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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 February 2002 (28.02.2002)

PCT

(10) International Publication Number
WO 02/17542 A2

(51) International Patent Classification[7]: H04L

(21) International Application Number: PCT/US01/26533

(22) International Filing Date: 23 August 2001 (23.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/228,008 23 August 2000 (23.08.2000) US
60/272,793 1 March 2001 (01.03.2001) US

(71) Applicant: **GEYSER NETWORKS, INC.** [US/US]; 555 Del Rey Avenue, Sunnyvale, CA 94085 (US).

(71) Applicants and
(72) Inventors: **SCOTT, jAMES** [US/US]; Cupertino, CA (US). **GORDON, Lee** [US/US]; 20567 Russell Lane, Saratoga, CA 95070 (US). **WEN-LUNG, Chen** [US/US]; 10455 Creston Drive, Los Altos, CA 94024 (US).

(74) Agents: **GUZMAN, Leonard, T.** et al.; McDermott, Will & Emery, 2700 Sand Hill Road, Menlo Park, CA 94025 (US).

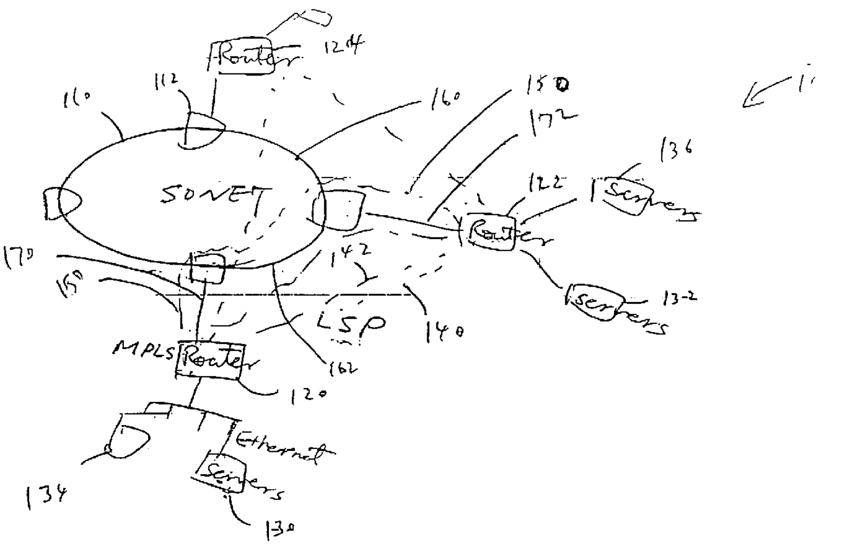
(81) Designated States *(national)*: AU, BR, CN, DE, DK, ES, GB, IN, JP, KR, MX, PT, SG.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *without international search report and to be republished upon receipt of that 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: SYSTEM AND METHOD OF BINDING MPLS LABELS TO VIRTUALLY CONCATENATED SONET/SDH TRANSPORT CONNECTIONS



(57) Abstract: A system and method of binding MPLS labels to virtually concatenated SONET/SDH transport connections is provided. A system and method of binding MPLS labels to nxVT connections is provided and includes (1) classifying packets; and (2) mapping packets with a same forward equialence class to a same nxVT connection. IN a further embodiment, the mapping includes: (a) associating a certain bandwidth of a connection between an external router and an add-drop multiplexer, between a source node and a destination node; and (b) creating an nxVT connection between the source node and the destination node, thereby creating a label switch path between the source node and the destination node. In another embodiment, a system and method of binding MPLS labels to nxVT connections is also provided and includes mapping packets with the same MPLS label to the same nxVT connection.

