(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 4 287 612 A1**

(12) EUROPEAN PATENT APPLICATION

published in accordance with Art. 153(4) EPC

(43) Date of publication:
06.12.2023 Bulletin 2023/49

(21) Application number: **22848094.3**

(22) Date of filing: **17.06.2022**

(51) International Patent Classification (IPC):
H04N 7/15 *(2006.01)* ***H04N 21/262*** *(2011.01)*

(52) Cooperative Patent Classification (CPC):
H04L 12/1822; H04L 12/1827; H04L 65/403;
H04N 7/15; H04N 21/262; H04L 51/10

(86) International application number:
PCT/CN2022/099302

(87) International publication number:
WO 2023/005487 (02.02.2023 Gazette 2023/05)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(30) Priority: **30.07.2021 CN 202110875339**

(71) Applicant: **Tencent Technology (Shenzhen) Company Limited**
Shenzhen, Guangdong 518057 (CN)

(72) Inventor: **XUE, Di**
Shenzhen, Guangdong 518057 (CN)

(74) Representative: **Nederlandsch Octrooibureau**
P.O. Box 29720
2502 LS The Hague (NL)

(54) AUDIO/VIDEO CONFERENCE IMPLEMENTATION METHOD, AUDIO/VIDEO CONFERENCE SYSTEM AND RELATED APPARATUS

(57) Disclosed in the present application are an audio/video conference implementation method, an audio/video conference system and a related apparatus. The present application relates to a plurality of device groups, each of the device groups comprises a plurality of interface machines, and each interface machine is connected to a conference access terminal, such that the conference access terminal accesses a conference room by means of the interface machine. Conference access terminals are distributed to different interface machines, and load sharing is performed by means of a plurality of interface machines, so as to support a large-scale conference. There is an origin-pull device in each device group, and the origin-pull device accesses a conference room by means of a second interface machine and generates a media data stream. An origin-pull device in each device group, e.g. a first device group, receives a stream pull application from a first interface machine in the first device group, and then forwards a media data stream corresponding to first identifier information to the first interface machine, such that the first interface machine forwards the media data stream to a corresponding conference access terminal. The data delivery pressure is shared by means of an origin-pull device, which is conducive to supporting a super-large-scale conference, such that microphones or cameras can be turned on at the same time, redirection is not required, and the experience is smooth.

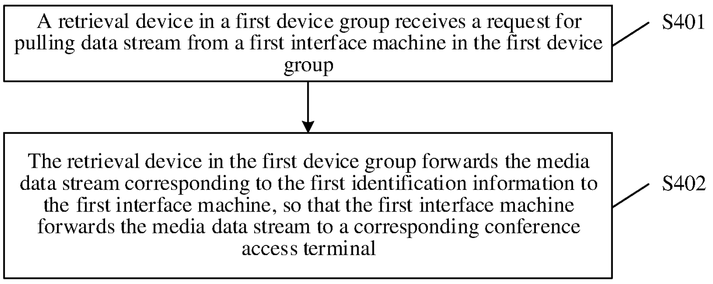

A retrieval device in a first device group receives a request for pulling data stream from a first interface machine in the first device group
S401
The retrieval device in the first device group forwards the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine forwards the media data stream to a corresponding conference access terminal
S402

FIG. 4a

EP 4 287 612 A1

Processed by Luminess, 75001 PARIS (FR)

Description

RELATED APPLICATION

[0001] This application claims priority to Chinese Patent Application No. 202110875339.6, entitled "AUDIO/VIDEO CONFERENCE IMPLEMENTATION METHOD, AUDIO/VIDEO CONFERENCE SYSTEM AND RELATED APPARATUS" filed with the China National Intellectual Property Administration on July 30, 2021, which is incorporated herein by reference in its entirety.

FIELD OF THE TECHNOLOGY

[0002] This application relates to the computer field, and in particular, to a method for implementing audio and video conference, an audio and video conference system, and a related apparatus.

BACKGROUND OF THE DISCLOSURE

[0003] With rapid development of network technologies, communication technologies, and streaming media technologies, and the increasing mobility of people to work and learn, more and more enterprises and individuals need video communication, and audio and video conference systems emerge.

[0004] In a current audio and video conference, a conference architecture based on a selective forwarding unit (SFU) is mainly used, and the conference architecture includes one server and multiple terminals. After receiving an audio and video stream (that is, a media data stream) shared by a terminal in a conference room, the server directly forwards the audio and video stream to another terminal in the conference room.

[0005] However, once the number of people entering the conference room in this conference architecture is too large, pressure on data distribution of the server is greatly increased. Therefore, in this conference architecture, the number of people attending a conference is limited, and the number of people who can enable microphones or videos at the same time is limited. A participant needs to raise hand to speak. This requires redirection, and the user experience is not smooth.

SUMMARY

[0006] To resolve the foregoing technical problem, this application provides a method for implementing audio and video conference, an audio and video conference system, and a related apparatus, which greatly reduces pressure on data distribution of a media source device such as a second interface machine. Therefore, more participants can be supported, and even millions of participants can be supported. At the same time, because the pressure on data distribution during media data stream transmission is greatly reduced, all participants can simultaneously enable microphones or videos without redirection, and the user experience is smooth.

[0007] Embodiments of this application disclose the following technical solutions:

[0008] According to a first aspect, an embodiment of this application provides a method for implementing audio and video conference, where the method includes:

[0009] receiving, by a retrieval device in a first device group, a pulling request for pulling data stream sent by a first interface machine in the first device group, the pulling request including first identification information of a media data stream, the first device group being one of multiple device groups each comprising a respective retrieval device and multiple interface machines, and each of the multiple interface machines being configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine; the first interface machine being an interface machine in the first device group; and forwarding, by the retrieval device in the first device group, the media data stream corresponding to the first identification information to the first interface machine, so as to enable forwarding the media data stream to a corresponding conference accessing terminal by the first interface machine; the media data stream being obtained by the retrieval device in the first device group from a second interface machine of a second device group, a conference accessing terminal that generates the media data stream accessing the conference room by using the second interface machine, and the second device group being a device group in the multiple device groups.

[0010] According to a second aspect, an embodiment of this application provides an apparatus for implementing audio and video conference, where the apparatus includes a receiving unit and a forwarding unit:

> the receiving unit being configured to receive a pulling request for pulling data stream sent by a first interface machine in a first device group, the pulling request including first identification information of a media data stream, the first device group being one of multiple device groups each comprising a respective retrieval device and multiple interface machines, and each of the multiple interface machines being configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine; the first interface machine being an interface machine in the first device group; and
>
> the forwarding unit being configured to forward the media data stream corresponding to the first identification information to the first interface machine, so as to enable forwarding the media data stream to a corresponding conference accessing terminal by the first interface machine; and the media data stream being obtained by the retrieval device in the first de-

vice group from a second interface machine of a second device group, a conference accessing terminal that generates the media data stream accessing the conference room by using the second interface machine, and the second device group being a device group in the multiple device groups.

[0011] According to a third aspect, an embodiment of this application provides an audio and video conference system, where the system includes a data transmission network and a room management subsystem:

the data transmission network being configured to transmit a media data stream generated in a conference room, the data transmission network including multiple device groups, and each device group including a retrieval device and multiple interface machines;

each of the multiple interface machines being configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses the conference room by using the interface machine;

a retrieval device in a first device group being configured to receive a pulling request for pulling data stream sent by a first interface machine in the first device group, the pulling request including first identification information of a media data stream, and forward the media data stream corresponding to the first identification information to the first interface machine, the media data stream being obtained by the retrieval device in the first device group from a second interface machine of a second device group, a conference accessing terminal that generates the media data stream accessing the conference room by using the second interface machine, and each of the first device group and the second device group being a device group in the multiple device groups;

the first interface machine being an interface machine in the first device group, and the first interface machine being configured to forward the media data stream to a corresponding conference accessing terminal; and

the room management subsystem including the multiple device groups and a central server, each device group including an intra-group server, and the intra-group server being configured to assist the central server in managing the conference room.

[0012] According to a fourth aspect, an embodiment of this application provides an electronic device for implementing an audio and video conference, where the electronic device includes a processor and a memory:

the memory being configured to store program code, and transmit the program code to the processor; and

the processor being configured to perform the method according to the first aspect according to instructions in the program code.

[0013] According to a fifth aspect, an embodiment of this application provides a computer readable storage medium, where the computer readable storage medium is configured to store program code, and when the program code runs on an electronic device, the electronic device performs the method according to the first aspect.

[0014] According to a sixth aspect, an embodiment of this application provides a computer program product. When the computer program product is executed, an electronic device performs the method according to the first aspect.

[0015] It can be learned from the foregoing technical solutions that this application relates to multiple device groups, each device group includes multiple interface machines, and each interface machine of the multiple interface machines is configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine. Compared with a case in which multiple conference accessing terminals access a conference room by using one server in a related technology, roles of the multiple interface machines are similar to that of the server in the related technology, and conference accessing terminals are distributed to different interface machines, and load is shared by multiple interface machines, so that more conference accessing terminals can be accessed, and a large-scale conference is supported. A retrieval device exists in each device group in the multiple device groups, and a media data stream is generated by accessing a conference room by using a second interface machine. After receiving a pulling request for pulling data stream sent by a first interface machine in a first device group, because the pulling request for pulling data stream includes first identification information of a media data stream, a retrieval device in each device group, such as the first device group, can forward the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine is configured to forward the media data stream to a corresponding conference accessing terminal. In this way, when each interface machine needs to pull a media data stream of the second interface machine, only the second interface machine needs to interact with a retrieval device in each device group, so that the retrieval device shares the pressure on data distribution for a media source device, such as the second interface machine, and interacts with another interface machine, and no interaction is required between the second interface machine and each interface machine, which greatly reduces the pressure on data distribution of the device, such as the second interface machine. Therefore, the method for implementing audio and video conference can support more participants and even millions of participants. In addition, because the pressure on data distribution during media data streaming is greatly reduced, all participants (that is, users) can simultaneously enable microphones or videos without redirection, and the user experience is smooth.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] To describe technical solutions in embodiments of this application or the related art more clearly, the following briefly introduces the accompanying drawings required for describing the embodiments or the related art. Apparently, the accompanying drawings in the following description show only some embodiments of this application, and a person of ordinary skill in the art may still derive other drawings from these accompanying drawings without creative efforts.

FIG. 1 is a diagram of an SFU-based conference architecture according to a related art.

FIG. 2 is a diagram of another SFU-based conference architecture according to related art.

FIG. 3 is a structural diagram of an audio and video conference system according to a related technology.

FIG. 4a is a flowchart of a method for implementing audio and video conference according to an embodiment of this application.

FIG. 4b is a structural diagram of a data transmission network according to an embodiment of this application.

FIG. 5 is an interface diagram of performing a video conference based on an audio and video conference system according to an embodiment of this application.

FIG. 6 is a structural diagram of a room management subsystem according to an embodiment of this application.

FIG. 7 is a diagram of an audio processing architecture based on a mixer engine according to an embodiment of this application.

FIG. 8 is an architectural diagram of interconnection with a third-party conference device according to an embodiment of this application.

FIG. 9 is a diagram of still another audio processing architecture according to an embodiment of this application.

FIG. 10 is a structural diagram of an apparatus for implementing audio and video conference according to an embodiment of this application.

FIG. 11 is a structural diagram of a terminal according to an embodiment of this application.

FIG. 12 is a structural diagram of a server according to an embodiment of this application.

DESCRIPTION OF EMBODIMENTS

[0017] The following describes the embodiments of this application with reference to the accompanying drawings.

[0018] Currently, an SFU-based conference architecture is mainly used for an audio and video conference. Referring to FIG. 1, in the figure, a black node is a server, a gray node is a terminal, multiple terminals are connected to the server, and a user A, a user B, and a user C separately access a conference room to participate in a conference by using corresponding terminals.

[0019] However, once the number of people entering the conference room in this conference architecture is too large, pressure on data distribution of the server is greatly increased. Therefore, in this conference architecture, the number of people attending a conference is limited, and the number of people who can enable microphones or videos at the same time is limited. An ordinary audience needs to hand up for speaking. This requires redirection, and the user experience is not smooth.

[0020] In addition, participants in this manner cannot simultaneously enable videos or microphones. Therefore, on a basis of the SFU architecture shown in FIG. 1, a diffusion agent is added (as shown by a node included in a dashed line box in FIG. 2), a user who does not speak is placed on the diffusion agent, and the user is redirected to a communication node (as shown by a node included in a solid line box in FIG. 2) when handing up for speaking, a number of participants in a conference is increased by using a distribution advantage of the diffusion agent. In this method, redirection is required, and the user experience is not smooth.

