

AUDIO/VIDEO CONFERENCE IMPLEMENTATION METHOD, AUDIO/VIDEO CONFERENCE SYSTEM AND RELATED APPARATUS

RELATED APPLICATION

[0001] This application claims priority to Chinese Patent Application No. 202110875339.6, entitled "METHOD FOR IMPLEMENTING AUDIO OR VIDEO CONFERENCING, SYSTEM AUDIO OR VIDEO CONFERENCING, 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] The present disclosure relates to the computer field, and in particular, to a method for implementing audio or video conferencing, a system for an audio or video conferencing, and a related apparatus.

BACKGROUND OF THE DISCLOSURE

[0003] Technology of networks, communications, and streaming media develops rapidly, and people's work and learning are becoming more and more mobile. Thus, enterprises and individuals have an increasing demand on video communications, from which audio/video conferencing systems emerge.

[0004] Current audio/video conferencing mainly adopt architecture based on a selective forwarding unit (SFU), and the architecture comprises one server and multiple terminals. In response to receiving an audio/video stream (that is, a media data stream) shared by a terminal in a conference room, the server forwards the audio/video stream directly to other terminal(s) in the conference room.

[0005] In such architecture, a load of data distribution on the server would be greatly increased when there are too many participants accessing one conference room, hence a quantity of participants of a conference is limited, and a quantity of participants who can enable microphones or cameras at the same time is also limited. A redirection process is necessary when an ordinary audience "raises hand" for speaking, which reduces smoothness of conferencing experience.

SUMMARY

[0006] In order to address at least the above technical issues, a method for implementing video or audio conferencing, a system for audio or video conferencing, and a related apparatus are provided accordingly to embodiments of the present disclosure. A load of data distribution on a media source device, such as a second interface machine, is greatly reduced, and hence more participants, even millions of participants, can be supported. Moreover, the reduced load of distribution during transmitting a media data stream allows all participants to enable their microphones or cameras without redirection, which renders conferencing experience smooth.

[0007] Following technical solutions are provided according to embodiments of the present disclosure.

[0008] In a first aspect, a method for implementing audio or video conferencing is provided according to an embodiment of the present disclosure. The method comprises: receiving, by an origin-pull device in a first device group, a stream pulling request from a first interface machine in the first device group, where the stream pulling request comprises a first identifier of a media data stream, the first device group is one of multiple device groups, each of which comprises a respective origin-pull device and multiple interface machines, each of the multiple interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access a conference room via said interface machine, and the first interface machine is one of the multiple interface machines in the first device group; and forwarding, by the origin-pull device in the first device group, the media data stream corresponding to the first identifier to the first interface machine to enable the first interface machine to forward the media data stream to the conference accessing terminal corresponding to the first interface machine, where the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the second device group is among the multiple device groups.

[0009] In a second aspect, an apparatus for implementing audio or video conferencing is provided according to an embodiment of the present disclosure. The apparatus comprises a receiving unit and a forwarding unit. The receiving unit is configured to receive a stream pulling request from a first interface machine in a first device group, where the stream pulling request comprises a first identifier of a media data stream, the first device group is one of multiple device groups, each of which comprises an origin-pull device and multiple interface machines, each of the multiple interface machines is configured to connect a conference accessing terminal to enable the

conference accessing terminal to access a conference room via said interface machine, and the first interface machine is one of the multiple interface machines in the first device group. The forwarding unit is configured to forward the media data stream corresponding to the first identifier to the first interface machine to enable the first interface machine to forward the media data stream to the conference accessing terminal corresponding to the first interface machine, where the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the second device group is among the multiple device groups.

[0010] In a third aspect, a system for audio or video conferencing is provided according to an embodiment of the present disclosure. The system comprises a data transmission network and a room management subsystem. 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, each of which comprises multiple interface machines. Each of the multiple interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access the conference room via said interface machine. Each of the multiple device groups comprises an origin-pull device, the origin-pull device in a first device group is configured to: receive a stream pulling request from a first interface machine in the first device group, where the stream pulling request comprises a first identifier of a media data stream; and forward the media data stream corresponding to the first identifier to the first interface machine, where the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the first device group and the second device group each is among the multiple device groups. The first interface machine is one of the multiple interface machines in the first device group, and the first interface machine is configured to forward the media data stream to the conference accessing terminal corresponding to the first interface machine. The room management subsystem comprises the multiple device groups and a central server, where each of the multiple device groups comprises an intra-group server, and the intra-group server is configured to assist the central server for managing the conference room.

[0011] In a fourth aspect, an electronic device for implementing audio or video conferencing is provided according to an embodiment of the present disclosure. The electronic device comprises a processor and a memory, where the memory is configured to store a program

code and transmit the program code to the processor, and the processor is configured to perform the method in the first aspect according to instructions in the program code.

[0012] In a fifth aspect, a computer-readable storage medium is provided according to an embodiment of the present disclosure. The computer readable storage medium is configured to store a program code, and the program code when executed on an electronic device causes the electronic device to perform the method in the first aspect.

[0013] In a sixth aspect, a computer program product is provided according to an embodiment of the present disclosure. The computer program product when being executed causes an electronic device to perform the method in the first aspect.

[0014] Herein the technical solutions are related to the multiple device groups, each device group comprises multiple interface machines, and each interface machine is configured to connect the conference accessing terminal, so that the corresponding conference accessing terminal can access the conference room via such interface machine. The multiple interface machines play a role similar to that of a server in conventional technology, in which multiple conference accessing terminals access a conference room via such server. The conference accessing terminals are allocated to different interface machines, and hence the interface machines can share a load, which facilitates accessing of more conference accessing terminals and supports a large conference. Each device group further comprises the origin-pull device. The media data stream is generated by the second interface machine accessing the conference room. Taking the first device group as an example, the origin-pull device is capable to forward the media data stream corresponding to the first identifier to the first interface machine after receiving the stream pulling request from the first interface machine, because the stream pulling request comprises the first identifier. Thereby, the first interface machine can forward the media data stream to its corresponding conference accessing terminal. Accordingly, when each interface machine needs to pull the media data stream from the second interface machine, the second interface machine is only necessary to interact with the origin-pull device in each device group. The origin-pull devices share the load of data distribution on a media source device, e.g., the second interface machine, and interact with other interface machines, and the media source device is not necessary to interact with each interface machine, which greatly reduces the load of data distribution pressure on the media source device, e.g., the second interface machine. Therefore, the method for implementing audio/video conferencing can support more participants, even millions of participants. In addition, the load of data distribution during streaming media data is greatly reduced, and hence all participants (i.e., users) can enable

microphones or cameras at the same time without redirection, which render the conferencing experience smooth.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] In order to clarify illustration of the technical solutions according to embodiments of the present disclosure or conventional technology, hereinafter briefly described are the drawings to be applied in embodiments of the present disclosure or conventional technology. Apparently, the drawings in the following descriptions are only some embodiments of the present disclosure, and other drawings may be obtained by those skilled in the art based on the provided drawings without creative efforts.

[0016] Figure 1 is a diagram of conference architecture based on an SFU in relevant technology.

[0017] Figure 2 is a diagram of another conference architecture based on an SFU in relevant technology.

[0018] Figure 3 is a structural diagram of a system for audio/video conferencing in relevant technology.

[0019] Figure 4a is a flowchart of a method for implementing audio/video conferencing according to an embodiment of the present disclosure.

