a third example, each of the N streams is subsampled by a factor of k in the video domain. The redundancy prediction structure may be a temporally scalable structure, where given an input stream, each of the k substreams capture a mutually exclusive subsampled portion of the video that is downsampled by a factor of k. As shown in Fig. 5, data is received from the network and processed at input stream processing components 54. Upon receiving the N redundant streams from the MCU 16, the MTU 14 decodes any k streams which corresponds to an (N,k) forward error correction code. This is sufficient to recover the k redundant sets of frames (e.g., 1, 2, 3, . . . k, as shown in Fig. 5). Furthermore, if one of these k substreams fails, then the other k-1 substreams may be used with an error concealment algorithm 56 to provide the reconstructed stream to all of the endpoints 10. As long as not all of the k substreams fail, then a version of the input stream may be provided to the appropriate endpoint 10 by the MTU 14.

Fig. 6 illustrates embodiments implemented at the MCU 16 to provide robustness. In a first example shown in Fig. 6, the MCU 16 offsets (i.e., staggers) delivery of packets 60 to decorrelate channels to the MTU 14 in case that the channels from MCU to MTU consist of paths that traverse similar routes. This prevents congestion or losses within the network.

In a second example, the MCU 16 is configured to only send non-redundant streams with appropriate side-information that enables the MTU 14 to choose streams appropriately for various endpoints 10. The communication bandwidth between the MCU 16 and MTU 14 is therefore reduced. For example, if S1 and S2 are redundant streams, the MCU 16 would only send one of the streams (e.g., only S1 or S2) to the MTU 14 (Fig. 6). In one embodiment, the MTU 14 communicates to the MCU 16 the number of streams to transmit.

In another embodiment, the MTU 14 and MCU 16 implementation is optimized for use with active speaker based stream selection. Fig. 7 is a flowchart illustrating active speaker based optimization at the MTU 14, in accordance with one embodiment. At step 70, the MTU 14 receives media streams from endpoints 10. The MTU 14 equalizes the audio components of the media streams received from various endpoints 10 to a common format that can be processed by the MCU 16 and sends the audio components of the streams to the MCU (step 72). The MCU 16 uses the audio components of the media streams from various endpoints 10 to identify the stream associated with the current active speaker at one of the endpoints. The MCU 16 identifies the current active speaker and communicates this information to the MTU 14. The MTU 14 receives an identifier for the active speaker stream at step 74. The MTU 14 processes (e.g., decodes, encodes) the video component for the active speaker stream identified by the MCU 16 (step 76). The MTU 14 preferably maintains in memory a list of one or more previously active speaker streams so that the MTU 14 can also process the previously active speaker streams for distribution to one or more of the endpoints 10.

The MTU 14 transmits the media streams to the endpoints without transmitting the video component to the multipoint conferencing unit (step 78). For example, the video components that are processed at the MTU 14 may follow path 42 shown in Fig. 3. The MTU 14 may choose to only perform processing on the uncompressed stream data corresponding to the currently active speaker stream and a specified number of previously active speaker streams (e.g., 0, 1, 2, . . .). The remaining video components (e.g., video components of media streams that are not identified as currently or previously active speaker streams) may be dropped at the MTU 14. The MTU 14 may transmit, for example, the media stream corresponding to the currently active speaker to all of the endpoints 10 except for the currently active speaker and transmit the stream corresponding to the last active speaker to the currently active speaker.

Transmission bit-rates for the audio component are significantly less than those for the video component and the active speaker stream identifier that is transmitted from the MCU 16 to the MTU 14 is an ultra-low bit-rate communication. Therefore, significant bandwidth savings are realized in the communication between the MTU 14 and MCU 16 by eliminating communication of the video components to the MCU. Furthermore, bypassing of additional encoding and decoding for the video components at the MTU 14, as shown in Fig. 3, and dropping video components of streams that are not currently or previously active speaker streams increase available capacity at the MTU.

The embodiments described herein may be implemented independent from one another or more than one embodiment may be implemented together.

Although the method and apparatus have been described in accordance with the embodiments shown, one of ordinary skill in the art will readily recognize that there could be variations made without departing from the scope of the embodiments. Accordingly, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

CLAIMS

WHAT IS CLAIMED IS:

1. A method comprising:

receiving at a multimedia transformation unit, media streams from a plurality of endpoints;

transmitting audio components of said media streams to a multipoint conferencing unit;

receiving an identifier from the multipoint conferencing unit identifying one of said media streams as an active speaker stream;

processing at the multimedia transformation unit, a video component of said active speaker stream; and

transmitting said active speaker stream to one or more of said plurality of endpoints without transmitting said video component to the multipoint conferencing unit.

