

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 December 2011 (22.12.2011)

(10) International Publication Number  
**WO 2011/159505 A2**

(51) International Patent Classification:  
*H04N 7/26* (2006.01) *G06F 15/16* (2006.01)  
*H04N 7/15* (2006.01)

(21) International Application Number:  
PCT/US2011/039245

(22) International Filing Date:  
6 June 2011 (06.06.2011)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
12/819,092 18 June 2010 (18.06.2010) US

(71) Applicant (for all designated States except US): **MICROSOFT CORPORATION** [US/US]; One Microsoft Way, Redmond, Washington 98052-6399 (US).

(72) Inventors: **LEE, Ming-Chieh**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **QIAN, Tin**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US). **MOORE, Timothy**; c/o Microsoft Corporation, LCA - International Patents, One Microsoft Way, Redmond, Washington 98052-6399 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ,

CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

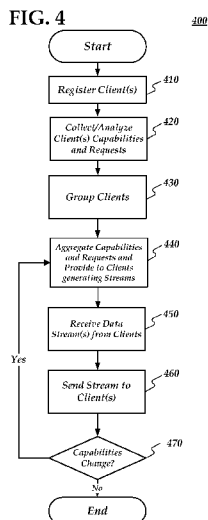
**Declarations under Rule 4.17:**

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

**Published:**

- without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: COMBINING MULTIPLE BIT RATE AND SCALABLE VIDEO CODING



(57) Abstract: Video streams are generated using a combination of Multiple Bit Rate (MBR) encoding and Scalable Video Coding (SVC). Capabilities and requests of the clients are used in determining the video streams to generate as well as what video streams to deliver to the clients. The clients are placed into groups based on a resolution capability of the client. For each resolution grouping, MBR is used for generating spatial streams and SVC is used for generating temporal and quality streams.

WO 2011/159505 A2

## COMBINING MULTIPLE BIT RATE AND SCALABLE VIDEO CODING

### BACKGROUND

[001] Clients participating in video conferencing typically have different processing capabilities and available bandwidth. In some situations, a video conferencing client may have a fluctuating amount of bandwidth while other clients have a stable amount of bandwidth. Some clients may not be able to receive the same quality video stream or encode video streams as other clients. For example, some clients may be able to encode and play back at 1080p resolution (1920x1080 pixels per frame) with a frame rate of 30 frames per second (fps), while other clients may only be able to encode and receive at CIF resolution (352x288) with a 15fps rate.

### SUMMARY

[002] This Summary is provided to introduce a selection of concepts in a simplified form that are further described below in the Detailed Description. This Summary is not intended to identify key features or essential features of the claimed subject matter. Nor is this Summary intended to be used to limit the claimed subject matter's scope.

[003] Video streams are generated using a combination of Multiple Bit Rate (MBR) encoding and Scalable Video Coding (SVC). Capabilities and requests of the clients are used in determining the video streams to generate as well as what video streams to deliver to the clients. The clients are placed into groups based on a resolution capability of the client. For each resolution grouping, MBR is used for generating spatial streams and SVC is used for generating temporal and quality streams.

[004] Both the foregoing general description and the following detailed description provide examples and are explanatory only. Accordingly, the foregoing general description and the following detailed description should not be considered to be restrictive. Further, features or variations may be provided in addition to those set forth herein. For example, embodiments may be directed to various feature combinations and sub-combinations described in the detailed description.

### BRIEF DESCRIPTION OF THE DRAWINGS

[005] The accompanying drawings, which are incorporated in and constitute a part of this disclosure, illustrate various embodiments of the present invention. In the drawings:

[006] FIG. 1 is a block diagram of an operating environment;

[007] FIG. 2 is a diagram illustrating different clients generating different streams that combine MBR and SVC in a video conference environment;

[008] FIG. 3 is a diagram illustrating combining SBC and MBR during video conferencing;

[009] FIG. 4 is a flow chart illustrating a video conference system that combines MBR and SVC in video streams;

[010] FIG. 5 is a flow chart illustrating the generation of streams that combines MBR and SVC; and

[011] FIG. 6 is a block diagram of a system including a computing device.

#### DETAILED DESCRIPTION

[012] The following detailed description refers to the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the following description to refer to the same or similar elements. While embodiments of the invention may be described, modifications, adaptations, and other implementations are possible. For example, substitutions, additions, or modifications may be made to the elements illustrated in the drawings, and the methods described herein may be modified by substituting, reordering, or adding stages to the disclosed methods. Accordingly, the following detailed description does not limit the invention. Instead, the proper scope of the invention is defined by the appended claims.

[013] FIG. 1 is a block diagram of an operating environment. The operating environment may comprise a video conference environment 100. Video conference environment 100 may comprise a variety of clients having a variety of capabilities, such as an intranet client 110 and a video conference server 120 connected via an intranet network connection 115. Video conference environment 100 may further comprise a first Internet client 130, a second Internet client 140, and a third Internet client 150. Each of the Internet clients may be connected to video conference server 120 via the Internet 160. Video conference server 120, intranet client 110, first Internet client 130, second Internet client 140, and third Internet client 150 may each comprise a computing device 600, described below in greater detail with respect to FIG. 6.

[014] Video conference environment 100 may be used when multiple clients want to share data streams amongst each other. Each client may connect to video conference server 120. Video conference server 120 may maintain a list of which clients are connected and each client's capabilities. The capabilities may be provided by the clients participating in the video conference and/or obtained by the video conference server. The capabilities for each of the clients may include all or a portion of the following, but are not limited to the: processing power, resolution capability, frame rate, bitrate, connection speed, available bandwidth, camera being used, and the like of a client. The capabilities may be different for some/all of

the clients. For example, some clients may be able to encode and play back at lower resolutions (e.g. CIF resolution (352x288) with a rate of 15 frames per second (fps) rate); some at medium resolutions (e.g. VGA resolution (640x480 pixels per frame) with a frame rate of 30 fps); and others at higher resolutions (e.g. 720p and 1080p with a 30fps rate). Additionally, some of the clients may be able to encode and play back a larger number of streams.

**[015]** Video conference server 120 may periodically update any changes in the client's capabilities. For example, video conference server 120 may either periodically automatically analyze or manually analyze network connections and determine that a client has more bandwidth available than previously determined. Video conference server 120 may also receive a message from a client that indicates a change in capability and store that information for use in facilitating the generation of the streams and the data sharing. Video conference server 120 may also be configured for determining which clients are responsible for generating data streams and request the generated data streams to be sent to video conference server 120 and/or some/all of the connected clients.

**[016]** Video conference server 120 is configured to aggregate client capabilities into different groups based on their resolution capability. Each group may include a resolution range that comprises one or more different resolutions and/or a range that consists of only one resolution. For example, there may be a group for 1080p clients; a group for 720p clients and 480p clients; and another group for CIF clients. The clients may also provide video conference server 120 with their desired preferences for viewing and playing video (e.g. full screen mode, full scale mode, small-window mode, bitrate and frame rate).