[0020] Figure 4b is a structural diagram of a data transmission network according to an embodiment of the present disclosure.

[0021] Figure 5 is a diagram of an interface of video conferencing in a system for audio/video conferencing according to an embodiment of the present disclosure.

[0022] Figure 6 is a structural diagram of a room management subsystem according to an embodiment of the present disclosure.

[0023] Figure 7 is a diagram of audio processing architecture based on a mixer engine according to an embodiment of the present disclosure.

[0024] Figure 8 is a diagram of architecture for interconnection with a third-party conference device according to an embodiment of the present disclosure.

[0025] Figure 9 is a diagram of audio processing architecture according to another embodiment of the present disclosure.

[0026] Figure 10 is a structural diagram of an apparatus for implementing audio/video conferencing according to an embodiment of the present disclosure.

[0027] Figure 11 is a structural diagram of a terminal according to an embodiment of the present disclosure.

[0028] Figure 12 is a structural diagram of a server according to an embodiment of the present disclosure.

DESCRIPTION OF EMBODIMENTS

[0029] Hereinafter embodiments of the present disclosure are described in conjunction with the drawings.

[0030] Currently, audio/video conferencing mainly adopts architecture based on a selective forwarding unit (SFU). Reference is made to Figure 1. A black node represents a server, each gray node represents a terminal, multiple terminals are connected to the server, and a user A, a user B, and a user C access a conference room via the terminals, respectively, to participate in a conference.

[0031] In such architecture, a load of data distribution on the server would be greatly increased when there are too many participants accessing one conference room, hence a quantity of participants of a conference is limited, and a quantity of participants who can enable microphones or cameras at the same time is also limited. A redirection process is necessary when an ordinary audience "raises hand" for speaking, which reduces smoothness of conferencing experience.

[0032] In addition, the participants cannot enable their cameras or microphones at the same time in such architecture. Hence, diffusion proxy nodes (e.g., nodes in a dashed box as shown in Figure 2) are added on a basis of the SFU architecture as shown in Figure 1. A user is directed to the diffusion proxy node when not speaking, and is redirected to a communication node (e.g., nodes in a solid box as shown in Figure 2) when raises hand for speaking. More participants can access a conference due to a distribution advantage of the diffusion proxy nodes. Such manner requires the redirection process which renders conferencing experience not smooth.

[0033] In order to address at least the above technical issues, a method for implementing audio/video conferencing is provided according to an embodiment of the present disclosure. In such method, when each interface machine needs to pull a media data stream from a second interface machine, the second interface machine is only necessary to interact with an origin-pull device in each device group. The origin-pull devices share a load of data distribution on a media source device, e.g., the second interface machine, and interact with other interface machines, and

the media source device is not necessary to interact with each interface machine, which greatly reduces the load of data distribution pressure on the media source device, e.g., the second interface machine. Therefore, the method for implementing audio/video conferencing can support more participants, even millions of participants. In addition, the load of data distribution during streaming media data is greatly reduced, and hence all participants (i.e., users) can enable microphones or cameras at the same time without redirection, which render the conferencing experience smooth.

[0034] Herein the method for implementing audio/video conferencing is applicable to various audio/video conferencing scenarios, and in particular, to super-large-scale video/audio conferencing in which millions of people participates in a single conference.

[0035] Hereinafter architecture of a system for audio/video conferencing according to an embodiment of the present disclosure is illustrated. Participants enter a conference room to participate in a conference based on the system of audio/video conferencing, and generally, the participants comprise guest(s), who should enable their microphones and/or cameras, and ordinary audience. A media data stream is generated when a guest enables his/her microphone and/or camera, and the media data stream is transmitted to conference accessing terminal(s) corresponding to the ordinary audience and conference accessing terminal(s) corresponding to other guest(s). The conference accessing terminal refers to a terminal for accessing the conference room, and it may be a user terminal. Therefore, a main factor of the audio/video conferencing is transmission of the media data stream when a participant enables audio/video generation, for example, enables a camera or a microphone. The system for audio/video conferencing system may support a large-scale conference when the transmission of the media data stream is not limited due to a quantity of participants in a conference room. Accordingly, an emphasis of embodiments of the present disclosure is improving architecture of a data transmission network.

[0036] Moreover, each conference corresponds to one conference room for storing room information, storing a list of participants, and maintaining a status (whether enabling audio/video generation, e.g., whether enabling a microphone, and whether enabling a camera, whether being muted, or whether exiting the room) of users. In a super-large-scale conference, it is possible that everyone keeps switching between enabling and disabling audio/video generation and keeps switching between entering and exiting the room, which generates a large quantity of messages concerning changes in the status. Hence, pulling and synchronizing the list of participants in the super-large audio/video conference puts a heavy load on a room management subsystem. Accordingly, improvements on the room management subsystem according to embodiments of the

present disclosure also enable a system for audio/video conferencing to support super-large-scale audio/video conferencing.

[0037] Accordingly, in an embodiment, architecture of a system for audio/video conferencing may comprise a data transmission network 301 and a room management subsystem 302, as shown in Figure 3. 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.

[0038] A method for implementing audio/video conferencing is first described with respect to the data transmission network 301 transmitting a media data stream. Reference is made to Figure 4a. The method comprises following steps S401 and S402.

[0039] In step S401, an origin-pull device in a first device group receives a stream pulling request from a first interface machine in the first device group.

[0040] The data transmission network 301 may be as shown in Figure 4b, and comprises multiple service groups 401, for example, as indicated by each dashed box in Figure 4b. Each device group 401 comprises multiple interface machines, and each interface machine is a node, for example, indicated by a circle in the device group 401 as shown in Figure 4b. Each interface machine is configured to connect a corresponding conference accessing terminal, so that the conference accessing terminal can access a conference room via such interface machine, and a participant corresponding to the conference accessing terminal can enter the conference room to participate in a conference. The device group may be denoted as SET, that is, the interface machines are grouped based on SET, and each interface machine may be a server. There is a respective origin-pull device in each device group 401.

[0041] The first device group may be an arbitrary one of the multiple device groups, and the first interface machine may be an arbitrary interface machine in the first device group.

[0042] When an interface machine (taking the first interface machine as an example) in a device group (taking the first device group as an example) needs to pull a media data stream, the interface machine may transmit a stream pulling request to the origin-pull device in the first device group. The stream pulling request may comprises a first identifier of the media data stream, and the first identifier identifies the media data stream, i.e., a stream of media data to which the first interface machine has subscribed. Thereby, the origin-pull device is aware of which media data stream is to be pulled. The first interface machine may be connected to the conference accessing terminal which corresponds to a user who has subscribed to the media data. That is, in a case that

a user on the first interface machine subscribes to media data of a certain interface machine, the first interface machine is triggered to transmit the stream pulling request to the origin-pull device in the first device group.

[0043] In step S402, the origin-pull device in the first device group forwards the media data stream corresponding to the first identifier to the first interface machine, so that the first interface machine forwards the media data stream to a corresponding conference accessing terminal.

[0044] After receiving the stream pulling request, the origin-pull device in the first device group forwards the media data stream corresponding to the first identifier to the first interface machine, and 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 via a second interface machine, and the second device group is an arbitrary one of the multiple device groups.