2. The method of claim 1 wherein said video components of said media streams not identified as active speaker streams are dropped at the multimedia transformation unit.

3. The method of claim 2 wherein said active speaker streams comprise currently and previously active speaker streams.

4. The method of claim 1 further comprising maintaining a list of previously active speaker streams at the multimedia transformation unit.

5. The method of claim 1 wherein processing said video component comprises decoding and encoding said video component.

6. The method of claim 5 wherein additional decoding and encoding processes are bypassed at the multimedia transformation unit.

7. An apparatus comprising:

a processor configured to transmit media streams received from a plurality of endpoints to a multipoint conferencing unit, receive a plurality of redundant streams from the multipoint conferencing unit, decode two or more substreams generated from the redundant streams, identify a failure of one of said substreams, and utilize another one of said substreams in place of said failed substream; and

memory for storing a redundancy parameter corresponding to a number of said substreams decoded at the apparatus.

8. The apparatus of claim 7 wherein the redundancy parameter is less than the number of redundant streams received from the multipoint conferencing unit.

9. The apparatus of claim 7 wherein each of the redundant streams received from the multipoint conferencing unit is subsampled by a factor corresponding to the redundancy parameter to generate said substreams.

10. The apparatus of claim 7 wherein the processor is configured to apply an error concealment algorithm to said substreams to generate a reconstructed stream for transmittal to said plurality of endpoints.

11. The apparatus of claim 7 wherein the processor is configured to perform forward error correction.

12. The apparatus of claim 7 wherein the apparatus is a multimedia transformation unit.

13. The apparatus of claim 7 wherein said substreams comprise a subset of the redundant streams received from the multipoint conferencing unit.

14. The apparatus of claim 7 wherein each of said substreams comprises a set of frames from the redundant streams.

15. An apparatus comprising:

an interface for receiving media streams from a multimedia transformation unit, said media streams transmitted from a plurality of endpoints; and

a processor for interconnecting said plurality of endpoints in a conference session;

wherein the processor is configured for offsetting delivery of packets to the multimedia transformation unit to decorrelate channels at the multimedia transformation unit.

16. The apparatus of claim 15 wherein the processor is configured to transmit only one of a plurality of redundant streams to the multimedia transformation unit.

17. The apparatus of claim 16 further comprising transmitting information to the multimedia transformation unit for use in choosing said media streams for each of said plurality of endpoints.

18. The apparatus of claim 15 wherein the apparatus is a multipoint conferencing unit.

19. The apparatus of claim 15 wherein the processor is configured for identifying one of said media streams as an active speaker stream.

20. The apparatus of claim 15 further comprising transmitting said packets to the multimedia transformation unit in a plurality of redundant streams.

