

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
5 April 2001 (05.04.2001)

PCT

(10) International Publication Number  
WO 01/24004 A2

(51) International Patent Classification: G06F 9/46

(21) International Application Number: PCT/US00/26879

(22) International Filing Date:  
29 September 2000 (29.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/410,204 30 September 1999 (30.09.1999) US

(71) Applicant: QUALCOMM INCORPORATED [US/US];  
5775 Morehouse Drive, San Diego, CA 92121-1714 (US).

(72) Inventors: PANKAJ, Rajesh; 7356 Park Village Road,  
San Diego, CA 92129 (US). GROB, Matthew, Stuart;  
2757 Bordeaux Avenue, La Jolla, CA 92037 (US). BEN-  
DER, Paul, E.; 2879 Angell Avenue, San Diego, CA 92122  
(US).

(74) Agents: WADSWORTH, Philip, R. et al.; Qualcomm In-  
corporated, 5775 Morehouse Drive, San Diego, CA 92121-  
1714 (US).

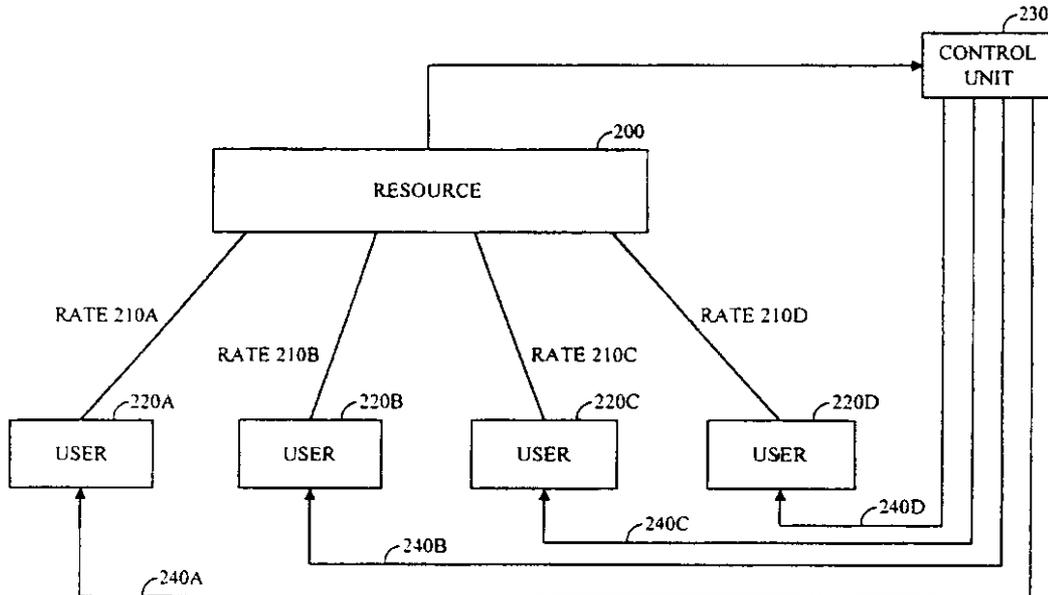
(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

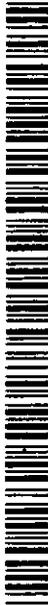
Published:  
— Without international search report and to be republished  
upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR PERSISTENCE-VECTOR-BASED MODIFICATION OF USAGE RATES



(57) Abstract: When a resource of limited capacity is shared by several users, it is possible for the usage rates of the users to exceed the resource's capacity, thereby causing an overload condition. In a system or method according to an embodiment of the invention, at least some of the users have a set of persistence vectors. When an overload condition is detected, the usage rate of at least one of these users is changed, at least in part according to the user's set of persistence vectors.



WO 01/24004 A2

39  
21: 5458

## SYSTEM AND METHOD FOR PERSISTENCE-VECTOR-BASED MODIFICATION OF USAGE RATES

### BACKGROUND OF THE INVENTION

#### 5 **Field of the Invention**

This invention relates to distribution of the use of a limited resource among multiple users. More specifically, this invention relates to the modification of usage rates according to a set of persistence vectors.

#### 10 **Description of Related Art and General Background**

A shared resource is one which may be used by multiple users. Shared resources which have limited availabilities or capacities include such diverse examples as electric power stations and other energy plants, water sources such as reservoirs and flowing bodies, supply systems for the distribution of goods and/or  
15 materiel, and data communications networks and pathways. Problems associated with allocating the use of a shared resource among multiple users may therefore arise in many different contexts. Regardless of the particular context, however, such resources may be found in many systems in which at least the following conditions hold:

- 20 • the capacity or availability of the shared resource may be expressed in terms of a finite rate  $R$  of units per measure of time (i.e. kilowatts/hour, gallons/minute, cartons/week, or bits/second);
- at any particular time, the resource is being used by  $n$  different users, where  $n$  is a nonnegative integer; and
- 25 • at any particular time, the usage of the  $i$ -th user (where  $1 \leq i \leq n$ ) may be characterized by a finite usage rate  $u_i$  of units per measure of time.

A basic model for such a system is shown in FIG. 1, where resource 100 is used by users 120a-d at rates 110a-d, respectively. Depending on the particular implementation, the rate R which characterizes the shared resource may indicate an actual or estimated limit of the capacity of the resource (e.g. in the case of a communications pathway) or, in the alternative, the rate R may be a threshold indicating a maximum safe or permissible load of the resource (e.g. in the case of a power generation facility or device). Likewise, the usage rates  $u_i$  may indicate actual use, expected use, or requests or demands for use.

An overload condition arises when the sum of the  $n$  usage rates  $u_i$  at any one time exceeds the value R. With respect to a power plant, for example, an overload condition may arise when the total current draw exceeds the rated capacity. With respect to a data communications pathway, an overload condition may arise when the total data transfer rate exceeds the pathway's actual capacity, thereby corrupting the data in transmission. In certain situations such as water supply or warehousing of materials, an overload condition may also indicate that although user demands are currently being met, reserve or buffer capacity is being depleted.

Depending on the nature of the resource, the consequences of an overload condition will vary, possibly including the need for an offline period for resource recovery (e.g. cooling of an power generation system or replenishment of a reservoir) or the need to expend present capacity in order to repeat a use that was attempted in the past but failed because of the overload (e.g. retransmission of a data packet corrupted by a collision). The resource may even become temporarily or permanently unable to regain its former capacity. In any case, it is generally desirable to avoid overload conditions whenever possible.

## SUMMARY OF THE INVENTION

A system according to an embodiment of the invention includes a resource and a number of users of the resource. Each user of the resource has a usage rate and a set of persistence vectors, and the user's use of the resource is determined at least in part  
5 by the user's usage rate. When a predetermined relationship arises between a sum of the usage rates and a certain measure of the capacity of the resource, then at least one of the users changes its usage rate according to at least its set of persistence vectors.

## BRIEF DESCRIPTION OF THE DRAWINGS

10 FIG. 1 shows a diagram of a system having a shared resource.

FIG. 2 shows a diagram of a system having a shared resource and a control unit.

FIG. 3 shows a diagram of a system having a consumer, a plurality of producers, and a common channel.

15 FIG. 4 shows a method according to a first embodiment of the invention.

FIG. 5 shows a method according to a second embodiment of the invention.

FIG. 6 shows a variation of the method of FIG. 5.

FIG. 7 shows an additional variation of the method of FIG. 5.

## 20 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

When an overload condition arises in a system according to FIG. 1, the users  
120 may not be aware that an overload has occurred, especially if the resource is consuming reserve capacity in order to meet user demands. Even if the overload condition causes the resource's availability to a user to drop below a user's  
25 expectation or demand, the user may be unable to verify whether the shortcoming is

16x20=380

due to a resource overload or to the failure of another component in the supply path. Moreover, in certain applications such as wireless data communications, it is possible that no feedback mechanism exists whereby a user may obtain timely notification of an overload. Therefore, the user may continue to use the resource, unaware of the  
5 problem. In such a situation, it is desirable for the system to include a capability for notifying the users of the overload condition via, e.g., a warning signal.

FIG. 2 shows an example of such a system, wherein control unit 230 receives information related to usage of resource 200 by users 220a-d and sends feedback information such as a warning signal to users 220a-d over respective communications  
10 pathways 240a-d. Note that it is possible for control unit 230 to be implemented as a part of resource 200 or alternatively as a part of one of the users 220a-d.

If a user becomes aware of an overload condition, then the possibility exists for user-driven remediation. In this case, if at least some of the users are able to communicate with each other, then a solution such as a reduction in usage rate may be  
15 negotiated. In many instances, however, such communication between users may be unavailable, impractical, or otherwise undesirable, in which case an alternate control mechanism may be provided for controlling usage of the resource. This alternate control mechanism may be centralized and/or decentralized.

If complete knowledge of the future usage requirements of the users were  
20 available, then it would be theoretically possible to construct an optimal usage schedule that would satisfy the users' requirements as much as possible while completely avoiding all overload conditions. In many practical systems, however, a user's future needs will be unknown even to the user itself. One way to prevent overload conditions in such systems would be on the basis of current usage  
25 requirements: for example, by granting usage rate allocations to users only on a

request basis. In order to convey usage requests from the users back to the control unit, however, such a scheme would require an upstream communication pathway which may not otherwise be necessary. Moreover, additional costs and delays are incurred in receiving, processing and responding to such requests.

5 In order to avoid some of the disadvantages of a request/grant scheme, a decentralized system may be designed wherein control is shared with the users. The control unit in such a system concentrates on the prediction and avoidance of overload conditions while issuing enough feedback information to allow the users to control their own usage to some extent.

10 A method according to an embodiment of the invention may be implemented in any system that fits the model of FIG. 1 wherein the users may obtain notification of an overload condition (as in the modified system of FIG. 2). An exemplary application of such a system is shown in FIG. 3 wherein users 320a-d are data producers, resource 300 is a common transmission channel linking the producers with  
15 data consumer 350, and control unit 330 receives usage information from the consumer. The producers use resource 300 by transmitting data to consumer 350 at or below rates 310a-d, respectively, and they receive respective signals 340a-d (which may include feedback and/or other control information) from the control unit.

One possible implementation of the exemplary application is the reverse link  
20 of a CDMA telecommunications system. In this case, each producer may comprise 1) a transmitter, such as a mobile telephone or a WLL (wireless local loop) station, connected to 2) a data-producing device, such as a laptop computer or a point-of-sale terminal, through a PCMCIA card or a similar interface, and outputting data encapsulated in packets over IP or any other suitable protocol. Consumer 350 and  
25 control unit 330 may be parts of a base station, and control signals 340 may be carried

over a forward link. Several generations and versions of CDMA telecommunications systems have already been implemented. While most of these CDMA systems have been designed to carry digitized voice communications, however, the method herein described is especially well-suited to a network serving producers with widely  
5 varying transmission rates, such as a data-only network or a mixed voice-data network.

A method according to a first embodiment of the invention is described in FIG. 4 with reference to the system of FIG. 2. In this method, a user's use of the resource at any particular time is determined in relation to a predetermined usage rate.  
10 As noted in block 400, a particular user is configured to have a usage rate  $r_j$ . The usage rate  $r_j$  is one among a set of  $m$  predetermined available rates  $r_1$  to  $r_m$ , where the relation  $a < b$  implies that  $r_a < r_b$ . It is not necessary for all users to have the same set of available rates, but the set for each user should be known to control unit 230 so that it may reliably predict the state of resource use and issue a warning signal  
15 appropriately. It is also possible for each user's set of available rates to be updated by control unit 230 whether periodically or otherwise. Schemes of rate selection, assignment, and allocation that may be used in systems incorporating an embodiment of the invention include those described in the copending Patent Applications Nos. 09/264,297, entitled "METHOD OF RATE ALLOCATION IN A DATA  
20 COMMUNICATIONS NETWORK," filed March 4, 1999 and assigned to the assignee of the present invention, and, 09/XXX, XXX entitled "METHOD AND APPARATUS FOR PERSISTENCE-VECTOR-BASED RATE ASSIGNMENT," filed concurrently herewith, assigned to the assignee of the present invention, and the disclosure of which application is hereby incorporated by reference.

11x23=253

Note that the usage rate  $r_j$  may indicate a maximum allowable rate, i.e. a permission rather than a requirement to use the resource at the given rate. The actual rate at which the user uses the resource may depend upon other factors in addition to the usage rate, such as a user's current need and/or ability to use the resource.

5 Likewise, note that the actual rate at which the user uses the resource need not be a member of the set of available rates.

In one particular implementation, each user has the same fixed set of available rates, wherein each rate is expressed in kilobits per second (Kb/s) and the set of rates is designed to increment in powers of two. Because a doubling in rate requires a

10 doubling in power to maintain the same ratio of energy per bit to noise power spectral density ( $E_b/N_0$ ), each rate step thus corresponds to a power step of 3 dB. The available rate values in this example include 4.8, 9.6, 19.2, 38.4, 76.8, 153.6, and 307.2 Kb/s.

In addition to a usage rate, each user also has a set of persistence vectors,

15 although it is possible to have other users in the system that lack a set of persistence vectors. The length of each such vector may be any integer greater than zero, and each vector element corresponds to one among the set of available rates and represents a probability that the usage rate will be the corresponding one among the set of available rates. In the exemplary application, each vector element is a

20 persistence value which represents a probability from 0 to 1. The set of persistence vectors may be unique to each user, or the same set may be assigned to all users in a particular class, or the same set may be assigned to all of the users in the system. Likewise, the set of persistence vectors may be a permanent aspect of the operation of the user, or it may be issued by control unit 230, in which case it may be updated

25 periodically or otherwise. Other relevant aspects of persistence vector distribution

and use are discussed in the co-pending Application No. 09/XXX,XXX entitled "METHOD AND APPARATUS FOR PERSISTENCE-VECTOR-BASED RATE ASSIGNMENT," the disclosure of which application is incorporated by reference above.

