



US008315168B2

(12) **United States Patent**
Davari

(10) **Patent No.:** **US 8,315,168 B2**
(45) **Date of Patent:** **Nov. 20, 2012**

(54) **PRIORITY-BASED HIERARCHICAL
BANDWIDTH SHARING**

(75) Inventor: **Shahram Davari**, Saratoga, CA (US)

(73) Assignee: **Broadcom Corporation**, Irvine, CA
(US)

(*) Notice: Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 306 days.

(21) Appl. No.: **12/759,471**

(22) Filed: **Apr. 13, 2010**

(65) **Prior Publication Data**

US 2011/0096666 A1 Apr. 28, 2011

Related U.S. Application Data

(60) Provisional application No. 61/255,841, filed on Oct.
28, 2009.

(51) **Int. Cl.**
G01R 31/08 (2006.01)

(52) **U.S. Cl.** **370/235; 370/232; 370/233**

(58) **Field of Classification Search** **370/235.1,**
370/230, 232, 233, 230.1
See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,412,647	A *	5/1995	Giroux et al.	370/232
6,947,418	B2	9/2005	Boura et al.	
6,970,426	B1	11/2005	Haddock	
7,176,813	B2	2/2007	Kawamata et al.	
7,327,682	B2 *	2/2008	Gandhi et al.	370/235.1
7,453,892	B2	11/2008	Buskirk et al.	
7,467,223	B2	12/2008	Wu et al.	
7,664,028	B1	2/2010	Gingras et al.	
7,680,049	B2 *	3/2010	Olsen et al.	370/235.1

2003/0174650	A1 *	9/2003	Shankar et al.	370/235
2005/0135378	A1	6/2005	Rabie et al.	
2006/0098572	A1	5/2006	Zhang et al.	
2006/0187839	A1	8/2006	Smith et al.	
2007/0070907	A1	3/2007	Kumar et al.	
2007/0153682	A1 *	7/2007	Swenson et al.	370/229
2008/0159135	A1 *	7/2008	Caram	370/230
2010/0054126	A1	3/2010	Kwan et al.	
2011/0112747	A1	5/2011	Downs et al.	
2011/0173015	A1	7/2011	Chapman et al.	

OTHER PUBLICATIONS

Clark, D. D., et al., "Explicit Allocation of Best-Effort Packet Deliv-
ery Service", IEEE/ACM Transactions on Networking, vol. 6, No. 4.,
(Aug. 1998), 12 pages.

Anker, Tal et al., "Hierarchical Bandwidth Sharing", Technical
Report HUI-CSE-LTR-2002-20, The Hebrew University of
Jerusalem, Computer Science, (Feb. 2002), 20 pages.

Blake, S. et al., "An Architecture for Differentiated Services", RFC
2475, (Dec. 1998), 37 pages.

(Continued)

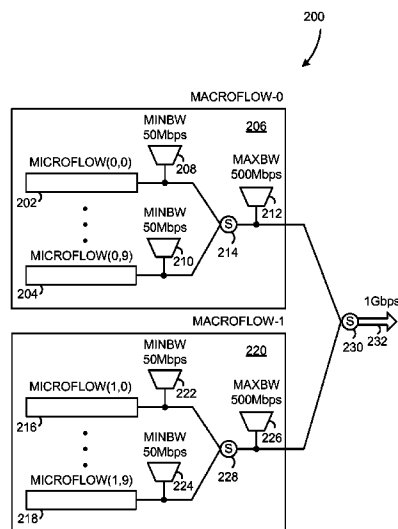
Primary Examiner — Albert T Chou

(74) *Attorney, Agent, or Firm* — Brake Hughes Bellermann
LLP

(57) **ABSTRACT**

Methods and apparatus for communicating data are dis-
closed. The example method includes allocating tokens to a
first token bucket of a first two-rate, three-color meter
(trTCM) at a first rate and allocating tokens to a second token
bucket of the first trTCM at a second rate. The example
method further includes allocating tokens to a first token
bucket of a second trTCM at a third rate and allocating tokens
to a second token bucket of the second trTCM at a fourth rate.
The example method also includes reallocating tokens allo-
cated to the first token bucket of the first trTCM to the first
token bucket of the second trTCM when a token count of the
first token bucket of the first trTCM exceeds a first capacity
and reallocating tokens allocated to the second token bucket
of the first trTCM to the second token bucket of the second
trTCM when a token count of the second token bucket of the
first trTCM exceeds a second capacity.

20 Claims, 19 Drawing Sheets



OTHER PUBLICATIONS

Heinanen, J. et al., "A Single Rate Three Color Marker", IETF RFC 2697, (Sep. 1999), 6 pages.

Heinanen, J. et al., "A Two Rate Three Color Marker", IETF RFC 2698, (Sep. 1999), 5 pages.

Aboul-Magd, O. et al., "A Differentiated Service Two-Rate, Three-Color Marker with Efficient Handling of in-Profile Traffic", IETF RFC 4115, (Jul. 2005), 6 pages.

Floyd, S. et al., "Link-Sharing and Resource Management Models for Packet Networks", IEEE/ACM Transactions on Networking, vol. 3., No. 4., (Aug. 1995), 22 pages.

Bennett, J. et al., "Hierarchical Packet Fair Queuing Algorithms", IEEE/ACM Transactions on Networking, vol. 5, No. 5, (Oct. 1997), 14 pages.

Bennett, J. et al., "WF2Q: Worst-case Fair Weighted Fair Queuing", IEEE Infocom 96, (Mar. 1996), pp. 120-128.

U.S. Appl. No. 12/198,640, Notice of Allowance mailed Jun. 24, 2010, 10 pages.

U.S. Appl. No. 12/881,966 Non-Final Office Action mailed Sep. 4, 2012, 59 pages.

* cited by examiner

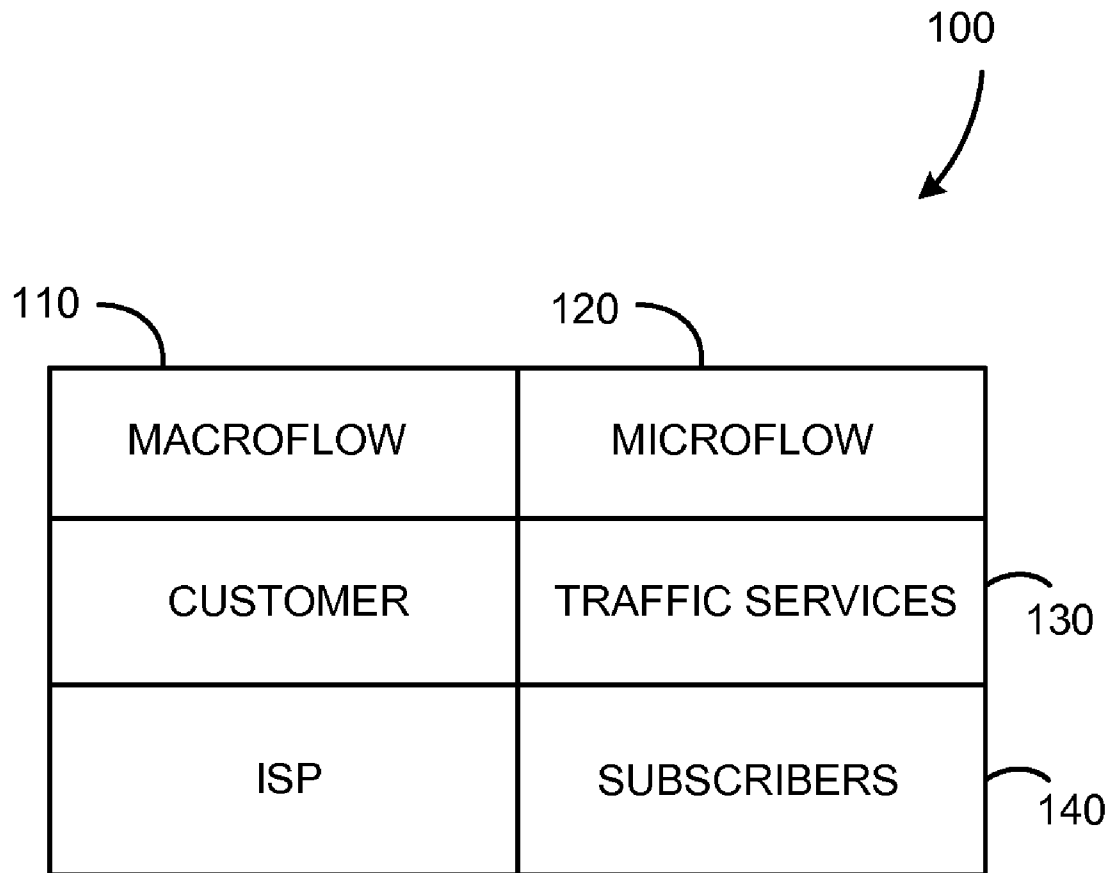


FIG. 1

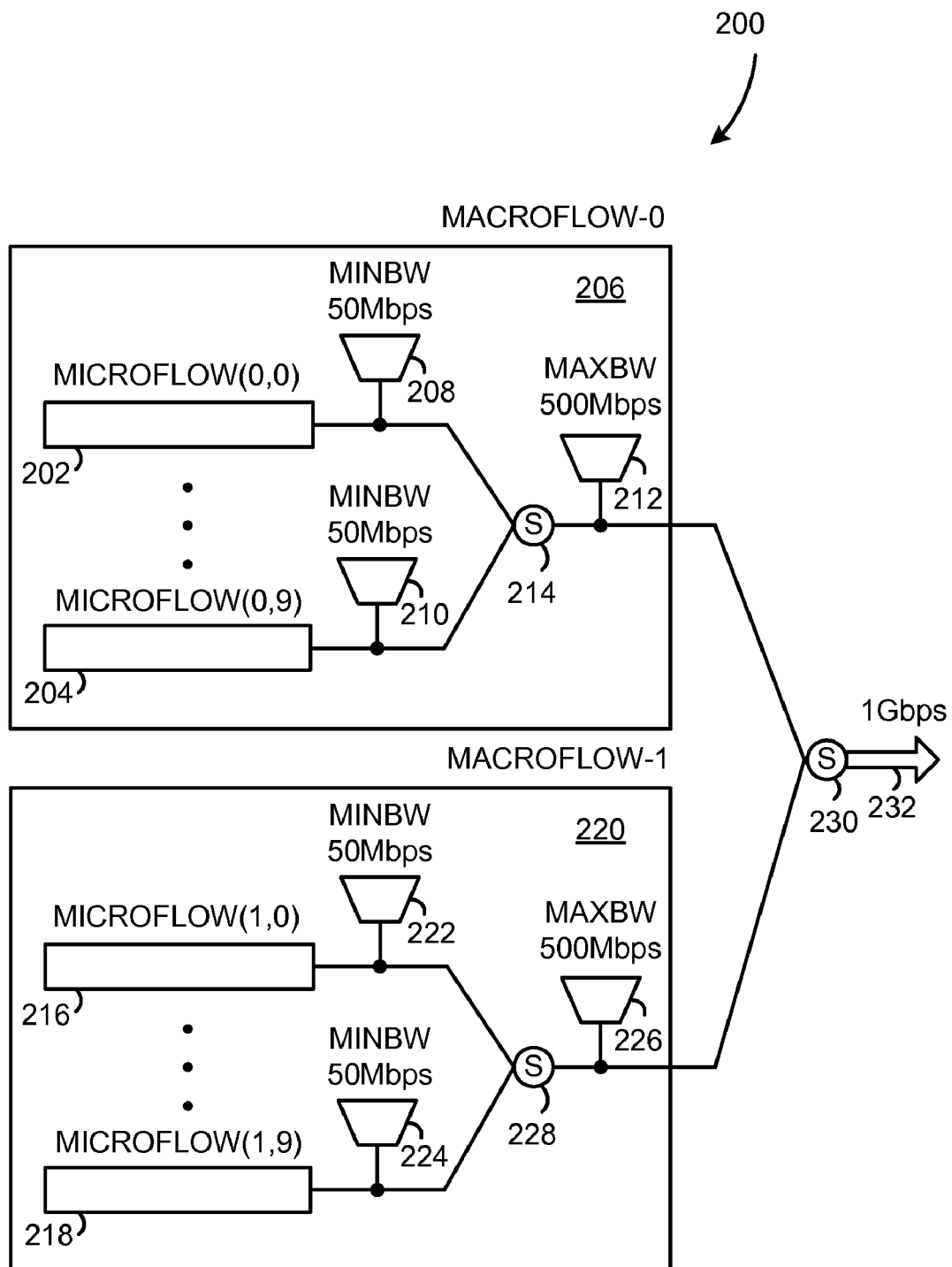


FIG. 2

300

	310 MICRO FLOW	320 MACRO FLOW	330 FINAL COLOR	340 MICRO FLOW METER UPDATE	350 MACRO FLOW METER UPDATE
360	G	G	G	UPDATE	UPDATE
370	G	R	G	UPDATE	UPDATE
380	R	G	G	DO NOT UPDATE	UPDATE
390	R	R	R	DO NOT UPDATE	DO NOT UPDATE

FIG. 3

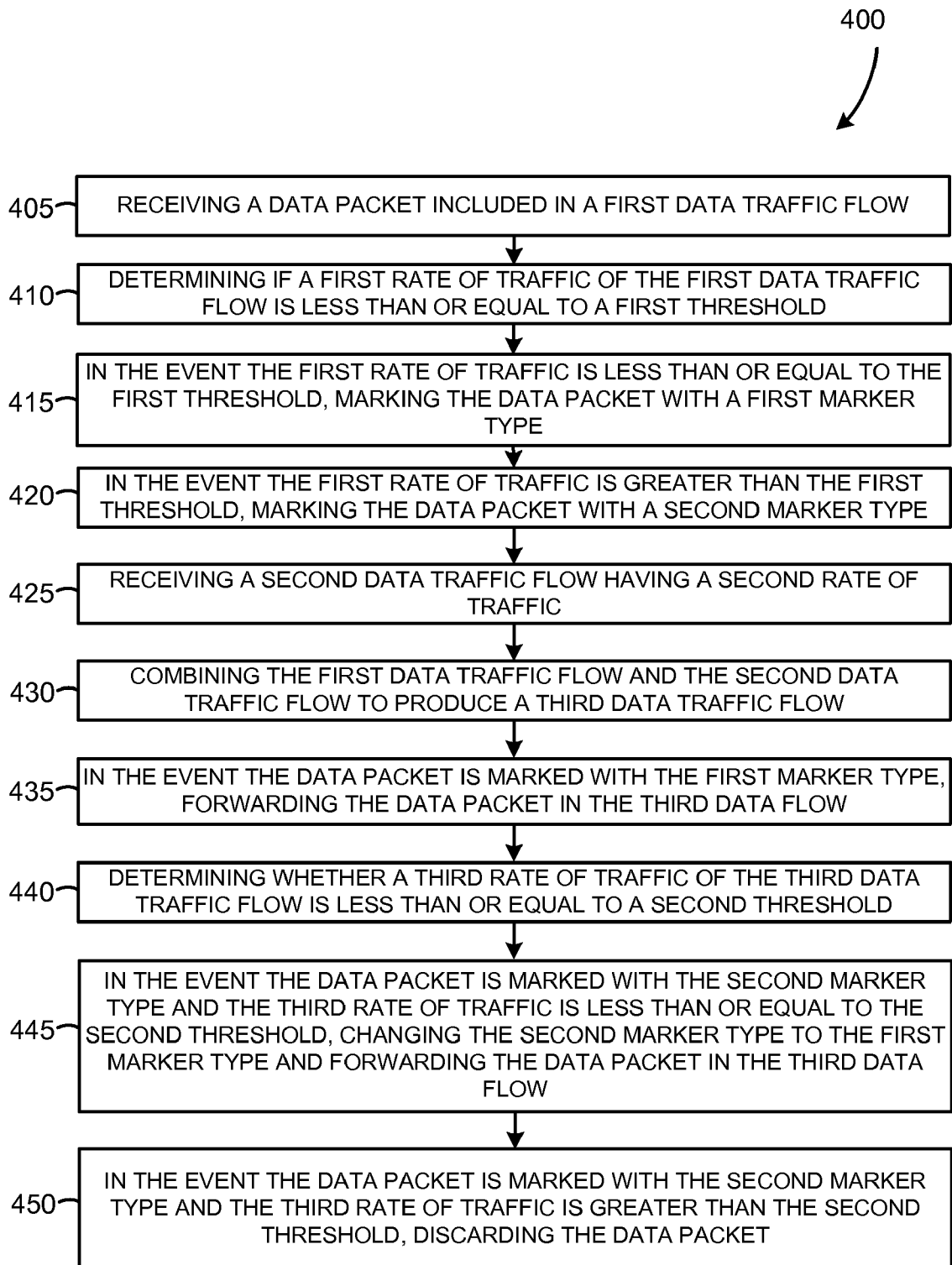


FIG. 4

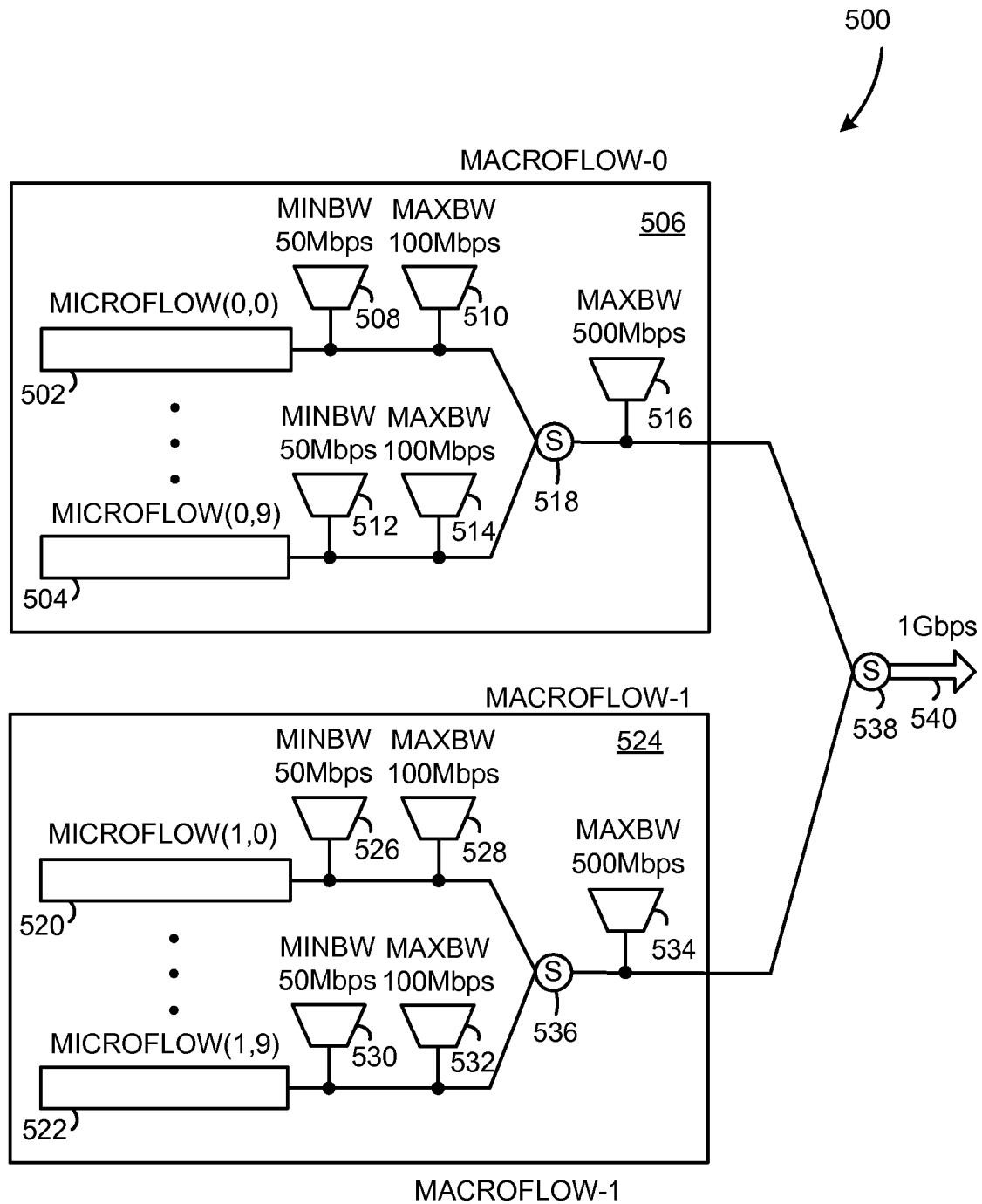


FIG. 5

600
↙

	605 MICRO FLOW	610 MACRO FLOW	615 FINAL COLOR	620 MICRO FLOW METER UPDATE	625 MACRO FLOW METER UPDATE
630	G	G	G	UPDATE COMMITTED BUCKET	UPDATE
635	G	R	G	UPDATE COMMITTED BUCKET	UPDATE
640	Y	G	G	UPDATE EXCESS BUCKET	UPDATE
645	Y	R	R	DO NOT UPDATE	DO NOT UPDATE
650	R	G	R	DO NOT UPDATE	DO NOT UPDATE
655	R	R	R	DO NOT UPDATE	DO NOT UPDATE

