



[12] 发明专利申请公开说明书

[21] 申请号 03814379.8

[43] 公开日 2005 年 8 月 31 日

[11] 公开号 CN 1663176A

[22] 申请日 2003.7.25 [21] 申请号 03814379.8

[30] 优先权

[32] 2002.7.30 [33] EP [31] 02255321.8

[86] 国际申请 PCT/GB2003/003308 2003.7.25

[87] 国际公布 WO2004/012393 英 2004.2.5

[85] 进入国家阶段日期 2004.12.20

[71] 申请人 安捷伦科技有限公司

地址 美国加利福尼亚州

[72] 发明人 安德鲁·罗伯特·莱哈恩

弗朗西斯科·雅维耶·加西亚

约瑟夫·斯文特克

[74] 专利代理机构 北京东方亿思知识产权代理有限
责任公司

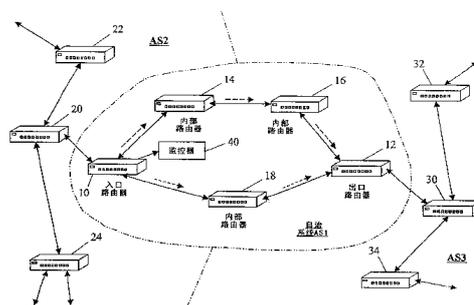
代理人 王 怡

权利要求书 3 页 说明书 19 页 附图 11 页

[54] 发明名称 标识网络路由器和路径

[57] 摘要

通过收集含有指示网络的互连及该网络与其它网络的互连的信息的分组，标识可能被通信网络中的分组采用的网络范围内的路径集。所收集分组的内容被用来标识网络范围内的路由器集和子网络集及它们的互连，上述路由器集和子网络集及它们的互连被网络中的通信所遍历。提供一个输出，用来指示网络范围内的路由器集和子网络集及它们的互连的任何所选择部分。



1. 一种用于标识可能被通信网络中的分组所采用的网络范围内的路径集的方法，包括下述步骤：

5 收集含有指示所述网络的互连及所述网络与其它网络的互连的信息的分组；

检测所选择分组的内容；

使用所述检测到的内容来标识网络范围内的路由器集和子网络集及它们的互连，所述路由器集和子网络集及它们的互连被所述网络中的通信所遍历；以及

10

提供一个输出，用来指示所述网络范围内的路由器集和子网络集及它们的互连的任何所选择部分。

15

2. 如权利要求 1 所述的方法，包括下述步骤：使用所述检测出的内容来确定与所标识的路由器和子网络相关联的功能性，以及所标识的互连的成本度量，所标识的路由器和子网络及互连被所述网络中的通信所遍历。

3. 如权利要求 2 所述的方法，包括下述步骤：使用所述检测出的内容来确定网络范围内的潜在路径集，所述潜在路径集包括通过所述网络的路径以及将所述网络连接到其它网络的路径，所述潜在路径集被所述网络中的通信所遍历。