**[017]** Video conference server 120 aggregates this information and sends it to each client that is generating streams. For example, there may be CIF, VGA, and HD requests from the clients to the video conference server 120 with CIF's bitrate range of 100 to 250Kbps with frame rate ranging from 10 to 30fps, VGA of 500 to 700K with 15 to 30fps, and HD of 1 to 1.5Mbps with 30fps. In this example, video conference server 120 sends the following information: (CIF, 150K to 250K, 10 to 30), (VGA, 500K to 700K, 15 to 30), and (720p, 1M to 1.5M, 30) to each of the clients that are requested to generate streams (e.g. active speaker and/or subscribed streams).

**[018]** Each client uses this information and its own capability to generate video streams. Video streams are encoded using a combination of Multiple Bit Rate (MBR) coding and Scalable Video Coding (SVC). Scalable Video Coding (SVC) is an extension of the H.264

standard. SVC allows for the transmission and decoding of partial bit streams to provide video services with lower temporal or spatial resolutions or reduced fidelity. SVC, however, can have high encoding and decoding complexity. Among the various scalability modes described by SVC, the spatial and quality scalabilities are the most complex modes to encode and decode. As such, not all of the clients in a video conference are configured to efficiently perform these operations. Most clients, however, are able to efficiently perform other SVC encoding/decoding that is not as complex. Further, the spatial scalability of SVC may not be very efficient when the resolution range is large. MBR encoding, however, is more computationally efficient at spatial scalability. Instead of using only SVC for the video encoding/decoding, aspects of SVC are combined with aspects of a MBR encoding process to lower the complexity of the encoding/decoding of the video. MBR refers to a data stream in which the same content is encoded at several resolutions and/or frame rates to optimize content delivery to one or more clients.

**[019]** Generally, the clients that are grouped within the higher resolution groups are capable of generating more streams than the clients that are grouped in the lower resolution groups. According to one embodiment, each client generating streams uses MBR for spatial scalability and SVC is used for temporal and quality scalability.

**[020]** FIG. 2 is a diagram illustrating different clients generating different streams that combine MBR and SVC in a video conference environment 100.

**[021]** Each client may generate one or more data streams to send to video conference server 120 and/or some other destination. Each client may have an audio/video input such as a webcam and/or microphone connected to the client. The input may be used in the generation of one or more video streams (that may include audio) and encoding of the video stream(s) before sending it to video conference server 120 and/or some other destination.

**[022]** For example, client 130 may encode a video stream as a series of video frames comprising a person currently talking. A video stream may be encoded as a series of video frames wherein each frame may comprise a single image. Each frame may be represented by bits of data. The video stream may be encoded so that not every bit of data in every frame needs to be sent in order to represent the source of the video stream. Different frame types may be used to encode and/or compress the video stream. Frame types may comprise I-frames, P-frames, and B-frames. I-frame stands for Intra-Frame, and may comprise a frame that may be decoded by itself without reference to any other frame in a video stream. P-frame stands for Predicted Frame and may be decoded by referencing at least one previous frame in

the video stream sequence. B-frame stands for Bidirectional Predicted Frame and may be decoded by referencing at least one previous frame and at least one subsequent frame in the video stream. A client may generate streams that combine MBR and SVC for more than one resolution. For example, client 130 may generate separate streams for 1080p, 720p, 480p, and CIF.

**[023]** Video conference server 120 is configured to determine the capabilities of the clients and direct the generated streams to the appropriate client(s). For example, video conference server 120 may receive streams of a current speaker that are generated by client 130 at different resolutions and deliver them to the requesting clients. In the current example, video from client 130 is delivered to client 110 at a CIF resolution, client 140 at a 720p resolution, and client 150 at a 480p resolution. Different frame rates/bit rates may be delivered to the clients.

**[024]** As discussed above, capabilities and requests of the clients are determined by video conference server 120. Generally, the requests relate to the desired manner in which to view the video sources and the capabilities relate to the capability of the device for displaying and encoding/decoding video. In the following example that is intended to be non-limiting, assume that client 110 is a CIF client; client 130 is a 1080p client; client 140 is a 720p client; and client 150 is a 480p client. Video conference server 120 receives the requests and capabilities from each of the clients and aggregates this information before sending it out to the clients that are requested to generate streams. For example, each client may send a subscription request to video conference server 120 that includes the desired viewing parameters as well as capabilities of the client.

**[025]** The subscription request may include requested parameters such as a source client, a resolution, a bitrate, and/or a frame rate. For example, client 110 may request a subscription to a video stream (230) from client 130 at a CIF resolution at 15fps and a bitrate of 150Kbps. A client may also wish to see video streams from all participating clients in the video conference. For example, client 140 may participate in a four client video conference and request to subscribe to video streams (210, 230, 240, 250) showing all four participating clients in 720p with a frame rate ranging from 15fps to 30ps and a bitrate ranging from 1Mbps to 1.5Mbps. Client 150 may request a subscription to a video stream (210, 230) from client 110 and client 130 shown in 480p using frame rates between 15fps and 30fps and a bitrate ranging from 500Kbps to 700Kbps. As illustrated, client 130 has requested to subscribe to

video streams (210, 240, 250) from client 110, client 140, and client 150 shown in 1080p at 30fps and a bitrate of 10Mbps.

**[026]** Video conference server 120 aggregates the requests and sends this information to each of the clients that are to generate video streams. In the current example, there are CIF, VGA, and HD requests from the clients to video conference server 120 with CIF's bitrate range of 100 to 250Kbps with frame rate ranging from 10 to 30fps, VGA of 500 to 700K with 15 to 30fps, and HD of 1 to 1.5Mbps with 30fps. MCU would then send (CIF, 150K, 15), (VGA, 500K to 700K, 15 to 30), (720p, 1M to 1.5M, 15 to 30), and (1080p, 10M, 30) to all the clients that are requested to generate streams (active speaker and/or subscribed ones).

**[027]** Each client uses this information and its own capability to generate MBR and SVC streams. According to one embodiment, each client decides how many MBR "groups" to generate. For each MBR group generated there are multiple SVC layers (one base + enhancement layers). Each SVC layer (within each MBR group) is sent as a separate RTP stream. For example, CIF client 110 may be requested to generate a CIF MBR stream and a CIF SVC stream resulting in 4 RTP streams (210). VGA client 150 may be requested to generate 480p streams plus any CIF streams (250). Some VGA clients may be able to generate all of the 480p and CIF SVC streams that result in 10 RTP streams, whereas other clients may only be able to generate a portion of the SVC layers thereby resulting in a lower number of streams generated. Client 140 may be requested to send 720p, 480p, and CIF SVC streams, resulting in a total of 16 RTP streams (240). Client 130 may be requested to send 1080p + 720p + 480p + CIF SVC streams, resulting in 22 RTP streams (230). For example, 1080p 30fps SVC stream with 6 layers (3 temporal and 2 quality scalability layers); 720p 30fps SVC stream with 6 layers; 480p 30fps SVC stream with 6 layers; and CIF 15fps SVC stream with 4 layers (2 temporal and 2 quality).

**[028]** Client 130 may send the streams 230 that it generates to video conference server 120. Client 140 may send the streams 240 that it generates to video conference server 120. Client 150 may send streams 250 that it generates to video conference server 120 and client 110 may send the video streams 210 it generates to video conference server 120. Video conference server 120 directs the received streams to the requested destination(s). Alternatively, a client may send all or a portion of the streams to the video conference server 120 and all or a portion of the streams directly to other destinations.