FIG. 6

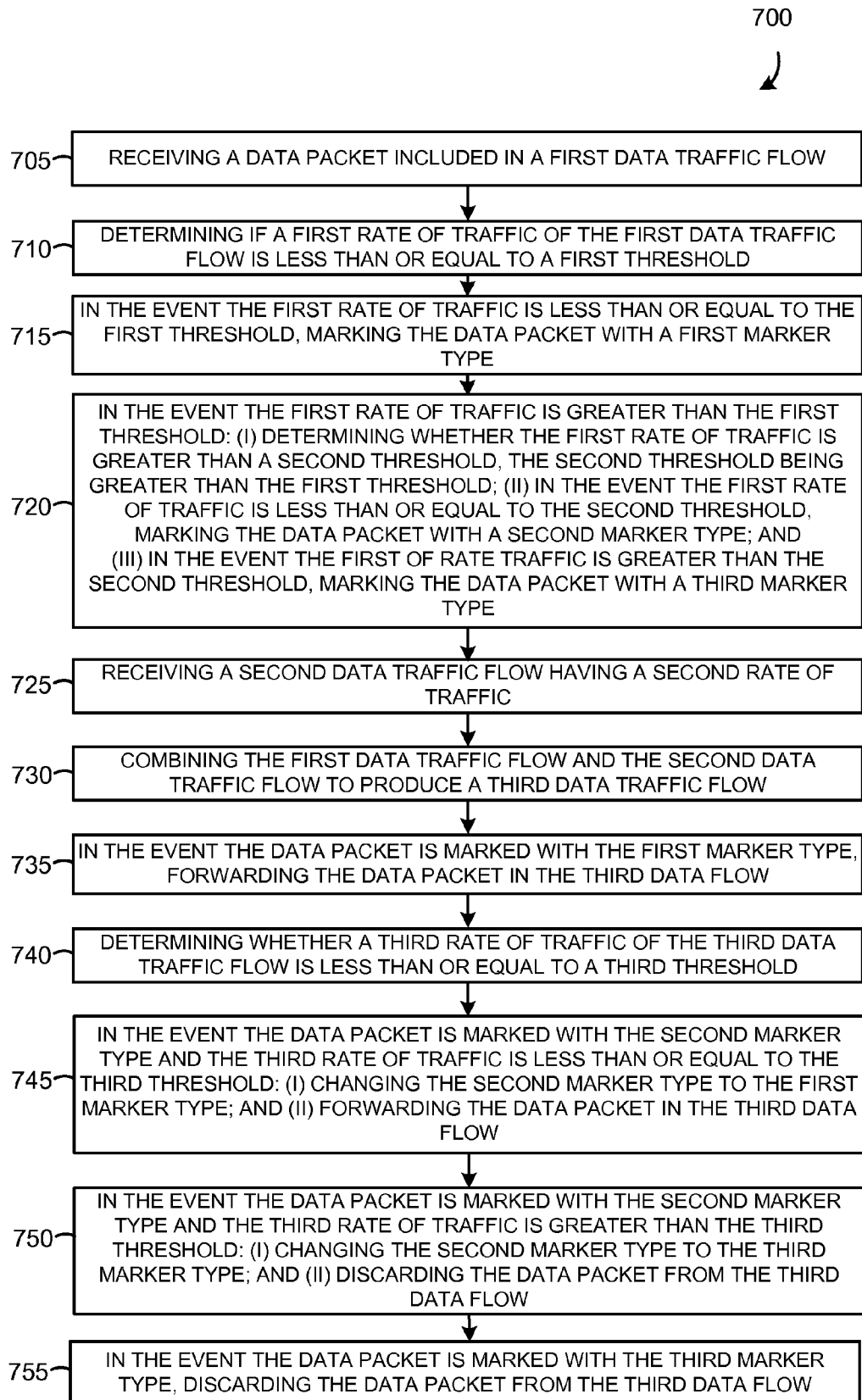


FIG. 7

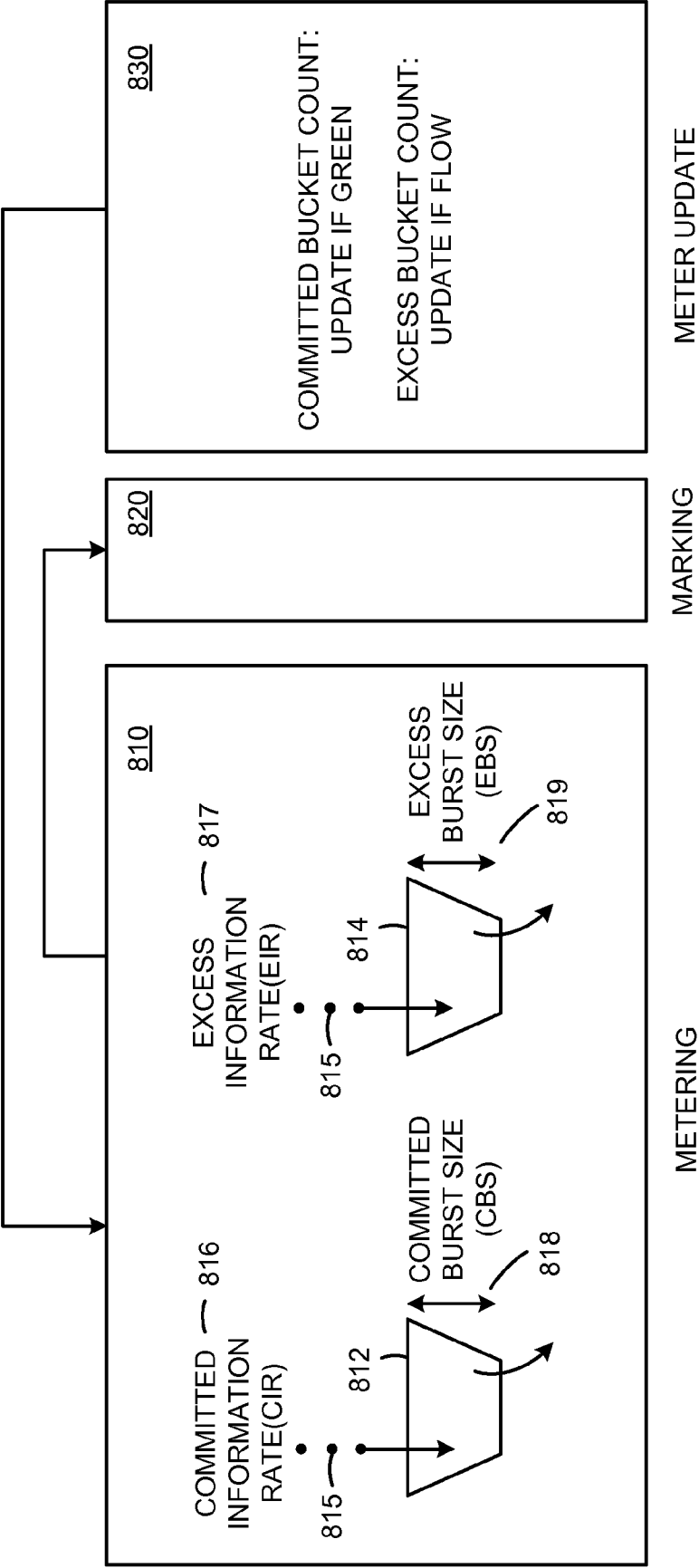


FIG. 8

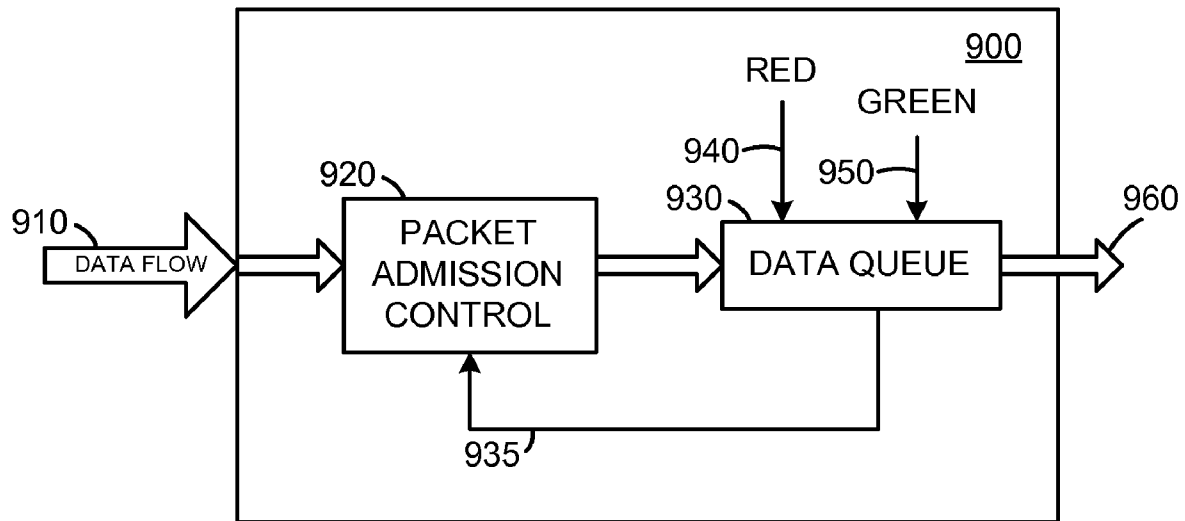


FIG. 9

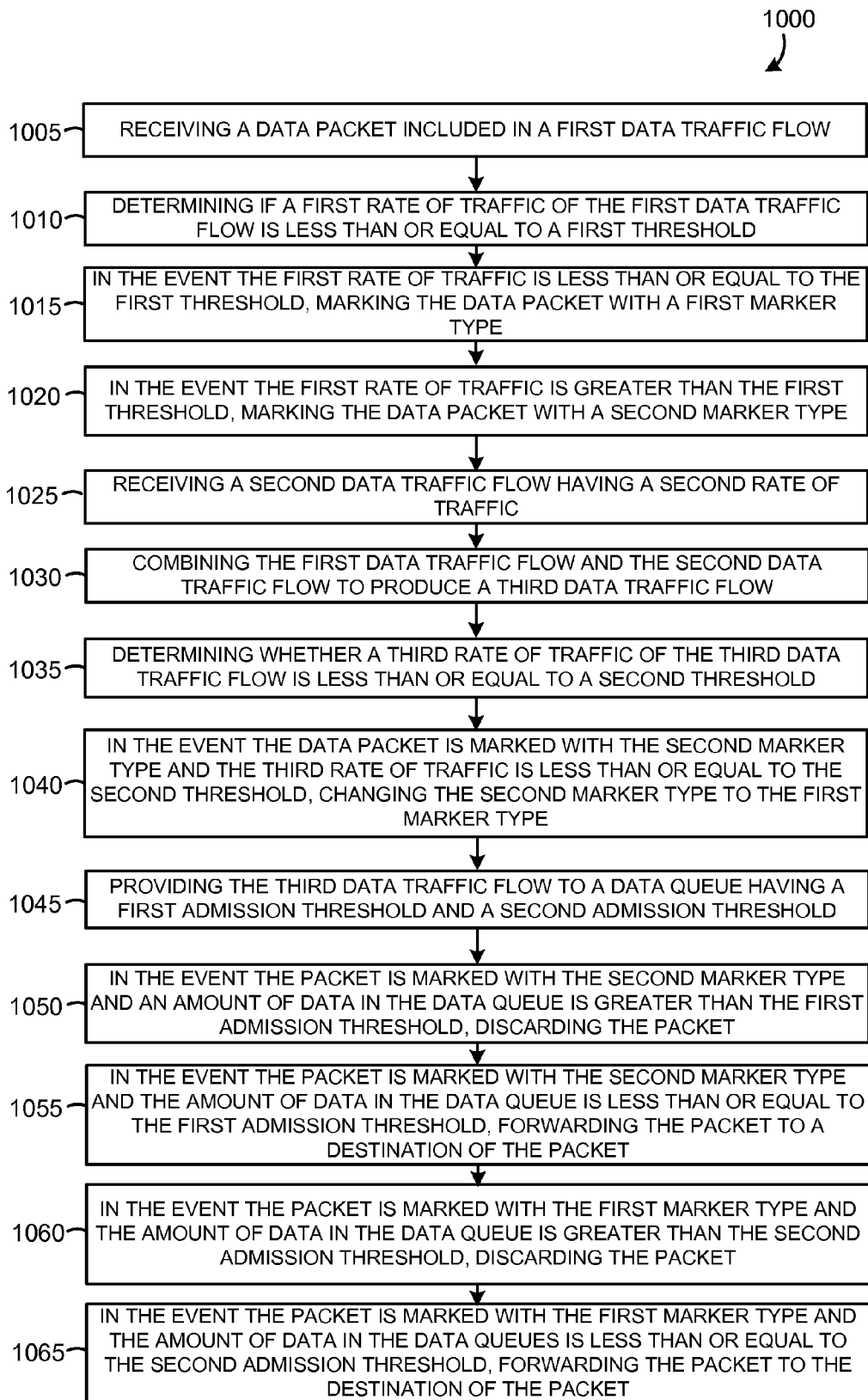


FIG. 10

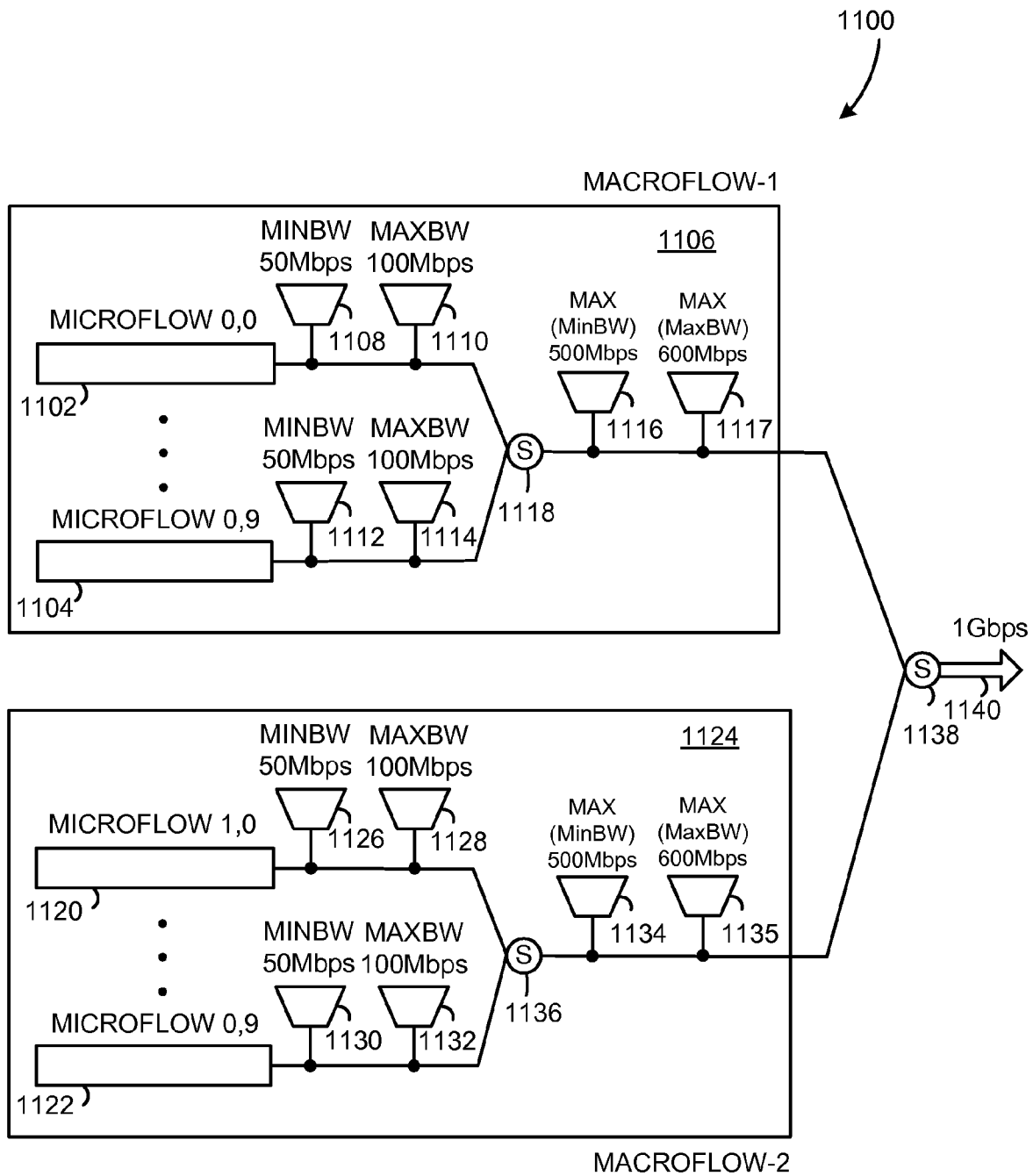


FIG. 11

1200
↙

	1205 MICRO FLOW	1210 MACRO FLOW	1215 FINAL COLOR	1220 MICRO FLOW METER UPDATE	1225 MACRO FLOW METER UPDATE
1230	G	G	G	UPDATE COMMITTED BUCKET	UPDATE COMMITTED BUCKET
1235	G	Y	G	UPDATE COMMITTED BUCKET	UPDATE COMMITTED BUCKET
1240	G	R	G	UPDATE COMMITTED BUCKET	UPDATE COMMITTED BUCKET
1245	Y	G	G	UPDATE EXCESS BUCKET	UPDATE COMMITTED BUCKET
1250	Y	Y	Y	UPDATE EXCESS BUCKET	UPDATE EXCESS BUCKET
1255	Y	R	R	DO NOT UPDATE	DO NOT UPDATE
1260	R	G	R	DO NOT UPDATE	DO NOT UPDATE
1265	R	Y	R	DO NOT UPDATE	DO NOT UPDATE
1270	R	R	R	DO NOT UPDATE	DO NOT UPDATE