[0021] To resolve the foregoing technical problem, an embodiment of this application provides a method for implementing audio and video conference. In the method, when each interface machine needs to pull a media data stream of a second interface machine, only the second interface machine needs to interact with a retrieval device in each device group, so that the retrieval device shares the pressure on data distribution for a second interface machine and interacts with another interface machine, and no interaction is required between the second interface machine and each interface machine, which greatly reduces the pressure on data distribution of a media source device such as the second interface machine. Therefore, the method for implementing audio and video conference can support more participants and even millions of participants. In addition, because the pressure on data distribution during media data streaming is greatly reduced, all participants can simultaneously enable microphones or videos without redirection, and the user experience is smooth.

[0022] The method for implementing audio and video conference provided in this embodiment of this application may be applied to various audio and video conference scenarios, in particular, a super-large-scale video and audio conference used for a single meeting of millions of people.

[0023] The following describes an architecture of an audio and video conference system according to an embodiment of this application. When participants enter a conference room to participate in a conference based on an audio and video conference system, the participants mainly include a guest who needs to enable audio and video, and an ordinary audience. When the guest enables audio and video, a media data stream is generated, and the media data stream is transmitted to a conference accessing terminal corresponding to the general audience, and a conference accessing terminal corresponding to another guest. The conference accessing terminal is a terminal used for accessing the conference room, and the conference accessing terminal may be a user terminal. Therefore, the audio and video conference mainly lies in that a participant enables transmission of a media data stream when the audio and video are enabled, for example, a video or a microphone is enabled. When transmission of the media data stream is not limited by a number of participants in a conference room, the audio and video conference system may support a large-scale conference. Therefore, this embodiment of this application focuses on improvement of a data transmission network architecture.

[0024] In addition, because each conference corresponds to one conference room for storing room information, a participant list, and maintaining a user status (enabling or disabling audio and video, i.e., enabling or disabling the microphone, enabling or disabling the video, mute, or entering or exiting the room). In the case of a super-large-scale conference, everyone keeps enabling or disabling the video or entering or exiting the room, resulting in a large number of notification messages that reflect status changes. Pulling and synchronization of a participant list of the super-large audio and video conference will also exert great pressure on a room management subsystem. Therefore, improvements to the room management subsystem in this embodiment of this application also enable the audio and video conference system to support a super-large-scale audio and video conference.

[0025] Therefore, as shown in FIG. 3, the architecture of the audio and video conference system in this embodiment mainly includes a data transmission network 301 and a room management subsystem 302. The data transmission network 301 is configured to transmit a media data stream generated in a conference room, and the room management subsystem 302 is configured to manage the conference room.

[0026] In this embodiment of this application, a method for implementing audio and video conference is first described from a perspective of the data transmission network 301 for transmitting a media data stream. Referring to FIG. 4a, the method includes the following steps:

[0027] S401. A retrieval device in a first device group receives a pulling request for pulling data stream sent by a first interface machine in the first device group.

[0028] The data transmission network 301 may be shown in FIG. 4b, and includes multiple device groups 401, for example, shown in each dashed line box in FIG. 4b. Each device group 401 includes multiple interface machines, and each interface machine is, for example, a node represented by a circle in the device group 401 in FIG. 4b. Each of the multiple interface machines is configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine, so that an participant corresponding to the conference accessing terminal can enter the conference room to participate in the conference. The device group may be represented by SET, that is, multiple interface machines are divided by SET, and the interface machine may be a server. A retrieval device exists in each of the plurality of device groups 401.

[0029] The first device group may be any device group in the multiple device groups, and the first interface machine may be any interface machine in the first device group.

[0030] When a media data stream needs to be pulled by an interface machine (for example, the first interface machine) in a device group (for example, the first device group), the first interface machine may send a pulling request for pulling data stream to the retrieval device in the first device group. The pulling request for pulling data stream may include first identification information of a media data stream, and the first identification information identifies a media data stream of media data subscribed by the first interface machine, so that the retrieval device knows which media data stream to pull. The first interface machine may be an interface machine connected to a conference accessing terminal corresponding to a user who subscribes to media data. That is, if a user on the first interface machine subscribes to media data of an interface machine, the first interface machine is triggered to send a pulling request for pulling data stream to the retrieval device in the first device group.

[0031] S402. The retrieval device in the first device group forwards the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine forwards the media data stream to a corresponding conference accessing terminal.

[0032] After receiving the pulling request for pulling data stream, the retrieval device in the first device group forwards the media data stream corresponding to the first identification information to the first interface machine, to prompt the first interface machine to forward the media data stream to the corresponding conference accessing terminal. In an example, the retrieval device, while forwarding the media data stream to the first interface machine, prompts the first interface machine to forward the media data stream to the corresponding conference accessing terminal. For example, the retrieval device may instruct the first interface machine to forward the media data stream to the corresponding conference accessing terminal. Then, the first interface machine forwards the media data stream to the corresponding conference accessing terminal. The conference accessing terminal that generates the media data stream accesses the conference room by using a second interface machine, and the second device group is any device group in the multiple device groups.

[0033] The second interface machine may be an interface machine connected to a conference accessing terminal corresponding to a user whose audio and video are enabled. For example, when a user, for example, a second user, enables video on a conference accessing terminal of the user, the conference accessing terminal corresponding to the second user accesses the conference by using the second interface machine, and the second interface machine is located in the second device group. If any user participating in the conference, for example, a first user wishes to watch the video of the second user (for example, subscribed to the video of the second user), the first interface machine connected to the conference accessing terminal corresponding to the first user may find a retrieval device in a device group in which the first interface machine is located, for example, the first device group, and send a pulling request for pulling data stream to the retrieval device in the first device group, so that the retrieval device forwards, according to the pulling request for pulling data stream, a media data stream obtained from the second interface machine to the first interface machine, and the first interface machine forwards the media data stream to the corresponding conference accessing terminal. The user in this embodiment of this application may be a participant.

[0034] The media data stream on the retrieval device is pulled from the second interface machine, and the pulled media data stream may be stored locally on the retrieval device. That is, generally, when receiving, for the first time, the application for pulling the media data stream, the retrieval device does not save the media data stream locally, and therefore needs to pull the media data stream from the second interface. If another interface machine has previously applied for the media data stream, the media data stream is locally saved, and the media data stream does not need to be pulled from the second interface machine.

[0035] Therefore, in this embodiment, different methods for pulling media data stream may be used according to whether the retrieval device in the first device group locally finds the media data stream.

[0036] In a possible implementation, a specific method in which the retrieval device in the first device group forwards the media data stream corresponding to the first identification information to the first interface machine may be that the retrieval device in the first device group receives the pulling request for pulling data stream sent by the first interface machine, searches for the media data stream of the corresponding second interface machine according to the first identification information, if the media data stream of the second interface machine is not found, forwards the pulling request for pulling data stream to the second interface machine, receives the media data stream forwarded by the second interface machine according to the first identification information in the pulling request for pulling data stream, and forwards the media data stream to the first interface machine.

[0037] It may be understood that if the pulling request for pulling data stream is sent when the first user subscribes to the video of the second user, after receiving the pulling request for pulling data stream, the second interface machine may locally record that the retrieval device subscribes to the second interface machine, and then the second interface machine sends the media data stream to the retrieval device. The retrieval device also stores the received media data stream locally, so that when another interface machine requests the media data stream, the retrieval device does not need to interact with the second interface machine, and may directly return the media data stream to the first interface machine.

[0038] For example, as shown in FIG. 4b, node A may be used as the second interface machine, and the second interface machine is used as an interface machine on which a guest is located, that is, a user corresponding to a conference accessing terminal on the second interface machine may enable the video and has an uplink video data stream. In this case, the video data stream is used as a media data stream. Node B may be used as the first interface machine, and node D may be used as a retrieval device of a device group such as the first device group in which node B is located. A user on node B subscribes to the uplink video data stream of node A. Node B first finds node D in the first device group, and sends a pulling request for pulling data stream to node D. After receiving the pulling request for pulling data stream sent by node B, node D first locally searches for the media data stream of node A. If not found, node D initiates a pulling request for pulling data stream to node A. After receiving the pulling request for pulling data stream sent by node D, node A locally records that node D subscribes to node A, and then forwards the video data stream to node D. Node D receives the video data stream and forwards the video data stream to node B. Node B finally sends the video data stream to a corresponding conference accessing terminal.

[0039] When the retrieval device in the first device group searches for the media data stream of the second interface machine according to the first identification information, if the media data stream of the second interface machine is found, the retrieval device directly forwards the found media data stream to the first interface machine. In this method, pull efficiency of a media data stream can be improved, and pressure on data distribution of a media source device such as the second interface machine is further reduced.

[0040] For example, as shown in FIG. 4b, node A may be used as the second interface machine, and the second interface machine is used as an interface machine on which a guest is located, that is, a user corresponding to a conference accessing terminal on the second interface machine may enable the video and has an uplink video data stream. In this case, the video data stream is used as a media data stream. Node C may be used as the first interface machine, and node D may be used as a retrieval device of a device group such as the first device group in which node C is located. A user on node C subscribes to the uplink video data stream of node A. Node C first finds node D in the first device group, and sends a pulling request for pulling data stream to node D. Node B and node C in the same device group find the same node D by using a consistent hash or another method, and send a pulling request for pulling data stream to node D. Node D finds that the video data stream already exists locally (previously applied for by node B). Node D directly forwards the found video data to node C, and node C receives the forwarded video data stream from node D, and finally forwards the video data stream to a corresponding conference accessing terminal.

[0041] It may be understood that the retrieval device in the first device group is an interface machine that needs to apply for a media data stream, for example, the first interface machine, and is determined according to a preset rule, and the preset rule causes the same media data stream to be routed, by using the same retrieval device, to an interface machine in a device group in which the retrieval device is located. For example, node B and node C in the same device group both find the same node D by using a consistent hash or another method. This facilitates unified distribution management on the same media data stream in the same device group.

[0042] The first device group and the second device group may be located in the same local area network (that is, an internal network), or may be located in different local areas (that is, an external network, including a domestic network and an external network of a network abroad). If the first device group and the second device group are located in the same local area network, the retrieval device in the first device group directly forwards the pulling request for pulling data stream to the second interface machine, and the second interface machine directly sends the media data stream to the retrieval device. If the first device group and the second device group are located in different local area networks, the data transmission network 301 further includes a transit agent device 402, for example, a device shown as 402 in FIG. 4b. The retrieval device in the first device group needs to forward the pulling request for pulling data stream to the second interface machine by using the transit agent device 402. Correspondingly, the second interface machine forwards the media data stream to the retrieval device in the first device group by using the transit agent device 402.

[0043] For example, as shown in FIG. 4b, node A may be used as the second interface machine, and the second interface machine is used as an interface machine on which a guest is located, that is, a user corresponding to a conference accessing terminal on the second interface machine may enable the video and has an uplink video data stream. In this case, the video data stream is used as a media data stream. Node G may be used as the first interface machine, and node F may be used as a retrieval device of a device group such as the first device group in which node G is located. In this case, the first device group in which node G is located and the second device group in which node A is located are not in the same local area network, that is, node G is an edge node. A user on node G subscribes to the video data stream upstream of node A (the video data stream is used as a media data stream), and a path of obtaining the video data stream is relatively long. Node G first finds node F in the first device group, and sends a pulling request for pulling data stream to node F. After receiving the pulling request for pulling data stream sent by node G, node F first locally searches for the video data stream of node A. If not found, node F initiates a pulling request for pulling data stream to node A by using the transit agent device 402 (for example, node E in FIG. 4b). Node A receives the pulling request for pulling data stream sent by node F, locally records that node F subscribes to node A, and then forwards the video data stream to node F by using node E, and then node E forwards the video data stream to node G. Finally, node G forwards the video data stream to a corresponding conference accessing terminal.

[0044] A process in which node H pulls the video data stream of node A in FIG. 4b is similar to that of node C, but node E also needs to be used in the middle. Details are not described herein again. In this process, search of the transit agent device 402, for example, node E, may be preconfigured, or may be dynamically allocated, which is not limited herein. Structurally, in this embodiment of this application, interface machines used for accessing an audio and video conference are divided by device group (SET). When an interface machine in a device group needs to pull a media data stream, a retrieval device may be determined in the device group in which the interface machine is located, and then the retrieval device initiates a pulling request for pulling data stream to an interface machine (that is, a media source device) that generates a media data stream. This design greatly reduces pressure on data distribution of the media source device, and greatly increases a number of participants in a conference room. For example, it is assumed that each SET has 100 interface machines, and the whole audio and video conference system has 100 SETs. Each 8-core 16G interface machine can support forwarding of

100 media data streams. If only one guest exists in a conference, and the rest of the conference are audiences, the maximum size of the conference is 100 (machine)* 100 (SET)* 100 (distribution) =1,000,000 people. If there is more than one guest, different interface machines of different SETs will be distributed based on the principle of proximity access. An uplink forwarding load will also be distributed, and a conference with millions of people will be easily supported.

