SYSTEM AND METHOD TO ADAPT TO NETWORK CONGESTION

I. Claim of Priority under 35 U.S.C. §119

[0001] The present Application for Patent claims priority to Provisional Application No. 61/020,368 entitled "METHOD AND APPARATUS FOR ADAPTING RATE-TRAJECTORY TO NETWORK CONGESTION" filed January 10, 2008 and assigned to the assignee hereof and hereby expressly incorporated by reference herein.

II. Related Application

[0002] The present application relates to U.S. Application No. 11/315,399, filed on December 21, 2005, entitled "Methods and Systems for Adaptive Encoding of Real-Time Information in Packet-Switched Wireless Communication Devices", and U.S. Application No. 11/972,594, filed on January 10, 2008, entitled "Content - and Link-Dependent Coding Adaptation for Multimedia Telephony", and assigned to the assignee hereof, and expressly incorporated by reference herein.

III. Field of the Disclosure

[0003] The present disclosure is generally directed to systems and methods to adapt to network congestion.

IV. Background

[0004] When a sending terminal in a network detects congestion or easing of congestion in the network, the sending terminal may determine how to adapt the transmission rate of data sent from the sending terminal. The problem of determining what transmission rate to choose based on feedback received from a receiver in the network may be challenging. Proper selection of the adaptation rate may improve the convergence of the adaptation control loop and may improve the quality of service. However, frequently oscillating rate adjustments towards convergence may degrade the service experience, particularly for real-time services. Another challenge of rate adaptation is to determine how quickly to increase the transmission rate when congestion has eased. Increasing the rate too aggressively may quickly introduce

further congestion if the sender is unaware of the channel condition, which may lead to a poor service experience because of the increase of the transmission rate followed by the sudden need to decrease the transmission rate due to the further congestion. Increasing the rate too conservatively may prevent the sender from making full use of the decongested channel as the decongested channel develops additional capacity.

[0005] Conventional approaches typically adapt to feedback by changing the send rate to a fixed value until another feedback message is received and the congestion status information is updated, involving multiple feedback messages. Such conventional approaches do not attempt to adapt to network congestion based on a single feedback message. During congestion, achieving a multi-phase adaptation of decongestion followed by transmission at the maximum sustainable rate requires multiple feedback messages from the receiver describing the status of the channel. During easing of congestion, the sender conventionally uses very conservative increases in rate along with waiting for feedback to ensure that the sender does not re-introduce congestion. Also, when congestion has eased, conventional approaches typically blindly probe the channel with additional data to get an estimate of the maximum sustainable rate of the channel. The blind probing may introduce additional delay if the blind probing re-introduces congestion and the channel is unable to transport the additionally inserted data in a timely manner.

V. Summary

[0006] In a particular embodiment, a method is disclosed that includes receiving a feedback message at a transmitter, the feedback message including an indication of a magnitude of congestion and a sustainable rate of data transmission at a receiver. The method also includes determining a decongestion rate and a decongestion time based on at least one of the sustainable rate of data transmission and the magnitude of congestion when the magnitude of congestion satisfies a threshold value. The method further includes sending data at the decongestion rate from the transmitter to the receiver for the decongestion time. The method also includes adjusting a data transmission rate at the transmitter to the sustainable rate of data transmission after the decongestion time is ended. In a particular embodiment, the sustainable rate of data transmission may be an estimated or predicted maximum sustainable rate.

[0007] In another embodiment, a method is disclosed that includes identifying congestion in a transmission path between a sender and a receiver using parameters provided in a feedback message from the receiver to the sender, the parameters including a magnitude of congestion and a sustainable rate of data transmission. The method also includes determining a decongestion rate and a decongestion time based on at least one of the sustainable rate of data transmission and the magnitude of congestion. The method further includes decongesting the transmission path by sending data at the decongestion rate from the sender to the receiver for the decongestion time. The method also includes adjusting a data transmission rate in the transmission path to the sustainable rate of data transmission after the decongestion time is ended.

[0008] In another embodiment, a method is disclosed that includes receiving a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR) in a feedback message at a transmitter. The method also includes increasing a sending rate at the transmitter to an increased sending rate determined based on at least one of the APTO value and the ARR. The method further includes continuing transmission at the increased sending rate after a detection time, wherein the detection time is determined based on at least one of the APTO value and the ARR.

[0009] In another embodiment, a computer-readable medium including computer executable instructions is disclosed. The computer executable instructions are operable to cause a computer to send a feedback message from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR). The computer executable instructions are operable to cause a computer to receive data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR. The computer executable instructions are operable to cause a computer to receive data at the receiver sent by the transmitter at a sustained data rate of the ARR after the decongestion time is ended. In alternative embodiments, the decongestion time may be set as a fixed constant T_decongest that is determined based on how quickly the service/user experience requires that decongestion be achieved. For example, the service requirement may be that decongestion be achieved within about 1000 milliseconds to avoid too much disruption to a video stream. In other alternative

embodiments, the decongestion time may be set adaptively based on other measurements or criteria aside from the APTO value received.