FIG.12

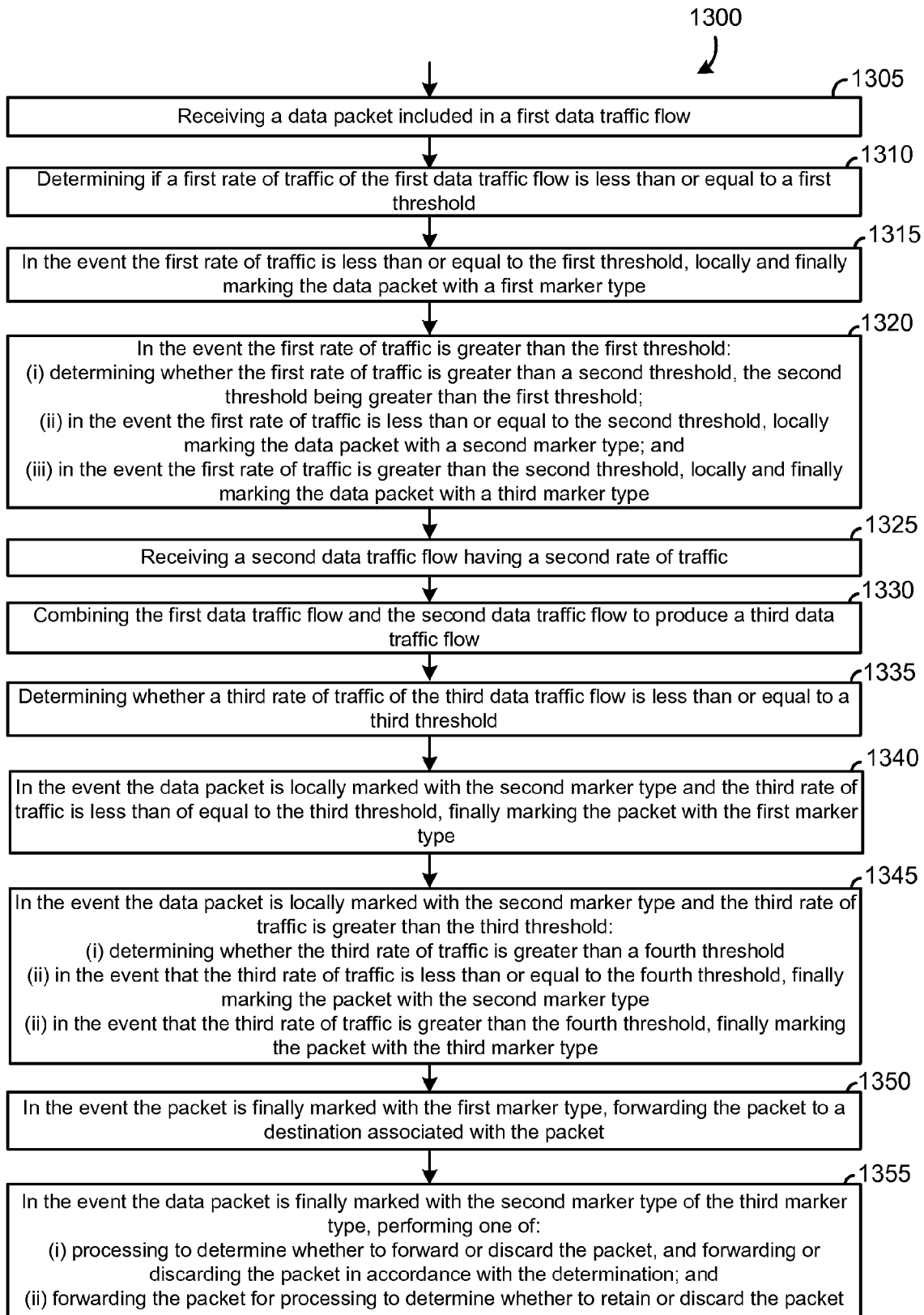


FIG. 13

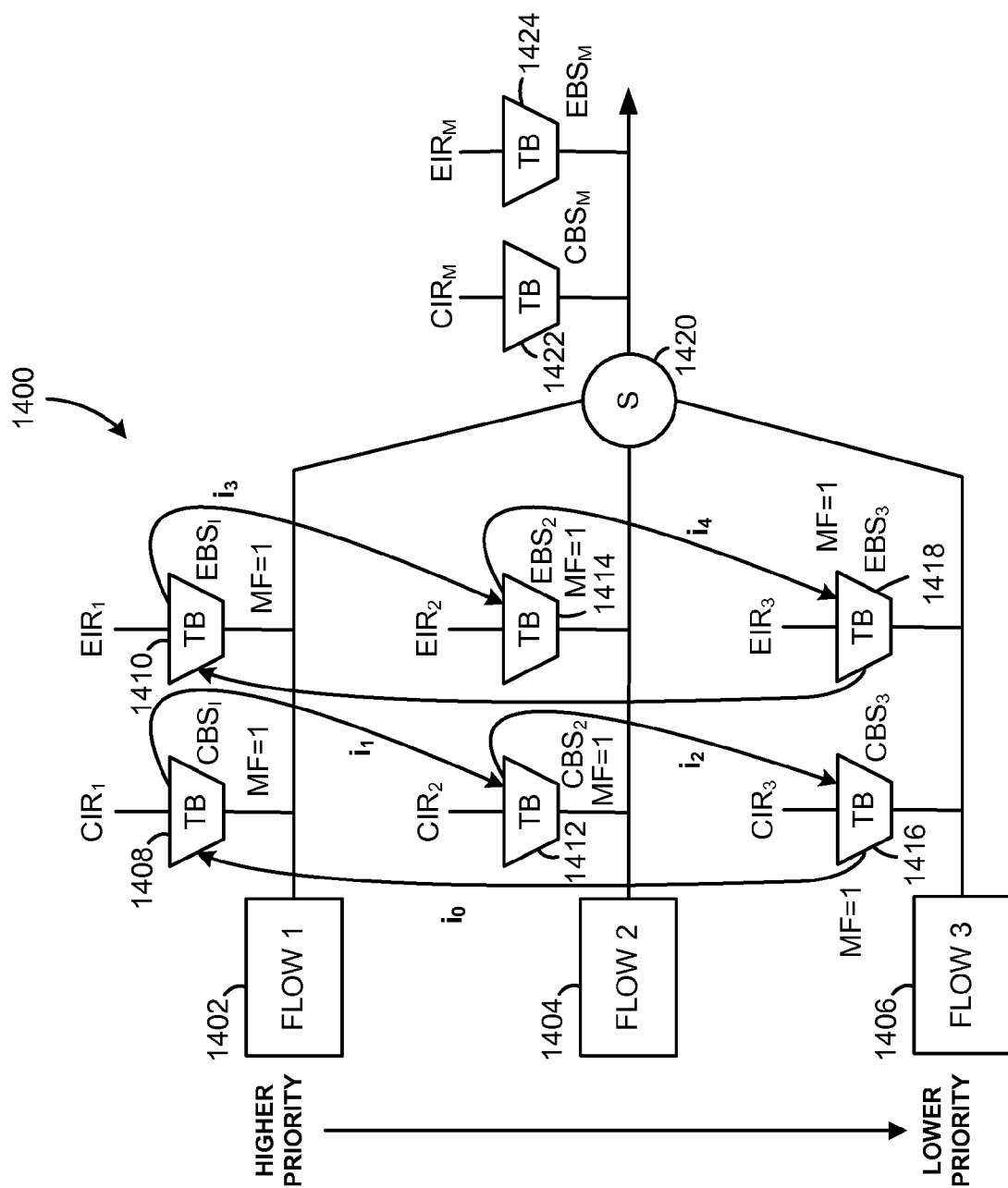


FIG. 14

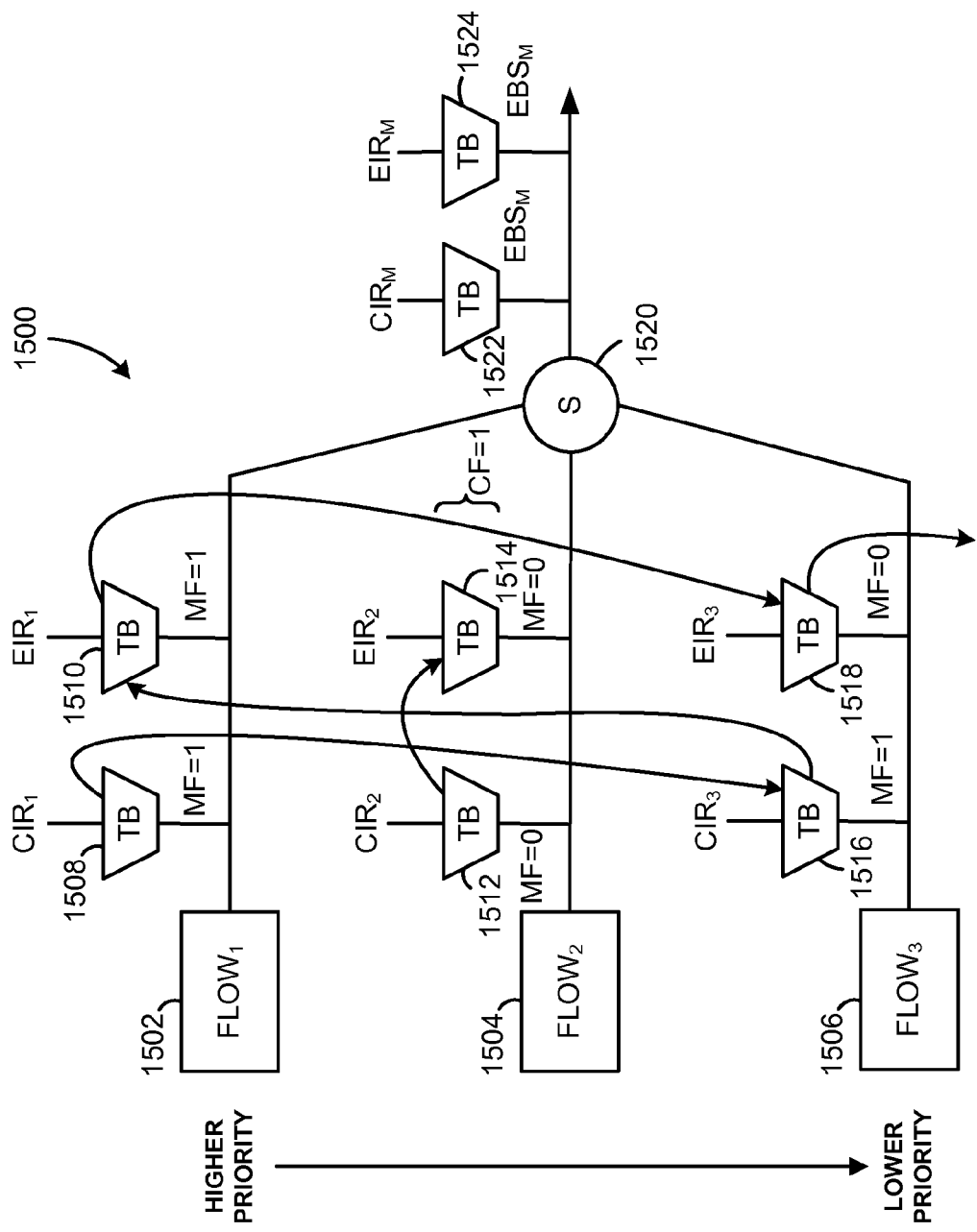


FIG. 15

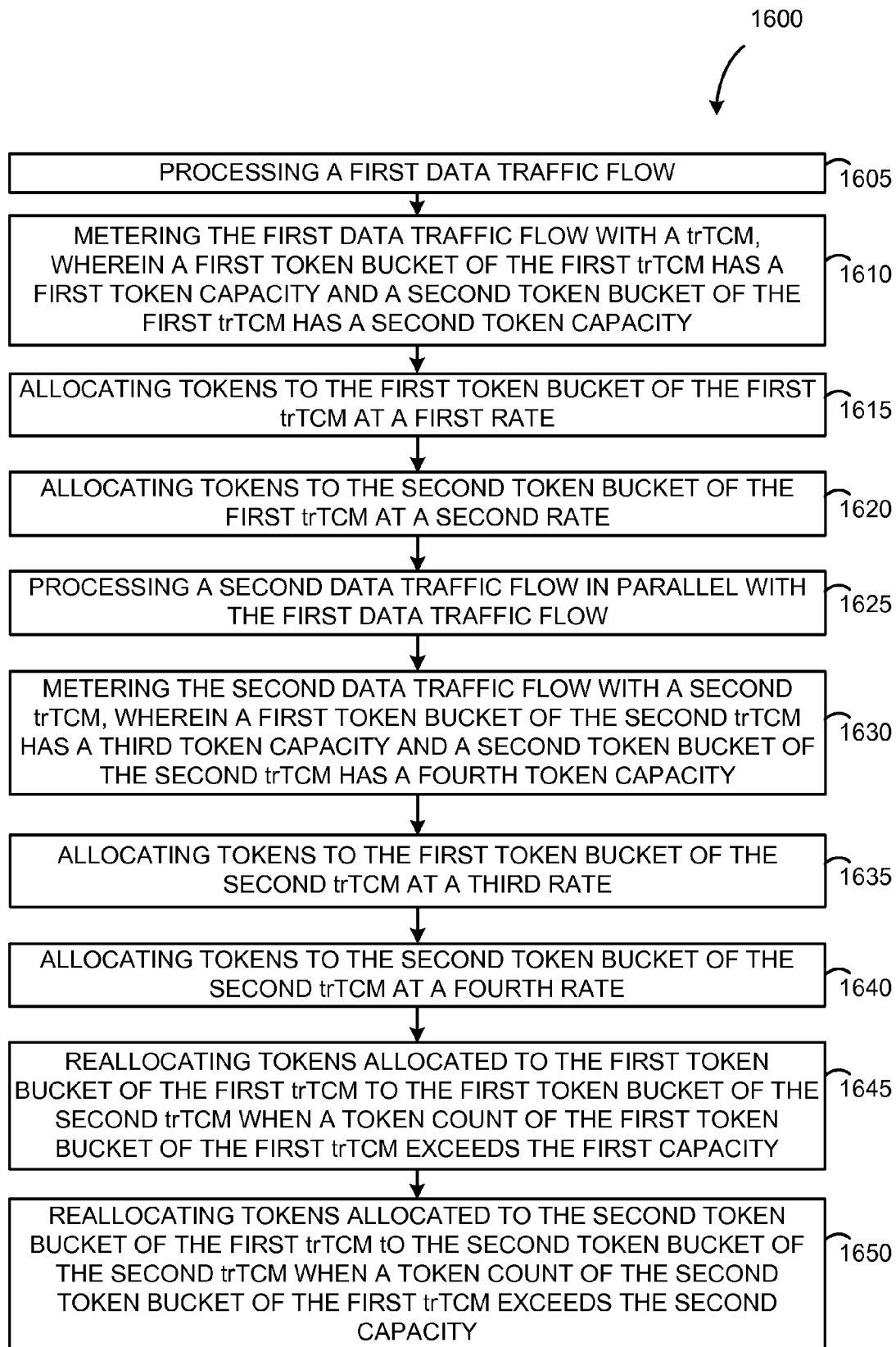


FIG. 16

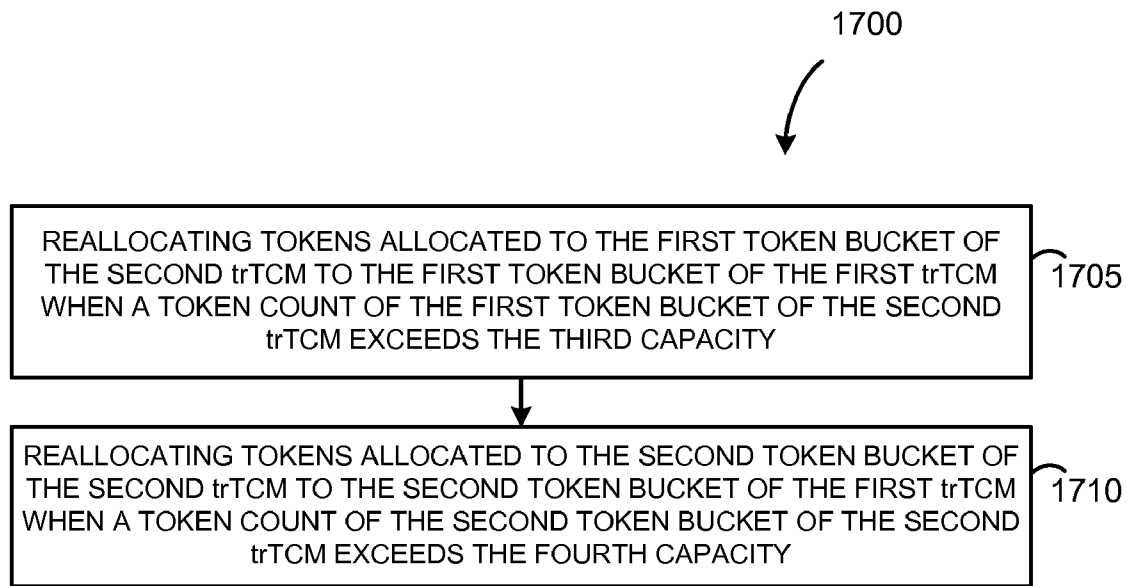


FIG. 17

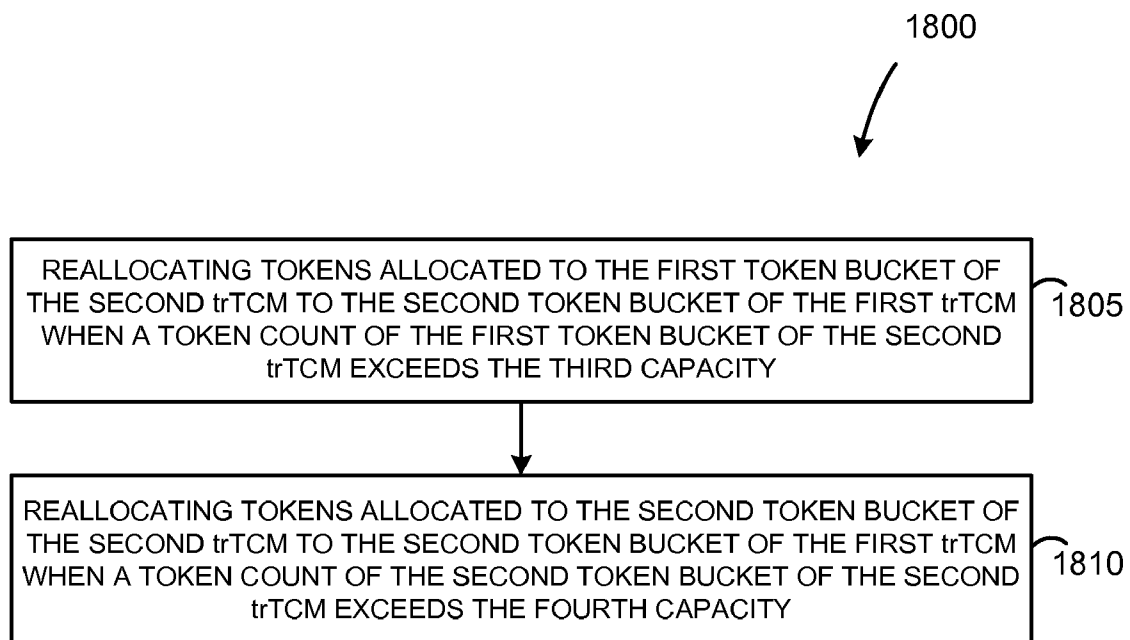


FIG. 18

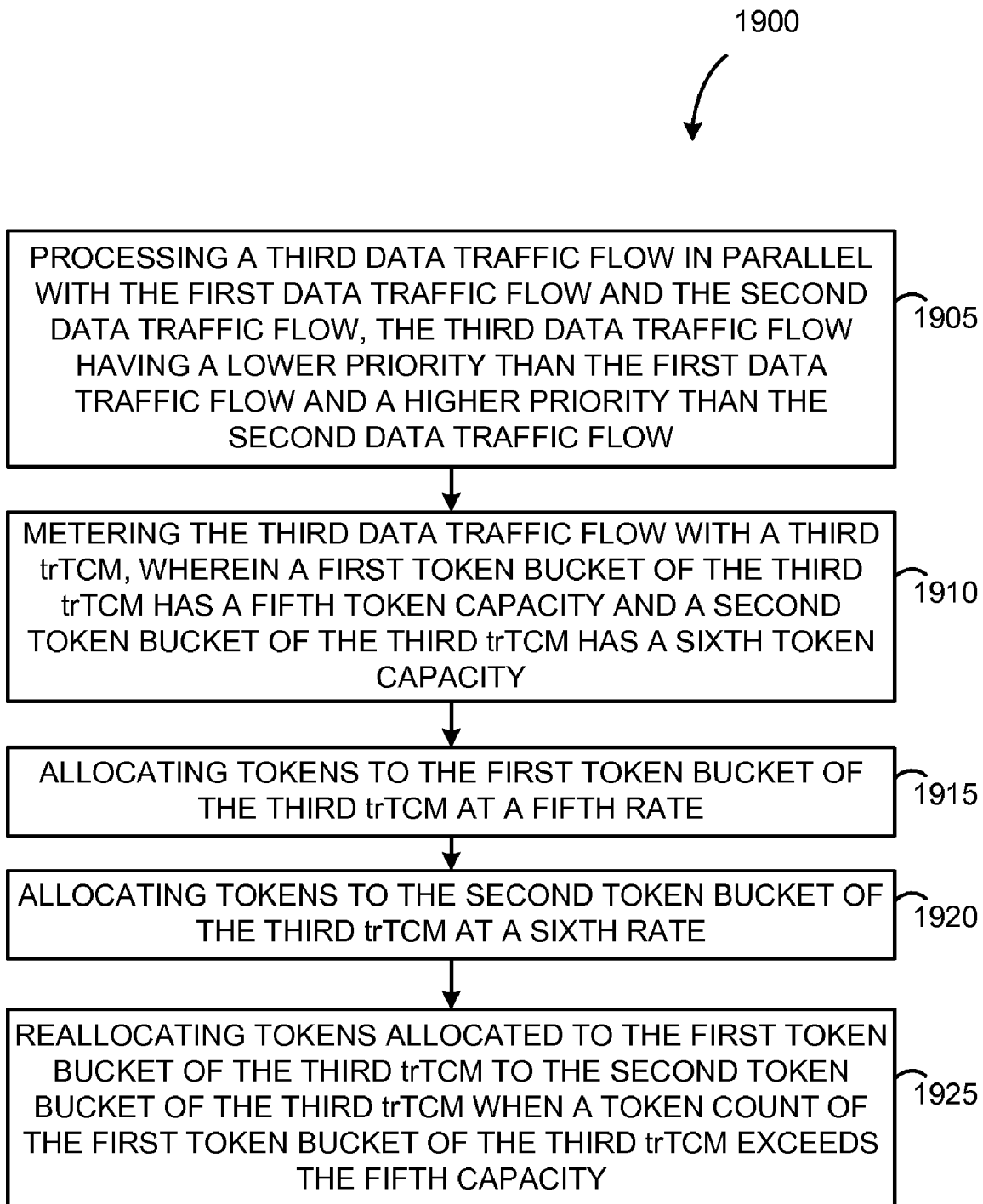


FIG. 19

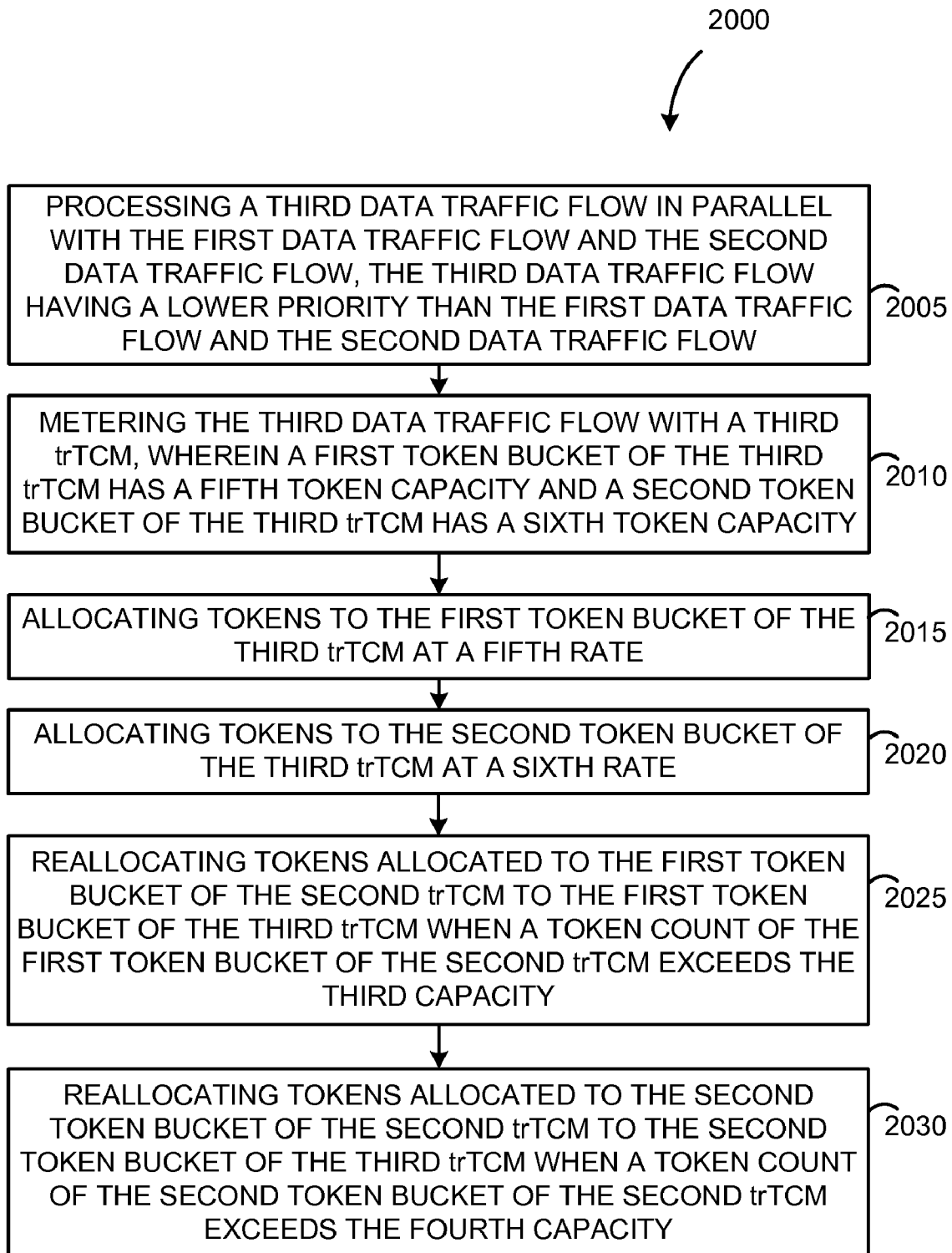


FIG. 20

PRIORITY-BASED HIERARCHICAL BANDWIDTH SHARING

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims benefit under 35 U.S.C. §119 to U.S. Provisional Patent Application Ser. No. 61/255,841, filed on Oct. 28, 2009. The disclosure of U.S. Provisional Patent Application Ser. No. 61/255,841 is incorporated herein reference in its entirety.

TECHNICAL FIELD

This description relates to data and network communications.

BACKGROUND

Data communication applications and the use of data networks continue to grow at a rapid pace. Often networks used for data communication are shared, where different users and/or subscribers communicate data traffic over a common, or shared network. In such situations, data traffic management is typically used to implement predictable bandwidth allocation across the various traffic flows (e.g., among users). Different bandwidth allocation policies may be implemented using such traffic management techniques. For instance, bandwidth may be equally shared across the various traffic flows or bandwidth may be allocated based on an associated class of service for each traffic flow, as two possible examples.

One technique that is used to implement data traffic management in network devices (e.g., network switches or routers), is the use of hierarchical data queues and associated schedulers to control the flow of data traffic. In such an arrangement, respective data queues are used to process each individual traffic flow. For instance, data traffic for each individual user (subscriber) of an Internet Service Provider (ISP) would be processed in a dedicated queue. In such an approach, the associated schedulers then combine the separate traffic flows for each of the individual users (microflows) into one or more larger (e.g., higher bandwidth) traffic flows (macroflows). In such an approach, the hierarchical queues and schedulers are configured to implement bandwidth allocation policies, or perform traffic management. Implementing such bandwidth allocation policies includes deciding which data packets are to be forwarded on to their destination and which packets are to be dropped. These decisions are made, at least in part, based on the specific bandwidth allocation policies being implemented.

However, implementing traffic management using such a hierarchical queuing approach requires implementing complex queuing structures and associated schedulers in network devices that use such techniques to implement bandwidth allocation policies and, therefore, may be cost prohibitive. Further, in network devices that have limited data queuing resources, implementing traffic management using such an approach may be technically impracticable and/or highly inefficient.

SUMMARY

A system and/or method for data communication, substantially as shown in and/or described in connection with at least one of the figures, as set forth more completely in the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a table illustrating example embodiments of data microflows and corresponding data macroflows.

FIG. 2 is a block diagram illustrating an example embodiment of a data communication apparatus for implementing meter-based hierarchical bandwidth sharing.

FIG. 3 is a table illustrating an example embodiment of data packet marking that may be employed in conjunction with the apparatus illustrated in FIG. 2.

FIG. 4 is a flowchart illustrating an example embodiment of a method for data communication that may be implemented in the apparatus illustrated in FIG. 2.

FIG. 5 is a block diagram illustrating another example embodiment of a data communication apparatus for implementing meter-based hierarchical bandwidth sharing.

FIG. 6 is a table illustrating an example embodiment of data packet marking that may be employed in conjunction with the apparatus illustrated in FIG. 5.

FIG. 7 is a flow chart illustrating an example embodiment of a method for data communication that may be implemented in the apparatus illustrated in FIG. 5.

FIG. 8 is a block diagram illustrating an example embodiment of an apparatus that may be used for metering data flow and associated marking of packets.

FIG. 9 is a block diagram of an example embodiment of a data queue with admission control.

FIG. 10 is a flowchart illustrating yet another example embodiment of a method for data communication.

FIG. 11 is a block diagram illustrating another example embodiment of a data communication apparatus for implementing meter-based hierarchical bandwidth sharing.

FIG. 12 is a table illustrating an example embodiment of data packet marking that may be employed in conjunction with the apparatus illustrated in FIG. 11.

FIG. 13 is a flow chart illustrating an example embodiment of a method for data communication that may be implemented in the apparatus illustrated in FIG. 12.

FIG. 14 is a block diagram illustrating a network device in accordance with an example embodiment.

FIG. 15 is a block diagram illustrating a network device in accordance with an example embodiment.

FIG. 16 is a flowchart illustrating a method in accordance with an example embodiment.

FIG. 17 is a flowchart illustrating a method in accordance with an example embodiment.

FIG. 18 is a flowchart illustrating a method in accordance with an example embodiment.

FIG. 19 is a flowchart illustrating a method in accordance with an example embodiment.

FIG. 20 is a flowchart illustrating a method in accordance with an example embodiment.

DETAILED DESCRIPTION

As indicated above, in certain applications, it may be desirable to process data traffic in microflows, where a number of microflows may be combined into one or more higher-bandwidth macroflows (e.g., higher bandwidth than the individual microflows). FIG. 1 is a table 100 illustrating two example situations in which such a microflow/macroflow arrangement might be advantageous when implementing a bandwidth allocation and sharing policy.

In the table 100, column 110 indicates the type of macroflow for each example, while column 120 indicates the associated type of microflows that may be combined to produce the macroflow. For instance, in row 130, the macroflow indi-

3

cated in column 110 is a data traffic flow for a customer, such as an individual network access customer. As shown in row 130, column 120, the microflows that may be combined to produce a customer macroflow are individual traffic services for the customer. Such individual traffic services may include voice data, streaming media, and Internet Protocol data, among any number of other possible traffic services.

In row 140, column 110, the indicated macroflow is a data traffic flow for an Internet Service Provider (ISP). In row 140, column 120, the associated microflows for the ISP macroflow are indicated as customer microflows. In such an embodiment, the microflows may each include data traffic for a respective customer of the ISP. The customer microflows may then be combined to produce the ISP macroflow.

As indicated above, microflows and macroflows may be processed using a hierarchical set of data queues and schedulers, where each microflow and macroflow is processed in a dedicated data queue. Such an approach allows for bandwidth allocation and sharing between the various data traffic flows. For example, bandwidth allocation and sharing may be implemented using hierarchical schedulers in such an approach. However, as was discussed above, such an approach may be cost prohibitive and complicated to implement.