5 In this method, the user's set of persistence vectors includes an  $(m-1)$ -element vector  $P$ , wherein  $P = \{P_k \text{ such that } 1 \leq k \leq m-1\}$  and  $m$  is the number of members of the user's set of available rates. (The vector  $P$  may be the only vector in the set of persistence vectors, or vector  $P$  may be selected from among others in the set according to such criteria as the most recent usage rate or the most recent actual rate  
10 for this user.) Vector  $P$  may (but is not required to) have the form of a probability density function, wherein the sum of its elements (or of the values represented by its elements) is equal or substantially equal to one.

In block 410, the user receives a warning signal from control unit 230. This warning signal may issue, for example, when an actual or impending overload  
15 condition is detected, and it may be sent to all users or only to a subset of the users (e.g. only to the users who have persistence vectors). Various embodiments and applications of a system wherein the warning signal is indicated by a busy bit in a reverse link signal are described in co-pending Application No. 09/346,882 entitled  
20 "METHOD AND APPARATUS FOR SIGNAL COMBINING IN A HIGH DATA RATE COMMUNICATION SYSTEM," filed July 2, 1999 and assigned to the assignee of the present invention.

Upon receiving the warning signal, the user generates a random number  $x$  as indicated in block 420. The range and distribution of  $x$  are limited only by the particular implementation: in an exemplary application,  $x$  represents a value drawn  
25 from a set having a uniform distribution over the range 0 to 1. In block 430, the value

92 24 = 216

of  $x$  is tested against the persistence value  $P_j$ , where  $P_j$  is the element of persistence vector  $P$  that corresponds to usage rate  $r_j$ . If the test fails (i.e.  $x$  is not less than  $P_j$ ), then the user's usage rate is not affected by the overload condition, as shown in block 440. If the test succeeds (i.e.  $x$  is less than  $P_j$ ), however, then the user's usage rate is  
5 decreased from  $r_j$  to  $r_{j-1}$ , as shown in block 450. If the user's usage rate is already the lowest rate in the user's set of available rates, then success in block 450 may indicate a reduction to a predetermined lower rate or even a denial of service. This method may be altered to allow the use of one among many other relations between the values of  $x$  and  $P_j$  in place of the test condition shown in block 430, depending on the  
10 particular characteristics of the values chosen for  $x$  and  $P_j$ .

Note that the values given to the elements of persistence vector  $P$  will in part influence how the redistribution of resource usage is biased among users starting with different usage rates. For example, a redistribution which is more equitable may be achieved by choosing large values for elements of persistence vector  $P$  which  
15 correspond to high usage rates and low values for elements of  $P$  which correspond to low usage rates. Such a scheme will make it more likely that a user currently having a high usage rate will reduce its rate, while making it less likely that a user already having a low usage rate will have to reduce its rate any further. Note as well that in a case where each persistence vector is associated with a particular member of the set of  
20 available rates, the relations between these vectors will also bias the redistribution of resource usage. Also note that use of the rate doubling scheme described above (or a similar non-constant distribution within the set of usage rates) will allow usage rate reductions by high-rate users to free up more resource capacity than usage rate reductions by low-rate users.

Numerous variations of the method described above may be used in applications of this embodiment. For example, the users may share the same set of persistence vectors, or different sets of persistence vectors may be assigned to allow the implementation of a priority scheme among the users. In another variation, the first element of each persistence vector may be eliminated (or set to represent a probability of 1) so that users already having the lowest usage rate will not suffer a further rate reduction. Likewise, more than one among the first elements of the persistence vectors may be so treated to protect users of other low rates.

Additional constraints on usage rate may exist as consequences of other aspects of the particular implementation. For example, the rate at which the user actually uses or accesses the shared resource may be limited by factors such as the user's present capacity or power. Therefore, it is possible that the user may use or may be permitted to use a rate lower than the usage rate granted by this or a similar method.

It may be desirable to choose rate  $R$  (a capacity measure of the shared resource) to be a threshold value rather than the actual capacity of the shared resource so that the warning signal is generated before an overload condition occurs, thereby allowing the system to react to avoid the condition. In this case, the threshold  $R$  should be selected to take into account at least (1) the longest possible delay in system response, as characterized by the maximum time between generation of the warning signal and the consequent reduction in total resource usage, and (2) the maximum possible increase in resource usage during the period of such delay.

A method according to a second embodiment of the invention is described in FIG. 5 with reference to FIG. 2. In contrast to the method described above, this method allows the user's usage rate to be reduced to any other rate in the set of

$$13 \times 24 = 312$$

available rates rather than to only one particular rate. As in the method described above, a user is configured to have a usage rate  $r_j$  from the user's set of available rates  $r_1$  to  $r_m$  (as noted in block 500) and an  $(m-1)$ -element persistence vector  $P$  which may be selected from a set according to, for example, the index  $j$ . In block 510, a warning signal is received from control unit 230, and in block 520 the user generates a random number  $x$  as described above. At this stage, the user also sets an index  $k$  to be equal to the index  $j$ .

In block 530, the value of  $x$  is tested against the persistence value  $P_k$ , where  $P_k$  is the element of persistence vector  $P$  that corresponds to usage rate  $u_k$ . If the test fails (i.e.  $x$  is not less than  $P_k$ ), then the index  $j$  is set to be equal to  $k$  in block 560, and the method ends in block 570 with the user being configured to have the usage rate  $r_j$ . In this case, in other words, the user's usage rate is not affected by the overload condition.

If the test in block 530 succeeds (i.e.  $x$  is less than  $P_k$ ), then the value of the index  $k$  is tested. If  $k$  is already at its minimum value (i.e. one in this example), then the procedure continues to blocks 560 and 570 as above. Otherwise, the value of  $k$  is decremented (i.e. reduced by one) and the test is repeated. Under this method, when block 570 is finally reached, the user may be configured to have any usage rate in the set which is equal to or less than the usage rate indicated in block 500. Again, this method may be altered to allow the use of one among many other relations between the values of  $x$  and  $P_k$  in place of the test condition shown in block 530, depending on the particular characteristics of the values chosen for  $x$  and  $P_k$ .

In a variation of this method as shown in FIG. 6, it is possible for the user to be denied usage of the shared resource. Block 540 is replaced with block 542, which allows the index  $k$  to reach a value of zero. When that event occurs, the user is

configured to have a null usage rate in block 580. This null usage rate may represent some predetermined rate outside the set of available rates (e.g. a minimal rate which draws from reserved capacity) or it may represent a usage rate of zero and thus a complete denial of usage. FIG. 7 shows an additional variation of the method of FIG. 5, wherein a new random number  $x$  is generated in block 526 at every iteration of the loop (in this variation, block 520 may be reduced as in block 522 to include only the initialization of index  $k$ ).

With respect to the methods shown in FIGs. 4–7, note that a minimum bound of the selected usage rate may be established by setting elements of the persistence vector which correspond to that rate and to any lesser rates to indicate a probability of 1 (i.e. setting these elements to zero in the examples of FIGs. 4–7). In such a case, the tests in blocks 430 and 530 will fail when that rate is reached (or when the procedure is called with the user already having a lower usage rate), and no further reduction in usage rate will occur.

The foregoing description of the preferred embodiments is provided to enable any person skilled in the art to make or use the present invention. Various modifications to these embodiments are possible, and the generic principles presented herein may be applied to other embodiments as well. For example, indices such as those for the set of available rates and the persistence vector being referenced may begin at zero, or at any other number or symbol, rather than beginning at one. Likewise, in a set of available rates, the relation  $a < b$  may imply that  $r_a > r_b$ , or the various rates may be arranged in some other order instead.

Additionally, the invention may be implemented in part or in whole as hard-wired circuits, as circuit configurations fabricated into application-specific integrated circuits, or as firmware programs loaded into non-volatile storage or software

14X23-322

programs loaded from or into data storage media as machine-readable code, such code being instructions executable by arrays of logic elements such as microprocessors or other digital signal processing units. Thus, the present invention is not intended to be limited to the embodiments shown above but rather is to be accorded the widest scope  
5 consistent with the principles and novel features disclosed in any fashion herein.

**What is claimed is:**

## CLAIMS

1. A system comprising:
  - 2 a resource having a capacity measure; and
  - 4 a plurality of users, each having a usage rate and a set of persistence vectors, wherein a use of the resource by each among the plurality of users is determined at least in part by the usage rate of the user; and
  - 6 wherein when a predetermined relation exists between a sum of the usage rates and the capacity measure, at least one among the plurality of users changes its usage
  - 8 rate according to at least the user's set of persistence vectors.
  
2. The system according to claim 1, wherein each among the plurality of users has a set of available rates, the user's usage rate being a member of the user's set of available rates.
  
3. The system according to claim 2, wherein each element of each vector in the set of persistence vectors of each among the plurality of users corresponds to a member of the user's set of available rates; and
- 4 wherein each element of each vector in the set of persistence vectors of each among the plurality of users indicates a probability that the user's usage rate will
- 6 change to be equal to the corresponding member of the user's set of available rates.
  
4. The system according to claim 3, wherein each vector in the set of persistence vectors of at least one among the plurality of users corresponds to a member of the user's set of available rates.

10X17=170

5. The system according to claim 4, wherein each among the plurality of users  
2 has the same set of available rates.

6. The system according to claim 3, wherein the predetermined relation  
2 between a sum of the usage rates and the capacity measure exists when the sum of the  
usage rates is not less than the capacity measure.

7. The system according to claim 3, wherein each among the plurality of users  
2 has a random number; and

wherein the usage rate of a user is determined at least in part by a  
4 predetermined relation between the random number and at least one of the elements  
of a vector in the user's set of persistence vectors.

8. The system according to claim 3, wherein each among the plurality of users  
2 has a random number; and

when the predetermined relation exists between a sum of the usage rates and  
4 the capacity measure, at least one among the plurality of users changes its usage rate  
according to at least a predetermined relation between the user's random number and  
6 a selected element of a vector in the user's set of persistence vectors,

wherein said selected element corresponds to the user's usage rate.

9. The system according to claim 3, wherein each among the plurality of users  
2 has a random number; and

BX 16 = 208

when the predetermined relation exists between a sum of the usage rates and  
4 the capacity measure, at least one among the plurality of users reduces its usage rate  
according to at least a predetermined relation between the user's random number and  
6 a selected element of a vector in the user's set of persistence vectors,

*wherein said selected element corresponds to the user's usage rate.*

10. The system according to claim 9, wherein the random number of each  
2 among the plurality of users is drawn from a set having a uniform distribution.

11. The system according to claim 9, wherein the predetermined relation  
2 between a sum of the usage rates and the capacity measure exists when the sum of the  
usage rates is not less than the capacity measure.

12. The system according to claim 3, said system further comprising a control  
2 unit, wherein the control unit sends a warning signal to at least one among the  
plurality of users when the predetermined relation exists between a sum of the usage  
4 rates and the capacity measure.

13. The system according to claim 12, wherein each among the plurality of  
2 users has a random number; and

wherein the usage rate of a user is determined at least in part by a  
4 predetermined relation between the random number and at least one of the elements  
of a vector in the user's set of persistence vectors.

$10 \times 17 = 170$

14. The system according to claim 12, wherein each among the plurality of  
2 users comprises a data producer, and each among the usage rates comprises a rate of  
data production.

15. The system according to claim 14, wherein the resource is a wireless  
2 channel for data communications; and

wherein use of the resource comprises transmitting data over the wireless  
4 channel.

16. The system according to claim 15, wherein the resource is the reverse link  
2 of a wireless CDMA channel for data communications.

17. The system according to claim 16, wherein the value of at least one  
2 member of a user's set of available rates is substantially equal to  $19,200 \times 2^i$   
bits/second, wherein  $i$  is an integer.

18. The system according to claim 15, wherein in the set of available rates of  
2 at least one among the plurality of users, the value of at least one member of the set is  
substantially equal to double the value of another member of the set.

19. The system according to claim 15, wherein the usage rate of at least one  
2 among the plurality of users is a null usage rate.

20. The system according to claim 15, wherein an actual use of the resource  
2 by at least one among the plurality of users is not greater than the user's usage rate.

$$15 \times 16 = 240$$

21. The system according to claim 15, wherein the control unit modifies the  
2 set of persistence vectors of at least one among the plurality of users at least  
indirectly.

22. The system according to claim 15, wherein the capacity measure is a  
2 predetermined threshold, said predetermined threshold being lower than an actual  
capacity of the resource.

23. The system according to claim 22, the predetermined threshold being  
2 determined by at least the actual capacity of the resource, a minimum delay between  
sending a warning signal and obtaining a resulting reduction in usage of the resource,  
4 and a maximum increase in resource usage over a period of the minimum delay.

24. The system according to claim 3, wherein each among the plurality of  
2 users has the same set of available rates.

25. The system according to claim 24, wherein each among the plurality of  
2 users has the same set of persistence vectors.

26. The system according to claim 25, wherein the predetermined relation  
2 between a sum of the usage rates and the capacity measure exists when the sum of the  
usage rates is not less than the capacity measure.

12 X 15 = 180

27. The system according to claim 25, wherein each among the plurality of  
2 users has a random number; and

wherein the usage rate of a user is determined at least in part by a  
4 predetermined relation between the random number and at least one of the elements  
of a vector in the user's set of persistence vectors.

