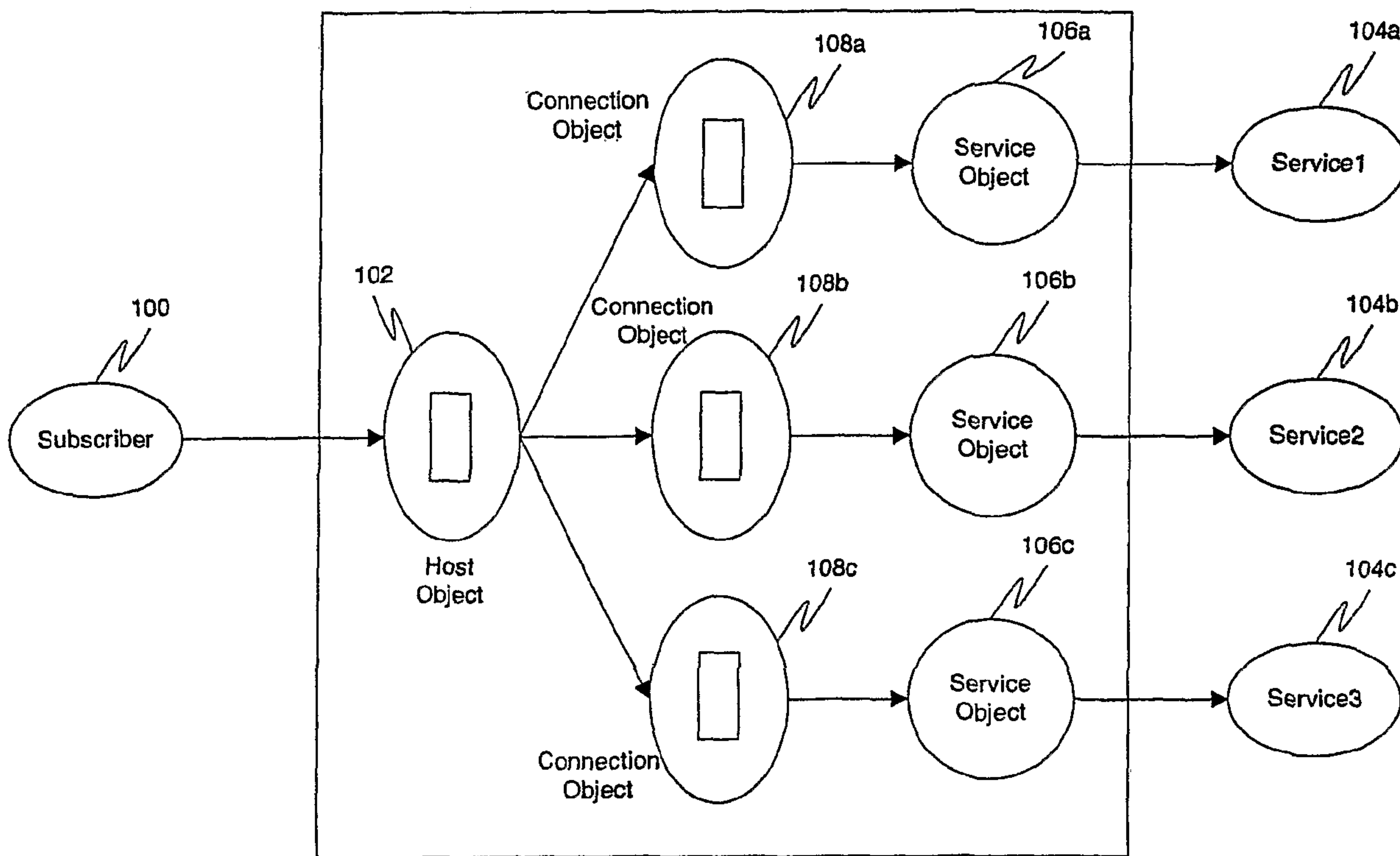




(86) Date de dépôt PCT/PCT Filing Date: 2003/09/23
(87) Date publication PCT/PCT Publication Date: 2004/04/08
(45) Date de délivrance/Issue Date: 2010/07/27
(85) Entrée phase nationale/National Entry: 2005/03/24
(86) N° demande PCT/PCT Application No.: US 2003/030308
(87) N° publication PCT/PCT Publication No.: 2004/030279
(30) Priorité/Priority: 2002/09/26 (US10/256,787)

(51) Cl.Int./Int.Cl. *H04L 12/24* (2006.01),
H04L 12/14 (2006.01), *H04L 12/56* (2006.01),
H04L 29/06 (2006.01)
(72) Inventeurs/Inventors:
ZHANG, SHUJIN, US;
JIANG, TIANJI, US
(73) Propriétaire/Owner:
CISCO TECHNOLOGY, INC., US
(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : APPROVISIONNEMENT DE TRAFIC PAR SERVICE ET PAR PERSONNE
(54) Title: PER USER PER SERVICE TRAFFIC PROVISIONING



(57) Abrégé/Abstract:

A host object (102) representing a user, a service object (106a, 106b, 106c) representing a service, and a connection object (108a, 108b, 108c) linking the two may be utilized in order to provide Quality of Service (QoS). The host object (102) and/or connection object (108a, 108b, 108c) may contain a provisioning scheme defining a limit on traffic, which can be provided through either a local or remote programming mechanism. When traffic flows through the host object (102) and/or connection object (108a, 108b, 108c), the appropriate provisioning scheme is utilized to determine which packets to discard.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
8 April 2004 (08.04.2004)