Various example embodiments are described herein for implementing meter-based hierarchical bandwidth sharing that may be implemented in devices with limited queuing and scheduling resources, as compared to a hierarchical queuing structure (such as in network devices used to process and route data traffic). For instance, the embodiments described herein may be implemented using a network device with a single data queue to process a plurality of microflows and, likewise, a macroflow or plurality of macroflows, rather than using a dedicated data queue per microflow and/or macroflow. In other embodiments, a plurality of data queuing structures may be used where one or more microflows and/or macro flows are processed in each queuing structure.

As described in detail below, hierarchical bandwidth sharing in such arrangements may be achieved using metering of data traffic flows in conjunction with marking (e.g., color marking) of packets (or any other appropriate data segment, such as a frame (hereafter collectively referred to as "packets")) and, in certain embodiments, preferential packet dropping.

In example embodiments, the packets from each microflow may include a field (such as in a packet header) indicating a particular microflow with which the packet is associated. Such an indication of an associated microflow allows for separate metering and marking of the packets for each individual microflow in order to implement a particular bandwidth sharing policy regardless of whether multiple dataflows (microflows and/or macroflows) are processed in the same data queue structure. Because microflows are combined to produce macroflows, a given packets macroflow may be determined from its microflow designation.

FIG. 2 is a block diagram illustrating an apparatus 200 in which meter-based hierarchical bandwidth sharing may be implemented. The apparatus 200 may be a network device with limited data queuing resources (e.g., fewer data queues than a number of individual dataflows being processed). In the apparatus 200, a first set of ten microflows MicroFlow(0, 0) 202 . . . MicroFlow(0,9) 204 may be combined, using a scheduler 214, to form a first macroflow MacroFlow-0 206. As discussed above, the microflows 202 . . . 204 may be processed in single data queue or in a plurality of queues. Also in the apparatus 200, in like fashion as discussed above for MacroFlow-1 206, a second set of ten microflows MicroFlow

4

(1,0) 216 . . . MicroFlow(1,9) 218 may be combined, using a scheduler 228, to produce a second macroflow MacroFlow-1 220.

Still further in the apparatus 200, the MacroFlow-0 206 and the MacroFlow-1 220 may be combined with one another, using a scheduler 230, to produce a data flow 232. In this instance, the MacroFlow-0 206 and the MacroFlow-1 220 may be considered to be microflows of the data traffic flow 232. For instance, the apparatus 200 may meter and mark packets of the MacroFlow-0 206 and the MacroFlow-1 220 in similar fashion as described below with respect to microflows 202, 204, 216 and 218. Likewise, the apparatus 200 may meter and mark data packets of the data flow 232 in similar fashion as discussed below for the macroflows 206 and 220.

In the apparatus 200, predictable bandwidth allocation for the microflows 202 . . . 204 and 216 . . . 218, as well as the macroflows 206 and 220 may be achieved using meters in conjunction with packet marking. For instance, metering of each microflow may be accomplished using respective token bucket meters (e.g., simple single token bucket meters and/or two-rate, three color token bucket meters) to determine whether each microflow is "in-policy" or "out-of-policy" with respect to a bandwidth allocation and sharing policy.

In an example embodiment, such as shown in FIG. 2, microflows 202 . . . 204 and 216 . . . 218 are metered, respectively, by single token bucket meters 208 . . . 210 and 222 . . . 224. In the apparatus 200, the meters 208, 210, 222 and 224 are used to ensure that bandwidth allocation of at least 50 megabits per second (Mbps) is provided for each of the corresponding microflows. It will be appreciated that this allocation is given merely by way of example and any number of other bandwidth allocation arrangements are possible. For instance, different bandwidths may be allocated to each of the microflows. Also in the apparatus 200, the macroflows 206 and 220 may be metered using simple single token bucket meters to determine whether a maximum bandwidth allocation for each macroflow is being exceeded. As described in detail below, such an arrangement provides for allocating a dedicated amount of bandwidth to each microflow and also provides for use (sharing) of unused macroflow bandwidth by microflows that exceed their dedicated bandwidth allocation.

In allocating data communication bandwidth to the microflows, the sum of the bandwidth allocations for a set of microflows (e.g., microflows 202 . . . 204) should be less than or equal to the bandwidth of the macroflow of which the particular set of microflows are a part of. For example, in FIG. 2, each of the ten microflows (202 . . . 204) that are part of the MacroFlow-0 206 may have a bandwidth allocation of 50 Mbps, while the MacroFlow-0 206 may have a bandwidth of 500 Mbps (i.e., ten times 50 Mbps). Therefore, the sum of the bandwidth allocations for the microflows 202 . . . 204 of MacroFlow-0 206 is equal to the bandwidth of the MacroFlow-0 206.

Other allocations are possible, of course. For instance, the MacroFlow-0 206 may include three microflows rather than ten microflows. In such a situation, the three microflows may have bandwidth allocations that, in total, are less than or equal to the bandwidth of MacroFlow-0 206. For instance, in this example, one of the microflows may have a bandwidth allocation of 250 Mbps, while the other two microflows may have bandwidth allocations of 125 Mbps each. Again, the sum of the microflow bandwidths would be equal to the bandwidth of the MacroFlow-0, 500 Mbps. Numerous other bandwidth allocations are possible for the respective microflows of the MacroFlow-0 206 and the MacroFlow-1 220, and the above arrangements are provided by way of example only.

In the apparatus **200**, the single token bucket meters **208** . . . **210** and **222** . . . **224** may be used to provide “minimum” bandwidth allocation for their corresponding microflows. For instance, if packets for MicroFlow(0,0) **202** (or any of the other microflows) arrive at a rate that is at or below an allocated data rate for the corresponding microflow (e.g., 50 Mbps in this example), the single token bucket meter **208** would indicate that each of the packets are in profile and the packets may be appropriately marked using a packet marker, such as described below. If, however, packets arrive at a rate that is above the allocated data bandwidth for the MicroFlow(0,0), e.g., above 50 Mbps, as least some of the packets will be identified, based on the state of the meter **208** when a given packet is received at the apparatus **200**, as being out-of-profile based on the allocated bandwidth. Such out-of-profile packets may be marked accordingly. In this situation, not every packet would be considered to out-of-profile, but only that portion of packets that corresponds with an amount of data traffic in the microflow that is above the allocated bandwidth for the microflow. That is, packets corresponding to the microflow **202**’s 50 Mbps bandwidth allocation would still be marked as in-profile.

For the single token bucket meters **208** . . . **210** and **222** . . . **224**, tokens (or credits) may be periodically added to respective token counts for each of the buckets. The tokens may be added at a rate that corresponds with the allocated bandwidth for the particular microflow, such as 50 Mbps in this example. This rate may be referred to as the Committed Information Rate (CIR) for the microflow. In an example embodiment, the total number of tokens that may be included in a given token count may be limited. This limit may be referred to as “token bucket depth.” The number of tokens corresponding with a token bucket depth may also be referred to as a Committed Burst Size (CBS), which represents the instantaneous bandwidth that a particular microflow may consume in such a data communication apparatus.