28. The system according to claim 25, wherein each among the plurality of  
2 users has a random number; and

when the predetermined relation exists between a sum of the usage rates and  
4 the capacity measure, at least one among the plurality of users reduces its usage rate  
according to at least a predetermined relation between the user's random number and  
6 a selected element of a vector in the user's set of persistence vectors,

wherein said selected element corresponds to the user's usage rate.

29. The system according to claim 28, wherein the predetermined relation  
2 between a sum of the usage rates and the capacity measure exists when the sum of the  
usage rates is not less than the capacity measure.

30. The system according to claim 29, said system further comprising a  
2 control unit, wherein the control unit sends a warning signal to at least one among the  
plurality of users when the predetermined relation exists between a sum of the usage  
4 rates and the capacity measure.

31. The system according to claim 24, said system further comprising a  
2 control unit, wherein the control unit sends a warning signal to at least one among the

13x19=247

plurality of users when the predetermined relation exists between a sum of the usage  
4 rates and the capacity measure.

32. The system according to claim 31, wherein each among the plurality of  
2 users comprises a data producer, and each among the usage rates comprises a rate of  
data production.

33. The system according to claim 32, wherein the resource is a wireless  
2 channel for data communications; and

wherein use of the resource comprises transmitting data over the wireless  
4 channel.

34. The system according to claim 33, wherein the resource is the reverse link  
2 of a wireless CDMA channel for data communications.

35. A method comprising:  
2 using a shared resource at a rate determined at least in part by a first usage  
rate;  
4 receiving a warning signal, said warning signal relating to use of the shared  
resource;  
6 obtaining a random number; and  
using the shared resource at a rate determined at least in part by a second  
8 usage rate, said second usage rate being determined at least in part by comparing the  
random number to at least one element of a persistence vector.

13 X 16 = 208

36. A data storage medium, said medium bearing machine-readable code,  
2 such code being instructions executable by an array of logic elements such as a  
microprocessor or other digital signal processing unit, said instructions defining a  
4 method comprising:

- 6 using a shared resource at a rate determined at least in part by a first usage  
rate;
- 8 receiving a warning signal, said warning signal relating to use of the shared  
resource;
- 10 obtaining a random number; and
- 12 using the shared resource at a rate determined at least in part by a second  
usage rate, said second usage rate being determined at least in part by comparing the  
random number to at least one element of a persistence vector.