20

4. 如权利要求 3 所述的方法，包括下述步骤：使用所述检测出的内容来确定如所述成本度量所定义的缺省路径集，所述缺省路径集被所述网络中的通信所遍历。

5. 如权利要求 3 所述的方法，包括下述步骤：

基于所述路由器的预定功能性来查询所述路由器；以及

25

使用所述查询结果来确定，在网络入口路由器处是否发生分组分类，以及所述缺省路径的任何替代性逻辑路径是否被所述网络中的通信所遍历。

6. 如权利要求 3 所述的方法，包括下述步骤：使用所述检测出的内容来确定可以被所述网络中的通信遍历的替代性逻辑路径。

7. 如权利要求 4、5 和 6 中的任意两个或所有三个所述的方法，包括下述步骤：生成所确定的路径之间的比较。

8. 如权利要求 5 所述的方法，包括下述步骤：向所述路由器查询与所确定的路径相关联的属性，所述属性指示所述路径的预定路由目标。

5 9. 如权利要求 7 和 8 所述的方法，包括下述步骤：使用所述比较来确定当前使用的逻辑路径的替代性逻辑路径，所述替代性逻辑路径满足所述预定路由目标。

10 10. 如权利要求 9 所述的方法，包括下述步骤：推荐满足所述预定目标的替代性逻辑路径集。

11. 如前述权利要求中的任何一个所述的方法，其中通过监控遍历所述网络中的至少一条链路的分组，来收集分组。

12. 一种用于标识可能被通信网络中的分组采用的网络范围内的路径集的装置，包括：

15 收集器，用于收集含有指示所述网络的互连及所述网络与其它网络的互连的信息的分组；

 检测器，用于检测所选择分组的内容；

 标识器，用于使用所述检测到的内容来标识网络范围内的路由器集和子网络集及它们的互连，所述路由器集和子网络集及它们的互连被网络中的通信所遍历；以及

20 输出部分，用来提供所述网络范围内的路由器集和子网络集及它们的互连的任何所选择部分的指示。

25 13. 如权利要求 12 所述的装置，其中所述标识器使用所述检测出的内容来确定与所标识的路由器和子网络相关联的功能性，以及所标识的互连的成本度量，所标识的路由器和子网络及互连被所述网络中的通信所遍历。

 14. 如权利要求 13 所述的装置，其中所述标识器使用所述检测出的内容来确定网络范围内的潜在路径集，所述潜在路径集包括通过所述网络的路径以及将所述网络连接到其它网络的路径，所述潜在路径集被所述网络中的通信所遍历。

15. 如权利要求 14 所述的装置，其中所述标识器使用所述检测出的内容来确定如所述成本度量所定义的缺省路径集，所述缺省路径集被所述网络中的通信所遍历。
16. 如权利要求 14 所述的装置，包括查询生成器，用于基于所述路由器的预定功能性来查询所述路由器，其中所述标识器使用所述查询结果来确定，在网络入口路由器处是否发生分组分类，以及所述缺省路径的任何替代性逻辑路径是否被所述网络中的通信所遍历。
17. 如权利要求 14 所述的装置，其中所述标识器使用所述检测出的内容来确定可以被所述网络中的通信遍历的替代性逻辑路径。
18. 如权利要求 15、16 和 17 中的任意两个或所有三个所述的装置，其中所述标识器生成所确定的路径之间的比较。
19. 如权利要求 16 所述的装置，其中所述查询生成器向所述路由器查询与所确定的路径相关联的属性，所述属性指示所述路径的预定路由目标。
20. 如权利要求 18 和 19 所述的装置，其中所述标识器使用所述比较来确定当前使用的逻辑路径的替代性逻辑路径，所述替代性逻辑路径将满足所述预定路由目标。
21. 如权利要求 20 所述的装置，其中所述标识器推荐满足所述预定目标的替代性逻辑路径集。
22. 如权利要求 12 到 21 中的任何一个所述的装置，包括监控器，用于监控遍历所述网络中的至少一条链路的分组。

标识网络路由器和路径

5 技术领域

本发明涉及用于标识通信网络中的路由器、相关联的互连通信链路以及遍历那些链路的分组所采用的路径的方法和装置，所述通信网络例如是使用因特网协议（IP）的分组网络。本发明特别适用于使用链路状态路由协议的网络，所述链路状态路由协议例如是开放最短路径优先（OSPF）或中间系统—中间系统（IS-IS）或与它们等价的任何协议。

背景技术

为了从众多竞争者中脱颖而出，并且为了在不损害现有成本结构和资金预算的情况下改善向用户提供的服务水平，因特网服务供应商（ISP）越来越多地采用成本优化、服务加强或服务区分机制，以在他们的网络中实现“流量管理”。这些机制包括流量工程（下文将描述）、服务质量（QoS）测量和服务水平协议（SLA）。存在可以帮助操作员实现这些“经流量管理”的网络的多种技术。对于 IP 网络，这包括多协议标记交换（MPLS），例如参见 Internet Engineering Task Force（因特网工程任务组）的 Request for comments (RFC) 3031, 以及区分的服务，例如参见 RFC 2474 和 2475。

这些技术的共同点是它们在数据分组最初进入分立网络（在因特网术语中通常被称为自治系统）的入口点处使用分组分类。相反，同一分组通常将在网络的出口点处取消分类，使得接收该分组的下一网络/自治系统如果想要的话，可以以自己的方式重新分类分组。分类确保每个分组在路由经过网络时得到适当的处理。取决于在入口路由器处给予分组的分类类型，分组经过网络时将得到的处理将会不同。

