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

(51) **International Patent Classification[7]:** H04L 12/26, 12/24, 12/56

(21) **International Application Number:**
PCT/EP2004/008523

(22) **International Filing Date:** 29 July 2004 (29.07.2004)

(25) **Filing Language:** English

(26) **Publication Language:** English

(71) **Applicant** *(for all designated States except US)*: **TELE-COM ITALIA S.P.A.** [IT/IT]; Piazza degli Affari, 2, I-20123 Milano (IT).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **VERCELLONE, Vinicio** [IT/IT]; Telecom Italia S.p.A., Via G. Reiss Romoli, 274, I-10148 Torino (IT). **FAURE RAGANI, Alessandro** [IT/IT]; Telecom Italia S.p.A., Via G. Reiss Romoli, 274, I-10148 Torino (IT). **SASSI, Massimo** [IT/IT]; Telecom Italia S.p.A., Via G. Reiss Romoli, 274, I-10148 Torino (IT). **ALESSI, Marzio** [IT/IT]; Telecom Italia S.p.A., Via G. Reiss Romoli, 274, I-10148 Torino (IT). **FANTOLINO, Luca** [IT/IT]; Telecom Italia S.p.A., Via G. Reiss Romoli, 274, I-10148 Torino (IT).

(74) **Agents: BATTIPEDE, Francesco** et al.; Pirelli & C. S.p.A., Viale Sarca, 222, I-20126 Milano (IT).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, 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, NA, 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, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

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.

(54) **Title:** METHOD AND SYSTEM FOR FAULT AND PERFORMANCE RECOVERY IN COMMUNICATION NETWORKS, RELATED NETWORK AND COMPUTER PROGRAM PRODUCT THEREFOR
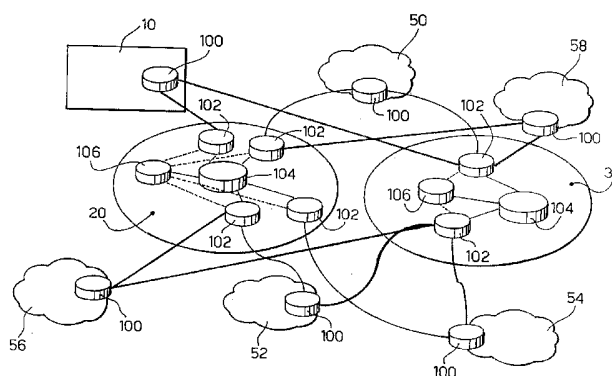
(57) **Abstract:** A system for fault and performance recovery in a communication network includes a set of network apparatuses (100, 102, 104, 106) as well as a plurality of sites (50, 52, 54, 56, 58) connected by means of connections having a given topology. The network also includes routers (100) for implementing given routing policies and has associated at least one database (370) maintaining information regarding the network apparatuses. The system includes test modules (represented by the - suitably configured - apparatuses and/or by specific hardware such as so-called "probes") for executing performance tests created automatically by exploiting the information available on the topology of the connections between the various sites and by accessing the database in question. The test modules are configured for analysing the results of the performance tests to detect possible fault and performance impairment conditions of said network. In the presence of a fault and performance impairment condition detected in the network, the routers (100) modify the routing policy to compensate for the fault and performance impairment condition.

"Method and system for fault and performance recovery
in communication networks, related network and computer
program product therefor"

5   * * *

Field of the invention

        The present invention relates to techniques for
10 fault and performance recovery in communication
    networks. The invention was devised in view of the
    preferred application to Virtual Private Network
    services provided through multiple backbones. However,
    reference to this preferred field of application is not
15 to be considered as limiting the scope of the
    invention.

Description of the related art

20   The significant development and ever increasing
    diffusion of telematic networks based upon the Internet
    Protocol (IP) and the increasingly important role that
    these assume in the context of company realities, more
    and more frequently leads to the use of these networks
25 as means of communication between the different
    entities in which an organization is structured. A
    typical solution in this connection consists in the use
    of Virtual Private Networks (VPNs), offered by a
    Service Provider on his own backbone network
30 infrastructure, with characteristics of privacy and
    security, for connection between the sites of a
    customer.
        In the context of the types of IP Virtual Private
    Networks offered as service managed by a Service
35 Provider, a model of embodiment of particular interest

CONFIRMATION COPY

2

is the technology known as IP Virtual Private Networks
in Multi Protocol Label Switching (MPLS). For this
reason, reference will be made, for example, to Border
Gateway Protocol / Multi Protocol Label Switching
5   Virtual Private Network (BGP/MPLS VPN) solutions,
hereinafter referred to as MPLS-VPN networks. The
corresponding architecture is defined in E. Rosen, Y.
Rekhter, "BGP/MPLS VPNs", IETF RFC 2547.

    All the standards/specifications referred to in
10 the following as ietf-… or RFC… are publicly available
at the date this application is filed with the IETF
Website at www.ietf.org.

    The elements of the Border Gateway Protocol /
Multi Protocol Label Switching Virtual Private Networks
15 architecture,   which   presents   advantageous
characteristics from the standpoint of scalability and
security, are illustrated in Figure 1.

    The elements shown are essentially a service
provider's Network Operating Center (NOC) 10, a Multi
20 Protocol Label Switching/IP backbone (MPLS/IP) 20 and
30, a set of edge routers of the customer (CE, Customer
Edge) 100, a set of edge routers of the Service
Provider (PE, Provider Edge) 102, a set of core routers
of the Service Provider (P, Provider Core) 104, and a
25 set of Route Reflector Servers (RRS) 106. These latter
elements, which are optional, have the function of
reducing the complexity inherent in the mesh of
sessions of the routing protocol Multi Protocol -
Border Gateway Protocol (MP-BGP) between the different
30 Provider Edges 102 of the network, as described in
RFC2547.

    Within the framework of the offer of Virtual
Private Networks services there are at times present
customer requirements for a high level of guarantee of
35 availability of the service, even during periods of

3

congestion or failures which introduce poor service or
degradation of the performance offered by the network.
These requirements can be satisfied, for example, by
means of a complete duplication of the backbone-network
5   infrastructures on which the Virtual Private Networks
service is supplied. Duplication can be either logical
or physical.

        If, on the one hand, this solution enables a
significant increase in the degree of availability and
10 quality of the network service, on the other, it
introduces the problem of appropriate management of
routing of the customer traffic on the various networks
that supply Virtual Private Network connectivity.

        The network scenario to which reference is made is
15 thus the one represented, in general terms, in Figure
1. The customers' sites 50, 52, 54, 56, or 58, are
connected through the respective Customer Edge 100
routers to two or more distinct backbones 20 or 30.

        A problem arises in detecting any possible
20 failure, or degradation of performance of the backbone
by means of which the sites exploit the service, while
also handling restoration of the service in very short
time intervals. Restoration is obtained by means of
control of re-routing of the customer traffic and must
25 be activatable also in the case where the poor service
is due to causes internal to the backbone network and
not detectable through mere monitoring of the access
links.

        Mere recourse to re-routing mechanisms obtained by
30 Border Gateway Protocol routing and managed by means of
the usual modalities made available by the protocol
itself does not provide a complete solution of the
problem. These mechanisms do not guarantee detection of
all the situations that cause a loss of connectivity or
35 a degradation in the quality of communication between

4

the sites. In fact, the Border Gateway Protocol supplies just an indication of reachability, but does not provide information on the quality of this reachability. For example, it does not typically enable detection of phenomena of malfunctioning that reduce the capacity of transfer of the network, and consequently the performance, below acceptable values.

Furthermore, in the framework of IP Multi Protocol Label Switching Virtual Private Networks, the Multi Protocol-Border Gateway Protocol (MP-BGP) is used within the backbone. Typically, a Service Provider configures the protocol with values of the timers of the same order of magnitude as the default values. In these conditions, a fault would be noted in the network (and hence it would be possible to react) in times of the order of a few minutes.

In addition, there may arise situations in which, within the backbone, one or more routers (typically Provider Edges) have routing tables, for the routing protocol Interior Gateway Protocol (IGP), that are not correct on account of Hardware/Software problems. In these cases, the undesired effect is that those Border Gateway Protocol advertisements relating to networks for which, instead, connectivity no longer exists, continue to be held valid.

Other solutions are known, which aim at optimising routing policies, overcoming some of the aforesaid limitations deriving from the criteria adopted by Border Gateway Protocol routing.

Documents such as WO-A-02/080462, US-A-2002/0078223, WO-A-03/001330, and WO-A-03/041342, basically deal with Routing Control for connection to the Internet. The application of these techniques of Intelligent Routing Control (IRC) in the context of Virtual Private Network IP presents, however,

peculiarities and problems that are not present in the case of application for Internet multihoming.

More specifically, WO-A-02/080462 describes a system for carrying out control of routing policies in a network including:

- a number of data collectors, which have the job of determining and probing paths in the network, for the purpose of detecting indications regarding the preferential routings, connected to a central server, entrusted with receiving and handling said data; and

- at least one routing control device, connected, in turn, to the central server, which by querying said server, receives indications on the preferential path towards a given destination network.

The functions can be implemented via stand-alone equipment or in the form of centralized and managed services.

US-A-2002/0078223 describes a system and corresponding methods for evaluating the routing paths on the basis of measurements of performance for the purposes of supporting criteria of intelligence in routing.

WO-A-03/001330 describes a system to be inserted alongside a router connected in a "multihomed" way to a number of Transit Autonomous Systems (TASs), which enable a plurality of Destination Autonomous Systems (DASs) to be reached. The mechanism includes a path-testing process, which performs an analysis of the traffic, by sending packets of tests to a set of IP addresses, towards different Destination Autonomous Systems, which can be selected by an operator via configuration files or other interface (e.g., Graphical User Interface, Command Line Interface). In order to carry out this path-testing process on particular links and Transit Autonomous Systems, the path-testing

6

process enters temporarily (with an appropriate configuration of the router) the test route through which the ping traffic is to be sent.

The data collected by the path-testing process is
5   then supplied to a path-evaluation process, consisting of a decision method which evaluates the quality of the paths for each Transit Autonomous Systems / Destination Autonomous Systems pair. A path for which the level of quality is below a certain (configurable) threshold is
10 a candidate for forcing a re-routing operation. The path-selection process proposes, or, if enabled, carries out, re-routing (change of path) by connecting up to the router and modifying the configuration of the Border Gateway Protocol policy. This operation has the
15 effect of re-evaluating, according to the new policy, all the routes learnt from the Transit Autonomous Systems selected.

WO-A-03/041342 describes a system called Flow-Control System, with functions of control of a
20 multihoming connection to the Internet.

However, these solutions are oriented primarily to handling routing of the traffic to a network such as the Internet, and their application to the context of interest illustrated in Figure 1 does not appear
25 immediate.

The known solutions for automation of the process of configuration of the measurements, in general, envisage manual configuration of the set of measurements or the possibility of allowing the system
30 to select automatically the destinations to be probed, according to the networks towards which there is a larger amount of traffic. However, neither of these approaches fits conveniently the context in question.

Another important aspect regards control of the
35 paths used by test traffic and user traffic. In this

7

connection, the known solutions refer principally to the context of an Internet multihomed company, i.e., one equipped with a number of connections to said network. In this framework, it is possible to control

5   exclusively the interface, i.e., the link, on which the test traffic is to be transmitted, without knowing from what link the return traffic will come back in. The control function is then generally achieved by temporary insertion of appropriate static routes in the

10 edge router.

Moreover, in the context of a company equipped with multihomed connection to the Internet the possibility exists of acting exclusively on the routing of the traffic directed to the Internet, by forcing,

15 for example, transmission on the backbone of one particular Internet Provider rather than on another, controlling the edge router or routers of the multihomed Internet site. It is not, instead, possible to act on the other end-point of the communication, in

20 so far as the other end-point does not belong to the same administrative domain. In the same way, no possibility exists of acting on the routers internal to the backbones of the Internet Providers.

The above limitations make it impossible to

25 control with sufficient accuracy the traffic entering the site.

In the context of company Virtual Private Networks coming under a number of backbones (such a term being used herein to designate, in general, a number of

30 different physical and/or virtual connections), it is mandatory to pay particular attention to the performance aspects. In this connection, it is essential for the values of the performance parameters (e.g., delays, losses, etc.) of the various backbones

35 that are compared by the analysis and optimisation

8

methods, to be obtained, as far as possible, in the same conditions, i.e., in the same time intervals. It is hence of basic importance to carry out the measurements on the various backbones in parallel.

This aspect is not addressed by the known solutions, which generally envisage making measurements in a cyclic way, examining one backbone/connection at a time.

Object and summary of the invention

The object of the invention is thus to provide an improved arrangement adapted to satisfy the needs previously described, while dispensing with the intrinsic drawbacks of the prior art arrangements discusses in the foregoing.

According to the present invention, that object is achieved by means of a method having the features set forth in the claims that follow. The invention also relates to a corresponding system, a related network as well as a related computer program product, loadable in the memory of at least one computer and including software code portions for performing the steps of the method of the invention when the product is run on a computer. As used herein, reference to such a computer program product is intended to be equivalent to reference to a computer-readable medium containing instructions for controlling a computer system to coordinate the performance of the method of the invention. Reference to "at least one computer" is evidently intended to highlight the possibility for the present invention to be implemented in a distributed/ modular fashion.