DX 10-120

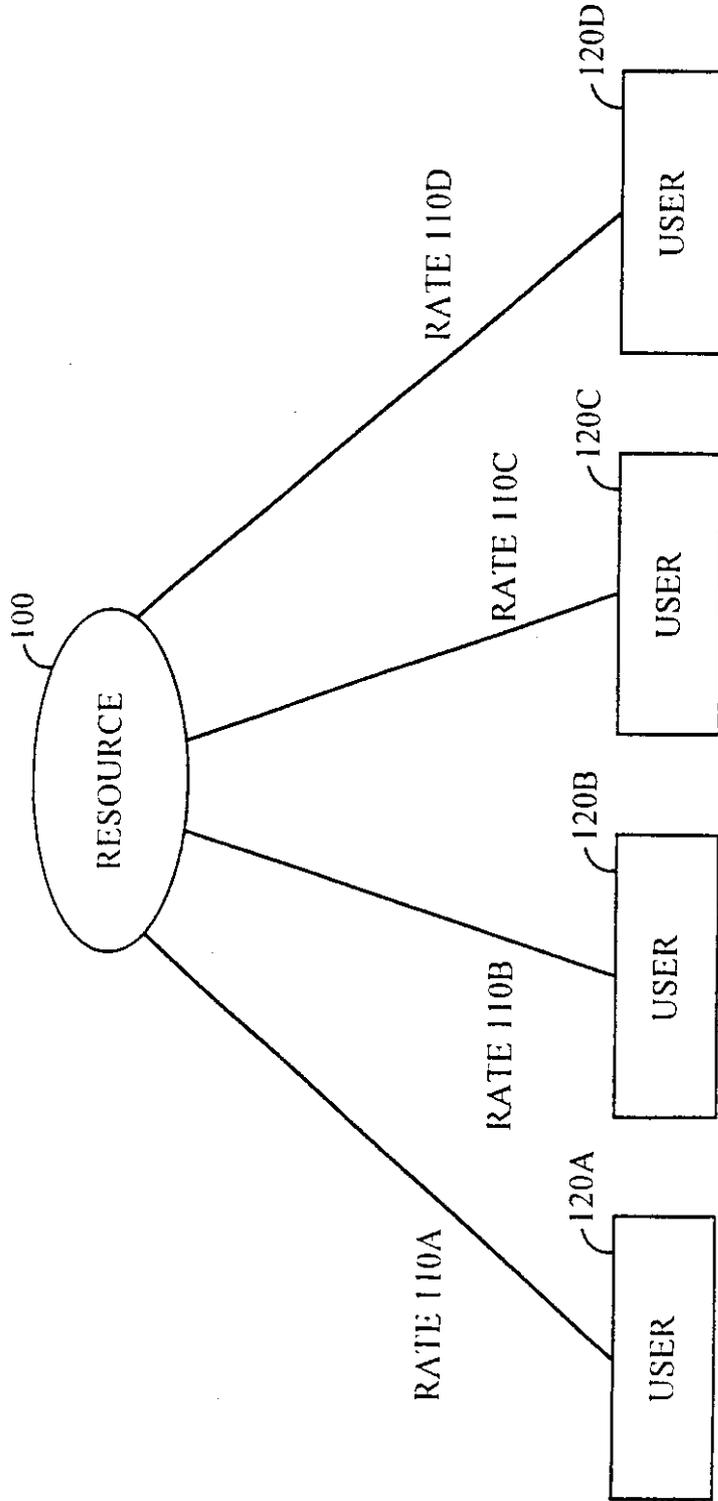


FIG. 1

9

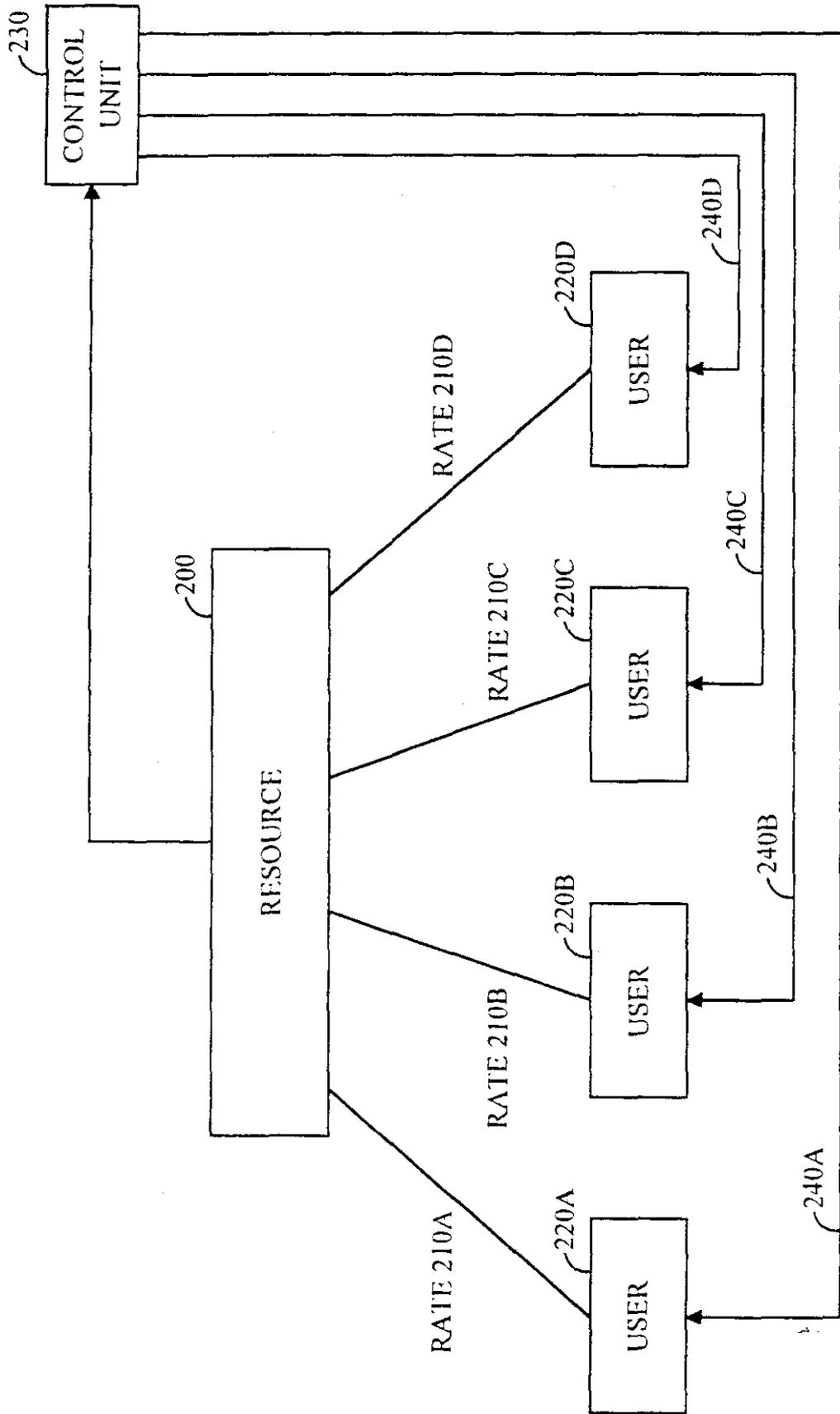


FIG. 2

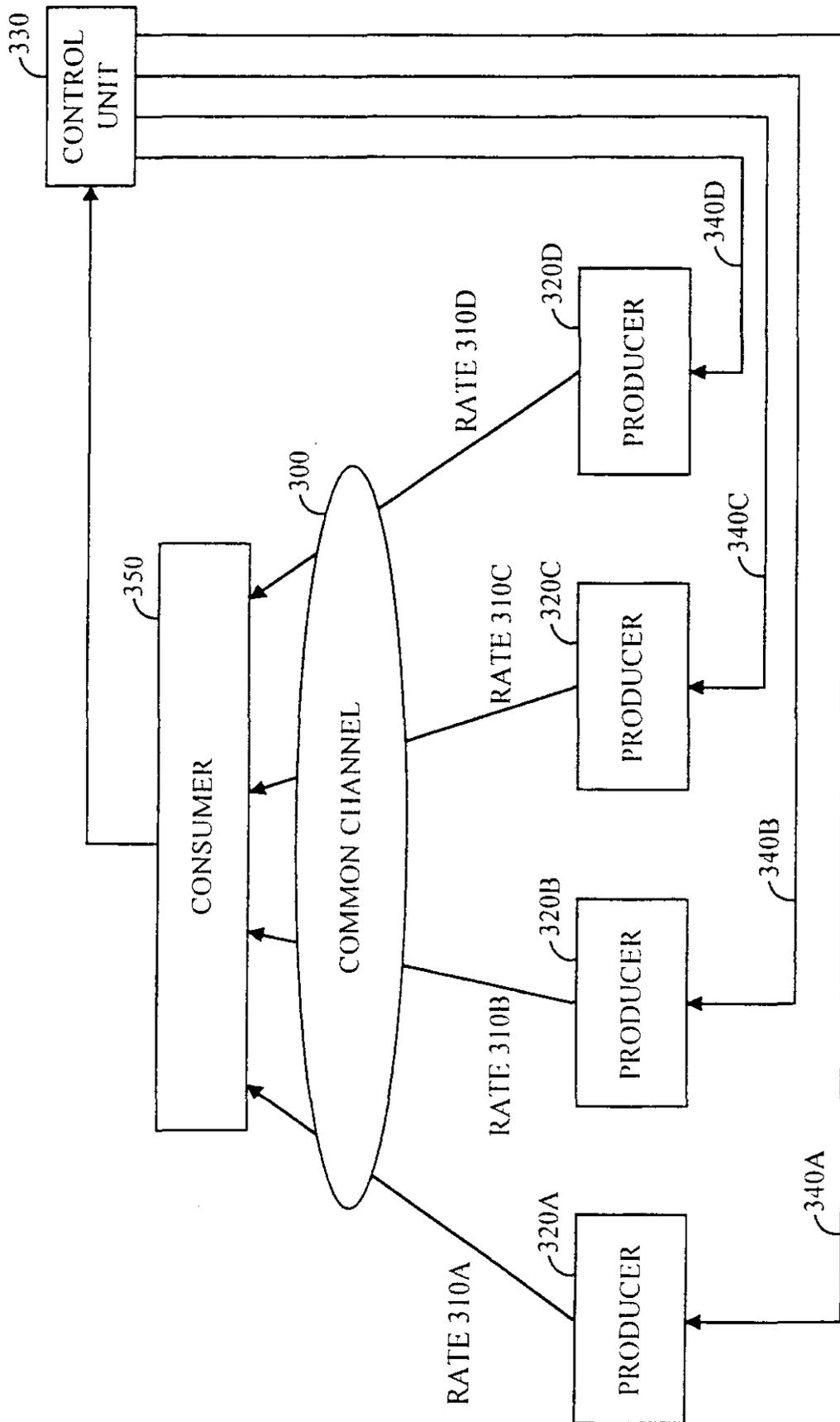


FIG. 3



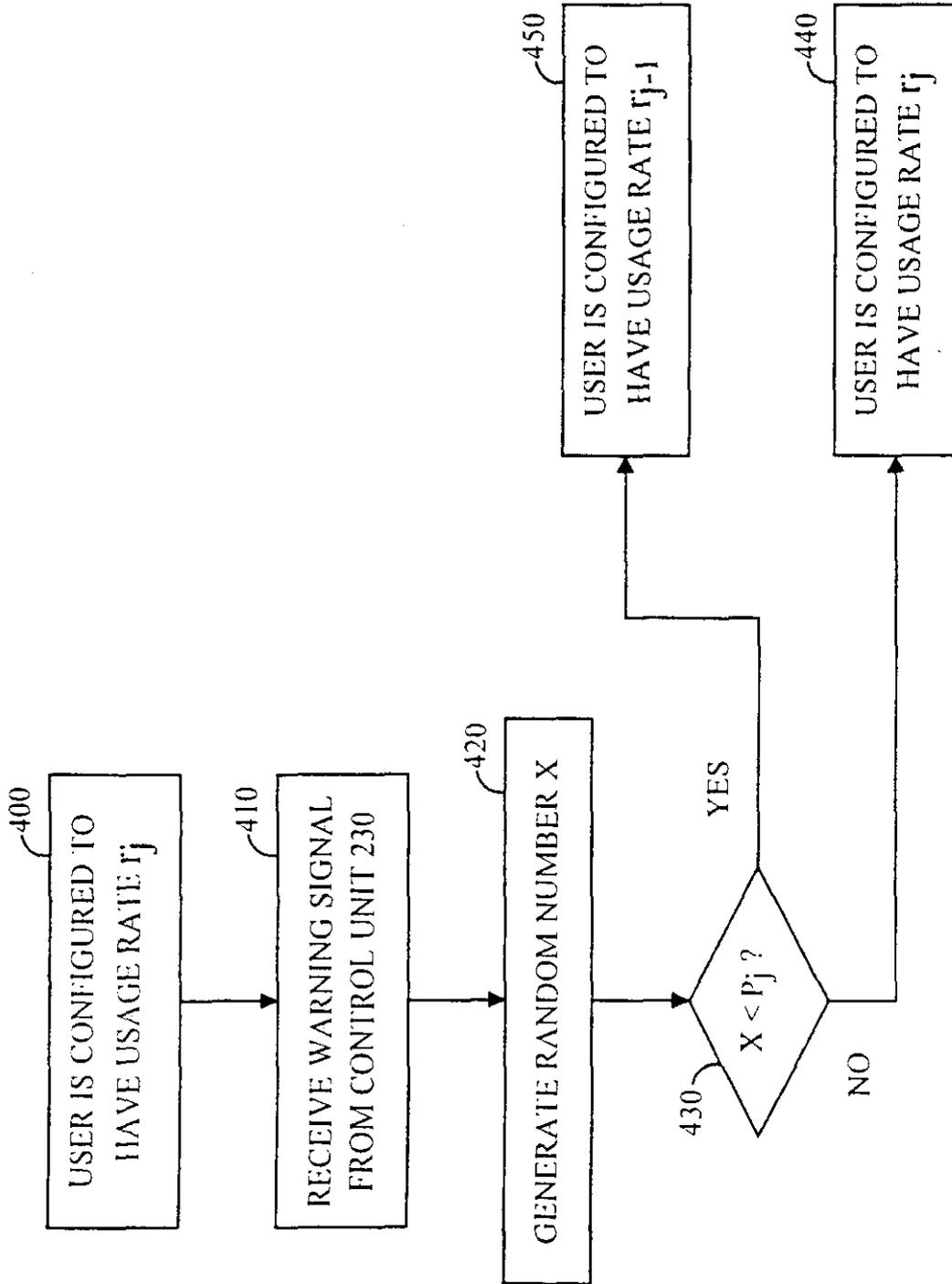


FIG. 4

9

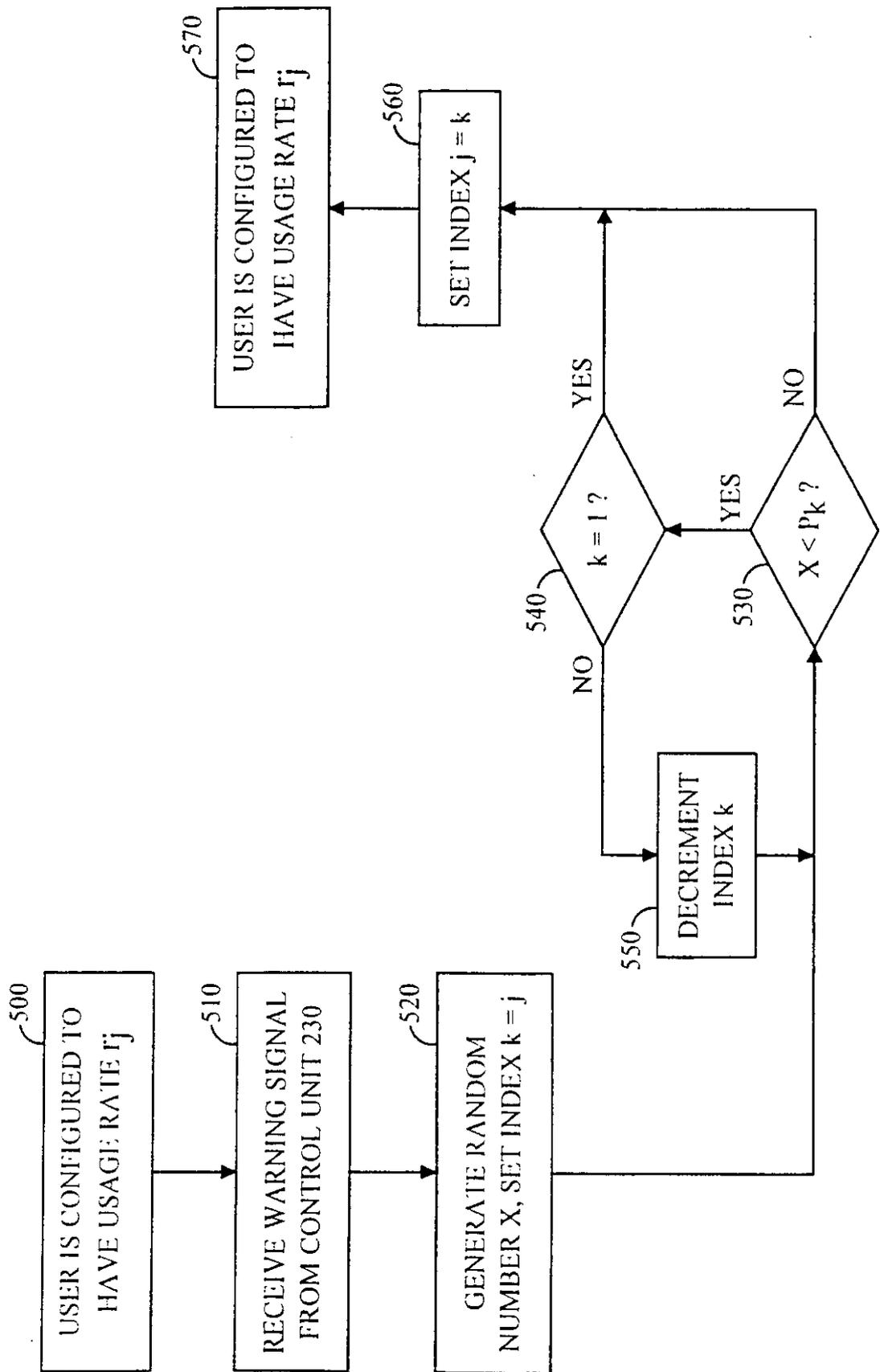


FIG. 5

70

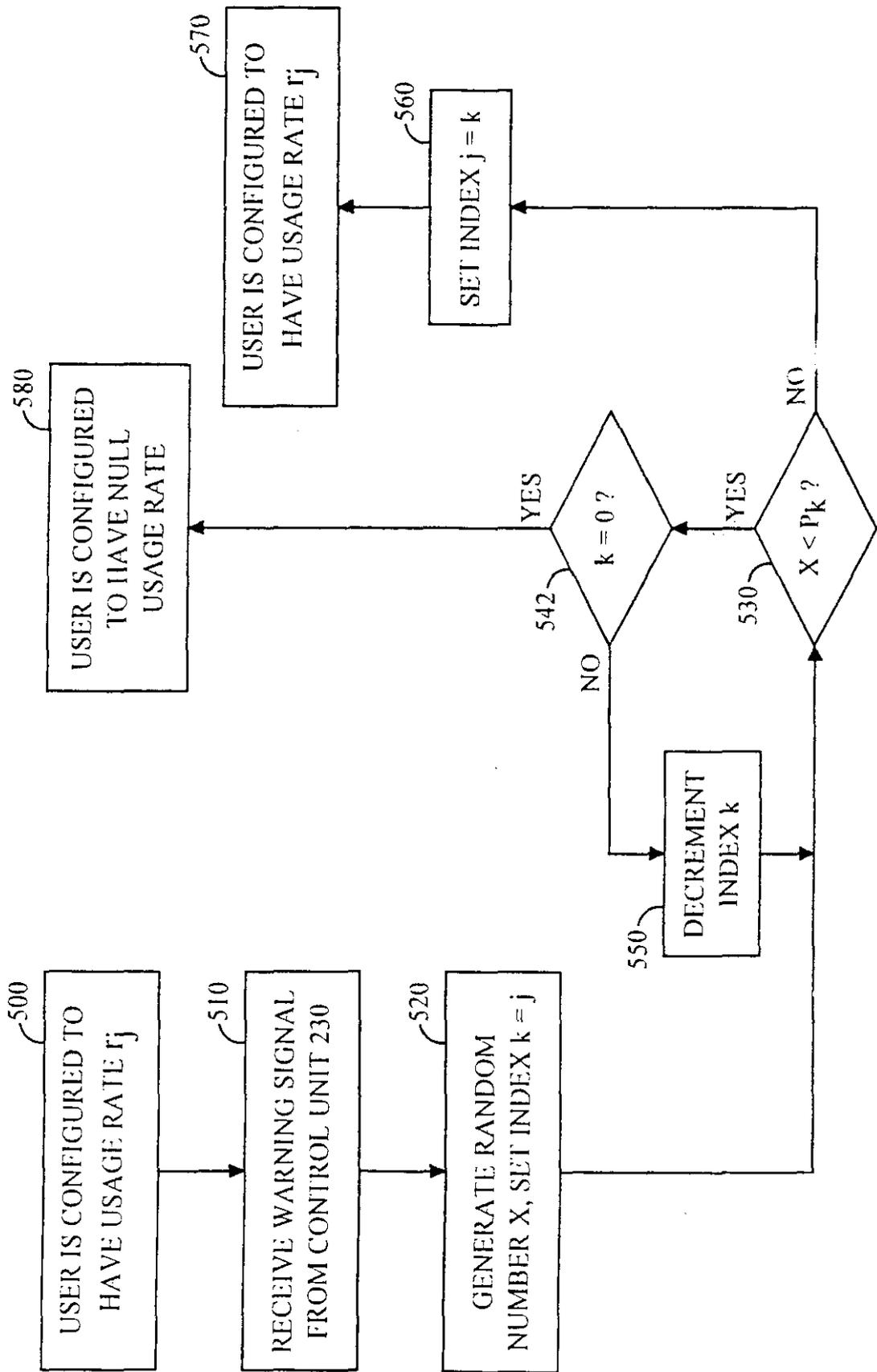


FIG. 6

20

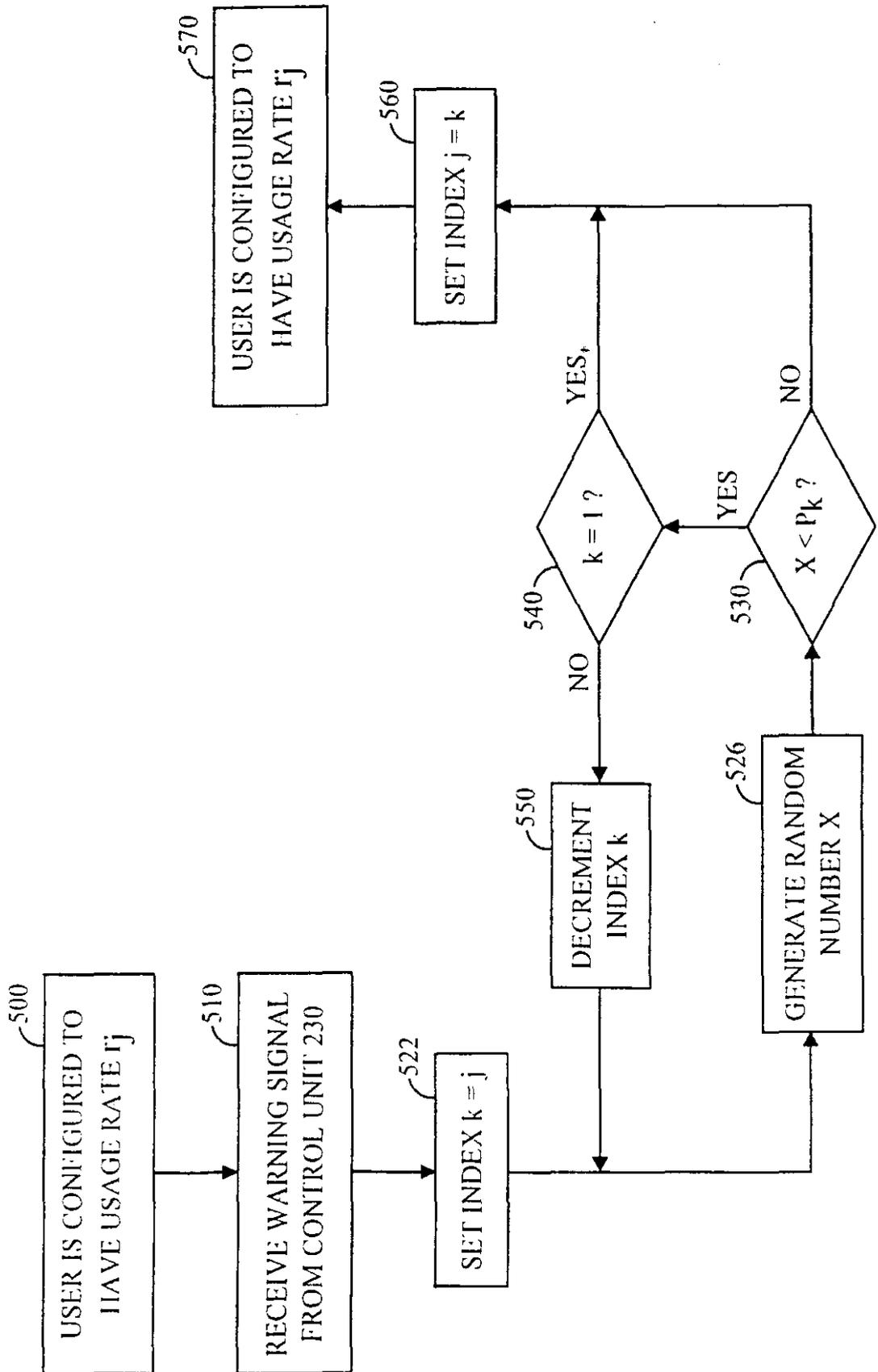


FIG. 7

26

[19] 中华人民共和国国家知识产权局

[51] Int. Cl<sup>7</sup>

G06F 9/46  
H04B 7/26

# [12] 发明专利申请公开说明书

[21] 申请号 00813567.3

[43] 公开日 2002 年 10 月 30 日

[11] 公开号 CN 1377484A

[22] 申请日 2000.9.29 [21] 申请号 00813567.3

[30] 优先权

[32] 1999.9.30 [33] US [31] 09/410,204

[86] 国际申请 PCT/US00/26879 2000.9.29

[87] 国际公布 WO01/24004 英 2001.4.5

[85] 进入国家阶段日期 2002.3.28

[71] 申请人 高通股份有限公司

地址 美国加利福尼亚州

[72] 发明人 R·潘卡 M·S·格罗布

P·E·本德

[74] 专利代理机构 上海专利商标事务所

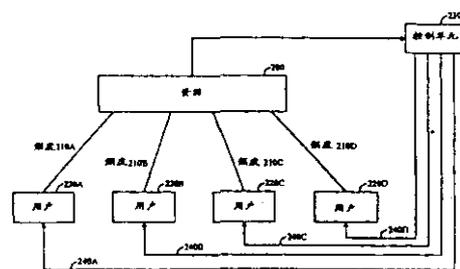
代理人 沈昭坤

权利要求书 4 页 说明书 8 页 附图页数 7 页

[54] 发明名称 基于存留矢量的使用率修正系统和方法

[57] 摘要

当具有有限容量的资源由多个用户共享时,用户的使用率就可能超过资源的容量,因而引起过载情况。在根据本发明的一种系统或方法中,至少一些用户具有存留矢量组。当检测到过载情况时,至少根据用户的存留矢量改变这些用户中至少一个用户的使用率。



# 权 利 要 求 书

---

1、一种系统，其特征在于，包括：

具有容量度量的资源；和

多个用户，每个用户具有一个使用率和一组存留矢量；

其中，所述多个用户中每个用户的资源使用至少部分由所述用户的使用率确定；并且

其中，当在使用率总和和容量度量之间存在预定关系时，所述多个用户中至少有一个用户根据所述用户的存留矢量组改变其使用率。

2、如权利要求 1 所述的系统，其特征在于，所述多个用户中的每个用户具有一组可用率，所述用户的使用率为所述用户的可用率组中的成员。

3、如权利要求 2 所述的系统，其特征在于，所述多个用户中的每个用户的存留矢量组中的每个矢量的每个元素对应于所述用户可用率组的一个成员；并且

其中，所述多个用户中的每个用户的存留矢量组中的每个矢量的每个元素指示了用户使用率将改变为与所述用户可用率组的对应成员相等。

4、如权利要求 3 所述的系统，其特征在于，在所述多个用户中至少一个用户的存留矢量组中的每个矢量对应于所述用户可用率组的一个成员。

5、如权利要求 4 所述的系统，其特征在于，所述多个用户中的每个用户具有相同的可用率组。

6、如权利要求 3 所述的系统，其特征在于，当使用率总和不小于容量度量时，在所述使用率总和和容量度量之间存在预定的关系。

7、如权利要求 3 所述的系统，其特征在于，所述多个用户中的每个用户具有一个随机数；并且

其中，用户的使用率至少部分由在随机数和所述用户存留矢量组中一个矢量的至少一个元素间的预定关系来确定。

8、如权利要求 3 所述的系统，其特征在于，所述多个用户中的每个用户具有一个随机数；并且

当在使用率总和和容量度量间存在预定关系时，所述多个用户中至少一个用户至少根据所述用户随机数和所述用户存留矢量组中所选的矢量元素间的预定关系来改变其使用率，

其中，所述所选元素对应于所述用户的使用率。

9、如权利要求 3 所述的系统，其特征在于，所述多个用户中的每个用户具有一个随机数；并且

当在使用率总和和容量度量间存在预定关系时，所述多个用户中至少一个用户至少根据所述用户随机数和所述用户存留矢量组中所选的矢量元素间的预定关系来减少其使用率，

其中，所述所选元素对应于所述用户的使用率。

10、如权利要求 9 所述的系统，其特征在于，所述多个用户中每个用户的随机数是从具有均匀分布的组中取出。

11、如权利要求 9 所述的系统，其特征在于，当使用率总和不小于容量度量时，在所述使用率总和和容量度量之间存在预定的关系。

12、如权利要求 3 所述的系统，其特征在于，所述系统进一步包括控制单元，其中当在使用率总和和容量度量之间存在预定的关系时，所述控制单元就向所述多个用户中的至少一个用户发送警告信号。

13、如权利要求 12 所述的系统，其特征在于，所述多个用户中的每个用户具有一个随机数；并且

其中，用户的使用率至少部分由在随机数和所述用户存留矢量组中一个矢量的至少一个元素间的预定关系来确定。

14、如权利要求 12 所述的系统，其特征在于，在所述多个用户中的每个用户包括数据生产者，并且使用率中的每个频度都包括数据生产率。

15、如权利要求 14 所述的系统，其特征在于，所述资源是用于数据通信的无线信道；并且

其中，所述资源的使用包括在无线信道上发送数据。

16、如权利要求 15 所述的系统，其特征在于，所述资源是用于数据通信的无线 CDMA 信道的反向链路。

17、如权利要求 16 所述的系统，其特征在于，用户可用率组中至少一个成员的值大致等于  $19,200 \times 2^i$  位/秒，其中  $i$  是整数。

18、如权利要求 15 所述的系统，其特征在于，在所述多个用户中至少一个用户的可用率组中，所述组至少一个成员的值大致等于所述组另一成员值的两倍。