[0045] The second interface machine may be connected to the conference accessing terminal which corresponds to a user who enables audio/video generation. For example, a user, e.g., a second user, enables video generation on his/her conference accessing terminal, the conference accessing terminal accesses the conference via the second interface machine, and the second interface machine belongs to the second device group. In a case that any user participating in the conference (e.g., a first user) wishes to watch a video of the second user (e.g. subscribes to the video of the second user), the first interface machine connected to the conference accessing terminal corresponding to the first user may determine the origin-pull device in the device group to which the first interface machine belongs (e.g., the first device group) and transmit the stream pulling request to the origin-pull device in the first device group. Thereby, the origin-pull device forwards the media data stream, which is obtained from the second interface machine, to the first interface machine according to the stream pulling request, and the first interface machine forwards the media data stream to the corresponding conference accessing terminal. Herein the user(s) may be participant(s).

[0046] The media data stream at the origin-pull device is pulled from the second interface machine, and the pulled media data stream may be stored locally at the origin-pull device. Generally, the origin-pull device needs to pull the media data stream from the second interface machine when receiving the stream pulling request for pulling the media data stream for the first time, because the media data stream has not been locally stored. In a case that another interface machine has previously requested such media data stream, the media data stream has been locally saved, and the media data stream does not need to pull it again from the second interface machine.

[0047] Therefore, in one embodiment, the media data stream may be pulled in different manners according to whether the origin-pull device in the first device group determines that the media data stream is locally stored.

[0048] In an embodiment, the origin-pull device in the first device group may forward the media data stream corresponding to the first identifier to the first interface machine in a following manner. The origin-pull device in the first device group receives the stream pulling request transmitted by the first interface machine, and searches for the corresponding media data stream of the second interface machine according to the first identifier. In case of not finding the media data stream of the second interface machine, the origin-pull device in the first device group forwards the stream pulling request to the second interface machine, receives the media data stream which is forwarded by the second interface machine according to the first identifier, and forwards the media data stream to the first interface machine.

[0049] In a case that the stream pulling request is transmitted when the first user has subscribed to the video of the second user, the second interface machine may record locally that the origin-pull device subscribes to the second interface machine after receiving the stream pulling request. Then, the second interface machine transmits the media data stream to the origin-pull device. The origin-pull device stores the received media data stream locally, so that it is not necessary to interact with the second interface machine again when another interface machine requests the media data stream. In such case, the origin-pull device may directly return the media data stream to the other interface machine.

[0050] Figure 4b is taken as an example. Node A may serve as the second interface machine, and a guest is on the second interface machine. That is, a user corresponding to the conference accessing terminal connecting the second interface machine may enable the video generation and generates an uplink video data stream, which serves as the foregoing media data stream. Node B may serve as the first interface machine, and node D may serve as the origin-pull device of the device group to which node B belongs, e.g., of the first device group. A user at node B subscribes to the uplink video data stream of node A. Node B first determines node D from the first device group, and transmits the stream pulling request to node D. After receiving the stream pulling request of node B, node D searches locally for the media data stream of node A first. In case of not finding the media data stream of node A, the node D initiates the stream pulling request towards node A. After receiving the stream pulling request from node D, node A locally records that node D subscribes to node A, and then forwards a video data stream to node D.

Node D receives the video data stream and forwards the video data stream to node B. As a final step, node B transmits the video data stream to the corresponding conference accessing terminal.

[0051] After searching for the media data stream of the second interface machine according to the first identifier, the origin-pull device in the first device group may find the media data stream of the second interface machine. In such case, the origin-pull device forwards the found media data stream directly to the first interface machine. Thereby, the media data stream can be pulled with higher efficiency, and a load of data distribution on a media source device such as the second interface machine is further reduced.

[0052] Figure 4b is further taken as another example. Node A may serve as the second interface machine, and a guest is on the second interface machine. That is, a user corresponding to the conference accessing terminal connecting the second interface machine may enable the video generation and generates an uplink video data stream, which serves as the foregoing media data stream. Node C may serve as the first interface machine, and node D may serve as the origin-pull device of the device group to which node C belongs, e.g., of the first device group. A user at node C subscribes to the uplink video data stream of node A. Node C first determines node D from the first device group, and transmits the stream pulling request to node D. Node B and node C belong to the same device group, and both determine node D through consistent hashing or another method to transmit the stream pulling request to node D. Node D determines the video data stream exists locally (since it is requested by node B), and forwards the video data directly to node C. Node C receives the video data stream forwarded from node D, and forwards the video data stream to the corresponding conference accessing terminal as a final step.

[0053] The origin-pull device in the first device group is determined by an interface machine that needs to request the media data stream, e.g., the first interface machine, according to a preset rule. The preset rule enables the same media data stream to be routed via the same origin-pull device to interface machine(s) in a device group to which the origin-pull device belongs. For example, the node B and the node C in the same device group both can find the same node D through consistent hashing or another method. Thereby, unified distribution management on the same media data stream within the same device group is facilitated.

[0054] The first device group and the second device group may be located in the same local area network (e.g., within an internal network), or may be located in different local area networks (e.g. concerning an external network which includes a domestic network and a foreign network). In a case that the first device group and the second device group are located in the same local area network, the origin-pull device in the first device group forwards the stream pulling request directly

to the second interface machine, and the second interface machine transmits the media data stream directly to the origin-pull device. In a case that the first device group and the second device group are located in different local area networks, the data transmission network 301 further comprises a transmit proxy device 402, for example, a device denoted by 402 in Figure 4b. The origin-pull device in the first device group needs to forward the stream pulling request to the second interface machine via the transmit proxy device 402. Correspondingly, the second interface machine forwards the media data stream to the origin-pull device in the first device group via the transmit proxy device 402.

[0055] Figure 4b is further taken as another example. Node A may serve as the second interface machine, and a guest is on the second interface machine. That is, a user corresponding to the conference accessing terminal connecting the second interface machine may enable the video generation and generates an uplink video data stream, which serves as the foregoing media data stream. Node G may serve as the first interface machine, and node F may serve as the origin-pull device of a device group to which node G belongs, for example, of the first device group. In such case, the first device group to which node G belongs and the second device group to which node A belongs are not in the same local area network, that is, node G is an edge node. A user on node G subscribes to the uplink video data stream of node A (which serves as the media data stream), and obtaining the video data stream requires a long route. Node G determines node F in the first device group first, and then transmits the stream pulling request to node F. After receiving the stream pulling request from node G, node F searches locally for the video data stream of node A first. In case of not finding the video data stream of node A, node F initiates the stream pulling request to node A via the transmit proxy device 402 (e.g., node E as shown in Figure 4b). Node A receives the stream pulling request from node F, records locally that node F subscribes to node A, and then forwards the video data stream to node F via node E. Afterwards, node E forwards the video data stream to node G. As a final step, node G forwards the video data stream to the corresponding conference accessing terminal.

[0056] As shown in Figure b, a process of node H pulling the video data stream of node A is similar to that of node C pulling the video data stream of node A, except that the node E serves as an intermediate node. Details are not repeated herein. Herein the transmit proxy device 402, e.g., node E, may be determined through pre-configuration, or may be dynamically allocated, which is not limited herein. The interface machines for accessing an audio/video conference are grouped based the device group (SET) in architecture. In a case that an interface machine in a device group needs to pull a media data stream, the origin-pull device may be determined from the device group to which the interface machine belongs, and then the origin-pull device initiates the stream pulling

request to an interface machine (that is, a media source device) that generates the media data stream. Such design reduces a load of distribution pressure on the media source device greatly, and increases a quantity of participants of a conference room greatly. For example, it is assumed that each SET has 100 interface machines, the whole system for audio/video conferencing has 100 SETs, and each 8-core 16G interface machine can support forwarding 100 channels of media data streams. In a case that there is only one guest in a conference and the rest are audience, a maximum scale of the conference is $100 \text{ (machines)} * 100 \text{ (SETs)} * 100 \text{ (distribution channels)} = 1,000,000$ participants. In a case that there is more than one guest, different interface machines of different SETs would be allocated based on a principle of proximity in accessing, and a load of uplink forwarding would also be distributed. A conference of millions of participants can be easily supported.