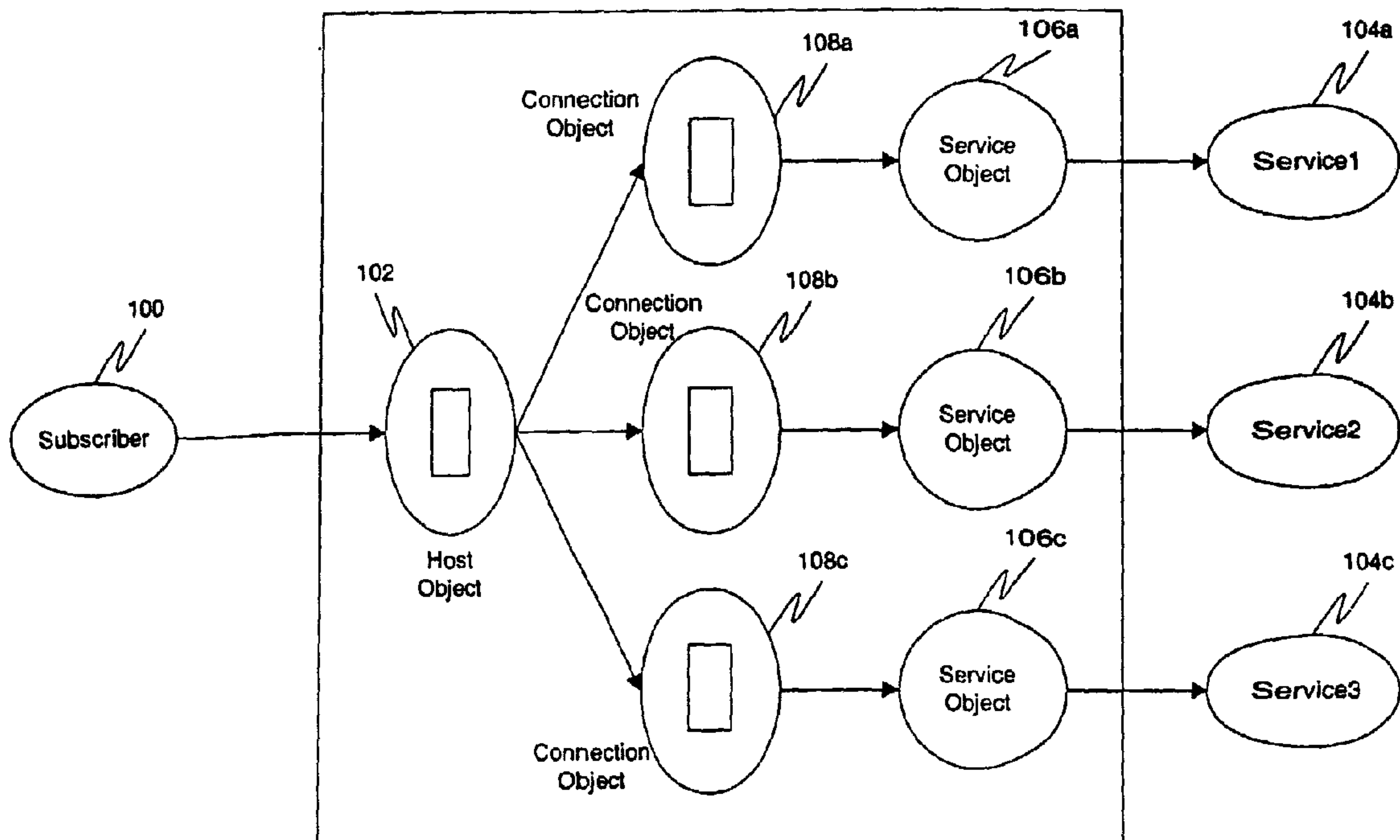
PCT

(10) International Publication Number
WO 2004/030279 A1

- (51) International Patent Classification⁷: **H04L 12/24**, 29/06
- (74) Agents: **RITCHIE, David, B.** et al.; Thelen, Reid & Priest, LLP, P.O. Box 640640, San Jose, CA 95134 (US).
- (21) International Application Number: PCT/US2003/030308
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, 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, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.
- (22) International Filing Date: 23 September 2003 (23.09.2003)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 10/256,787 26 September 2002 (26.09.2002) US
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant: **CISCO TECHNOLOGY, INC.** [US/US]; 170 West Tasman Drive, MS SJC 10/2/1, San Jose, CA 95134-1706 (US).
- (72) Inventors: **ZHANG, Shujin**; 3625 Cisco Way, San Jose, CA 95134 (US). **JIANG, Tianji**; 3625 Cisco Way, San Jose, CA 95134 (US).
- Published: — with international search report

[Continued on next page]

(54) Title: PER USER PER SERVICE TRAFFIC PROVISIONING



(57) Abstract: A host object (102) representing a user, a service object (106a, 106b, 106c) representing a service, and a connection object (108a, 108b, 108c) linking the two may be utilized in order to provide Quality of Service (QoS). The host object (102) and/or connection object (108a, 108b, 108c) may contain a provisioning scheme defining a limit on traffic, which can be provided through either a local or remote programming mechanism. When traffic flows through the host object (102) and/or connection object (108a, 108b, 108c), the appropriate provisioning scheme is utilized to determine which packets to discard.

WO 2004/030279 A1

WO 2004/030279 A1



-
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

S P E C I F I C A T I O N

TITLE OF INVENTION

PER USER PER SERVICE TRAFFIC PROVISIONING

FIELD OF THE INVENTION

[0001] The present invention relates to the field of computer network gateways. More specifically, the present invention relates to a solution to provide quality of service (QoS) to gateway users through a traffic provisioning scheme.

BACKGROUND OF THE INVENTION

[0002] Subscriber aggregation systems allow computer network users to simultaneously connect to multiple service destination networks. In order to obtain the services provided by different service networks, a subscriber will exchange data or control packets with the networks. This traffic may be received from the subscriber by a gateway in the upstream direction, then received from the service network by the gateway in the downstream direction.

[0003] Through the user of a specialized gateway, a subscriber may dynamically choose one or more types of services. Each type of service may have its own bandwidth requirements, as well as different pricing levels, in order to provide corresponding quality of service levels to its subscribers. Thus the bandwidth consumed by a subscriber in either the upstream or downstream direction may fluctuate with the services selected. Accordingly, the price a subscriber would pay for its subscription would also vary. For example, a subscriber may pay \$19.95 a month to an ISP for basic service, while paying \$39.95 for premier service. Additionally, sometimes a user may simultaneously subscribe to multiple services offered by an ISP, such as video service and Internet access, at different prices. The video service usually requires the higher bandwidth and the ISP has to provision the transmission media, such as digital subscriber line (DSL) to the acceptable high speed. Unfortunately, this allows the user to enjoy Internet at a higher speed as well,

and an ISP may wish to charge more money for high speed Internet than high speed video services. These may collectively be known as Quality of Service (QoS).

[0004] The most common approach to handling these problems has been to use an Asynchronous Transfer Mode (ATM) virtual circuit (VC) for each subscriber and to set a limit on the VC. This is known as traffic provisioning on a per subscriber line basis. However, there are several limitations to this approach. First, there may be more than one user on a subscriber line. For example, when the subscriber is a small company, there will be multiple employees who share the same line. This approach cannot differentiate between the users. Second, a user may have access to more than one service provider. Each service provider may wish to set its own limit. For example, one service provider may set a 1.5Mb limit and another 0.5Mb. This approach cannot apply different limits to different services, it is only able to apply an aggregated limit of 2Mb to the VC as a whole.

[0005] What is needed is a solution that overcomes these limitations.

BRIEF DESCRIPTION

[0006] A host object representing a user, a service object representing a service, and a connection object linking the two may be utilized in order to provide Quality of Service (QoS). The host object and/or connection object may contain a provisioning scheme defining a limit on traffic, which can be provided through either a local or remote programming mechanism. When traffic flows through the host object and/or connection object, the appropriate provisioning scheme is utilized to determine which packets to discard.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] The accompanying drawings, which are incorporated into and constitute a part of this specification, illustrate one or more embodiments of the present invention and, together with the detailed description, serve to explain the principles and implementations of the invention.

[0008] In the drawings:

FIG. 1 is a diagram illustrating a gateway architecture in accordance with a specific embodiment of the present invention.

FIG. 2 is a diagram illustrating token bucket provisioning in accordance with a specific embodiment of the present invention.

FIG. 3 is a flow diagram illustrating a method for providing quality of service for a user in a gateway in accordance with a specific embodiment of the present invention.

FIG. 4 is a flow diagram illustrating a method for providing quality of service for a service in a gateway in accordance with a specific embodiment of the present invention.

FIG. 5 is a flow diagram illustrating a method for providing quality of service for a user subscribing to one or more services in a gateway in accordance with a specific embodiment of the present invention.

FIG. 6 is a block diagram illustrating a gateway for providing quality of service for a user in accordance with a specific embodiment of the present invention.

FIG. 7 is a block diagram illustrating a gateway for providing quality of service for a service in accordance with a specific embodiment of the present invention.