**[029]** Each client that is generating streams determines the streams to generate based on the received information from video conference server 120 that includes the SVC and MBR

scalability types, range or number of MBR streams and SVC layers, and bitrate for each layer based on the client's capability. According to one embodiment, the bitrate and frame rate ranges are associated with each resolution and the number of streams that are generated are based on the requests from the clients.

**[030]** The determined stream configurations may change during the video conference. The video conference server 120, or client, may send updated information at any time during the video conference to allow the other clients adapt to the change of conditions/configurations of the clients (such as bandwidth, video viewing mode of a particular stream from big/full screen to sub-window mode, or the addition/removal of clients). According to one embodiment, when there is a status change of a client, the client notifies the video conference server 120. Video conference server 120 then aggregates the information and sends the updated info to the current senders that are being requested to send out streams.

**[031]** FIG. 3 is a diagram illustrating combining SVC and MBR during video conferencing. As illustrated, video conference system 300 comprises a video conference server 120 and a variety of clients, such as client 301, client 303, and client 304.

**[032]** Clients in video conference system 300 may connect with different bandwidths and/or different video encoding and/or decoding capabilities. Each client in video conference system 300 may register with video conference server 120 and establish capabilities such as available bandwidth, a maximum encoding resolution, a maximum decoding resolution, and available data encoding and/or decoding algorithms. For example, client 301 may register with video conference server 120 and establish an available bandwidth of 2 Mbps with capability to encode/decode a video stream with VGA (640x480) resolution at 30 fps and CIF (352x288) resolution at 15 fps. Client 303 may register with video conference server 120 and establish an available bandwidth of 150 Kbps with capability to encode/decode a video stream with 720p (1280x720) resolution at 30 fps, VGA (640x480) resolution at 30 fps, and CIF (352x288) resolution at 15 fps. Client 304 may register with video conference server 120 and establish an available bandwidth of 200 Kbps with capability to encode/decode a video stream only at CIF (352x288) resolution at 15 fps.

**[033]** Clients may be capable of encoding a video stream they may be incapable of decoding, and vice versa. For example, client 304 may be capable of receiving and decoding VGA resolution video streams, but may only be able to encode a CIF resolution video stream. In this scenario, client 301 and client 303 may only encode streams at VGA resolution that

video conference server 120 may send to client 304. The CIF resolution encoded video stream from client 304 may be sent to client 301 and client 303.

**[034]** After a video conference call is established and the capabilities for each client are registered with video conference server 120, video conference server 120 may communicate an encoding request to each client. For example, client 301 and client 303 may be requested to each encode two video resolutions of video streams that combine MBR and SVC and send them to video conference server 120 - a first version of the video stream may be encoded as streams with VGA resolution at 30 fps and a second version of the video stream may be encoded as streams with CIF resolution at 15 fps. Client 304 may be requested to encode one set of video streams with CIF resolution at 15 fps. According to one embodiment, the encoding request is placed within a message that includes the aggregated information created from the requests and capabilities of each of the clients.

**[035]** Video conference server 120 may then receive the stream(s) that are generated from each client and choose at least one received stream to send to other client(s) based on the registered and/or determined capabilities of each client. For example, video conference server 120 may determine that client 304 is only capable of decoding a CIF resolution video stream while client 301 and client 303 are capable of decoding both VGA resolution and CIF resolution video streams. Video conference server 120 may then send the CIF resolution stream generated by client 304 to client 301 and client 303. Video conference server 120 may send the VGA resolution stream generated by client 301 to client 303 and the CIF stream generated by client 301 to client 304. Similarly, video conference server 120 may send the VGA resolution stream from client 303 to client 301 and the CIF stream from client 303 to client 304.

**[036]** Video conference server 120 may determine whether each client has sufficient bandwidth to receive a video stream and may choose a different stream based on that determination. Consistent with further embodiments of the invention, video conference server 120 and/or one of the clients such as client 303 may periodically re-evaluate the amount of bandwidth available. For example, video conference server 120 may determine that bandwidth available to client 303 has dropped from 150 Kbps to 75 Kbps and may begin sending the CIF resolution video stream instead of the VGA resolution video stream from client 301 to client 303.

**[037]** Video conference server 120 may also periodically re-evaluate each client's encoding/decoding capabilities, and dynamically alter what encoding algorithms are requested

for video streams for each client. The re-evaluation may be performed automatically at predetermined times and/or manually. For example, client 304 may be the only client receiving the CIF resolution video streams from client 301 and client 303. If client 304 drops out of the video conference call, video conference server 120 may request that client 301 and client 303 stop encoding and sending a CIF resolution version of their respective video streams. According to one embodiment, this updated information is sent to each of the clients generating streams as the aggregated information as described above.

**[038]** Video conference server 120 may instruct at least one of the clients associated with the video conference call to encode a video stream determined to be in-use by one of the other clients associated with the video conference call. For example, video conference server 120 may determine that client 304 is currently using CIF decoding; video conference server 120 may instruct client 303 to encode a CIF resolution stream to send to video conference server 120 that may be relayed to second Internet server 304.

**[039]** Video conference server 120 may alter a video stream before sending it to a client. For example, client 301 may be capable of encoding and sending video streams at a 720p, VGA, and CIF resolutions and have a large, stable amount of bandwidth available. Client 303 may be capable of receiving and decoding video streams at a 720p, VGA, and CIF resolutions, but have a highly variable amount of bandwidth available. As the bandwidth available to client 303 reduces, video conference server 120 may alter the video stream relayed from client 301 to client 303 as best suited to the available amount of bandwidth.

**[040]** The choice of streams sent to each client may depend on each client's viewing preferences as well as the client's capabilities. For example, client 303 may be configured to view the other two participants in video conference system 300 side by side, simultaneously, while client 304 may be configured to only show a current speaker. In that case, video conference server may send the VGA resolution stream from client 301 and the CIF resolution stream from client 304 to client 303. Client 304 may alternately receive the CIF resolution streams from client 301 and client 303 as each of those clients speak on the video conference call.

**[041]** Referring now to FIGURES 3-4, illustrative processes for combining MBR and SVC for a video conference will be described. When reading the discussion of the routines presented herein, it should be appreciated that the logical operations of various embodiments are implemented (1) as a sequence of computer implemented acts or program modules running on a computing system and/or (2) as interconnected machine logic circuits or circuit

modules within the computing system. The implementation is a matter of choice dependent on the performance requirements of the computing system implementing the invention. Accordingly, the logical operations illustrated and making up the embodiments described herein are referred to variously as operations, structural devices, acts or modules. These operations, structural devices, acts and modules may be implemented in software, in firmware, in special purpose digital logic, and any combination thereof.

**[042]** FIG. 4 is a flow chart illustrating a video conference system that combines MBR and SVC in video streams.

**[043]** After a start block, the operation flows to operation 410, where at least one client is registered. For example, a client may be operatively connected to computing device 600 via a network connection and may request participation in a video conference call hosted by computing device 600.