[0010] In another embodiment, a method is disclosed that includes sending a feedback message from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR). The method also includes receiving data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR. The method further includes receiving data at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

[0011] In another embodiment, an apparatus is disclosed that includes means for sending a feedback message from a receiver to a transmitter. The feedback message includes a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR). The apparatus also includes means for receiving data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR. The apparatus further includes means for receiving data at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

[0012] In another embodiment, an apparatus is disclosed that includes a processor configured to generate a feedback message to send from a receiver to a transmitter. The feedback message includes a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR). Data is received at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR. Data is received at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

[0013] In another embodiment, an apparatus is disclosed that includes a transmitter configured to transmit data, to receive a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR) in a feedback message, to initiate decongestion by reducing a sending rate at the transmitter to a decongestion rate

determined based on at least one of the APTO value and the ARR, and to end decongestion transmission at the decongestion rate after a decongestion time determined based on at least one of the APTO value and the ARR.

[0014] In another embodiment, an apparatus is disclosed that includes means for initiating decongestion by reducing a sending rate at a transmitter to a decongestion rate determined based on at least one of a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR) that are received in a feedback message. The apparatus also includes means for ending decongestion transmission at the decongestion rate after a decongestion time determined based on at least one of the APTO value and the ARR.

[0015] One particular advantage provided by the disclosed embodiments is that channel decongestion and then transmission at the maximum sustainable rate under congestion conditions may be achieved.

[0016] Another advantage provided by the disclosed embodiments is that a ramp-up of rate above the current transmission rate that does not reintroduce congestion under conditions of congestion easing may be achieved.

[0017] Other aspects, advantages, and features of the present disclosure will become apparent after review of the entire application, including the following sections: Brief Description of the Drawings, Detailed Description, and the Claims.

VI. Brief Description of the Drawings

[0018] FIG. 1 is a diagram of a particular illustrative embodiment of modifying a transmission rate to adapt to network congestion;

[0019] FIG. 2 is a diagram of a particular illustrative embodiment of a fluid model of network congestion to determine an amount of congested data to remove from a network;

[0020] FIG. 3 is a timing diagram of a particular illustrative embodiment of adapting a transmission rate to network congestion to decongest the transmission path;

[0021] FIG. 4 is a diagram of a particular illustrative embodiment of using a fluid model of congestion to determine an amount of data to insert into a network;

[0022] FIG. 5 is a timing diagram of a particular illustrative embodiment of adapting a transmission rate to network congestion to ramp-up data transmission;

[0023] FIG. 6 is a flow diagram of a first illustrative embodiment of a method to adapt to network congestion;

[0024] FIG. 7 is a flow diagram of a second illustrative embodiment of a method to adapt to network congestion;

[0025] FIG. 8 is a flow diagram of a third illustrative embodiment of a method to adapt to network congestion;

[0026] FIG. 9 is a flow diagram of a fourth illustrative embodiment of a method to adapt to network congestion; and

[0027] FIG. 10 is a diagram of a particular illustrative embodiment of a system that is adaptable to network congestion.

VII. Detailed Description

[0028] Referring to FIG. 1, a diagram of a particular illustrative embodiment of modifying a transmission rate to adapt to network congestion is depicted and generally designated 100. The diagram 100 depicts a transmission rate of a transmitter as a function of time. Initially, the transmitter sends data at a first transmission rate 101. A feedback message may be received at the transmitter, as indicated at 102. The feedback message includes an indication of a magnitude of congestion of a transmission network and a sustainable rate of data transmission 108 at a receiver that receives data from the transmitter. A decongestion rate 110 and a decongestion time 112 may be determined based on at least one of the sustainable rate of data transmission 108 and the magnitude of congestion when the magnitude of congestion satisfies a threshold value. After determining the magnitude of congestion and the sustainable rate of data transmission 108, data may be sent at the decongestion rate 110 from the transmitter to the receiver for the decongestion time 112, as indicated at 104. A data transmission rate at the transmitter may be adjusted to the sustainable rate of data transmission 108 after

the decongestion time 112 is ended, as indicated at 106. In a particular embodiment, the transmitter transmits the data to the receiver via a wireless network. In a particular embodiment, the sustainable rate of data transmission may be substantially the maximum sustainable rate of data transmission. As used herein, the term maximum sustainable rate of data transmission is an estimated or predicted maximum sustainable rate of data transmission.

[0029] Including the magnitude of congestion and the maximum sustainable rate of data transmission enables the sender or transmitter to determine rate-trajectories based on a single feedback message. Either the sender or the transmitter may estimate the amount of congested bits in the transmission path. Based on this estimate and the estimate of the maximum sustainable rate of data transmission provided in the single feedback message, the sender or transmitter may determine a set of rates that may achieve channel decongestion and then transmission at the maximum sustainable rate under congestion conditions.

[0030] In a particular embodiment, the magnitude of congestion may be measured by an arrival-to-playout time offset (APTO) value and the maximum sustainable rate of data transmission 108 may be measured by an average received rate (ARR) of data received at the receiver. In a particular embodiment, congestion in an uplink, a downlink, and a core network may be reflected in a standardized APTO_ARR feedback message from the receiver to the sender or transmitter. The APTO_ARR is defined in the 3GPP2 C.P0055-A standard, and has been proposed for the 3GPP TS 26.114 standard. Information in the standardized APTO_ARR feedback message may indicate congestion at the receiver where data packets are not arriving at the receiver in time for properly scheduled playout. Information in the standardized APTO_ARR feedback message may also provide an estimate for guidance on what rate may be sustainable for the end-to-end transmission path. The standardized APTO_ARR feedback message may indicate the average received rate (ARR) at the receiver and also the receiver's request to advance or delay arrival times of data packets (APTO) based on playout needs. The delay may serve as an outer loop variable of a real-time packet (RTP) service.