The single token bucket meters **208** . . . **210** and **222** . . . **224** may determine whether packets of their corresponding microflows are in profile or out-of-profile based on their token counts. For example, when a packet arrives at the apparatus **200** that is identified (e.g., in its header) as being part of the MicroFlow(0,0) **202**, the single token bucket meter **208** may be examined to determine if there is a positive token count. If the meter **208** has a positive token count, the received packet may be marked as being in-profile and a number of tokens corresponding to the size of the packet may be subtracted from the token count for the meter **208**.

Conversely, if a packet arrives at the apparatus **200** that is associated with the microflow **202** and the associated token bucket meter **208** has a zero token count, or a negative token count, the received packet may be marked as being out-of-profile. For this particular embodiment, the token count of the meter **208** would not be modified to produce a negative, or further negative token count if the packet is marked as being out-of-profile. However, in other instances, a token count for a meter may be modified (producing a negative, or further negative token count) when a packet is marked as being out-of-profile, such as is some of the example embodiments described below.

In the example apparatus **200**, while a limit for dedicated bandwidth allocation for the microflows, e.g., using the meters **208** . . . **210** and **222** . . . **224** is imposed, a corresponding limit for a maximum bandwidth is not imposed for the individual microflows. In such an arrangement, an upper bandwidth for the microflows may then be limited by the bandwidth allocation of the associated macroflow and the bandwidth usage of related microflows (e.g., microflows of

the same macroflow). For instance, for the microflows **202** . . . **204**, the associated MacroFlow-0 **206** has maximum bandwidth allocation of 500 Mbps, which is monitored in similar fashion as the microflows using a single token bucket meter **212**. The meter **212** may be used to determine whether the MacroFlow-0 is in-profile or out-of-profile with respect to its 500 Mbps bandwidth allocation. Such an arrangement allows for bandwidth sharing of unused bandwidth. For instance, if only a single microflow, MicroFlow(0,0) **202** is operating at 400 Mbps in the apparatus **200**, the apparatus **200** may allow all of the packets in the microflow **202** to be communicated to their destination because the associated MacroFlow-0 **206** would remain in profile (e.g., below its 500 Mbps bandwidth allocation).

In such a situation, some packets of the microflow **202** would be marked as in-profile (e.g., packets corresponding with the 50 Mbps bandwidth allocation), while the remaining packets in the microflow **202** would be marked as being out-of-profile (e.g., packets corresponding with the 350 Mbps of bandwidth usage above the 50 Mbps bandwidth allocation). In this instance, it is advantageous to allow the microflow **202** to use the excess bandwidth of the macroflow **206** that is not being used and would be otherwise wasted.

In an example embodiment, such as in the apparatus **200** for the above scenario, the packets in the microflow **202** that were determined (and marked) as being out-of-profile by the meter **208** may be upgraded to being “in-profile” based on the macroflow meter **212**. In this example, when a packet arrives at the meter **212**, the meter **212** may be examined to determine if a positive token count is present. Because the microflow **202** is the only microflow communicating data (at 400 Mbps), a positive token count would typically exist in the meter **212**. Therefore, the packet’s marking may be changed from being marked as out-of-profile to being in-profile based on the state of the macroflow meter **212**. Accordingly, for the apparatus **200**, packets that are marked as out-of-profile by a microflow meter (e.g., **208**, **210**, **222** and **224**) may be upgraded to being in-profile by the corresponding macroflow meters **212** and **226** as long as the respective macroflows **206** and/or **220** remain in-profile, e.g., below their bandwidth allocation.

In the above example, if the remaining microflows that constitute the macroflow **206** began communicating packets at a rate of 50 Mbps (their allocated bandwidths) while the microflow **202** continued to communicate packets at a rate of 400 Mbps, the macroflow meter **212** would go out-of-profile and, in this example, discontinue upgrading the packets from the microflow **202** that exceed its 50 Mbps bandwidth allocation. Using such an arrangement, excess bandwidth may be advantageously used by microflows, even though the microflows using the excess bandwidth may be operating above their individual bandwidth allocation. Further, such an arrangement provides a way to ensure that each microflow has uncontested access to its allocated bandwidth.

In the above situation, where a group of microflows begin operating in a substantially simultaneous fashion, the available data queuing resources in an apparatus such as the apparatus **200** should be sufficient to process the packets in such an instance. For example, the amount of data queuing resources available in the apparatus **200** for handling such a situation may be a product of the bucket depths for each of token bucket meters for the associated microflows (i.e., the CBSs of the microflow token bucket meters).

FIG. 3 is a table **300** illustrating an example embodiment for packet marking that may be used in conjunction with the apparatus **200** of FIG. 2 to indicate in-profile and out-of-profile packets and upgrade packets to facilitate unused bandwidth sharing. The example in the table **300** may be applied to

individual packets to determine whether a packet is in-profile, or out-of-profile and, if the packet is marked as out-of-profile by a microflow meter, whether the packet should be upgraded.

The packet marking approach illustrated in FIG. 3 will be discussed with respect to an individual packet. In the table 300, column 310 indicates the state of a microflow meter when the packet is received. If the microflow meter indicates the microflow is in-profile (e.g., has a positive token count) when the packet is received, the packet is marked as "green" (G). However, if the microflow meter indicates that the microflow is out-of-profile when the packet is received, the packet is marked as "red" (R). The markings in column 310 may be referred to as the microflow "local color," as those markings indicate whether the microflow (e.g., based on its meter) is in or out-of-profile. Therefore the packet markings in column 310 indicate the local state of an associated microflow meter.

Column 320 in FIG. 3 indicates the state of the macroflow meter when the packet is received and may be referred to as the macroflow local color. As with the microflow local color, if the macroflow meter indicates the macroflow is in profile when the packet is received, the packet is marked (locally) as green "G." If the macroflow meter indicates that the macroflow is out-of-profile when the packet is received, the Packet is marked locally as red "R."

Column 330 in FIG. 3 indicates what the final color of a packet would be in each instance illustrated in the table 300. In this example, the final color of a packet may depend on its microflow local color and its macroflow local color, as described in further detail below. Columns 340 and 350 indicate, respectively, whether the microflow meter and the macroflow meter are updated (token counts reduced) for a given packet in the various situations illustrated in FIG. 3.

Row 360 of the table 300 illustrates a situation where a packet is marked as "G" by both a microflow meter and a macroflow meter, indicating that the microflow and the macroflow are both in profile. The final color of the packet is marked as "G." In this situation, both the microflow meter and the macroflow meter are updated, e.g., by reducing their token counts by an amount corresponding with the size of the packet.

Row 370 of the table 300 illustrates a situation where a packet is marked as "G" by a microflow meter and "R" by a macroflow meter. The final color of the packet is marked as "G" even though the macroflow meter indicates the macroflow is out-of-profile. Such an outcome may ensure the microflow's guaranteed bandwidth allocation. Because the packet was locally marked as "G" by the microflow meter, that indicates that the microflow is in profile and the packet should be forwarded on in the macroflow, not discarded. In this situation, both the microflow meter and the macroflow meter are updated. Because the microflow meter is in profile, the token count would be positive when the packet arrives. However, the macroflow meter is out-of-profile when the packet arrives, as the packet is marked "R" locally by the macroflow meter. Therefore, updating the macroflow meter will cause it to go negative or further negative in this instance. Such an approach may be advantageous in the situation described above where a single microflow is using excess bandwidth when multiple other microflows begin transmitting data at their allocated rates. By allowing the macroflow meter's token count to go negative, this will prevent the macroflow meter from upgrading any packets until the macroflow meter's token count becomes positive again. In this situation, each operating microflow would be allowed to transmit at data rates up to their respective allocated bandwidths, in accordance with the marking arrangement illustrated in row 370 of the table 300. Depending on the particular embodi-

ment, the extent to which the macroflow meter's token count can go negative may be bounded to prevent the macroflow meter's token count from going further negative indefinitely.

Row 380 of the table 300 illustrates the situation where a packet receives an upgrade. As indicated in column 310, the packet is locally marked "R" by the microflow meter. As indicated in column 320, the packet is then locally marked "G" by the macroflow meter. This indicates that the individual microflow is out-of-profile (e.g., the microflow meter has a zero or negative token count), but that the macroflow is in profile (e.g., the macroflow meter has a positive count). In this example, the packet is then "upgraded" and marked with a final color of "G," indicating that the packet should be forwarded on in the macroflow and not discarded. Techniques for discarding packets based on their final color are discussed in further detail below.

In the situation illustrated in row 380 of the table 300, the microflow meter is not updated and the macroflow meter is updated. Because the packet is upgraded to allow the opportunistic use of excess bandwidth by the out-of-profile microflow, updating the out-of-profile microflow meter could unnecessarily penalize that microflow for utilizing the excess bandwidth. For instance, allowing the microflow meter token count to go negative, or further negative may prevent the microflow from accessing its allocated (e.g., guaranteed) bandwidth once the other microflows begin to transmit until the microflow meter's token count is restored to a positive value by the periodic adding of tokens to its token count.

Row 390 of the table 300 illustrates the situation where both the microflow and the macroflow are out-of-profile. In this situation, the packet would be marked locally "R" for both the microflow and macroflow, as indicated in columns 310 and 320. As also shown in column 330, this would result in a final color of "R" for the packet. In this situation, neither the microflow meter nor the macroflow meter would be updated and the packet, typically would be dropped as being out-of-profile and not upgraded due to the unavailability of excess bandwidth in the macroflow.

FIG. 4 is a flowchart illustrating an example embodiment of a method 400 for data communication that may be implemented in the apparatus 200 of FIG. 2 using the packet marking illustrated in FIG. 3. The method 400 may, of course, be implemented in any number of other data communication apparatus, such as the apparatus 500 illustrated in FIG. 5, for example. For the below discussion, the method 400 will be described with reference to FIG. 4 and additional reference to FIGS. 2 and 3.

At block 405 of the method 400, a data packet is received as part of a first data traffic flow. The packet may be, for example, part of a first microflow, such as the microflow 202. At block 410, it is determined whether a first rate of traffic of the first data traffic flow is less than or equal to a first threshold. For instance, the token bucket meter 208 may be examined. If a positive token count is present in the meter 208, the first rate of traffic would be determined to be less than or equal to the first threshold (e.g., an allocated or "guaranteed" data rate). If the meter 208 has a zero or negative token count, the first rate of traffic Would be determined to be greater than the first threshold.

At block 415, in the event the first-rate of traffic is determined to be less than or equal to the first threshold, the packet may be marked with a first marker type, e.g., "G" as the local microflow color, as discussed above. At block 420, in the event the first rate of traffic is greater than the first threshold, the packet may be marked with a second marker type, e.g., "R" as the local microflow color, as described above with respect to FIG. 3.

The method **400** further includes, at block **425**, receiving a second data traffic flow having a second rate of traffic, such as a second microflow **204**. At block **430**, the first data traffic flow (microflow **202**) may be combined with the second data traffic flow (microflow **204**) to produce a third data traffic flow. For instance, in the apparatus of FIG. 2, the microflows **202** and **204** may be combined using the scheduler **214** to produce the MacroFlow-**0 206**.

At block **435**, in the event the data packet was marked with the first marker type (e.g., locally "G" by the microflow meter **208**) the packet may be forwarded as part of the third data flow regardless of a rate of traffic for the third data traffic flow (macroflow **206**). This situation is represented by rows **360** and **370** of the table **300** illustrated in FIG. 3. In this situation, the local macroflow color could be determined, as described below, and meter updates performed in accordance with the table **300** illustrated in FIG. 3.

At block **440**, it may be determined whether a third rate of traffic of the third data traffic flow is less than or equal to a second threshold (e.g., the upper bandwidth limit for the macroflow **206**, in this case 500 Mbps). In the apparatus **200**, this determination could be made based on the state of the macroflow meter **212**. If the meter **212** has a positive token count when the packet is received, that would indicate that the third rate of traffic is less than or equal to the second threshold (e.g., the macroflow **206** is in-profile). If the meter **212** has a zero or negative token count, that would indicate that the third rate of traffic is greater than the second threshold (e.g., the macroflow **206** is out-of-profile).

At block **445**, in the event the data packet is marked with the second marker type (e.g., locally "R" for the microflow **202**) and the third rate of traffic is less than or equal to the second threshold (e.g., locally "G" for the macroflow **206**), the marker type for the packet may be changed from the second marker type (local microflow color of "R") to the first marker type (final color of "G"). This illustrates the situation in row **380** of the table **300**, where a packet marked "R" from an out-of-profile microflow is upgraded to "G" in order to opportunistically take advantage of unused bandwidth, such as is in the apparatus **200**, as discussed above. Further at block **445**, the upgraded packet is forwarded as part of the third data flow (e.g., macroflow **206**).

At block **450**, in the event the data packet is marked with the second marker type (e.g., locally "R" for the microflow **202**) and the third rate of traffic is greater than the second threshold (e.g., locally "R" for the macroflow **206**), the packet may be discarded. Various approaches exist for discarding the packet. For instance, the packet may be immediately discarded when the determination is made to mark the packet with a final color "R." Alternatively, for example, the packet may be forwarded to a data queuing structure with admission control and be discarded by the queuing structure. Other alternatives also exist. For instance, if congestion is not present at the data queuing structure, packets marked "R" as their final color may still be admitted to the data queuing structure. For instance, if the data queue occupancy is below a "red" threshold, packets with a final color marking of "R" may be admitted. If the queue occupancy is above the red threshold, the packets would be discarded in this example. A functionally similar threshold could be used for green packets, where the green threshold is higher than the red threshold. As an example, a red threshold may be set at twenty-five percent queue occupancy, while a green threshold may be set at ninety percent queue occupancy, as one example. An embodiment of such a data queuing structure with admission control is described in further detail below with respect to FIG. 9

FIG. 5 is a block diagram that illustrates another apparatus **500** for data communication that may be used to implement meter-based hierarchical bandwidth sharing. The apparatus **500** is similar to the apparatus **200** in a number of respects. For instance, a set of microflows **502 . . . 504** is combined, using scheduler **518** to form a first macroflow **506**. Also, a second set of microflows **520 . . . 522** are combined, using a scheduler **536** to form a second macroflow **524**. The macroflows **506** and **524** are combined, using a scheduler **538** to form a data flow **540**. The macroflow **506** is metered in similar fashion as the macroflow **206** using a single token bucket meter **516**. Likewise, the macroflow **524** is metered in similar fashion as the macroflow **220** using a single token bucket meter **534**.

In the apparatus **500**, the microflows are metered using two token buckets, or using dual-token bucket meters. For instance, the microflow **502** is metered using a first token bucket **508**, which is used to ensure that the microflow **502** has access to its allocated ("minimum") bandwidth. The microflow **502** is also metered by a second token bucket **510**, which is used to ensure that the microflow **502** does not exceed an upper ("maximum") bandwidth limit. The microflows **504**, **520** and **522** are similarly metered, respectively, by the token bucket pairs of **512/514**, **526/528** and **530/532**. For purposes of illustration, the metering of microflow **502** will be discussed below.

As with the token bucket meter **208** for the microflow **202** in the apparatus **200**, the token bucket **508** is used to ensure that the microflow **502** has access to its allocated "guaranteed" bandwidth. In an example embodiment, tokens may be periodically added to a token count of the meter **508** at a rate proportional to the allocated bandwidth and up to a bucket depth corresponding with a CBS for the microflow **502**. As discussed above with respect to the meter **208**, the allocated bandwidth metered by the token bucket **508** may be referred to as the CIR of the microflow **502**.

In the apparatus **500**, the token bucket **510** meters the use of excess bandwidth by the microflow **502** above its CIR and up to an upper limit, in this case 100 Mbps. The difference between the CIR of the microflow **502** and the upper bandwidth limit may be referred to as the excess information rate (EIR), which in this case would be 50 Mbps (i.e., 100 Mbps–50 Mbps). In an example embodiment, tokens may be periodically added to a token count of the meter **510** at a rate proportional to the EIR and up to a bucket depth corresponding with an excess bucket size of the token bucket **510**. Because two rates (e.g., the CIR and the EIR) are being monitored for the microflows in the apparatus **500**, packet marking may be accomplished using a three color scheme for microflows, as is described below.

FIG. 6 is a table **600** illustrating an example embodiment for packet marking that may be employed with the apparatus **500** illustrated in FIG. 5. The packet marking illustrated in FIG. 6 accounts for metering of both the CIRs and EIRs for the microflows of the apparatus **500**. In the table **600**, column **605** indicates the state of a microflow (e.g., dual-bucket) meter when a packet is received. If the token bucket **508** indicates that the microflow is operating within its CIR (e.g., the token bucket **508** has a positive token count) when the packet is received, the packet is marked as green ("G"). However, if the token bucket **508** indicates that the microflow is operating above its CIR when the packet is received (e.g., the token count of the bucket **508** is zero or negative), the token bucket **510** may be examined to determine if the microflow **502** is operating within its EIR (e.g., the token bucket **510** has a positive token count). If the microflow is operating within its EIR, the packet is marked as yellow ("Y"). Further, if the

11

token buckets **508** and **510** indicate that the microflow **502** is operating above the EIR (e.g., the token counts of the buckets **508** and **510** are either zero or negative), the packet is marked as red ("R").

As with column **310** of the table **300**, the color markings in column **605** may be referred to as the microflow "local color," as those markings indicate whether the microflow (e.g., based on its dual token bucket) is operating within its CIR, operating within its EIR, or is out-of-profile. Therefore the packet markings in column **605** indicate the local state of an associated microflow dual token bucket meter.

Column **610** in FIG. **6** indicates the state of the macroflow meter when the packet is received and may be referred to as the macroflow local color. As discussed above with respect to FIG. **3**, if the macroflow meter indicates the macroflow is in profile when the packet is received, the packet is marked (locally) as green "G." If the macroflow meter indicates that the macroflow is out-of-profile when the packet is received, the packet is marked locally as red "R."

Column **615** in FIG. **6** indicates what the final color of a packet would be in each instance illustrated in the table **600**. In this example, the final color of a packet may depend on its microflow local color and its macroflow local color, such as described in further detail below. Columns **620** and **625** indicate, respectively, whether token counts of the microflow CIR bucket (e.g., token bucket **508**), the EIR bucket (e.g., the token bucket **510**) and/or the macroflow meter (e.g., single token bucket meter **516**) are updated (e.g., token counts reduced) for a given packet in the various situations illustrated in FIG. **6**.

Row **630** of table **600** illustrates the situation where a microflow is operating within its CIR and an associated macroflow is operating in-profile when a given packet of the microflow is received. Accordingly, the packet would be marked locally as "G" for the microflow and locally as "G" for the associated macroflow. In this instance, as shown in column **615**, the final color of the packet would be "G" as well. Typically, the packet would be forwarded onto to its destination. In some embodiments, however, the packet could still be dropped due to congestion at, for example, an egress data queuing structure. In this situation, the CIR bucket (e.g., bucket **508**) and the macroflow bucket (e.g., **516**) would be updated, or have their token counts reduced by an amount corresponding with the size of the packet.

Row **635** of table **600** illustrates the situation where a microflow is operating within its CIR and an associated macroflow is operating out-of-profile when a given packet of the microflow is received. Accordingly, the packet would be marked locally as "G" for the microflow and locally as "R" for the associated macroflow. In this instance, as shown in column **615**, the final color of the packet would be "G" as the microflow is operating within its CIR (e.g., at or below its guaranteed bandwidth). As shown in columns **620** and **625**, both the committed bucket and the macroflow meter would be updated. In this instance, the token count of the macroflow meter would go negative, or further negative. As discussed above, this outcome is desirable in certain embodiments as it may ensure that upgrades are not available for a period of time when additional microflows begin communicating in situations where one or more other microflows have been operating above their CIR and using excess bandwidth. As previously discussed, using such an approach would allow each microflow to have access to its CIR. Packet upgrades would not be available again until the macroflow meter achieved a positive token count.