[0057] Herein the technical solutions are related to the multiple device groups, each device group comprises multiple interface machines, and each interface machine is configured to connect the conference accessing terminal, so that the corresponding conference accessing terminal can access the conference room via such interface machine. The multiple interface machines play a role similar to that of a server in conventional technology, in which multiple conference accessing terminals access a conference room via such server. The conference accessing terminals are allocated to different interface machines, and hence the interface machines can share a load, which facilitates accessing of more conference accessing terminals and supports a large conference. Each device group further comprises the origin-pull device. The media data stream is generated by the second interface machine accessing the conference room. Taking the first device group as an example, the origin-pull device is capable to forward the media data stream corresponding to the first identifier to the first interface machine after receiving the stream pulling request from the first interface machine, because the stream pulling request comprises the first identifier. Thereby, the first interface machine can forward the media data stream to its corresponding conference accessing terminal. Accordingly, when each interface machine needs to pull the media data stream from the second interface machine, the second interface machine is only necessary to interact with the origin-pull device in each device group. The origin-pull devices share the load of data distribution on a media source device, e.g., the second interface machine, and interact with other interface machines, and the media source device is not necessary to interact with each interface machine, which greatly reduces the load of data distribution pressure on the media source device, e.g., the second interface machine. Therefore, the method for implementing audio/video conferencing can support more participants, even millions of participants. In addition, the load of data distribution during streaming media data is greatly reduced, and hence all participants (i.e., users) can enable

microphones or cameras at the same time without redirection, which render the conferencing experience smooth.

[0058] In the above solutions, operation and maintenance are simpler and are not limited by performances and costs of hardware, in comparison with conference architecture based on a multipoint control unit (MCU). A quantity of participants which can be actually supported is greatly increased, which facilitates implementing a super-large-scale conference.

[0059] An interface of a video conference implemented through the method and the system according to embodiments of the present disclosure may refer to Figure 5. Images of participants who enabling the video generation, a list of participants, icons of relevant functional keys (for example, a microphone key, a camera key, an administrator key, and an exit key) may be displayed in the interface. The icon of the microphone may be displayed along with a slash on it to indicate a participant who does not enable the microphone. Herein the system for audio/video conferencing system can implement a super-large-scale conference. For example, the maximum quantity of participants entering the conference is 50000, that is, 50000 participants can be managed.

[0060] Hereinafter the method for implementing audio/video conferencing is illustrated with respect to the room management subsystem 302 managing a conference room.

[0061] Reference is made to Figure 6. The room management subsystem 302 may comprise the foregoing multiple device groups 601 and a central server (RoomSvc) 602. Here the device group 601 also comprises the foregoing multiple interface machines, and is similar to the device group as shown in Figure 4b. In addition, each device group 601 further comprises an intra-group server (RoomSvcInSet), and the intra-group server is configured to assist the central server 602 for managing the conference room. The intra-group server is denoted by a circle in each device group 601 in Figure 6, for example, node B, node C, node D, node E, node F, and node G in Figure 6. As an example, the central server 602 is node A in Figure 6.

[0062] The intra-group server and the origin-pull device in a device group may be the same device, or may be different devices, which is not limited herein.

[0063] Generally, managing the conference room includes managing room information, messages notifying a change in a status, and pull of a list of participants, of the conference room. The room information includes, for example, a name, a creation time, participants, and a conference period of the conference room.

[0064] Herein the room management system has the central server and the intra-group servers. There is only one central server, and the intra-group servers are deployed based on SETs as the interface machines. Thereby, synchronizing the notification messages and pulling the list of participants can be implanted through the intra-group servers, hence a load of management on the central server is reduced, and a super-large-scale conference can be realized.

[0065] For the same conference room, all interface machines in a device group register the same RoomSvcInSet through a certain algorithm, and maintain a long connection with the RoomSvcInSet through heartbeats. Afterwards, the media data stream can be pulled in the conference room. The RoomSvcInSet would be utilized when the notification messages need to be synchronized (for example, when a status of a user changes, such as a user entering the conference room to attend the conference, a user exiting the conference room, a user enabling/disabling the microphone, a user enabling/disabling the video) or the list of participants needs to be pulled in the conference room.

[0066] First, a registration flow for entering a conference room to attend the conference is described. When an arbitrary interface machine, such as a third interface machine, in an arbitrary 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 transmitted by the third interface machine. The registration request may comprise a room identifier for indicating a conference room which is to be entered. In a case that the intra-group server in the third device group does not find information of the conference room locally, the intra-group server transmits the registration request to the central server. The central server records a second identifier of the intra-group server in the third device group into a first conference-registration list, and returns successful-registration information to the intra-group server in the third device group. The intra-group server in the third device group records a third identifier of the third interface machine into a second conference-registration list, and returns the information indicating successful registration to the third interface machine. Thereby, registration of the third interface machine is completed. Both the second identifier and the third identifier are unique identifiers of corresponding devices. As an example, the identifier may be an address. That is, the second identifier may be an address of the intra-group server, and the third identifier may be an address of the third interface machine. The third device group may be an arbitrary one of the multiple device groups, the third interface machine may be an arbitrary 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 embodiments, and the third interface machine may be the first interface machine or the second interface machine in the foregoing embodiments.

[0067] Reference is made to Figure 6 for an example. Node a may serve as the third interface machine, and the third interface machine requests to register for a conference, that is, a user corresponding to the conference accessing terminal connecting the third interface machine requests to enter a conference room to attend the conference. Node A may serve as the central server, and node B may be serve as the intra-group server of a device group to which the third interface machine belongs, for example, of the third device group. Node a may determine node B according to a certain rule (such as consistent hashing or scheduled allocation). Node a transmits 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, searches for node A, through a scheduling system (Scheduler). Node B transmits the registration request to node A. Node A receives the registration request transmitted by node B, records an address of node B into a registration list of the conference (i.e., the first conference-registration list stored at node A), and returns information indicating successful registration to node B. After receiving the information indicating successful registration returned by node A, node B records an address of node a into another registration list of the conference (i.e., the second conference-registration list stored at node B), and returns the information indicating successful registration to node a.

[0068] Herein after the successful registration, long connection may be maintained between the intra-group server and the central server and between the intra-group server and the registered interface machine, such as the third interface machine, so as to facilitate subsequent interaction therebetween. For example, node B may further transmit periodical heartbeats to node A, so as keep a channel of long connection with node A. After receiving the information indicating successful registration returned by the node B, the node a transmit periodical heartbeats to node B, so as to keep a channel of long connection with node B.

[0069] The intra-group server in the third device group may find the information of the conference room locally after searching for the information according to the registration request. In such case, the intra-group server may record the third identifier of the third interface machine into the second conference-registration list, and return information indicating successful registration to the third interface machine.