[0031] In a particular embodiment, sending the data from the transmitter to the receiver at the decongestion rate 110 for the decongestion time 112 may remove an amount of congested data, the amount of congested data removed being substantially equal to a product of the decongestion time 112 and a difference between the ARR 108 and the decongestion rate 110, as shown in FIG. 1. In a particular embodiment, the amount of congested data removed may also be equal to a product of the APTO value and the ARR 108. In a particular embodiment, during congestion the sender or transmitter rate adapts down to enable the data packets to arrive at the receiver in time for playout, initially by removing the backlog caused by congestion and then by operating at the maximum sustainable rate of the system that does not introduce further congestion or another backlog. If the backlog is not removed initially, then just operating at the maximum sustainable rate may not correct the current congestion state and the data packets may still be delayed. In a particular embodiment, during easing of congestion the sender or transmitter rate adapts up to improve the quality of the data transmission while ensuring that data packets arrive at the receiver in time for proper playout, improving the quality of the data transmission. When the data packets are from a video encoder, the resulting video quality may be thereby improved, for example.

[0032] Referring to FIG. 2, a diagram of a particular illustrative embodiment of a fluid model of network congestion to determine an amount of congested data to remove from a network is depicted and generally designated 200. A spout 201 directs a data packet stream 202 into a bucket 203. An amount of data 205 accumulates in the bucket 203 before being drained out of the bucket 203 through a drain 207. When the inflow of the data through the spout 201 equals the outflow of the data through the drain 207, the amount of data 205 in the bucket 203 remains constant. When the inflow of the data through the spout 201 is greater than the outflow of the data through the drain 207, the amount of the data 205 accumulating in the bucket 203 increases, indicating a state of congestion in the data transmission path. When the inflow of the data through the spout 201 is less than the outflow of the data through the drain 207, the amount of the data 205 accumulating in the bucket 203 decreases, which decongests the data transmission path.

[0033] Data packets from a video encoder may enter the data packet stream, as indicated at 202. An amount of congested data to remove 208 may be equal to the

product of a delay_to_adjust time 212 and a drain_rate 210 to the receiver. In a particular embodiment, the delay_to_adjust time 212 may be equal to the APTO value and the drain_rate 210 may be equal to the ARR 108 of FIG. 1. The encoder rate of data transmission may be reduced to the decongestion rate 110 for the decongestion time 112 to remove an amount of data 206 from the bucket 203, where the amount of data 206 removed from the bucket 203 may be equal to the amount of congested data to remove 208. Once the backlog of the amount of data 206 has been removed, data packets 204 arriving after the decongestion time 112 may leave the bucket 203 at the proper playout time. In a particular embodiment, the ARR 108 may serve as an estimate of the drain_rate 210 of the bucket and the sender or transmitter may remove a number of octets of congested data from the bucket 203 equal to the product of the APTO value and the ARR 108. As described above, when the inflow of the data through the spout 201 (the encoder rate of data transmission) is less than the outflow of the data through the drain 207 (the drain_rate 210), the amount of the data 205 accumulating in the bucket 203 decreases, which decongests the data transmission path. The ARR 108 may also serve as an estimate of the maximum sustainable rate through the system and may be used to transmit at the maximum sustainable rate of the ARR 108 once congestion is removed. In a particular embodiment, the ARR 108 may be measured for media throughput, real-time packet (RTP) throughput, which allows a video encoder, for example, to use the ARR 108 for a rate control target rate.

[0034] Referring to FIG. 3, a diagram of a particular illustrative embodiment of a method to adapt to network congestion by showing how fast to decongest a transmission path is depicted and generally designated 300. In a particular embodiment, the diagram 300 illustrates timing associated with decongestion according to FIG. 1. Initially, as indicated at 304, a transmission rate may be greater than the ARR 108. The receiver 301 may measure an unacceptable APTO value, as indicated at 324, and send an APTO_ARR feedback message 310 to the sender or transmitter 303. The APTO_ARR feedback message 310 may be received at the sender or transmitter 303, which may take a time T_react to react to the APTO_ARR feedback message 310, as indicated at 312. The sender or transmitter 303 may react to the APTO_ARR feedback message 310 by reducing the transmission rate to the decongestion rate 110, as indicated at 306, for the decongestion time 112, as indicated at 302. The start of the decongestion phase propagates to the receiver 301, as indicated at 314, and the end of the

decongestion phase propagates to the receiver 301, as indicated at 316. After the decongestion time 112 has ended, the transmission rate is increased to the ARR 108, as indicated at 308. A group delay of the APTO measurement filter is indicated at 320. A minimum feedback sending interval is indicated at 322. Finally, as indicated at 326, the receiver 301 checks the APTO measurement to determine whether another APTO_ARR feedback message 318 needs to be sent to the sender or transmitter 303.

[0035] In a particular embodiment, the channel may be decongested before the receiver 301 makes the next decision about the state of congestion. In a particular embodiment, the transmission path may be decongested as fast as possible to quickly ease congestion at the receiver 301. However, very aggressive decongestion requires reducing the data transmission to a much lower decongestion rate 110, which may unacceptably degrade video performance during the decongestion phase when the data packets contain encoded video data. In a particular embodiment, the decongestion rate 110 may be at least equal to a rate to achieve an acceptable frame quality and an acceptable frame rate. The rate to achieve an acceptable frame quality and an acceptable frame rate may be a rate that produces video frames that maintain an acceptable peak signal-to-noise ratio (PSNR), because otherwise the video frame may be skipped in accord with a variable frame rate (VFR). In another particular embodiment, the decongestion rate 110 may be at least equal to a rate to achieve a minimum frame quality and a minimum frame rate. The rate to achieve a minimum frame quality and a minimum frame rate may be a rate that produces video frames that maintain a minimum peak signal-to-noise ratio (PSNR), because otherwise the video frame may be skipped in accord with a variable frame rate (VFR).