19、如权利要求 15 所述的系统，其特征在于，所述多个用户中至少一个用户的使用率为无效使用率。

20、如权利要求 15 所述的系统，其特征在于，由所述多个用户中至少一个

用户使用的实际资源使用不大于所述用户的使用率。

21、如权利要求 15 所述的系统，其特征在于，所述控制单元至少间接对所述多个用户中至少一个用户的存留矢量组进行修正。

22、如权利要求 15 所述的系统，其特征在于，所述容量度量是预定的阈值，所述预定阈值小于资源的实际容量。

23、如权利要求 22 所述的系统，其特征在于，所述预定阈值至少由资源实际容量、发送警告信号和获得所引起的在资源使用中减少之间的最小延迟，以及在最小延迟周期上资源使用中最大增加来确定。

24、如权利要求 3 所述的系统，其特征在于，所述多个用户中的每个用户具有相同的可用率组。

25、如权利要求 24 所述的系统，其特征在于，所述多个用户中的每个用户具有相同的存留矢量组。

26、如权利要求 25 所述的系统，其特征在于，当使用率总和不小于容量度量时，在所述使用率总和和容量度量之间存在预定的关系。

27、如权利要求 25 所述的系统，其特征在于，所述多个用户中的每个用户具有一个随机数；并且

用户的使用率至少部分由在随机数和所述用户存留矢量组中一个矢量的至少一个元素间的预定关系来确定。

28、如权利要求 25 所述的系统，其特征在于，所述多个用户中的每个用户具有一个随机数；并且

当在使用率总和和容量度量间存在预定关系时，所述多个用户中至少一个用户至少根据所述用户随机数和所述用户存留矢量组中所选的矢量元素间的预定关系来减少其使用率，

其中，所述所选元素对应于所述用户的使用率。

29、如权利要求 28 所述的系统，其特征在于，当使用率总和不小于容量度量时，在所述使用率总和和容量度量之间存在预定的关系。

30、如权利要求 29 所述的系统，其特征在于，所述系统进一步包括控制单元，其中当在使用率总和和容量度量之间存在预定的关系时，所述控制单元就向所述多个用户中的至少一个用户发送警告信号。

31、如权利要求 24 所述的系统，其特征在于，所述系统进一步包括控制单元，其中当在使用率总和和容量度量之间存在预定的关系时，所述控制单元就

向所述多个用户中的至少一个用户发送警告信号。

32、如权利要求 31 所述的系统，其特征在于，在所述多个用户中的每个用户包括数据生产者，并且使用率中的每个频度都包括数据生产率。

33、如权利要求 32 所述的系统，其特征在于，所述资源是用于数据通信的无线信道；并且

其中，所述资源的使用包括在无线信道上发送数据。

34、如权利要求 33 所述的系统，其特征在于，所述资源是用于数据通信的无线 CDMA 信道的反向链路。

35、一种方法，其特征在于，包括：

以至少部分由第一使用率确定的频度使用共享资源；

接收警告信号，所述警告信号涉及共享资源的使用；

获得随机数；并且

以至少部分由第二使用率确定的频度使用共享资源；所述第二使用率至少部分是通过将所述随机数与存留矢量中至少一个元素作比较来确定。

36、一种数据存储媒体，所述媒体承载机器可读代码，这种代码是可以由逻辑阵列元件例如微处理器或其他数字信号处理单元执行的指令，其特征在于，定义方法的所述指令包括：

以至少部分由第一使用率确定的频度使用共享资源；

接收警告信号，所述警告信号涉及共享资源的使用；

获得随机数；并且

以至少部分由第二使用率确定的频度使用共享资源；所述第二使用率至少部分是通过将所述随机数与存留矢量中至少一个元素作比较来确定。

# 说明书

---

## 基于存留矢量的使用率修正系统和方法

### 技术领域

本发明涉及在多个用户之间的有限资源使用的分配。具体说，本发明涉及根据一组持续矢量的使用率修正。

### 背景技术

共享资源是一种可以由多个用户使用的资源。具有有限利用率或容量的共享资源包括例如电站以及其他能源站、供水系统例如水库和流动体，分配货物和/或材料的供给系统，以及数据通信网络和路径的各种实例。因而，与在多个用户间分配共享资源的使用所关联的问题可以在许多不同的范围中产生。无论是在何种范围中，在许多系统中可以找到至少具有下述情况的这种资源：

- 共享资源的容量或利用率可以以每时间度量的有限单位频度 (rate)  $R$  的形式来表示 (例如千瓦/小时、加仑/分钟、箱/星期或位/秒)；
- 在任意特定时间，该资源由  $n$  个不同用户使用，其中  $n$  是非负整数；以及
- 在任意特定时间，第  $i$  位用户 (其中  $1 \leq i \leq n$ ) 的使用率可以用每时间度量的有限单位使用率  $u_i$  来表征。

在图 1 中示出了这种系统的基本模型，其中资源 100 由用户 120a-d 分别以 110a-d 的频度进行使用。依据特定的实行过程，表征共享资源的频度  $R$  可以指示资源的实际或估计容量极限 (例如在通信路径的情况下) 或者，频度  $R$  可以是指示资源最大安全或可容许负载的阈值 (例如在发电设备或装置情况下)。类似地，使用率  $u_i$  可以指示实际使用、预期使用或使用请求或需求。

当  $n$  个使用率  $u_i$  的总和在任意时间超过值  $R$  时，就会引起过载情况。例如，对于电站，当总电流抽取 (current draw) 超出额定容量时，就会产生过载情况。对于数据通信路径，当总数据传输率超过路径实际容量就会引起过载情况，因而破坏了传输中的数据。在某些情况中例如供水系统或材料仓库的情况下，过载情况也可以指示出虽然目前满足了用户的需求，但保留或缓冲容量正在被耗尽。

依据资源特性，过载情况的后果各不相同，可能包括需要一段资源恢复的离

线周期（例如使发电系统冷却或补充水库）或为了能让由于过载而使先前尝试过且失败的使用重新进行，需要对容量进行扩展。该资源甚至会变得临时或永远也不能恢复到其先前的容量。在任何情况下，通常如果可能的话，总是希望避免过载的情况。

### 发明内容

根据本发明实施例的一种系统包括资源和许多资源用户。资源的每个用户具有一个使用率和一组存留矢量，并且资源的用户使用至少部分取决于用户的使用率。当在使用率总数和资源容量的某种度量之间出现预定的关系，那么，至少这些用户中的一个将至少根据其存留矢量组改变其使用率。