FIG. 8 is a block diagram illustrating a gateway for providing quality of service for a user subscribing to one or more services in accordance with a specific embodiment of the present invention.

DETAILED DESCRIPTION

[0009] Embodiments of the present invention are described herein in the context of a system of computers, servers, and software. Those of ordinary skill in the art will realize that the following detailed description of the present invention is illustrative only and is not intended to be in any way limiting. Other embodiments of the present invention will readily suggest themselves to such skilled persons having the benefit of this disclosure.

Reference will now be made in detail to implementations of the present invention as illustrated in the accompanying drawings. The same reference indicators will be used throughout the drawings and the following detailed description to refer to the same or like parts.

[0010] In the interest of clarity, not all of the routine features of the implementations described herein are shown and described. It will, of course, be appreciated that in the development of any such actual implementation, numerous implementation-specific decisions must be made in order to achieve the developer's specific goals, such as compliance with application- and business-related constraints, and that these specific goals will vary from one implementation to another and from one developer to another. Moreover, it will be appreciated that such a development effort might be complex and time-consuming, but would nevertheless be a routine undertaking of engineering for those of ordinary skill in the art having the benefit of this disclosure.

[0011] In accordance with one embodiment of the present invention, the components, process steps, and/or data structures may be implemented using various types of operating systems (OS), computing platforms, firmware, computer programs, computer languages, and/or general-purpose machines. The method can be run as a programmed process running on processing circuitry. The processing circuitry can take the form of numerous combinations of processors and operating systems, or a stand-alone device. The process can be implemented as instructions executed by such hardware, hardware alone, or any combination thereof. The software may be stored on a program storage device readable by a machine. In addition, those of ordinary skill in the art will recognize that devices of a less general purpose nature, such as hardwired devices, field programmable gate arrays (FPGAs), application specific integrated circuits (ASICs), or the like, may also be used without departing from the scope and spirit of the inventive concepts disclosed herein.

[0012] A host object representing a user, a service object representing a service, and a connection object linking the two may be utilized in order to provide Quality of Service (QoS). In a specific embodiment of the present invention, when a user logs in, the gateway may create the host object. The user can then log in, either sequentially or concurrently, to one or more subscribed services. When a user logs into a subscribed

service, a service object may be created to describe the service (unless the service object has already been created due to actions of other subscribers). A connection object may then be created to link the user's host object with the service object. A packet from a user to the Internet may then flow from the user to the host object, then to the connection object, then to a service object, and then to the Internet. A packet from the Internet to the user may flow in the reverse order.

[0013] In a specific embodiment of the present invention, a provisioning scheme may be placed inside the host object and/or the connection object. The provisioning scheme for a particular service may first be stored in a service profile, and then copied to each connection object created that is linked to the service. The provisioning scheme may monitor the traffic rate and discard certain packets when the rate exceeds the limit accorded to packets meeting criteria defined in the scheme. That is an example of a policing scheme. However, other provisioning schemes such as shaping may also be applied.

[0014] Thus, several different provisioning cases may be implemented using the present invention. These include:

1. Aggregated Provisioning. If a user has subscribed to multiple service, the network access provider can set a provisioning scheme on the host object so that the user cannot send or receive traffic at a higher rate than the limits.
2. Service Provisioning. Each service provider can set its own limit by set a provisioning scheme in the appropriate service object.
3. Per user and per service provisioning. The network access provider can set the limits on the user and the service provider can also set the limits on each service. For example, the limit on the user may be set at 2Mb. The limit on service A may be set at 1.5Mb and the limit on service B at 1Mb. If the user subscribes to both services at the same time, the user will be limited by both the service limits and the total 2Mb limit. If the user only subscribes to one service, the user will only be limited by the total 2Mb limit as it is higher than the service limit.

[0015] FIG. 1 is a diagram illustrating a gateway architecture in accordance with a specific embodiment of the present invention. A subscriber 100 may have an associated host object 102. The user may be subscribed to three services 104a, 104b, 104c. Each service may have a corresponding service object 106a, 106b, 106c. A connection object 108a, 108b, 108c may associated each service object 106a, 106b, 106c with the appropriate host object 102.

[0016] An additional benefit of a specific embodiment of the present invention is that the Remote Access Dial-In User Service (RADIUS) access control protocol may be utilized to automate configuration of the provisioning schemes. In the past, any provisioning scheme had to be manually configured through a command line interface (CLI). By modifying the user profile and service profile to include an attribute describing the provisioning parameters, the provisioning schemes may be easily modified through a remote subscriber management protocol such as RADIUS. This may include information on upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream burst size, and downstream excess burst size. In the user profile, this defines the average bandwidth the user is entitled to obtain and the normal/excess burst tolerance the user can have. In the service profile, this defines the average rate a service has to achieve and the normal excess/burst size the service can tolerate to provide corresponding quality of service.

[0017] In a specific embodiment of the present invention, the provisioning scheme may be implemented as a token bucket mechanism. FIG. 2 is a diagram illustrating token bucket provisioning in accordance with a specific embodiment of the present invention. A token rate TR 200 is set, which corresponds to either the aggregated bandwidth a subscriber is entitled to achieve in the per user provisioning or the bandwidth an individual service requires to provide acceptable quality in the per user per service provisioning. A token bucket size or the normal burst size TBS 202 may be used to restrict the burst size a stream of traffic may reach.

[0018] When a packet with length L, either from the upstream or downstream direction, arrives at the policer (whether in the host object or the connection object), the packet may be passed on if its token bucket has enough tokens ($\geq L$) and the number of

tokens in the bucket will be updated accordingly. If the bucket does not have sufficient tokens ($<L$), the packet will be dropped and the number of tokens in the bucket will not be updated. Thus:

```

When a packet with length L arrives at the gateway,
Update token bucket: Token# = Min (Token# +
TR*Elapsed_time_since_last_update, TBS),
If      (Token #  $\geq$  L) then
    Pass the packet,
    Update token bucket: Token# = Token# - L;
Else
    Drop the packet without updating token bucket

```

[0019] When the optional parameter, extended burst size ETBS, is set, it should be equal to or greater than TBS. When ETBS is equal to TBS, the provisioning scheme is the token bucket scheme. When ETBS is greater than TBS, the extended burst limit may be applied. Unlike a standard token bucket scheme, a token bucket with an extended burst size allows a stream to borrow more tokens. The need for a token bucket with extended burst capability is not to immediately enter into a tail-drop scenario such as the standard token bucket, but rather to gradually drop packets in a more random early detection (RED)-like fashion.

[0020] The scheme may make use of the comparison results among actual debt, compound debt, and extended burst size to determine the provisioning result of a packet. Actual debt is a count of how many tokens the flow has currently borrowed. Compounded debt is the sum of all actual debt since the last time a packet was dropped. If the compounded debt is greater than the extended burst size, the packet would be dropped. After the dropping, the compounded debt is effectively set to zero, and a new value will be computed the next time a new packet is dropped. If the actual debt is greater than the extended burst value, all packets will be dropped until the actual debt is reduced through accumulation of tokens in the token bucket. Dropped packets do not count against any rate or burst limit. That is, when packet is dropped, no tokens are removed from the token bucket.