[0045] It can be learned from the foregoing technical solutions that this application relates to multiple device groups, each device group includes multiple interface machines, and each interface machine of the multiple interface machines is configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine. Compared with a case in which multiple conference accessing terminals access a conference room by using one server in a related technology, roles of the multiple interface machines are similar to that of the server in the related technology, and conference accessing terminals are distributed to different interface machines, and load is shared by multiple interface machines, so that more conference accessing terminals can be accessed, and a large-scale conference is supported. A retrieval device exists in each device group in the multiple device groups, and a media data stream is generated by accessing a conference room by using a second interface machine. After receiving a pulling request for pulling data stream sent by a first interface machine in a first device group, because the pulling request for pulling data stream includes first identification information of a media data stream, a retrieval device in each device group, such as the first device group, can forward the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine is configured to forward the media data stream to a corresponding conference accessing terminal. In this way, when each interface machine needs to pull a media data stream of the second interface machine, only the second interface machine needs to interact with a retrieval device in each device group, so that the retrieval device shares pressure on data distribution for a media source device, such as the second interface machine, and interacts with another interface machine, and no interaction is required between the second interface machine and each interface machine, which greatly reduces pressure on data distribution of the device, such as the second interface machine. Therefore, the method for implementing audio and video conference can support more participants and even millions of participants. In addition, because pressure on data distribution during media data streaming is greatly reduced, all participants (that is, users) can simultaneously enable microphones or videos without redirection, and the user experience is smooth.

[0046] Compared with a conference architecture based on a multipoint control unit (MCU), operation and maintenance are simple, and are not limited by performance and hardware price. Therefore, a number of actual supported people is greatly increased, and a super-large-scale conference is implemented.

[0047] For an interface diagram of a video conference performed based on the method for implementing audio and video conference and the audio and video conference system provided in this embodiment of this application, references may be made to FIG. 5. An image of a participant who enables the video, a participant list, an icon of a related function key (for example, a microphone icon, a camera icon, an administrator icon, and an icon indicating leaving a conference) may be displayed on the interface, and the icon of the microphone may be displayed with a slash for a participant who does not enable the microphone. In the audio and video conference system provided in this embodiment of this application, a super-large-scale conference is implemented. For example, the maximum number of people entering the conference may be 50000, that is, 50000 management members.

[0048] Next, in this embodiment of this application, the method for implementing audio and video conference is described from a perspective of the room management subsystem 302 in managing a conference room.

[0049] Referring to FIG. 6, the room management subsystem 302 may include the foregoing multiple device groups 601 and a central server (RoomSvc) 602. The central server 601 herein also includes the foregoing multiple interface machines, which is similar to the device group shown in FIG. 4b. In addition, each device group 601 further includes an intra-group server (RoomSvcInSet), and the intra-group server is configured to assist the central server 602 in managing a conference room. The intra-group server is shown in a circle in each device group 601 in FIG. 6, for example, node B, node C, node D, node E, node F, and node G in FIG. 6. The central server 602 is, for example, node A in FIG. 6.

[0050] In this embodiment, an intra-group server and a retrieval device in the same device group may be the same device, or may be different devices. This is not limited to this embodiment.

[0051] Managing the conference room mainly includes managing room information of the conference room, a notification message of a status change, and pull of a participant list. The room information includes, for example, a name of the conference room, a creation time, a participant, and a conference time.

[0052] In this embodiment of this application, a room management system is divided into a central server and an intra-group server. There is only one central server, and the intra-group server and the interface machine are deployed by SET, so that synchronization of a notification message and pull of a participant list can be realized by using the intra-group server to release management pressure of the central server, thereby implementing a super-large-scale conference.

[0053] In the device group, for the same conference room, all interface machines register the same RoomSvcInSet by using a certain algorithm, and maintain a long connection through a heartbeat before a media data stream can be pulled in the conference room. When there is a notification message in the conference room to be synchronized (such as a status change of a user: enter the conference room to attend the conference, exit the conference room, enable or disable the microphone, enable or disable the video), or the participant list needs to be pulled, those are done through this RoomSvcInSet.

[0054] First, a registration flow of entering a conference room to attend the conference is described. When any interface machine such as a third interface machine in any device group such as a third device group requests to enter the conference room, an intra-group server in the third device group receives a registration request sent by the third interface machine, where the registration request may include a room identifier used for indicating a conference room that needs to be entered. If the intra-group server in the third device group does not locally find information about the conference room, the intra-group server sends the registration request to the central server. The central server records second identification information of the intra-group server in the third device group to a first conference registration list, and returns registration success information to the intra-group server in the third device group. The intra-group server in the third device group records third identification information of the third interface machine into second conference registration list, and returns registration success information to the third interface machine, to complete registration of the third interface machine. Both the second identification information and the third identification information are identification information, and are unique identifiers of corresponding devices. The identification information may be, for example, an address. That is, the second identification information may be an address of the intra-group server, and the third identification information may be an address of the third interface machine. The third device group is any device group in multiple device groups, the third interface machine is any interface machine in the third device group, the third device group may be the first device group or the second device group in the foregoing embodiment, and the third interface machine may be the first interface machine or the second interface machine in the foregoing embodiment.

[0055] For example, as shown in FIG. 6, node a may be used as the third interface machine, and the third interface machine is an interface machine that wants to register for a conference, that is, a user corresponding to a conference accessing terminal on the third interface machine wants to enter a conference room to attend the conference. Node A may be used as the central server, and node B may be used as the intra-group server of a device group in which the third interface machine is located, for example, the third device group. Node B may be determined by node a according to a rule (consistent hash or allocation scheduling). Node a sends a registration request to node B. After receiving the registration request from node a, node B searches for RoomSvc corresponding to the conference, that is, node A. Node B sends a registration request to node A. Node A receives the registration request sent by node B, records an address of node B in the registration list of the conference (the first conference registration list saved on node A), and returns registration success information to node B. After receiving the registration success information returned by node A, node B records an address of node a in the registration list of the conference (the second conference registration list saved in node B), and returns registration success information to node a.

[0056] It may be understood that in this embodiment, after the registration succeeds, the intra-group server and the central server, the intra-group server and the registered interface machine, such as the third interface machine, can maintain a long connection, to facilitate subsequent interaction between the two. For example, node B may also periodically send a heartbeat to node A, and keep a long connection channel with node A. After receiving the registration success information returned by node B, node a periodically sends a heartbeat, and keeps a long connection channel with node B.

[0057] If the intra-group server in the third device group locally searches for information about the conference room according to the registration request, and locally finds the information about the conference room, the intra-group server may record the third identification information of the third interface machine in the second conference registration list, and return registration success information to the third interface machine.

[0058] For example, as shown in FIG. 6, node b may be used as the third interface machine, and the third interface machine is an interface machine that wants to register for a conference, that is, a user corresponding to a conference accessing terminal on the third interface machine wants to enter a conference room to attend the conference. Node A may be used as the central server, and node B may be used as the intra-group server of a device group in which the third interface machine is located, for example, the third device group. Node B may be determined by node b according to a rule (consistent hash or allocation scheduling). Node b sends a registration request to node B. Node b receives the registration request from node b, finds that information about the conference room exists locally, records an address of node b in the registration list of the conference (the second conference registration list saved on node B), and returns registration success information to node b. After receiving the registration success information returned by node B, node b periodically sends a heartbeat and keeps a long connection channel with node B.

[0059] Similarly, when a user located on an edge node (that is, the device group and the central server are not in the same local area network, and the interface machine in the device group and the central server are not in the same local area network) registers in the same conference room, the registration process needs to be forwarded through a transit agent device such as node O, node P, and node Q in FIG. 6. The rest of the process is similar.

[0060] When a status change occurs in a conference accessing terminal that accesses a conference room, for example, when the user enables or disables audio and video, the user needs to notify another user in the conference room of the status change of enabling or disabling the audio and video. For example, a conference accessing terminal whose status changes is located in the third interface machine in the third device group. In this case, the intra-group server in the third device group receives a notification message sent by the third interface, and sends the notification message to the central server. The central server sends the notification message to an intra-group server of each device group, and the intra-group server of each device group sends the notification message to a successfully registered interface machine.

[0061] Still using FIG. 6 as an example, node a may be used as the third interface machine, node a is an interface machine in which a guest is located, node A may be used as the central server, and node B may be used as the intra-group server in the device group such as the third device group in which the third interface machine is located. If the guest on node a enables the video, that is, a status change occurs, node a sends a notification message (msg-a-openvideo) indicating the status change to node B, and node B sends the notification message to node A. Node A traverses a local registration list (the first conference registration list stored in node A), and sends the notification message to all intra-group servers, for example, node B, node C, node D, node E, node F, and node G in FIG. 6. Node B, node C, node D, node E, node F, and node G receive the notification message, traverse the local registration list (each serving as the second conference registration list stored in the intra-group server), and send the notification message to all interface machines, for example, node a, node b, node c, node d, node e, node f, node g, node o, node p, node q, node r, and node s in FIG. 6.

[0062] A participant list is maintained in an audio and video conference, and the participant list includes all participants. In the super-large-scale audio and video conference implemented by the audio and video conference system provided in this embodiment of this application, there are a large number of participants, and a pull by page manner may be used for updating the local participant list. When an interface machine such as the third interface machine in a device group, for example, the third device group, requests to pull the participant list, the intra-group server in the third device group receives a list pull request from the third interface machine, and the intra-group server in the third device group locally searches for the participant list. If the intra-group server in the third device group does not locally find the participant list, the intra-group server sends the list pull request to the central server. The intra-group server in the third device group receives the participant list sent by the central server, and sends the participant list to the third interface machine.

[0063] For example, as shown in FIG. 6, if node a is used as the third interface machine, a user on node a requests the participant list from node B, and node B searches for the list in a local cache. If not found, node B initiates a list pull request to node A. After receiving the list pull request, node A returns the corresponding participant list to node B. After caching the participant list locally, node B returns the participant list to node a. The participant list may be a full participant list, or may be a page-by-page participant list such as Page1.

[0064] When the intra-group server in the third device group locally searches for the participant list, if the intra-group server in the third device group locally finds the participant list, and the participant list does not expire, the intra-group server returns the participant list to the third interface machine.

[0065] In the previous example, if node b shown in FIG. 6 is used as the third interface machine, if another node already requested the foregoing participant list Page1 before node b, in this case, the participant list already exists on node B, and therefore the participant list does not need to be requested from node A. Therefore, node a sends a list pull request to node B. Node B finds Page1 in a local cache and Page1 does not expire, and node B directly returns Page1 to node b.

[0066] Next, a disaster recovery process is described. Because all information of RoomSvc comes from RoomSvcInSet and the interface machine, automatic reconstruction can be conveniently performed through RoomSvcInSet after RoomSvc is down. Similarly, all data of RoomSvcInSet comes from the interface machine. When RoomSvcInSet is down, automatic reconstruction can be performed through the interface machine in the device group (SET).

[0067] According to the method for implementing audio and video conference provided in this embodiment of this application, management pressure of the central server is relieved by using the intra-group server, to implement a super-large-scale conference.

[0068] An SFU-based audio and video conference provided by a related technology requires audio routing based on energy values when transmitting media data streams generated in conference rooms, such as audio data streams. In this method, a delay generated in a long-distance call greatly affects experience. When there are a relatively large number of participants, a lot of bandwidth is wasted, and a network performance loss of a conference accessing terminal is increased. In addition, in a scenario that requires interactive voice response (IVR) and precise recording in a conference, precise control is not easy to be performed under SFU transformation (for example, determining of start and end, by using signaling or by using a media data stream). In addition, in the case of voice recording, some industries with strict requirements cannot accept the case where more or fewer words are recorded in the recording process. If there is no centralized audio processing service, it is easy to record the content that is not to be recorded or not to record the content that needs to be recorded. These cases are absolutely unacceptable to more demanding users.