### 附图说明

图 1 示出一种具有共享资源的系统的图例。

图 2 示出一种具有共享资源和控制单元的系统的图例。

图 3 示出一种具有一个消费者、多个生产者和一个公共通道的系统的图例。

图 4 示出根据本发明第一实施例的方法。

图 5 示出根据本发明第二实施例的方法。

图 6 示出图 5 方法的变化。

图 7 示出图 5 方法的附加变化。

### 具体实施方式

当在依据图 1 的系统中出现过载情况时，用户 120 可能不会知道发生了过载，特别是如果资源为了满足用户的需要而消耗了保留容量。即使过载情况引起可供用户的资源可用性下降到用户预期或需求之下，用户可能也不能验证这种不足是由于资源过载还是由于供给路径中另一部件的故障而引起的。而且，在某些应用中，例如无线数据通信，可能也不存在用户可以被及时告知过载的反馈机制。因而用户可能会在不知道过载问题的情况下继续使用该资源。在这种情况下，就需要一种包括通过警告信号告知用户过载情况的能力的系统。

图 2 示出这种系统的实例，其中控制单元 230 接收有关由用户 220a-d 使用的资源 200 的信息，并且在各自的通信路径 240a-d 上向用户 220a-d 发送

反馈信息例如警告信号。注意控制单元 230 可以作为资源 200 的一部分或作为用户 220a-d 之一的一部分来实现

如果用户知道了过载情况，那么，就存在由用户驱动的补救的可能性。在这种情况下，如果至少某些用户能彼此进行通信，那么就可以协商例如减少使用率的解决方案。然而，在许多情况下，用户间的这种通信可能难以获得、不切实际或者其它情况下是不希望的，在这种情况下，一种备择的控制机构可提供用于控制资源的使用。这种备择控制机构可以是集中式和/或非集中式的。

如果可以获得对用户将来使用需求的完整了解，那么，在理论上就可能构建一种优化的使用调度，这种调度将尽可能满足用户的需求，而又完全避免所有过载情况。然而，在许多实际系统中，即使对于用户本身也不知道他将来的需要。一种在这种系统中避免过载情况的途径是依据当前的使用需求：例如，通过仅以请求为基础给予用户使用率分配。然而，为了能将来自用户的使用请求传回控制单元，这种方案将需要向上通信路径，而这种路径在别的方案中并不需要。而且，在接收、处理和响应这种请求中会引起额外的代价和延迟。

为了避免请求/给予方案的一些缺点，可能设计一种非集中式系统，在该系统中由用户共享控制。在这种系统中的控制单元集中精力对过载情况进行预测和避免过载情况，而产生足够的反馈信息以允许用户将他们自己的使用控制在某种程度。

根据本发明实施例的方法可以在符合图 1 模式的任意系统中实现，其中用户可以获得过载情况的通知（如图 2 的经修改系统）。在图 3 中示出这种系统的示范应用，其中用户 320a-d 是数据生产者、资源 300 是将生产者与数据消费者 350 链接的公共传输信道，并且控制单元 330 从消费者接收使用信息。生产者通过以速率 310a-d 或低于它们的速率分别向消费者 350 发送数据来使用资源 300，并且它们从控制单元接收各自的信号 340a-d（这些信号可以包括反馈和/或其他控制信息）。

一种可能的示范应用实现是 CDMA 电信系统的反向链路。在这种情况下，每个生产者可以包括 1) 发射机，例如移动电话或 WLL（无线本地环路）站，连接到 2) 数据产生装置，例如膝上电脑或销售点终端，通过 PCMCIA 卡或类似接口，并输出以 IP 或任何其他适合的协议封装在包中的数据。消费者 350 和

控制单元 330 可以是基站的一部分，并且控制信号 340 可以在前向链路上上传送。已经实现了几代 CDMA 电信系统和其几种版本。然而，大多数这些 CDMA 系统已经设计用于传送数字化语音通信，在此所述的该方法特别适用于为具有宽变化传输率的生产者服务的网络，例如纯数据网络或语音-数据混合网络。

参照图 2 的系统，在图 4 中描述了根据本发明第一实施例的方法。在该方法中，在任意给定时间用户的资源使用由有关预定使用率确定。如框 400 所示，特定用户配置为具有使用率  $r_j$ 。使用率  $r_j$  是一组  $m$  个预定可用率  $r_1$  到  $r_m$  之一，其中关系  $a < b$  表示  $r_a < r_b$ 。对于用户并不需要具有相同的可用率组，但用于每个用户的组应该为控制单元 230 所知，以便其能可靠预测资源使用的状态，并适当地发出警告信号。对于每个用户可用率组也可能由控制单元 230 进行周期性或其他形式的更新。频度选择、指定和分配的方案，该方案可以用于结合本发明实施例的系统中，包括在共同待批专利申请号 09/264,297 名为“METHOD OF RATE ALLOCATION IN A DATA COMMUNICATIONS NETWORK”（1999.4 申请，并已转让给本发明的受让人）以及 09/XXX,XXX 名为“METHOD AND APPARATUS FOR PERSISTENCE-VECTOR-BASED RATE ASSIGNMENT”（在此共同申请，已已转让给本发明的受让人，并引入该申请所揭示的内容作为参考）中所描述的那些内容。

注意使用率  $r_j$  可以指示最大可用率，即允许以给定频度使用资源，而不是请求。除使用率外，用户使用资源的实际频度可以依赖其他因素，例如用户的当前需要和/或使用资源的能力。同样，注意用户使用资源的实际频度不需要是可用率组中的一员。

在一个特定实施例中，每个用户具有系统固定的可用率组，其中每个频度表示为千位每秒 (Kb/s) 并且将频度组设计为以 2 的幂递增。因为频度的加倍需要功率的加倍以保持相同的每位的能量与噪声功率频谱密度的比值 ( $E_b/N_0$ )，这样，每个频度阶跃对应于 3dB 的功率跃阶。在本实例中的可用率值包括 4.8、9.6、19.2、38.4、76.8、153.6 以及 307.2Kb/s。

除使用率外，每个用户还具有存留矢量组，尽管在系统也可以有缺少存留矢量组的其他用户。每个这种矢量的长度可以是大于 0 的任意整数，并且代表了一种概率，就是使用率将是可用率组中对应的一个。在示范应用中，每个矢量元素是代表从 0 到 1 概率的连续值。存留矢量组可以对每个用户唯

一，或可以为特定一类中的所有用户分配同一组，或为系统中的所有用户分配同一组。同样，存留矢量组可以是用户操作的永久特征，或它可以由控制单元 230 产生，在这种情况下它可以进行周期性或以其他方式更新。存留矢量分配和使用的其他相关方面在共同待批的申请号 09/XXX,XXX 名为“METHOD AND APPARATUS FOR PERSISTENCE-VECTOR-BASED RATE ASSIGNMENT”（引入该申请所揭示的内容作为参考）中有讨论。

在这种方法中，用户的存留矢量组包括一个具有  $(m-1)$  个元素的矢量  $P$ ，其中  $P = \{P_k, \text{因而 } 1 \leq k \leq m-1\}$ ，而  $m$  是用户可用率组成员的数目。（矢量  $P$  可以是存留矢量组中唯一的矢量，或可以根据如最近使用率或最近该用户的实际频度为标准从组中的其他矢量中选择矢量  $P$ 。）矢量  $P$  可以（但不一定要）具有概率密度函数形式，其中其元素（或由其元素代表的值）的总和等于或实质上等于 1。

在框 410 中，用户从控制单元 230 接收警告信号。这种警告信号会在，例如，当检测到实际或即将发生过载情况时产生，并且它会被发送给所有的用户或仅发送给用户的子集（例如仅发送给具有存留矢量的用户）。一种系统的各种实施例和应用，其中警告信号由反向链路信号中的忙碌位表示，在共同待批的申请号 09/346,882 名为“METHOD AND APPARATUS FOR SIGNAL COMBINING IN A HIGH DATA RATE COMMUNICATION SYSTEM”（1999.7.2 申请，并已转让给本发明的受让人）中有描述。

依据接收警告信号，如框 420 所示，用户产生随机数  $x$ 。 $x$  的范围和分布仅由特定的实现来限定；在示范应用中， $x$  代表从范围在 0 到 1 上具有均匀分布的组中得出的值。在框 430 中， $x$  的值以连续值  $P_j$  为对照进行测试，其中  $P_j$  是对应于使用率  $r_j$  的存留矢量的元素。如果测试失败（即  $x$  不小于  $P_j$ ），那么如框 440 所示，用户的使用率不受过载情况的影响。如果测试成功（即  $x$  小于  $P_j$ ），那么如框 450 所示，用户的使用率从  $r_j$  减少为  $r_{j-1}$ 。如果用户的使用率已经是用户可用率组中最小频度，那么框 450 中的成功可能指示将预定的较低频度减少，或甚至是服务拒绝。该方法也可以进行修改以允许依据为  $x$  和  $P_j$  所选值的特定特征，使用  $x$  和  $P_j$  间的许多其他关系中的一种代替如框 430 所示的测试情况。

注意给予存留矢量  $P$  的元素的值将部分影响如何使资源使用的重新分配在以不同使用率开始的用户中进行偏置。例如，更加合理的重新分配可以通

过为对应于高使用率的存留矢量  $P$  元素选择较大值，而为对应于低使用率的  $P$  元素选择较小值来实现。这种方案使得当前具有高使用率的用户更加可能会减少其使用率，而使得已经具有低使用率的用户更加不可能进一步减少其使用率。注意在每个存留矢量与可用率组的特定成员关联的情况下，这些矢量之间的关系也将偏置资源使用的重新分配。也要注意上述频度加倍方案的使用（或在使用率组中类似的非恒定分配）将使得高使用率的用户使用率减少以释放超过低使用率用户使用率减少的更多资源容量。

上述方法的许多变化可以用于本实施例的应用中。例如，用户可以共享同一存留矢量组，或可以分配不同的存留矢量组以允许在用户中实现优先权方案。在另一变化中，每个存留矢量的第一个元素可以消除（或设定为代表概率为 1），以便已经具有最小使用率的用户将不会进一步遭减少。同样，在存留矢量的第一元素中有多于 1 个的元素也会这样进行处理以保护其他低使用率用户。

对使用率的附加限制可以作为特定实施例的其他问题方面的后果存在。例如，用户实际使用或访问共享资源的频度可以由例如用户目前容量和概率的因素来限制。因而，用户就可能使用或可能允许使用低于由本方法或类似方法给予的使用率的使用率。

可能希望选择频度  $R$ （共享资源的容量度量）作为阈值，而不是共享资源的实际容量，以便在发生过载情况之前产生警告信号，因而允许系统反应以避免这种情况。这种情况下，阈值  $R$  的选择至少要考虑到（1）系统响应中的最长可能延迟，如由警告信号生成和在总资源使用中跟着发生的减少之间的最大时间所表征，以及（2）在这种延迟周期期间资源使用中的最大可能增量。

参照图 2，在图 5 中描述了一种根据本发明第二实施例的方法。相对于上述方法，该方法允许用户使用率减少到可用率组中任意其他的频度，而不是仅是某个特定频度。如在上述方法中，配置用户具有来自可用率  $r_1$  到  $r_m$  用户组中的使用率  $r_j$ （如框 500 所述），并且具有  $(m-1)$  个元素的存留矢量  $P$ ，该矢量可以根据例如下标  $j$  从组中进行选择。在框 510 中，从控制单元 230 接收警告信号，并且在框 520 中，用户如上所述产生随机数  $x$ 。在此阶段，用户也将下标  $k$  设定为与下标  $j$  相等。

在框 530 中， $x$  的值与连续值为对照进行测试，其中  $P_k$  是对应于使用率  $u_k$  的存留矢量  $P$  的元素。如果测试失败（即  $x$  不小于  $P_k$ ），那么在框 560 中下

标  $j$  设定等于  $k$ ，并且通过给用户配置具有使用率  $r_j$ ，本方法在框 570 结束。在这种情况下，换句话说，用户使用率没有受过载情况的影响。

如果在框 530 中的测试成功（即  $x$  小于  $P_k$ ），那么就测试下标  $k$  的值，如果  $k$  已经是其最小值（即在本实施例中为 1），那么该过程在上述框 560 和 570 中继续。否则，就将  $k$  的值递减（即减小 1），并且重复测试。在这种方法下，当最终到达框 570，用户可以配置为具有组中任意的使用率，该使用率等于或小于在框 500 中示出的使用率。对该方法再次进行改变以允许依据为  $x$  和  $P_k$  所选值的特定特征，使用  $x$  和  $P_k$  值之间的许多其他关系中的一种代替如框 530 所示的测试条件。

在如图 6 所示的本方法的变化中，可能拒绝用户对共享资源的使用。框 540 由框 542 代替，在框 542 中允许下标  $k$  达到值零。当这种情况发生时，在框 580 中用户就配置为具有无效使用率。该无效使用率可以代表在可用率组外的某些预定频度（例如从保留容量中得到的最小频度）或它可以代表零使用率，因而是完全的使用拒绝。图 7 示出图 5 方法的附加变化，其中在每个循环迭代的框 526 中产生一个新的随机数  $x$ （在本变化中，如在框 522 中，可以减少框 520 以仅包括下标  $k$  的初始值）。

相对于在图 4-7 中所示出的方法，注意所选使用率的最小限度可以通过将对应于该频度的存留矢量元素设定为任意较小的频度以指示概率为 1 来建立（即在图 4-7 实例中设定这些元素为 0）。在这种情况下，当达到该频度时（或当由已经具有较低使用率的用户调用该过程时）在框 430 和 530 中的测试将失败，并且不会发生使用率中的进一步减少。