WO 02/17542 A2

SYSTEM AND METHOD OF BINDING MPLS LABELS TO VIRTUALLY CONCATENATED SONET/SDH TRANSPORT CONNECTIONS

SPECIFICATION

RELATED APPLICATIONS

This application is related to U.S. Provisional Application No. 60/228,008, filed on August 23, 2000, to U.S. Provisional Application No. 60/272,793 , filed on March 1, 2001, to co-pending and commonly assigned U.S. Patent Application No. (Number to be assigned) with Attorney Docket Number 55369-014, filed on August 23, 2001, and to co-pending and commonly assigned U.S. Patent Application No. (Number to be assigned) with Attorney Docket Number 55369-015, filed on August 23, 2001. The contents of U.S. Provisional Application No. 60/228,008, filed on August 23, 2000, of U.S. Provisional Application No. 60/272,793, filed on March 1, 2001, of co-pending and commonly assigned U.S. Patent Application No. (Number to be assigned) with Attorney Docket Number 55369-014, filed on August 23, 2001, and of co-pending and commonly assigned U.S. Patent Application No. (Number to be assigned) with Attorney Docket Number 55369-015, filed on August 23, 2001, are hereby incorporated by reference. This application claims priority to U.S. Provisional Application No. 60/228,008, filed on August 23, 2000, and to U.S. Provisional Application No. 60/272,793, filed on March 1, 2001.

FIELD OF THE INVENTION

The present invention relates to optical networking. More particularly, the invention relates to a system and method of binding MPLS labels to virtually concatenated SONET/SDH transport connections.

BACKGROUND OF THE INVENTION

Both SONET (See Synchronous Optical Network (SONET) Transport Systems: Common Generic Criteria. GR-253-CORE, Issue 2, Revision 1. December, 1997.) and SDH (See International Telecommunication Union. Network Node Interface for the Synchronous Digital Hierarchy. Recommendation G.707. March, 1996.) enable the use of virtual concatenation to support both the dynamic resizing of transport trunks and the grooming of traffic. More recently, advances in the transport of routed datagram traffic leveraging the research and experience of ATM has resulted in the standardization of MPLS (See Internet Engineering Task Force. Multiprotocol Label Switching Architecture. IETF Draft Document. August, 1999 and http://www.ietf.org/internet-drafts/draft-ietf-mpls-arch-06.txt.). This work allows network devices to employ a standards-based method by which packet traffic can traverse a network, while receiving a previously agreed upon Quality of Service.

Additionally, there is currently an effort underway by IETF (See Generalized MPLS - Signaling Functional Description; ftp://ftp.ietf.org/internet-drafts/draft-ietf-mpls-generalized-signaling-05.txt), which attempts to adapt the control plane semantics of generalized MPLS to allow vendor-independent provisioning of services at the transport layer.

Traditionally, design approaches for large-scaled packet networks embrace the use of multiple control planes. A typical example would be mapping an IP overlay on top of a Layer-2 ATM network. Such networks typically suffer a discontinuity between the higher layer connectionless IP control plane running an IP routing protocol (whether it be an interior gateway protocol, such as OSPF, or an exterior gateway protocol, such as BGP), and the lower layer connection-oriented ATM control plane running its own routing protocol (PNNI). Numerous approaches have been submitted to standards bodies to address this issue, each with its various merits and demerits. In addition to the discontinuity between the Layer 2 and Layer 3 control planes, the underlying transport layer connections, using such technologies as SONET, SDH, or DWDM, have typically been provisioned using a directed approach, where each cross-connect in the transport path is individually established using a vendor-specific element management system (EMS). As such, end-to-end provisioning of circuits has traditionally been a laborious and time-consuming procedure for service providers.

In a prior art system in Figure 1, labels have been used to interconnect each forwarding engine so as to be distributed throughout the MPLS domain. Traffic associated with these labels would traverse the quasi-static label switched paths (LSPs) 140 and 142 used to interconnect devices within the access network, so there are no issues related to label proliferation and usage of VT space on the SONET ring. In Figure 1, prior art system 100 includes routers 120, 122, and 124, servers 130, 132, 134, and 136, an add/drop multiplexer (MUX) 112, logically interconnected as shown, in a ring 110. Servers 130 and 132 are logically connected via LSP 140, while servers 134 and 136 are logically connected via LSP 142. Router 120 is a MPLS router. There is a physical link 150.