[0069] Therefore, as shown in FIG. 3, the audio and video conference system provided in this embodiment of this application further includes a mixer engine 303. The mixer engine 303 is located between the media transmission network 301 and a public switched telephone network (PSTN) 305, and serves as a bridge for communication between the two, thereby greatly improving conference experience in a super-large-scale room. The mixer engine 303 connects to the public switched telephone network 305 by using a protocol such as the Session Initiation Protocol (SIP) or the Real-Time Transport Protocol (RTP).

[0070] The mixer engine 303 includes a mixer and a selector, and each device group has a corresponding mixer engine. When a media data stream is an audio data stream, the selector in the mixer engine sends acquired multiple audio data streams to a mixer located in the same device group for sending, and the multiple audio data streams respectively come from fourth interface machines in different device groups. The mixer selects a target audio data stream from the multiple audio data streams, and sends a selection notification to a target interface machine connected to a conference accessing terminal that generates the target audio data stream. The target interface machine sends the target audio data stream to the mixer, the mixer mixes the target audio data stream, and forwards the mixed audio data stream to a selector corresponding to the fourth interface machine, where the fourth interface machine includes the target interface machine. The selector corresponding to the fourth interface machine sends the mixed audio data stream to the fourth interface machine, and the fourth interface machine forwards the mixed audio data stream to another interface machine in a device group in which the fourth interface machine is located.

[0071] The mixer and the selector are deployed by SET. To minimize the delay, the mixer and the selector of a SET are deployed in the same available area (the transmission delay within the available area can be controlled at 2 ms). In addition, audio data streams of the same conference will be processed in one SET, and eventually will be mixed in the same mixer.

[0072] It is assumed that FIG. 7 represents a real conference, where three participants enable the microphone. Interface machines to which the three participants belong are node a, node f, and node g, and all other interface machines are audiences. To simplify the example diagram, it is assumed that two audio data streams are selected for each conference (generally, four to six audio data streams are selected for a real conference). In this case, a processing procedure of the audio data stream is as follows:

[0073] Node a, node f, and node g request a selector from the scheduling system in FIG. 6 (to minimize the delay, the scheduling system allocates selectors in the same available area to audio data streams in the same conference room), and respectively obtain selectors represented by node B, node C, and node D in FIG. 7. Node a, node f, and node g forward the audio data streams to respective corresponding nodes B, C, and D. Node B, node C, and node D cache the uplink audio data streams, and report energy values to the mixer. The mixer sorts the energy values, selects two audio data streams from node f and node g as target audio data streams, and then notifies node C and node D, where node C and node D are used as target interface machines. After receiving the selection notification, node C and node D forward the corresponding target audio data streams to the mixer (the two steps have a delay within 5ms). The mixer mixes the received target audio data streams, and then forwards the mixed audio data stream to node B, node C, and node D. Node B, node C, and node D send the mixed audio data stream to node a, node f, and node g. Node a, node f, and node g then perform SET internal forwarding.

[0074] The stream to be sent to the guest is different from the stream to the general audience. The mixed audio data stream to the guest needs to rule out the audio data stream from the guest. Therefore, the selected interface machine receives two audio data streams. One audio data stream is forwarded to the speaker, and the other audio data stream is forwarded to the audience in the SET.

[0075] The mixer engine 303 is a real-time media processing system, because audio data streams processed by the mixer engine 303 need to be forwarded to all participants in the conference room in real time. In many cases, an asynchronous stream processing system may be further required, that is, after the system obtains the audio data stream, the system does not need to send the audio data stream back to a participant in a conference room, but uses it for another purpose (for example, recording, porn detecting, and live streaming push). Therefore, the system generally does not need to be used in a real time. Based on this, referring to FIG. 3, the audio and video conference system provided in this embodiment of this application further includes a bypass media processing system 304. The bypass media processing system 304 is configured to pull, by using a robot terminal, a media data stream generated in a conference room, and process the media data stream. Referring to FIG. 3, the robot terminal includes a live stream robot, a record robot, and an MRA robot.

[0076] In this embodiment, a robot terminal is used for entering a conference room to pull a stream to implement a bypass media processing function. Recording is used as an example. A user service background starts a recording task in an application programming interface (API) manner, and a recording task system instantiates a record robot to simulate a participant to enter a conference. The record robot pulls the media data stream in the conference to the local place, mixes and transcodes the stream, and records it. After the conference ends, the record robot uploads a recorded file to specified storage (or periodically uploads segments of the recorded file).

[0077] The solutions of porn detecting and live streaming push are similar to recording, and the difference lies only in service logic. During the interconnection with a third-party conference device, two-way communication is required, which is slightly different from the recording solution.

[0078] During the interconnection with a third-party conference device, referring to FIG. 8, a user service background starts an MRA (conference connector) robot. The MRA robot enters a conference ** (in this case, other participants can see this MRA robot), and the MRA robot is connected to the third-party conference device by using a connection protocol (such as the SIP or H323 protocol). After pulling all media data streams in the conference ** to the local place and mixing the streams, the MRA robot forwards them to the third-party conference device through the connection protocol. The MRA robot converts the media data stream from the third-party conference device into the private protocol of the conference ** and forwards it to other participants in the conference room through the media transmission network. In FIG. 8, 801 is a mobile conference terminal, and 802 is a conference room intelligent conference room ** for accessing the conference **.

[0079] If the audio and video conference system does not introduce a mixer engine, the architecture shown in FIG. 9 may further be used for implementing interconnection between the data transmission network 301 and the public switched telephone network 305. That is, the function of the mixer engine is implemented by using a selector and a robot involved in a bypass media processing system. The selector performs selection based on reported energy and a corresponding selection decision.

[0080] Based on the method for implementing audio and video conference provided above, an embodiment of this application further provides an apparatus for implementing audio and video conference. Referring to FIG. 10, the apparatus 1000 includes a receiving unit 1001 and a forwarding unit 1002:

> the receiving unit 1001 being configured to receive a pulling request for pulling data stream sent by a first interface machine in a first device group, the pulling request for pulling data stream including first identification information of a media data stream, the first device group being any device group in multiple device groups, a retrieval device existing in each device group in the multiple device groups, each device group including multiple interface machines, and each interface machine in the multiple interface machines being configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine; the first interface machine being any interface machine in the first device group; and
>
> the forwarding unit 1002 being configured to forward the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine forwards the media data stream to a corresponding conference accessing terminal; and the media data stream being obtained by the retrieval device in the first device group from a second interface machine in a second device group, a conference accessing terminal that generates the media data stream accessing the conference room by using the second interface machine, and the second device group being any device group in the multiple device groups.

[0081] In a possible implementation, the forwarding unit 1002 is configured to:

> search for a media data stream of the corresponding second interface machine according to the first identification information; and
>
> forward the pulling request for pulling data stream to the second interface machine when the media data stream of the second interface machine is not found;
>
> the receiving unit 1001 is further configured to receive the media data stream forwarded by the second interface machine according to the first identification information in the pulling request for pulling data stream; and
>
> the forwarding unit 1002 is further configured to forward the media data stream to the first interface machine.

[0082] In a possible implementation, the forwarding unit 1002 is further configured to:
forward the media data stream to the first interface machine when the media data stream of the corresponding second interface machine is found.

[0083] In a possible implementation, if the first device group and the second device group are located in different local area networks, the forwarding unit 1002 is configured to:

> forward the pulling request for pulling data stream to the second interface machine by using a transit agent device; and
>
> receive, by using the transit agent device, the media data stream forwarded by the second interface machine.

[0084] In a possible implementation, the retrieval device in the first device group is determined by the first interface machine according to a preset rule, and the preset rule causes a same media data stream to be routed, by using a same retrieval device, to an interface machine in a device group in which the retrieval device is located.

[0085] In a possible implementation, the apparatus further includes a sending unit, a recording unit, and a returning unit:

> the receiving unit 1001 is further configured to: receive, when a third interface machine in the third device group requests to enter the conference room, a registration request sent by the third interface machine, the third device group being any device group in the multiple device groups, and the third interface machine being any interface machine in the third device group;
>
> the sending unit is configured to: send the registration request to a central server when information about the conference room is not found locally, so that the central server records second identification information of the intra-group server in the third device group to a first conference registration list, and returns registration success information to the intra-group server in the third device group;
>
> the recording unit is configured to record third identification information of the third interface machine to a second conference registration list; and
>
> the returning unit is configured to return registration success information to the third interface machine.

[0086] In a possible implementation, if the information about the conference room is locally found, the recording unit is further configured to record the third identification information of the third interface machine to the second conference registration list; and
the returning unit is further configured to return registration success information to the third interface machine.

[0087] In a possible implementation, the receiving unit 1001 is further configured to receive, when a status of a conference accessing terminal connected to the third interface machine in the third device group changes, a notification message sent by the third interface machine; and
the sending unit is configured to send the notification message to the central server, so that the central server sends the notification message to an intra-group server in each device group, the intra-group server in each device group being configured to send the notification message to a successfully registered interface machine.

[0088] In a possible implementation, the receiving unit 1001 is further configured to receive a list pull request of the third interface machine when the third interface machine in the third device group requests to pull a participant list;

> the sending unit is configured to: send the list pull request to the central server when the participant list is not found locally;
>
> the receiving unit 1001 is further configured to receive the participant list sent by the central server; and
>
> the sending unit is configured to send the participant list to the third interface machine.

[0089] In a possible implementation, the sending unit is further configured to: return the participant list to the third interface machine when the participant list is found locally, and the participant list does not expire.

[0090] In a possible implementation, the apparatus further includes a selection unit and a mixing unit:

> the sending unit is further configured to: send, when the media data stream includes an audio data stream, acquired multiple audio data streams to a mixer located in a same device group by using a selector in a mixer engine, the multiple audio data streams respectively coming from fourth interface machines in different device groups;
>
> the selection unit is configured to select a target audio data stream from the multiple audio data streams by using the mixer;
>
> the sending unit is configured to send a selection notification to a target interface machine, the target interface machine being an interface machine connected to a conference accessing terminal that generates the target audio data stream;
>
> the receiving unit 1001 is further configured to receive, by using the mixer, the target audio data stream sent by the target interface machine;
>
> the mixing unit is configured to mix the target audio data stream;
>
> the forwarding unit 1002 is further configured to forward the mixed audio data stream to a selector corresponding to the fourth interface machine; the fourth interface machine including the target interface machine; and
>
> the sending unit is further configured to send the mixed audio data stream to the fourth interface machine by using the selector corresponding to the fourth interface machine, so that the fourth interface machine forwards the mixed audio data stream to another interface machine in a device group in which the fourth interface machine is located.

[0091] In a possible implementation, the apparatus further includes a processing unit:
the processing unit is configured to: pull, by using a robot terminal, a media data stream generated in the conference room, and process the media data stream.

[0092] An embodiment of this application further provides an electronic device for implementing an audio and video conference. The electronic device may be a terminal, and the terminal is a smartphone for example.

[0093] FIG. 11 is a block diagram of a partial structure of a smartphone related to a terminal according to an embodiment of this application. Referring to FIG. 11, the smartphone includes: a radio frequency (RF) circuit 1110, a memory 1120, an input unit 1130, a display unit 1140, a sensor 1150, an audio circuit 1160, a wireless fidelity (Wi-Fi) module 1170, a processor 1180, a power supply 1190, and the like. The input unit 1130 may include a touch panel 1131 and another input device 1132, the display unit 1140 may include a display panel 1141, and the audio circuit 1160 may include a speaker 1161 and a microphone 1162. A person skilled in the art may understand that the structure of the smartphone shown in FIG. 11 does not constitute a limitation on the smartphone, and the smartphone may include more components or fewer components than those shown in the figure, or some components may be combined, or a different component deployment may be used.

[0094] The memory 1120 may be configured to store a software program and module. The processor 1180 runs the software program and module stored in the memory 1120, to implement various functional applications and data processing of the smartphone. The memory 1120 may mainly include a program storage area and a data storage area. The program storage area may store an operating system, an application program required by at least one function (for example, a sound playing function and an image playing function), or the like. The data storage area may store data (such as audio data and an address book) created according to use of the smartphone. In addition, the memory 1120 may include a high speed random access memory, and may also include a non-volatile memory, such as at least one magnetic disk storage device, a flash memory, or another volatile solid-state storage device.

[0095] The processor 1180 is a control center of the smartphone, and is connected to various parts of the entire smartphone by using various interfaces and lines. By running or executing a software program and/or module stored in the memory 1120, and invoking data stored in the memory 1120, the processor 1180 executes various functions of the smartphone and performs data processing, thereby monitoring the entire smartphone. Optionally, the processor 1180 may include one or more processing units. Preferably, the processor 1180 may integrate an application processor and a modem. The application processor mainly processes an operating system, a user interface, an application program, and the like. The modem mainly processes wireless communication. It may be understood that the foregoing modem may not be integrated into the processor 1180.