前面所提供的对较佳实施例的描述使得本领域的熟练技术人员能够制造或使用本发明。可以对这些实施例进行各种变化，并且在此所出现的基本原理可以也应用于其他实施例。例如，不是从 1 开始，下标例如可用率组的这些下标以及参照存留矢量可以以零或任意其他数字或符号开始。同样，在可用率组中，关系  $a < b$  可以表示  $r_a > r_b$ ，或可以以某种其他顺序对各种频度进行分配来代替。

另外，本发明可以作为硬件电路、构成专用集成电路的电路配置或加载在非易失存储器中的固件程序或从数据存储媒体中加载或载入其的软件程序，例如机器可读代码（这种作为指令的代码可以由逻辑阵列元件例如微处理器或其他数字信号处理单元执行）部分或完全实现。这样，本发明并不是

要限制在所示的实施例中，而是根据与在此以任意形式揭示的原理和新颖特征相符合的最大范畴。

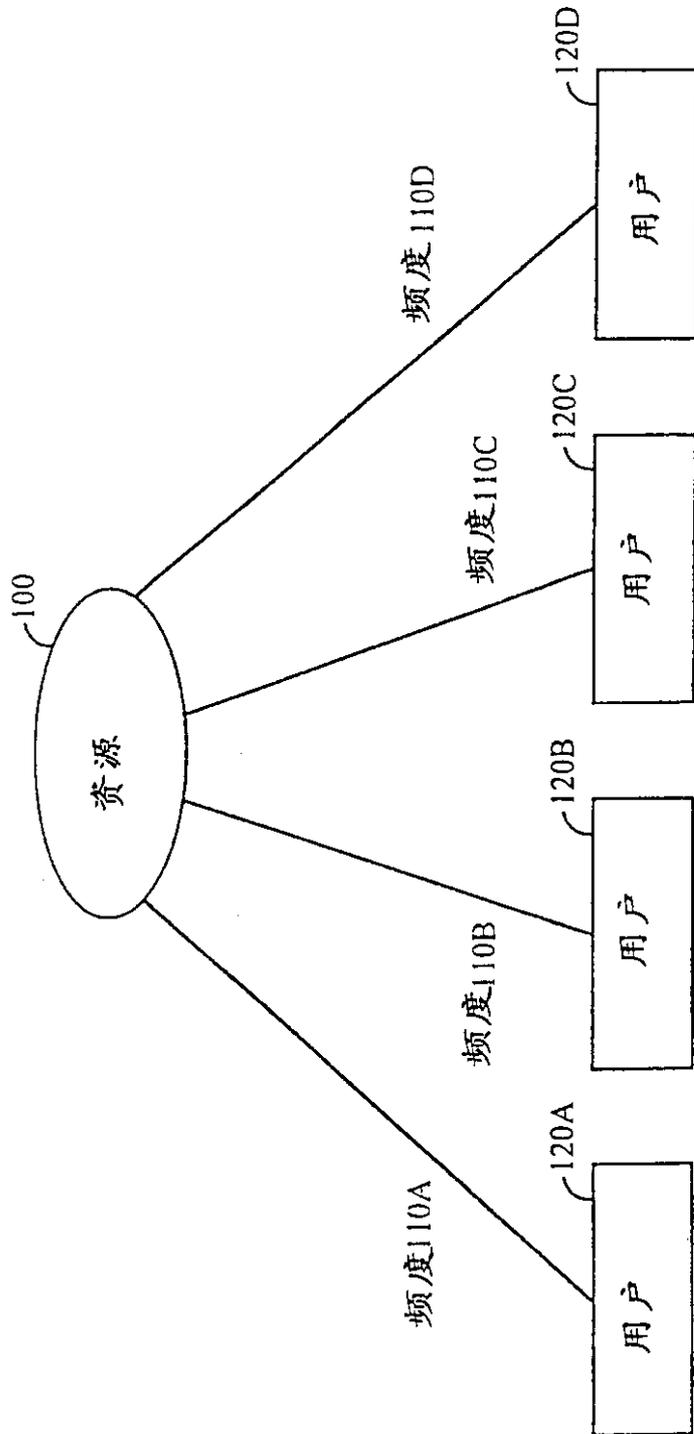
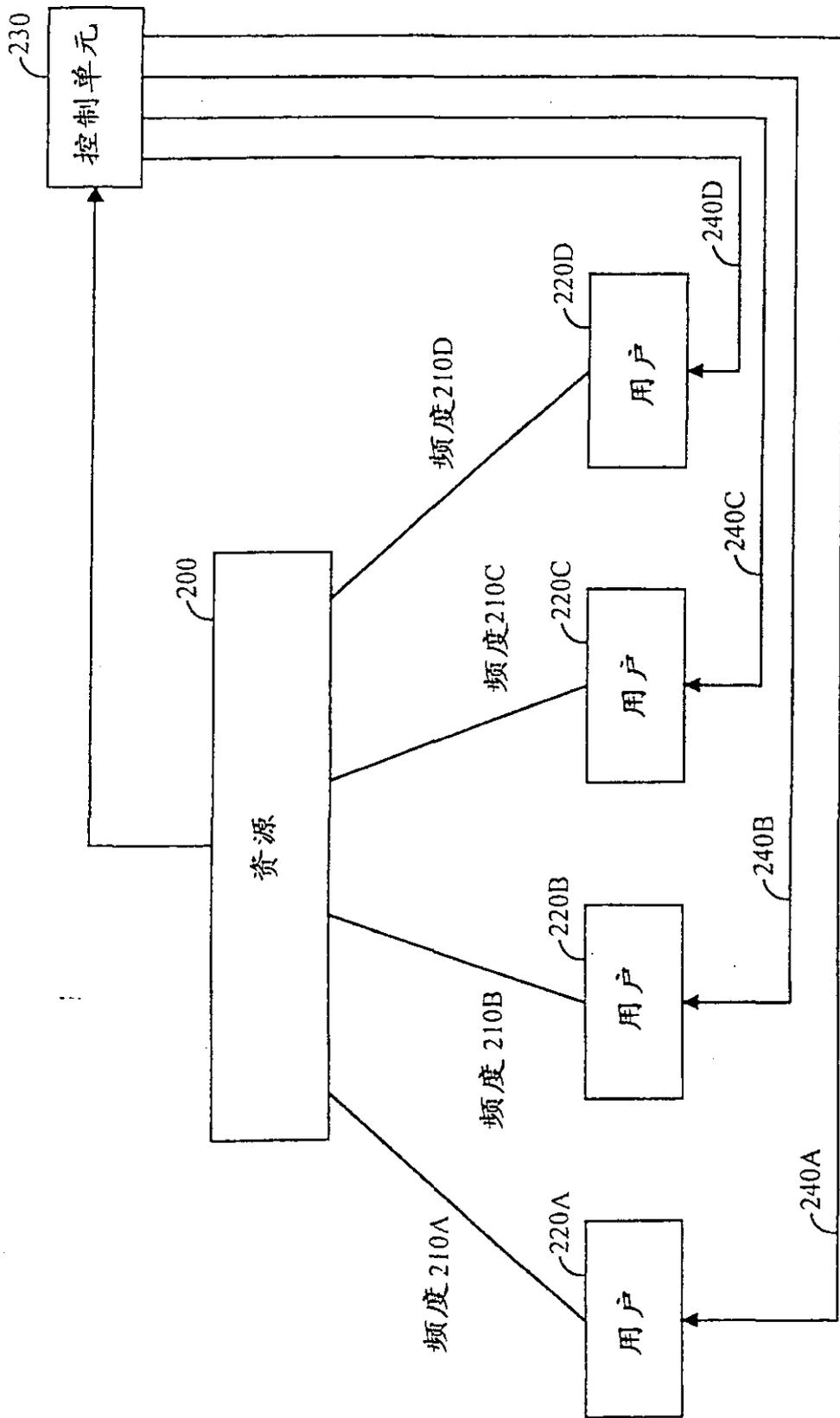