In this scenario, nodes in the access network receiving L3 packets from attached routers would process the IP header or MPLS label and select an outgoing label, used to direct the packet toward the appropriate line card at the destination node. This label is then pushed onto the label stack. Next, the label used to forward the packet to the destination node itself is selected and pushed onto the stack.

At the destination node, the switch card, which is an MPLS LSR (Label Switched Router), will pop the top label off of the stack and examine the underlying label to determine which line card the packet is to be switched to. In this way, an interoperable method of forwarding traffic in the access network is achieved. However, the cost is the overhead of having the switch card pop the top label off of the label stack. The functionality necessary for this capability is essentially identical to that required for supporting access network topologies defined by sparsely connected .

In other prior art systems, any connections afforded service guarantees were point-to-point (Transparent Lan Service received Best Effort service). As such, the admission control decisions were fairly simple. However, in the cases described in this document, anypoint-to-anypoint connectivity with service guarantees must now be supported. As a result, any port in the system can be significantly oversubscribed, due to multiple ports directing traffic to it. As such, when connections are established, a method of admission control is required to ensure that the level of oversubscription does not reach a point at which the associated service guarantees may be compromised.

Thus, two levels of admission control are required: the lowest level determines whether the connection can be admitted, based on the existing port configuration and resources reserved for the currently established connections on those ports. This level of admission control can be implemented in the CR-LDP/RSVP Traffic Engineering module, local to each network element (residing on the present invention's switch card). Additionally, this level of admission control should embrace a more sophisticated algorithm than simply accounting for the summation of CIR or PIR.

The higher level of admission control is needed to determine whether admitting (or deleting) the connection will cause the bandwidth associated with the quasi-static node-interconnect LSPs to change. If so, the bandwidth of the relevant LSP is modified according to the two standards-based methods described above. This level of admission control is most easily placed in an element that has a global view of the topology (such as the centralized EMS, in the current software architecture).

SUMMARY OF THE INVENTION

The present invention provides a system and method of binding MPLS labels to virtually concatenated SONET/SDH transport connections. A system and method of binding MPLS labels to nxVT connections is provided and includes (1) classifying packets, and (2) mapping packets with a same forward equivalence class to a same nxVT connection. In a further embodiment, the mapping includes (a) associating a certain bandwidth of a connection between an external router and an add-drop multiplexer, between a source node and a destination node, and (b) creating a nxVT conection between the source node and the destination node, thereby creating a label switch path between the source node and the destination node.

In another embodiment, a system and method of binding MPLS labels to nxVT connections is also provided and includes mapping packets with the same MPLS label to the same nxVT connection.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a prior art MPLS labeling system.

Figure 2 illustrates MPLS labeling in accordance with an exemplary embodiment of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Introduction

The invention described in co-pending and commonly assigned U.S. Patent Application No. (Number to be assigned) with Attorney Docket Number 55369-014 provides a system and method of virtually concatenating VT1.5s and STS-1s over SONET and SDH and WDM. The virtual concatenation invention allows users to setup connections or pipes with configurable bandwidth over either nxSTS-1/nxAU-3/nxAU-4 or nxVT1.5/nxTU-11/nxTU-12 within a nxSTS-1/nxAU-3/nxAU-4 pipe on an existing SONET/SDH network. This provides a connection or pipe of adjustable bandwidth with a granularity of close to 1.5 Mbps to fit the needs of applications. The resulting connection can be treated as a TDM like connection.