[0096] In this embodiment, the steps performed by the terminal in the foregoing embodiment may be implemented based on the structure shown in FIG. 2.

[0097] The electronic device may further include a server. As shown in FIG. 12, an embodiment of this application further provides a server. Referring to FIG. 12, FIG. 12 is a structural diagram of a server 1200 according to an embodiment of this application. The server 1200 may vary greatly according to configuration or performance, and may include one or more central processing units (CPU) 1222 (for example, one or more processors) and a memory 1232, and one or more storage media 1230 (for example, one or more massive storage devices) storing an application program 1242 or data 1244. The memory 1232 and the storage medium 1230 may be used for transient storage or permanent storage. A program stored in the storage medium 1230 may include one or more modules (which are not marked in the figure), and each module may include a series of instruction operations on the server. Still further, the central processing unit 1222 may be configured to communicate with the storage medium 1230, and execute a series of instruction operations in the storage medium 1230 on the server 1200.

[0098] The server 1200 may further include one or more power supplies 1226, one or more wired or wireless network interfaces 1250, one or more input/output interfaces 1258, and/or one or more operating systems 1241, such as Windows Server™, Mac OS X™, Unix™, Linux™, and FreeBSD™.

[0099] In this embodiment, the central processing unit 1222 in the server 1200 may perform the following steps:

> receiving a pulling request for pulling data stream sent by a first interface machine in a first device group, the pulling request for pulling data stream including first identification information of a media data stream, the first device group being any device group in multiple device groups, a retrieval device existing in each device group in the multiple device groups, each device group including multiple interface machines, and each interface machine in the multiple interface machines being configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine; the first interface machine being any interface machine in the first device group; and
> forwarding the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine forwards the media data stream to a corresponding conference accessing terminal; and the media data stream being obtained by the retrieval device in the first device group from a second interface machine in a second device group, a conference accessing terminal

that generates the media data stream accessing the conference room by using the second interface machine, and the second device group being any device group in the multiple device groups.

[0100] According to an aspect of this application, a computer readable storage medium is provided, where the computer readable storage medium is configured to store program code, and when the program code runs on an electronic device, the electronic device performs the method for implementing audio and video conference described in the foregoing embodiments.

[0101] An aspect of this application provides a computer program product or a computer program, the computer program product or the computer program including computer instructions, the computer instructions being stored in a computer readable storage medium. A processor of an electronic device reads the computer instructions from the computer readable storage medium, and the processor executes the computer instructions, so that the electronic device performs the method provided in the foregoing optional implementations of the embodiments.

[0102] In the specification and accompanying drawings of this application, the terms "first", "second", "third", "fourth", and so on (if existing) are intended to distinguish between similar objects rather than describe a specific order or sequence. It is to be understood that the terms used in such a way are interchangeable in a proper circumstance, so that the embodiments of this application described herein can be implemented in orders except the order illustrated or described herein. Moreover, the terms "include", "contain" and any other variants mean to cover the non-exclusive inclusion. For example, a process, method, system, product, or device that includes a list of steps or units is not necessarily limited to those steps or units, but may include other steps or units not expressly listed or inherent to such a process, method, product, or device.

[0103] In the several embodiments provided in this application, it is to be understood that the disclosed system, apparatus, and method may be implemented in other manners. For example, the described apparatus embodiment is merely an example. For example, the unit division is merely a logical function division and may be other division during actual implementation. For example, a plurality of units or components may be combined or integrated into another system, or some features may be ignored or not performed. In addition, the displayed or discussed mutual couplings or direct couplings or communication connections may be implemented by using some interfaces. The indirect couplings or communication connections between the apparatuses or units may be implemented in electronic, mechanical, or other forms.

[0104] The units described as separate components may or may not be physically separated, and the components displayed as units may or may not be physical units, and may be located in one place or may be distributed over a plurality of network units. Some or all of the units may be selected according to actual needs to achieve the objectives of the solutions of the embodiments.

[0105] In addition, functional units in the embodiments of this application may be integrated into one processing unit, or each of the units may be physically separated, or two or more units may be integrated into one unit. The integrated unit may be implemented in a form of hardware, or may be implemented in a form of a software functional unit.

[0106] When the integrated unit is implemented in the form of a software functional unit and sold or used as an independent product, the integrated unit may be stored in a computer-readable storage medium. Based on such an understanding, the technical solutions of this application essentially, or the part contributing to the related art, or all or some of the technical solutions may be implemented in the form of a software product. The computer software product is stored in a storage medium and includes several instructions for instructing a computer device (which may be a personal computer, a server, a network device, or the like) to perform all or some of the steps of the methods described in the embodiments of this application. The foregoing storage medium includes any medium that can store program code, such as a USB flash drive, a removable hard disk, a read-only memory (ROM), a random access memory (RAM), a magnetic disk, or an optical disc.

[0107] The foregoing embodiments are merely used for describing the technical solutions of this application, but are not intended to impose limitations thereto. Although this application is described in detail with reference to the foregoing embodiments, a person of ordinary skill in the art shall understand that: modifications may still be made to the technical solutions described in the foregoing embodiments, or equivalent replacements may be made to the part of the technical features; However, these modifications or replacements do not depart the essence of the corresponding technical solutions from the spirit and scope of the technical solutions in the embodiments of this application.

Claims

1. A method for implementing audio and video conference, comprising:

   receiving, by a retrieval device in a first device group, a pulling request for pulling data stream sent by a first interface machine in the first device group, the pulling request comprising first identification information of a media data stream, wherein the first device group is one of multiple device groups each comprising a respective retrieval device and multiple interface machines, and each of the multiple interface machines is configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine; wherein the first interface machine is an interface machine in the first device group; and

forwarding, by the retrieval device in the first device group, the media data stream corresponding to the first identification information to the first interface machine, so as to enable forwarding the media data stream to a corresponding conference accessing terminal by the first interface machine; wherein the media data stream is obtained by the retrieval device in the first device group from a second interface machine of a second device group, a conference accessing terminal that generates the media data stream accesses the conference room by using the second interface machine, and the second device group is a device group in the multiple device groups.

2. The method according to claim 1, wherein forwarding, by the retrieval device in the first device group, the media data stream corresponding to the first identification information to the first interface machine comprises:

searching for a media data stream of the corresponding second interface machine according to the first identification information;

forwarding the pulling request to the second interface machine in response to the media data stream of the second interface machine being not found; and

receiving the media data stream forwarded by the second interface machine according to the first identification information in the pulling request, and forwarding the media data stream to the first interface machine.

3. The method according to claim 2, further comprising: forwarding the media data stream to the first interface machine in response to the media data stream of the corresponding second interface machine being found.

4. The method according to claim 2, wherein when the first device group and the second device group are located in different local area networks, forwarding the pulling request to the second interface machine comprises:

forwarding the pulling request to the second interface machine by using a transit agent device; and

receiving the media data stream forwarded by the second interface machine according to the first identification information in the pulling request comprises:

receiving, by using the transit agent device, the media data stream forwarded by the second interface machine.

5. The method according to any one of claims 1 to 4, wherein the retrieval device in the first device group is determined by the first interface machine according to a preset rule, and the preset rule causes a same media data stream to be routed, by using a same retrieval device, to an interface machine in a device group in which the retrieval device is located.

6. The method according to any one of claims 1 to 4, further comprising:

receiving, by an intra-group server in a third device group in response to a third interface machine in the third device group requesting to enter the conference room, a registration request sent by the third interface machine, wherein the third device group is a device group in the multiple device groups, and the third interface machine is an interface machine in the third device group;

sending the registration request to a central server in response to the intra-group server in the third device group not finding information about the conference room locally, so that the central server records second identification information of the intra-group server in the third device group to a first conference registration list, and returns registration success information to the intra-group server in the third device group; and

recording, by the intra-group server in the third device group, third identification information of the third interface machine into a second conference registration list, and returning the registration success information to the third interface machine.

7. The method according to claim 6, further comprising: recording the third identification information of the third interface machine into the second conference registration list in response to the intra-group server in the third device group finding the information about the conference room locally, and returning the registration success information to the third interface machine.

8. The method according to claim 6, further comprising:

receiving, by the intra-group server in the third device group in response to a status of a conference accessing terminal connected to the third interface machine in the third device group changing, a notification message sent by the third interface machine; and

sending, by the intra-group server in the third device group, the notification message to the central server, so that the central server sends the notification message to an intra-group server in each device group, wherein the intra-group server in each device group is configured to send the notification message to a successfully registered interface machine.

9. The method according to claim 6, further comprising:

receiving, by the intra-group server in the third device group, a list pull request of the third interface machine in response to the third interface machine in the third device group requesting to pull a participant list;

sending the list pull request to the central server in response to the intra-group server in the third device group not finding the participant list locally; and

receiving, by the intra-group server in the third device group, the participant list sent by the central server, and sending the participant list to the third interface machine.

10. The method according to claim 6, further comprising: returning the participant list to the third interface machine in response to the intra-group server in the third device group finding the participant list locally, and the participant list does not expire.

11. The method according to claims 1 to 4, further comprising:

sending, in response to the media data stream comprising an audio data stream, acquired multiple audio data streams to a mixer located in a same device group by using a selector in a mixer engine, wherein the multiple audio data streams respectively come from fourth interface machines in different device groups;

selecting a target audio data stream from the multiple audio data streams by using the mixer, and sending a selection notification to a target interface machine, wherein the target interface machine is an interface machine connected to a conference accessing terminal that generates the target audio data stream;

receiving, by using the mixer, a target audio data stream sent by the target interface machine, and mixing the target audio data stream;

forwarding the mixed audio data stream to a selector corresponding to the fourth interface machine; the fourth interface machine comprising the target interface machine; and

sending the mixed audio data stream to the fourth interface machine by using the selector corresponding to the fourth interface machine, so that the fourth interface machine forwards the mixed audio data stream to another interface machine in a device group in which the fourth interface machine is located.

12. The method according to claims 1 to 4, further comprising:
pulling, by using a robot terminal, a media data stream generated in the conference room, and processing the media data stream.

13. An apparatus for implementing audio and video conference, comprising a receiving unit and a forwarding unit, wherein:

the receiving unit is configured to receive a request for pulling data stream sent by a first interface machine in a first device group, the pulling request comprising first identification information of a media data stream, wherein the first device group is one of multiple device groups each comprising a respective retrieval device and multiple interface machines, and each of the multiple interface machines is configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses a conference room by using the interface machine; wherein the first interface machine is an interface machine in the first device group; and

the forwarding unit is configured to forward the media data stream corresponding to the first identification information to the first interface machine, so as to enable forwarding the media data stream to a corresponding conference accessing terminal by the first interface machine, wherein the media data stream is obtained by the retrieval device in the first device group from a second interface machine of a second device group, a conference accessing terminal that generates the media data stream accesses the conference room by using the second interface machine, and the second device group is a device group in the multiple device groups.

14. An audio and video conference system, comprising a data transmission network and a room management subsystem, wherein:

the data transmission network is configured to transmit a media data stream generated in a conference room, the data transmission network comprises multiple device groups, and each device group comprises a retrieval device and multiple interface machines;

each of the multiple interface machines is configured to connect to a conference accessing terminal, so that the conference accessing terminal accesses the conference room by using the interface machine;

a retrieval device in a first device group is configured to: receive a request for pulling data stream sent by a first interface machine in the first device group, the pulling request comprising first identification information of a media data stream; and forward the media data stream corresponding to the first identification information to the first interface machine, wherein the media data stream is obtained by the retrieval device in the first device group from a second interface machine of a second device group, a conference accessing terminal that generates the media data stream accesses the conference room by using the second interface machine, and each of the first device group and the second device group is a device group in the multiple device groups;

the first interface machine is an interface machine in the first device group, and the first interface machine is configured to forward the media data stream to a corresponding conference accessing terminal; and

the room management subsystem comprises the multiple device groups and a central server, each device group comprises an intra-group server, and the intra-group server is configured to assist the central server in managing the conference room.

15. An electronic device for implementing an audio and video conference, comprising a processor and a memory:

the memory being configured to store program code, and transmit the program code to the processor; and

the processor being configured to perform the method according to any one of claims 1 to 12 according to instructions in the program code.

16. A computer readable storage medium, configured to store program code, when running on an electronic device, causing the electronic device to perform the method according to any one of claims 1 to 12.

17. A computer program product, when executed, causing an electronic device to perform the method according to any one of claims 1 to 12.