[0070] Reference is made to Figure 6 for another example. Node b may serve as the third interface machine, and the third interface machine requests to register for a conference, that is, a user corresponding to the conference accessing terminal connecting the third interface machine requests to enter a conference room to attend the conference. Node A may serve as the central server, and node B may be serve as the intra-group server of a device group to which the third

interface machine belongs, for example, of the third device group. Node b may determine node B according to a certain rule (such as consistent hashing or scheduled allocation). Node b transmits a registration request to node B. Node B receives the registration request from node b, determines that information of the conference room has been stored locally, records an address of node b in a registration list of the conference (i.e., the second conference-registration list stored at node B), and returns information indicating successful registration to node b. After receiving the information indicating successful registration returned by node B, node b transmits periodical heartbeats to keep a channel of long connection with node B.

[0071] When a user on an edge node (that is, a 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 such conference room, a process of the registration is similar except that forwarding via a transmit proxy device, such as node O, node P, and node Q as shown in Figure 6, is required.

[0072] When a status changes at the conference accessing terminal that accesses the conference room, for example, when a user enables or disables audio/video generation, other users in the conference room need to be notified of the change in the status. For example, the conference accessing terminal subject to the change is connected to the third interface machine in the third device group. In such case, the intra-group server in the third device group receives a notification message transmitted by the third interface machine, and forwards the notification message to the central server. The central server forwards the notification message to the intra-group server of each device group, and the intra-group server of each device group forwards the notification message to corresponding interface machine(s) which have successfully registered.

[0073] Figure 6 is further taken as an example. Node a may serve as the third interface machine, and a guest is on node a. Node A may serve as the central server, and node B may be serve as the intra-group server of a device group to which the third interface machine belongs, for example, of the third device group. In a case that the guest on node a enables his/her camera, that is, a status changes, node a transmits the notification message (msg-a-openvideo) indicating a change of the status to node B, and node B transmits the notification message to node A. Node A searches a local registration list (i.e., the first conference-registration list stored at node A), and transmits the notification message to all intra-group servers, for example, node B, node C, node D, node E, node F, and node G as shown in Figure 6. Node B, node C, node D, node E, node F, and node G receive the notification message, search respective local registration lists (i.e. the second conference-registration lists stored at the intra-group servers, respectively), and transmit 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 as shown in Figure 6.

[0074] A list of participants is maintained in audio/video conferencing, and the list records all participants. There are a lot of participants in a super-large-scale audio/video conference, which is implemented through the system according to embodiments of the present disclosure, and hence the local list of participants may be updated through page-by-page pulling. When an interface machine (e.g., the third interface machine) in a device group (e.g., the third device group) requests to pull the list of participants, the intra-group server in the third device group receives a list pulling request from the third interface machine, and the intra-group server in the third device group searches locally for the list of participants. In case of not finding the list of participants, the intra-group server in the third device group transmits a list pulling request to the central server. The intra-group server in the third device group receives the list of participants transmitted by the central server, and forwards the list of participants to the third interface machine.

[0075] Reference is made to Figure 6 for an example. Node a may serve as the third interface machine, and a user on node a requests the list of participants from node B. In such case, node B searches a local cache for the list. In case of not finding the list, node B initiates the list pulling request to node A. After receiving the list pulling request, node A returns the corresponding list of participants to node B. After caching the list of participants locally, node B returns the list of participants to node a. The list of participants may be a list of all participants, or may be a page, such as Page1, of the list of participants due to the page-by-page pulling.

[0076] The intra-group server in the third device group may find the list of participants locally after searching for the list, and the list of participants does not expire. In such case, the intra-group server in the third device group returns the list of participants to the third interface machine.

[0077] Reference is further made to the foregoing example. Node b as shown in Figure 6 may serve the third interface machine. In a case that another node has already requested the foregoing list of participants, e.g., Page1, before node b, there is the list of participants at node B, and hence it is not necessary to request the list of participants from node A. In such case, node b transmits the list pulling request to node B. Node B searches the local cache for Page1, finds Page1 which has not expired, and then returns Page1 directly to node b.

[0078] Second, a process of disaster recovery process is described. Since all data at RoomSvc comes from RoomSvcInSet(s) and the interface machine(s), it is convenient to perform automatic reconstruction based on RoomSvcInSet(s) after RoomSvc crashes. Similarly, all data at

RoomSvcInSet comes from the interface machine(s). When RoomSvcInSet crashes, automatic reconstruction can be performed based on the interface machine(s) in the device group (SET).

[0079] The method according to embodiments of the present disclosure relieves a load of management on the central server through the intra-group server(s), which facilitates implementation of super-large-scale conferences.

[0080] In conventional technology, audio routing based on energy is required when media data streams, such as audio data streams, generated in the conference room are transmitted through audio/video conferencing based on the SFU. In such scheme, a delay in a remote call hampers conferencing experience greatly. A large bandwidth is wasted and network performances of the conference accessing terminals degrade greatly when there are a lot of participants. Moreover, precise control (for example, whether a start and/or an end is determined based on signaling or based on a media data stream) is difficult for the SFU-based architecture in a scenario which requires interactive voice response (IVR) and precise recording in a conference. Further, in some scenarios having strict requirements, it is unacceptable that speech recording is subject to missing words or unwanted words. In a case that there is no centralized audio processing service, it is highly likely to record content which is not intended to record or miss content which should be recorded, which is absolutely unacceptable to strict users.

[0081] In view of the above, the system for audio/video conferencing according to embodiments of the present disclosure further comprises a mixer engine 303. The mixer engine 303 is an intermediate between the media transmission network 301 and a public switched telephone network (PSTN) 305, and serves as a bridge for communication between the two. The mixer engine 303 improves conferencing experience of a super-large-scale room greatly. The mixer engine 303 connects to the public switched telephone network 305 through a protocol, such as the Session Initiation Protocol (SIP) or the Real-Time Transport Protocol (RTP).

[0082] The mixer engine 303 comprises 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 transmits multiple acquired audio data streams to the mixer in the same device group, and the multiple audio data streams come from fourth interface machines in different device groups, respectively. The mixer selects a target audio data stream from the multiple audio data streams, and transmits a selection notification to a target interface machine connecting to the conference accessing terminal that generates the target audio data stream. The target interface machine transmits the target audio data stream to the mixer. The mixer performs audio mixing on the target audio data stream, and then forwards the audio data stream to the selector(s) corresponding to the

fourth interface machines, where the fourth interface machines comprise the target interface machine. The selector(s) corresponding to the fourth interface machines transmit the audio data stream after the audio mixing to the fourth interface machines. The fourth interface machine forwards the audio data stream after the audio mixing to other interface machines in the device group to which the fourth interface machine belongs.

[0083] The mixer and the selector are deployed based on the SET. The mixer and the selector of the same SET are deployed in the same available region to minimize a delay (where a transmission delay within the available area is capable to be controlled smaller than 2ms). In addition, audio data streams of the same conference are processed in one SET, and are concentrated into one mixer for audio mixing.

[0084] It is assumed that Figure 7 shows a real conference, in which three participants enable their microphones. Interface machines to which the three participants belong are node a, node f, and node g, and all other interface machines are for audience. In order to simplify the example, it is further assumed that two audio data streams are selected for each conference (in general practice, four to six audio data streams are selected for a real conference). In such case, the audio data stream is process as follows.