Row **640** illustrates the situation where a microflow is operating above its CIR but within its EIR and an associated

12

macroflow is operating in-profile when a given packet of the microflow is received. Here the packet would be marked locally as "Y" for the microflow and locally as "G" for the macroflow. In this instance, the final color of the packet would be marked as "G", or the packet would be upgraded to allow the microflow to utilize the excess bandwidth of the macroflow. As shown in columns **620** and **625**, the token counts for an excess bucket (e.g., bucket **510**) and the macroflow bucket (e.g., token bucket meter **516**) would be updated by reducing their token counts by an amount corresponding with the size of the packet.

Row **645** illustrates the situation where a microflow is operating above its CIR but within its EIR and an associated macroflow is operating out-of-profile when a given packet of the microflow is received. Here the packet would be marked locally as "Y" for the microflow and locally as "R" for the macroflow. In this instance, upgrades would not be available as the macroflow is operating out-of-profile. Accordingly, the packet is marked with a final color of "R" and none of the token buckets for the microflow or macroflow are updated. Typically the packet in this situation is discarded, though in some embodiments the packet may be forwarded to its destination if congestion does not exist downstream, such as described above and in further detail below.

Rows **650** and **655** illustrate situations where a microflow is operating above both its EIR and CIR. In the situation of row **650**, an associated macroflow is operating in-profile, while in the situation of row **655**, the macroflow is operating out-of-profile. In both of these situations, the final color of the packet would be marked as "R" because the microflow is operating above its EIR, which represents an upper bandwidth limit for the macroflow. Therefore, even if the associated macroflow is operating in-profile, upgrades would not be given to packets that are marked locally as "R" for microflows in this embodiment. In these situations, none of the token buckets for the microflow or macroflow would be updated.

FIG. **7** illustrates an example embodiment of a method **700** of meter-based hierarchical bandwidth sharing. The method **700** may be implemented in the apparatus **500** of FIG. **5** using the packet marking illustrated in FIG. **6**. Of course, the method **700** could be implemented in any number of data communication apparatus and is not limited to the approaches illustrated in FIGS. **5** and **6**. However, for purposes of illustration, the method **700** will be described with additional reference to FIGS. **5** and **6**.

At block **705** of the method **700**, a data packet is received as part of a first data traffic flow. The packet may be, for example, part of a first microflow, such as the microflow **502**. At block **710**, it is determined whether a first rate of traffic of the first data traffic flow is less than or equal to a first threshold. For instance, the token bucket **508** may be examined to determine if the microflow **502** is operating at or below its CIR. If a positive token count is present in the meter **508**, the first rate of traffic would be determined to be less than or equal to the first threshold (e.g., the CIR or "guaranteed" data rate). If the meter **508** has a zero or negative token count, the first rate of traffic would be determined to be greater than the first threshold, indicating the microflow is operating above its CIR. At block **715**, in the event the first rate of traffic is determined to be less than or equal to the first threshold (the CIR), the packet may be marked with a first marker type (e.g., "G") as the local microflow color, such as discussed above.

At block **720**, in the event the first rate of traffic is greater than the first threshold (e.g., the microflow **502** is operating above its CIR), a determination may then be made as to whether the first rate of traffic is greater than a second threshold (e.g., an EIR for the microflow), where the second thresh-

13

old (EIR) is greater than the first threshold (CIR). In the event the first rate of traffic is less than or equal to the second threshold, the data packet may be marked with a second marker type (e.g., "Y") as the local microflow color, such as discussed in the above example with respect to FIG. 6. However, in the event the first of rate traffic is greater than the second threshold (EIR), the data packet may then be marked with a third marker type ("R") as the local color for microflow, as also discussed above with respect to FIG. 6.

At block 725 of the method 700, a second data traffic flow having a second rate of traffic may be received. At block 730, the first data traffic flow may be combined with the second data traffic flow to produce a third data traffic flow. Of course, the first and second data traffic flows may simply be combined with each other, or may be combined with additional data traffic flows to form the third data traffic flow.

At block 735, in the event the data packet is marked with the first marker type (locally "G" for the microflow), the data packet may be forwarded in the third data flow. This may be done regardless of the state of the third traffic flow (e.g., the macroflow) because the first data traffic flow (e.g., microflow) is operating below the first threshold (e.g., within its CIR). These situations are illustrated by rows 630 and 635 of the table 600 shown in FIG. 6.

At block 740, it may be determined whether a third rate of traffic of the third data traffic flow is less than or equal to a third threshold (e.g., a macroflow bandwidth limit). At block 745, in the event the data packet is marked with the second marker type (e.g., locally as "Y" for the microflow) and the third rate of traffic is less than or equal to the third threshold (e.g., the macroflow is in profile), the packet's marker may be changed from the second marker type to the first marker type (e.g., the packet may be upgraded and given a final color of "G," such as illustrated by row 640 of the table 600). In this instance, the packet may be forwarded in the third data flow, such as in the fashions described herein.

At block 750, in the event the data packet is marked with the second marker type (e.g., marked locally "Y" for the microflow) and the third rate of traffic is greater than the third threshold (e.g., the macroflow is out-of-profile), the marker of the packet may be changed from the second marker type to the third marker type (e.g., marked with a final color of "R" as no upgrades are available due the out-of-profile state of the macroflow). In this situation, the packet may be discarded from the third data flow. Alternatively, the packet may be sent to a data queuing structure with admission control as described above and in further detail below.

At block 755, if the packet is marked with the third marker type (e.g., locally as "R" for the microflow), the packet may be discarded regardless of the rate of traffic of the third data flow (e.g., the macroflow). Such examples were discussed above with regard to rows 650 and 655 of the table 600 illustrated in FIG. 6.

FIG. 8 is block diagram of a two-rate three-color meter (trTCM) 800 that may be used for metering microflows in the apparatus 500 illustrated in FIG. 5. The trTCM 800 includes dual-token buckets 810, packet marking 820 and meter updating 830. The dual-token bucket 810 includes a CIR bucket 812 and an EIR bucket 814. As was discussed above, tokens 815 are added to the CIR bucket 812 at rate that is proportional with a CIR 816 for an associated microflow. Likewise, tokens 815 are added to the EIR bucket 814 at a rate that is proportional with an EIR 817 for the associated microflow. As was also discussed above, the CIR bucket 812 is limited in its token count by the CBS 818 (e.g., the CIR bucket 812's

14

depth). Likewise, the EIR bucket 814 is limited in its token count by an excess burst size (EBS) 810 (e.g., the EIR bucket 814's depth).

The packet marking block 820 may mark packets in accordance with the embodiments illustrated and described above with respect to FIGS. 5-7. Also, the meter update block 830 may update token counts of the CIR bucket 812 and the EIR bucket 814 in accordance with the embodiments illustrated and described above with respect to FIGS. 5-7.

FIG. 9 illustrates a data queuing structure 900 that includes packet admission control 920. In the queuing structure 900, packets that are admitted by the admission control 920 may be queued in a data queue 930 for transmission to their respective destinations.

The data queuing structure 900 may receive a data flow 910 that includes packets that have been marked in accordance with the packet marking embodiments illustrated in FIGS. 3 and 6, or using any other packet marking approach. The packet admission control 920 may admit or discard packets based only on their final color. In such an approach, any packet marked with a final color "R" would be discarded, while any packet marked with a final color "G" would be admitted to the data queuing structure 900 and placed in data queue 930 to be transmitted to its final destination.

Alternatively, packets may be admitted to the data queuing structure 900 by the packet admission control 920 based on their final color and on data occupancy of the data queue 930. For instance if a packet with a final color of "R" arrives at the packet admission control 920, the packet admission control 920 may determine the amount of data presently in the data queue via line 935. If the occupancy of the data queue 930 is below a red threshold (indicating there is very little or no data congestion) the packet admission control 920 may admit the packet and place it in the data queue 930 for delivery. Conversely, if the data occupancy is above the red threshold 940, the packet may be discarded. Green packets may be similarly admitted and discarded based on a green threshold 950 for queue occupancy, where the green threshold 950 is higher than the red threshold 940. The admission control 920 may operate without the use of the thresholds, using both thresholds or using only a single threshold. For instance, only the red threshold 940 may be used, while all packets marked with a final color "G" are admitted to the queue 930.

FIG. 10 is a flowchart illustrating an example embodiment of a method 1000 for meter-based hierarchical bandwidth sharing including preferential dropping of packets using the data queuing structure 900 illustrated in FIG. 9. It will be appreciated that a preferential packet dropper that is not part of a data queuing structure may alternatively, perform the packet admission and discard functions of admission control 920. As another alternative, the packet dropping functions may be carried out as part of the packet marking process. Still other alternatives may exist.

At block 1005 of the method 1000, a data packet may be received that is included in a first data traffic flow. At block 1010, it may be determined if a first rate of traffic of the first data traffic flow is less than or equal to a first threshold (e.g., a microflow's CIR). At block 1015, in the event the first rate of traffic is less than or equal to the first threshold, the data packet may be marked with a first marker type (e.g., locally as "G" for the microflow). At block 1020, in the event the first rate of traffic is greater than the first threshold, the data packet may be marked with a second marker type (e.g., locally "R" for the microflow).

At block 1025, a second data traffic flow having a second rate of traffic may be received. At block 1030, the first and second data traffic flows may be combined to produce a third

15

data traffic flow (e.g., a macroflow). At block **1035**, it may be determined whether a third rate of traffic of the third data traffic flow is less than or equal to a second threshold (e.g., the macroflow's bandwidth limit). At block **1040**, in the event the data packet is marked with the second marker type ("R") and the third rate of traffic is less than or equal to the second threshold, the second marker type may be changed to the first marker type (e.g., the packet may be upgraded, such as illustrated in row **380** of the table **300** in FIG. **3**).

At block **1045**, the third data traffic flow is provided to a data queue having a first admission threshold (e.g., red threshold **940**) and a second admission threshold (e.g., green threshold **950**). At block **1050**, in the event the packet is marked with the second marker type ("R") and an amount of data in the data queue is greater than the first admission threshold (red threshold), the packet may be discarded. At block **1055**, in the event the packet is marked with the second marker type ("R") and the amount of data in the data queue is less than or equal to the first admission threshold (red threshold), the packet may be forwarded to a destination of the packet.

At block **1060**, in the event the packet is marked with the first marker type ("G") and the amount of data in the data queue is greater than the second admission threshold (green threshold), the packet may be discarded. At block **1065**, in the event the packet is marked with the first marker type ("G") and the amount of data in the data queues is less than or equal to the second admission threshold (green threshold), the packet may be forwarded to its destination.

FIG. **11** is a block diagram that illustrates another apparatus **1100** for data communication that may be used to implement meter-based hierarchical bandwidth sharing. The apparatus **1100** is similar to the apparatus **500** in a number of respects. For instance, a set of microflows **1102** . . . **1104** is combined, using scheduler **1118** to form a first macroflow **1106**. Also, a second set of microflows **1120** . . . **1122** are combined, using a scheduler **1136** to form a second macroflow **1124**. The macroflows **1106** and **1124** are combined, using a scheduler **1138** to form a data flow **1140**. The data flow **1140** is a macroflow that includes the macroflows **1106** and **1124**.

In the apparatus **1100**, the microflows are metered using two token buckets, or using dual-token bucket meters. For instance, the microflow **1102** is metered using a first token bucket **1108**, which is used to ensure that the microflow **1102** has access to its allocated ("minimum") bandwidth. The microflow **1102** is also metered by a second token bucket **1110**, which is used to ensure that the microflow **1102** does not exceed an upper ("maximum") bandwidth limit. The microflows **1104**, **1120** and **1122** are similarly metered, respectively, by the token bucket pairs of **1112/1114**, **1126/1128** and **1130/1132**. For purposes of illustration, the metering of microflow **1102** will be discussed below.

As with the token bucket meter **508** for the microflow **502** in the apparatus **500**, the token bucket **1108** is used to ensure that the microflow **1102** has access to its allocated "guaranteed" bandwidth. In an example embodiment, tokens may be periodically added to a token count of the meter **1108** at a rate proportional to the allocated bandwidth and up to a bucket depth corresponding with a CBS for the microflow **1102**. As discussed above with respect to the meter **508**, the allocated bandwidth metered by the token bucket **1108** may be referred to as the CIR of the microflow **1102**.

In the apparatus **1100**, the token bucket **1110** meters the use of excess bandwidth by the microflow **1102** above its CIR and up to an upper limit, in this case 100 Mbps. The difference between the CIR of the microflow **1102** and the upper bandwidth limit may be referred to as the excess information rate

16

(EIR). In an example embodiment, tokens may be periodically added to a token count of the meter **1110** at a rate proportional to the EIR and up to a bucket depth corresponding with an excess bucket size of the token bucket **1110**. Because two rates (e.g., the CIR and the EIR) are being monitored for the microflows in the apparatus **1100**, packet marking may be accomplished using a three color scheme for microflows, as is described below.

In the apparatus **1100**, the macroflow **1106** is metered using two token buckets, or using dual-token bucket meters **1116** and **1117**. Likewise, the macroflow **1124** is metered in similar fashion as the macroflow **1106** using two token buckets, or using dual-token bucket meters **1134** and **1135**. For purposes of illustration, the monitoring of the macroflow **1106** using the dual-token bucket meters **1116** and **1117** is discussed below.

The token bucket **1116** is used to ensure that the macroflow **1106** has access to its allocated "guaranteed" bandwidth. In an example embodiment, tokens may be periodically added to a token count of the meter **1116** at a rate proportional to the allocated bandwidth and up to a bucket depth corresponding with a CBS for the macroflow **1106**. The allocated bandwidth metered by the token bucket **1116** may be referred to as the CIR of the macroflow **1106**.

In the apparatus **1100**, the token bucket **1117** meters the use of excess bandwidth by the macroflow **1106** above its CIR and up to an upper limit, in this case 600 Mbps. The difference between the CIR of the macroflow **1106** and the upper bandwidth limit may be referred to as the excess information rate (EIR). In an example embodiment, tokens may be periodically added to a token count of the meter **1117** at a rate proportional to the EIR and up to a bucket depth corresponding with an excess bucket size of the token bucket **1117**. Because two rates (e.g., the CIR and the EIR) are being monitored for the macroflows in the apparatus **1100**, packet marking may be accomplished using a three color scheme for macroflows, as is described below.

Because the token bucket meters **1116** and **1117** monitor the combination of microflows **1102** . . . **1104** in the macroflow **1106**, the token bucket meters **1116** and **1117** may be used to determine whether the microflows **1102** . . . **1104** collectively exceed the minimum (CIR) bandwidth (e.g., 500 Mbps) for the macroflow **1106** and also whether the microflows **1102** . . . **1104** collectively exceed the maximum (EIR) bandwidth (e.g., 600 Mbps) for the macroflow **1106**. Thus, the token bucket meter **1116** may be used to determine that the sum of the data rates of the microflows **1102** . . . **1104** do not exceed the MinBW threshold for the macroflow **1106**. This is indicated in FIG. **11** by the token bucket meter **1116** determining Max(MinBW), such as by summing the data rates of the individual microflows **1102** . . . **1104**. The token bucket meter **1117** may be used to determine that the sum of the data rates of the microflows **1102** . . . **1104** do not exceed the MaxBW threshold for the macroflow **1106**. The MaxBW threshold for the macroflow **1106** (shown as Max(MaxBW)) should be greater than or equal to the largest MaxBW threshold for the microflows **1102** . . . **1104**.

FIG. **12** is a table **1200** illustrating an example embodiment for packet marking that may be employed with the apparatus **1100** illustrated in FIG. **11**. The packet marking illustrated in FIG. **12** accounts for metering of both the CIRs and EIRs for the microflows of the apparatus **1100** and both the CIRs and EIRs for the macroflows **1106** and **1124** of the apparatus **1100**.

In the table **1200**, column **1205** indicates the state of a microflow (e.g., dual-bucket) meter when a packet is received. For example, if the token bucket **1108** indicates that

17

the microflow **1102** is operating within its CIR (e.g., the token bucket **1108** has a positive token count) when the packet is received, the packet is marked as green (“G”). However, if the token bucket **1108** indicates that the microflow **1102** is operating above its CIR when the packet is received (e.g., the token count of the bucket **1108** is zero or negative), the token bucket **1110** may be examined to determine if the microflow **1102** is operating within its EIR (e.g., the token bucket **1110** has a positive token count). If the microflow **1102** is operating within its EIR, the packet is marked as yellow (“Y”). Further, if the token buckets **1108** and **1110** indicate that the microflow **1102** is operating above the EIR (e.g., the token counts of the buckets **1108** and **1110** are either zero or negative), the packet is marked as red (“R”).

As with column **605** of the table **600**, the color markings in column **1205** may be referred to as the microflow “local color,” as those markings indicate whether the microflow (e.g., based on its dual token bucket) is operating within its CIR, operating within its EIR, or is out-of-profile (e.g., operating above its EIR). Therefore the packet markings in column **1205** indicate the local state of an associated microflow dual token bucket meter.

Column **1210** in FIG. **12** indicates the state of the macroflow dual-token bucket meters **1116** and **1117** when the packet is received and may be referred to as the macroflow local color. The local color marking, e.g., for macroflow **1106**, in column **1210** is determined in similar fashion as the local color marking for microflows, as indicated in column **1205**. For instance, if the token bucket **1116** indicates that the macroflow **1106** is operating within its CIR (e.g., the token bucket **1116** has a positive token count) when the packet is received, the packet is locally marked as green (“G”). However, if the token bucket **1116** indicates that the macroflow **1106** is operating above its CIR when the packet is received (e.g., the token count of the bucket **1116** is zero or negative), the token bucket **1117** may be examined to determine if the macroflow **1106** is operating within its EIR (e.g., the token bucket **1117** has a positive token count). If the microflow is operating within its EIR, the packet is locally marked as yellow (“Y”). Further, if the token buckets **1116** and **1117** indicate that the macroflow **1106** is operating above the EIR (e.g., the token counts of the buckets **1116** and **1117** are both zero or negative), the packet is locally marked as red (“R”).

Column **1215** in FIG. **12** indicates what the final color of a packet would be in each instance illustrated in the table **1200**. In this example, the final color of a packet may depend on its microflow local color and its macroflow local color, such as described in further detail below. Columns **1220** and **1225** indicate, respectively, whether token counts of the microflow CIR bucket (e.g., token bucket **1108**), the EIR bucket (e.g., the token bucket **1110**), the macroflow CIR bucket (e.g., token bucket **1116**) and/or the macroflow EIR bucket (e.g., token bucket **1117**) are updated (e.g., token counts reduced) for a given packet in the various situations illustrated in FIG. **12**.

Row **1230** of table **1200** illustrates the situation where a microflow is operating within its CIR and an associated macroflow is also operating within its CIR when a given packet of the microflow is received. Accordingly, the packet would be marked locally as “G” for the microflow and locally as “G” for the associated macroflow. In this instance, as shown in column **1215**, the final color of the packet would be “G” as well. Typically, the packet would then be forwarded onto its destination. In some embodiments, however, the packet could still be dropped due to congestion at, for example, an egress data queuing structure. In this situation, for the microflow **1102**, the microflow CIR (committed) bucket (e.g., bucket **1108**) and the macroflow CIR (committed) bucket (e.g.,

18

bucket **1116**) would be updated, or have their token counts reduced by an amount corresponding with the size of the packet.

Row **1235** of table **1200** illustrates the situation where a microflow is operating within its CIR and an associated macroflow is operating above its CIR but within its EIR when a given packet of the microflow is received. Accordingly, the packet would be marked locally as “G” for the microflow and locally as “Y” for the associated macroflow. In this instance, as shown in column **1215**, the final color of the packet would be “G” as the microflow is operating within its CIR (e.g., at or below its guaranteed bandwidth). As shown in columns **1220** and **1225**, in this situation, both the microflow committed bucket and the macroflow committed bucket that correspond with the particular microflow would be updated. In this instance, the token count of the macroflow committed bucket would go negative, or further negative. As discussed above, this outcome is desirable in certain embodiments as it may ensure that upgrades are not available for a period of time when additional microflows begin communicating in situations where one or more other microflows have been operating above their CIR and using excess bandwidth. As previously discussed, using such an approach would allow each microflow to have access to its CIR. However, packet upgrades would not be available again until the macroflow committed bucket again achieved a positive token count.