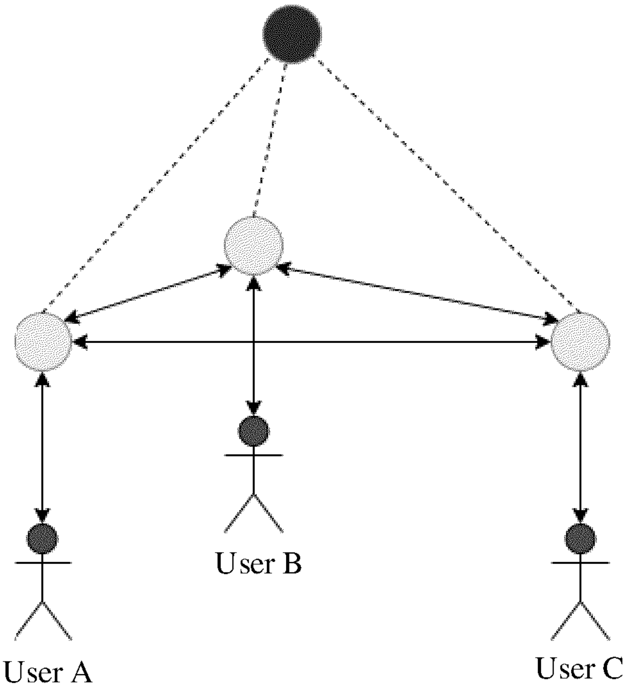


FIG. 1

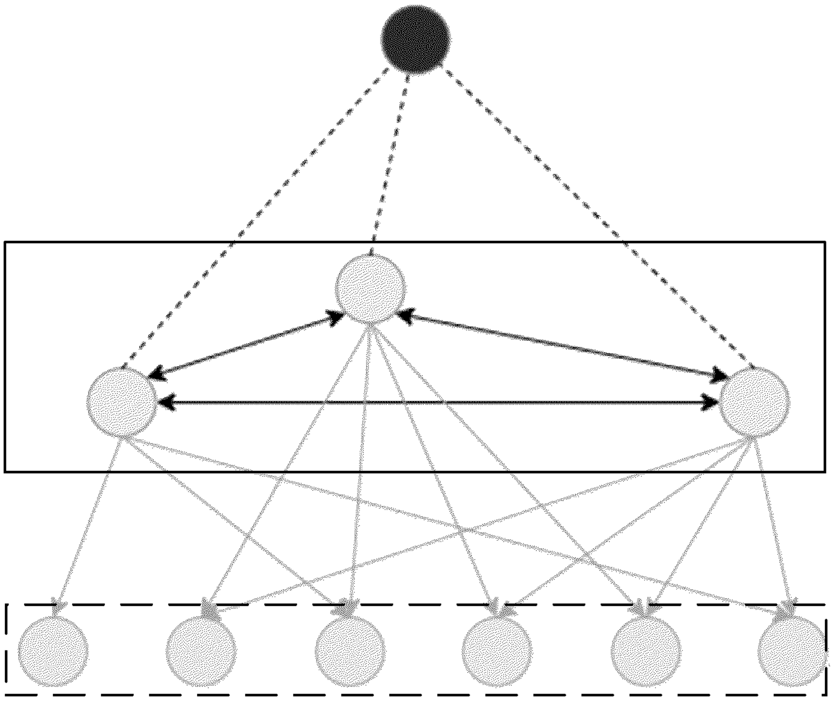


FIG. 2

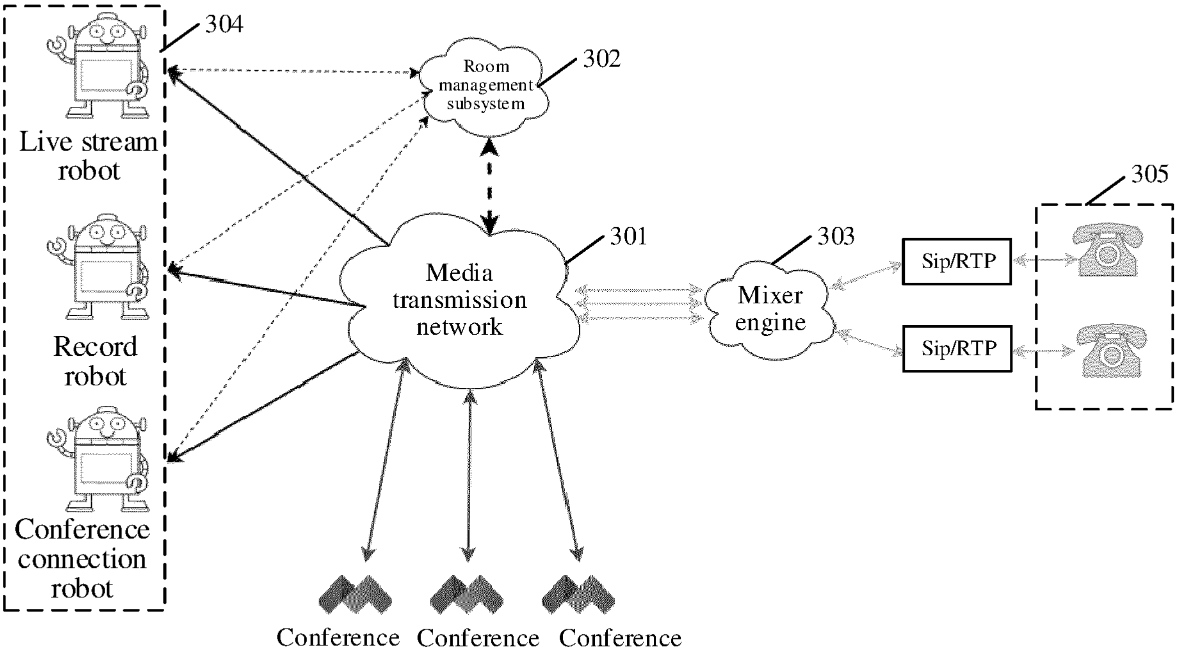

304
Live stream robot
Record robot
Conference connection robot
Room management subsystem
302
Media transmission network
301
303
Mixer engine
Sip/RTP
Sip/RTP
305
Conference
Conference
Conference

FIG. 3

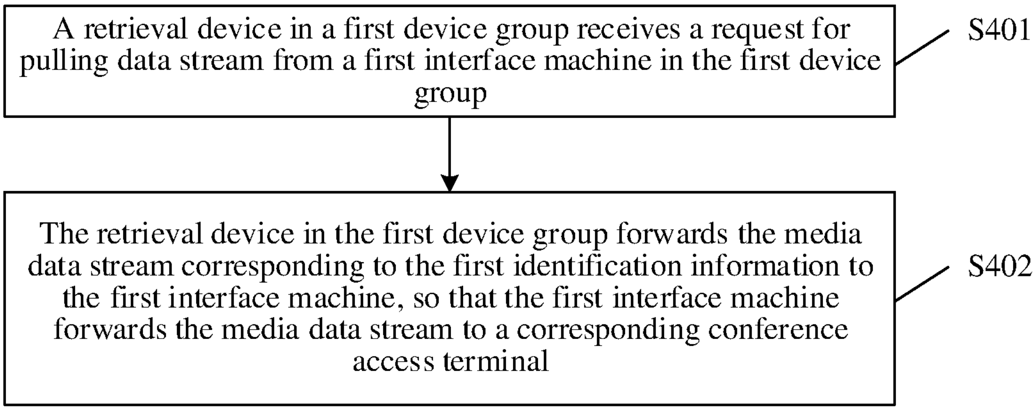

A retrieval device in a first device group receives a request for pulling data stream from a first interface machine in the first device group
S401
The retrieval device in the first device group forwards the media data stream corresponding to the first identification information to the first interface machine, so that the first interface machine forwards the media data stream to a corresponding conference access terminal
S402

FIG. 4a

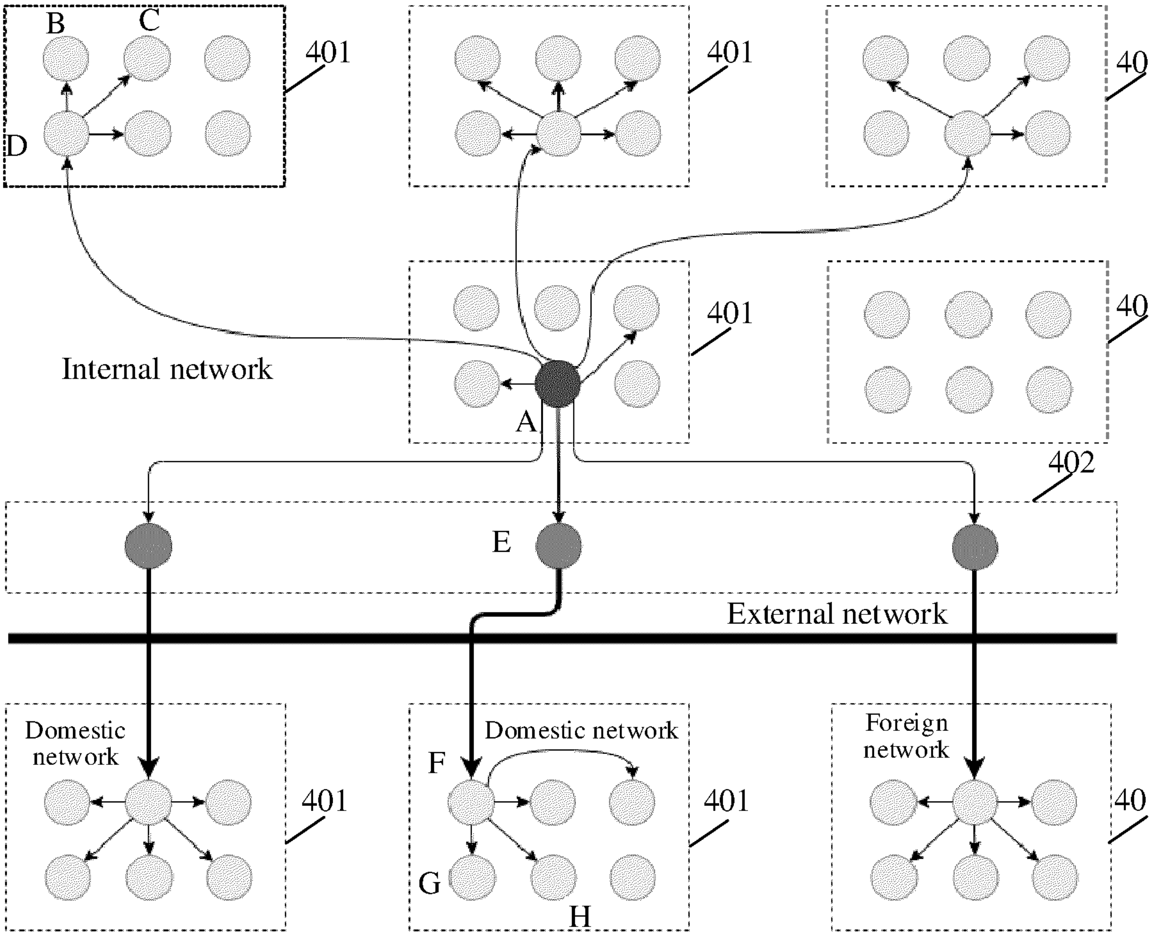
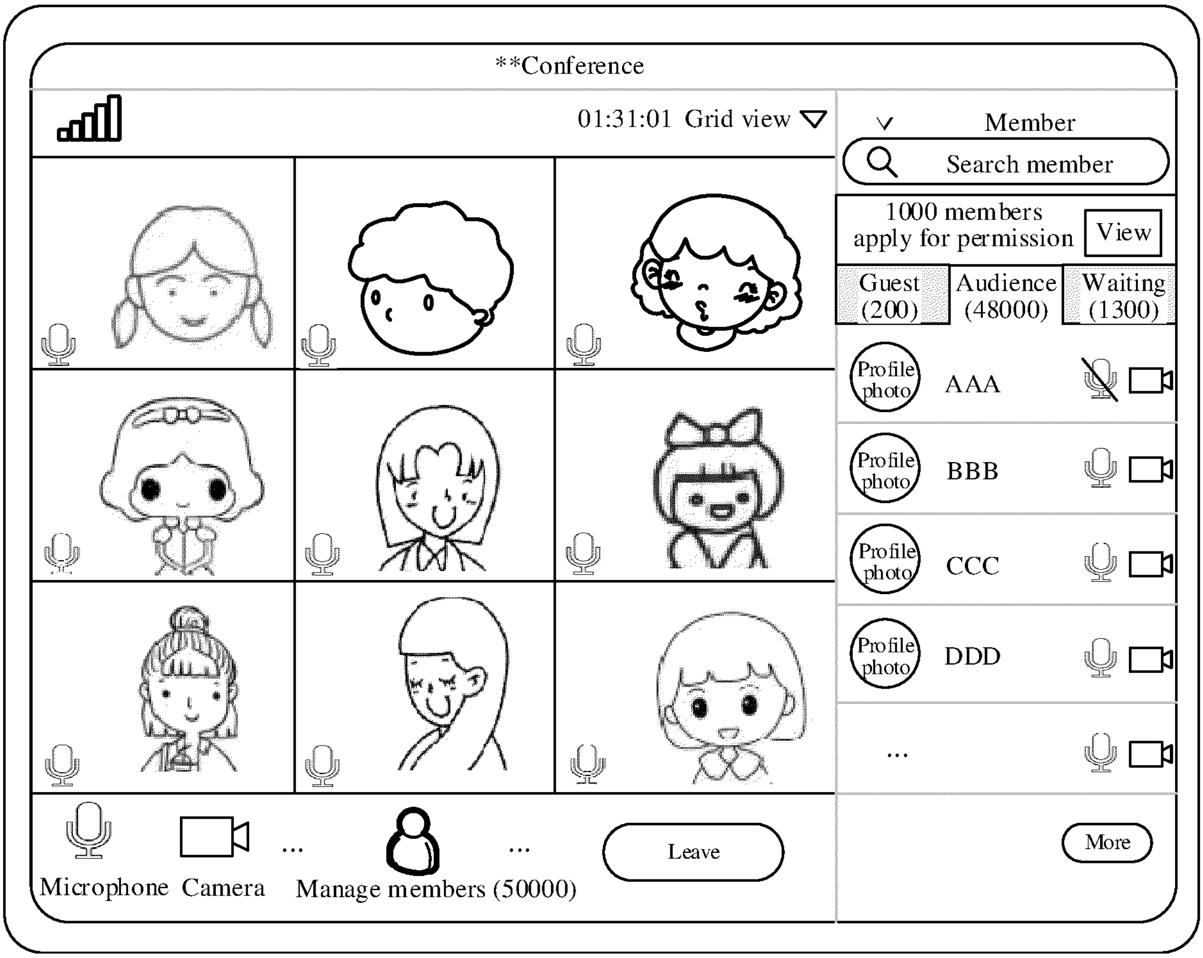