**[044]** Moving to operation 420, capabilities and requests of each registered client are collected and/or analyzed. For example, computing device 600 may determine an amount of bandwidth available to each client. Computing device 600 may also determine data processing capabilities for each client, such as video encoding and/or decoding capability and/or data processing speed. Video encoding/decoding capability may include a resolution capability, such as a maximum resolution capability.

**[045]** A subscription request may also be received from at least one client in the video conference. For example, computing device 600 may receive a subscription request from one client for a VGA encoded, 30 fps stream from a second client. The request may include information such as a desired resolution, bitrate and frame rate. Consistent with embodiments of the invention, computing device 600 may receive multiple subscription requests from any and/or all of the clients participating in the video conference. For example, one client may request lower resolution versions of all the other participating clients; another client may request a high resolution version from whichever client may be designated an active speaker in the video conference and low resolution versions from some or all of the other clients.

**[046]** Moving to operation 430, the clients are grouped based on their capabilities. According to one embodiment, the clients are grouped according to the resolution they support. For example, clients may be placed into groups relating to standard resolution formats (e.g. 1080p, 720p, VGA, CIF, and the like). Each group may include one or more clients that support a resolution range that comprises one or more different resolutions and/or

a range that consists of only one resolution. For example, there may be a group for 1080p clients; a group for 720p clients and 480p clients; and another group for CIF clients.

**[047]** Flowing to operation 440, the capabilities and requests are aggregated and then sent to at least one client that is requested to generate video streams. The aggregated information may be represented in different ways, such as by creating a vector index having entries of resolution, bitrate range and frame rate range. For example, when there are CIF, VGA, and HD requests from the clients having CIF's bitrate range of 100 to 250Kbps with frame rate ranging from 10 to 30fps, VGA of 500 to 700K with 15 to 30fps, and HD of 1 to 1.5Mbps with 30fps, the following vector index may be generated: (CIF, 150K to 250K, 10 to 30), (VGA, 500K to 700K, 15 to 30), (720p, 1M to 1.5M, 30). This information is then provided in a message to the clients that are requested to generate streams (active speaker and/or subscribed ones). Each client uses this information and its own capability to generate the proper MBR and SVC streams (See FIGURE 5).

**[048]** Transitioning to operation 450, one or more data streams are received from at least one client that is requested to generate streams. For example, computing device 600 may receive a plurality of streams for one or more resolutions from a client. According to one embodiment, each resolution is represented by an MBR stream and one or more SVC streams.

**[049]** Flowing to operation 460, streams are sent to at least one client. For example, computing device 600 may send a first resolution of a stream received from a first client to a second client and send a second resolution of a stream received from the first client to a third client. Computing device 600 may also determine that a client does not have sufficient bandwidth to receive a requested high resolution version of a video stream from a second client and computing device 600 may choose a lower resolution version of the video stream to send to the client.

**[050]** Transitioning to decision operation 470, a determination is made as to whether the capabilities of a client have changed. When the capabilities of a client and/or the requirements of the system have changed, method 400 may return to operation 440 where the aggregated information is updated that represents a request for different video streams to be generated by one or more of the clients. For example, computing device 600 may determine that the only client in a video conference call utilizing CIF decoding has dropped out. Computing device 600 may then request by sending out the aggregated information that any other clients on the video conference call stop generating CIF streams. Consistent with embodiments of the invention, a client's available bandwidth may decrease, and the client may

inform computing device 600 of this occurrence. Computing device 600 may choose a lower resolution data stream for that client and/or alter the currently chosen data stream for that client to take the reduced bandwidth into account. This determination may be made periodically throughout the conference.

**[051]** When the capabilities do not change and the conference call ends, the process flows to an end operation and returns to processing other actions.

**[052]** FIG. 5 is a flow chart illustrating the generation of streams that combines MBR and SVC.

**[053]** After a start operation, the process flows to operation 510 where the client receives the aggregated information relating to the desired streams to be generated. According to one embodiment, the aggregated information includes the desired resolutions, bitrates, and frame rates that are grouped according to a resolution. The aggregated information may be received/updated at different times. For example, the information may be received after a client joins a video conference and updated when the conditions of the video conference change. For example, a client may be added/dropped from the video conference, a client's capabilities/requests may change during the video conference, and the like.

**[054]** Moving to operation 520, the streams for a resolution are generated. The client device determines the MBR streams and SVC layer streams to encode. For example, one 1080p client may encode streams for all of the resolutions that utilize all of the SVC layers, whereas another 1080p client may encode streams for all of the resolutions but only generate streams for a portion of the SVC layers. Generally, a client attempts to encode streams for its supported resolution and any resolutions that it supports that is a lower resolution than its maximum resolution.

**[055]** Moving to decision operation 530, a determination is made as to whether there are any more resolutions to generate. For example, the client may be requested to generate streams for many different resolutions, such as a 720p resolution video stream, a VGA resolution video stream, and a CIF resolution video stream. When there are more streams to generate, the process returns to operation 520. When there are not more streams to generate, the process flows to operation 540.

**[056]** At operation 540, the generated streams are sent from the client that created the streams. The generated streams may be sent to a video server and/or to a different destination. For example, a client may send some of the streams to a video server and other streams directly to a client device.

**[057]** FIG. 6 is a block diagram of a system including computing device 600. Consistent with an embodiment of the invention, the aforementioned memory storage and processing unit may be implemented in a computing device, such as computing device 600 of FIG. 6. Any suitable combination of hardware, software, or firmware may be used to implement the memory storage and processing unit. For example, the memory storage and processing unit may be implemented with computing device 600 or any of other computing devices 618, in combination with computing device 600. The aforementioned system, device, and processors are examples and other systems, devices, and processors may comprise the aforementioned memory storage and processing unit, consistent with embodiments of the invention. Furthermore, computing device 600 may comprise an operating environment for system 100 as described above. System 100 may operate in other environments and is not limited to computing device 600.

**[058]** With reference to FIG. 6, a system consistent with an embodiment of the invention may include a computing device, such as computing device 600. In a basic configuration, computing device 600 may include at least one processing unit 602 and a system memory 604. Depending on the configuration and type of computing device, system memory 604 may comprise, but is not limited to, volatile (e.g. random access memory (RAM)), non-volatile (e.g. read-only memory (ROM)), flash memory, or any combination. System memory 604 may include operating system 605, one or more programming modules 606, and may include a program 607 relating to determining capabilities and the encoding/decoding of video. Operating system 605, for example, may be suitable for controlling computing device 600's operation. In one embodiment, programming modules 606 may include, for example, a data analysis module and/or a video conference server 620. Furthermore, embodiments of the invention may be practiced in conjunction with a graphics library, other operating systems, or any other application program and is not limited to any particular application or system. This basic configuration is illustrated in FIG. 6 by those components within a dashed line 608.

**[059]** Computing device 600 may have additional features or functionality. For example, computing device 600 may also include additional data storage devices (removable and/or non-removable) such as, for example, magnetic disks, optical disks, or tape. Such additional storage is illustrated in FIG. 6 by a removable storage 609 and a non-removable storage 610.