对于每个经分类的分组，中间路由器将该分组强制路由到网络上的不同逻辑路径，该路径远离在分组未被分类的情况下通常会采取的预定缺省

路径。为网络内的每个源头与目的地定义至少一个缺省路径。该缺省路径通常是由内部网关协议（IGP）成本度量（metric）为每个互连定义的最小成本路径（下面将参考图 4-7 在 OSPF 的上下文中描述）。

5 因此，逻辑路径是得到不同路由（分组转发）处理的所有分组所采取的替代性非缺省路径。逻辑路径例如可以是不同于分组在未被分类的情况下通常所采取的路径的单独路径。类似地，可以由中间路由器处的不同排队处理来定义逻辑路径。在两个实施例中，根据所接收的分类，经分类的分组将得到不同的一组处理，并且与同一分组在缺省路径上路由的情况相比，给予该分组不同的一组传输特性。每个逻辑路径都具有一组给定属性，用来确定遍历路径的分组所需的传输特性，例如物理互连上的多少带宽被预留给该逻辑路径、服务水平（“铜”“银”或“金”）、最大容许抖动（jitter）或者该逻辑路径必须经过的所有特定路由器。

10 例如，应用流量工程的网络操作员可以决定经由专用逻辑路径将对抖动敏感的视频会议流量传输通过 MPLS 使能的网络。该路径不同于非视频会议流量将采取的其它缺省路径。尽管潜在地具有更多的路由器跳（hop），但是该专用路径（在这种情况下为单独物理路径）不承载任何其它流量，因而可以方便地容纳结合的话音和视频，而不会引入有害的抖动。所有其它流量都采取缺省路径，例如由 IGP 定义的具有最小总成本度量的路由。不管采取哪种路由，所有流量最终到达出口路由器，并且分组随后被取消分类以备传送到下一网络。如果没有这个负载平衡机制，将使用缺省路径来路由所有的网络分组，在峰值期间，这会使网络变得过载，进而丢弃或延迟分组，使得视频会议不可用并且给其它数据流量用户引起问题。

25 可以在多个不同的级别上应用流量工程处理，例如针对不同用户、针对不同服务或者针对两者的结合。同样，可以采用具有不同商业目标的其它流量管理工具，例如 QoS 和 SLA 机制。QoS 和 SLA 两者都需要在入口点和出口点处的分组分类，并且都造成其它路由策略和逻辑路径的使用，所述逻辑路径与缺省（通常是最小成本）路径不同以在网络内同时路由流量。

许多流量工程技术涉及网络操作的监控，例如核查是否符合采用的 QoS 或 SLA 标准，以及如果未达到该标准则及时启动补救措施或（次理想的）补偿。尝试对经流量管理的网络进行这种监控的网络管理系统的问题是，发现分组从哪里进入和离开网络以及分组的分类和后续处理是否正确。网络管理系统应当检测不正确的分组分类或内部路由器故障，其中不正确的分组分类可能引起流量被错误地路由，内部路由器故障可能引起所有流量使用同一路径而不考虑分类，无论那种情形，都会导致分组被延迟或丢弃，并且可能违反 SLA。

整个因特网被划分成多个管理域。例如，因特网服务供应商可能构成单个管理域。通过与邻近域（其它 ISP 等）订立协议以形成对等或转接关系来承载每个其它方的流量以及实现用户期望的连接，每个管理域都形成因特网的一部分。管理域含有一个或多个自治系统（AS）。AS 是通常在单个技术管理（例如一个 ISP）之下的一组路由器，其中：

- 对外部显示出具有单个一致的内部路由计划（使用一个和可能几个内部网关协议以及一个或多个公共度量来在该 AS 内路由分组）；
- 呈现通过它能到达哪些目的地的一致画面；以及
- 使用外部网关协议将分组路由到其它 AS。

这里，因特网上下文中使用的词语“网络”意指这种自治系统。在其它种类的通信系统的上下文中，词语网络应当被理解为意为操作元件集，其在概念和功能上类似于因特网 AS，而不论该元件集包括整个系统还是仅仅为其一部分。