By replacing "STS-1" with "AU-3" or "AU-4" and "VT" or "VT1.5" with "TU-11" or "TU-12", the virtual concatenation invention applies to nxAU-3/nxAU-4 and nxTU-11/nxTU-12 for SDH networks. For simplicity, these connections are called "nxVT" for both SONET and SDH networks. By replacing "STS-1" with "VT" or "VT1.5", the virtual concatenation invention applies to nxSTS-1 and nxAU-3/nxAU-4.

On top of the virtual concatenation invention, a dynamic bandwidth allocation (DBA) protocol, which is described in co-pending and commonly assigned U.S. Patent Application No. (Number to be assigned) with Attorney Docket Number 55369-015, allows for dynamically changing the throughput of all nxVT connections, based on the real-time traffic loads of applications using the nxVT connections. The DBA protocol allows for the efficient use of the SONET/SDH bandwidth through statistical multiplexing. The same dynamic bandwidth allocation protocol applies to nxSTS-1 and nxAU-3/nxAU-4.

The virtual concatenation invention provides for virtual concatenation, which includes creating a logical connection or pipe by combining multiple, n (where n is a positive integer), STS-1 or VT connections or pipes, which may be contiguous or non-contiguous, into a single connection or pipe, nxSTS-1 or nxVT, respectively, in order to support a connection or pipe with a higher throughput than the throughput of the original STS-1 or VT pipes.

By using virtual concatenation, the present invention uses dynamic bandwidth allocation protocol used to dynamically manage bandwidth on a ring (or graph of rings) topology. This technology allows for more efficient capacity utilization on a ring (or multiple rings), by rapidly providing ring capacity to individual network elements on demand. As such, high capacity utilization is achieved for arbitrary traffic loading patterns.

The present invention provides a system and method of binding MPLS labels to virtually concatenated SONET/SDH transport connections.

By leveraging the control plane semantics of MPLS, the present invention effects a more seamless management solution by allowing a device to participate as a peer in the topology exchange between attached routers at Layer-3 and core transport devices at Layer-2. Embracing this philosophy, the present invention employs MPLS label switched paths (LSP) to establish connections between devices in the transport network (See Internet Engineering Task Force. Extensions to RSVP for LSP Tunnels. IETF Draft Document. September, 1999. http://www.ietf.org/internet-drafts/draft-ietf-mpls-rsvp-lsp-tunnel-04.txt.) (See Internet Engineering Task Force. Constraint Based LSP Setup Using LDP. IETF Draft Document. September, 1999. http://www.ietf.org/internet-drafts/draft-ietf-mpls-cr-ldp-03.txt.), and provides a gateway functionality between attached IP or MPLS devices and the internal transport network connections carrying traffic, whose available ring capacity is managed by the DBA protocol. Operating as an MPLS peer in the topology exchange at Layer-3 allows the present invention employs components of MPLS (See Internet Engineering Task Force. LSP Modification Using CR-LDP. IETF Draft Document. September, 1999. http://www.ietf.org/internet-drafts/draft-ietf-mpls-crlsp-modify-01.txt.) to allow network operators to dynamically create/destroy transport level connections and/or modify their associated Service Level Agreements (SLA).

Label Binding Issues

When discussing label binding in the context of the present invention, the present inventions does the following:

(1) the pre-establishment of label switched paths (LSPs) used to interconnect devices within the metropolitan access network, where these devices support the ability to dynamically modify the bandwidth associated with these LSPs;

(2) the aggregation of connectionless IP flows or connection-oriented MPLS LSP traffic arriving at the edge device (from attached routers) onto paths that are routed through the access network on one or more of the above pre-established LSPs;

(3) dynamic modification of bandwidth associated with both the pre-established LSPs and the Service Level Agreements (SLA) associated with the connections from attached Layer-3 devices; and

(4) binding of labels used for LSPs that provide connectivity to core transport devices being used to interconnect various metro access networks.