[0036] In a particular embodiment, the decongestion time 112 may be proportional to the APTO value with a constant of proportionality F. The decongestion time 112 may also be proportional to the amount or magnitude of congestion. The decongestion rate 110 may be proportional to the ARR 108 with a constant of proportionality equal to a ratio of (F-1) to F. For video data, the decongestion rate 110 may be achieved by encoding each frame at a target rate of the ARR 108 and then skipping every Fth frame. For example, when F equals two, the decongestion time 112 may be twice the APTO value and the decongestion rate 110 may be half the ARR 108. For video data, the

video encoder may skip sending every other frame encoded at the target rate of the ARR 108 for a time period of twice the APTO value.

[0037] In alternative embodiments, the decongestion time may be set as a fixed constant T_decongest that is determined based on how quickly the service/user experience requires that decongestion be achieved. For example, the service requirement may be that decongestion be achieved within about 1000 milliseconds to avoid too much disruption to a video stream. The decongestion rate may then be equal to the product of the ARR with (1 – APTO/T_decongest). In other alternative embodiments, the decongestion time may be set adaptively to T_decongest_adapt based on other measurements or criteria aside from the APTO value received. For example, the measurements or criteria may include quality of service considerations, total network usage, the number of users, the amount of usage per user, and the like. The decongestion rate may then be equal to the product of the ARR with (1 - APTO/T_decongest_adapt).

[0038] Referring to FIG. 4, a diagram of a particular illustrative embodiment of a method to adapt to network decongestion/easing of congestion using the fluid model of congestion is depicted and generally designated 400. A spout 401 directs a data packet stream 402 into a bucket 403. An amount of data 405 accumulates in the bucket 403 before being drained out of the bucket 403 through a drain 407. When the inflow of the data through the spout 401 equals the outflow of the data through the drain 407, the amount of data 405 in the bucket 403 remains constant. When the inflow of the data through the spout 401 is greater than the outflow of the data through the drain 407, the amount of the data 405 accumulating in the bucket 403 increases, indicating a state of congestion in the data transmission path. When the inflow of the data through the spout 401 is less than the outflow of the data through the drain 407, the amount of the data 405 accumulating in the bucket 403 decreases, indicating that there is room to insert additional data into the data packet stream 402 without increasing the congestion of the system.

[0039] Data packets from a video encoder may enter the data packet stream, as indicated at 402. An amount of insertable data to insert 408 may be equal to the product of a delay_to_adjust time 412 and a drain_rate 410 to the receiver. As described above,

when the inflow of the data through the spout 401 is less than the outflow of the data through the drain 407, the amount of the data 405 accumulating in the bucket 403 decreases, indicating that there is room to insert additional data into the data packet stream 402 without causing additional congestion. In a particular embodiment, the delay_to_adjust time 412 may be equal to the absolute value of the APTO value and the drain_rate 410 may be greater than or at least equal to the ARR 108. The encoder rate of data transmission may be increased to another rate of sending data long enough to insert an amount of data 406 into the bucket 403, where the amount of data 406 inserted into the bucket 403 may be equal to the amount of insertable data to insert 408. Once the amount of data 406 has been inserted, data packets 404 arriving after the insertion phase may leave the bucket 403 at the proper playout time. In a particular embodiment, the transmission rate of the transmitter may be increased to a rate higher than a current rate of data transmission when the magnitude of congestion is less than a threshold value. In a particular embodiment, the transmission rate of the transmitter may be increased to a rate higher than the ARR 108 when the APTO value is less than a threshold value.

[0040] In a particular embodiment, the APTO value indicates a difference in the arrival statistic of data packets at the receiver compared to when the data packets are scheduled for proper playout without jitter, for video data. The APTO value may be an indication of the amount of delay the receiver would like to adjust in the end-to-end transmission path between the video sender and receiver. If there is congestion, then the APTO value may be positive and backlog packets in the bucket may be drained by the APTO value to remove the delay, as shown in FIG. 2. If there is an easing of congestion, then the APTO value may be negative and the encoder may increase the transmission rate to use the additional bandwidth, where an amount of delay may be introduced that is equal to the absolute value of the APTO value, as shown in FIG. 4.

[0041] Referring to FIG. 5, a diagram of a particular illustrative embodiment of a method to adapt to network congestion by showing how fast to ramp-up data transmission is depicted and generally designated 500. Initially, as indicated at 504, a transmission rate may be less than the ARR 108. In another embodiment, the transmission rate may not be less than the ARR 108. The receiver 501 may measure an APTO value that is less than a threshold value, as indicated at 524, and send an

APTO_ARR feedback message 510 to the sender or transmitter 503. The APTO_ARR feedback message 510 may be received at the sender or transmitter 503, which may take a time T_react to react to the APTO_ARR feedback message 510, as indicated at 512. The sender or transmitter 503 may react to the APTO_ARR feedback message 510 by increasing the transmission rate to another rate of sending data, as indicated at 508, for at least a detection time, as indicated at 502. The start of the rate increase propagates to the receiver 501, as indicated at 514, and the end of the detection time, when the rate increase is detected at the receiver 501, propagates to the receiver 501, as indicated at 516. A group delay of the APTO measurement filter is indicated at 520. A minimum feedback sending interval is indicated at 522. Finally, as indicated at 526, the receiver 501 checks the APTO measurement to determine whether another APTO_ARR feedback message 518 needs to be sent to the sender or transmitter 503.