B
C
401
D
401
401
401
Internal network
401
A
402
E
External network
Domestic network
F
Domestic network
Foreign network
401
401
401
G
H

FIG. 4b



1000 members apply for permission

FIG. 5

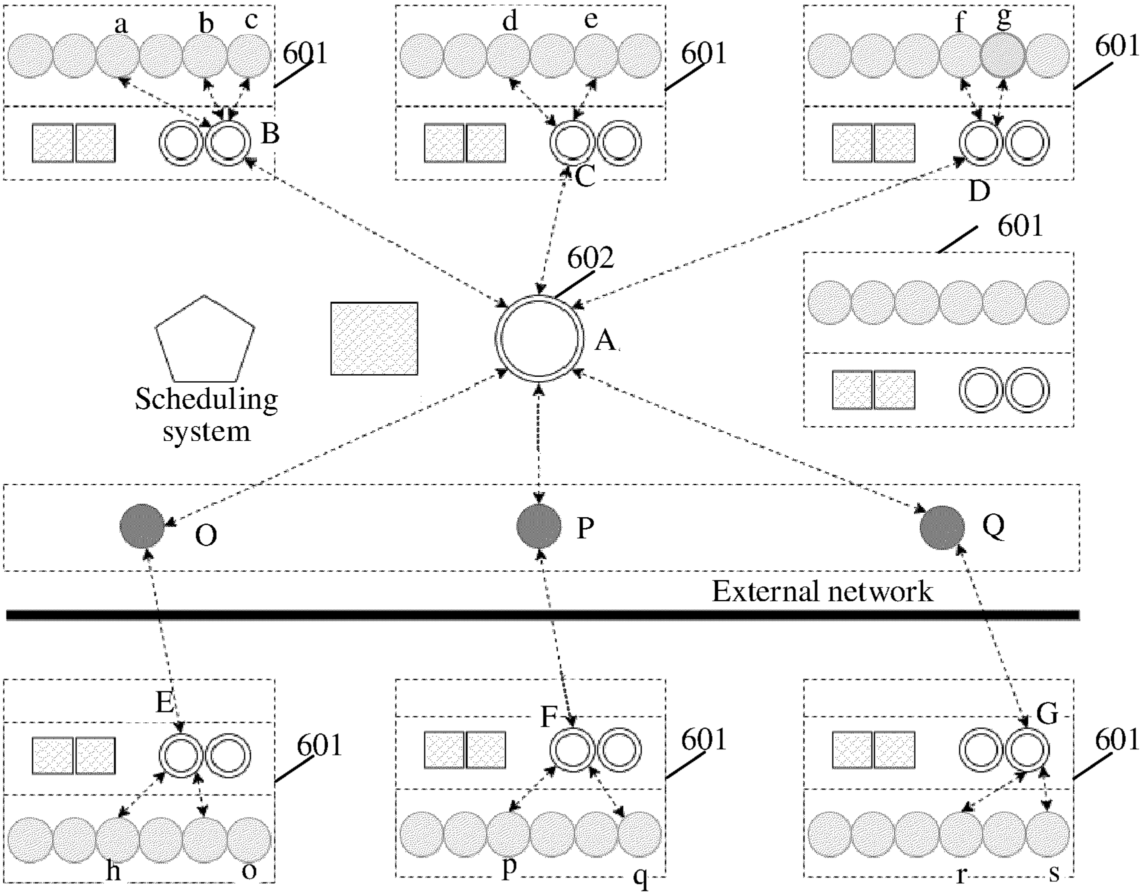

a
b
c
601
B
d
e
601
C
f
g
601
D
601
602
A
Scheduling system
O
P
Q
External network
E
601
h
o
F
601
p
q
G
601
r
s

FIG. 6

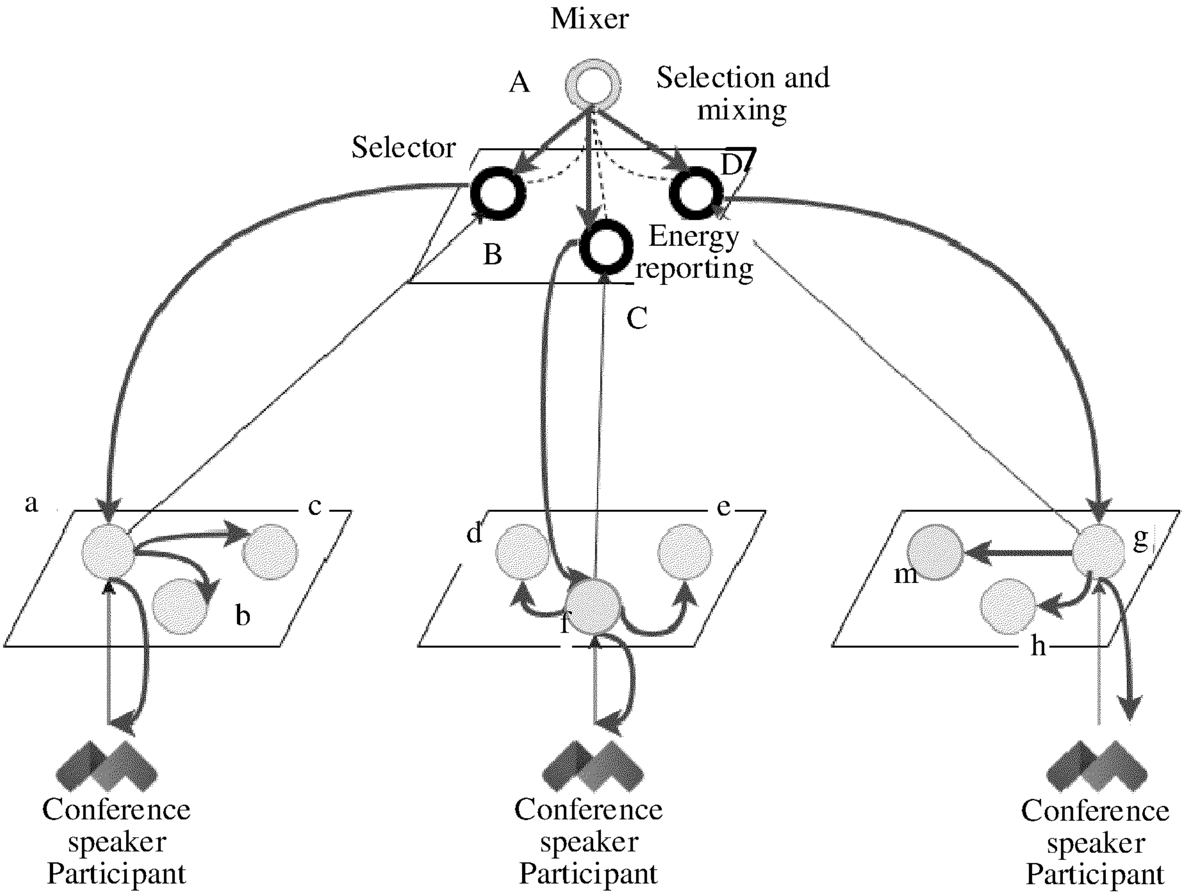

Mixer
A
Selection and mixing
Selector
D
B
Energy reporting
C
a
c
b
d
e
f
g
m
h
Conference speaker Participant
Conference speaker Participant
Conference speaker Participant

FIG. 7

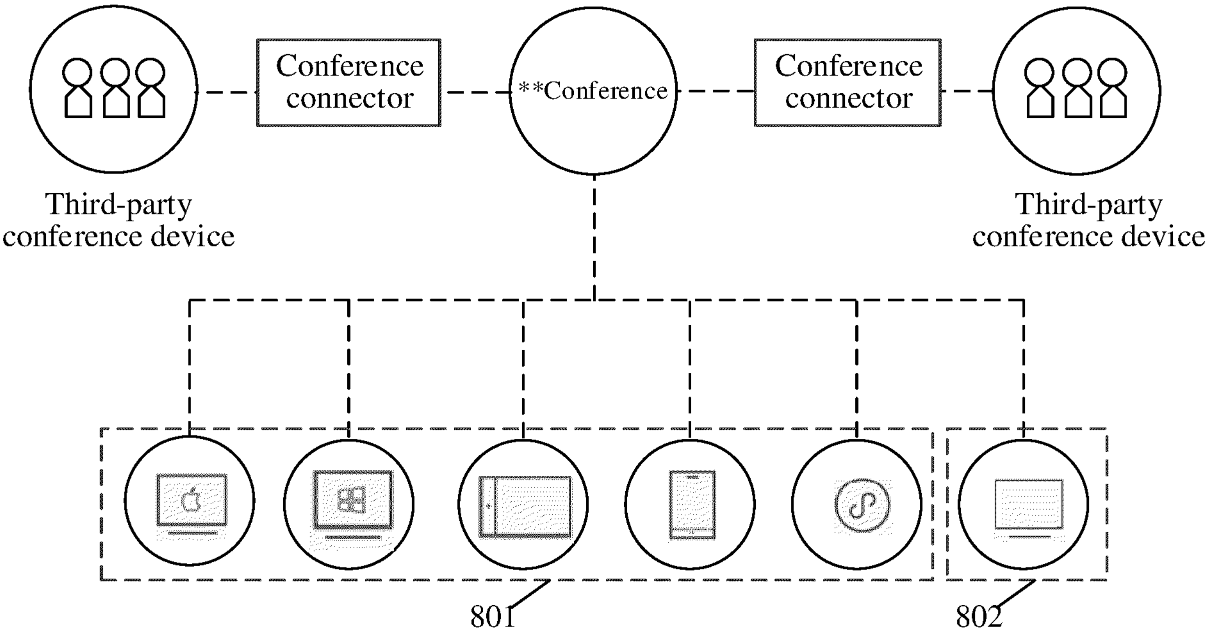

Conference connector
**Conference
Conference connector
Third-party conference device
Third-party conference device
801
802