Referrring to Figure 2, by embracing the work underway in the IETF, edge devices in transport level access networks can exist as peers in the topology exchange of core transport devices via an interoperable, standards-based method. Therefore, the label bindings used to resolve the last bullet item in the list above is considered beyond the scope of this document. What is of particular relevance to the content of this document is the resolution of the issues described in the first two bullet items. For the purpose of this discussion, consider the network topology described by N network elements, L line cards per N, P interfaces (physical or logical) per L, and F queueing subclasses (denoting an MPLS Forward Equivalence Class, or FEC) per P.

LSPs Used to Interconnect NEs

In the present invention, the worst-case number of LSP required to interconnect network elements on each ring is calculated according to the well-known formula $(((N * (N - 1)) / 2)$. Note that the graph can contain fewer vertices, in order to save VT space on the ring. However, in the case where a connected graph is not established, the network operator must establish an explicitly routed LSP tunnel to route traffic between endpoints of a connection that do not share an adjacency. Additionally, in the case where the graph is not connected, packets traversing these LSP tunnels will endure increased delay and delay variance as a result of the increased hop count. Perhaps more importantly, this case requires that the present invention push and pop labels on the label stack, which will require CPU intervention. Because the present invention's line and switch cards provide QoS-capable forwarding of packet traffic, no dilation of the basic node interconnect graph is required. Interconnecting network elements to those of other vendors may require the use of additional LSPs; one per FEC, for example.

These LSPs can be configured using either RSVP-LDP or CR-LDP. The essential behavior of these LSPs should be similar in nature to those of ATM Soft-PVCs. For the general case, the bandwidth allocated for each LSP is calculated according to the following formula:

$$BW_{LSP} = \max\left(\left(\sum_{i=1}^{L} \sum_{j=1}^{P} BW_{i,j} \right)_{TxEndpoint1}, \left(\sum_{i=1}^{L} \sum_{j=1}^{P} BW_{i,j} \right)_{TxEndpoint2} \right)$$

For the case where only symmetric connections (where the specified bandwidth must be identical in each direction) are supported, the above expression is constrained such that each endpoint must have equivalent values. This constraint alleviates the necessity for any proprietary messaging to pass between the endpoints of an LSP, used to signal each endpoint's corresponding component of the above expression. Thus, for the deconstructed case, where only symmetric virtually concatenated connections are supported, the above expression reduces to:

$$BW_{LSP} = \left(\sum_{i=1}^{L} \sum_{j=1}^{P} BW_{i,j} \right)$$

In the simple case, the BW component of the summation above is simply the PIR associated with each connection on each line card of a network element. The accuracy of the value BW could be improved by employing a slightly more sophisticated approach, such as the well-known equivalent bandwidth calculation [Guerin91], which accounts for the effects of statistical multiplexing and aggregate utilization, as well as employing a more realistic source model. However, because DBA is managing the bandwidth usage within these LSPs, the amount of oversubscribing actually occurring is not of particular concern.

<u>Dynamic Modification LSP Modification</u>

It should be noted that, when connections are added or removed from the network, such that the bandwidth derived from the above expressions changes, the bandwidth associated with the LSP is dynamically adjusted.

<u>LSP Modification Using CR-LDP</u>

Modification of the bandwidth associated with an existing LSP begins with the issuance of a LDP Label Request Message by the node requiring more bandwidth. Within this message, the encoded Type-Length-Value (TLV) will contain the newly requested bandwidth values, and the LSP identifier TLV (LSPID TLV). On transmission of the Label Request Message, the sending node will retain the original entry in its FEC to Next Hop Label Forwarding Entry (FTN) table.

On receipt of the Label Request Message, the receiving node will detect that the LSPID is identical to that of one of the entries in its Incoming Label Map (ILM). If the receiving node did not find such an entry in its ILM, it would simply treat the Label Request Message as a new label request, and process the request in the typical fashion. However, in this case, it will compute the new bandwidth required. The receiving node will reserve only the difference between the original and new bandwidth values, in order to prevent double booking the capacity. As such, this node will temporarily have two labels in its ILM with the same LSPID.