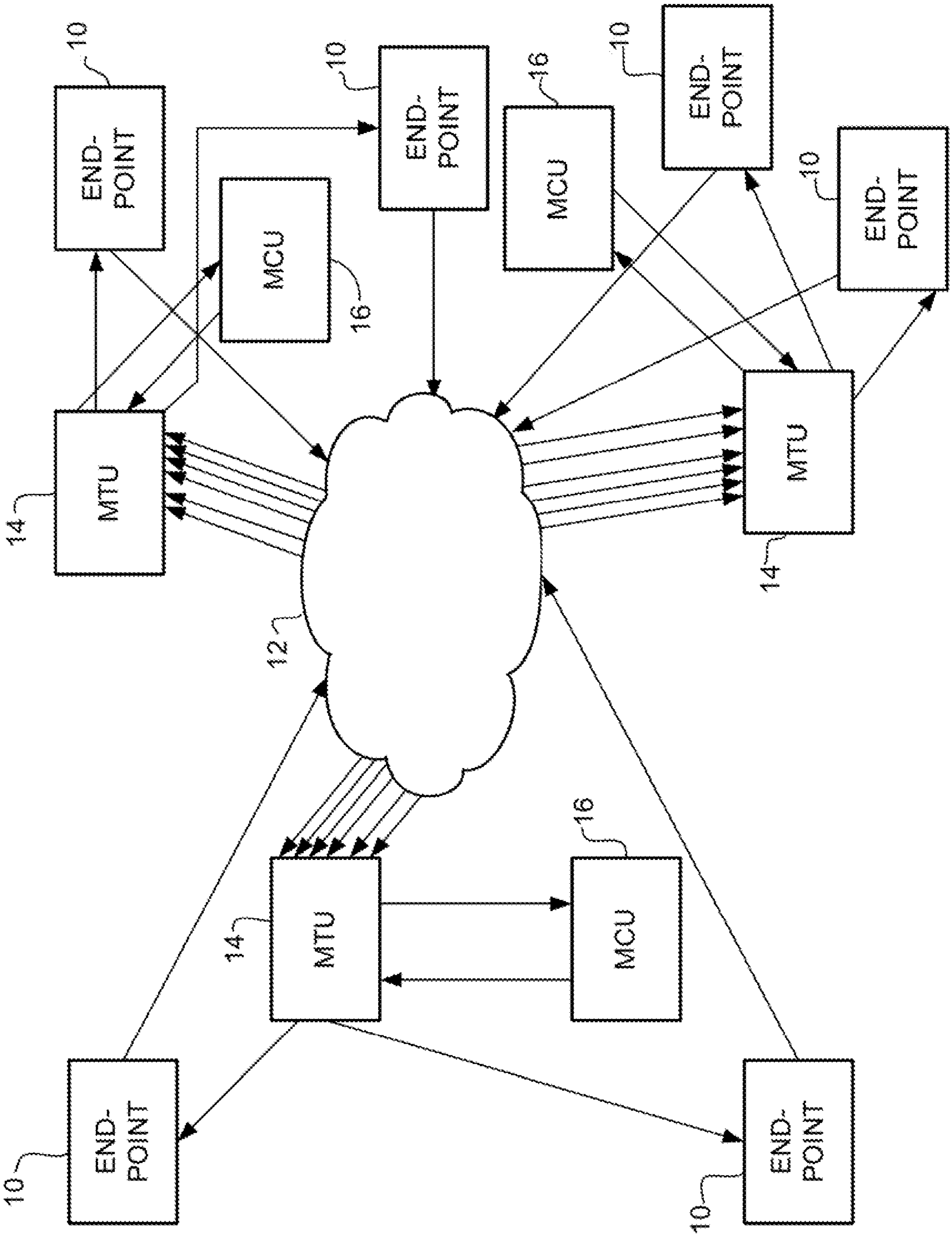

FIGURE 1

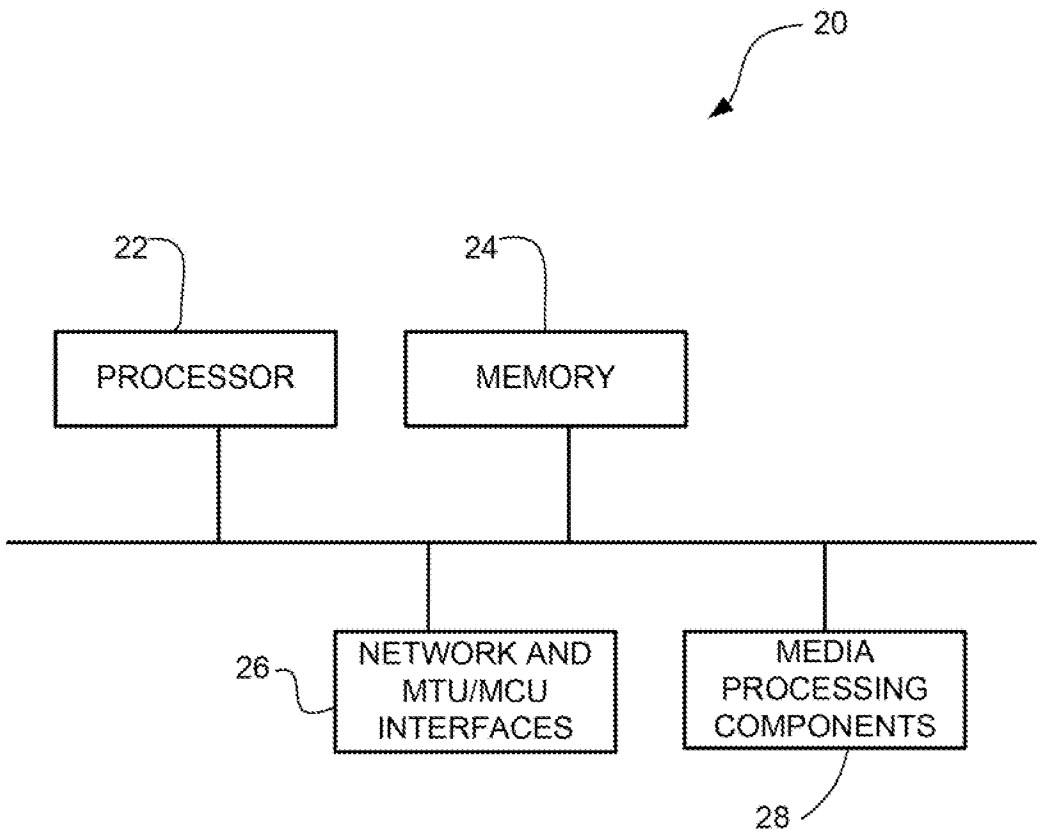
20
22
24
PROCESSOR
MEMORY
26
NETWORK AND MTU/MCU INTERFACES
MEDIA PROCESSING COMPONENTS
28