[0021] FIG. 3 is a flow diagram illustrating a method for providing quality of service for a user in a gateway in accordance with a specific embodiment of the present invention. At 300, a host object corresponding to the user is defined, the host object having a provisioning scheme indicating a limit on traffic sent to or from the user. This may occur when a user logs in. The host object may be defined using a user profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. At 302, traffic sent to or from the user that exceeds the limit is discarded.

[0022] FIG. 4 is a flow diagram illustrating a method for providing quality of service for a service in a gateway in accordance with a specific embodiment of the present invention. At 400, a service object corresponding to the service is defined. This may occur when a user logs in to the service. At 402, a connection object linking the service object to a host object is created, the host object corresponding to a user subscribing to the service, wherein the connection object has a provisioning scheme indicating a limit on traffic sent to or from the service. The service object may be defined using a service profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. At 404, traffic sent to or from the service that exceeds the limit is discarded.

[0023] FIG. 5 is a flow diagram illustrating a method for providing quality of service for a user subscribing to one or more services in a gateway in accordance with a specific embodiment of the present invention. At 500, a host object corresponding to the user is defined, the host object having a provisioning scheme indicating a limit on traffic sent to or from the user. This may occur when a user logs in. The host object may be defined using a user profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess

burst size. The provisioning scheme may be implemented as token bucket provisioning. At 502, a service object corresponding to the service is defined. This may occur when a user logs in to the service. At 504, a connection object linking the service object to the host object is created, wherein the connection object has a provisioning scheme indicating a limit on traffic sent to or from the service. The service object may be defined using a service profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. At 506, traffic sent to or from the user that exceeds the limit on traffic sent to or from the user and traffic sent to or from the service that exceeds the limit on traffic sent to or from the service is discarded.

[0024] FIG. 6 is a block diagram illustrating a gateway for providing quality of service for a user in accordance with a specific embodiment of the present invention. A host object definer 600 defines a host object corresponding to the user, the host object having a provisioning scheme indicating a limit on traffic sent to or from the user. This may occur when a user logs in. The host object may be defined using a user profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. A traffic discarder 602 coupled to the host object definer 600 may discard traffic sent to or from the user that exceeds the limit.

[0025] FIG. 7 is a block diagram illustrating a gateway for providing quality of service for a service in accordance with a specific embodiment of the present invention. A service object definer 700 defines a service object corresponding to the service. This may occur when a user logs in to the service. A connection object creator 702 coupled to the service object definer 700 may create a connection object linking the service object to a host object, the host object corresponding to a user subscribing to the service, wherein the connection object has a provisioning scheme indicating a limit on traffic sent to or from the service. The service object may be defined using a service profile having an attribute

describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. A traffic discarder 704 coupled to the connection object creator 702 may discard traffic sent to or from the service that exceeds the limit.

[0026] FIG. 8 is a block diagram illustrating a gateway for providing quality of service for a user subscribing to one or more services in accordance with a specific embodiment of the present invention. A host object definer 800 may define a host object corresponding to the user, the host object having a provisioning scheme indicating a limit on traffic sent to or from the user. This may occur when a user logs in. The host object may be defined using a user profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. A service object definer 802 may define a service object corresponding to the service. This may occur when a user logs in to the service. A connection object creator 804 coupled to the host object definer 800 and to the service object definer 802 may create a connection object linking the service object to the host object, wherein the connection object has a provisioning scheme indicating a limit on traffic sent to or from the service. The service object may be defined using a service profile having an attribute describing provisioning parameters. These parameters may include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size and downstream excess burst size. The provisioning scheme may be implemented as token bucket provisioning. A traffic discarder 806 coupled to the host object definer 800 and to the connection object creator 804 may discard traffic sent to or from the user that exceeds the limit on traffic sent to or from the user and traffic sent to or from the service that exceeds the limit on traffic sent to or from the service.

[0027] While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art having the benefit of this disclosure that many more modifications than mentioned above are possible without

departing from the inventive concepts herein. The invention, therefore, is not to be restricted except in the spirit of the appended claims.

What is claimed is:

1. A method in a gateway for providing quality of service for a user subscribing to a plurality of subscription services, comprising:
 - defining a host object corresponding to the user, the host object having a provisioning scheme indicating a first limit on traffic sent to or from the user;
 - defining a plurality of service objects each corresponding to a respective distinct one of the subscription services;
 - creating a plurality of connection objects each linking a corresponding service object to said host object, each connection object having a provisioning scheme indicating a respective second limit on traffic of the user sent to or from the corresponding subscription service;
 - discarding traffic of the user that exceeds said first limit; and
 - for each of the subscription services, discarding traffic of the user sent to or from the subscription service that exceeds said respective second limit .
2. The method of claim 1, wherein said defining a host object occurs when the user logs in to said gateway.
3. The method of claim 1, wherein said host object is defined using a user profile having an attribute describing provisioning parameters.
4. The method of claim 1, wherein said defining a plurality of service objects occurs when the user logs in to the subscription services.
5. The method of claim 1, wherein each of said connection objects is created using a respective service profile having an attribute describing provisioning parameters.
6. The method of claim 3, wherein said provisioning parameters include upstream bandwidth, upstream normal burst size, upstream excess burst size,

downstream bandwidth, downstream normal burst size, and downstream excess burst size.

7. The method of claim 5, wherein said provisioning parameters include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size, and downstream excess burst size.

8. The method of claim 1, wherein said provisioning scheme is implemented as token bucket provisioning.

9. The method of claim 1, further including receiving said provisioning schemes through a remote subscriber management system.

10. A gateway operative to provide quality of service for a user subscribing to a plurality of subscription services, comprising:

- a host object definer configured to define a host object corresponding to the user, the host object having a provisioning scheme indicating a first limit on traffic sent to or from the user;

- a service object definer configured to define a plurality of service objects each corresponding to a respective distinct one of the subscription services;

- a connection object creator coupled to said host object definer and said service object definer and configured to create a plurality of connection objects each linking a corresponding service object to said host object, each connection object having a provisioning scheme indicating a respective second limit on traffic of the user sent to or from the corresponding subscription service; and

- a traffic discarder coupled to said host object definer and said connection object creator and configured to (1) discard traffic of the user that exceeds said first limit, and (2) for each of the subscription services, discard traffic of the user sent to or from the subscription service that exceeds said respective second limit.

11. An apparatus in a gateway for providing quality of service for a user subscribing to a plurality of subscription services , comprising:

means for defining a host object corresponding to the user, the host object having a provisioning scheme indicating a first limit on traffic sent to or from the user;

means for defining a plurality of service objects each corresponding to a respective distinct one of the subscription services;

means for creating a plurality of connection objects each linking a corresponding service object to said host object, each connection object having a provisioning scheme indicating a respective second limit on traffic of the user sent to or from the corresponding subscription service;

means for discarding traffic of the user that exceeds said first limit ; and

means for discarding, for each of the subscription services, traffic of the user sent to or from the subscription service that exceeds said respective second limit .

12. The apparatus of claim 11, wherein said defining a host object occurs when the user logs in to the gateway.

13. The apparatus of claim 11, wherein said host object is defined using a user profile having an attribute describing provisioning parameters.

14. The apparatus of claim 11, wherein said defining a plurality of service objects occurs when the user logs in to the subscription services.

15. The apparatus of claim 11, wherein each of said connection objects is created using a respective service profile having an attribute describing provisioning parameters.