Similarly, when the initiating node receives the label from its LDP peer, two sets of labels will exist for the same LSPID. As such, the initiating node will now have two outgoing labels for the same FEC. The initiating node can then activate the new label in its FTN and begin the process of releasing the old label, by sending an LDP Label Release Message to its peer.

On receipt of the Label Release Message, the egress node will detect that the LSPID has another active label associated with it. As a result, this node will release the old label without deallocating the resources being used by that label.

LSP Modification Using RSVP-LDP

The method of LSP bandwidth modification using RSVP-LDP is virtually identical to that used by CR-LDP, as this portion of CR-LDP is based on previous work in RSVP. However, the specific details of this functionality in RSVP are left for future study.

Anypoint-to-Anypoint Layer-3 Connectivity

This type of connectivity is provided through the use of Layer-3 routing and forwarding. As the L3 forwarding functionality resides on line cards in the present invention, each packet must contain information that the switch card can use to determine the appropriate line card at the destination endpoint.

The present invention appends a tag to the packet by the Layer-3 forwarding engine to specify the line card. This is essentially a tag switching approach. This method will work well on homogeneous rings employing only devices of the present invention. In this approach, the number of tags needed is calculated as the product of the total number of NEs (N - 1), line cards (L), and FECs (F):

$$NUM_{TAGS} = F \times ((N - 1) \times L)$$

For the current design, this implies (4 x (31 x 8)) = 992 proprietary tags (10 bits).

Dynamic Resizing of Virtually Concatenated Connections

One of the essential benefits of the present invention functioning as an MPLS LSR lies in its ability to leverage CR-LDP or RSVP-LDP for dynamically resizing the dimension N of NxVT, NxTU11, or NxSTS-1 connections. This capability provides a mechanism by which attached LSRs residing at the service provider's Point-of-Presence (POP) can signal the creation, destruction, or resizing of connections on an end-to-end (POP-to-POP) basis. This capability is afforded by the ability of the present invention to function as an LDP peer with attached LSRs.

Note that, in the case where the attached router is not an LSR, the method available for dynamic resizing of virtually concatenated connections will leverage native RSVP over IP, rather than CR-LDP or RSVP-LDP.

Conclusion

The present invention relates to optical networking. More particularly, the invention relates to a system and method of binding MPLS labels to virtually concatenated SONET/SDH transport connections.

Having fully described a preferred embodiment of the invention and various alternatives, those skilled in the art will recognize, given the teachings herein, that numerous alternatives and equivalents exist which do not depart from the invention. It is therefore intended that the invention not be limited by the foregoing description, but only by the appended claims.

CLAIMS

We claim:

1. A method of binding MPLS labels to nxVT connections comprising:

classifying packets; and

mapping packets with a same forward equivalence class to a same nxVT connection.

2. The method of claim 1 wherein the classifying comprises identifying a packet by a packet field of the packet.

3. The method of claim 2 wherein the identifying comprises, if the packet is a Internet Protocol packet, examining the DHCP byte of the packet.

4. The method of claim 2 wherein the identifying comprises, if the packet is an PPP packet, examining PPP information of the packet.

5. The method of claim 2 wherein the identifying comprises, if the packet is an ATM packet, examining the VPI or VCI of the packet.

6. The method of claim 2 wherein the identifying comprises, if the packet is a Frame Relay packet, examining the DLCI of the packet.

7. The method of claim 1 wherein the mapping comprises:

associating a certain bandwidth of a connection between an external router and an add-drop multiplexer, between a source node and a destination node; and

creating a nxVT conection between the source node and the destination node, thereby creating a label switch path between the source node and the destination node.

8. A method of binding MPLS labels to nxVT connections comprising mapping packets with the same MPLS label to the same nxVT connection.

