CLAIMS

1. A computer-implemented method for streaming video from a server to a client comprising:

receiving at the server a request for playing a low latency video game or application from the client over a network which includes public network components;

in response to the request, establishing a video game or application session with the client, determining the hardware/software configuration of the client and executing the requested video game or application on the server;

generating and/or selecting a temporary decoder based on the hardware/software configuration of the client;

transmitting the temporary decoder to the client, the client installing the temporary decoder;

receiving control signals transmitted from the client indicating user input to the client as the user is playing the video game or using the application and responsively executing the video game or application on the server to generate uncompressed video images output from the video game or application;

receiving feedback information over the network, the feedback information usable to determine current channel characteristics of a communication channel between the first server and the client over the network;

encoding the uncompressed video images using encoding parameters selected based on the detected channel characteristics and the capabilities of the temporary decoder on the client to generate low latency compressed video images, the encoding parameters selected to achieve a latency such that the user has the perception that the selected video game or application is responding instantly to the control signals, wherein the encoding comprises:

subdividing each uncompressed video image into a first set of one or more frame portions and a second set of one or more frame portions, wherein the total number of frame portions depends upon the encoding parameters, and the particular frame portions in the first set and particular frame portions in the second set depend upon the encoding parameters and the number of video images that have been compressed since a particular frame portion was in the first set of frame portions;

encoding frame portions in the first set using a first type of encoding, wherein the first type of encoding is not dependent on prior images or frame portions;

encoding frame portions in the second set using a second type of encoding, wherein the second type of encoding is dependent on prior images or frame portions;

wherein subdividing each video image into frame portions and encoding the first set of frame portions using the first encoding format and the second set of frame portions using the second encoding format produces a more consistent bitrate for the video images and lower latency for the end user than if entire video images were compressed using the first type of encoding and the second type of encoding;

transmitting the low latency compressed video images from the server to the client over the network, the low latency video stream being decoded by the temporary decoder and rendered on the client;

wherein the operations of receiving control signals transmitted from the client, executing the video game or application to generate the uncompressed video images, encoding the uncompressed video images to generate low latency compressed video images and streaming the low latency compressed video images to the client over the network which includes public network components, and decoding the low latency compressed video images is performed with a combined latency such that the user has the perception that the selected video game or application is responding instantly to the control signals received from the client, the combined latency being achieved, at least in part, based on the encoding parameters;

detecting that the client has ended the session with the server; and

in response to detecting that the client has ended the session, temporarily disabling and/or removing the temporary decoder from the client.

2. The method as in claim 1 wherein the encoding parameters include a specified image resolution, frame rate, bitrate, and/or compression ratio.

3. The method as in claim 2 wherein one of the characteristics of the communication channel comprises the round-trip latency of the communication channel.

4. The method as in claim 2 wherein one of the characteristics of the communication channel comprises the data rate of the communication channel.

5. The method as in claim 2 wherein one of the characteristics of the communication channel comprises the reliability of the communication channel.

6. The method as in claim 5 wherein the reliability of the communication channel is characterized based on a percentage of lost packets or packets received with errors.

7. The method as in claim 6 wherein, in response to detecting a percentage of lost packets or packets received with errors below a first specified threshold value, then encoding the video content without error correction; and in response to detecting a percentage of lost packets or packets received with errors above the first specified threshold value, then encoding the video content with error correction.

8. The method as in claim 7 wherein, in response to detecting a percentage of lost packets or packets received with errors above a second specified threshold value, then attempting to establish a different communication channel with the client.

9. The method as in claim 8 wherein the different communication channel is between the client and a different server.

10. The method as in claim 3 wherein, in response to detecting a round-trip latency above a specified threshold value, then attempting to establish a different communication channel with the client with a relatively lower round-trip latency.

11. The method as in claim 2 wherein a particular image resolution, frame rate, bitrate, and/or compression ratio is selected in order to achieve a latency of less than or equal to a specified maximum value between the operations of generating the control signals responsive to user input and rendering the video images on the client.