A preferred embodiment of the invention is thus a system for fault and performance recovery in a

9

communication network including a set of network apparatuses as well as a plurality of sites connected by means of connections having a given topology. The network also includes routers for implementing given

5   routing policies and has associated at least one database maintaining information regarding the network apparatuses. The system includes test modules (represented by the - suitably configured - apparatuses and/or by specific hardware such as so-called "probes")

10 for executing performance tests created automatically by exploiting the information available on the topology of the connections between the various sites and by accessing the database in question. The test modules are configured for analysing the results of the

15 performance tests to detect possible fault and performance impairment conditions of said network. In the presence of a fault and performance impairment condition detected in the network, the routers are instructed to modify the routing policy in order to

20 compensate for the fault and performance impairment condition.

A particularly preferred embodiment of the arrangement described herein overcomes the limitations of the solution based exclusively on the usual Border

25 Gateway Protocol modalities, by verifying continuously the effective continuity of the service, identifying also any significant degradation in performance and undertaking any appropriate action for the purpose of correcting, in real time, the routing policies.

30 The strategies used by the arrangement described herein, enable advantage to be drawn from all the possibilities of correlation of the indications detected and of co-ordination of the corresponding corrective action. In a context such as the one in

35 question, these may be exploited in order to obtain a

fast and optimal response in the case of failure or degradation of performance, for the customers of the Multi Protocol Label Switching Virtual Private Networks service. The network architecture and the context of

5   service that the arrangement described herein addresses are suited, in fact, to exploiting of certain prerogatives, in favour of rapidity of intervention of the system and of optimisation of the strategies of intervention.

10   The main preferred features of the arrangement described herein can be grouped together in a number of areas, which regard, respectively:

        - creation and configuration of the performance tests necessary for detection of any service

15   deficiencies;

        - co-ordinated control of re-routing of the traffic; and

        - management of the measurement processes.

20   Brief description of the annexed drawings

        The invention will now be described, by way of example only, with reference to the enclosed figures of drawing, wherein:

25   - Figure 1 shows a typical network scenario adapted for use of the arrangement described herein;

        - Figure 2 shows a networks scenario detailing the typical paths used by test traffic;

        - Figure 3 is a basic block diagram of the

30   arrangement described herein;

        - Figure 4 illustrates pre-configuration of the network elements considered in the foregoing;

        - Figure 5 presents a possible embodiment whereby, on the basis of the topology of the customers

35   configured, the arrangement configures automatically a

11

minimum number of measurements;

    - Figure 6 is a flowchart of an operating cycle in the arrangement described herein;

    - Figure 7 is a flowchart of a control step in the operating cycle considered in the foregoing;

    - Figure 8 is a further flowchart related to an analysis step in the operating cycle considered in the foregoing;

    - Figure 9 illustrates an implementation of an asynchronous mechanism of detection of degradation; and

    - Figure 10 is a flowchart of an updating step in the operating cycle considered in the foregoing.

Detailed description of preferred embodiments of the invention

    A first aspect to be analysed in the arrangement described herein relates to the process of configuration of the measurements between apparatuses present in the various sites of a Virtual Private Network customer. In this connection, bearing in mind that the system is directed typically to a Service Provider, the aim is to prevent having to insert a probe apparatus in each customer site. Instead, by keeping the aspects of analysis and control centralized, the aim is to exploit the functions offered by the apparatuses themselves, such as, for example, the Service Assurance Agent (SAA) commercially available with Cisco and the MIBs (Management Information Bases) developed by the IETF DISMAN working group, i.e. the Ping MIB, the Traceroute MIB, the Script MIB and the Scheduled MIB. Alternatively, the same functions can be provided via specific hardware components in the form of so-called "probes".

12

Contrary to the manual configuration of the set of measurements envisaged by some prior art solutions, the arrangement described creates automatically and in a co-ordinated way the appropriate set of performance

5   tests to be executed by the various apparatuses, thus avoiding manual configuration. This function is performed by exploiting the information available on the topology of the connections between the various sites and by accessing the databases of the Service

10 Provider, which maintains the information regarding the network apparatuses (frequently designated as Network Inventory).

Once again in the context of reference, it is not advantageous to condition (i.e. relate) the creation of

15 the tests to the networks with a greater amount of traffic. This aspect is illustrated by the example shown in Figure 2. This figure shows a first site 200, with networks (i.e., IP networks represented by the corresponding address prefixes) represented by a first

20 and second LAN, 202 and 204, respectively and a second site 210, with a third LAN 212. In such a case, if the measurements are configured on the basis of the criterion of the networks to which the majority of the traffic is directed, it would typically be necessary to

25 configure on a first router 206 tests towards the LAN 212 and on a second router 214 tests towards the LAN 202 and the LAN 204.

However, using this criterion, not only would the potential tests to be conducted increase considerably,

30 but, in the calculation of performance, also segments of network internal to the LAN of the customer sites would be considered, a fact that lies outside the purpose of the arrangement described.

The architecture of the proposed arrangement is

35 illustrated in the schematic logic diagram represented

13

in Figure 3. A centralized system 300 gathers statistics for the purpose of characterizing from the standpoint of performance the reachability and the quality of each connection path between the customer

5 sites (Customer Edges). When detecting a problem on a network, the system 300 can operate immediately by issuing, to the apparatuses, appropriate instructions for choosing, for example, from among the routes received from the other network.

10 The arrangement examines the performance between sites by carrying out tests between the edge routers and by associating the same performance to all the internal networks of a site. This approach enables a reduction in the number of measurements. This is a fact

15 of no little importance since, given the requirements of performance, the tests must be carried out as much as possible simultaneously on the various backbones, as will be illustrated hereinafter.

In the context of optimisation of the set of

20 measurements, it is expedient to adapt the set of tests to the Virtual Private Network type of the customer. For this reason, the arrangement described is able to activate different sets of measurements, according to whether the Virtual Private Network of the customer

25 has, for example, a "Hub & Spoke" or else "Full Mesh" topology, in order to optimise the testing process.

In the case of a Virtual Private Network of a "Hub & Spoke" type, where all the communications between the multi-backbone sites are directed to or travel from a

30 central site, it is convenient to limit the test to the stretches between the central site and each generic site i. This option thus envisages configuring on the central site the measurements directed to all the other sites (with possible problems of performance linked to

35 the load of the routers of the central site), or else

14

configuring from the remote sites the measurement
directed to the central site.

5

In the case of "Full Mesh" Virtual Private
Networks, instead, it is advantageous to test the end-
to-end connections, configuring a complete mesh of
measurements. In this latter case, or in situations of
Virtual Private Network topologies with incomplete-mesh
connections ("Partial Mesh" connections), it is
possible to use a method of automatic construction of
the measurements that guarantees completeness and, at
the same time, optimisation of the test set to be
performed. A specific exemplary algorithm for that
purpose will be introduced later in connection with
Figure 5.

10

15

In the context of control of the routing of the
test traffic, i.e., that of Multi Protocol Label
Switching Virtual Private Networks, there is the
possibility of a complete control of the two directions
of communication, via the configuration of the routing
policies implemented by both of the sites involved in
the communication. There is thus the possibility of
making estimations on the connectivity, and on the
corresponding performance, referred to a routing path
that selectively involves a particular backbone, both
out and back (round-trip).

20

25

Referring to Figure 2, it therefore appears
expedient to configure the tests in such a way that the
traffic between the site A 200 and the site B 210
always travels on the same backbone 220, in both
directions (Path A 240). So the backbone 230 is not
utilized for the communications between the sites in
question (200, 210). This, in particular, enables an
analysis of the behaviour of one backbone at a time,
and, since it is consequently possible to report
precisely the results of the performance tests, it

30

35

15

enables possible corrective actions to be taken in a co-ordinated and targeted way.

In order to gain control over routing of the test traffic, on each router there are configured (starting from a public address space of the provider, which is assumed as being dedicated to this purpose) two addresses that are advertised respectively on the two backbones. These addresses constitute the targets of the measurements made using the Service Assurance Agent.

Moreover, in the context of company Virtual Private Networks coming under a number of backbones, as addressed by the system in question, the solutions adopted fully exploit the capability of acting on both end points of communication in order to control routing policy. In fact, the sites situated at both ends of a communication path belong to the same administrative domain; this makes it possible to influence both the path of the incoming traffic and the path of the outgoing traffic of the customer sites through control of the routing policies on the edge routers (CEs) located at the two ends of the communication. Furthermore, this possibility is maintained also by performing reconfigurations on the edge routers of the backbone of the Service Provider (PEs).

The arrangement considered sets itself, then, the target (see Figure 1) of supplying to a given first site, indicated 50, a service capable of guaranteeing an efficient communication with a certain second site 2, indicated 52,, namely:

    - on the one hand, an efficient transmission of the outgoing traffic, directed towards the destination 52,; and

    - on the other hand, an efficient reception of the incoming traffic, coming from the site 52.

16

In order to achieve the above purpose, it is possible to act on the edge router of a given site and influence:

- i) the outgoing traffic, using the Border

5   Gateway Protocol attribute "Local Preference";

- ii) the incoming traffic, using the Border Gateway Protocol attribute "Autonomous System-Path length" (via the so-called pre-pend action, i.e., increasing fictitiously the list of the Autonomous

10 Systems in the Autonomous System-Path for the routes announced to the backbone that are not to be preferred).

In this way, in effect, for the traffic leaving said site, the desired result is obtained. However, for

15 the incoming traffic, the problem is that the action ii) entails the following limitations:

- in general, on the Internet, this action does not guarantee that the traffic will follow the desired path, because the Autonomous Systems traversed could,

20 in turn, alter fictitiously the Border Gateway Protocol attribute "Autonomous System-Path length", or else apply policies based upon the attribute of "Local Preference", (which has priority over the "Autonomous System-Path length" in the Border Gateway Protocol

25 method for path choice);

- in addition, in relation to the specific context of interest for the solution described herein (MPLS VPN offered by the Service Provider via multiple connection to its own backbones) the action ii) affects all of the

30 incoming traffic in the given site and, consequently, it would also have the undesirable effect of re-routing the flows of traffic coming from other sites (other than Site 2) to the given Site 1 (hence, it would not enable optimisation of the relations of traffic in a

selective way on the basis of the choices of the decision method).

Conversely, the preferred embodiment of the arrangement described herein envisages acting
5 exclusively on the traffic leaving the customer site adopting the following method (see Figure 1).

For the flow from the site 52 to the site 50, instead of acting on the site 50, by controlling the incoming traffic by increasing the "Autonomous System-
10 Path length" by means of the pre-pend technique, the arrangement will act on the traffic leaving the site 52, by configuring the attribute "Local Preference" at a value that is greater for the routes corresponding to the site 50, learned from the preferred backbone.
15   As an alternative to the technique of acting on the values of the Border Gateway Protocol attribute "Local Preference" corresponding to the various networks of destination of the traffic, on the various connections/backbones, the arrangement described also
20 envisages another technique for exclusive control of the traffic leaving the customer site. In fact, using this second option, it is possible to act on the PE routers, under which the sites come, modifying the Border Gateway Protocol attribute Multi-Exit
25 Discriminator (MED). This second way is useful in a context of CEs unmanaged by the Service Provider.

Finally, thanks to the criteria of co-ordination provided by the arrangement, which is able to correlate the results of the measurements with the Virtual
30 Private Network ,topology, it is possible to manage re-routing of communications, by modifying the routing policies applied in a site (e.g., the site 50 in Figure 1) also on the basis of the indications of the measurements made starting from the remote site (e.g.,
35 the site 52 in Figure 1).

18

The arrangement in question meets the requirement of carrying out the measurements in parallel on the various backbones, on the one hand, with an appropriate configuration of the tests (probes) performed via e.g.

5 Service Assurance Agent. On the other hand, it operates with a corresponding architecture that exploits creation in parallel, wherever advisable, of different processes/threads.

The arrangement described is directed at meeting

10 stringent performance requirements. In the event of marked degradation of performance, limiting intervention to carrying out appropriate measurements periodically may prove insufficient for reacting within the times envisaged. For this reason, the system

15 implements a dual mechanism of detection of degradation of performance: a synchronous-polling mechanism of a periodic type, and a mechanism of asynchronous signalling by the apparatuses. The latter mechanism is obtained by exploiting the possibility of creating

20 asynchronous events (e.g. Simple Network Management Protocol traps) when given thresholds have been exceeded within the configurations of the probes (e.g. Service Assurance Agent).

The architecture of the arrangement described

25 herein and represented schematically in Figure 3 envisages a centralized controller 300 (coming under the various backbone networks), which gathers statistics on reachability and connection quality between the routers of the various sites of the

30 customer's Virtual Private Network and, when it detects a problem on a network, can operate immediately by instructing the apparatuses involved for choosing from amongst the routes received by a backbone other than the one in use. The centralized controller 300 will

19

reach the apparatuses by exploiting the interface
connected to the best backbone at a given moment.