图 1



图

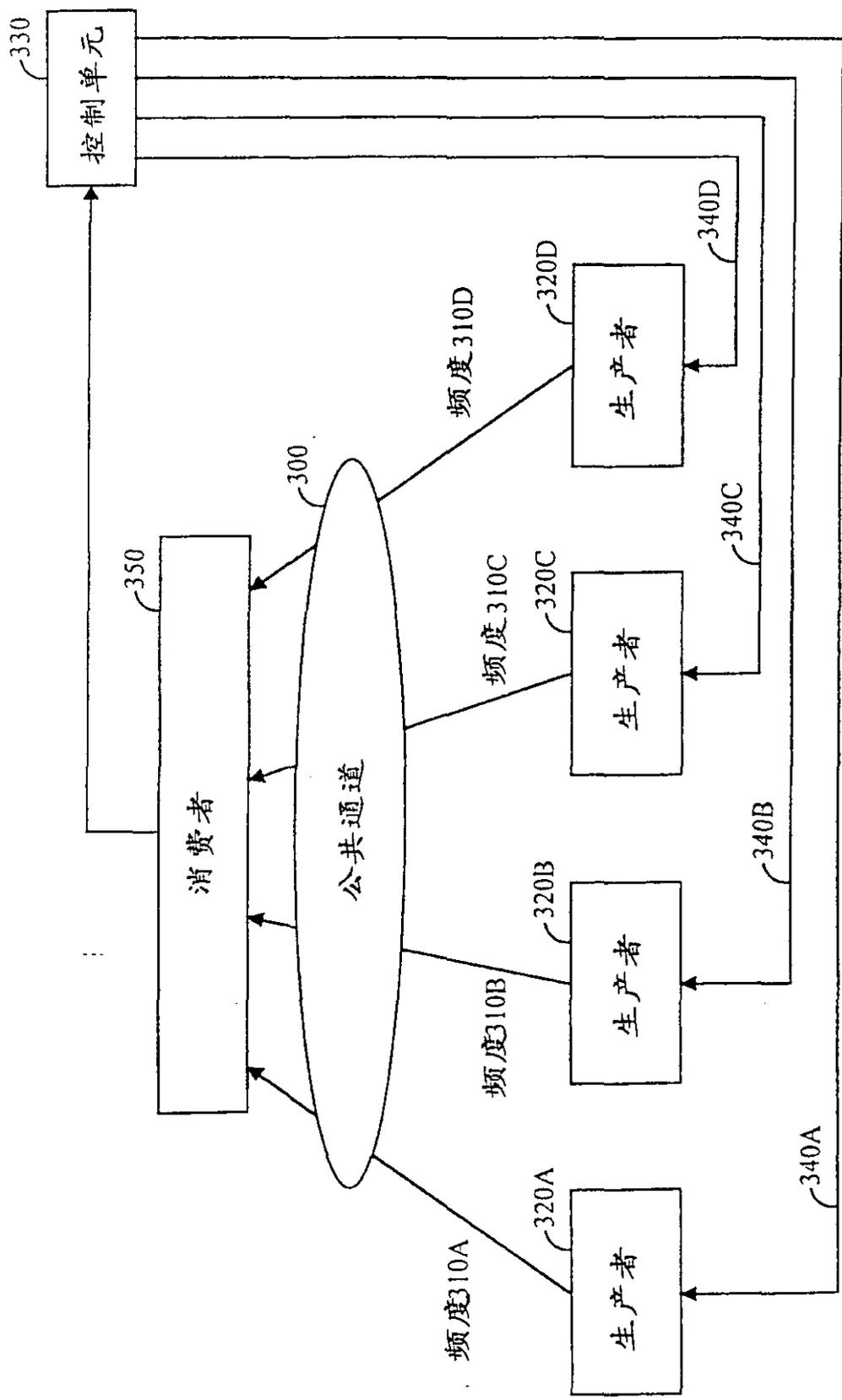


图 3

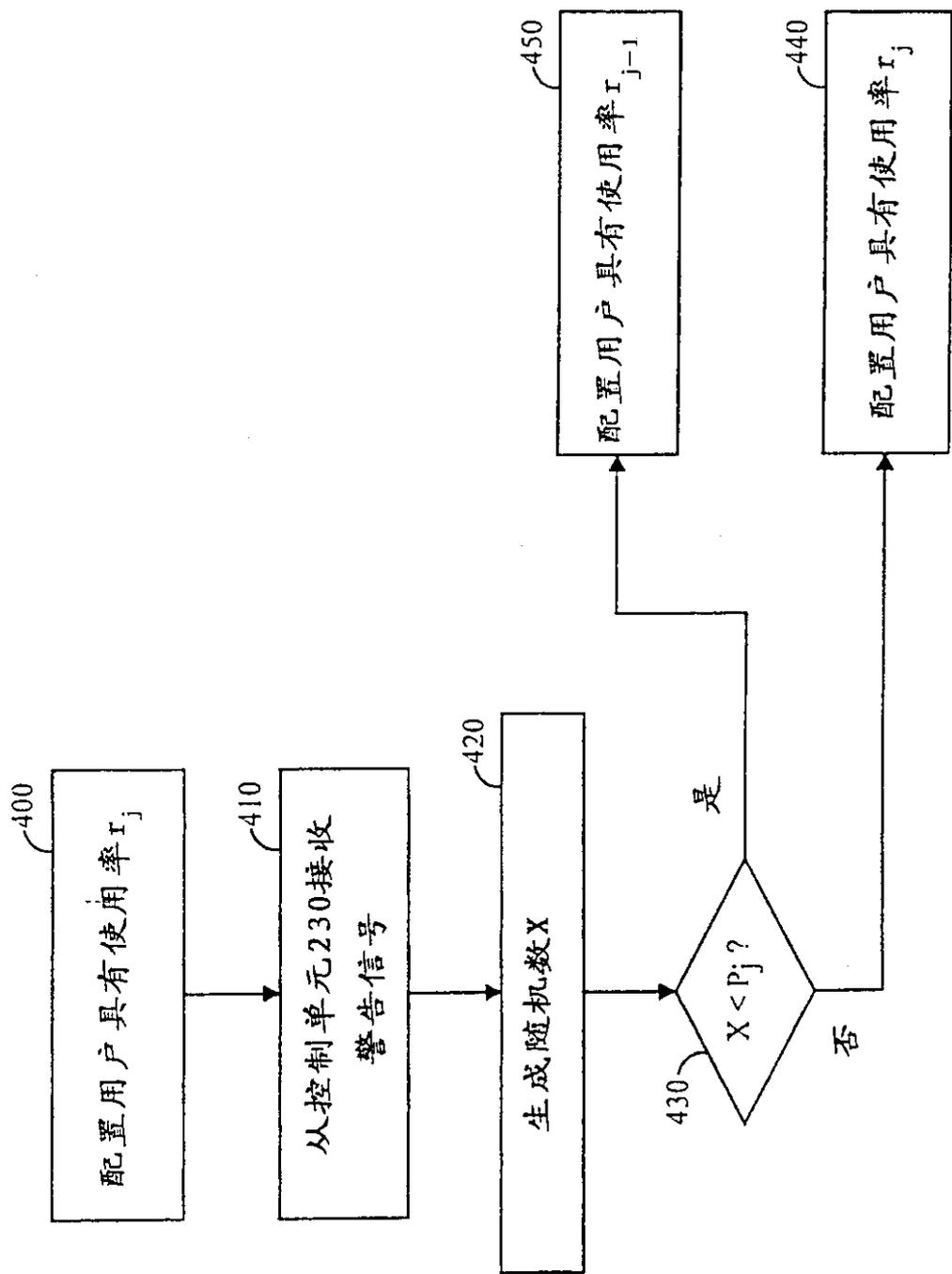


图 4

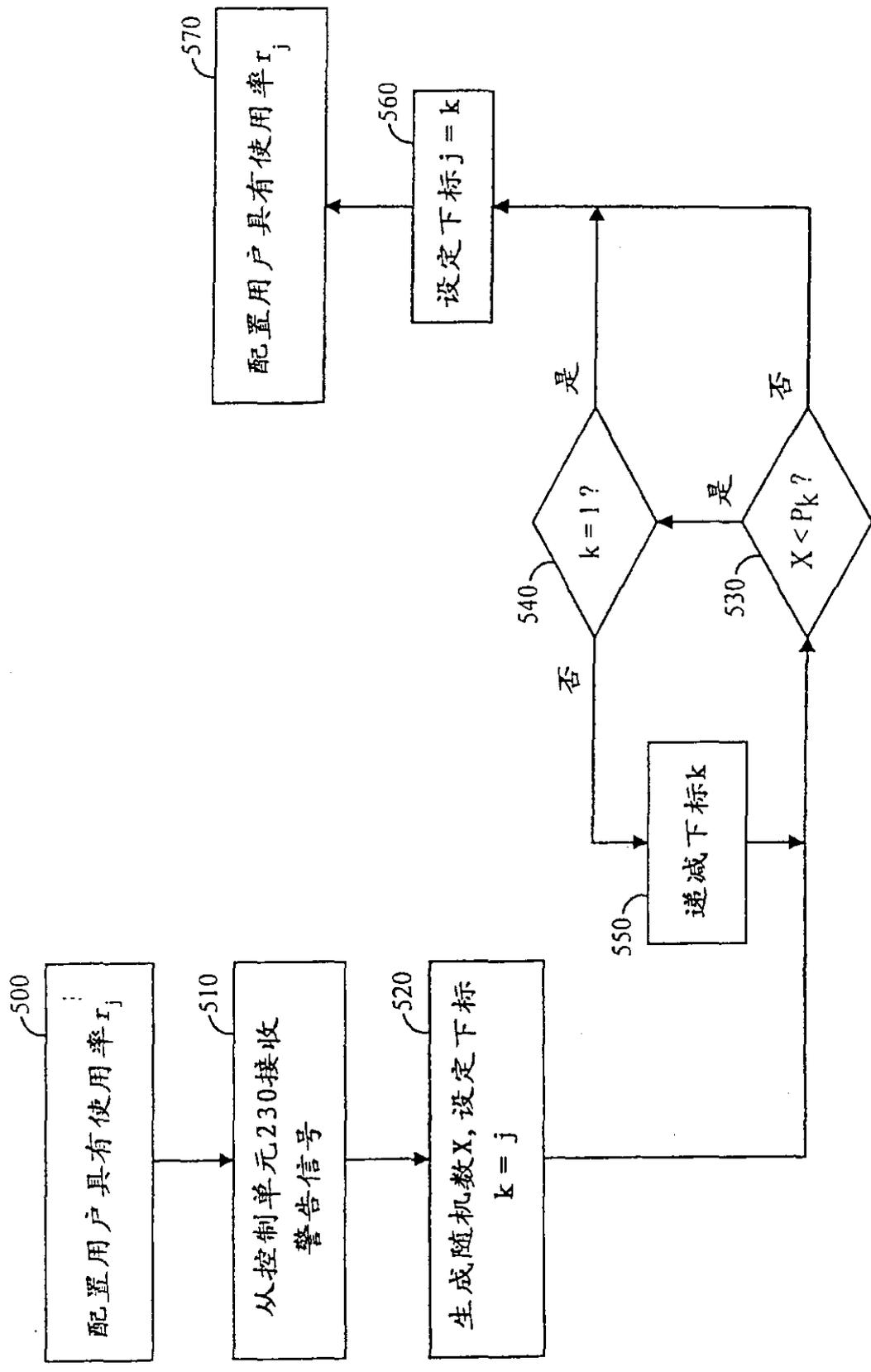


图 5

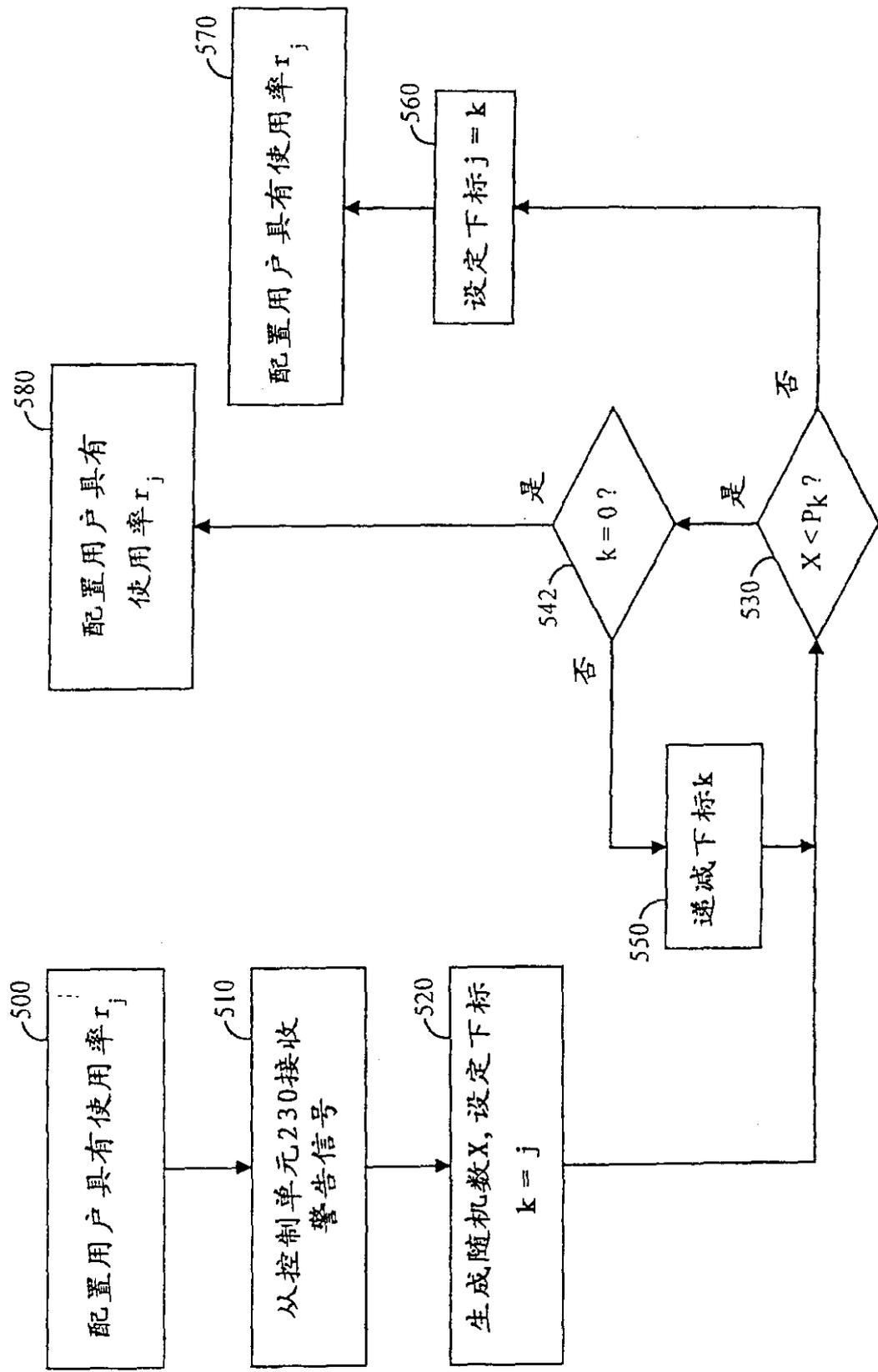


图 6

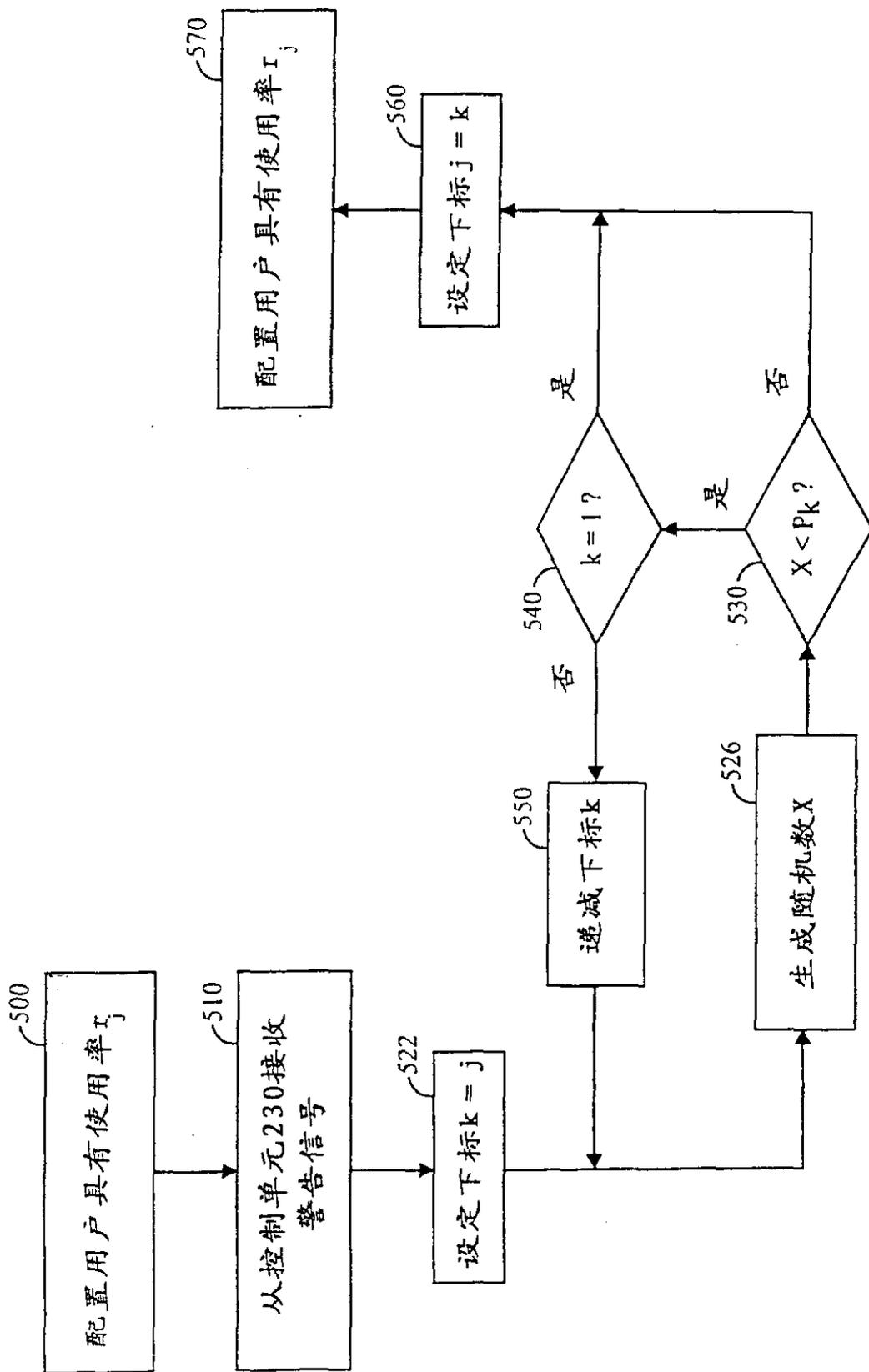


图 7