12. The method as in claim 11 further comprising: specifying a maximum latency requirement for the low latency video game, the maximum latency requirement indicating the maximum allowable latency between a first user input to generate the control signals and a rendering of a video image on the client resulting from the first user input.

13. The method as in claim 11 wherein the specified maximum value comprises 80 ms.

14. A system for hosting low-latency streaming interactive video games or applications, comprising:

a plurality of servers, one or more of the servers to establish a video game or application session with a client and execute the low-latency streaming interactive video games or applications, one or more of the servers further generating and/or selecting a temporary decoder for decoding video of the low-latency streaming interactive video games based on the hardware/software configuration of a client and transmitting the temporary decoder to the client, the client installing the temporary decoder;

an inbound routing network to receive packet streams over a network including public network components from users playing the video games or applications on the client and routing the packet streams to one or more of the servers, the packet streams including user control input for the video games, one or more of the servers being operable to compute video game or application video and audio output data responsive to the user control input comprising sequences of video images output from the video game or application;

the one or more of the servers to receive feedback information over the network, the feedback information usable to determine current channel characteristics of a communication channel between the one or more servers and the client over the network;

a compression unit coupled to receive the video game or application video and audio output data from the one or more of the servers and generate low-latency streaming compressed audio/video (A/V) data therefrom in a format decodable by the temporary decoder, the compression unit generating the low-latency streaming

compressed audio/video (A/V) data by compressing the sequences of video images using encoding parameters selected based on the detected channel characteristics and the capabilities of the temporary decoder on the client to achieve a latency such that the user has the perception that the selected video game or application is responding instantly to the control signals, wherein the video encoding comprises:

subdividing each uncompressed video image into a first set of one or more frame portions and a second set of one or more frame portions, wherein the total number of frame portions depends upon the encoding parameters, and the particular frame portions in the first set and particular frame portions in the second set depend upon the encoding parameters and the number of video images that have been compressed since a particular frame portion was in the first set of frame portions;

encoding frame portions in the first set using a first type of encoding, wherein the first type of encoding is not dependent on prior images or frame portions;

encoding frame portions in the second set using a second type of encoding, wherein the second type of encoding is dependent on prior images or frame portions;

wherein subdividing each video image into frame portions and encoding the first set of frame portions using the first encoding format and the second set of frame portions using the second encoding format produces a more consistent bitrate for the video images and lower latency for the end user than if entire video images were compressed using the first type of encoding and the second type of encoding; and

an output routing network to route the low-latency streaming compressed A/V data over the network including public network components to the client over a corresponding communication channel;

wherein the temporary decoder on the client responsively decodes the streaming compressed A/V data in real time to render video and audio at the client representing the video game or application video and audio output data;

wherein the operations of receiving packet streams over the public network from the client, executing the interactive video games or applications, generating and routing the low latency streaming compressed A/V data over the public network to the client, and decoding the low latency streaming compressed A/V data is performed with a combined latency such that the user has the perception that the selected video game or application is responding instantly to the control signals

received from the client, the combined latency being achieved, at least in part, based on the encoding parameters;

detecting that the client has ended the session with the server; and

in response to detecting that the client has ended the session, temporarily disabling and/or removing the temporary decoder from the client.

15. The system as in claim 14 wherein the latency associated with receiving packet streams, generating the low-latency streaming compressed A/V data, and streaming the output low-latency streaming compressed A/V data back to the user with a finite latency less than 80 ms.

16. The system as in claim 14 wherein the latency associated with receiving packet streams, generating the streaming compressed A/V data, and streaming the output streaming compressed A/V data back to the user is tolerable for the user, given the nature of the video game, computer application or interactive linear video experience being interactively streamed.

17. The system as in claim 14 wherein the encoding parameters include a specified image resolution, frame rate, bitrate, and/or compression ratio.

18. The system as in claim 17 wherein a particular image resolution, frame rate, bitrate, and/or compression ratio is selected in order to achieve a latency of less than or equal to a specified maximum value between the operations of generating the user control input responsive to user input and decoding the low latency streaming compressed A/V data to render the video images on the client.