**[060]** The term computer readable media as used herein may include computer storage media. Computer storage media may include volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information, such

as computer readable instructions, data structures, program modules, or other data. System memory 604, removable storage 609, and non-removable storage 610 are all computer storage media examples (i.e. memory storage.) Computer storage media may include, but is not limited to, RAM, ROM, electrically erasable read-only memory (EEPROM), flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to store information and which can be accessed by computing device 600. Any such computer storage media may be part of device 600. Computing device 600 may also have input device(s) 612 such as a camera, keyboard, a mouse, a pen, a sound input device, a touch input device, etc. Output device(s) 614 such as a display, speakers, a printer, etc. may also be included. The aforementioned devices are examples and others may be used.

**[061]** The term computer readable media as used herein may also include communication media. Communication media may be embodied by computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport mechanism, and includes any information delivery media. The term “modulated data signal” may describe a signal that has one or more characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media may include wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency (RF), infrared, and other wireless media.

**[062]** Computing device 600 may also contain a communication connection 616 that may allow device 600 to communicate with other computing devices 618, such as over a network in a distributed computing environment, for example, an intranet or the Internet. Communication connection 616 is one example of communication media.

**[063]** As stated above, a number of program modules and data files may be stored in system memory 604, including operating system 605. While executing on processing unit 602, programming modules 606 (e.g. video conference server 620) may perform processes including, for example, one or more of method stages as described above. The aforementioned process is an example, and processing unit 602 may perform other processes. Other programming modules that may be used in accordance with embodiments of the present invention may include electronic mail and contacts applications, word processing

applications, spreadsheet applications, database applications, slide presentation applications, drawing or computer-aided application programs, etc.

[064] Generally, consistent with embodiments of the invention, program modules may include routines, programs, components, data structures, and other types of structures that may perform particular tasks or that may implement particular abstract data types. Moreover, embodiments of the invention may be practiced with other computer system configurations, including hand-held devices, multiprocessor systems, microprocessor-based or programmable consumer electronics, minicomputers, mainframe computers, and the like. Embodiments of the invention may also be practiced in distributed computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed computing environment, program modules may be located in both local and remote memory storage devices.

[065] Furthermore, embodiments of the invention may be practiced in an electrical circuit comprising discrete electronic elements, packaged or integrated electronic chips containing logic gates, a circuit utilizing a microprocessor, or on a single chip containing electronic elements or microprocessors. Embodiments of the invention may also be practiced using other technologies capable of performing logical operations such as, for example, AND, OR, and NOT, including but not limited to mechanical, optical, fluidic, and quantum technologies. In addition, embodiments of the invention may be practiced within a general purpose computer or in any other circuits or systems.

[066] Embodiments of the invention, for example, may be implemented as a computer process (method), a computing system, or as an article of manufacture, such as a computer program product or computer readable media. The computer program product may be a computer storage media readable by a computer system and encoding a computer program of instructions for executing a computer process. The computer program product may also be a propagated signal on a carrier readable by a computing system and encoding a computer program of instructions for executing a computer process. Accordingly, the present invention may be embodied in hardware and/or in software (including firmware, resident software, micro-code, etc.). In other words, embodiments of the present invention may take the form of a computer program product on a computer-usable or computer-readable storage medium having computer-usable or computer-readable program code embodied in the medium for use by or in connection with an instruction execution system. A computer-usable or computer-readable medium may be any medium that can contain, store, communicate, propagate, or

transport the program for use by or in connection with the instruction execution system, apparatus, or device.

[067] The computer-usable or computer-readable medium may be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific computer-readable medium examples (a non-exhaustive list), the computer-readable medium may include the following: an electrical connection having one or more wires, a portable computer diskette, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, and a portable compact disc read-only memory (CD-ROM). Note that the computer-usable or computer-readable medium could even be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via, for instance, optical scanning of the paper or other medium, then compiled, interpreted, or otherwise processed in a suitable manner, if necessary, and then stored in a computer memory.

[068] Embodiments of the present invention, for example, are described above with reference to block diagrams and/or operational illustrations of methods, systems, and computer program products according to embodiments of the invention. The functions/acts noted in the blocks may occur out of the order as shown in any flowchart. For example, two blocks shown in succession may in fact be executed substantially concurrently or the blocks may sometimes be executed in the reverse order, depending upon the functionality/acts involved.

[069] While certain embodiments of the invention have been described, other embodiments may exist. Furthermore, although embodiments of the present invention have been described as being associated with data stored in memory and other storage mediums, data can also be stored on or read from other types of computer-readable media, such as secondary storage devices, like hard disks, floppy disks, or a CD-ROM, a carrier wave from the Internet, or other forms of RAM or ROM. Further, the disclosed methods' stages may be modified in any manner, including by reordering stages and/or inserting or deleting stages, without departing from the invention.

[070] All rights including copyrights in the code included herein are vested in and the property of the Applicant. The Applicant retains and reserves all rights in the code included herein, and grants permission to reproduce the material only in connection with reproduction of the granted patent and for no other purpose.

[071] While the specification includes examples, the invention's scope is indicated by the following claims. Furthermore, while the specification has been described in language specific to structural features and/or methodological acts, the claims are not limited to the features or acts described above. Rather, the specific

[072] features and acts described above are disclosed as example for embodiments of the invention.

WHAT IS CLAIMED IS:

1. A method for encoding video streams using a combination of Multiple Bit Rate (MBR) encoding and Scalable Video Control (SVC) encoding, comprising:
  - determining a capability associated with clients in a video conference; wherein the capability relates to an ability of each of the clients to encode one or more video streams;
  - placing each of the clients into a group that is selected based on the determined capability for each of the clients; and
  - receiving from at least one of the clients in the video conference, streams that include an MBR stream that is used for spatial scalability and one or more SVC streams relating to at least one of temporal and quality scalability.
2. The method of Claim 1, further comprising aggregating capabilities and requests from of each of the clients into a message and providing the message to clients in the video conference that are requested to generate streams within the video conference.
3. The method of Claim 1, wherein placing each of the clients into the group comprises determining a supported resolution for each of the clients and grouping the clients according to the supported resolution.
4. The method of claim 1, wherein each layer of a SVC stream is packetized into a separate steam.
5. The method of Claim 2, wherein the request that is received comprises a desired resolution, a frame rate and a bitrate.
6. The method of Claim 2, further comprising determining when conditions change within the video conference and updating the message such that different streams are generated by the one or more clients.
7. A computer-readable medium which stores a set of instructions which when executed performs a method for generating video streams that combine Multiple Bit Rate (MBR) encoding and Scalable Video Coding (SVC), the method executed by the set of instructions comprising:
  - determining capabilities associated with clients in a video conference; wherein the capabilities include determining for each client: a resolution and an encoding ability;
  - placing each of the clients into a group that is selected based on the determined resolution for each of the clients;
  - aggregating the capabilities for the clients into a message;

sending the message to clients that are requested to generate streams for the video conference; and

receiving from the clients in the video conference that are requested to generate streams, streams that include an MBR stream that is used for spatial scalability and one or more SVC streams relating to at least one of temporal and quality scalability.