Consequently, the arrangement described herein, on
the one hand, carries out functions of analysis and
5 updating, and, on the other hand, drives the control
module. In order to obtain the data necessary for the
processing operations, the control module will exploit
the functions of the apparatuses in the customer site,
thus preventing the need to envisage specific hardware.
10 The system is thus made up of three main sub-
systems, which correspond to three main activities:
      - a subsystem, 310, responsible for the activity
of providing the service and collecting user
information
15   - a subsystem, 320, responsible for the activity
of configuring and provisioning network apparatuses
      - a subsystem, 330, responsible for the continuous
activity.

The subsystem 310 has the purpose of interfacing
20 with the user 340 and making available, through a
graphic interface 342 having associated a check module
343, functions for reading the configuration
corresponding to a customer/site/router, modifying the
service parameters for sites/routers already present in
25 the system, erasing routers/sites/customers, and for
the entry of routers/sites/customers. The user
information will be entered into a database and, by
means of appropriate mechanisms of syntactic and
semantic validation, may be used for the step of
30 configuration of the network apparatuses. Some of the
databases used in this step are the repositories used
by Service Providers for keeping the network inventory.

In the case where the administrator modifies the
configuration or else some router needs to be re-
35 configured, the subsystem 320 may have a character of

20

on-demand service of modification of the system-
configuration parameters, e.g. customers, devices,
topology, performance thresholds, QoS requirements.

In this step, starting from the information
5   contained in the various databases, the system
configures, on the network apparatuses, whatever is
necessary for activation of the Intelligent Routing
Control monitoring/control (for example, loop back
interfaces and corresponding addressing operations,
10 route-maps for outgoing and incoming advertisements,
measurements, etc.). The subsystem is equipped with a
method for optimisation of the set of measurements to
be made on the basis of the topology configured by the
customers via the subsystem 310.

15   The subsystem 330 includes a runtime block 331 and
implements an operating step that is basically made up
of the following three modules:

- a Control Module 332: this carries out
operations of monitoring the different connectivity
20 paths between sites and deriving therefrom measurements
of at least one of use and performance;

- an Analysis Module 334: this selects the best
paths on the basis of the data registered by the
preceding module and the policies established by the
25 customer or by the operator; and

- an Updating Module 336: on the basis of what is
decided by the Analysis module 334, this imparts the
directives of customer-router updating necessary for
modifying its routing decisions.

30   Consequently, with the character of round-the-
clock service, the subsystem 330 periodically:

- activates all the measurements configured
previously;

- traces the performance data from the routers;

21

- analyses the measurements according to the methods and the parameters entered during the stage of configuration of the service; and

- re-configures routing, i.e., takes the necessary measures to improve the traffic-routing operations.

In the following a possible implementation is described of the modules previously designated by subsystem 320 and subsystem 330. In fact, these steps concentrate quite interesting elements of the exemplary arrangement described herein.

Reference will be made to the case of Cisco routers, and two backbones will be considered by way of example. The system will act on the Costumer Edges in the customer site. The modalities of dialogue can come about in different ways for the specific requirements. The mechanisms indicated in what follows are Telnet 350 and Simple Network Management Protocol 360. The solution can however be implemented also using routers of other manufacturers (such as Juniper) or other protocols different from Simple Network Management Protocol 360 and Telnet 350 for interfacing therewith (e.g., Common Open Policy Service, COPS or eXtensible Markup Language, XML).

In Figure 3, references 370, 380, 390 designate three databases collecting information items related to:

- configuration information of the Intelligent Routing Control (IRC) system (clients, times of intervention, sites, routers, topological links, interfaces and IPs used, thresholds, passwords,…) – database 370;

- measurements/tests configured by the system on the various apparatuses in order to prompt them whenever necessary – database 380; and

22

- templates of the SAA instances as used for
configuring the various tests (these are essentially
templates having default values for various user-
configurable parameters such as the duration of the
5 tests, protocol types, packet sizes,…) – database 390.

In the sequel of the description, reference will
be made to and use will be made of the following Cisco
mechanisms used in the embodiment proposed herein for
filtering advertisements and for applying policy-
10 routing techniques (similar mechanisms are envisaged
also for other technologies).

- Prefix-list: this is a mechanism for definition
and configuration of a filter on the basis of IP
addresses. In the embodiment described, it is used for
15 filtering the advertisements of the IP networks (also
referred to as "routes" and, in what follows, referred
to also as IP "networks") to be made through one or
more conditions that enable or prevent advertisement of
the addresses (or range of addresses) indicated in the
20 condition.

- Route-map: this is a mechanism that operates by
evaluating in sequence clauses/conditions to be
verified (e.g., one or more prefix-list clauses),
enabling specific actions to be taken using constructs
25 of the type: if <condition> then <action>. In the
embodiment described herein, it is used for deciding
exactly which IP networks/routes to announce on the
different backbones (block 320) and for making
selections on the advertisements received (updating
30 block 336 of subsystem 330).

As regards the aspects of pre-configuration linked
to the measurements, the subsystem 320 described in
Figure 4, in order to carry out its operations
correctly, needs to make certain pre-configurations on
35 the customer apparatuses that must be set under

control. The contexts involved regard the aspect of the
measurements and the aspect of the modalities whereby
re-routing operations are to be carried out. In order
to make the necessary measurements of quality of the
communication, while maintaining the capacity of
determining selectively the performance provided
through the different backbones, the use is envisaged
of separate loopback interfaces on the Customer Edges
100, dedicated to monitoring the path on the respective
network backbones 420 and 430. On the generic CE 100 in
the customer site, two different loopback addresses
402, 403, 404 must be configured, which will be
advertised in the Virtual Private Network, exclusively
on the respective backbones to which they are
associated, for the purposes of reachability for the
measurements. The system will:

    - select the two addresses starting from a range
of public addresses that the provider will dedicate to
this purpose;

    - configure on the apparatus in the customer site
the two loopback interfaces (with the addresses
previously selected), which are to be the targets of
the measurements on the two backbones 420 and 430 (for
example, the address L1 400 is the target address for
the measurements on Backbone 1 420);

    - configure appropriate route-maps and prefix-
lists for the loopback interfaces referred to above so
as to obtain the certainty that a given measurement
will traverse the desired backbone (for example, with
reference to the network scenario illustrated in Figure
4, configuring two prefix-lists, one for each loopback
address, which will enable passage of all the IP
addresses except for the loopback address, and then
configuring route-maps, which will prevent
advertisement of the addresses of a given loopback

24

address, such as, for example, the address L1 400, on the backbone that is not associated thereto, for example, Backbone 2 430).

Exemplary pseudo-code implementing the procedure
5 just described is reproduced below.

```
    interface L1
     ip address<ipaddress L1><netmask>
     !
10  interface L2
      ip address<ipaddress L2><netmask>
     !
    router bgp 100
    ...
15  redistribute connected
    neighbor<ipaddress Neighbor Backbone 1>remote-as 1
    neighbor<ipaddress Neighbor Backbone 1>route-map toBB1 out
    neighbor<ipaddress Neighbor Backbone 2>remote-as 2
    neighbor<ipaddress Neighbor Backbone 2>route-map toBB2 out
20  no auto-summary
     !
    ip prefix-list Loop_BB1 seq5 deny <ipaddress L2>
    ip prefix-list Loop_BB1 seq 333 permit 0.0.0.0./0 le 32
     !
25  ip prefix-list Loop_BB2 seq5 deny <ipaddress L1>
    ip prefix-list Loop_BB2 seq 333 permit 0.0.0.0./0 le 32
     !
    route-map toBB2 permit 15
     match ip address prefix-list Loop_BB2
30   match route-type local
     !
    route-map toBB1 permit 15
     match ip address prefix-list Loop_BB1
     match route-type local
35  !
```

As regards the aspects of pre-configuration linked to the re-routing operations, the subsystem 320 corresponding sets the network apparatuses involved so that the variations of configuration of the system
40 decided will have the desired effects in modifying the routings. In general, these modifications are aimed at handling the traffic between two generic sites of a Virtual Private Network, modifying the values of the Border Gateway Protocol parameters of the networks
45 announced by the routers of the different sites.

25

For the purposes of operation, it is advantageous to provide appropriate route-maps, used for manipulating the metrics regarding the incoming Border Gateway Protocol advertisements for given networks,
5 defined, in turn, via the configuration of appropriate prefix-lists.

The system then, in its normal operation, will populate the prefix-lists, each time entering the IP networks on the basis of the decisions regarding their
10 routing.

Consequently, for the purposes of operation of the system, the two route-maps associated to the respective backbones, are previously configured in such a way that it will be possible to modify, for example, the value
15 of the Border Gateway Protocol attribute "Local Preference" for all and exclusively the networks that will progressively be entered, in the different prefix-lists, during the updating step 336 of subsystem 330. The number of prefix-lists necessary will correspond,
20 then, to the number of different backbones on which re-routing is handled.

As regards the configuration of the performance-test probes, once the operations described above have been completed, the subsystem 310 corresponding,
25 configures, on the customer apparatuses to be handled, appropriate Service Assurance Agent measurements. The Service Assurance Agent, provided by the Internetwork Operating System (IOS) of Cisco routers is formally defined as an application-aware synthetic operation
30 that monitors network performance and can measure, between a Cisco router and any other apparatus (Cisco router, IP Host, etc.), a number of key metrics on the basis of modern Service Level Agreements. Amongst these metrics, there may be recalled response time,

26

availability, jitter, packet loss, and application performance.

The working principle of the Service Assurance Agent can be exemplified as follows:

i. artificial packets are generated by the Service Assurance Agent, which simulate line traffic;

ii. once these packets have arrived at the Target Device, by exploiting the Echo mechanism they are forwarded to the Source Router;

iii. once they have reached the Source Router, the Service Assurance Agent calculates the Round-Trip Time (RTT).

Amongst the measurements made available by the Service Assurance Agent, there is envisaged, in the first place, recourse to the mechanism of "Internet Control Message Protocol Echo", in order to evaluate the size of the possible packet loss in the network and determine the round-trip delay. Should it be deemed expedient to acquire indications on additional quantities, such as jitter, it is possible to use further mechanisms (e.g., Up Down Pair jitter) made available by the Service Assurance Agent. The measurements will typically be made between pairs of customer sites coming under the "dual-backbone" modality, and can be performed via the configuration of the appropriate Service Assurance Agent test probes on the Customer Edge routers present in said sites. The target addresses of the measurements will be the loopback interfaces configured previously. The modalities of configuration and activation of the measurements and the periodicity with which they will be executed must take into account, on one hand, targets fixed for the time of reaction of the system to any possible service deficiency, and, on the other hand

27

constraints in terms of test traffic generated and band occupation.

Since on one and the same apparatus there may be a number of measurements to be triggered, the exemplary
5 embodiment described uses the following technique: it pre-configures a measurement that will be defined as start-trigger; this measurement is to all effects a measurement of an Internet Control Message Protocol Echo type towards a non-reachable address (outside the
10 addressing space). In fact, by exploiting the function present in the Service Assurance Agent of conditioning execution of a measurement to the outcome of another measurement, referred to as start-trigger measurement, the implementation is configured so as to cause the
15 various test probes to start exclusively following upon a time-out occurring on the start-trigger measurement. This approach, on the one hand, guarantees the simultaneous execution of the measurements, and, on the other, it limits of the interactions between the
20 centralized system and the routers for activation of the various measurements (the number of operations is constant for each router irrespective of the number of measurements configured, so enabling a high level of scalability).
25   The set of test-probe measurements to be configured depends upon the topology of the Virtual Private Network of the customer considered.

The following is a description of a possible embodiment wherein, on the basis of the topology of the
30 customer Virtual Private Networks configured, the system automatically configures a minimum number of measurements. This is sufficient for guaranteeing monitoring of the performance with coverage of all the communication stretches.

28

Assume that inside the Virtual Private Network there exist N sites connected and consider as being known the topology of the connections between the sites. The aim is to configure just one measurement for

5   each pair of sites connected, evenly distributing (taking also into account the characteristics of the sites) the measurements that each site is to make.

Then, for each apparatus, a value is obtained that indicates (in proportion to the total of all the

10 measurements to be made) how many measurements can be entrusted to this apparatus. This value could be obtained as a function of a number of variables that can be measured statically on the apparatus, such as, for example, a value associated to the type of

15 platform, the hardware equipment (for example, the amount of RAM possessed by the device), and possibly of variables assigned manually on the basis of particular requirements of use of the sites.

An example of such a function could be the

20 following:

Function(router) = platform coefficient(router) * RAM(router) * coefficient of use(router).

As regards the coefficients of use, if, for example, it is intended to balance the measurements in

25 an equal manner for all the routers, the value zero may be assigned to all of said coefficients.

Finding the minimum number of measurements to be configured involves essentially the assignment of the measurements between sites starting from the visit of a

30 graph that stores the topology of the connections between the sites and the policies on the basis of which it is intended to assign the measurements to the sites. A site responsible for the measurements for each pair of sites is assigned. If between two sites the

35 measurements are not to be made, it is sufficient to

29

enter previously fictitious arcs connecting said sites,
so that the sites will be already connected, so
guaranteeing proper functioning thereof. Setting as N
the number of sites and indicating the i-th site as
site [i], with i ranging from 1 to N, the flowchart
that describes the method appears in Figure 5.