16. The apparatus of claim 13, wherein said provisioning parameters include

upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size, and downstream excess burst size.

17. The apparatus of claim 15, wherein said provisioning parameters include upstream bandwidth, upstream normal burst size, upstream excess burst size, downstream bandwidth, downstream normal burst size, and downstream excess burst size.

18. The apparatus of claim 11, wherein said provisioning scheme is implemented as token bucket provisioning.

19. The apparatus of claim 11, further including means for receiving said provisioning schemes through a remote subscriber management system.

20. A program storage device readable by a machine, tangibly embodying a program of instructions executable by the machine to perform a method for providing quality of service for a user subscribing to a plurality of services, the method comprising:

- defining a host object corresponding to the user, the host object having a provisioning scheme indicating a first limit on traffic sent to or from the user;

- defining a plurality of service objects each corresponding to a respective distinct one of the services;

- creating a plurality of connection objects each linking a corresponding service object to said host object, each connection object having a provisioning scheme indicating a respective second limit on traffic of the user sent to or from the corresponding subscription service;

- discarding traffic of the user that exceeds said first limit ; and

- for each of the subscription services, discarding traffic of the user sent to or from the subscription service that exceeds said respective second limit .

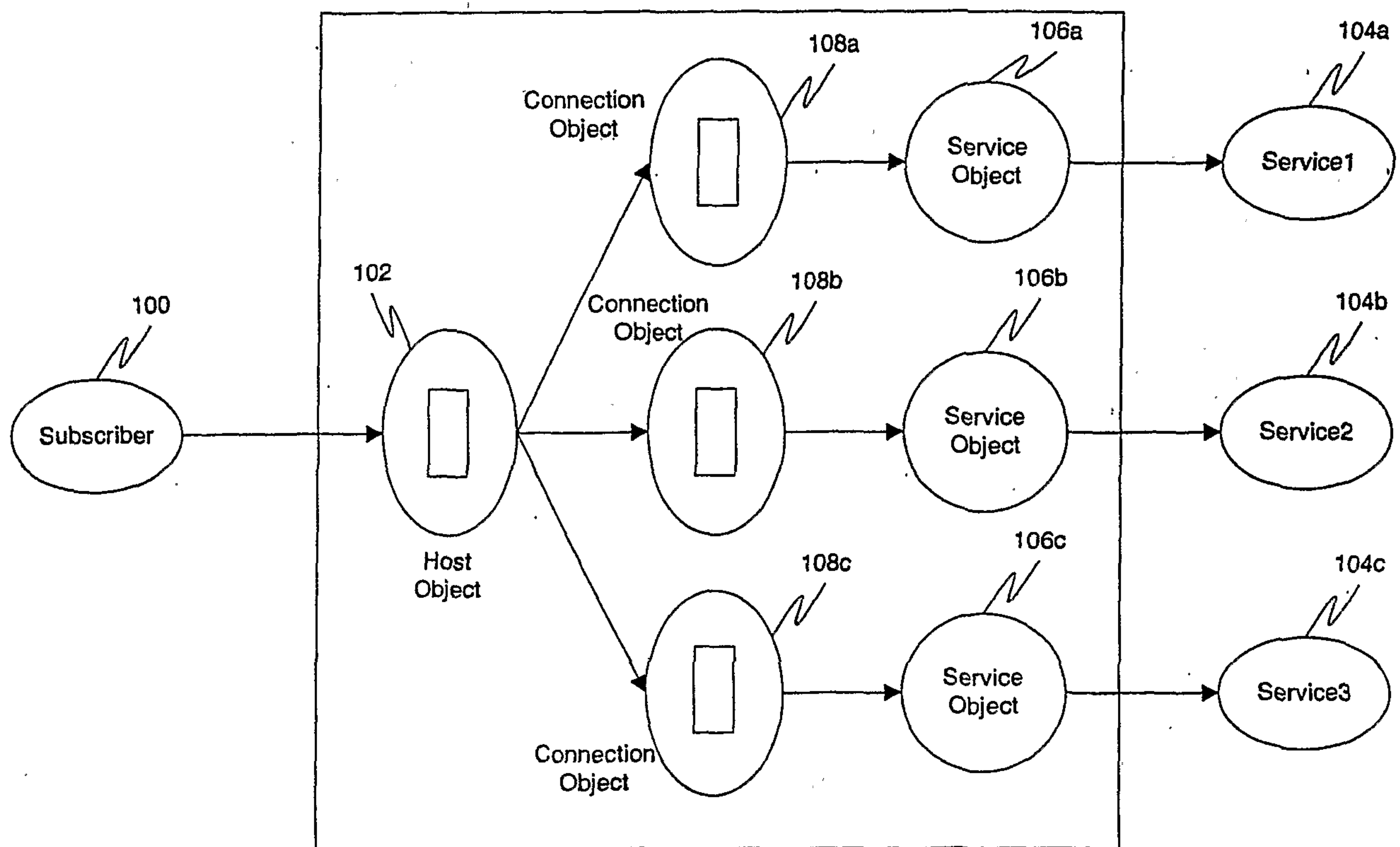


FIG. 1

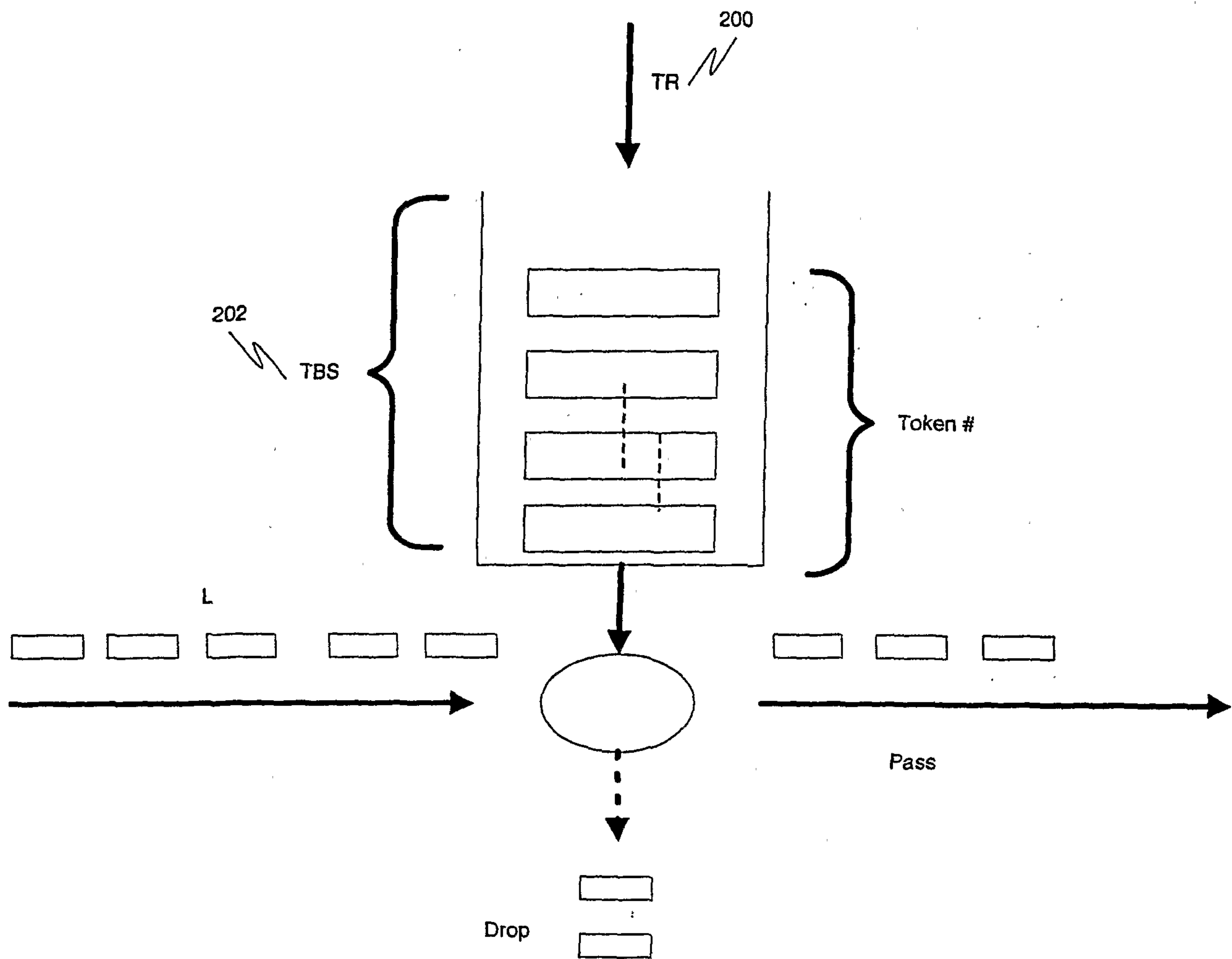


FIG. 2

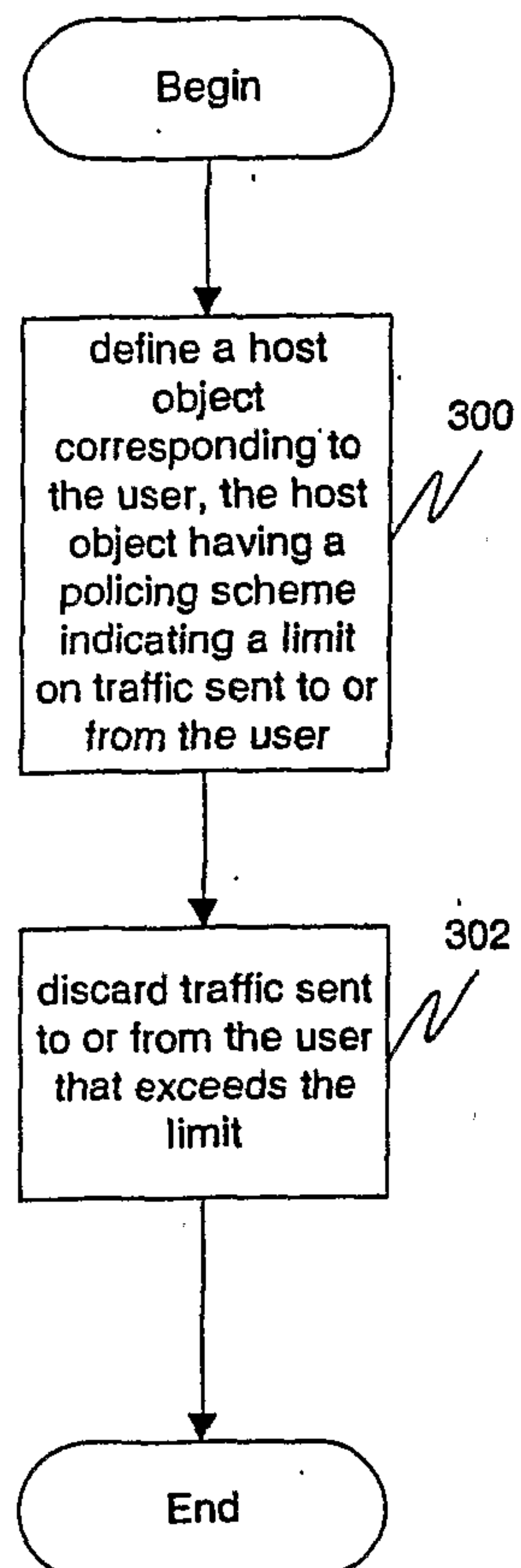


FIG. 3

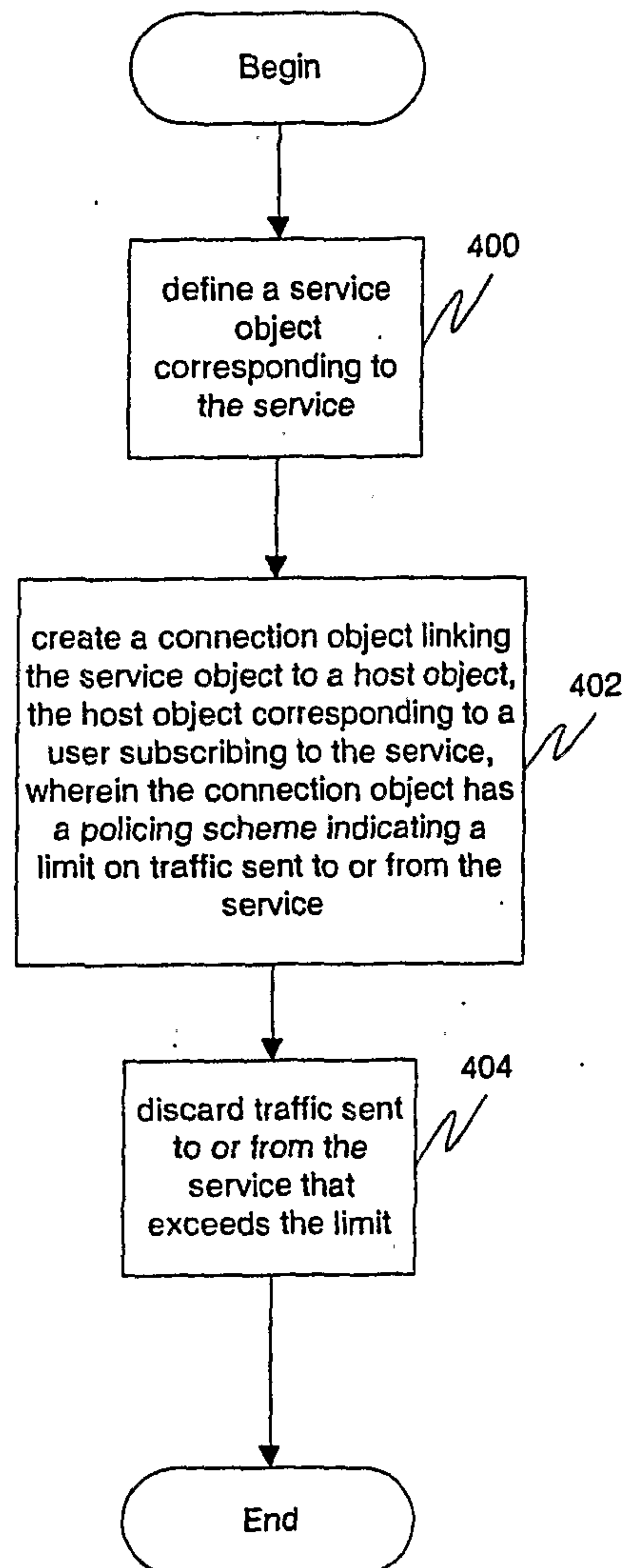


FIG. 4

5/8

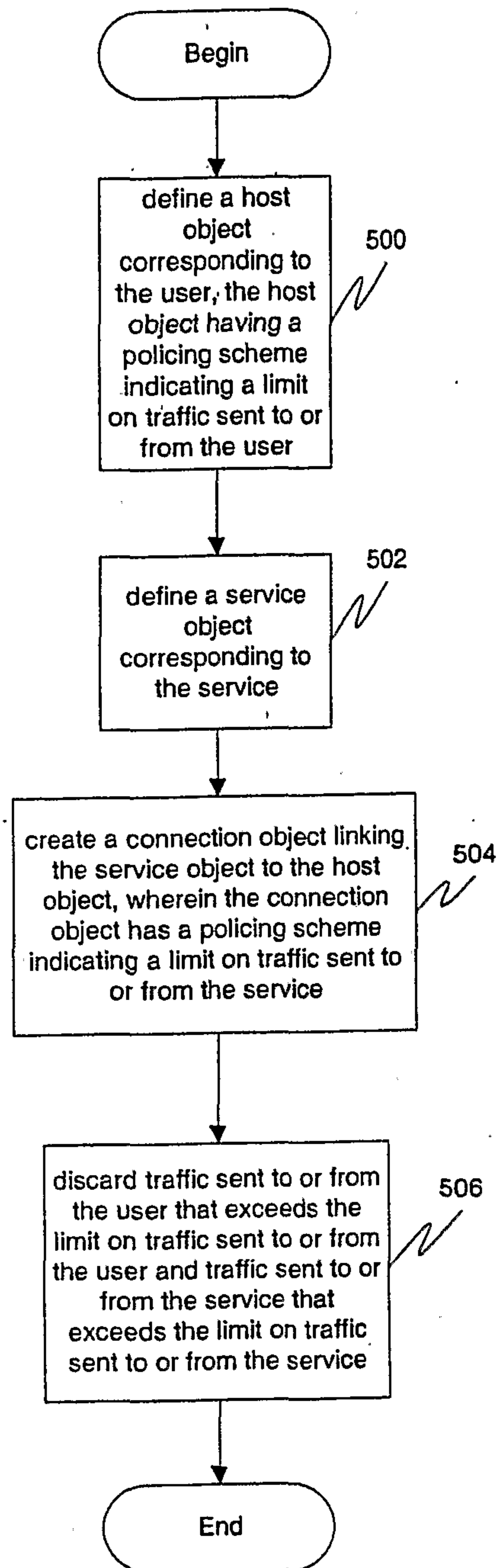


FIG. 5

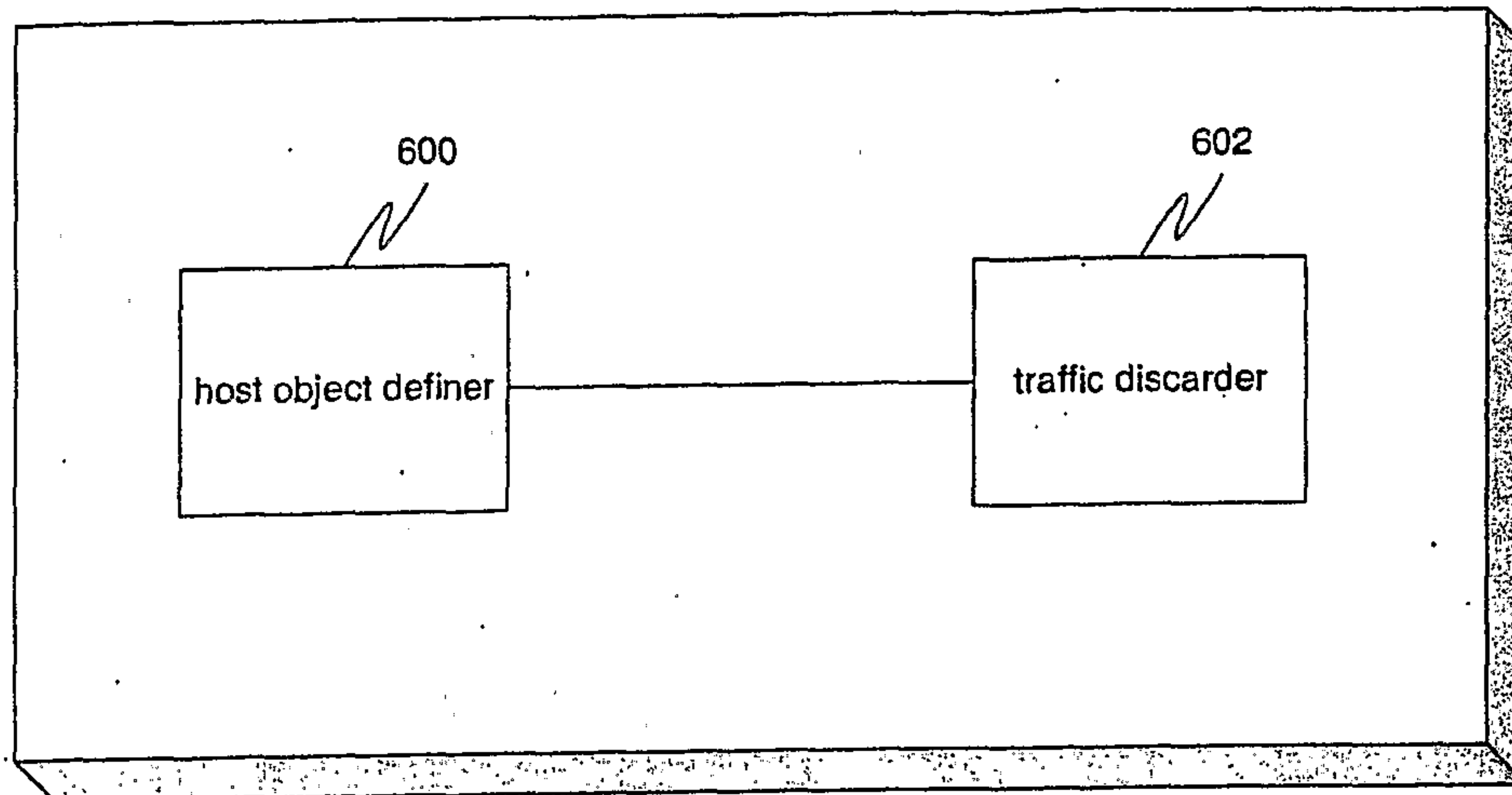


FIG. 6

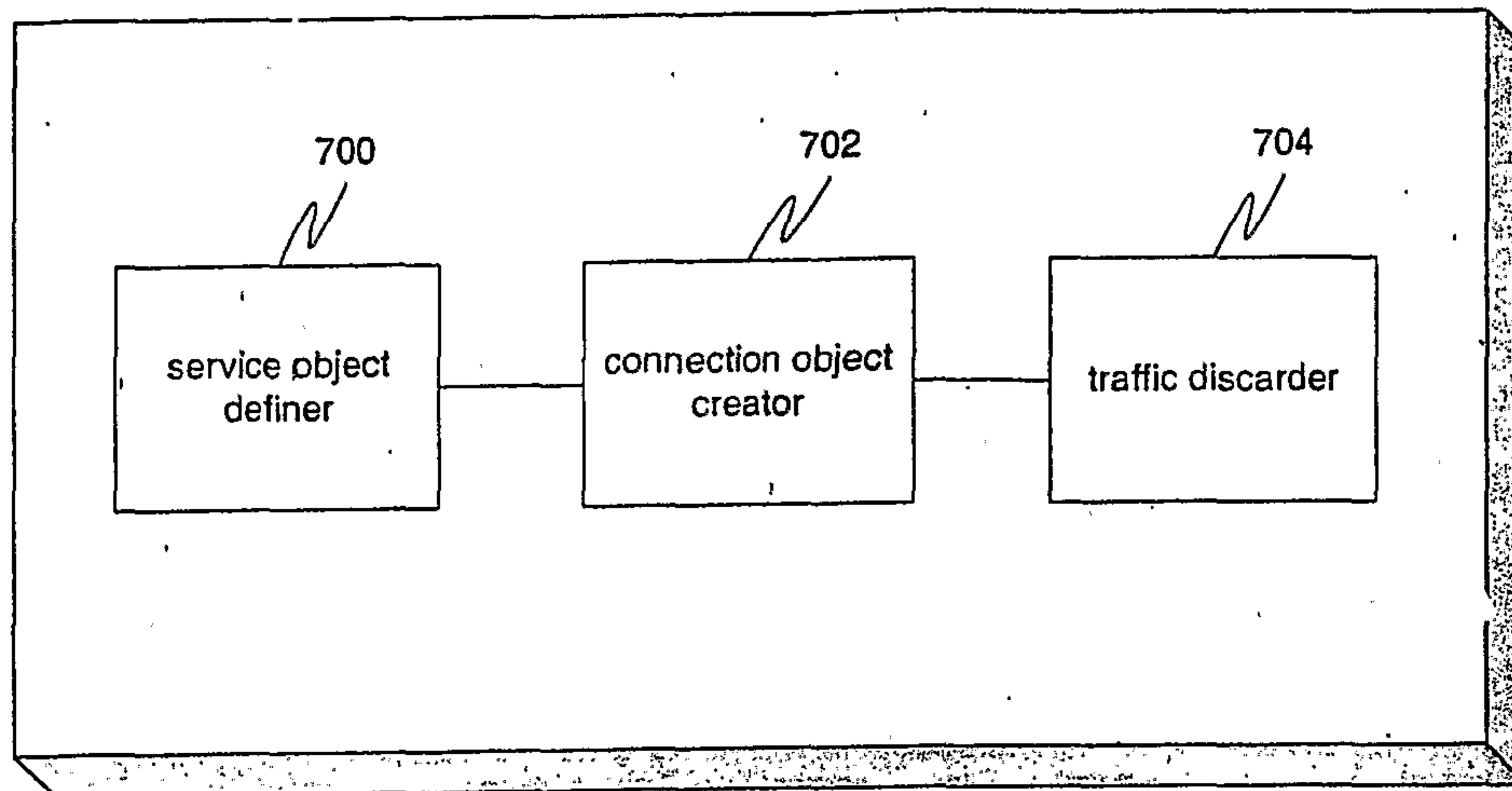


FIG. 7

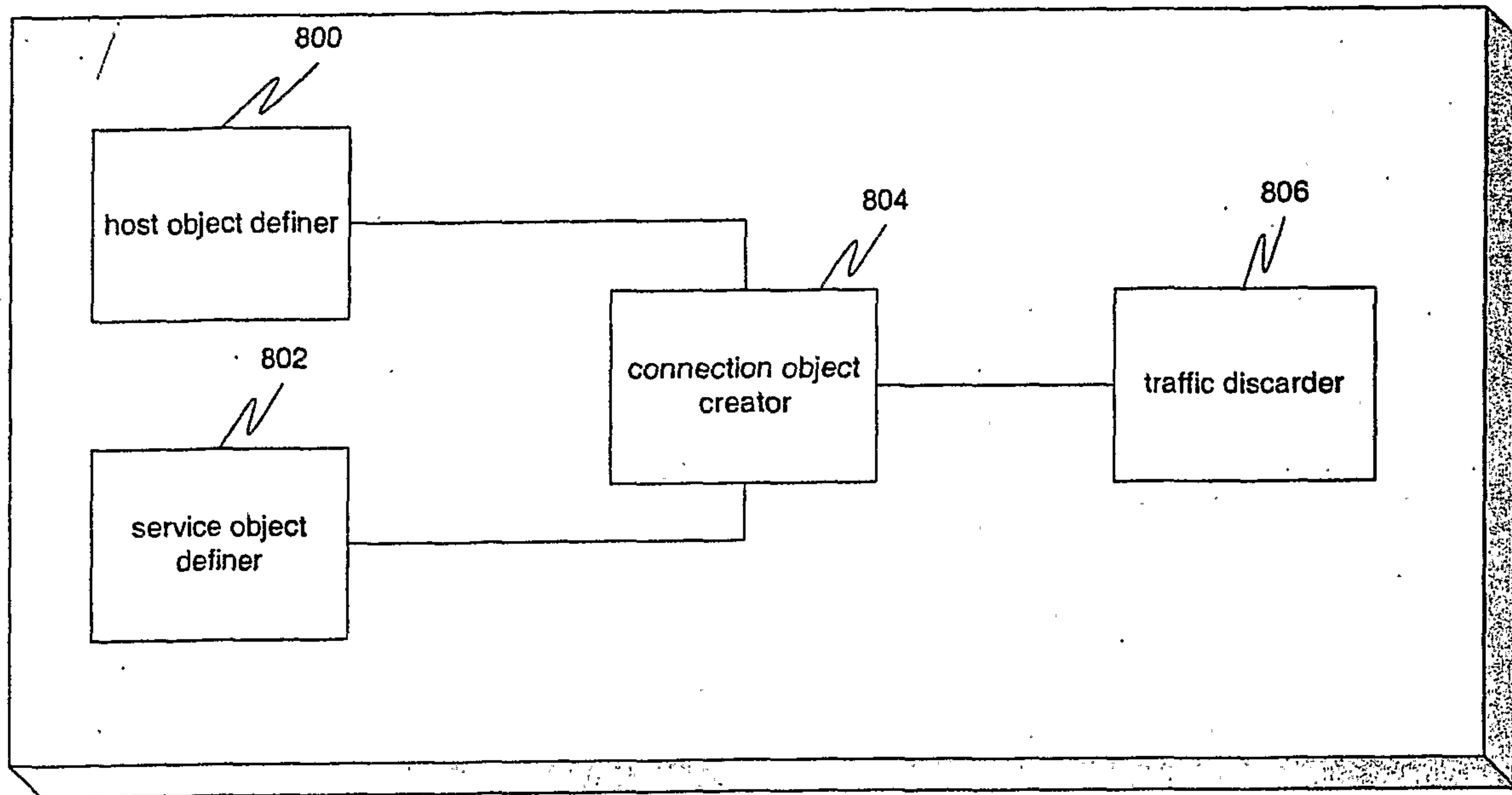


FIG. 8