8. A system for generating streams that combines Multiple Bit Rate (MBR) encoding and Scalable Video Coding (SVC) encoding in a video conference, the system comprising:

a memory storage;

a network connection device; and

a processing unit coupled to the memory storage, wherein the processing unit is operative to:

register at least one first video conference client and at least one second video conference client operatively connected to the system via the network connection device, wherein being operative to register the at least one first video conference client and the at least one second video conference client comprises being operative to establish capabilities for the at least one first video conference client and the at least one second video conference client, wherein the capabilities comprise: a resolution; a frame rate; and a bit rate;

receive a plurality of video streams from the at least one first video conference client, wherein the plurality of video streams are generated by the first video conference client and comprise a first set of streams that comprise an MBR stream and at least one SVC stream and a second set of video streams that comprise an MBR stream and at least one SVC stream, wherein the at least one first set of streams comprises higher resolution video streams than the second set of video streams.

9. The system of Claim 8, further comprising determining when capabilities change within the video conference.

10. The system of Claim 8, further comprising aggregating the capabilities and requests from of each of the clients into a message and provide the message to the clients in the video conference that are requested to generate streams within the video conference.

1/6

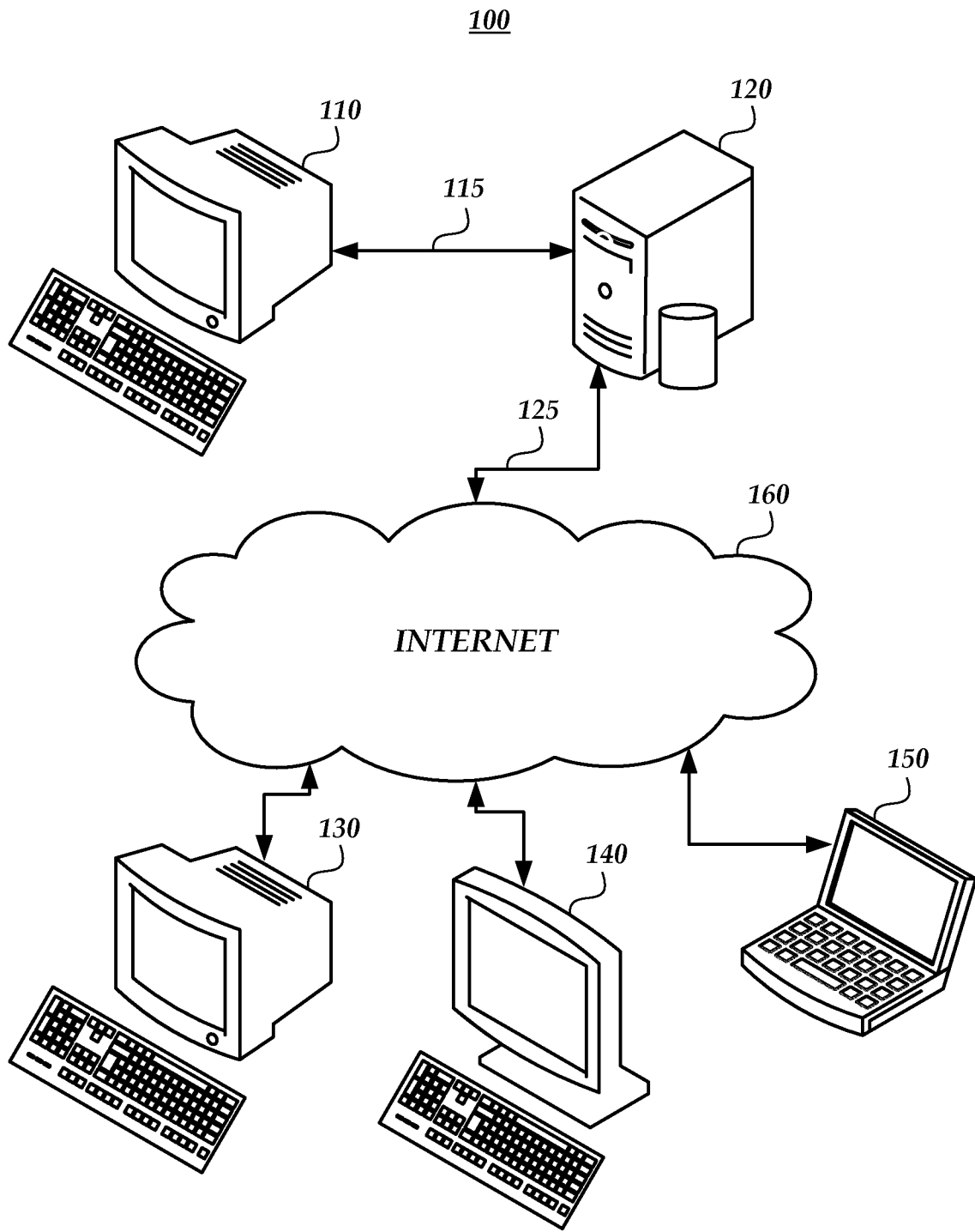


FIG. 1

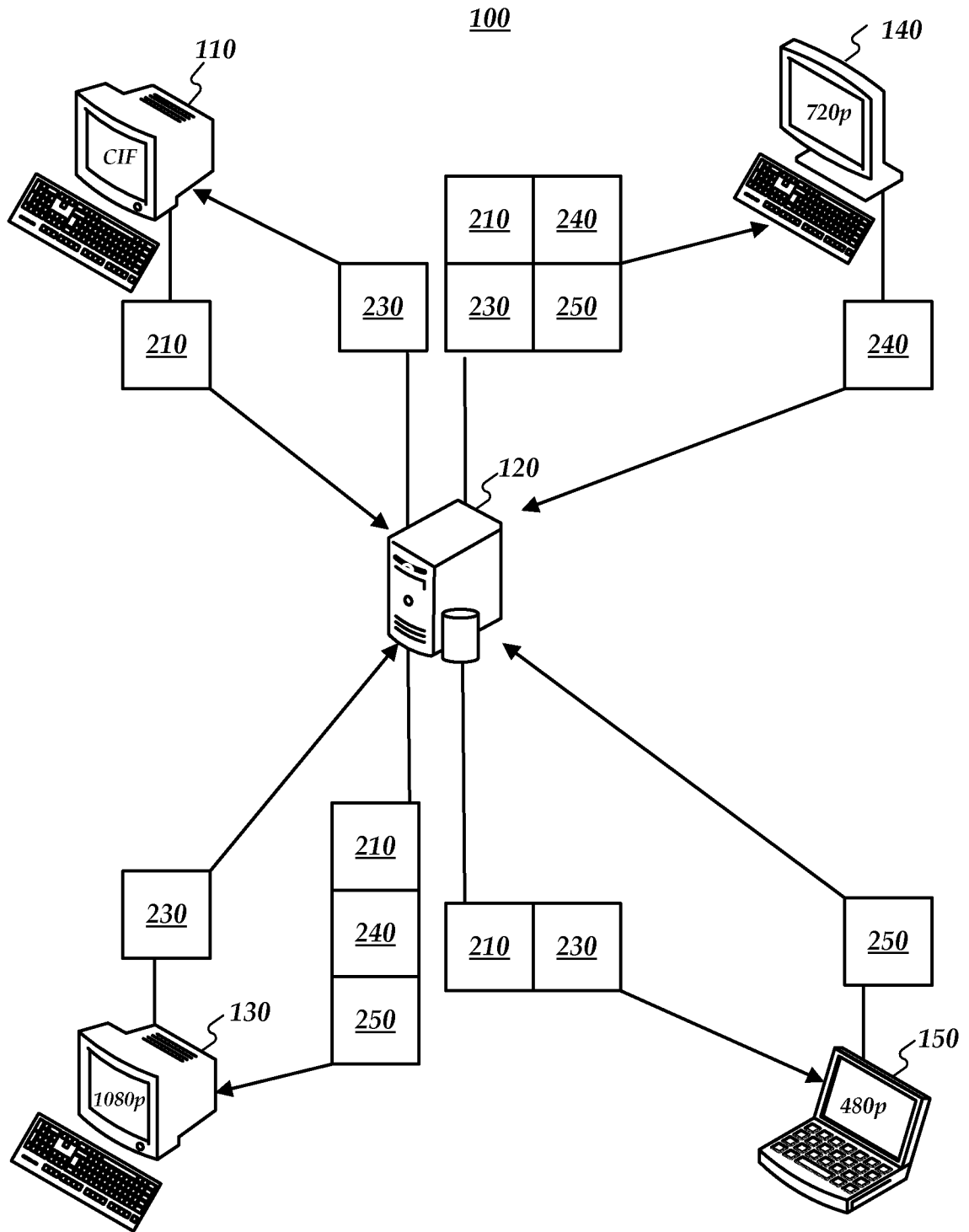


FIG. 2

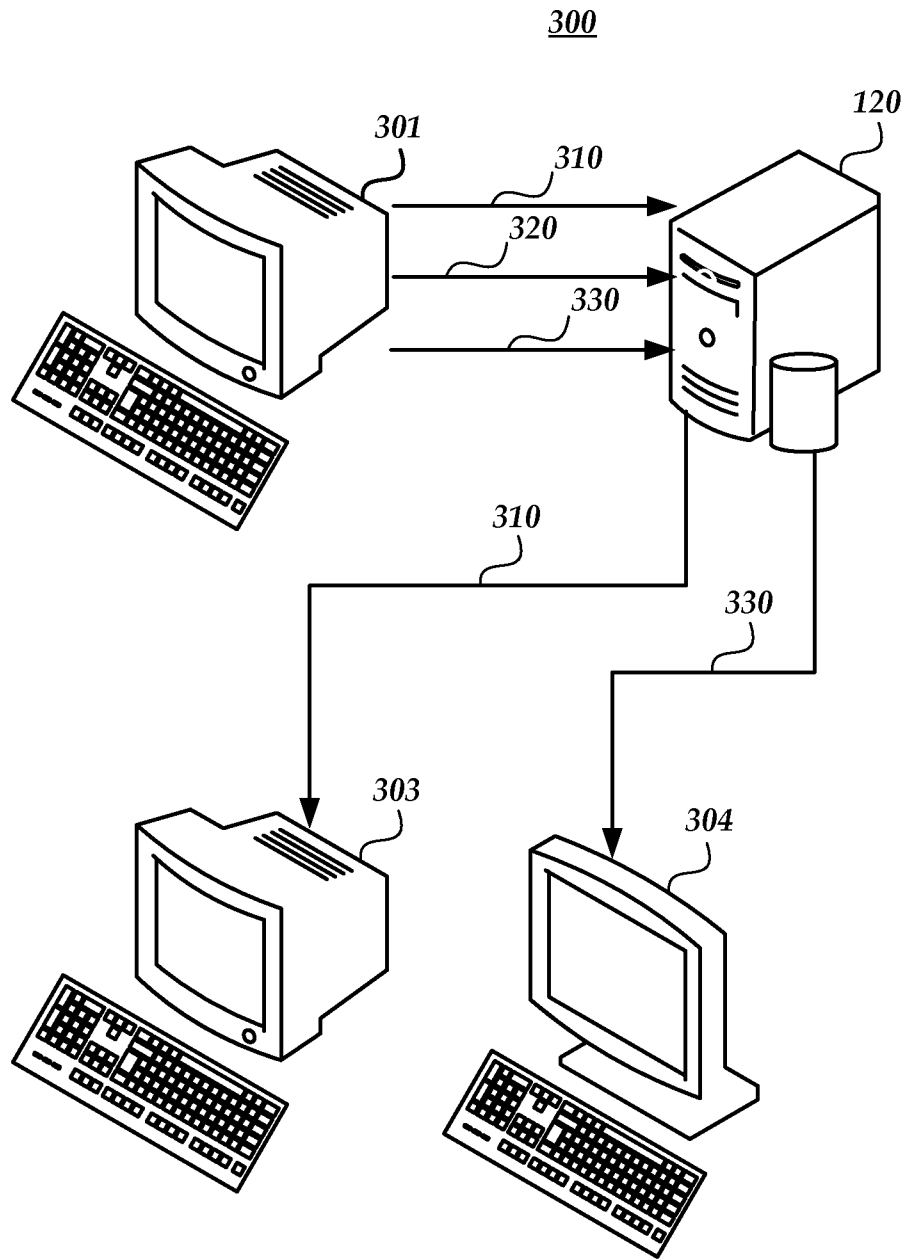


FIG. 3

4/6

400

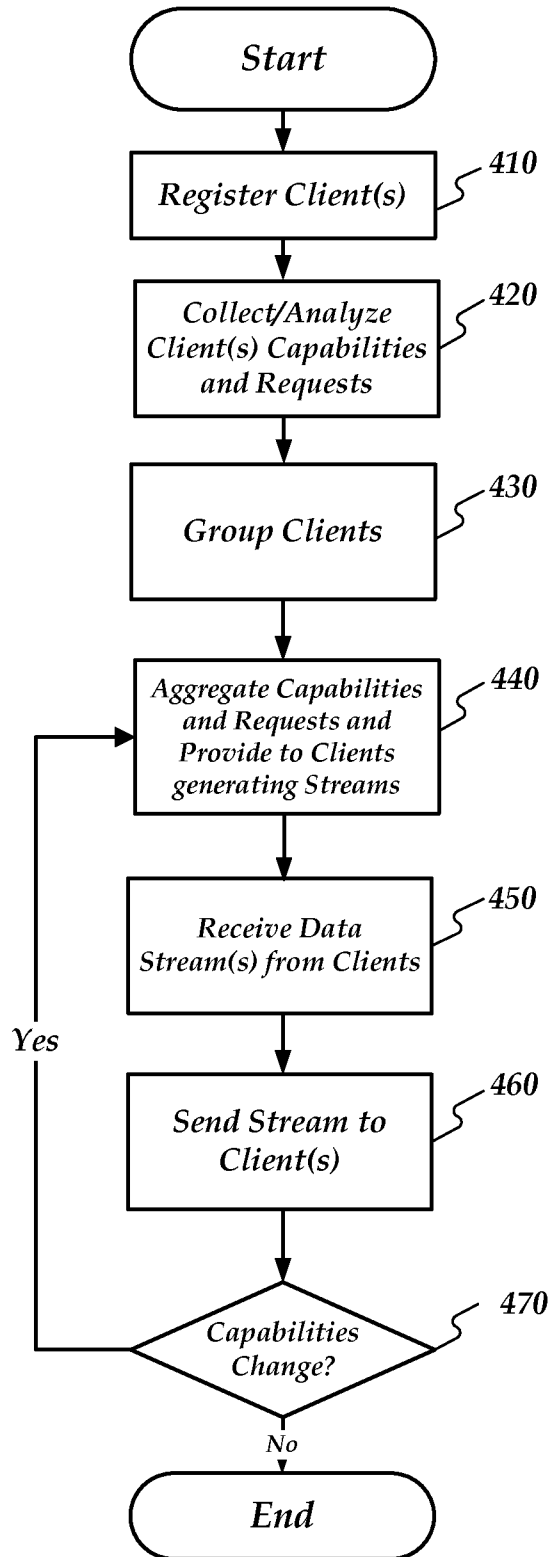


FIG. 4

5/6

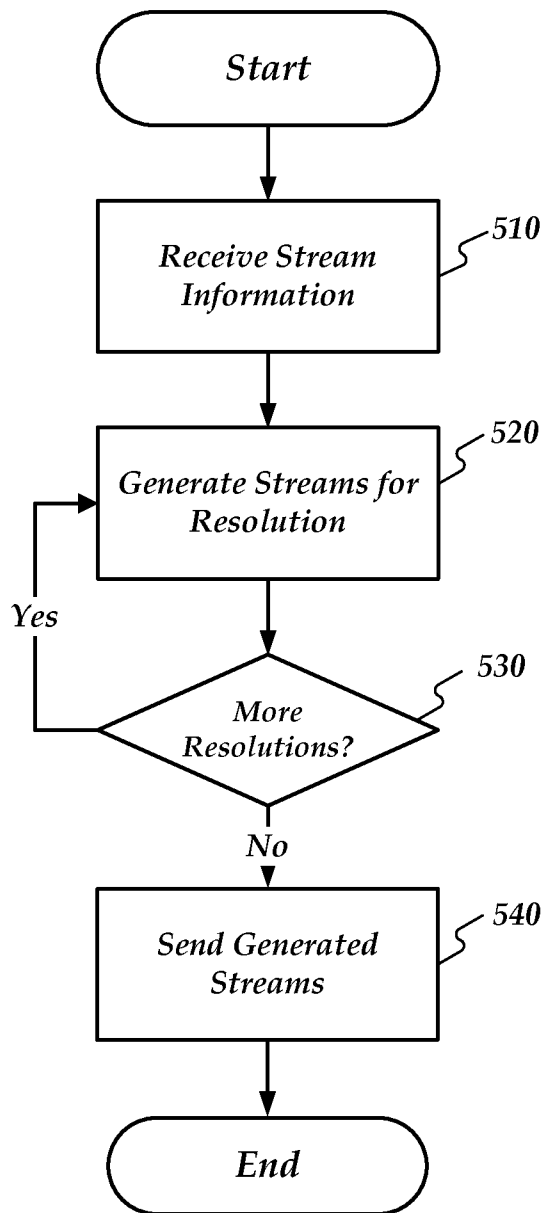


FIG. 5

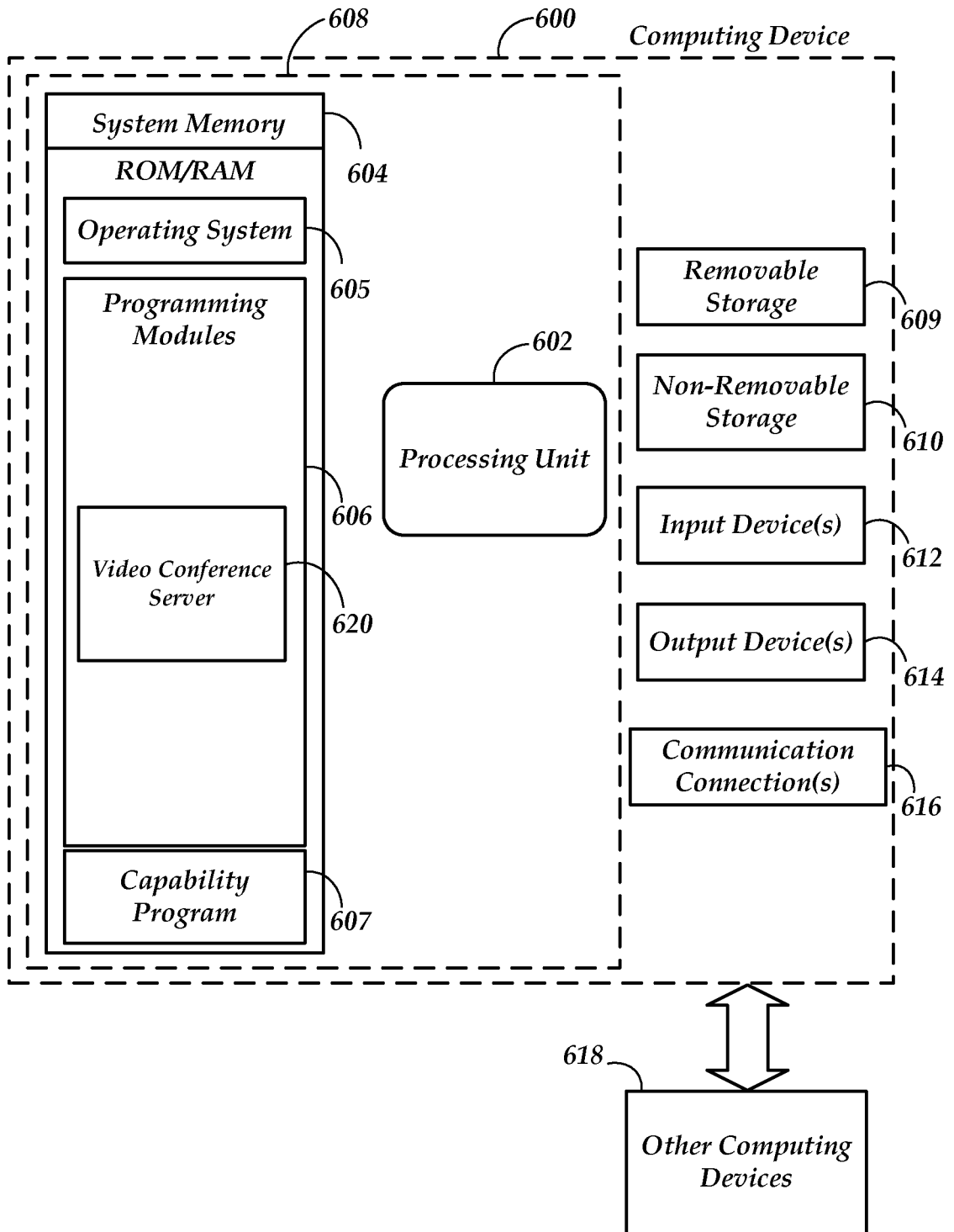


FIG. 6