In Figure 5, reference numeral 500 indicates an
input step that performs a control operation for each
pair of linked sites (site [i], site [j]), between
which is not necessary to make measurements. In a step
502 every site that does not require a measurements is
marked. After that, in a decision step 504, a decision
is taken in order to control if every site is marked.
If the condition result true, in a step 506, the
procedure terminates. Conversely, if a number of sites
are not marked the procedure continues in a step 508.
Specifically, in the step 508 a site [k] is selected.
The site [k] is not marked and his Function(site [k])
is maximum. Furthermore, in a step 510 a site [j] is
selected if there is not a link between site [j] and
site [k] and the Function(site [j]) is minimum. The
site [k] is responsible for the measurements for the
pair site [j] and site [k], and in a step 512 a link is
created between these two sites. In a step 514 the
value of the Function(site [k]) is update to a value
equal to Function(site [k]) – the value of every single
link. In a further decision step 516 the connection
between the site [k] and every other sites is
evaluated. If the condition is true, the procedure
continues with a step 518 where the site [k] is marked.
On the other hand, if the condition is false the
procedure continues with a decision step 520 where the
connection between the site [j] and every other sites
is evaluated. If the condition is true the procedure
continue with the step 522 where the site [j] is

30

marked. After that, and in the case where the condition
is false, the procedure returns to the decision step
504.

In Figure 5, for each pair of sites there is thus
assigned the site responsible for the measurements for
that pair. Furthermore, the measurements are assigned
to the sites in proportion to the value of the
Function(site [i]). In the case where a site (because
it is particularly "important" or much used) needs to
be kept out of the mechanism of the measurements, it is
possible to set the coefficient of use at a value zero
and mark the site prior to assignment of the
responsibility of the measurements. In this way, it is
certain that the site will not make any measurement,
but this will not modify the behaviour of the method
(the coefficient of use having being set at 0).

A particular case of topology is that of Hub&Spoke
Virtual Private Networks: if the customer Virtual
Private Networks has a Hub&Spoke topology, the System
will configure probes from all the sites towards the
hub (or central site). In particular, if there are
spoke sites with a router coming under both backbones
it will be necessary to:

i. configure the start-trigger measurement;

ii. configure a measurement towards the loopback
for Backbone 1 (BB1) of the central site;

iii. configure a measurement towards the loopback
for Backbone 2 (BB2) of the central site;

iv. configure scheduling of the three measurements
(i.e., condition the execution of the measurements to
points 2 and 3 above at time-out of the start-trigger
measurement); and

v. save the parameters for configuration of the
measurements referred to in the previous points (e.g.,
ID of the measurements) in the database of the system,

so as to be able to use them subsequently in normal
operation.

In order to handle the case of spoke and hub sites
with a number of apparatuses it is possible to envisage
5 appropriate solutions.

Figure 6 presents a possible implementation at the
level of flowchart of the operations of Subsystem 330
in Figure 3.

For some operations, the various operations are
10 triggered in parallel.

As may be noted, the runtime step may be
considered as being split into three steps, highlighted
in Figure 6, with different requirements: whereas an
analysis part 610 is handled by a single process
15 (thread) for a given customer, control 600 and updating
620 parts are performed in parallel, by creating one
thread for each router 605. In Figure 6, reference 611
designates an update check that, in the case of a
positive outcome, leads from the step 610 to the step
20 620 and, in the case of a negative outcome, leads
directly to the final stop of the process.

During the control step 600, the system triggers
on the different routers 605 the measurements
configured previously and samples their results,
25 ascertaining that the values recovered refer in effect
to the latest measurements. Considering interacting
with the apparatus via Simple Network Management
Protocol, this can be readily obtained by storing the
SysUpTime of the various apparatuses, as shown in
30 Figure 7.

In Figure 7 a first step 700 implements a
GetSysUpTime operation. In a further step 710 the
operation Start the Trigger Probe is performed. In a
step 720 the system waits for the test completion.
35 After that in a step 730 Read Data operation is

32

performed. In a decision step 740 the condition
ProbeTime > SysUpTime is evaluated. If the condition is
true, in a step 750 the Save Data operation is execute
and the procedure is finished. Conversely, if the
5 condition is false, in a step 760 the Delete Data
operation is performed and in the next step 770 the
Send Warning operation is execute.

In the analysis step 610 of Figure 6 there is only
one main process assigned to each customer.

10   The strategies that can be adopted by the analysis
module 610 for making the decisions of reconfiguration
of the routing policies of the traffic leaving the
site, on the two backbones, can in general be
summarized in the main alternatives listed in what
15 follows.

- i) According to -- customer, for all the sites:
this entails processing indicators of performance
corresponding to the backbone, as a whole, regardless
of the individual site of origin and destination, for
20 example, the weighted average of the data of the
measurements on the parameter of interest over the
entire set of the "probes"; the indicator, which is one
for each backbone, will be used to decide whether to
direct all the traffic of the Virtual Private Network
25 sites of the customer onto the alternative backbone.

- ii) according to -- customer site: this entails
processing of performance indicators corresponding to
transit on the backbone of the traffic originated by
each customer site, for example, the weighted average
30 of the data of the measurements on the parameter of
interest over the set of the "probes" performed on the
Customer Edges of the site considered; said indicator,
which is one for each site, will be used to decide
whether to direct all the traffic of the individual
35 customer site onto the alternative backbone. This mode

33

can be applied, for example, to the case where the communication that it is intended to handle through the Intelligent Routing Control (IRC) system is between the different sites and a central site.

5   - iii) According to -- pair of customer sites: this entails processing performance indicators corresponding to the transit, on the backbone, of the traffic originated by each customer site A towards each of the other destination sites B(i), for example,

10   weighted averages of the data of the measurements on the parameter of interest over the set of the "probes" made on the Customer Edges A to the Customer Edges B(i); said indicators will be used to decide whether to direct the traffic of the communications between A and

15   B(i) onto the alternative backbone. This strategy enables highly selective control of the routings among the various sites and is the one to which reference is made in what follows and to which the pre-configurations given previously refer.

20   Figure 8 describes a flowchart with which the operating steps of a possible implementation of the analysis step of the subsystem 330 are summarized. The operations can be summed up as follows:

          - in steps 800 and 820, information is obtained on

25   the backbone active between two sites by accessing the database;

          - in a step 810 the performance value is calculated with reference to the cost function evaluated on the basis of the parameters entered in the

30   configuration;

          - in a step 830 the value of performance corresponding to the active backbone is compared with the absolute threshold contained in the database;

34

- if the value is below the threshold, no other controls or changes are performed in a step 870, which is in fact a fictitious step;

- if the value is above the threshold,

5   - in a step 840 a calculation is made of whether and in what percentage the backbone in stand-by would improve the performance; an example of function that may be used for said calculation could be the following:

10          $$D = \left| \frac{Sample\_BB1 - Sample\_BB2}{\max(Sample\_BB1, Sample\_BB2)} \right|$$

- in a step 850 the value obtained is compared with the comparative threshold of the given customer present in the database of the system;

- if it is above the threshold, in a step 15 860 a decision for re-routing is taken and signalled in the updating step.

The embodiment described can be further refined to take also into account the appropriateness of envisaging a mechanism of asynchronous notification, 20 capable of duly signalling any possible significant degradation in performance. This enables prompt reaction to situations such as the ones presented in Figure 9. The centralized system will then be equipped with an "Event Listener" mechanism 960 (see also figre 25 3). Upon reception of an event (e.g., of the Trap Simple Network Management Protocol or SysLog message type), the mechanism 960 will interrupt the normal flow of operations, triggering an immediate reaction/re-configuration.

30   Specifically, a diagram indicated 970 in figure 9 presents a possible time behaviour over time (abscissa scale) of an observed parameter such as e.g. the Round Trip Time (RTT - ordinate scale). The diagram in question represents an exemplary situation where the

35

parameter RTT - or more precisely, a significant number of samples thereof - exceeds a given threshold (e.g. 800 ms.) over a time interval A during which performance is significantly impaired.

5   As regards the updating module 336 described, this envisages acting exclusively on the outbound traffic from the customer site. This corresponds to the decision of acting exclusively on the outgoing traffic, because co-ordinating the configurations at the two

10   end-points of a communication it is possible to manage all the traffic of the different sites. The configuration of the end-points is thus modified in a co-ordinated manner. One possibility is to act on the CE routers 100 to modify the values of the Border

15   Gateway Protocol attribute Local Preference associated to the different connections/backbones, for the various networks for destination of the traffic.

As an alternative, the possibility exists of acting on the PE routers 102, under which the different

20   sites come, and modifying the parameter MED.

With reference to the case where it is necessary to act exclusively on the Customer Edges, the system executes the steps presented in Figure 10.

In a step 1200 the sites network of the customer

25   AAA is read. In a step 1202 the updates for the customer router are created, and in the steps 1204, 1206, 1208 and 1210 the router A, the router B, the router C, and the router D are respectively configured.

The set of commands are sent in parallel on the

30   different apparatuses through Telnet/CLI connections (or other types of protocol).

The detailed description refers, by way of example only, to the case of the Cisco router and considers two backbones. The mechanisms indicated in the detailed

35   description use, for interaction with the apparatuses,

36

the protocols Telnet and Simple Network Management Protocol.

The arrangement described herein can in any case be implemented also

5   - using routers of other manufacturers (e.g., Juniper);

- using other protocols different from Simple Network Management Protocol and Telnet for interfacing with the apparatus (e.g., COPS or XML);

10   - with a number of separate IP backbones that is greater than two.

Again it is recalled that the backbones in question may be either logically or physically distinct.

15   The system can operate both as a centralized system and as a local arrangement at the customer site.

Consequently, without prejudice to the underlying principles of the invention, the details and the embodiments may vary, also appreciably, with reference

20 to what has been described by way of example only, without departing from the scope of the invention as defined by the annexed claims.

CLAIMS

1. A method of fault and performance recovery in a
communication network including a set of network
apparatuses (100, 102, 104, 106), said network
including a plurality of sites (50, 52, 54, 56, 58)
connected by means of connections (20, 30) having a
given topology and adopting a given routing policy, the
network having associated at least one database (370)
maintaining information regarding said network
apparatuses, the method including the steps of:
    - executing at least one set of performance tests,
wherein said at least one set of performance tests is
created automatically by exploiting the information
available on the topology of the connections (20, 30)
between the various sites and by accessing said at
least one database (370),
    - analysing the results of said at least one set
of performance tests to detect possible fault and
performance impairment conditions of said network, and
    - in the presence of a fault and performance
impairment condition detected in said network,
modifying said routing policy to compensate for said
fault and performance impairment condition.

2. The method of claim 1, characterized in that
said at least one set of performance tests includes
performance tests between sites of said plurality (50,
52, 54, 56, 58), said tests being carried out between
edge routers (100) and by associating the same
performance to all the internal networks of a site.

3. The method of claim 1, wherein said network
includes at least two backbones (20, 30), the method

38

characterized in that said at least one set of performance tests includes performance tests carried out simultaneously on said at least two backbones (20, 30).

4. The method of claim 1, wherein said network includes at least two backbones (20, 30) and corresponding multibackbone sites (50, 52, 54, 56, 58), and wherein the communications between the multibackbone sites (50, 52, 54, 56, 58) are directed to or travel from a central site, the method characterized in that said performance tests are limited to the stretches from a site in said plurality and said central site.

5. The method of claim 1, wherein said network includes at least two backbones (20, 30) and corresponding multibackbone sites (50, 52, 54, 56, 58), and wherein said network is a full mesh network, the method characterized in that said performance tests involve all end-to-end connections (20, 30) in the network, thus configuring a complete mesh of performance tests.

6. The method of claim 1, wherein said network includes at least two backbones (20, 30) and corresponding multibackbone sites (50, 52, 54, 56, 58), and wherein said network is a partial mesh network, the method characterized in that said performance tests involve all end-to-end connections in a test subset in the network.

7. The method of claim 1, characterized in that it includes the steps of:
       - generating test traffic in said network,

39

- controlling the routing of said test traffic in both possible directions of communication, by configuring the routing policies implemented by both of the sites (200, 210) involved in the communication,
5 causing said test traffic between a first site (200) and a second site (210) of a given pair always to travel on the same backbone (220), in both directions;
- making estimations on the connectivity referred to a routing path that selectively involves a given
10 network backbone (20, 30), both out and back (round-trip).

8. The method of claim 7, applied to a network including at least two backbones (20, 30), the method
15 characterized in that it includes the step of gaining control over said routing of the test traffic, by configuring on each router (100) two addresses that are advertised respectively on said at least two backbones (20, 30), whereby said addresses constitute the targets
20 of the performance tests made.

9. The method of claim 1, characterized in that it includes the step of modifying the routing policy applied at a site (50) in the network based on
25 performance tests made starting from a remote site (52).

10. The method of claim 1, characterized in that it includes the step of implementing a mechanism of
30 detection of performance impairment by means of at least one of:
- a synchronous-polling mechanism of a periodic type, and
- a mechanism of asynchronous signalling by said
35 network apparatuses (100, 102).

40

11. The method of claim 10, characterized in that said asynchronous signalling is obtained by creating asynchronous events (960) in the network when given thresholds have been exceeded within predetermined
5   probe configurations.