Row **1240** of table **1200** illustrates the situation where a microflow is operating within its CIR and an associated macroflow is operating above its CIR and its EIR when a given packet of the microflow is received. Accordingly, the packet would be marked locally as “G” for the microflow and locally as “R” for the associated macroflow. In this instance, as shown in column **1215**, the final color of the packet would be “G” as the microflow is operating within its CIR (e.g., at or below its guaranteed bandwidth). As shown in columns **1220** and **1225**, both the microflow committed bucket and the macroflow committed bucket would be updated. In like fashion as discussed above with respect to row **1235**, the token count of the macroflow committed bucket would go negative, or further negative in this situation.

Row **1245** illustrates the situation where a microflow is operating above its CIR but within its EIR and an associated macroflow is operating within its CIR when a given packet of the microflow is received. In this situation, the packet would be marked locally as “Y” for the microflow and locally as “G” for the macroflow. In this instance, as shown in column **1215**, the final color of the packet would be marked as “G”, or the packet would be upgraded from “Y” to “G” to allow the microflow to utilize the excess bandwidth of the macroflow. As shown in columns **1220** and **1225**, the token counts for a microflow excess bucket (e.g., bucket **1110**) and the macroflow committed bucket (e.g., bucket **1116**) would be updated by reducing their token counts by an amount corresponding with the size of the packet.

Row **1250** illustrates the situation where a microflow is operating above its CIR but within its EIR and an associated macroflow is also operating above its CIR but within its EIR when a given packet of the microflow is received. Here the packet would be marked locally as “Y” for the microflow and locally as “Y” for the macroflow. In this instance, further processing may be done on the packet to determine whether to upgrade the packet to “G” or downgrade the packet to “R.” This decision may depend on a number of factors, such as, for example, an amount of downstream congestion in an associated network device and/or network.

Row **1255** illustrates the situation where a microflow is operating above its CIR but within its EIR and an associated

macroflow is operating above its EIR when a given packet of the microflow is received. Here the packet would be marked locally as "Y" for the microflow and locally as "R" for the macroflow. In this instance, upgrades would not be available as the macroflow is operating above its EIR. Accordingly, the packet is marked with a final color of "R" and none of the token buckets for the microflow or macroflow are updated. Typically the packet in this situation is discarded, though in some embodiments the packet may still be forwarded to its destination if congestion does not exist downstream, such as described above and in further detail below.

Rows 1260, 1265 and 1270 illustrate situations where a microflow is operating above both its CIR and EIR. In the situation of row 1260, an associated macroflow is operating within its CIR. In the situation of row 1265, the macroflow is operating above its CIR but within its EIR. In the situation of row 1270, the macroflow is also operating above its EIR. In each of these situations, the final color of the packet would be marked as "R" because the microflow is operating above its EIR, which represents an upper bandwidth limit for the microflow. Therefore, even if the associated macroflow is operating in-profile (e.g., below its CIR and/or EIR), upgrades would not be given to packets that are marked locally as "R" for microflows in this embodiment. In these situations, none of the token buckets for the microflow or macroflow would be updated.

FIG. 13 illustrates an example embodiment of a method 1300 for meter-based hierarchical bandwidth sharing. The method 1300 may be implemented in the apparatus 1100 of FIG. 11 using the packet marking illustrated in FIG. 12. Of course, the method 1300 could be implemented in any number of data communication apparatus and is not limited to the approaches illustrated in FIGS. 11 and 12. However, for purposes of illustration, the method 1300 will be described with additional reference to FIGS. 11 and 12.

At block 1305 of the method 1300, a data packet is received as part of a first data traffic flow. The packet may be, for example, part of a first microflow, such as the microflow 1102. At block 1310, it is determined whether a first rate of traffic of the first data traffic flow is less than or equal to a first threshold. For instance, the committed token bucket 1108 may be examined to determine if the microflow 1102 is operating at or below its CIR. If a positive token count is present in the committed bucket 1108, the first rate of traffic would be determined to be less than or equal to the first threshold (e.g., the CIR or "guaranteed" data rate). If the meter 1108 has a zero or negative token count, the first rate of traffic would be determined to be greater than the first threshold, indicating the microflow 1102 is operating above its CIR. At block 1315, in the event the first rate of traffic is determined to be less than or equal to the first threshold (the CIR), the packet may be locally and finally marked with a first marker type (e.g., "G") as the local microflow color and final macroflow color, such as discussed above. The final marking may be done by the committed bucket 1108, or alternatively may be done by one of the dual-token buckets 1116 and 1117 associated with the macroflow 1106.

At block 1320, in the event the first rate of traffic is greater than the first threshold (e.g., the microflow 1102 is operating above its CIR), a determination may then be made as to whether the first rate of traffic is greater than a second threshold (e.g., an EIR for the microflow), where the second threshold (EIR) is greater than the first threshold (CIR). In the event the first rate of traffic is less than or equal to the second threshold (but greater than the CIR), the data packet may be marked with a second marker type (e.g., "Y") as the local microflow color, such as discussed in the above example with

respect to FIG. 12. However, in the event the first of rate traffic is greater than the second threshold (EIR), the data packet may then be marked with a third marker type ("R") as the local color for the microflow 1102 and finally marked as red for the macroflow 1106, such as in the fashions discussed above.

At block 1325 of the method 1300, a second data traffic flow having a second rate of traffic may be received. At block 1330, the first data traffic flow may combined with the second data traffic flow to produce a third data traffic flow. Of course, the first and second data traffic flows may simply be combined with each other, or may be combined with additional data traffic flows to form the third data traffic flow.

At block 1335, it may be determined whether a third rate of traffic of the third data traffic flow is less than or equal to a third threshold (e.g., a macroflow minimum bandwidth limit). At block 1340, in the event the data packet is marked with the second marker type (e.g., locally as "Y" for the microflow) and the third rate of traffic is less than or equal to the third threshold (e.g., the macroflow 1106 is operating within its CIR), the packet may be finally marked with the first marker type (e.g., the packet may be upgraded from "Y" and given a final color of "G," such as illustrated by row 1245 of the table 1200).

At block 1345, in the event the data packet is marked with the second marker type (e.g., marked locally "Y" for the microflow 1102) and the third rate of traffic is greater than the third threshold (e.g., the macroflow 1106 is operating above its CIR) a determination may then be made as to whether the third rate of traffic is greater than a fourth threshold (e.g., an EIR for the macroflow 1106), where the fourth threshold (macroflow EIR) is greater than the third threshold (macroflow CIR). In the event the third rate of traffic is less than or equal to the fourth threshold (less than the macroflow EIR but greater than the macroflow CIR), the data packet may be finally marked with the second marker type (e.g., "Y"), such as discussed in the above example with respect to FIG. 12. However, in the event the third rate of traffic is greater than the fourth threshold (macroflow EIR), the data packet may then be finally marked with the third marker type ("R") as the final color for the macroflow 1106, such as in the fashions discussed above.

At block 1350, in the event the packet is finally marked with the first marker type (e.g., "G"), the packet may be forwarded to a destination associated with the packet, such as indicated by a destination address included in the packet. At block 1355, in the event the packet is finally marked with one of the second marker type (e.g., "Y") and the third marker type (e.g., "R"), the method 1300 may include one of (i) processing the packet to determine whether to forward or discard the packet, and forwarding or discarding the packet in accordance with the determination; and (ii) forwarding the packet for processing to determine whether to retain or discard the packet. The determination whether to forward the packet to its destination or discard the packet may be based, at least in part, on down stream data traffic congestion in a network and/or network device implementing such a method.

In the example method 1300 shown in FIG. 13, where the data packet is locally and finally marked green, forwarding the data packet in the third data flow may include decrementing a token count of a committed information rate token bucket of a dual-token bucket meter associated with the first data flow by an amount corresponding with a size of the data packet. In this situation, forwarding the data packet may also include decrementing a token count of a committed informa-

tion rate token bucket of a dual-token bucket meter associated with the third data flow by the amount corresponding with the size of the data packet.

Also in the example method **1300**, where the data packet is locally marked yellow and finally marked green, forwarding the data packet in the third data flow may include decrementing a token count of an excess information rate token bucket of a dual-token bucket meter associated with the first data flow by an amount corresponding with a size of the data packet. In this situation, forwarding the data packet may also include decrementing a token count of a committed information rate token bucket of a dual-token bucket meter associated with the third data flow by the amount corresponding with the size of the data packet.

Further in the example method **1300**, where the data packet is locally and finally marked yellow, forwarding the data packet in the third data flow may include decrementing a token count of an excess information rate token bucket of a dual-token bucket meter associated with the first data flow by an amount corresponding with a size of the data packet. In this situation, forwarding the data packet may also include decrementing a token count of an excess information rate token bucket of a dual-token bucket meter associated with the third data flow by the amount corresponding with the size of the data packet.

Priority-Based Hierarchical Bandwidth Sharing

FIG. **14** is a block diagram that illustrates an apparatus **1400** in accordance with an example embodiment that may be used to implement priority-based hierarchical bandwidth sharing. FIG. **15** is a block diagram that illustrates an apparatus **1500** in accordance with another example embodiment that may be used to implement priority-based hierarchical bandwidth sharing. In various embodiments, such priority-based bandwidth sharing may be implemented as an alternative to the techniques described above. In other embodiments, the techniques for priority-based bandwidth sharing may be implemented in conjunction with the techniques described above.

While only a single level of data traffic flows is shown in FIG. **14** (and in FIG. **15**), the approaches for priority-based bandwidth sharing discussed herein may be applied to multiple levels of hierarchy in a data processing device. For instance, each incoming flow (e.g., microflow) shown in FIGS. **14** and **15** may be a combination of flows from a previous level of hierarchy (e.g., each incoming flow may be a macroflow of microflows from a previous level of hierarchy). In such an approach, the previous level of hierarchy may also implement priority-based bandwidth sharing using the techniques discussed herein in any appropriate combination. Additionally, the combination of the individual flows that are illustrated in FIGS. **14** and **15** (e.g., the macroflows that include the individual microflows) may be processed in parallel with other flows (e.g., other macroflows) and those flows may also implement priority-based bandwidth sharing using the techniques described herein and, likewise, be combined to produce a macroflow for a subsequent level of hierarchy.

Further, the particular arrangements of the various techniques for priority-based bandwidth sharing described herein are shown by way of illustration and example. These techniques may be applied in a number of different combinations and arrangements, and a given embodiment is not limited to implementing priority-based bandwidth sharing using only the particular arrangements illustrated in the FIGs. and as described herein.

Referring to FIG. **14**, an apparatus **1400** is shown that illustrates an example embodiment of priority-based bandwidth sharing. In the apparatus **1400**, three data traffic flows,

FLOW₁ **1402**, FLOW₂ **1404** and FLOW₃ **1406** are processed in parallel. As indicated in FIG. **14**, FLOW₁ **1402** has a higher priority than FLOW₂ **1404** and FLOW₂ **1404** has a higher priority than FLOW₃. Such an arrangement may be used to implement priority-based bandwidth sharing for a set of COS queues. Such COS queues were previously discussed. It is noted that the apparatus **1400** may process additional data flows (e.g., COS flows) in parallel with those shown in FIG. **14**, and the techniques for priority-based bandwidth sharing described herein may be applied equally to those additional flows in any appropriate arrangement.

The apparatus **1400** includes a first trTCM that may be configured to meter the FLOW₁ **1402**, such as in a manner as described herein. The first trTCM includes a first token bucket **1408** that has a first capacity (bucket depth), which may be referred to as CBS₁ (committed burst size, as previously discussed) for the FLOW₁ **1402**. The first trTCM also includes a second token bucket **1410** that has a second capacity. The second capacity, for this embodiment, represents an EBS₁ (excess burst size, as previously discussed) for the FLOW₁ **1402**.

The first trTCM may be configured to allocate tokens for use in metering the FLOW₁ **1402** to the token buckets **1408** and **1410**. Tokens may be allocated to token bucket **1408** at a rate of CIR₁ (committed information rate, such as previously discussed) and allocated to token bucket **1410** at a rate of EIR₁ (excess information rate, such as previously discussed).

A second trTCM that includes token buckets **1412** and **1414** may be used to meter the FLOW₂ **1404** in like fashion. The second trTCM may, however, allocate tokens to the token buckets **1412** and **1414** at the same or different rates (CIR₂ and EIR₂) than the first trTCM allocates tokens to its token buckets. Likewise, a third trTCM that includes token buckets **1416** and **1418** may be used to meter the FLOW₃ **1406** in like fashion. The third trTCM may allocate tokens to the token buckets **1416** and **1418** at rates of CIR₃ and EIR₃, which may be the same or different than the rates used by the first and second trTCMs.

As illustrated in FIG. **14**, excess bandwidth may be shared between the data traffic flows FLOW₁ **1402**, FLOW₂ **1404** and FLOW₃ **1406**, from higher-priority flows to lower-priority flows by “overflowing” tokens” from higher-priority token buckets to lower-priority buckets when the higher priority buckets are full (e.g., have token counts at or above their capacity or “bucket depth”). For purposes of this disclosure, overflowing of tokens from one token bucket to another is also referred to as reallocating tokens. However, other terms may be used in addition to overflowing and reallocating, such as reassigning, diverting, distributing, among a number of other possible terms. However, for purposes of clarity and consistency, the balance of this disclosure refers to the process of overflowing tokens from one bucket to another as reallocating. Additionally, the term recycling is used to describe instances where excess tokens are fed back from a token bucket associated with a lower priority data traffic flow to a token bucket associated with a higher priority data traffic flow.

As shown in FIG. **14**, the tokens from the token bucket **1408** of the first trTCM may be reallocated to the token bucket **1412** of the second trTCM. This reallocation may occur when the token bucket **1408** has a token count that is equal to or greater than its capacity CBS₁, which indicates that the token bucket **1408** is full (e.g., that the FLOW₁ **1402** is not using its committed bandwidth CIR₁ allocation).

In an example embodiment, the token bucket **1408** (and/or the first trTCM) may utilize an index *i*₁ as a pointer to access the token bucket **1412** in order to reallocate the excess tokens

23

that are not being used by the FLOW₁ 1402. The index may be a fixed index (e.g., where the token bucket 1408 has a fixed relationship with the token bucket 1412 for reallocation of excess token). Alternatively, the index i_1 for reallocating tokens may be programmable and the token bucket 1408 may be configured to reallocate excess tokens to other token buckets of the apparatus 1400 by using different indices. Such an example is described in further detail with respect to FIG. 15.

In like fashion as the token bucket 1408, the token bucket 1410 may be configured to reallocate excess tokens to the token bucket 1414 of the second trTCM. This reallocation may occur when the token bucket 1410 has a token count that is equal to or greater than its capacity, EBS₁, which indicates that the token bucket 1410 is full (e.g., that the FLOW₁ 1402 is not using its excess bandwidth EIR₁ allocation). The token bucket 1410 may use an index i_3 to access the token bucket 1414 to reallocate its excess tokens. As with the index i_1 , the index i_3 may be a fixed index or may be programmable and operate as a pointer to the token bucket 1414.

As further shown in FIG. 14, the token buckets 1412 and 1414 of the second trTCM may also be configured to respectively reallocate excess tokens to the token buckets 1416 and 1418 of the third trTCM of the apparatus 1400. As with the first trTCM, the second trTCM may use fixed or programmable indices, i_2 and i_4 , as pointers to the token buckets 1416 and 1418 of the third trTCM.

As may be inferred from the foregoing discussion, when tokens are reallocated from one token bucket to another, the tokens that are reallocated may originate from different sources. First, tokens that are reallocated from one token bucket to another may be tokens that are periodically allocated to it in accordance with its bandwidth guarantee (e.g., at CIR or EIR for the particular token bucket). Second, tokens that are reallocated from one token bucket to another may be tokens that were reallocated from yet another token bucket. In this situation, if a token bucket is full and attempts to reallocate tokens to another full token bucket, the receiving token bucket may then attempt to again reallocate those tokens to yet another token bucket that is represented by an associated index.

As is also shown in FIG. 14, the token buckets 1416 and 1418 may recycle excess tokens back to a token bucket that is associated with a data traffic flow having a higher priority. For instance, the token bucket 1416 may recycle excess tokens back to the token bucket 1408 (e.g., using a pointer/index i_2 to the token bucket 1408). Likewise, the token bucket 1418 may recycle excess tokens back to the token bucket 1410 using a corresponding index/pointer i_6 .

As was discussed above, the techniques for reallocating excess tokens described herein are shown by way of example. As is discussed below with respect to FIG. 15, excess tokens may be reallocated (including recycling of tokens) in other manners, such as from a committed token bucket associated with a data traffic flow of one priority to an excess bucket associated with a data traffic flow of another priority.

FLOW₁ 1402, FLOW₂ 1404 and FLOW₃ 1406 may be combined using a scheduler 1420 and that combined data traffic flow may then be metered by another trTCM that includes the token buckets 1422 and 1424. The token bucket 1422 may have a bucket depth of CBS_M, which may be the sum of CBSs for FLOW₁ 1402, FLOW₂ 1404 and FLOW₃ 1406. Tokens may be allocated to the token bucket 1422 at a rate CIR_M, which may be the sum of CIRs for FLOW₁ 1402, FLOW₂ 1404 and FLOW₃ 1406. Likewise, the token bucket 1424 may have a bucket a bucket depth of EBS_M, which may be the sum of EBSs for FLOW₁ 1402, FLOW₂ 1404 and FLOW₃ 1406. Tokens may be allocated to the token bucket

24

1422 at a rate EIR_M, which may be the sum of EIRs for FLOW₁ 1402, FLOW₂ 1404 and FLOW₃ 1406. As was discussed above, the combined flow including FLOW₁ 1402, FLOW₂ 1404 and FLOW₃ may be processed in parallel with other combined flows and priority-based bandwidth sharing may be used for those combined flows.

As is also shown in FIG. 14, each of the token buckets 1408-1418 may have a mode flag (MF) associated with it. The respective MFs may be used to indicate whether priority-based bandwidth sharing is enabled for a specific token bucket. In embodiments that employ such a MF, setting a token bucket's MF (assigning it a binary value of '1') may indicate that priority-based bandwidth sharing is enabled for that token bucket. In such an approach, the corresponding token bucket may then reallocate (and/or recycle) tokens to a token bucket that is indicated by a fixed or programmable pointer/index, such as discussed above. In the apparatus 1400, each of the MFs is set (e.g., assigned a binary value of '1') for the token buckets 1408-1418. Accordingly, priority-based bandwidth sharing token reallocation is enabled for each of the token buckets 1408-1418 in the apparatus 1400.

FIG. 15 is a block diagram that illustrates an apparatus 1500 in accordance with another example embodiment that may be used to implement priority-based hierarchical bandwidth sharing. In the apparatus 1500, the arrangement of the data flows FLOW₁ 1502, FLOW₂ 1504 and FLOW₃ 1506 (e.g., from highest to lowest priority), the trTCMs and associated token buckets 1508-1518, 1522 and 1524, and the scheduler 1520 are identical to the arrangement of their corresponding elements in FIG. 14, which are referenced with like 1400 series reference numbers. For the sake of brevity, the elements of the apparatus 1500 are not discussed again here except to describe the techniques for priority-based bandwidth sharing illustrated in FIG. 15.