因特网由许多管理域中的许多 AS 构成。在每个 AS 之间的每个连接处存在“边缘”路由器，并且每个边缘路由器具有实现某些类型流量管理的能力。大 ISP 可能具有与许多其它 ISP 进行交互的许多入口和出口路由器，并且具有不同的端用户。每个入口和出口路由器可以使用多种不同策略来分类并路由流量。配置、监控和管理流量管理技术中的巨大挑战是显而易见的。

具有整个网络拓扑知识（操纵分组遍历网络的中间路由器和活动边缘路由器的身份）对于应对这些挑战有很大帮助。然而，现有分组网络技术

没有以显式的形式提供可被外部工具容易地访问的这种知识，其中外部工具可以被用来促进流量管理。可以帮助解决上述挑战的作为网络补充的系统将需要从可能非常大的网络近乎实时地发送拓扑信息，并且理想地没有对现有网络和路由器性能产生明显影响。路由器通常含有网络中的路由器和链路状态的完整数据库。该信息被称为链路状态数据库，并且被用于生成每个路由器内部的路由表，以确定数据分组向着它的最终目的地而要被转发到的最优邻近路由器。例如通过 RFC 2328 中描述的 OSPF 链路状态协议（并且下文被称为 OSPF 协议）来产生路由表。路由器的链路状态数据库中含有的信息描述了足以达到路由器操作要求程度上的拓扑；这样的数据原则上可以从路由器逐个提取，并且可以用来产生完整拓扑描述。令人遗憾的是，使用当前的技术，在确保维护网络完整性的同时，不能以满足规模、精确度和及时的必需要求的方式来访问所需数据。

例如，尽管使用简单网络管理协议（SNMP）的查询理论上可以用来收集所需数据，但是该方法并不适用于含有众多路由器的大网络。在确定网络完整拓扑时所需的大量 SNMP 查询将在路由器上加上大的处理负荷，并且产生相当大量的网络流量。此外，为了避免必须查询网络上的每个地址，其中多数地址将是诸如个人计算机或工作站之类的终端，需要预先知道路由器地址，进而强加了相当的管理负荷并且损害了网络拓扑的自动发现或验证的好处。类似的问题适用于从操作支持系统（OSS）或其它外部数据源提取信息；该信息可能不可用，可能没提供，或者可能被认为太敏感而不准通过直接访问来获取。此外，即使信息是可用的，在没有独立验证的条件下也没有对该信息是准确的任何保证。

本发明的目标是通过帮助提供网络拓扑的描述来促进对流量管理的监控。例如，通过使用本发明获得的网络拓扑描述可以帮助网络操作员管理采用流量管理技术（例如 MPLS 和区分服务）的网络，或者可以用于采用核心 MPLS 使能的 IP 网络（参见 RFC 2917）、IP 上语音服务（还被称为因特网电话）、SLA 和 QoS 机制。具体地说，本发明促进对在网络中的各种物理互连、路由器和子网络上实现的不同逻辑路径和任何相关联传输特性的监控。

发明内容

根据本发明的一个方面，提供了一种用于标识可能被通信网络中的分组采用的网络范围内的路径集的方法，包括下述步骤：

- 5 收集含有指示该网络的互连及该网络与其它网络的互连的信息的分组；

检测所收集分组的内容；

- 使用所述检测到的内容来标识网络范围内的路由器集和子网络集及它们的互连，所述路由器集和子网络集及它们的互连被网络中的通信所遍历；以及
- 10

提供一个输出，用来指示网络范围内的路由器集和子网络集及它们的互连的任何所选择部分。

- 上下文中的“网络范围”意为产生的网络描述并不集中于网络中任何特定的路由器或其它节点。例如 OSPF 在概念上在每个路由器中产生经过
- 15 网络的路径的树状描述，其中以该路由器作为树的根，以作为产生期望路由表的过渡步骤。该树中不包括从该“根”路由器转发分组所不需要的路由器之间的路径。相反，本发明生成网络拓扑的描述，其中所有路由器同等重要，并且在一个典型的实现中，提供了所有路由器之间的不仅限于缺省路径的所有路径的全面视图。