12. The method of claim 1, applied to networks including at least two backbones (20,30), the method characterized in that it includes the steps of
10   performing in a centralized manner (300) the operations of:
        - gathering statistics on reachability and associated quality between the routers (100) of various sites in said plurality, and
15   - in the presence of a fault and performance impairment condition detected in said network, instructing the apparatuses involved for choosing from amongst the routes advertised by a backbone (20, 30) other than the one in use.
20

13. The method of claim 12, characterized in that it includes the steps of reaching said apparatuses instructed by exploiting the interface/address connected to the best backbone (20, 30) at a given
25   moment.

14. The method of claim 1, characterized in that it includes at least one of the steps of:
        - providing service and collecting user and
30   topology information by means of a first subsystem (310);
        - configuring and provisioning network apparatuses by means of a second subsystem (320); and
        - ensuring continuous network activity by means of
35   a third subsystem (330).

15. The method of claim 14, characterized in that it includes the step of providing via said second subsystem (320) on-demand modification of configuration parameters by configuring, starting from the
5 information contained in said at least one database (370), functions for performance monitoring and routing control.

16. The method of claim 15, characterized in that
10 said functions for performance monitoring and routing control includes at least one of: loop back interfaces and corresponding addressing operations, route-maps for outgoing and incoming advertisements, and performance tests.
15

17. The method of claim 14, characterized in that it includes the step of providing via said third subsystem (330) at least one of the operations of:
    - monitoring (332) the different connectivity
20 paths between sites, and deriving therefrom measurements of at least one of use and performance;
    - selecting (334) best paths in said network based on at least one of data obtained during said monitoring (332) and pre-established policies; and
25   - imparting (336) directives for router (100, 102) updating for modifying routing decisions based on selections made during said selecting (334) .

18. The method of claim 14, characterized in that
30 it includes the step of providing via said third subsystem (330) at least one of the operations of:
    - activating all previously configured performance tests;
    - tracing performance data from the routers (100,
35 102);

42

- analysing the performance tests according to
service configuration; and
- re-configuring routing.

19. The method of claim 1, characterized in that
it includes the step of optimising said set of
performance tests by automatically configuring a
minimum number of performance tests to guarantee
performance monitoring of said network with coverage of
all the communication stretches therein by configuring
only one measurement for each pair of sites (50, 52)
connected in said topology.

20. The method of claim 1, characterized in that
it includes the step of distributing the performance
tests to be performed at said sites (50, 52, 54, 56,
58) in said network by generating for each apparatus
involved a value representative of the performance
tests to be entrusted to said apparatus.

21. The method of claim 20, characterized in that
it includes the step of obtaining said value as a
function of at least one of a first set of variables
adapted to be measured statically on the apparatus, and
a second set of variables assigned based on
requirements of use of respective site (50, 52, 54, 56,
58).

22. The method of claim 21, characterized in that
said first set of variables includes variables selected
out of the group consisting of the type of platform,
the hardware equipment, the memory capabilities of the
apparatus.

43

23. The method of claim 1, applied to a network including at least two backbones (20, 30), the method characterized in that it includes the steps of reconfiguring the traffic leaving a site (50, 52, 54, 56, 58) over said at least two backbones (20, 30) according to customer, for all the sites associated with a given customer in said network.

24. The method of claim 23, characterized in that it includes the step of processing performance indicators corresponding to a backbone (20, 30) as a whole regardless of the individual site of origin and destination.

25. The method of claim 24, characterized in that said indicators include a weighted average of the data of the performance tests on the parameters of interest over a set of probes.

26. The method of claim 1, applied to a network including at least two backbones (20, 30), the method characterized in that it includes the steps of reconfiguring the traffic leaving a site (50, 52, 54, 56, 58) over said at least two backbones (20, 30) according to customer site, irrespective of any other site possibly associated to the same customer in said network.

27. The method of claim 26, characterized in that it includes the step of processing performance indicators corresponding to transit on a backbone (20, 30) of the traffic originated by each customer site.

28. The method of claim 27, characterized in that said indicators include a weighted average of the data

44

of the performance tests on the parameters of interest over a set of probes performed on the customer edges of the site considered.

5   29. The method of claim 1, applied to a network including at least two backbones (20, 30), the method characterized in that it includes the step of reconfiguring the traffic leaving a site (50, 52, 54, 56, 58) over said at least two backbones (20, 30)
10 selectively for different pairs sites associated with the same customer in said network.

        30. The method of claim 29, characterized in that it includes the step of processing performance
15 indicators corresponding to the transit, on a backbone (20, 30), of the traffic originated by each said customer site towards any other destination sites.

        31. The method of claim 30, characterized in that
20 said indicators include weighted averages of the data of the performance tests on the parameters of interest over the set of probes made on the customer edges (100) of each said customer site to the customer edges (100) of any other destination sites.
25
        32. A system for fault and performance recovery in a communication network including a set of network apparatuses (100, 102, 104, 106), said network including a plurality of sites (50, 52, 54, 56, 58)
30 connected by means of connections (20, 30) having a given topology and including routers (100) for implementing given routing policies, the network having associated at least one database (370) maintaining information regarding said network apparatuses, the
35 system including:

        - test modules (100, 102, 104, 106) for executing
at least one set of performance tests, wherein said at
least one set of performance tests is created
automatically by exploiting the information available
5   on the topology of the connections (20, 30) between the
various sites and by accessing said at least one
database (370), said test modules configured for
analysing the results of said at least one set of
performance tests to detect possible fault and
10 performance impairment conditions of said network, and
        - said routers (100) configured for modifying, in
the presence of a fault and performance impairment
condition detected in said network, said routing policy
to compensate for said fault and performance impairment
15 condition.

        33. The system of claim 32, characterized in that
said at least one set of performance tests includes
performance tests between sites of said plurality (50,
20 52, 54, 56, 58), said test modules (100, 102, 104,
106) configured for carrying out said tests between
edge ' routers (100) and by associating the same
performance to all the internal networks of a site.

25   34. The system of claim 32, wherein said network
includes at least two backbones (20, 30), the system
characterized in that said test modules (100, 102, 104,
106) are configured for carrying out said at least one
set of performance tests as a set including performance
30 tests carried out simultaneously on said at least two
backbones (20, 30).

        35. The system of claim 32, wherein said network
includes at least two backbones (20, 30) and
35 corresponding multibackbone sites (50, 52, 54, 56, 58),

46

and wherein the communications between the multi-backbone sites (50, 52, 54, 56, 58) are directed to or travel from a central site, the system characterized in that said test modules (100, 102, 104, 106) are
5   configured for carrying out performance tests limited to the stretches from a site in said plurality and said central site.

36. The system of claim 32, wherein said network
10   is a full mesh network, the system characterized in that said test modules (100, 102, 104, 106) are configured for carrying out performance tests involving all end-to-end connections (20, 30) in the network, thus configuring a complete mesh of performance tests.
15

37. The system of claim 32, wherein said network is a partial mesh network, the system characterized in that said test modules (100, 102, 104, 106) are configured for carrying out performance tests involving
20   all end-to-end connections in a test subset in the network.

38. The system of claim 32, characterized in that the system is configured for:
25   - generating test traffic in said network,
        - controlling via said routers (100, 102) the routing of said test traffic in both possible directions of communication, by configuring the routing policies implemented by both of the sites (200, 210)
30   involved in the communication, said test modules (100, 102, 104, 106) being configured for causing said test traffic between a first site (200) and a second site (210) of a given pair always to travel on the same backbone (220), in both directions, and

47

- making estimations on the connectivity referred to a routing path that selectively involves a given network backbone (20, 30), both out and back (round-trip).

5

39. The system of claim 38, applied to a network including at least two backbones (20, 30), the system characterized in that said routers (100) are configured for gaining control over said routing of the test
10 traffic, wherein on each router (100) two addresses are configured that are advertised respectively on said at least two backbones (20, 30), whereby said addresses constitute the targets of the performance tests made.

15   40. The system of claim 32, characterized in that said routers (100) are configured for modifying the routing policy applied at a site (50) in the network based on performance tests made starting from a remote site (52).
20
       41. The system of claim 32, characterized in that it includes a mechanism of detection of performance impairment by means of at least one of:
       - a synchronous-polling mechanism of a periodic
25 type, and
       - a mechanism of asynchronous signalling by said network apparatuses (100, 102).

       42. The system of claim 41, characterized in that
30 said asynchronous signalling is obtained by creating asynchronous events (960) in the network when given thresholds have been exceeded within predetermined probe configurations.

35   43. The system of claim 32, applied to networks including at least two backbones (20,30), the system

48

characterized in that it includes a centralized module (300) configured for performing the operations of:

- gathering statistics on reachability and associated quality between the routers (100) of various
5 sites in said plurality, and

- in the presence of a fault and performance impairment condition detected in said network, instructing the apparatuses involved for choosing from amongst the routes advertised by a backbone (20, 30)
10 other than the one in use.

44. The system of claim 43, characterized in that said centralized module (300) id configured for reaching said apparatuses instructed by exploiting the
15 interface/address connected to the best backbone (20, 30) at a given moment.

45. The system of claim 32, characterized in that it includes at least one of:
20   - a first subsystem (310) providing service and collecting user and topology information;

- a second subsystem (320) configuring and provisioning network apparatuses; and

- a third subsystem (330) ensuring continuous
25 network activity.

46. The system of claim 45, characterized in that said second subsystem (320) is configured for providing on-demand modification of configuration parameters by
30 configuring, starting from the information contained in said at least one database (370), functions for performance monitoring and routing control.

47. The system of claim 46, characterized in that
35 said functions for performance monitoring and routing

49

control includes at least one of: loop back interfaces
and corresponding addressing operations, route-maps for
outgoing and incoming advertisements, and performance
tests.

48. The system of claim 45, characterized in that
said third subsystem (330) is configured for performing
at least one of the operations of:
    - monitoring (332) the different connectivity
paths between sites, and deriving therefrom
measurements of at least one of use and performance;
    - selecting (334) best paths in said network based
on at least one of data obtained during said monitoring
(332) and pre-established policies; and
    - imparting (336) directives for router (100, 102)
updating for modifying routing decisions based on
selections made during said selecting (334).

49. The system of claim 45, characterized in that
said third subsystem (330) is configured for performing
at least one of the operations of:
    - activating all previously configured performance
tests;
    - tracing performance data from the routers (100,
102);
    - analysing the performance tests according to
service configuration; and
    - re-configuring routing.

50. The system of claim 32, characterized in that
said test modules (100, 102, 104, 106) are configured
for optimising said set of performance tests by
automatically configuring a minimum number of
performance tests to guarantee performance monitoring
of said network with coverage of all the communication

50

stretches therein by configuring only one, measurement for each pair of sites (50, 52) connected in said topology.

5   51. The system of claim 32, characterized in that said test modules (100, 102, 104, 106) are configured for distributing the performance tests to be performed at said sites (50, 52, 54, 56, 58) in said network by generating for each apparatus involved a value

10 representative of the performance tests to be entrusted to said apparatus.

    52. The system of claim 51, characterized in that said test modules (100, 102, 104, 106) are configured

15 for obtaining said value as a function of at least one of a first set of variables adapted to be measured statically on the apparatus, and a second set of variables assigned based on requirements of use of respective site (50, 52, 54, 56, 58).

20
    53. The system of claim 52, characterized in that said first set of variables includes variables selected out of the group consisting of the type of platform, the hardware equipment, the memory capabilities of the

25 apparatus.

    54. The system of claim 32, applied to a network including at least two backbones (20, 30), the system characterized in that said routers (100) are configured

30 for reconfiguring the traffic leaving a site (50, 52, 54, 56, 58) over said at least two backbones (20, 30) according to customer, for all the sites associated with a given customer in said network.

51

55. The system of claim 54, characterized in that the system is configured for processing performance indicators corresponding to a backbone (20, 30) as a whole regardless of the individual site of origin and destination.

56. The system of claim 55, characterized in that said indicators include a weighted average of the data of the performance tests on the parameters of interest over a set of probes.

57. The system of claim 32, applied to a network including at least two backbones (20, 30), the system characterized in that said routers (100) are configured for reconfiguring the traffic leaving a site (50, 52, 54, 56, 58) over said at least two backbones (20, 30) according to customer site, irrespective of any other site possibly associated to the same customer in said network.

58. The system of claim 57, characterized in that the system is configured for processing performance indicators corresponding to transit on a backbone (20, 30) of the traffic originated by each customer site.

59. The system of claim 58, characterized in that said indicators include a weighted average of the data of the performance tests on the parameters of interest over a set of probes performed on the customer edges of the site considered.

60. The system of claim 32, applied to a network including at least two backbones (20, 30), the system characterized in that said routers (100) are configured for reconfiguring the traffic leaving a site (50, 52,

52

54, 56, 58) over said at least two backbones (20, 30) selectively for different pairs sites associated with the same customer in said network.

5   61. The system of claim 60, characterized in that the system is configured for processing performance indicators corresponding to the transit, on a backbone (20, 30), of the traffic originated by each said customer site towards any other destination sites.