[0042] In a particular embodiment, for the sake of convenience, the detection time may be set to be equal to the decongestion time 112. In a particular embodiment, the detection time may be equal to the decongestion time 112 and may also be proportional to the absolute value of the last received APTO value with a constant of proportionality F. In a particular embodiment, another rate of sending data may be determined to accomplish an addition of insertable data into the network that may be distributed over a time period that is proportional to the decongestion time 112 with a constant of proportionality R. In a particular embodiment, another rate of sending data may be equal to the product of the ARR 108 and the sum of one and the reciprocal of the product of R and F. In another particular embodiment, another rate of sending data may be equal to the product of the sender's current transmission rate and the sum of one and the reciprocal of the product of R and F. In other particular embodiments, the detection time may be a constant, or the detection time may be set adaptively. The detection time may therefore be determined based on at least one of the APTO value, the ARR, a constant value, or an adaptively set value

[0043] Referring to FIG. 6, a flow diagram of a particular illustrative embodiment of a method to adapt to network congestion is depicted and generally designated 600. The method 600 includes receiving a feedback message at a transmitter, the feedback message including a magnitude of congestion and a sustainable rate of data transmission at a receiver, as indicated at 602. For example, the APTO_ARR feedback message 310

may be received at a transmitter, the APTO_ARR feedback message 310 including a magnitude of congestion, the APTO value, and the sustainable rate of data transmission, the ARR 108, as depicted in FIG. 1 and FIG. 3. The method 600 includes determining whether the magnitude of congestion satisfies a threshold value, as indicated at 604. For example, the APTO value may be positive, indicating congestion, and may be larger than a predetermined amount, indicating too much congestion. If the magnitude of congestion does not satisfy a threshold value, then the method 600 includes continuing data transmission at the same or a higher rate, as indicated at 606. For example, if the APTO is less than a threshold amount, then the transmission rate may be increased to another rate of sending data, as depicted in FIG. 5.

[0044] If the magnitude of congestion does satisfy the threshold value, then the method 600 includes determining a decongestion rate and a decongestion time based on at least one of the sustainable rate of data transmission and the magnitude of congestion, as indicated at 608. For example, the decongestion time 112 of FIG. 1 may be proportional to the APTO value with a constant of proportionality F, and the decongestion rate 110 may be proportional to the ARR 108 with a constant of proportionality equal to a ratio of (F-1) to F. The method 600 includes sending data at the decongestion rate from the transmitter to the receiver for the decongestion time, as indicated at 610. For example, data may be sent at the decongestion rate 110 from the transmitter to the receiver for the decongestion time 112, as depicted in FIG. 1 and FIG. 3. The method 600 also includes adjusting a data transmission rate at the transmitter to the sustainable rate of data transmission after the decongestion time is ended, as indicated at 612. For example, the data transmission rate at the transmitter may be adjusted to the ARR 108 after the decongestion time 112 is ended, as depicted in FIG. 1 and FIG. 3.

[0045] Referring to FIG. 7, a flow diagram of another particular illustrative embodiment of a method to adapt to network congestion is depicted and generally designated 700. The method 700 includes identifying congestion in a transmission path between a sender and a receiver using parameters provided in a feedback message from the receiver to the sender, the parameters including a magnitude of congestion and a sustainable rate of data transmission, as indicated at 702. For example, the APTO_ARR feedback message 310 may be received at a transmitter, the APTO_ARR feedback

message 310 including a magnitude of congestion, the APTO value, and the sustainable rate of data transmission, the ARR 108, as depicted in FIG. 1 and FIG. 3. The method 700 includes determining a decongestion rate and a decongestion time based on at least one of the sustainable rate of data transmission and the magnitude of congestion, as indicated at 704. For example, the decongestion time 112 may be proportional to the APTO value with a constant of proportionality F, and the decongestion rate 110 may be proportional to the ARR 108 with a constant of proportionality equal to a ratio of (F-1) to F. The method 700 includes decongesting the transmission path by sending data at the decongestion rate from the sender to the receiver for the decongestion time, as indicated at 706. For example, data may be sent at the decongestion rate 110 from the transmitter to the receiver for the decongestion time 112, as depicted in FIG. 1 and FIG. 3. The method 700 includes adjusting a data transmission rate in the transmission path to the sustainable rate of data transmission after the decongestion time is ended, as indicated at 708. For example, the data transmission rate at the transmitter may be adjusted to the ARR 108 after the decongestion time 112 is ended, as depicted in FIG. 1 and FIG. 3.

[0046] In a particular embodiment, the magnitude of congestion is an arrival-to-playout time offset (APTO) value measured at the receiver. In a particular embodiment, the sustainable rate of data transmission 108 is an average received rate (ARR) of data transmission of data received at the receiver, as shown in FIG. 1. In a particular embodiment, the sustainable rate of data transmission 108 is the maximum rate of data transmission that does not cause congestion of the transmission path. In a particular embodiment, decongesting the transmission path removes an amount of backlogged data, the amount of backlogged data being substantially equal to a product of the decongestion time 112 and a difference between the ARR 108 and the decongestion rate 110, as shown in FIGs. 1-3.