FIGURE 2

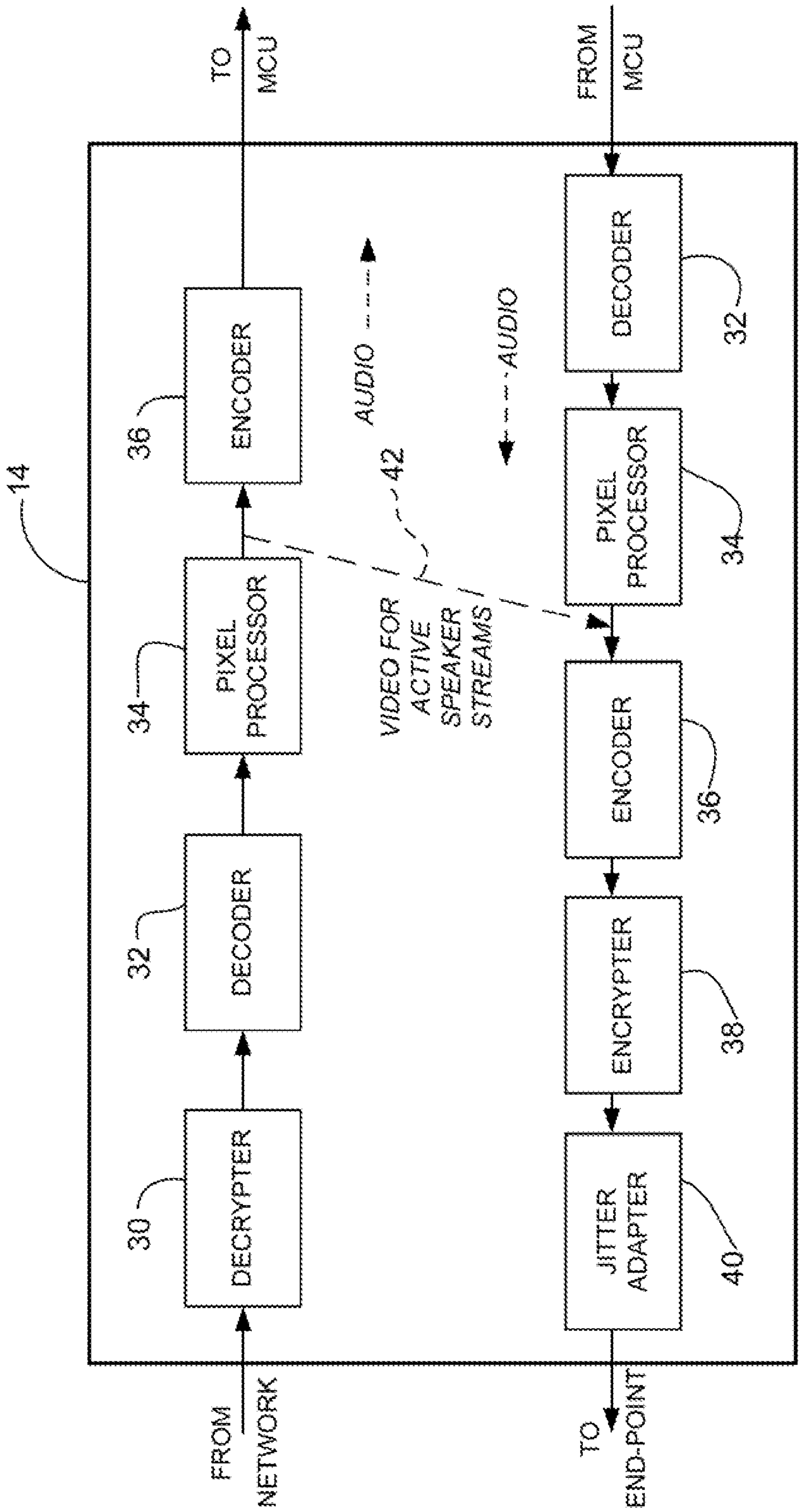
14
FROM NETWORK
DECRYPTER
30
DECODER
32
PIXEL PROCESSOR
34
ENCODER
36
TO MCU
AUDIO
AUDIO
42
VIDEO FOR ACTIVE SPEAKER STREAMS
FROM MCU
DECODER
32
PIXEL PROCESSOR
34
ENCODER
36
ENCRYPTER
38
JITTER ADAPTER
40
TO END-POINT