10

62. The system of claim 61, characterized in that said indicators include weighted averages of the data of the performance tests on the parameters of interest over the set of probes made on the customer edges (100)

15 of each said customer site to the customer edges (100) of any other destination sites.

63. A communication network including:

- a set of network apparatuses (100, 102, 104,

20 106),

- a plurality of sites (50, 52, 54, 56, 58) connected by means of connections (20, 30) having a given topology and adopting a given routing policy,

- at least one associated database (370)

25 maintaining information regarding said network apparatuses,

- the network having associated a system according to any of claims 32 to 72.

30   64. A computer program product loadable in the memory of at least one computer and including software code portions for performing the method oaf nay of claims 1 to 31.
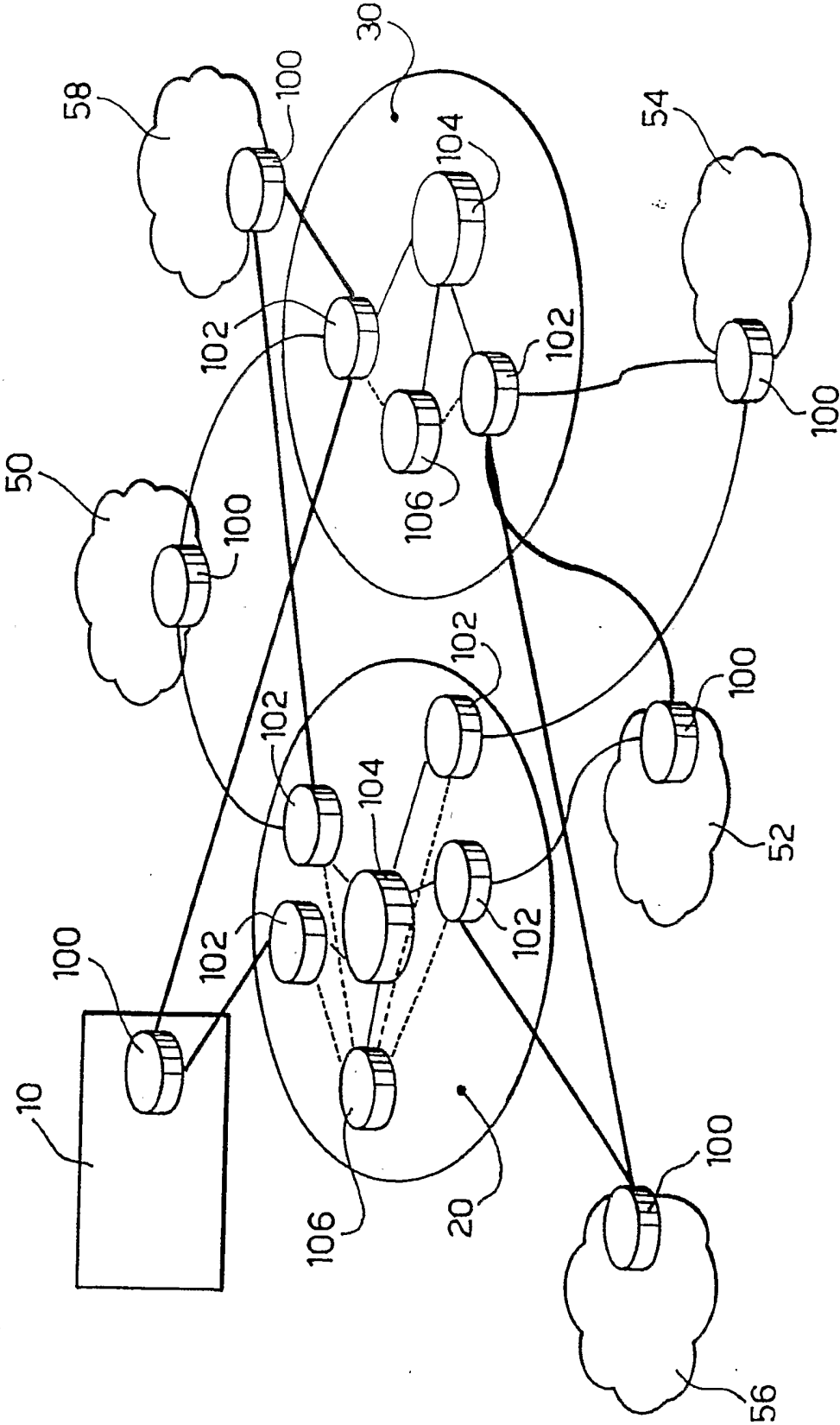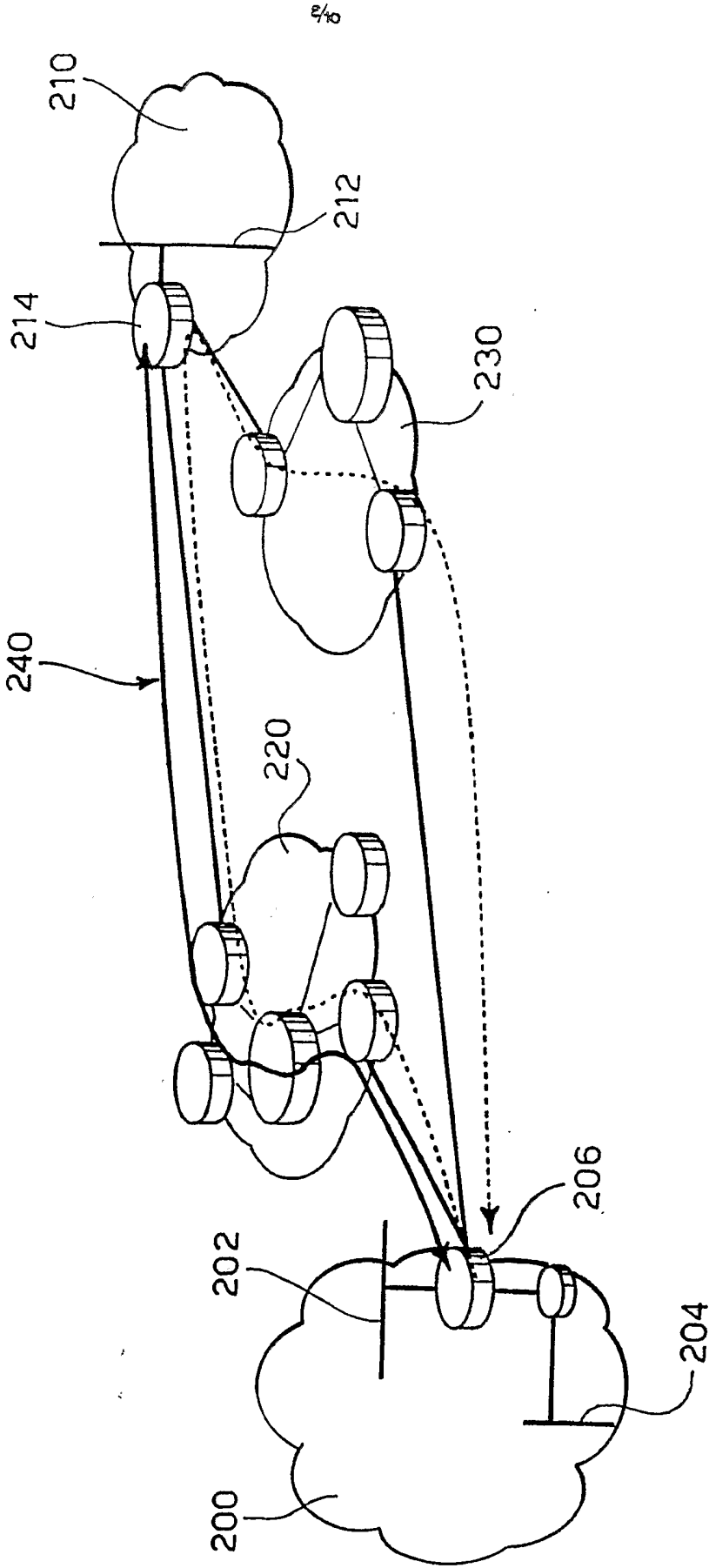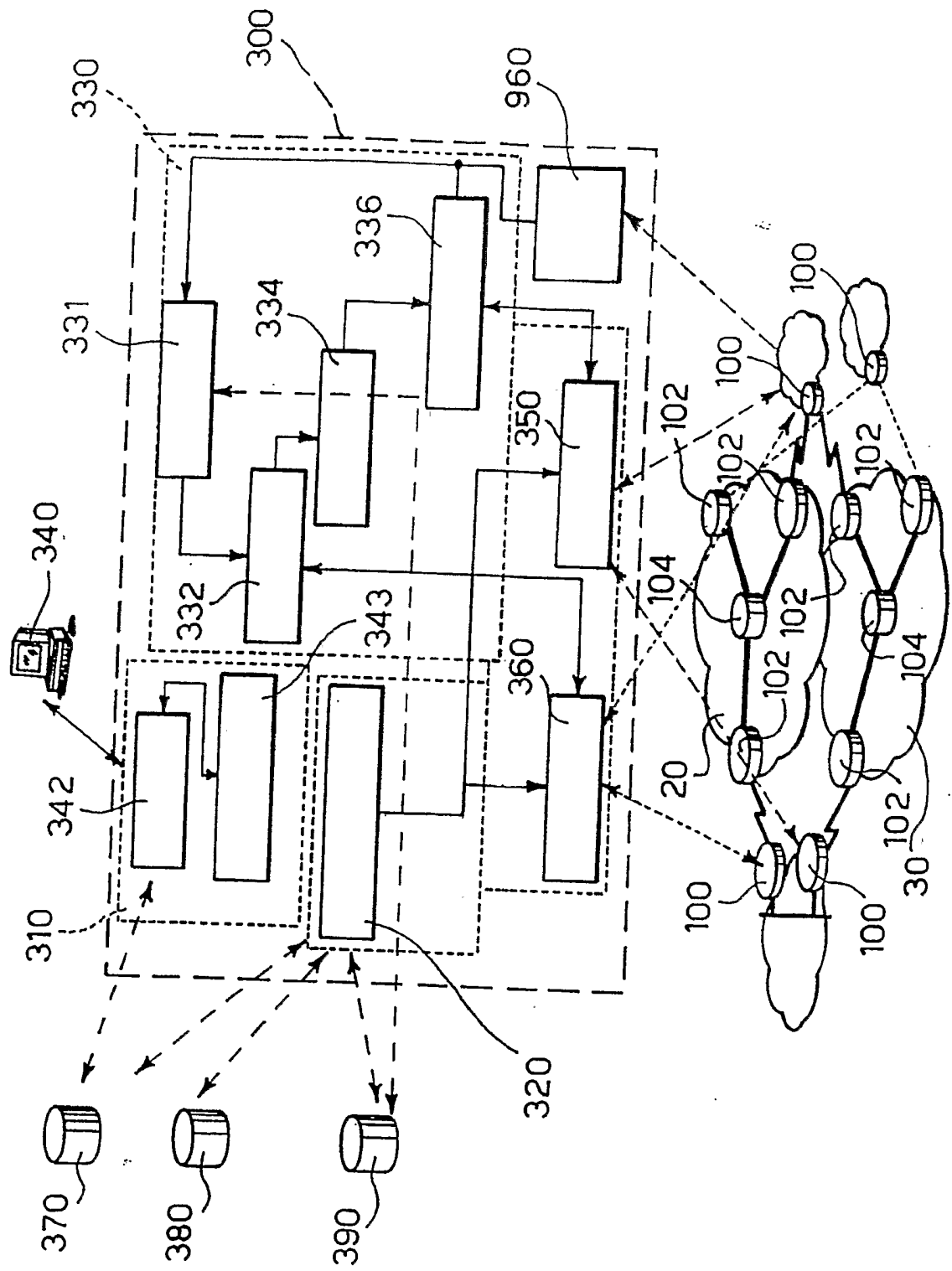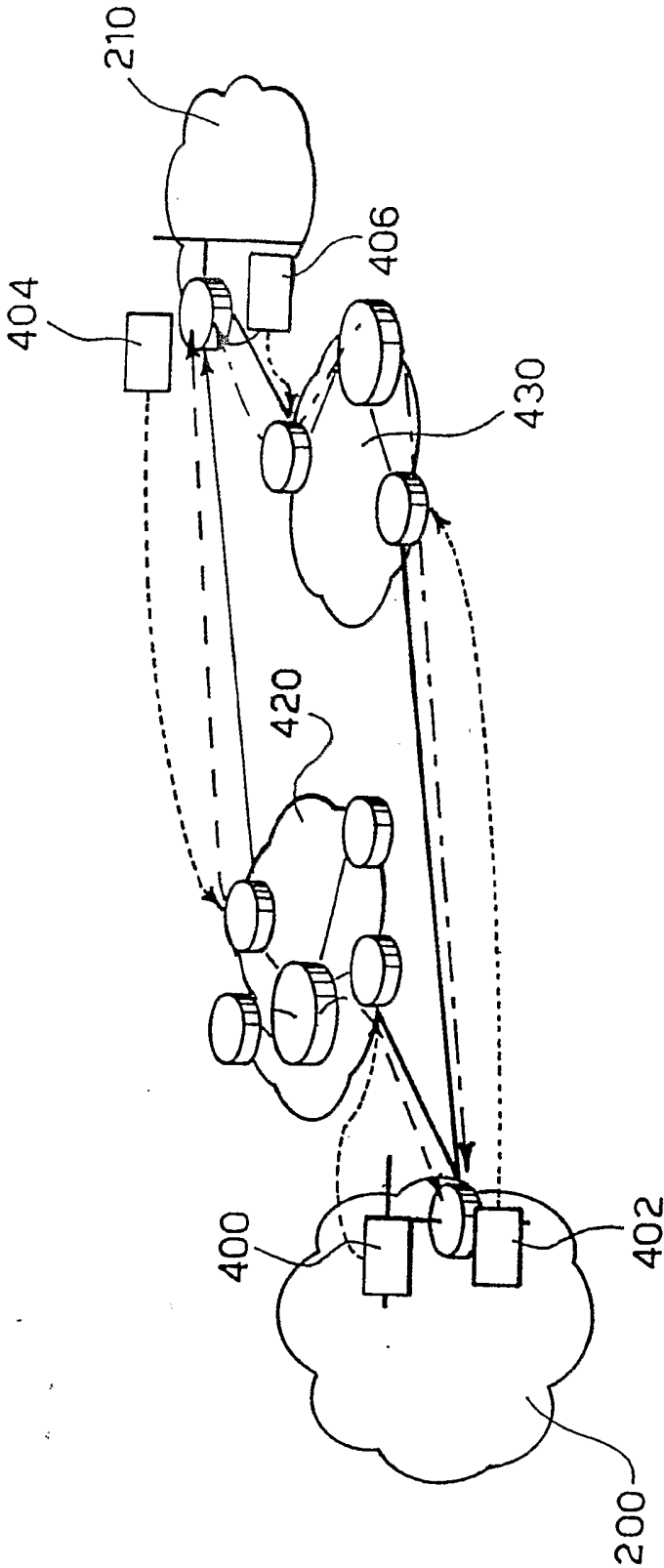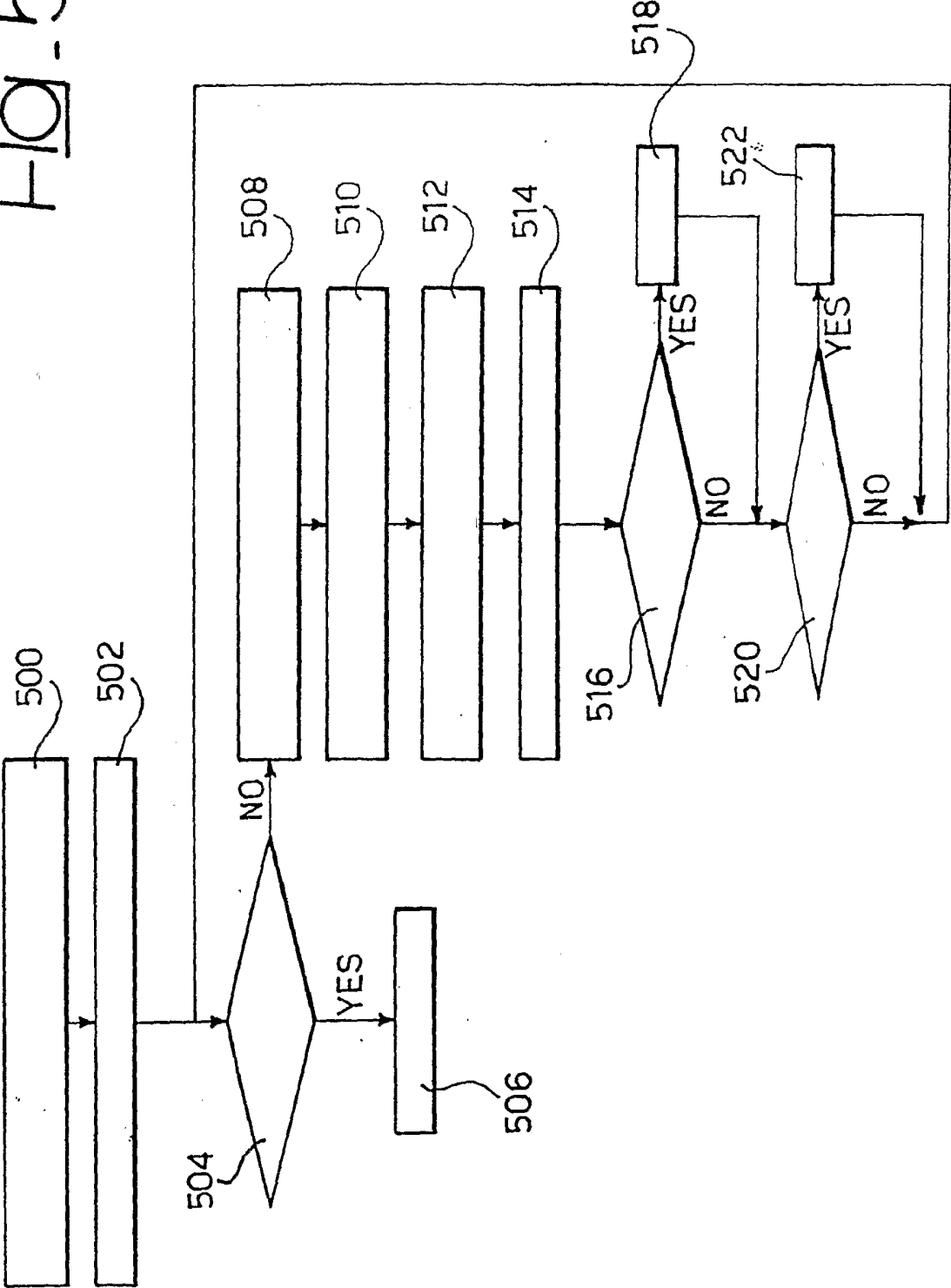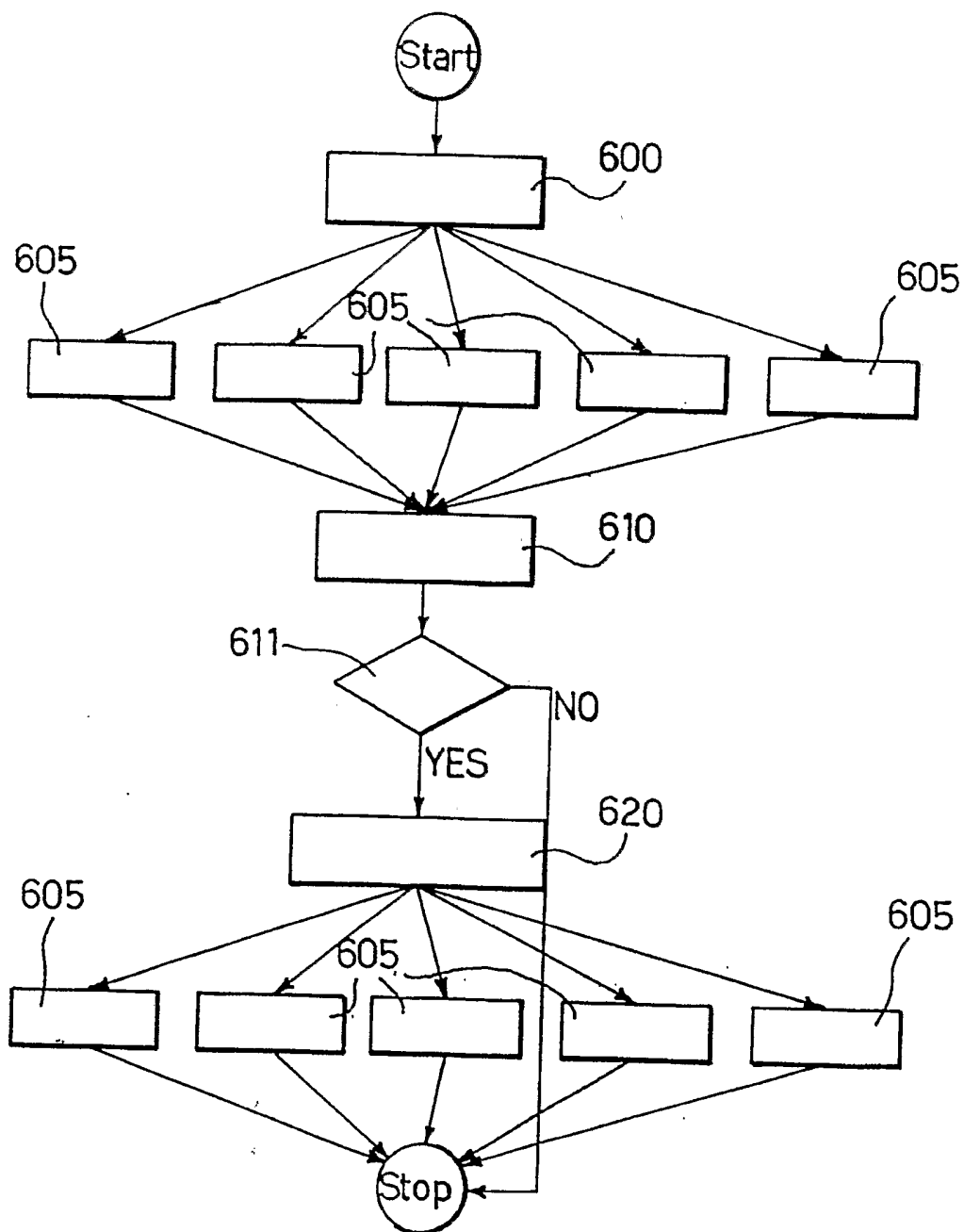
1/10