- 20 根据本发明的另一方面，提供了一种用于标识可能被通信网络中的分组采用的网络范围内的路径集的装置，包括：

收集器，用于收集含有指示该网络的互连以及该网络与其它的网络的互连的信息的分组；

检测器，用于检测所收集分组的内容；

- 25 标识器，用于使用所述检测到的内容来标识网络范围内的路由器集和子网络集及它们的互连，所述路由器集和子网络集及它们的互连被网络中的通信所遍历；以及

输出部分，用来提供网络范围内的路由器集和子网络集及它们的互连的任何所选择部分的指示。

附图说明

通过示例并结合附图，现在将说明根据本发明的用于标识通信网络中被通信链路互连的路由器的功能的方法和装置，附图中：

- 5 图 1 示出了因特网的概念段；
- 图 2 示出了示例性网络拓扑描述；
- 图 3 至图 7 示出了 OSPF 协议中定义的链路状态通告的格式；
- 图 8 至 13 示出了用于导出网络拓扑描述的过程；以及
- 图 14 示出了图 1 中的两个路由器发生故障后的图 1 的概念段。

10

具体实施方式

参考图 1，示出的因特网概念段包括自治系统 AS1 和连接到它的两个其它自治系统 AS2 和 AS3 的一部分。系统 AS1 含有两个边缘路由器 10 和 12 和三个内部路由器 14、16 和 18，两个边缘路由器分别提供到系统 AS2 和 AS3 的外部连接，三个内部路由器只是连接到其自己 AS 内部的其它路由器。系统 AS2 和 AS3 类似地分别包括边缘路由器 20 和 30 和内部路由器 22、24、30 和 34，边缘路由器用于提供到系统 AS1 的连接。

每个 AS 需要转发 AS 本地的以及 AS 之间全局的信息，从而数据分组可以通过节点或路由器被路由到正确的目的地。在 AS 之间，使用被称为外部网关协议例如 RFC 1771 中描述的边界网关协议（BGP）的一类协议来静态或动态配置路由器（和路由）。在 AS 内部，使用被称为内部网关协议（IGP）例如 OSPF、IS-IS 或路由信息协议（RIP）的一类协议来静态或动态配置路由器（和路由）。为方便起见，下面的说明将假定使用 OSPF，但是本发明可以结合实现与 OSPF（包括 IS-IS）类似概念和功能的其它协议来使用。

在例如 OSPF 的链路状态路由协议中，每个路由器负责发布并维护一个数据库，该数据库描述用于形成含有该路由器的 AS 的全部或部分的区段（zone）或区域拓扑。该数据库被称为链路状态数据库。在开始工作时，基于由路由器的管理员预配置的信息，该路由器仅仅知道它自身的本

地状态、它连接的接口和网络。学习并发布更多网络状态信息例如连接性的过程，是通过与 AS 中的其它路由器交换由 OSPF 协议定义的特殊数据分组来实现的。

5 最初使用例如分组多播技术来形成与邻近路由器的“邻接”。邻接是为交换路由信息的目的而与路由器的每个活动邻居形成的相互关系。一旦邻接已形成，邻接的路由器就使用根据 OSPF 协议进行格式化的 OSPF 链路状态描述分组来交换有关它们状态的信息。该过程一直持续到两个路由器共享它们在 AS 中的区段的拓扑的公共视图，从而在每个路由器中创建链路状态数据库。

10 在整个 AS 中完成邻接形成过程后，AS 中的每个路由器结合其自身的链路状态数据库拷贝来执行相同的算法，以构造包括最小成本路径树的唯一路径表，如由 IGP 度量定义的那样，其中路由器的最小成本路径树是以其自身为根到每个目的地的。所得的最小成本路径变成遍历网络的所有未经分类的分组所采用的缺省路径。

15 如上所述，AS 中的多组网络可以被一起分组成路由区域或区段。区段的拓扑不与含有该区段的 AS 的其它部分共享，以提供路由流量的显著减少。在区段之间交换总结分组（summary packet）以确保区段之间的连接性。

20 在初始生成其链路状态数据库和路由表后，如果在其网络区段内发生变化，则每个路由器重复信息交换和路由计算过程。变化可能涉及链路或路由器的添加或去除，或者链路成本的变化。为了避免链路状态数据库变为过时的可能性，在没有新的更新的情况下，分组被周期性地重新广播，通常间隔为 30 分钟。