[0085] Node a, node f, and node g request the scheduling system, as shown in Figure 6, to allocate selectors (where the scheduling system allocates selectors in the same available area to audio data streams in the same conference room, in order to minimize a delay), and obtain selectors represented by node B, node C, and node D, respectively, as shown in Figure 7. Node a, node f, and node g forward the audio data streams to their corresponding nodes B, C, and D, respectively. Node B, node C, and node D cache the uplink audio data streams, and report values of energy of the audio data streams to the mixer. The mixer sorts the values of energy, and selects two audio data streams, which are from node f and node g, as the target audio data streams, and then notifies node C and node D. Node C and node D serve as the target interface machines. After receiving the selection notification, node C and node D forward the corresponding target audio data streams to the mixer (where a delay between the two operations is within 5ms). The mixer performs audio mixing on the received target audio data streams to obtain a mixed audio data stream, and then forwards the mixed audio data stream to node B, node C, and node D. Node B, node C, and node D transmit the mixed audio data stream to node a, node f, and node g. Then, nodes a, node f, and node g forward the the mixed audio data stream within their respective SETs.

[0086] Here a stream transmitted to the guest is different from a stream transmitted to the ordinary audience. An audio data stream from a guest should be removed from the mixed audio

data stream transmitted to the guest. Therefore, the selected interface machine receives two audio data streams. One audio data stream is forwarded to the guest who is speaking, and the other audio data stream is forwarded to the audience belong to such SET.

[0087] The audio data streams processed by the mixer engine 303 need to be forwarded to all participants in the conference room in real time, and hence the mixer engine 303 is a system for real-time media processing. In various scenarios, a system for asynchronous stream processing may be required. That is, after obtaining an audio data stream, the system does not need to transmit the audio data stream to a participants in the conference room, process the audio data stream for another purpose (for example, recording, porn detecting, and live streaming push). In such case, the real-time processing is not highly demanded. On such basis, as shown in Figure 3, a system for audio/video conferencing according to an embodiment of the present disclosure further comprises a bypass media processing system 304. The bypass media processing system 304 is configured to pull a media data stream generated in a conference room and process the media data stream by using a robot terminal. The robot terminal is as shown in Figure 3, and includes a live stream robot, a record robot, and an MRA robot.

[0088] Herein a function of bypass media processing is implemented though the robot terminal entering the conference room to pull a stream. Taking recording as an example, a user service background activates a recording task via an application programming interface (API), and a recording-task system instantiates a record robot, which simulates a participant, for entering a conference. The record robot pulls the media data streams of the conference to a local machine, mixes the streams, and transcodes the streams for recording. After the conference ends, the record robot uploads a recording file to a specified storage (or uploads fragments periodically).

[0089] Scenarios of porn detecting and live streaming push are similar to that of recording, and they only differ in logic of service. Bi-directional communication is required when the system is interconnected with a third-party conference device, and hence a process is slightly different from the recording scenario.

[0090] Reference is made to Figure 8. When connecting to a third-party conference device, a user service background activates an MRA (conference connector) robot. The MRA robot enters a conference ** (at such time, other participants can see this MRA robot), and the MRA robot is connected to the third-party conference device through a connection protocol (such as the SIP or H323 protocol). After pulling all media data streams of the conference ** to a local machine and mixing the streams, the MRA robot forwards the mixed stream to the third-party conference device through the connection protocol. The MRA robot converts a media data stream transmitted from

the third-party conference device into a stream complying a private protocol of the conference **, and forwards the stream to other participants in the conference room through the media transmission network. In Figure 8, 801 denotes mobile conference terminals, and 802 denotes an intelligent conference room for accessing the conference **.

[0091] In a case that the system for audio/video conferencing does not comprise the mixer engine, architecture as shown in Figure 9 may be used for implementing the interconnection between the data transmission network 301 and the public switched telephone network 305. That is, a function of the mixer engine is implemented by using a selector and a robot involved in the system for bypass media processing. The selector performs selection based on reported values of energy and a corresponding strategy of selection.

[0092] On a basis of the forgoing method for implementing audio/video conferencing, an apparatus for implementing audio/video conferencing is provided according to an embodiment of the present disclosure. Reference is made to Figure 10. The apparatus 1000 comprises a receiving unit 1001 and a forwarding unit 1002:

[0093] The receiving unit 1001 is configured to receive a stream pulling request from a first interface machine in a first device group, where the stream pulling request comprises a first identifier of a media data stream, the first device group is one of multiple device groups, each of which comprises an origin-pull device and multiple interface machines, each of the multiple interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access a conference room via said interface machine, and the first interface machine is one of the multiple interface machines in the first device group.

[0094] The forwarding unit 1002 is configured to forward the media data stream corresponding to the first identifier to the first interface machine to enable the first interface machine to forward the media data stream to the conference accessing terminal corresponding to the first interface machine, where the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the second device group is the one or another of the multiple device groups.

[0095] In an embodiment, the forwarding unit 1002 is configured to: search for the media data stream, which corresponds to the second interface machine, according to the first identifier; and forward the stream pulling request to the second interface machine in response to not finding the media data stream corresponding to the second interface machine.

[0096] The receiving unit 1001 is further configured to receive the media data stream, which is forwarded by the second interface machine according to the first identifier.

[0097] The forwarding unit 1002 is further configured to forward the media data stream to the first interface machine.

[0098] In an embodiment, the forwarding unit 1002 is further configured to forward the media data stream to the first interface machine in response to finding the media data stream corresponding to the second interface machine.

[0099] In an embodiment, the first device group and the second device group are in different local area networks. The forwarding unit 1002 is configured to: forward the stream pulling request to the second interface machine via a transmit proxy device; and receive the media data stream, which is forwarded by the second interface machine, via the transmit proxy device.

[0100] In an embodiment, the first interface machine determines the origin-pull device in the first device group according to a preset rule, and the preset rule configures that any media data stream being routed to interface machines in a device group is routed via an origin-pull device belonging to the device group.

[0101] In an embodiment, the apparatus further comprises a transmitting unit, a recording unit, and a returning unit.

[0102] The receiving unit 1001 is further configured to: receive a registration request transmitted by a third interface machine, when the third interface machine requests to enter the conference room, where the third device group is among the multiple device groups, and the third interface machine is one of the multiple interface machines in the third device group.

[0103] The transmitting unit is configured to, in response to determining that information of the conference room is not locally stored, transmit the registration request to a central server to enable the central server to: record a second identifier of an intra-group server in the third device group into a first conference-registration list, and return information indicating successful registration to the intra-group server in the third device group.

[0104] The recording unit is configured to record a third identifier of the third interface machine into a second conference-registration list.

[0105] The returning unit is configured to forward the information indicating successful registration to the third interface machine.

[0106] In an embodiment, the recording unit is further configured to record the third identifier of the third interface machine into the second conference-registration list, in response to determining that the information of the conference room is locally stored.

[0107] The returning unit is further configured to forward the information indicating successful registration to the third interface machine.

[0108] In an embodiment, the receiving unit 1001 is further configured to receive a notification message transmitted by the third interface machine, when a status of the conference accessing terminal connecting the third interface machine in the third device group changes.

[0109] The transmitting unit is configured to transmit the notification message to the central server to enable the central server to transmit the notification message to an intra-group server in each of the multiple device groups, where the intra-group server in each of the multiple device groups is configured to forward the notification message to interface machines that register to the conference room successfully.