Fig_1

2/10

Fig. 2

3/10

Fig-3

Fig. 4

Fig.5

6/10

Fig.6

Fig. 7
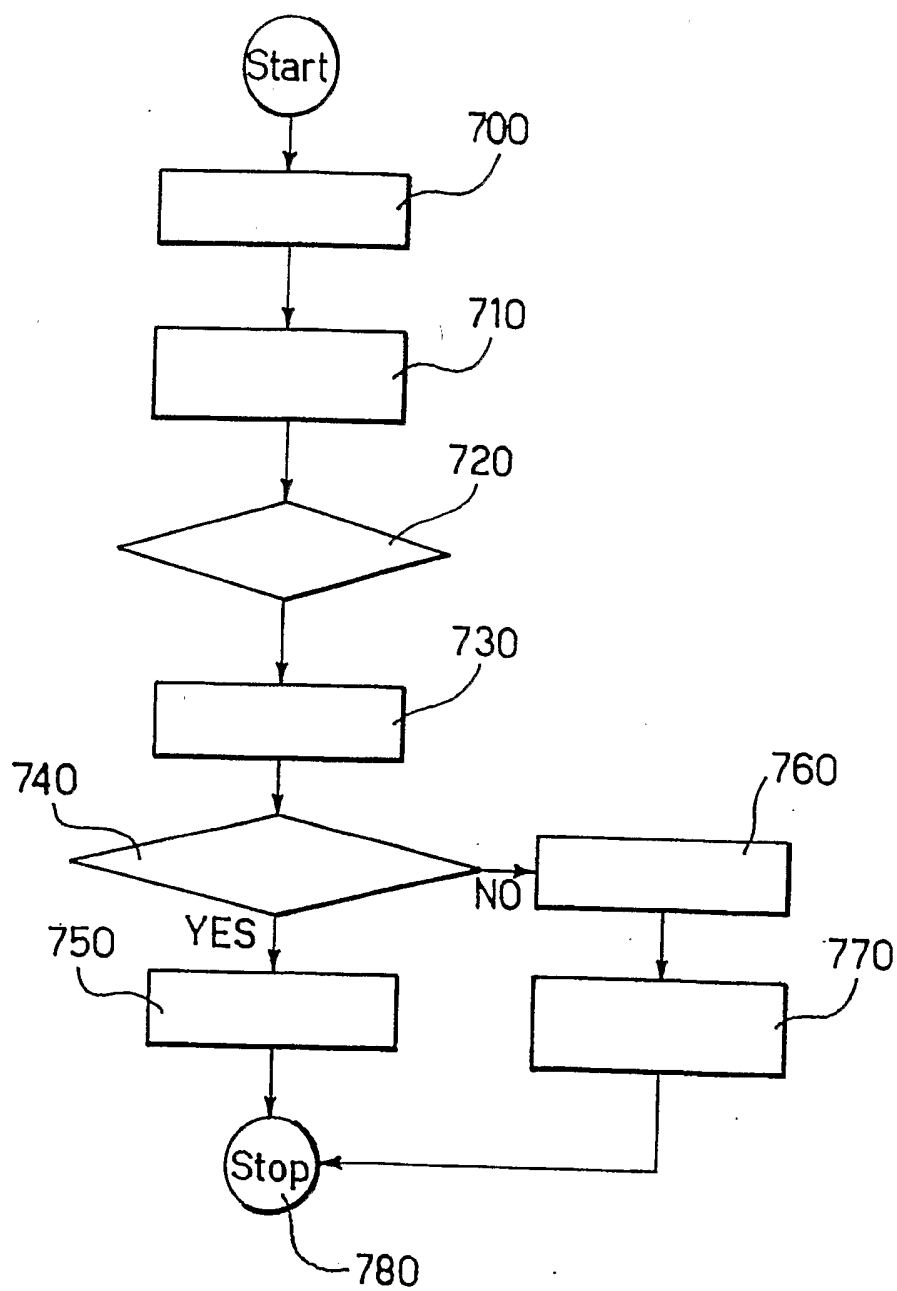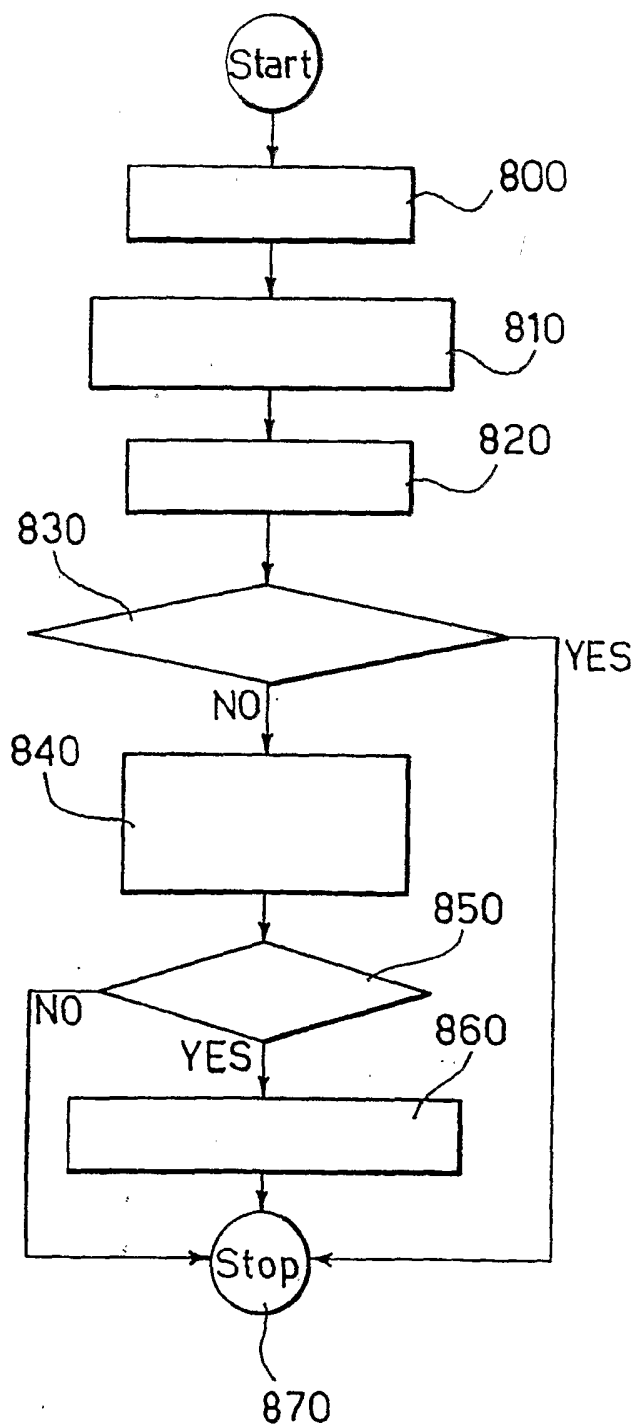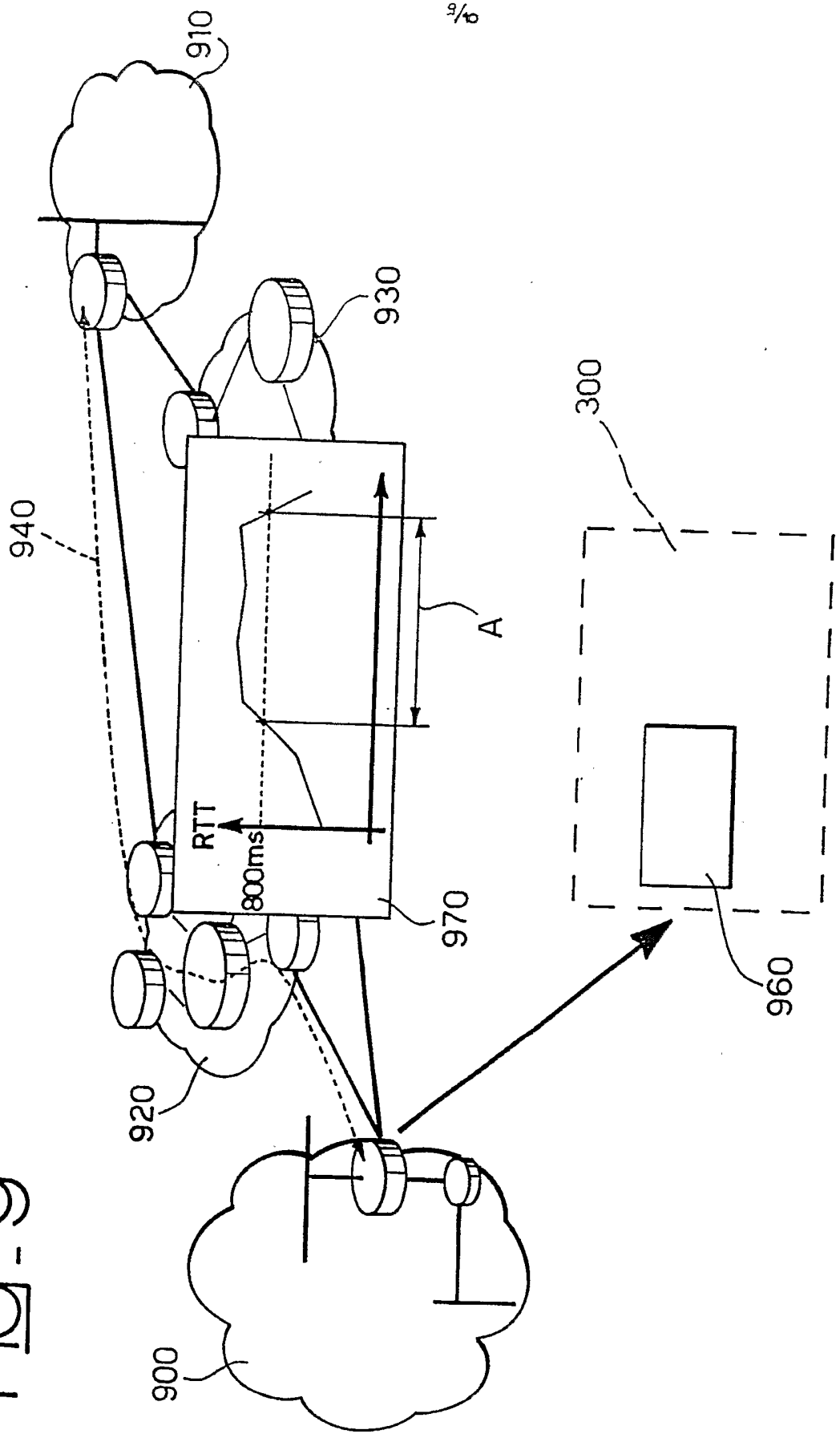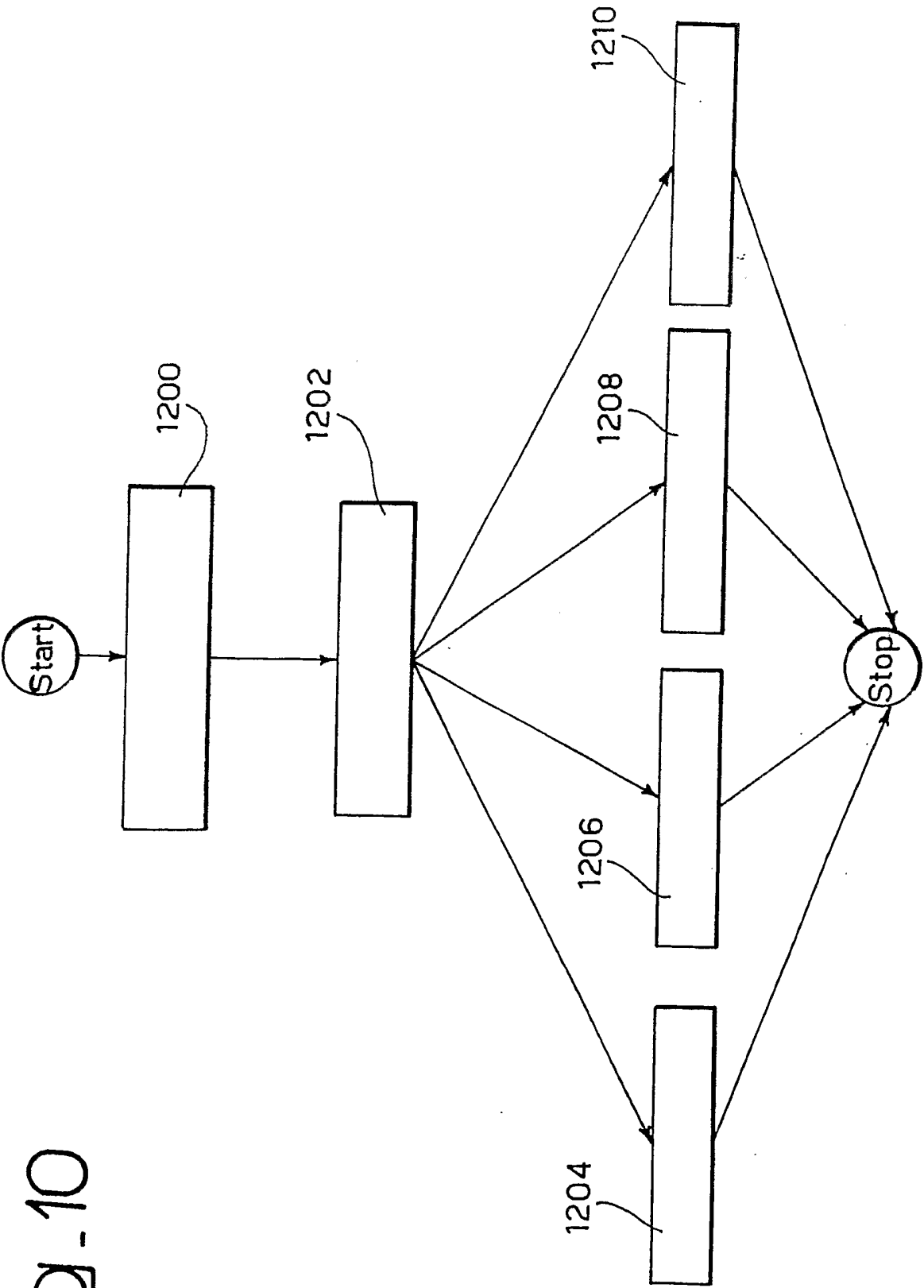
Fig - 8

9/10

Fig. 9

10/10

Fig - 10

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   H04L12/26   H04L12/24   H04L12/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   H04L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 2002/087680 A1 (CERAMI RICHARD S ET AL) 4 July 2002 (2002-07-04) paragraphs '0024!, '0026!, '0041! - '0052!, '0066!, '0071! - '0086!; figures 1-5	1,32,63, 64 2-31, 33-62
Y	WO 02/080462 A1 (PROFICIENT NETWORKS, INC; BAYS, ROBERT, JAMES; PINSKY, BRUCE, ERIC; LE) 10 October 2002 (2002-10-10) cited in the application abstract page 4, line 7 - page 6, line 9 page 20, line 15 - page 22, line 13	1-64

-/--

[X] Further documents are listed in the continuation of box C.	[X] Patent family members are listed in annex.

° Special categories of cited documents :

'A' document defining the general state of the art which is not considered to be of particular relevance

'E' earlier document but published on or after the international filing date

'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

'O' document referring to an oral disclosure, use, exhibition or other means

'P' document published prior to the international filing date but later than the priority date claimed

'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

'&' document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 February 2005	28/02/2005

Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Authorized officer Ceccarini, G

4

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	THOMAS D. NADEAU ET AL: "OAM Requirements for MPLS Networks draft-ietf-mpls-oam-requirements-02.txt" IETF INTERNET DRAFT, June 2003 (2003-06), XP002316305 the whole document	1-64
Y	DINESH MOHAN ET AL: "VPLS OAM Requirements and Framework draft-mohan-sajassi-l2vpn-oam-req-frmk-00.txt" IETF INTERNET DRAFT, June 2004 (2004-06), XP002316306 the whole document	1-64
A	E. ROSEN ET AL: "BGP/MPLS VPNs" IETF RFC 2547, March 1999 (1999-03), XP015008330 cited in the application the whole document	1-64

4

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2002087680	A1	04-07-2002	AU	7913001 A	13-02-2002
			WO	0210944 A1	07-02-2002
			US	2002111883 A1	15-08-2002
			US	2002073062 A1	13-06-2002
			US	2002078017 A1	20-06-2002
			US	2002099841 A1	25-07-2002
			US	2002073355 A1	13-06-2002
			US	2002071440 A1	13-06-2002
WO 02080462	A1	10-10-2002	US	2002141378 A1	03-10-2002
			US	2002141343 A1	03-10-2002