FIGURE 3

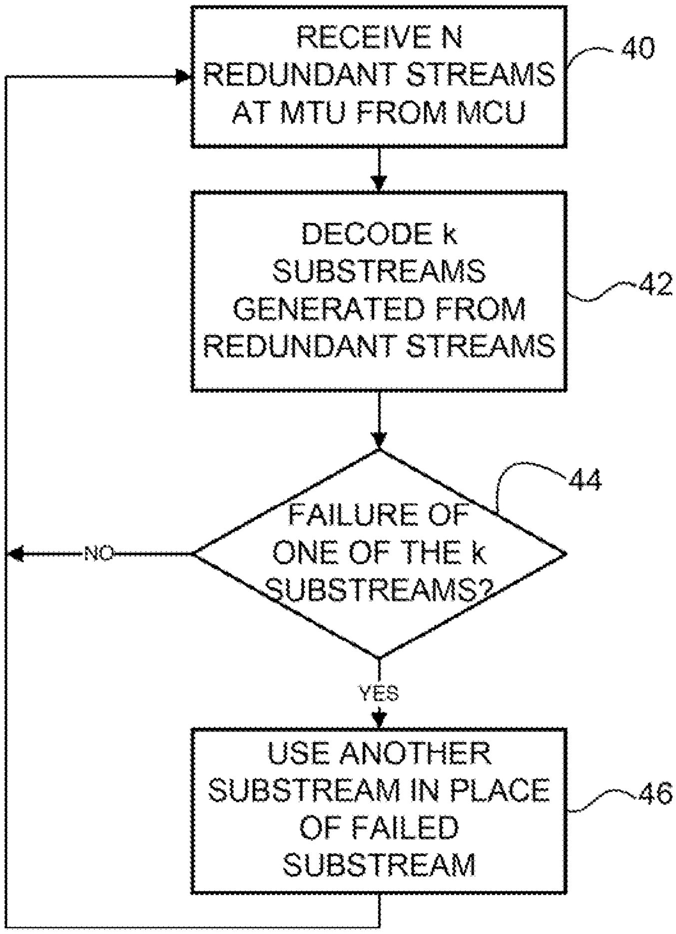
RECEIVE N REDUNDANT STREAMS AT MTU FROM MCU
40
DECODE k SUBSTREAMS GENERATED FROM REDUNDANT STREAMS
42
FAILURE OF ONE OF THE k SUBSTREAMS?
44
NO
YES
USE ANOTHER SUBSTREAM IN PLACE OF FAILED SUBSTREAM
46

FIGURE 4

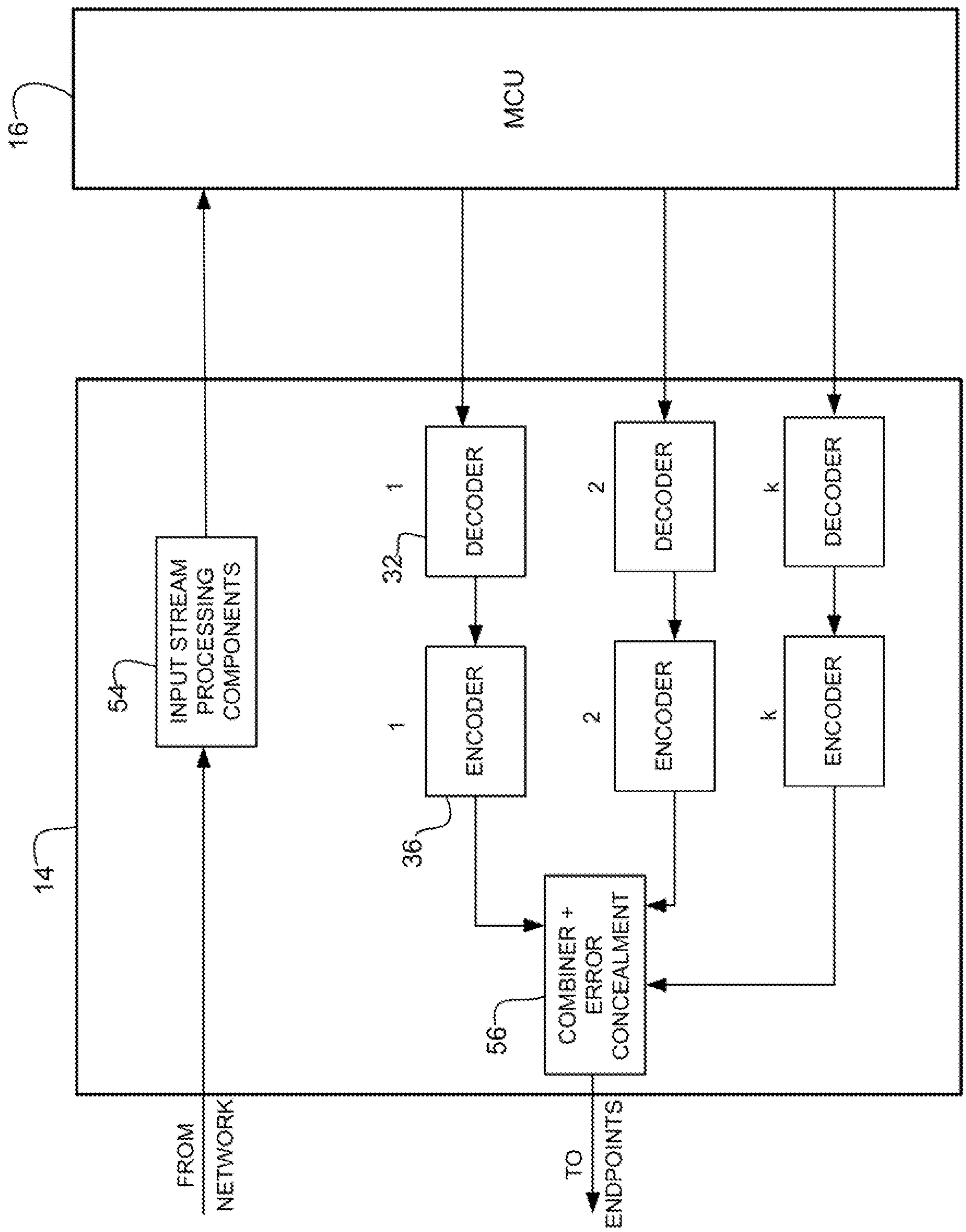
16
MCU
14
54
INPUT STREAM PROCESSING COMPONENTS
FROM NETWORK
1
DECODER
32
2
DECODER
k
DECODER
1
ENCODER
36
2
ENCODER
k
ENCODER
56
COMBINER + ERROR CONCEALMENT
TO ENDPOINTS