[0110] In an embodiment, the receiving unit 1001 is further configured to receive a list pulling request of the third interface machine when the third interface machine requests to pull a list of participants;

[0111] The transmitting unit is configured to transmit the list pulling request to the central server, in response to determining that the list of participants is not locally stored.

[0112] The receiving unit 1001 is further configured to receive the list of participants transmitted by the central server.

[0113] The transmitting unit is configured to forward the list of participants to the third interface machine.

[0114] In an embodiment, the transmitting unit is further configured to: transmit the list of participants to the third interface machine, in response to determining that the list of participants is locally stored and does not expire.

[0115] In an embodiment, the apparatus further comprises a selecting unit and a mixing unit.

[0116] The transmitting unit is further configured to transmit, in response to the media data stream comprising audio data streams, the audio data streams to a mixer belonging to the same device group via a selector, where the mixer and the selector are in a mixer engine and belong to the

same device group, and the audio data streams are from fourth interface machines in different device groups, respectively.

[0117] The selecting unit is configured to select a target audio data stream from the multiple audio data streams through the mixer.

[0118] The transmitting unit is configured to transmit a selection notification to a target interface machine, which connects the conference accessing terminal generating the target audio data stream.

[0119] The receiving unit 1001 is further configured to receive the target audio data stream, which is transmitted by the target interface machine, via the mixer.

[0120] The mixing unit is configured to perform audio mixing on the target audio data stream.

[0121] The forwarding unit 1002 is further configured to forward the mixed audio data stream to a selector corresponding to the fourth interface machines, where the fourth interface machines comprise the target interface machine.

[0122] The transmitting unit is further configured to transmit the mixed audio data stream to the fourth interface machines via the selector corresponding to the fourth interface machines to enable the fourth interface machines to forward the mixed audio data stream to other interface machines in the different device groups to which the fourth interface machines belong.

[0123] In an embodiment, the apparatus further comprises a processing unit.

[0124] The processing unit is configured to pull, through a robot terminal, a media data stream generated in the conference room and process the media data stream generated in the conference room.

[0125] An electronic device for implementing an audio/video conferencing is further provided according to an embodiment of the present disclosure. The electronic device may be a terminal. A smartphone is taken as an example of the terminal.

[0126] Figure 11 is a block diagram of partial structure of a smartphone serving as a terminal device according to an embodiment of the present disclosure. As shown in Figure 11 the smartphone comprises components such as 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 (WiFi) module 1170, a processor 1180, and a power supply 1190. A person skilled in the art can understand that the structure of the mobile phone as shown in Figure 11 does not constitute a

limitation on the mobile phone. The mobile phone may comprise more or fewer components than what is depicted in the figure, some of the components may be combined, or the components may have different arrangement.

[0127] The memory 1120 may be configured to store a software program and a software module. The processor 1180 runs the software program and the software module stored in the memory 1120, so as to implement various functional applications and data processing of the smartphone. The memory 1120 may mainly comprise a program storing region and a data storing region. The program storing region 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 storing area may store data (such as audio data and an address book), which are created according to application of the smartphone. In addition, the memory 1120 may include a high speed random access memory, and may include a non-volatile memory, such as at least one magnetic disk storage device, a flash memory, or another volatile solid-state storage device.

[0128] The processor 1180 is a control core of the smartphone, and is connected to various parts of the entire smartphone via various interfaces and lines. The processor 1180 runs or executes the software program and/or the software module stored in the memory 1120, and invokes data stored in the memory 1120, so as to implement various functions of the smartphone and performs data processing, and thereby monitors the entire smartphone. Optionally, the processor 1180 may include one or more processing units. Preferably, an application processor and a modem may be integrated in the processor 1180 may integrate. The application processor is mainly configured to operate an operating system, a user interface, an application program, and the like. The modem mainly is configured to operate wireless communication. The foregoing modem may be alternatively not integrated into the processor 1180.

[0129] Herein the steps performed by the terminal in the foregoing embodiments may be implemented based on a structure shown in Figure 2.

[0130] The electronic device may further include a server. As shown in Figure 12, a server is further provided according to an embodiment of the present disclosure. Figure 12 is a schematic structural diagram of a server 1200 according to an embodiment of the present disclosure. The server 1200 may vary greatly due to different configurations or performances, and may comprise one or more central processing units (CPU) 1222 (for example, one or more processors), a memory 1232, and one or more storage media 1230 (for example, one or more mass storage devices) that store application programs 1242 or data 1244. The memory 1232 and the storage medium 1230

may be transient or persistent storages. A program stored in the storage medium 1230 may comprise one or more modules (not depicted), and each module may comprise a series of operation instructions for the server. Moreover, the CPU 1222 may be configured to communicate with the storage medium 1230, and execute the series of operation instructions in the storage medium 1230 on the server 1200.

[0131] 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, for example, Windows ServerTM, Mac OS XTM, UnixTM, LinuxTM, or FreeBSDTM.

[0132] Herein the central processing unit 1222 in the server 1200 may perform the following steps.

[0133] A stream pulling request is received from a first interface machine in the first device group where the stream pulling request comprises a first identifier of a media data stream, the first device group is one of multiple device groups, each of which comprises a respective origin-pull device and multiple interface machines, each of the multiple interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access a conference room via said interface machine, and the first interface machine is one of the multiple interface machines in the first device group.

[0134] The media data stream corresponding to the first identifier is forwarded to the first interface machine to enable the first interface machine to forward the media data stream to the conference accessing terminal corresponding to the first interface machine, where the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the second device group is the one or another of the multiple device groups.

[0135] A computer-readable storage medium is further provided according to an embodiment of the present disclosure. The computer-readable storage medium is configured to store program codes, and the program codes when executed on an electronic device perform the method for implementing audio/video conferencing according to any foregoing embodiment.

[0136] A computer program product or a computer program is further provided according to an embodiment of the present disclosure. The computer program product or the computer program comprises computer instructions, and the computer instructions are stored in a computer

readable storage medium. A processor of an electronic device reads the computer instructions from the computer readable storage medium and executes the computer instructions to cause the electronic device to perform the method for implementing audio/video conferencing according to any foregoing embodiment.

[0137] Herein the terms "first", "second", "third", "fourth", and the like (if any) in the specification and the drawings are intended to distinguish between similar objects rather than describing a specific order or sequence. It is to be understood that these terms are interchangeable in a proper circumstance, so that the embodiments described herein can be implemented by using an order other than the one 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 comprises a list of steps or units is not necessarily limited to those steps or units, but may comprise other steps or units not expressly listed or inherent to such process, method, product, or device.

[0138] In several embodiments of the present disclosure, the disclosed system, apparatus, and method may be implemented in other manners. For example, the described apparatus embodiments are only exemplary. For example, the division of the units is only based on logical functions and there may be another manner of division in practice. For example, multiple units or components may be combined or integrated into another system, or some features may be ignored or not performed. In addition, the shown or discussed mutual couplings, direct couplings, or communication connections may be implemented through some interfaces. The indirect couplings or communication connections among apparatuses or units may be electrical, mechanical, or implemented in other forms.

[0139] The units described as separate parts may or may not be physically separate. Parts displayed as units may or may not be physical units, and may be located in one position or distributed among network units. Some or all of the units may be selected according to an actual requirement to achieve an objective of the solutions according to embodiments.

[0140] In addition, function units in embodiments of the present disclosure may be integrated into one processing unit. Each of the units may exist alone physically, 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.