9. The method of claim 8 wherein the mapping comprises:

associating a certain bandwidth of a connection between an external router and an add-drop multiplexer, between a source node and a destination node; and

creating a nxVT conection between the source node and the destination node, thereby creating a label switch path between the source node and the destination node.

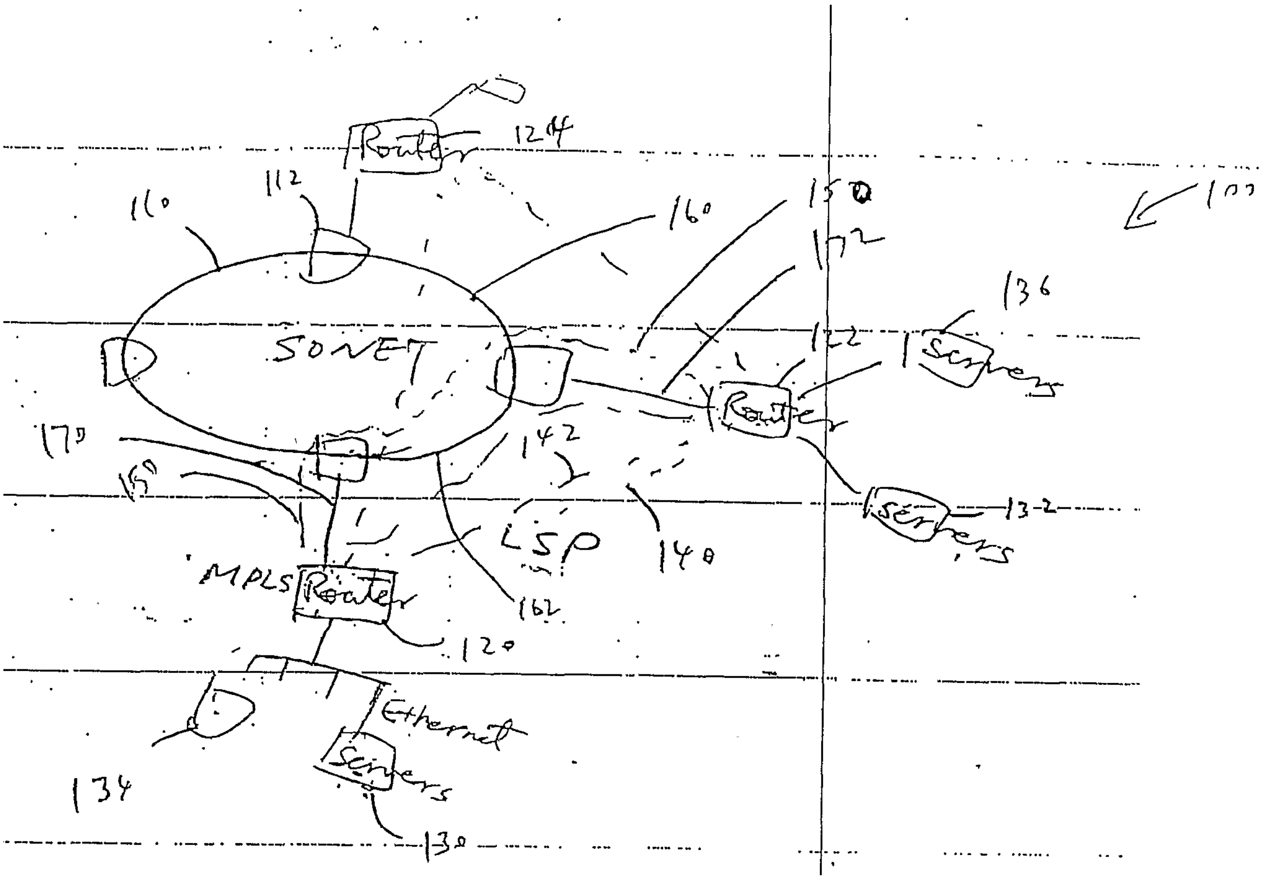

SONET
Router
Router
MPLS Router
Servers
Servers
Servers
Ethernet
LSP

FIG. 1

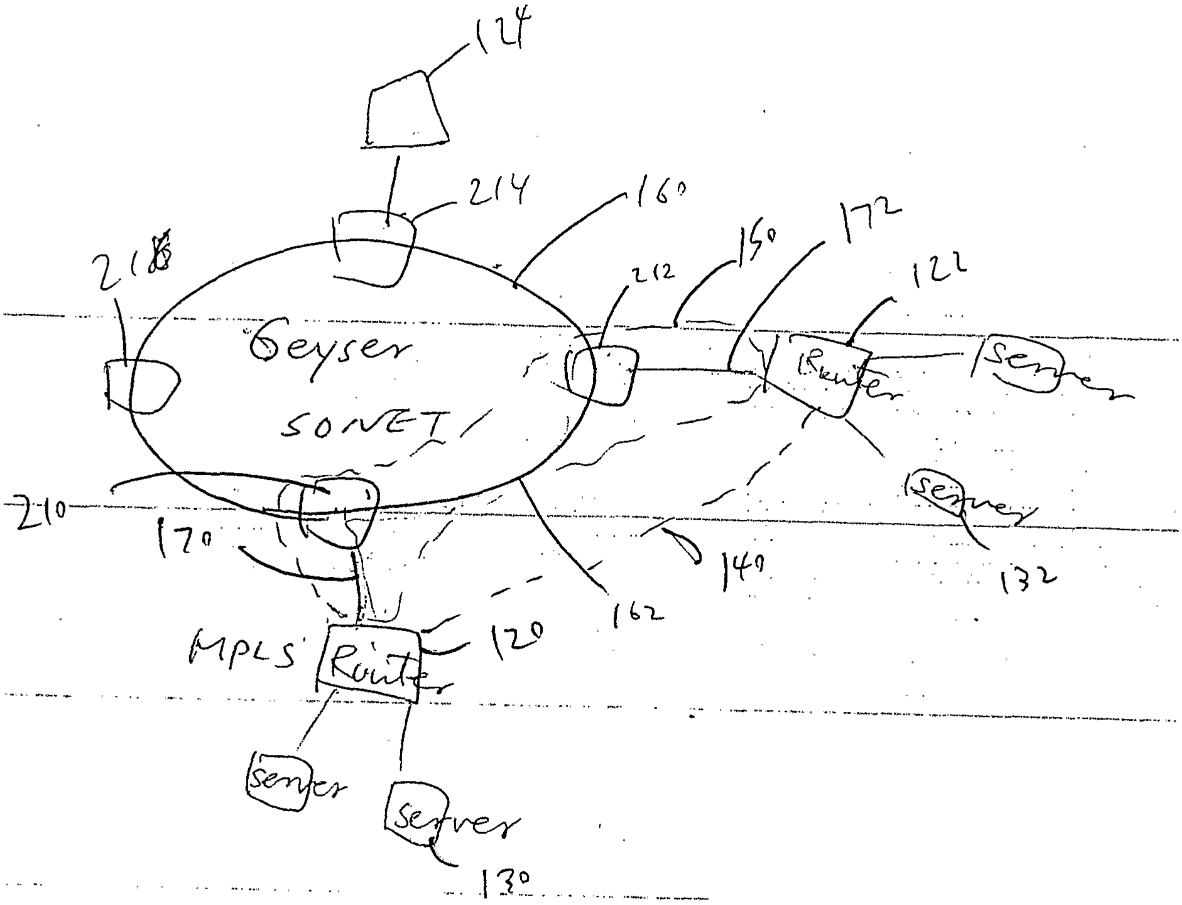

124
214
160
218
Geyser
SONET
212
150
172
122
Router
Server
Server
132
210
170
140
162
MPLS Router
120
Server
Server
130

FIG. 2