FIGURE 5

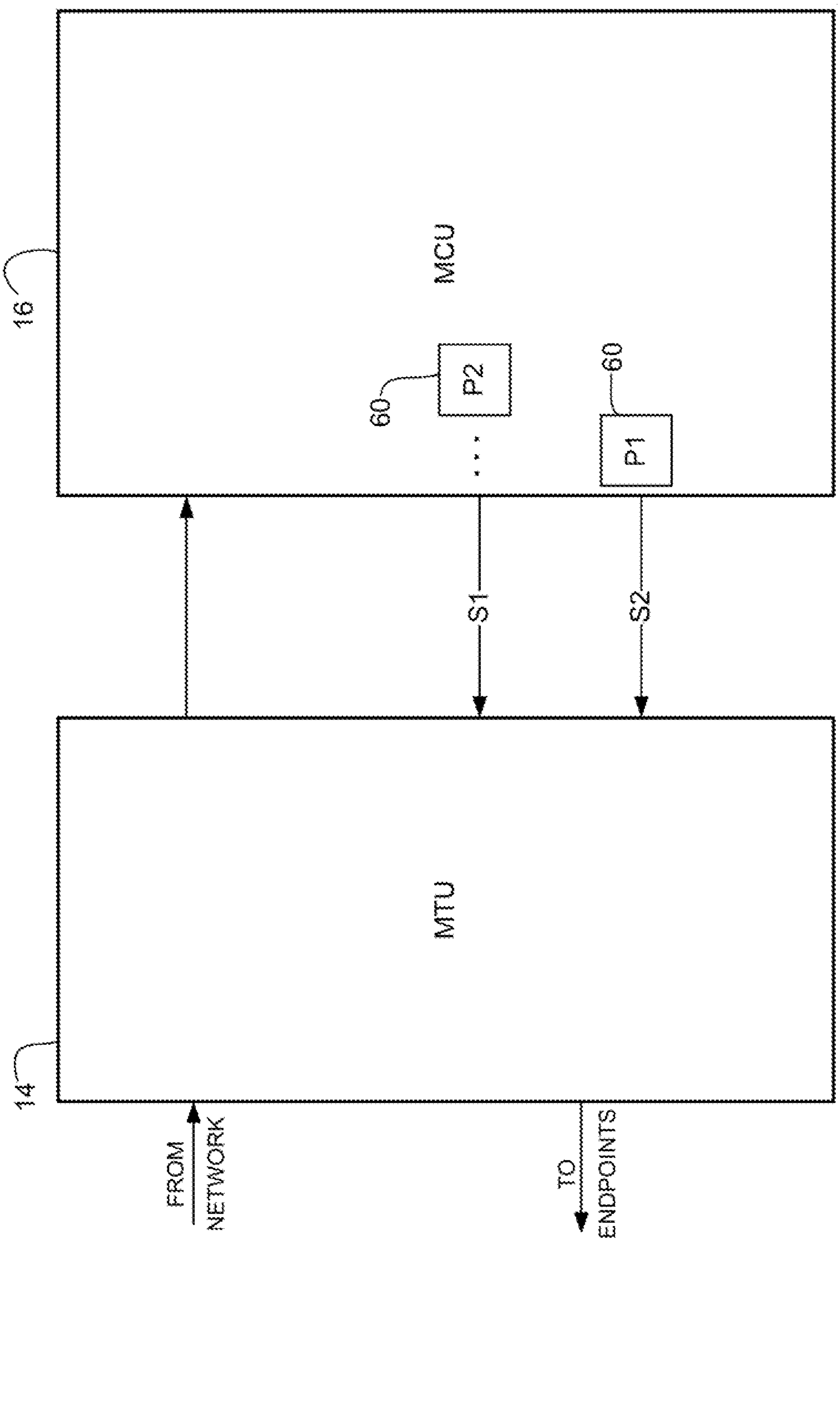

FIGURE 6

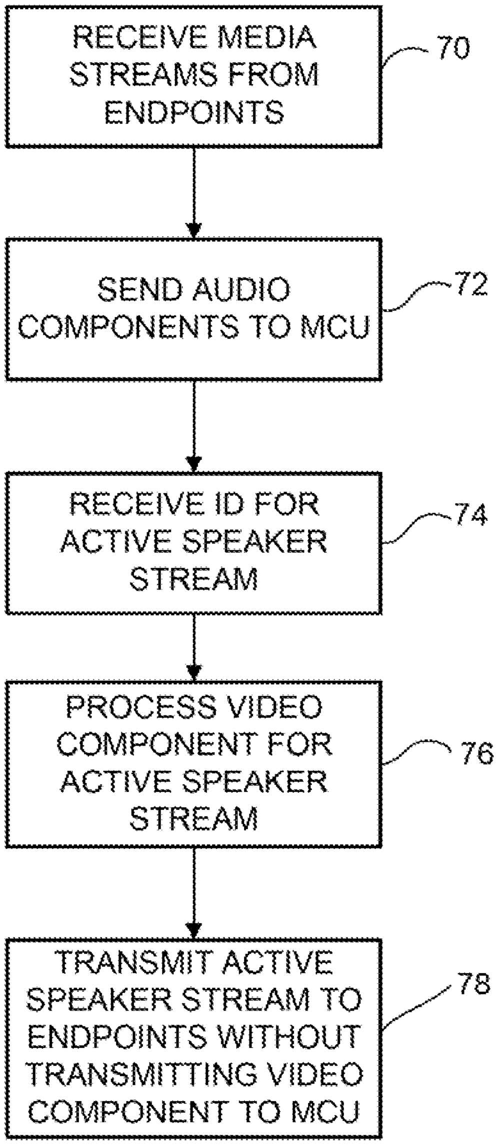
RECEIVE MEDIA STREAMS FROM ENDPOINTS
70
SEND AUDIO COMPONENTS TO MCU
72
RECEIVE ID FOR ACTIVE SPEAKER STREAM
74
PROCESS VIDEO COMPONENT FOR ACTIVE SPEAKER STREAM
76
TRANSMIT ACTIVE SPEAKER STREAM TO ENDPOINTS WITHOUT TRANSMITTING VIDEO COMPONENT TO MCU
78

FIGURE 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/031606

A. CLASSIFICATION OF SUBJECT MATTER
INV. H04N21/2343 H04N7/15 H04L12/18 H04L29/06 H04L1/00
H04L12/56 H04N21/262 H04N21/2365
ADD.

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)
H04N H04L H04M

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

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

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 2006/092269 A1 (BAIRD RANDALL [US] ET AL) 4 May 2006 (2006-05-04) paragraphs [0059], [0067] - [0072], [0076] -----	1-6 12,15-20
A	US 7 627 629 B1 (WU FANG [US] ET AL) 1 December 2009 (2009-12-01) the whole document -----	1-6
A	EP 1 039 734 A2 (SIEMENS INF & COMM NETWORKS [US] SIEMENS COMM INC [US]) 27 September 2000 (2000-09-27) paragraphs [0011] - [0012], [0021] -----	1-6
A	US 2007/156924 A1 (RAMALINGAM THIYAGESAN [US] ET AL) 5 July 2007 (2007-07-05) paragraphs [0024], [0029], [0036], [0042], [0046], [0053], [0065] ----- -/--	1-6

[X] Further documents are listed in the continuation of Box C. [X] 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	Date of mailing of the international search report
27 July 2012	07/08/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	McGrath, Simon