25 通过使用链路状态 IGP，例如 OSPF 或 IS-IS，本发明实现了 AS 中的网络拓扑的非干扰式（non-intrusive）的发现，并且实现了该拓扑的带注释表示的创建，以促进通过网络的网络范围内的路径集的后续发现。带注释的表示借助有向图（directed graph）来描述 AS，其中顶点（vertice）代表路由器或网络，边（edge）代表连接到路由器的链路。注释指示所发现的与由每个顶点代表的路由器或网络有关的数据。对于路由器的情形，注释

指示关联的 IP 地址、由 IP 地址指示的一组接口以及类型或功能（区段内、区段之间或自治系统之间）。对于网络而言，示出了由 IP 地址指示的关联的网络地址和网络掩码以及网络类型（存根（stub）、转接或外部）。

5 网络。存根网络类似于死胡同（cul-de-sacs），而外部网络是到 AS 之外其它网络的目的地。

图 2 中示出了根据本发明产生的图形示例的可视化表示。图形的边连接个体顶点。当两个路由器通过物理点到点链路附着到一起时，边连接两个路由器，而将路由器连接到网络的边指示该路由器具有网络上的接口。

10 如 IGP 所定义的那样，以使用该接口进行分组转发的成本来对每条边进行注释。在 OSPF 中，这被称为链路度量。

本发明的一个目的是生成一个拓扑，同时对网络或路由器的正常操作影响有限。该拓扑发现过程在下述意义上是非干扰性的，即，借助和路由器或其它网络元件的有限主动交互来获得所需要的信息，而不产生显著的

15 额外网络流量。为此目的如图 1 所示，在存在 OSPF 分组的点处，至少一个探测器或监控器 40 被连接到 AS。探测器例如可以是运行专用软件程序并通过以太网卡连接到 AS 的低成本计算机，例如“个人计算机”。对连接到网络的“逻辑”点进行选择，以确保可以收集由路由器广播的 OSPF 分组。物理上，该连接点可以是例如经由集线器或开关来自子网络的两个

20 路由器之间的链路中的抽头，或者路由器上的端口。在 OSPF 条件下，在被 OSPF 分组所遍历的网络中的任何点处都需要连接。对于这些物理连接，探测器 40 中的软件程序以“混合模式（promiscuous mode）”开通到所选网络区段的网络链路或段的连接。混合模式允许探测器接收需要的 OSDF 分组而不必考虑它们的 LAN 目的地地址。所接收的分组被允许在没

25 有干扰的情况下继续它们的旅程通过网络（而不是被接收并被从网络上去除）。或者，OSPF 和其它 IGP 路由分组也可以被路由器自己存储在它们自身的 LSDB 中，并且为调试目的可以通过 SNMP MIB 以原始字节格式来利用。最初，对于拓扑中的每次后续改变，还可以以有限的网络影响和开销从 MIB 收集分组。对来自 MIB 的原始分组的收集需要很少数量的

SNMP 呼叫而不是多个 SNMP 呼叫，其中多个 SNMP 呼叫是使用 SNMP 的现有方法的特征。可以使用 SNMP 陷阱 (trap) 来跟踪 MIB 数据的改变，SNMP 陷阱一旦被设置，就将向外部应用通知目标 MIB 数据的改变。

探测器 40 没有实现如 RFC 2328 描述的状态机来建立与任何路由器的邻接，因为那将需要探测器变成 OSPF 路由协议中的主动参与者，从而在该区段的其它路由器中创建虚假的链路状态数据库条目。对于通过监控链路来收集分组，探测器 40 保持是被动的，并且依赖于区段或 AS 中的路由器进行的 OSPF 分组的洪泛 (flooding) 过程。探测器 40 等待 OSPF 分组到达被监控的接口，而不是使用正常的 OSPF 机制来请求它们。在接收到每个 OSPF 分组之后，执行拓扑导出过程 (下面将描述并在图 8-13 中举例说明)，以增量创建期望的拓扑描述。在确定完整的拓扑描述之前，启动过程需要已经经过缺省链路状态刷新时间间隔，该间隔通常是三十分
10 钟。变成 OSPF 路由协议中的主动参与者可以加速分组的收集并减少完整发现过程的时间；这种情况下的副作用很明显是增加了对网络和路由器的影响。或者，在探测器 40 从 MIB 中含有的原始分组收集 OSPF 路由数据
15 时，可以在不需要启动周期的情况下获得完整且最新的拓扑。该方法还限制了对路由器或网络的影响。因此，通过继续跟踪 OSPF 分组，探测器可以保持与网络状态步调一致的拓扑描述。