FIG. 8

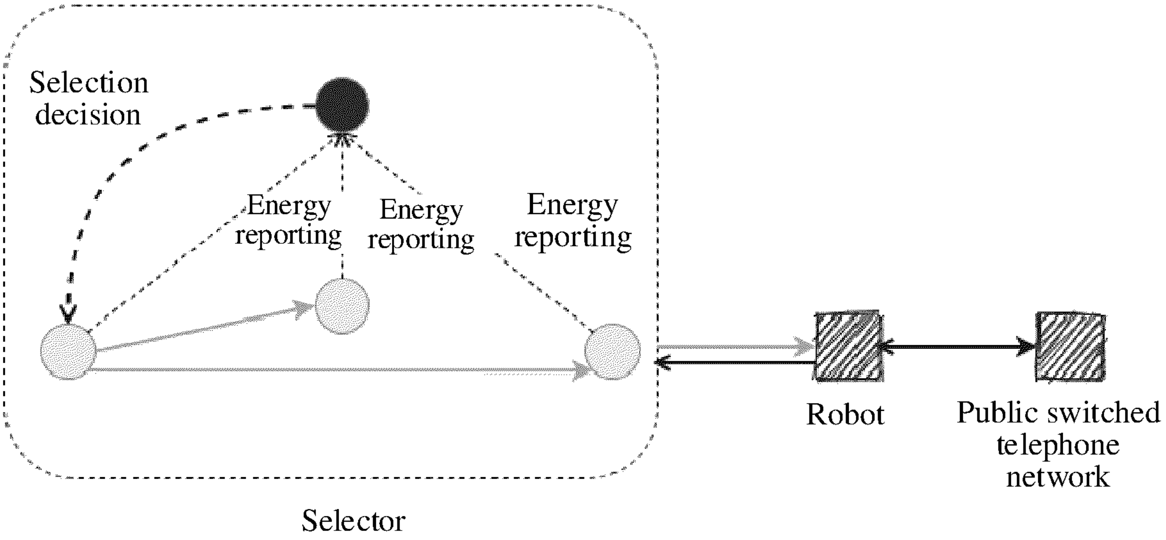

Selection decision
Energy reporting
Energy reporting
Energy reporting
Robot
Public switched telephone network
Selector

FIG. 9

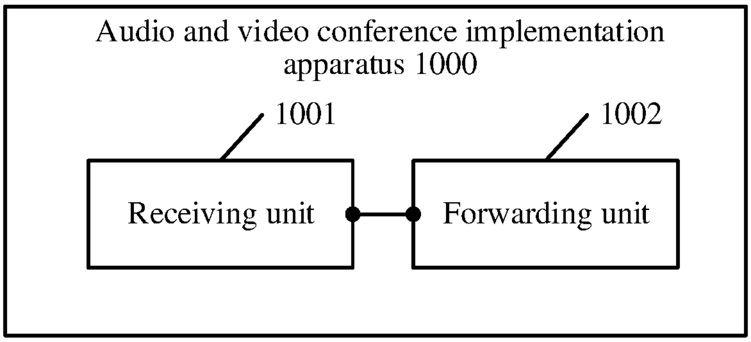

Audio and video conference implementation apparatus 1000
1001
1002
Receiving unit
Forwarding unit

FIG. 10

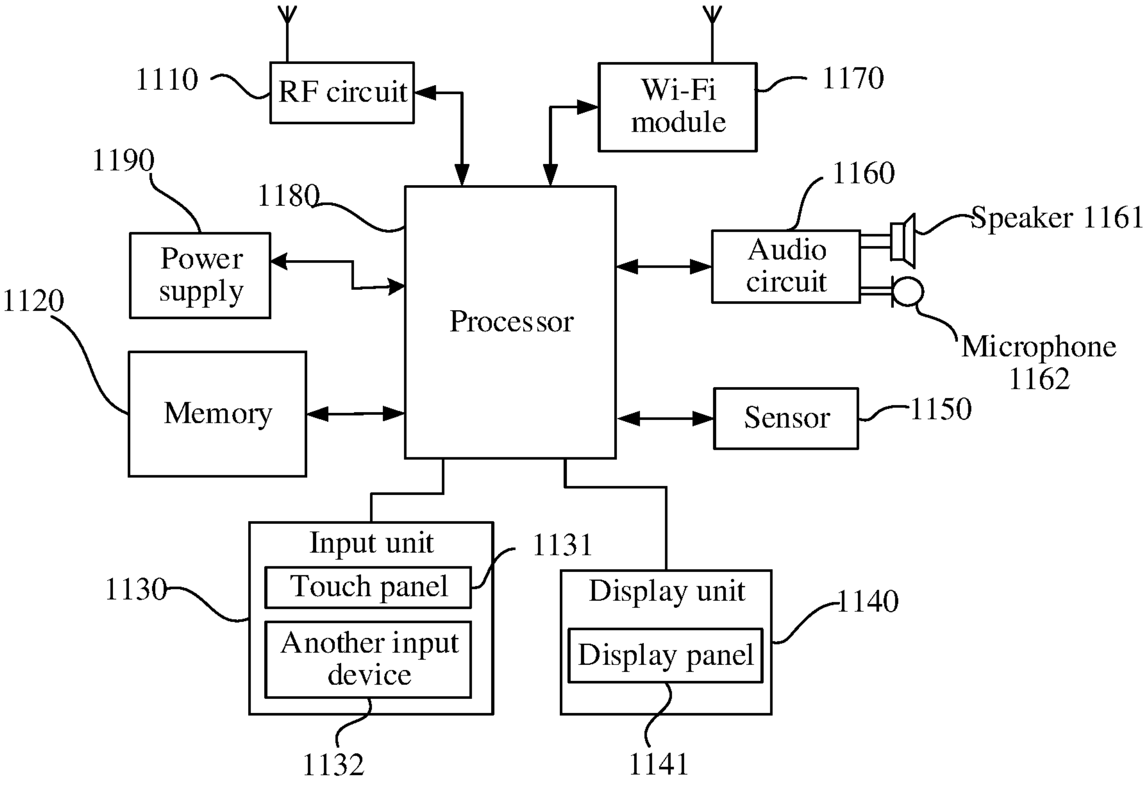

1110
RF circuit
Wi-Fi module
1170
1190
1180
1160
Speaker 1161
Power supply
Audio circuit
1120
Processor
Microphone 1162
Memory
Sensor
1150
Input unit
1131
1130
Touch panel
Display unit
1140
Another input device
Display panel
1132
1141

FIG. 11

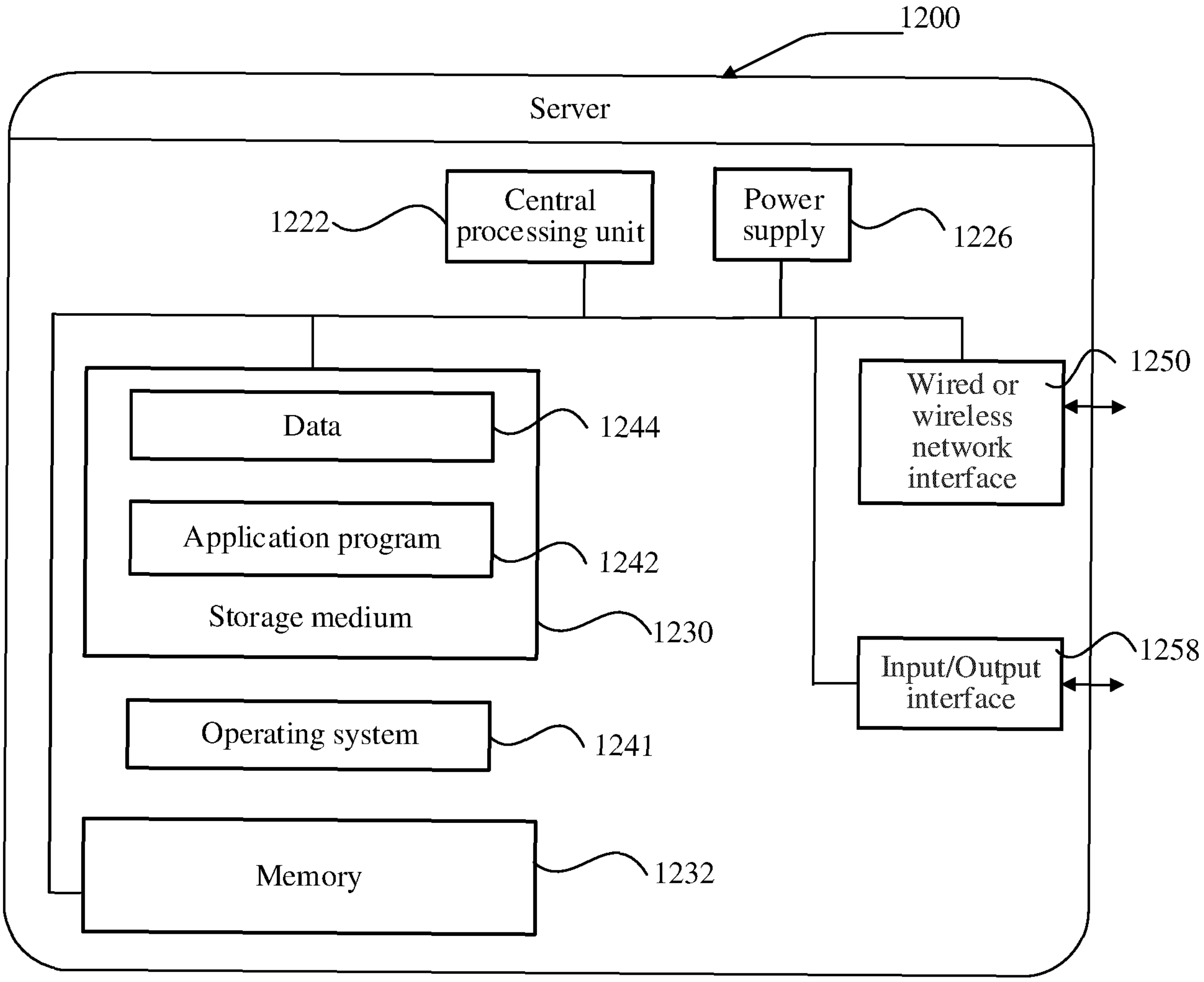

1200
Server
1222
Central processing unit
Power supply
1226
Wired or wireless network interface
1250
Data
1244
Application program
1242
Storage medium
1230
Input/Output interface
1258
Operating system
1241
Memory
1232

FIG. 12

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/099302

A. CLASSIFICATION OF SUBJECT MATTER

H04N 7/15(2006.01)i；H04N 21/262(2011.01)i

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

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)

CNABS; CNTXT; CNKI; VEN; WOTXT; USTXT; EPTXT; 3GPP: 音频, 视频, 音视频, 会议, 远程, 分组, 负载, 群组, 均衡, 平衡, 压力, 分担, 接口机, 终端, 回源, 数据流, 视频流, 标识, 标记, 转发, 拉流, 拉取, 提取, 跟踪, 选择性转发, 调度, audio, video, meeting, conference, group, room, set, load, balance, stress, terminal, interface, headstream, data, flow, stream, identity, ID, track, SFU, selective forwarding unit, scheduler, live

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 108307198 A (GUANGZHOU KUGOU TECHNOLOGY CO., LTD.) 20 July 2018 (2018-07-20) description, paragraphs [0007]-[0317]	1-17
A	CN 111447401 A (ZHUZHOU HUATONG TECHNOLOGY CO., LTD.) 24 July 2020 (2020-07-24) entire document	1-17
A	CN 108833823 A (VTRON GROUP CO., LTD.) 16 November 2018 (2018-11-16) entire document	1-17
A	CN 110113557 A (VISIONVERA INFORMATION TECHNOLOGY CO., LTD.) 09 August 2019 (2019-08-09) entire document	1-17
A	CN 111787269 A (NANJING BAIJIAYUN TECHNOLOGY CO., LTD.) 16 October 2020 (2020-10-16) entire document	1-17

☑ 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 **09 August 2022**	Date of mailing of the international search report **31 August 2022**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088, China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2022/099302

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2016255126 A1 (SARRIS WILLIAM) 01 September 2016 (2016-09-01) entire document	1-17
A	US 2021082068 A1 (OBRIEN BEATRICE T) 18 March 2021 (2021-03-18) entire document	1-17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/CN2022/099302

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	108307198	A	20 July 2018	None	
CN	111447401	A	24 July 2020	None	
CN	108833823	A	16 November 2018	None	
CN	110113557	A	09 August 2019	None	
CN	111787269	A	16 October 2020	None	
US	2016255126	A1	01 September 2016	None	
US	2021082068	A1	18 March 2021	None	

REFERENCES CITED IN THE DESCRIPTION

Patent documents cited in the description

- CN 202110875339 **[0001]**