[0047] In a particular embodiment, the decongestion time 112 is proportional to the APTO value, with a constant of proportionality F, and the decongestion rate 110 is proportional to the ARR, with a constant of proportionality equal to a ratio of (F-1) to F. The method 700 may also include determining a quantity of insertable data that can be inserted into the transmission path before the transmission path starts to become congested, where insertion of the insertable data into the transmission path is distributed

over a time period that is proportional to the decongestion time 112 with a constant of proportionality R. In a particular embodiment, another rate of sending data may be determined based on the ARR 108 and a product of R and F. In a particular embodiment, another rate of sending data may be equal to the product of the ARR 108 and the sum of one and the reciprocal of the product of R and F. In other particular embodiments, the period of inserting the insertable data may be a constant or may be set adaptively based on other parameters or criteria, such as quality of service considerations, total network usage, the number of users, the amount of usage per user, and the like.

[0048] Referring to FIG. 8, a flow diagram of yet another particular illustrative embodiment of a method to adapt to network congestion is depicted and generally designated 800. The method 800 includes receiving a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR) in a feedback message at a transmitter, as indicated at 802. For example, the APTO_ARR feedback message 510 may be received at a transmitter, the APTO_ARR feedback message 510 including the APTO value and the ARR, as depicted in FIG. 5. The method 800 includes increasing a sending rate at the transmitter to an increased sending rate determined based on at least one of the APTO value and the ARR, as indicated at 804. For example, the sending rate at the transmitter may be increased to another rate 508, as shown in FIG. 5, and another rate 508 may be proportional to the ARR.

[0049] The method 800 includes determining a detection time based on at least one of the APTO value and the ARR, as indicated at 806. For example, the detection time 502 of FIG. 5 may be proportional to the absolute value of the APTO value with a constant of proportionality F. The method 800 includes continuing transmission at the increased sending rate after the detection time, as indicated at 808. For example, data may be sent at another rate 508 from the transmitter 503 to the receiver 501 after the detection time 502, as depicted in FIG. 5.

[0050] The method 800 may also include determining a quantity of insertable data that can be inserted into a transmission path before the transmission path starts to become congested, wherein insertion of the insertable data into the transmission path is distributed over a time period that is proportional to the detection time 502 of FIG. 5

with a constant of proportionality R. The method 800 may also include adding insertable data distributed over a data insertion time period that is proportional to the detection time 502 of FIG. 5 with a constant of proportionality R, where the detection time 502 is proportional to the absolute value of the APTO value with a constant of proportionality F, where the increased sending rate 508 is determined based on the ARR and a product of R and F. In a particular embodiment, the increased sending rate 508 of sending data may be equal to the product of the ARR and the sum of one and the reciprocal of the product of R and F. In another particular embodiment, the insertable data may be inserted over a fixed or constant detection time value.

[0051] Referring to FIG. 9, a flow diagram of still another particular illustrative embodiment of a method to adapt to network congestion is depicted and generally designated 900. The method 900 includes sending a feedback message from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR), as indicated at 902. For example, the APTO_ARR feedback message 310 may be sent to a transmitter, the APTO_ARR feedback message 310 including the APTO value and the ARR 108, as depicted in FIG. 1 and FIG. 3.

[0052] The method 900 includes receiving data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR, as indicated at 904. For example, data may be received at the receiver sent by the transmitter at the decongestion rate 110 for the decongestion time 112, as depicted in FIG. 1 and FIG. 3. The decongestion time 112 may be proportional to the APTO value with a constant of proportionality F, and the decongestion rate 110 may be proportional to the ARR 108 with a constant of proportionality equal to a ratio of (F-1) to F. The method 900 includes, after the decongestion time is ended, receiving data at the receiver sent by the transmitter at a sustained data rate equal to the ARR, as indicated at 906. For example, the data transmission rate at the transmitter may be adjusted to the ARR 108 after the decongestion time 112 is ended, as depicted in FIG. 1 and FIG. 3.

[0053] In a particular embodiment, the decongestion time 112 is proportional to the APTO value with a constant of proportionality F, and the decongestion rate 110 is proportional to the ARR 108 with a constant of proportionality equal to a ratio of (F-1) to F. The method 900 may also include receiving additional inserted data at the receiver, the additional inserted data received during an insertion time period. In a particular embodiment, the insertion time period may be proportional to the decongestion time 112 with a constant of proportionality R. In other particular embodiments, the insertion time period may be constant or may be set adaptively based on other measurements or criteria, such as quality of service considerations, total network usage, the number of users, the amount of usage per user, and the like.