AS 所需要的探测器的数量取决于 AS 的大小及其组织的方式。单个探测器可以为它连接到的区段生成完整的带注释的拓扑。OSPF 网络通常具有至少一个区段，其被称为主干 (backbone)，到这个主干的连接是优选的。经验表明，许多网络在设计上是分层次的，连接到主干的单个探测器将提供非常有用的带注释的拓扑。为了发现多区段 AS 的完整的带注释的拓扑，需要到每个活动区段的连接。然而，除了提供所选区段的完整的带
20 注释的拓扑外，即使单个连接也将提供 AS 中的其它区段中的网络总结信息，加上通过 AS 的边缘路由器到外部网络的任何连接。

每个探测器 40 收集下面描述的分组并且拷贝所选类型的分组。它随后从这些拷贝中提取数据并处理该数据以得到用于带注释的拓扑的信息。

在 OSPF 协议中定义了五种类型的分组，如下表所示。为了本发明的

目的，使用了这些 OSPF 分组中的一种，链路状态更新分组，第四种类型。

	<u>类型</u>	<u>描述</u>
	1	打招呼 (Hello)
5	2	数据库描述
	3	链路状态请求
	4	链路状态更新
	5	链路状态确认

打招呼分组也存在于 OSPF 网络上，例如在诸如因特网的广播媒介上，并且被最频繁地传输，其在给定网络段上定期出现。打招呼分组因而可以被用于向探测器 40 提供网络时间的准确指示。在打招呼分组到达探测器时被施加准确的时间戳。例如，基于个人计算机的探测器可以从其内部时钟或从全球定位系统 (GPS) 结合网络时间协议 (NTP) 来获得时间的准确指示。经验显示，大多数网络操作员提供可以用于此目的的准确时间服务。通过测量多个打招呼分组的间隔之间的时间 (inter-interval time) 并存储该结果，可以建立由正常 OSPF 分组序列驱动的时间流逝 (passage of time) 的准确内部表示。这种形式的定时器机制是提供用于确保从探测器 40 清除过时信息的时间基准的简便方式。然而，可以提供时间流逝准确指示的任何其它形式的定时器机制也是足够用的。

链路状态更新分组 (OSPF 类型 4) 含有一个或多个链路状态通告 (LSA)，其描述网络或路由器的状态 (包括路由器接口和邻接的状态)。一个区段的 LSA 的收集构成链路状态数据库。如下所示，存在几种类型的 LSA，每个 LSA 类型表述 AS 或网络区段中的不同元件。

	<u>LS 类型</u>	<u>描述</u>
25	1	路由器-LSA
	2	网络-LSA
	3	总结-LSA (IP 网络)
	4	总结-LSA (ASBR)
	5	AS-外部-LSA

只要发生网络配置中的改变，就定期广播 LSA，以确保过时的信息不会存在于网络中。每个 LSA 具有一个头部部分（图 3 中示出），该头部具有关键字（包括头部中的字段的组合）和年龄信息，其给出 AS 中的 LSA 的唯一身份。确定 LSA 是否应当被收入到链路状态数据库中的方法在 RFC 2328，13.1 和 13.2 节中有描述，并被探测器 40 用来确定其接收的 LSA 与其已经接收的现有 LSA 相比是否更新，以及该 LSA 是否应当被收入到其自身的链路状态数据库中。

随着探测器的内部时钟被更新，新的时间值被用来增加链路状态数据库中的每个 LSA 的年龄字段。如果 LSA 的年龄值因此变得大于 OSPF 标准体系结构常数“最大年龄”（其通常被设置为一小时），则该 LSA 被从链路状态数据库中去除（如图 8 所示）。该方法提供了一种安全措施，确保过时的 LSA 被从探测器的链路状态数据库中去除，使得如果更新的 LSA 被探测器错过或者因临时链路故障而丢失时，由探测器 40 提供的拓扑描述将不会被不当地破坏。