[0141] When the integrated unit is implemented in the form of the 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 understanding, all or a part, which contributes

to the related art in essence, of the technical solutions of the present disclosure may be implemented in a form of a software product. The computer software product is stored in a storage medium, and comprises 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 method according to embodiments of the present disclosure. The foregoing storage medium includes any medium that can store program codes, 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.

[0142] The foregoing embodiments are merely intended for describing the technical solutions of the present disclosure, and not for limiting the present disclosure. Although the present disclosure is described in detail with reference to the foregoing embodiments, a person skilled in the art can appreciate that modifications may be made to the technical solutions described in the foregoing embodiments, or equivalent replacements may be made to a part of the technical features of the technical solutions. These modifications or replacements would not make the essence of corresponding technical solutions depart from the spirit and scope of the technical solutions in embodiments of the present disclosure.

CLAIMS

What is claimed is:

1. A method for implementing audio or video conferencing, comprising:

receiving, by an origin-pull device in a first device group, a stream pulling request from a first interface machine in the first device group, wherein the stream pulling request comprises a first identifier of a media data stream, the first device group is among a plurality of device groups, each of which comprises a respective origin-pull device and a plurality of interface machines, each of the plurality of interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access a conference room via said interface machine, and the first interface machine is one of the plurality of interface machines in the first device group; and

forwarding, by the origin-pull device in the first device group, the media data stream corresponding to the first identifier to the first interface machine to enable the first interface machine to forward the media data stream to the conference accessing terminal corresponding to the first interface machine, wherein the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the second device group is among the plurality of device groups.

2. The method according to claim 1, wherein forwarding, by the origin-pull device in the first device group, the media data stream corresponding to the first identifier to the first interface machine comprises:

searching for the media data stream, which corresponds to the second interface machine, according to the first identifier; and

in response to not finding the media data stream corresponding to the second interface machine,

forwarding the stream pulling request to the second interface machine;

receiving the media data stream, which is forwarded by the second interface machine according to the first identifier; 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 finding the media data stream corresponding to the second interface machine.

4. The method according to claim 2, wherein the first device group and the second device group are in different local area networks,

wherein forwarding the stream pulling request to the second interface machine comprises:

forward the stream pulling request to the second interface machine via a transmit proxy device; and

wherein receiving the media data stream, which is forwarded by the second interface machine according to the first identifier comprises:

receiving the media data stream, which is forwarded by the second interface machine, via the transmit proxy device.

5. The method according to any one of claims 1 to 4, wherein the first interface machine determines the origin-pull device in the first device group according to a preset rule, and the preset rule configures that any media data stream being routed to interface machines in a device group is routed via an origin-pull device belonging to the device group.

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, a registration request transmitted by a third interface machine, when the third interface machine requests to enter the conference room, wherein the third device group is among the plurality of device groups, and the third interface machine is among the plurality of interface machines in the third device group;

in response to the intra-group server in the third device group determining that information of the conference room is not locally stored,

transmitting, by the intra-group server in the third device group, the registration request to a central server to enable the central server to:

record a second identifier of an intra-group server in the third device group into a first conference-registration list, and

return information indicating successful registration to the intra-group server in the third device group;

recording, by the intra-group server in the third device group, a third identifier of the third interface machine into a second conference-registration list; and

forwarding, by the intra-group server in the third device group, the information indicating successful registration to the third interface machine.

7. The method according to claim 6, further comprising: in response to the intra-group server in the third device group determining that the information of the conference room is locally stored,

recording, by the intra-group server in the third device group, the third identifier of the third interface machine into the second conference-registration list, and

forwarding, by the intra-group server in the third device group, the information indicating successful registration 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, a notification message transmitted by the third interface machine, when a status of the conference accessing terminal connecting the third interface machine in the third device group changes; and

transmit, by the intra-group server in the third device group, the notification message to the central server to enable the central server to transmit the notification message to an intra-group server in each of the plurality of device groups, wherein the intra-group server in each of the plurality of device groups is configured to forward the notification message to interface machines that register to the conference room successfully.

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

receiving, by the intra-group server in the third device group, a list pulling request of the third interface machine when the third interface machine requests to pull a list of participants;

in response to the intra-group server in the third device group determining that the list of participants is not locally stored,

transmitting, by the intra-group server in the third device group, the list pulling request to the central server;

receiving, by the intra-group server in the third device group, the list of participants transmitted by the central server; and

forwarding, by the intra-group server in the third device group, the list of participants to the third interface machine.

10. The method according to claim 6, further comprising: in response to the intra-group server in the third device group determining that the list of participants is locally stored and does not expire,

transmitting the list of participants to the third interface machine.

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

transmitting, in response to the media data stream comprising audio data streams, the audio data streams to a mixer belonging to the same device group via a selector, wherein the mixer and the selector are in a mixer engine and belong to the same device group, and the audio data streams are from fourth interface machines in different device groups, respectively;

selecting a target audio data stream from the plurality of audio data streams through the mixer;

transmitting a selection notification to a target interface machine, which connects the conference accessing terminal generating the target audio data stream;

receiving the target audio data stream, which is transmitted by the target interface machine, via the mixer;

performing audio mixing on the target audio data stream;

forwarding the mixed audio data stream to a selector corresponding to the fourth interface machines, wherein the fourth interface machines comprise the target interface machine; and

transmitting the mixed audio data stream to the fourth interface machines via the selector corresponding to the fourth interface machines to enable the fourth interface machines to forward the mixed audio data stream to other interface machines in the different device groups to which the fourth interface machines belong.

12. The method according to any one of claims 1 to 4, wherein the method further comprises:

pulling, through a robot terminal, a media data stream generated in the conference room and process the media data stream generated in the conference room.

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

the receiving unit is configured to receive a stream pulling request from a first interface machine in a first device group, wherein the stream pulling request comprises a first identifier of a media data stream, the first device group is among a plurality of device groups, each of which comprises an origin-pull device and a plurality of interface machines, each of the plurality of

interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access a conference room via said interface machine, and the first interface machine is one of the plurality of interface machines in the first device group; and

the forwarding unit is configured to forward the media data stream corresponding to the first identifier to the first interface machine to enable the first interface machine to forward the media data stream to the conference accessing terminal corresponding to the first interface machine, wherein the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the second device group is among the plurality of device groups.

14. A system for audio or video conferencing, 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, and the data transmission network comprises a plurality of device groups, each of which comprises a plurality of interface machines;

each of the plurality of interface machines is configured to connect a conference accessing terminal to enable the conference accessing terminal to access the conference room via said interface machine;

each of the plurality of device groups comprises an origin-pull device, the origin-pull device in a first device group is configured to:

receive a stream pulling request from a first interface machine in the first device group, wherein the stream pulling request comprises a first identifier of a media data stream; and

forward the media data stream corresponding to the first identifier to the first interface machine;

the media data stream is obtained by the origin-pull device in the first device group from a second interface machine in a second device group, the conference accessing terminal that generates the media data stream accesses the conference room via the second interface machine, and the first device group and the second device group each is among the plurality of device groups;

the first interface machine is one of the plurality of interface machines in the first device group, and the first interface machine is configured to forward the media data stream to the conference accessing terminal corresponding to the first interface machine; and

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

15. An electronic device for implementing audio or video conferencing, comprising a processor and a memory, wherein:

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

the processor is 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 a program code, wherein:

the program code when executed on an electronic device causes the electronic device to perform the method according to any one of claims 1 to 12.

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