[0054] Referring to FIG. 10, a diagram of a particular illustrative embodiment of a system that is adaptable to network congestion is depicted and generally designated 1000. The system 1000 includes a transmitter 1002 configured to transmit data 1006. The system 1000 also includes means for receiving the data 1006, such as a receiver 1004 configured to receive the data 1006. The means for receiving may include a digital data receiver, a digital data packet receiver, a digital video data receiver, a digital video data packet receiver, a wireless data receiver, a wireless video data receiver, and the like. The transmitter 1002 is further configured to receive a measured arrival-to-playout time offset (APTO) value 1005 and an average received rate (ARR) 1007 in a feedback message 1008 from the receiver 1004, initiate decongestion by reducing a sending rate at the transmitter 1002 to a decongestion rate determined based on at least one of the APTO value 1005 and the ARR 1007, and end decongestion transmission at the decongestion rate after a decongestion time determined based on at least one of the APTO value 1005 and the ARR 1007. The APTO value 1005 may be measured at the receiver 1004 using an APTO measurement device 1016. The ARR 1007 may be measured at the receiver 1004 using an ARR measurement device 1018. The feedback message 1008 may be sent from the receiver 1004 using a feedback message sender 1020. The feedback message 1008 may be received at the transmitter 1002 using a feedback message receiver 1014. The sending rate at the transmitter 1002 may be reduced to the decongestion rate using a sending rate adjuster 1010. Ending the decongestion transmission at the decongestion rate after the decongestion time may be determined using a decongestion timer 1012.

[0055] In a particular embodiment, the decongestion time 112 of FIG. 1 may be proportional to the APTO value 1005 with a constant of proportionality F, and the decongestion rate 110 may be proportional to the ARR 1007 with a constant of proportionality equal to a ratio of (F-1) to F. In a particular embodiment, the transmitter 1002 may be further configured to receive the APTO value 1005 and the ARR 1007 in a single feedback message 1008 from the receiver 1004, where the transmitter 1002 is further configured to add insertable data distributed over an insertion time period that is proportional to the decongestion time 112 with a constant of proportionality R, and where another rate of sending data is determined based on the ARR 1007 and a product of R and F. In other particular embodiments, the insertion time period may be constant or may be set adaptively based on other measurements or criteria, such as quality of service considerations, total network usage, the number of users, the amount of usage per user, and the like.

[0056] The transmitter 1002 may include a memory 1028 and a cache memory 1030, which are coupled to a processor, such as a digital signal processor (DSP) 1024. The memory 1028 or the cache memory 1030 may include computer executable instructions that are operative to cause a computer, such as the digital signal processor 1024, to perform various operations. A coder/decoder (CODEC) 1026 may also be coupled to the digital signal processor 1024. A wireless controller 1022 may be coupled to the digital signal processor 1024 and to a wireless antenna (not shown). An input/output device 1032 may also be coupled to the digital signal processor 1024. The sending rate adjuster 1010 and the decongestion timer 1012 may be implemented in hardware, such as by dedicated circuitry, or may be executed by the digital signal processor 1024.

[0057] The receiver 1004 may include a memory 1040 and a cache memory 1042, which are coupled to a processor, such as a digital signal processor (DSP) 1036. The memory 1040 or the cache memory 1042 may include computer executable instructions that are operative to cause a computer, such as the digital signal processor 1036, to perform various operations, such as generating the feedback message 1008. A coder/decoder (CODEC) 1038 may also be coupled to the digital signal processor 1036. A wireless controller 1034 may be coupled to the digital signal processor 1036 and to a wireless antenna (not shown). An input/output device 1044 may also be coupled to the

digital signal processor 1036. The APTO measurement device 1016 and the ARR measurement device 1018 may be implemented in hardware, such as by dedicated circuitry, or may be executed by the digital signal processor 1036.

[0058] Those of skill would further appreciate that the various illustrative logical blocks, configurations, modules, circuits, and algorithm steps described in connection with the embodiments disclosed herein may be implemented as electronic hardware, computer software, or combinations of both. To clearly illustrate this interchangeability of hardware and software, various illustrative components, blocks, configurations, modules, circuits, and steps have been described above generally in terms of their functionality. Whether such functionality is implemented as hardware or software depends upon the particular application and design constraints imposed on the overall system. Skilled artisans may implement the described functionality in varying ways for each particular application, but such implementation decisions should not be interpreted as causing a departure from the scope of the present disclosure.

[0059] The steps of a method or algorithm described in connection with the embodiments disclosed herein may be embodied directly in hardware, in a software module executed by a processor, or in a combination of the two. A software module may include instructions executed by a processor that reside in random access memory (RAM), flash memory, read-only memory (ROM), programmable read-only memory (PROM), erasable programmable read-only memory (EPROM), electrically erasable programmable read-only memory (EEPROM), registers, hard disk, a removable disk, a compact disk read-only memory (CD-ROM), or any other form of storage medium known in the art. An exemplary storage medium is coupled to the processor such that the processor can read information from, and write information to, the storage medium. In the alternative, the storage medium may be integral to the processor. The processor and the storage medium may reside in an application-specific integrated circuit (ASIC). The ASIC may reside in a computing device or a user terminal. In the alternative, the processor and the storage medium may reside as discrete components in a computing device or user terminal.

[0060] The previous description of the disclosed embodiments is provided to enable any person skilled in the art to make or use the disclosed embodiments. Various

modifications to these embodiments will be readily apparent to those skilled in the art, and the generic principles defined herein may be applied to other embodiments without departing from the spirit or scope of the disclosure. Thus, the present disclosure is not intended to be limited to the embodiments shown herein but is to be accorded the widest scope possible consistent with the principles and novel features as defined by the following claims.

CLAIMS:

1. A method comprising:
receiving a feedback message at a transmitter, the feedback message including an indication of a magnitude of congestion and a sustainable rate of data transmission at a receiver;
determining a decongestion rate and a decongestion time based on at least one of the sustainable rate of data transmission and the magnitude of congestion when the magnitude of congestion satisfies a threshold value;
sending data at the decongestion rate from the transmitter to the receiver for the decongestion time; and
adjusting a data transmission rate at the transmitter to the sustainable rate of data transmission after the decongestion time is ended.