As shown in FIG. 15, the token buckets 1512 and 1514 of the second trTCM that is used to process the data flow FLOW₂ 1504 have their MFs cleared (are assigned a binary value of '0'), which indicates that priority-based bandwidth sharing is disabled for the token buckets 1512 and 1514. For the apparatus 1500, the second trTCM has another flag associated with it, which may be referred to a coupling flag (CF). Such a CF may be used for data processing devices that operate in accordance with the Metro Ethernet Forum 10.1 Standard. As shown in FIG. 15, when the CF is set for the second trTCM, the first token bucket 1512 of the second trTCM may overflow its excess tokens to the second token bucket 1514 of the second trTCM when the token bucket 1512 is full. Such an approach does not provide for reallocating tokens to token buckets that are used to meter data traffic for other flows (e.g., other levels of data traffic priority or other COSs). Also, in such an approach, if the token bucket 1514 is full, the overflow tokens from the token bucket 1512 are simply discarded and that excess bandwidth may be wasted.

As is further illustrated in FIG. 15, the token bucket 1508 of the first trTCM may be configured to reallocate excess tokens to the token bucket 1516 of the third trTCM. As compared with the approach illustrated in FIG. 14, the token buckets 1512 and 1514 of the second trTCM have their MFs cleared, indicating that priority-based bandwidth sharing is disabled. Therefore, the token bucket 1508 skips over the token bucket 1512 when reallocating excess tokens, in this example, and reallocates its excess tokens to the token bucket 1516 of the third trTCM. In similar fashion as discussed above with respect to the apparatus 1400, token bucket 1508 may use a fixed or programmable index/pointer when reallocating tokens to the token bucket 1516.

25

As is also shown in FIG. 15, the token bucket 1516 (committed bucket of the third trTCM) recycles/reallocates excess tokens to the token bucket 1510 (excess of the first trTCM). Again, a fixed or programmable index may be used to reallocate excess tokens from the token bucket 1516 to the token bucket 1510.

In similar fashion as the token bucket 1508 reallocating excess tokens to the token bucket 1516, the token bucket 1510 may reallocate excess tokens to the token bucket 1518 (excess bucket of the third trTCM), skipping over the token bucket 1514, which has its MF cleared.

As is further shown in FIG. 15, the token bucket 1518 also has its MF cleared. In this situation, the token bucket 1518 may discard excess tokens when the token bucket 1518 is full. Excess tokens of the token bucket 1518 may be tokens that are allocated to it by the third trTCM in accordance with its EIR₃.

As was previously noted the particular approaches for priority-based bandwidth sharing discussed herein, such as those illustrated in FIGS. 14 and 15, are given by way of example. These techniques may be modified and combined in a number of appropriate manners. Accordingly, embodiments of network devices implementing priority-based bandwidth sharing and related methods are not limited to the particular arrangements discussed herein.

FIG. 16 is flowchart illustrating a method 1600 for priority-based bandwidth sharing in accordance with an example embodiment. Such a method may be implemented, for example, using the apparatus illustrated in FIGS. 14 and 15, though such a method may be implemented in other data processing devices. The method 1600 may also be implemented using the techniques described above with respect to FIGS. 14 and 15. For the sake of brevity, those techniques will not be described again in detail here except as they relate to the method operations.

The method 1600 may include, at block 1605, processing a first data traffic flow and, at block 1610, metering the first data traffic flow with a first two-rate, three-color meter (trTCM), where a first token bucket of the first trTCM has a first token capacity and a second token bucket of the first trTCM has a second token capacity. The method 1600 may further include, at block 1615, allocating tokens to the first token bucket of the first trTCM at a first rate and, at block 1620, allocating tokens to the second token bucket of the first trTCM at a second rate.

The method 1600 may still further include, at block 1625, processing a second data traffic flow in parallel with the first data traffic flow and, at block 1630, metering the second data traffic flow with a second trTCM, where a first token bucket of the second trTCM has a third token capacity and a second token bucket of the second trTCM has a fourth token capacity. Method 1600 may also include, at block 1635, allocating tokens to the first token bucket of the second trTCM at a third rate and, at block 1640, allocating tokens to the second token bucket of the second trTCM at a fourth rate.

The method 1600 may still further include, at block 1645, reallocating tokens allocated to the first token bucket of the first trTCM to the first token bucket of the second trTCM when a token count of the first token bucket of the first trTCM exceeds the first capacity, such as in accordance with the priority-based bandwidth sharing techniques described herein. At block 1650, the method 1600, may also include reallocating tokens allocated to the second token bucket of the first trTCM to the second token bucket of the second trTCM when a token count of the second token bucket of the first trTCM exceeds the second capacity, such as in accordance with the priority-based bandwidth sharing techniques described herein.

26

FIGS. 17-20 illustrate methods 1700, 1800, 1900 and 2000 for priority-based bandwidth that may implemented in conjunction with the method 1600. As with the method 1600, the methods 1700, 1800, 1900 and 2000 may be implemented, for example, using the apparatus illustrated in FIGS. 14 and 15, though such methods may be implemented in other data processing devices. The methods 1700, 1800, 1900 and 2000 may also be implemented using the techniques described above with respect to FIGS. 14 and 15. For the sake of brevity, those techniques will not be described again in detail here except as they relate to the method operations. Also, for purposes of illustration, each of the methods 1700-2000 will be described as being implemented in conjunction with the method 1600 and will, accordingly, make reference to the operations of the method 1600.

The method 1700, when implemented in conjunction with the method 1600, may include, at block 1705, reallocating tokens allocated to the first token bucket of the second trTCM to the first token bucket of the first trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity. The method 1700, may also include, at block 1710, reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

The method 1800, when implemented in conjunction with the method 1600, may include, at block 1805, reallocating tokens allocated to the first token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity. The method 1800, may also include, at block 1810, reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

The method 1900, when implemented in conjunction with the method 1600, may include, at block 1905, processing a third data traffic flow in parallel with the first data traffic flow and the second data traffic flow, the third data traffic flow having a lower priority than the first data traffic flow and a higher priority than the second data traffic flow. The method 1900 may further include, at block 1910, metering the third data traffic flow with a third trTCM, wherein a first token bucket of the third trTCM has a fifth token capacity and a second token bucket of the third trTCM has a sixth token capacity. The method 1900 may still further include, at block 1915, allocating tokens to the first token bucket of the third trTCM at a fifth rate and, at block 1920, allocating tokens to the second token bucket of the third trTCM at a sixth rate. The method 1900 may still further include, at block 1925, reallocating tokens allocated to the first token bucket of the third trTCM to the second token bucket of the third trTCM when a token count of the first token bucket of the third trTCM exceeds the fifth capacity.

The method 2000, when implemented in conjunction with the method 1600, may include, at block 2005, processing a third data traffic flow in parallel with the first data traffic flow and the second data traffic flow, the third data traffic flow having a lower priority than the first data traffic flow and the second data traffic flow and, at block 2010, metering the third data traffic flow with a third trTCM, wherein a first token bucket of the third trTCM has a fifth token capacity and a second token bucket of the third trTCM has a sixth token capacity. The method 2000 may further include, at block 2015, allocating tokens to the first token bucket of the third trTCM at a fifth rate and, at block 2020, allocating tokens to the second token bucket of the third trTCM at a sixth rate.

27

The method **2000** may still further include, at block **2025**, reallocating tokens allocated to the first token bucket of the second trTCM to the first token bucket of the third trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity. The method **2000** may also include, at block **2030**, reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the third trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

Implementations of the various techniques described herein may be implemented in digital electronic circuitry, or in computer hardware, firmware, software, or in combinations of them. Implementations may be implemented as a computer program product, i.e., a computer program tangibly embodied in an information carrier, e.g., in a machine-readable storage device, for execution by, or to control the operation of, data processing apparatus, e.g., a programmable processor, a computer, or multiple computers. A computer program, such as the computer program(s) described above, can be written in any form of programming language, including compiled or interpreted languages, and can be deployed in any form, including as a stand-alone program or as a module, component, subroutine, or other unit suitable for use in a computing environment. A computer program can be deployed to be executed on one computer or on multiple computers at one site or distributed across multiple sites and interconnected by a communication network.

Method steps may be performed by one or more programmable processors executing a computer program to perform functions by operating on input data and generating output. Method steps also may be performed by, and an apparatus may be implemented as, special purpose logic circuitry, e.g., an FPGA (field programmable gate array) or an ASIC (application-specific integrated circuit).

Processors suitable for the execution of a computer program include, by way of example, both general and special purpose microprocessors, and any one or more processors of any kind of digital computer. Generally, a processor will receive instructions and data from a read-only memory or a random access memory or both. Elements of a computer may include at least one processor for executing instructions and one or more memory devices for storing instructions and data. Generally, a computer also may include, or be operatively coupled to receive data from or transfer data to, or both, one or more mass storage devices for storing data, e.g., magnetic, magneto-optical disks, or optical disks. Information carriers suitable for embodying computer program instructions and data include all forms of non-volatile memory, including by way of example semiconductor memory devices, e.g., EPROM, EEPROM, and flash memory devices; magnetic disks, e.g., internal hard disks or removable disks; magneto-optical disks; and CD-ROM and DVD-ROM disks. The processor and the memory may be supplemented by, or incorporated in special purpose logic circuitry.

To provide for interaction with a user, implementations may be implemented on a computer having a display device, e.g., a cathode ray tube (CRT) or liquid crystal display (LCD) monitor, for displaying information to the user and a keyboard and a pointing device, e.g., a mouse or a trackball, by which the user can provide input to the computer. Other kinds of devices can be used to provide for interaction with a user as well; for example, feedback provided to the user can be any form of sensory feedback, e.g., visual feedback, auditory feedback, or tactile feedback; and input from the user can be received in any form, including acoustic, speech, or tactile input.

28

Implementations may be implemented in a computing system that includes a back-end component, e.g., as a data server, or that includes a middleware component, e.g., an application server, or that includes a front-end component, e.g., a client computer having a graphical user interface or a Web browser through which a user can interact with an implementation, or any combination of such back-end, middleware, or front-end components. Components may be interconnected by any form or medium of digital data communication, e.g., a communication network. Examples of communication networks include a local area network (LAN) and a wide area network (WAN), e.g., the Internet.

While certain features of the described implementations have been illustrated as described herein, many modifications, substitutions, changes and equivalents will now occur to those skilled in the art. It is, therefore, to be understood that the appended claims are intended to cover all such modifications and changes as fall within the true spirit of the embodiments of the invention.

What is claimed is:

1. A data communication apparatus comprising:
a processor;

a non-transitory machine readable medium having instructions stored therein, the instructions, when executed by the processor, result in the data communication apparatus:

metering a first data traffic flow with a first two-rate, three-color meter (trTCM), wherein a first token bucket of the first trTCM has a first token capacity and a second token bucket of the first trTCM has a second token capacity;

allocating tokens to the first token bucket of the first trTCM at a first rate and allocating tokens to the second token bucket of the first trTCM at a second rate;

metering a second data traffic flow with a second trTCM, wherein a first token bucket of the second trTCM has a third token capacity and a second token bucket of the second trTCM has a fourth token capacity;

allocating tokens to the first token bucket of the second trTCM at a third rate and allocating tokens to the second token bucket of the second trTCM at a fourth rate;

reallocating tokens allocated to the first token bucket of the first trTCM to the first token bucket of the second trTCM when a token count of the first token bucket of the first trTCM exceeds the first capacity; and

reallocating tokens allocated to the second token bucket of the first trTCM to the second token bucket of the second trTCM when a token count of the second token bucket of the first trTCM exceeds the second capacity.

2. The data communication apparatus of claim 1, wherein the first data traffic flow is assigned a higher priority than the second data traffic flow.

3. The data communication apparatus of claim 1, wherein the instructions, when executed by the processor, further result in the data communication apparatus:

reallocating tokens allocated to the first token bucket of the second trTCM to the first token bucket of the first trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity; and

reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

29

4. The data communication apparatus of claim 3, wherein: reallocating tokens allocated to the first token bucket of the second trTCM to the first token bucket of the first trTCM comprises at least one of:

reallocating at least a portion of the tokens allocated to the first token bucket of the second trTCM at the third rate; and

reallocating at least a portion of the tokens reallocated from the first token bucket of the first trTCM to the first token bucket of the second trTCM; and

reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the first trTCM comprises at least one of:

reallocating at least a portion of the tokens allocated to the second token bucket of the second trTCM at the third rate; and

reallocating at least a portion of the tokens reallocated from the second token bucket of the second trTCM to the second token bucket of the second trTCM.

5. The data communication apparatus of claim 1, wherein the instructions, when executed by the processor, further result in the data communication apparatus reallocating tokens allocated to the first token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity.

6. The data communication apparatus of claim 5, wherein the instructions, when executed by the processor, further result in the data communication apparatus reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

7. The data communication apparatus of claim 1, wherein the instructions, when executed by the processor, further result in the data communication apparatus:

processing a third data traffic flow in parallel with the first data traffic flow and the second data traffic flow, the third data traffic flow having a lower priority than the first data traffic flow and a higher priority than the second data traffic flow;

metering the third data traffic flow with a third trTCM, wherein a first token bucket of the third trTCM has a fifth token capacity and a second token bucket of the third trTCM has a sixth token capacity;

allocating tokens to the first token bucket of the third trTCM at a fifth rate;

allocating tokens to the second token bucket of the third trTCM at a sixth rate; and

reallocating tokens allocated to the first token bucket of the third trTCM to the second token bucket of the third trTCM when a token count of the first token bucket of the third trTCM exceeds the fifth capacity.

8. The data communication apparatus of claim 7, wherein: reallocating tokens allocated to the first token bucket of the first trTCM to the first token bucket of the second trTCM is done in response to a first flag being set, the first flag being associated with the first token bucket of the first trTCM;

reallocating tokens allocated to the second token bucket of the first trTCM to the second token bucket of the second trTCM is done in response to a second flag being set, the second flag being associated with the second token bucket of the first trTCM; and

reallocating tokens allocated to the first token bucket of the third trTCM to the second token bucket of the third trTCM is done in response to a third flag being cleared

30

and a fourth flag being set, the third flag being associated with the first token bucket of the third trTCM being and the fourth flag being associated with the third trTCM.

9. The data communication apparatus of claim 1, wherein the instructions, when executed by the processor, further result in the data communication apparatus:

processing a third data traffic flow in parallel with the first data traffic flow and the second data traffic flow, the third data traffic flow having a lower priority than the first data traffic flow and the second data traffic flow;

metering the third data traffic flow with a third trTCM, wherein a first token bucket of the third trTCM has a fifth token capacity and a second token bucket of the third trTCM has a sixth token capacity;

allocating tokens to the first token bucket of the third trTCM at a fifth rate;

allocating tokens to the second token bucket of the third trTCM at a sixth rate;

reallocating tokens allocated to the first token bucket of the second trTCM to the first token bucket of the third trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity; and

reallocating tokens allocated to the second token bucket of the second trTCM to the second token bucket of the third trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

10. The data communication apparatus of claim 9, wherein the instructions, when executed by the processor, further result in the data communication apparatus:

reallocating tokens allocated to the first token bucket of the third trTCM to the first token bucket of the first trTCM when a token count of the first token bucket of the third trTCM exceeds the fifth capacity; and

reallocating tokens allocated to the second token bucket of the third trTCM to the second token bucket of the first trTCM when a token count of the second token bucket of the third trTCM exceeds the sixth capacity.

11. The data communication apparatus of claim 1, wherein:

the first rate comprises a committed information rate (CIR) of the first data traffic flow;

the second rate comprises an excess information rate (EIR) of the first data traffic flow;

the third rate comprises a CIR of the second data traffic flow; and

the fourth rate comprises an EIR of the second data traffic flow.

12. The data communication apparatus of claim 1, wherein the instructions, when executed by the processor, further result in the data communication apparatus:

in response to a flag associated with the first token bucket of the second trTCM being cleared, discarding tokens reallocated to the first token bucket of the second trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity; and

in response to a flag associated with the second token bucket of the second trTCM being cleared, discarding tokens reallocated to the second token bucket of the second trTCM when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

13. The data communication apparatus of claim 1, wherein:

reallocating tokens allocated to the first token bucket of the first trTCM to the first token bucket of the second trTCM comprises reallocating tokens allocated to the first token

31

bucket of the first trTCM to the first token bucket of the second trTCM based on an index of the first token bucket of the second trTCM; and
 reallocating tokens allocated to the second token bucket of the first trTCM to the second token bucket of the second trTCM comprises reallocating tokens allocated to the second token bucket of the first trTCM to the second token bucket of the second trTCM based on an index of the second token bucket of the second trTCM.

14. A network device comprising:
 a first two-rate, three-color meter (trTCM) configured to meter a first data traffic flow, the first trTCM comprising:
 a first token bucket having a first capacity; and
 a second token bucket having a second capacity;
 a second trTCM configured to meter a second data traffic flow, the second trTCM comprising:
 a first token bucket having a third capacity; and
 a second token bucket having a fourth capacity,
 wherein:
 the first token bucket of the first trTCM is configured to reallocate tokens allocated to the first token bucket of the first trTCM to the first token bucket of the second trTCM when a token count of the first token bucket of the first trTCM exceeds the first capacity; and
 the second token bucket of the first trTCM is configured to reallocate tokens allocated to the second token bucket of the first trTCM to the second token bucket of the second trTCM when a token count of the second token bucket of the first trTCM exceeds the second capacity.

15. The network device of claim 14, further comprising:
 a third trTCM configured to meter a third data traffic flow, the third trTCM comprising:
 a first token bucket having a fifth capacity; and
 a second token bucket having a sixth capacity,
 wherein:
 the third data traffic flow has a lower priority than the first data traffic flow and a higher priority than the second data traffic flow; and
 the first token bucket of the third trTCM is configured to reallocate tokens allocated to the first token bucket of the third trTCM to the second token bucket of the third trTCM when a token count of the first token bucket of the third trTCM exceeds the fifth capacity.

16. The network device of claim 14, wherein the first token bucket of the second trTCM is configured to reallocate tokens allocated to the first token bucket of the second trTCM to the second token bucket of the first trTCM when a token count of the first token bucket of the second trTCM exceeds the third capacity.

17. The network device of claim 14;
 a third trTCM configured to meter a third data traffic flow, the third trTCM comprising:

32

a first token bucket having a fifth capacity; and
 a second token bucket having a sixth capacity,
 wherein:
 the third data traffic flow has a lower priority than the first data traffic flow and a lower priority than the second data traffic flow; and
 the first token bucket of the third trTCM is configured to reallocate tokens allocated to it to the second token bucket of the first trTCM when a token count of the first token bucket of the third trTCM exceeds the fifth capacity.

18. The network device of claim 17, wherein the tokens allocated to the first token bucket of the third trTCM comprise:
 tokens allocated to the first token bucket of the third trTCM by the third trTCM; and
 tokens reallocated to the first token bucket of the third trTCM by a token bucket of a trTCM that is configured to process a data traffic flow with a higher priority than the third data traffic flow and a lower priority of the first data traffic flow.

19. The network device of claim 14, wherein:
 the first token bucket of the second trTCM is configured to reallocate tokens allocated to it to a token bucket of a trTCM that is configured to process a data traffic flow having a higher priority than the second data traffic flow when a token count of the first token bucket of the second trTCM exceeds the third capacity; and
 the second token bucket of the second trTCM is configured to discard tokens allocated to it when a token count of the second token bucket of the second trTCM exceeds the fourth capacity.

20. A network device comprising:
 a first two-rate, three-color meter (trTCM) comprising a first token bucket having a first capacity and a second token bucket having a second capacity, wherein the first trTCM is configured to meter a first data traffic flow; and
 a second trTCM comprising a first token bucket and a second token bucket, wherein the second trTCM is configured to meter a second data traffic flow,
 wherein:
 the first token bucket of the first trTCM is configured to reallocate tokens allocated to it to the first token bucket of the second trTCM when a token count of the first token bucket of the first trTCM exceeds the first capacity; and
 the second token bucket of the first trTCM is configured to reallocate tokens allocated to it to the second token bucket of the second trTCM when a token count of the second token bucket of the first trTCM exceeds the second capacity.

* * * * *