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2012/031606

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2008/266384 A1 (TRIPLICANE VIJAYARAGHAVAN R [US] ET AL) 30 October 2008 (2008-10-30) the whole document -----	1-6
A	US 2008/267282 A1 (KALIPATNAPU RAJAH K V R [US] ET AL) 30 October 2008 (2008-10-30) the whole document -----	1-6
X	US 6 944 123 B1 (MOON BILLY G [US]) 13 September 2005 (2005-09-13)	7,8,10, 13,14
Y	column 2, line 38 - column 11, line 3; figures 1,4,8,9 -----	9,11,12
X	KON F ET AL: "Using dynamic configuration to manage a scalable multimedia distribution system", COMPUTER COMMUNICATIONS, ELSEVIER SCIENCE PUBLISHERS BV, AMSTERDAM, NL, vol. 24, no. 1, 1 January 2001 (2001-01-01), pages 105-123, XP004227545, ISSN: 0140-3664, DOI: 10.1016/S0140-3664(00)00293-0	7,8,13, 14
Y	page 106 - page 113 -----	12
Y	US 2008/022340 A1 (HANNUKSELA MISKA [FI] ET AL) 24 January 2008 (2008-01-24) paragraphs [0005], [0034] - [0064] -----	9,11
A	US 2008/267078 A1 (FARINACCI DINO [US] ET AL) 30 October 2008 (2008-10-30) the whole document -----	7-14
A	REHA CIVANLAR M ET AL: "IP-Networked Multimedia Conferencing", IEEE SIGNAL PROCESSING MAGAZINE, IEEE SERVICE CENTER, PISCATAWAY, NJ, US, vol. 17, no. 4, 1 July 2000 (2000-07-01), pages 31-43, XP011089870, ISSN: 1053-5888 the whole document -----	7-14
Y	US 2005/204242 A1 (CHOU PHILIP A [US] ET AL CHOU PHILIP [US] ET AL) 15 September 2005 (2005-09-15) paragraphs [0041] - [0044], [0067]; figures 2,6 -----	15-20
Y	US 2009/327842 A1 (LIU HANG [US] ET AL) 31 December 2009 (2009-12-31) paragraphs [0039], [0044] -----	15-20

INTERNATIONAL SEARCH REPORT	International application No. PCT/US2012/031606

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 additional sheet

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 an 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.:

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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-6

   Methods in multipoint conferencing which identify and transmits an active speaker stream without the video component to multipoint conferencing units (MCUs).

   ---

2. claims: 7-14

   Apparatuses for receiving multiple redundant streams, identifying the "failure" of one (sub)stream and replacing the failed (sub)stream with another (sub)stream.

   ---

3. claims: 15-20

   Apparatuses which offset delivery of packets to a multimedia transformation unit to decorrelate channels.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2012/031606

Patent document cited in search report		Publication date	Patent family member(s)			Publication date
US 2006092269	A1	04-05-2006	EP	1958443	A1	20-08-2008
			US	2006092269	A1	04-05-2006
			WO	2007067236	A1	14-06-2007
US 7627629	B1	01-12-2009	NONE			
EP 1039734	A2	27-09-2000	DE	60038516	T2	31-07-2008
			EP	1039734	A2	27-09-2000
			US	6775247	B1	10-08-2004
US 2007156924	A1	05-07-2007	NONE			
US 2008266384	A1	30-10-2008	AT	497289	T	15-02-2011
			CN	101675623	A	17-03-2010
			EP	2143234	A1	13-01-2010
			US	2008266384	A1	30-10-2008
			WO	2008137373	A1	13-11-2008
US 2008267282	A1	30-10-2008	CN	101669324	A	10-03-2010
			EP	2140608	A1	06-01-2010
			US	2008267282	A1	30-10-2008
			WO	2008134301	A1	06-11-2008
US 6944123	B1	13-09-2005	US	6944123	B1	13-09-2005
			US	2005243714	A1	03-11-2005
US 2008022340	A1	24-01-2008	NONE			
US 2008267078	A1	30-10-2008	CN	101669105	A	10-03-2010
			EP	2140363	A1	06-01-2010
			US	2008267078	A1	30-10-2008
			WO	2008134292	A1	06-11-2008
US 2005204242	A1	15-09-2005	US	6996097	B1	07-02-2006
			US	2005204242	A1	15-09-2005
			US	2005249211	A1	10-11-2005
US 2009327842	A1	31-12-2009	BR	PI0621900	A2	20-12-2011
			CN	101512989	A	19-08-2009
			EP	2067318	A1	10-06-2009
			JP	2009545228	A	17-12-2009
			US	2009327842	A1	31-12-2009
			WO	2008013528	A1	31-01-2008