2. The method of claim 1, wherein the magnitude of congestion is measured by an arrival-to-playout time offset (APTO) value and the sustainable rate of data transmission is measured by an average received rate (ARR) of data received at the receiver.

3. The method of claim 2, wherein sending the data from the transmitter to the receiver at the decongestion rate for the decongestion time removes an amount of congested data, the amount of congested data removed being substantially equal to a product of the decongestion time and a difference between the ARR and the decongestion rate.

4. The method of claim 2, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F.

5. The method of claim 1, further comprising:
increasing the transmission rate of the transmitter to a rate higher than the sustainable rate of data transmission when the magnitude of congestion is less than the threshold value.

6. A method comprising:
identifying congestion in a transmission path between a sender and a receiver using parameters provided in a feedback message from the receiver to the sender, the parameters including a magnitude of congestion and a sustainable rate of data transmission;
determining a decongestion rate and a decongestion time based on at least one of the sustainable rate of data transmission and the magnitude of congestion;
decongesting the transmission path by sending data at the decongestion rate from the sender to the receiver for the decongestion time; and
adjusting a data transmission rate in the transmission path to the sustainable rate of data transmission after the decongestion time is ended.

7. The method of claim 6, wherein the magnitude of congestion is an arrival-to-playout time offset (APTO) value measured at the receiver.

8. The method of claim 7, wherein the sustainable rate of data transmission is an average received rate (ARR) of data transmission of data received at the receiver.

9. A method comprising:
receiving a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR) in a feedback message at a transmitter;
increasing a sending rate at the transmitter to an increased sending rate determined based on at least one of the APTO value and the ARR; and
continuing transmission at the increased sending rate after a detection time, wherein the detection time is determined based on at least one of the APTO value, the ARR, a constant value, or an adaptively set value.

10. The method of claim 9, further comprising:
determining a quantity of insertable data that can be inserted into a transmission path before the transmission path starts to become congested, wherein insertion of the insertable data into the transmission path is distributed over a time period that is proportional to the detection time with a constant of proportionality R.

11. The method of claim 9, further comprising:
adding insertable data distributed over a data insertion time period that is proportional to the detection time with a constant of proportionality R, wherein the detection time is proportional to the APTO value with a constant of proportionality F, wherein the increased sending rate is determined based on the ARR and a product of R and F.

12. A computer-readable medium including computer executable instructions that are operative to cause the computer to:
send a feedback message from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR);
receive data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR; and
receive data at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

13. The computer-readable medium of claim 12, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F, and the decongestion rate is proportional to the ARR with a constant of proportionality equal to a ratio of (F-1) to F.

14. The computer-readable medium of claim 13, wherein the computer executable instructions are further operative to cause the computer to receive additional inserted data at the receiver, the additional inserted data received during an insertion time period.

15. A method comprising:
sending a feedback message from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR);
receiving data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR; and
receiving data at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

16. The method of claim 15, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F, and the decongestion rate is proportional to the ARR with a constant of proportionality equal to a ratio of (F-1) to F.

17. An apparatus comprising:
means for sending a feedback message from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR);
means for receiving data at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR; and
means for receiving data at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

18. The apparatus of claim 17, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F, and the decongestion rate is proportional to the ARR with a constant of proportionality equal to a ratio of (F-1) to F.

19. An apparatus comprising:

a processor configured to generate a feedback message to send from a receiver to a transmitter, the feedback message including a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR);
wherein data is received at the receiver sent by the transmitter at a decongestion rate determined based on at least one of the APTO value and the ARR for a decongestion time determined based on at least one of the APTO value and the ARR; and
wherein data is received at the receiver sent by the transmitter at the maximum sustained data rate of the ARR after the decongestion time is ended.

20. The apparatus of claim 19, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F, and the decongestion rate is proportional to the ARR with a constant of proportionality equal to a ratio of (F-1) to F.

21. An apparatus comprising:
a transmitter configured to transmit data, to receive a measured arrival-to-playout time offset (APTO) value and an average received rate (ARR) in a feedback message, to initiate decongestion by reducing a sending rate at the transmitter to a decongestion rate determined based on at least one of the APTO value and the ARR, and to end decongestion transmission at the decongestion rate after a decongestion time determined based on at least one of the APTO value and the ARR.

22. The apparatus of claim 21, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F, and the decongestion rate is proportional to the ARR with a constant of proportionality equal to a ratio of (F-1) to F.

23. The apparatus of claim 22, wherein the transmitter is further configured to receive the APTO value and the ARR in a single feedback message, wherein the transmitter is further configured to add insertable data distributed over an insertion time period that is proportional to the decongestion time with a constant of proportionality R, and wherein another rate of sending data is determined based on the ARR and a product of R and F.

24. An apparatus comprising:
means for initiating decongestion by reducing a sending rate at a transmitter to a decongestion rate determined based on at least one of a measured arrival-to-playout

time offset (APTO) value and an average received rate (ARR) that are received in a feedback message; and

means for ending decongestion transmission at the decongestion rate after a decongestion time determined based on at least one of the APTO value and the ARR.

25. The apparatus of claim 24, wherein the decongestion time is proportional to the APTO value with a constant of proportionality F, and the decongestion rate is proportional to the ARR with a constant of proportionality equal to a ratio of (F-1) to F.