当探测器已经首先汇编了它的链路状态数据库时，以及检测出已发生该探测器的链路状态数据库中的后续改变之后（例如，跟在接收到新的或更新的 LSA 之后），必须创建或刷新探测器的当前网络拓扑的带注释的描述。现在将参考图 8-13 来描述实现上述内容的过程。尽管第六步必须最后执行，但是所涉及的大多数步骤的准确顺序并不是关键性因素。同等地，可以为每个链路状态数据库的改变来完全重新计算拓扑，或者仅就所处理的最近 LSA 改变来增量计算。两种方法同等有效，并且可以选择在具体实施方式中实现起来更为简单或更合适的方法。在下面描述的例子中，首先标识拓扑顶点是方便的并且遵照常规图形构造技术。

步骤 1（图 9）：标识区段内的活动子网络以及这些子网络中的活动路由器；这通过使用含有当前区段内的网络元件信息的 LSA 来完成，具体而言是第二类网络-LSA 和第一类路由器-LSA 的子集。网络-LSA 指定附着到支持不只一个路由器的子网络的路由器。网络-LSA 中的网络掩码字段（见图 5）描述了子网络的地址空间的大小和范围，并且 LSA 头部的链路状态标识符字段（图 3）中的 IP 地址标识了子网络中的第一 IP 地

址。还提供了该子网络上的活动路由器的列表，路由器由附加路由器字段（图 5）中的 IP 地址来指示。每个 LSA 为子网络上的每个活动路由器含有一个条目。

- 可以根据所描述的链路类型来细分路由器—LSA，并且每个路由器—LSA 可以描述不同类型的几种链路。连接的类型如下所示：

<u>类型</u>	<u>描述</u>
1	与另一路由器的点到点连接
2	到转接网络的连接
3	到存根网络的连接
10 4	虚拟链路

在此步骤中仅考虑含有关于第三类到存根网络的链路的信息的那些路由器—LSA。对于用于描述到存根网络的连接的每个路由器—LSA，每个链路标识符字段（图 4）和随后的链路数据字段给出路由器上用于到存根网络的连接的 IP 地址和网络掩码，所述路由器由 LSA 头部的通告路由器字段来标识。如果子网络上的倒数第二个路由器发生故障，使得子网络不再具有两个或更多路由器，则对应的网络—LSA 可能不会从链路状态数据库中主动撤销。在这种情形下，尽管网络—LSA 仍旧存在，但是其被含有描述到存根网络的连接的条目的第 3 类路由器—LSA 所取代。因此，为了确保在本步骤中仅考虑活动网络上的活动路由器，在这两类 LSA 中含有的信息被组合在一起，使得由第 3 类路由器—LSA 中的条目定义的路由器优先于有关在网络 LSA 中定义的相同路由器的信息。

步骤 2（图 9）：指定拓扑的内部网络顶点。为步骤 1 中得到的活动子网络的列表中的每个活动网络创建顶点。使用 IP 地址和网络掩码来对顶点加以注释，从而指定由顶点标识的子网络的身份和地址范围。这些顶点还带有类型为“内部网络”的注释。

步骤 3（图 9）：指定拓扑的路由器顶点和他们的关联接口：

步骤 3.1：分析含有用于描述到转接网络的连接的条目的第 2 类路由器—LSA。这些 LSA 描述具有到子网络的连接的路由器，这些子网络具有不只有一个入口/出口点。对于每个 LSA，由 LSA 头部中的通告路由器字段

标识的路由器的 IP 地址被添加到顶点列表中。由链路数据字段中的 IP 地址标识的活动路由器接口列表与该顶点条目相关联。在此上下文中，路由器上的接口是进行网络连接或链路的端口的同义词。

5 步骤 3.2：分析含有用于描述作为点到点连接的链路的条目的第 1 类路由器—LSA。如前所述，路由器 IP 地址被加入到顶点列表中，并且由链路数据字段标识的路由器接口的 IP 地址也被加入。

步骤 3.3：类似的过程也被用于含有用于描述虚拟链路（虚拟链路在 RFC 2328 的 3.1 和 15 节中进行了描述）的条目的第 4 类路由器—LSA 中。

10 步骤 3.4：接下来分析含有关于到存根网络的连接的信息的第 3 类路由器—LSA。该过程与用于含有描述到转接网络的连接的条目的第 2 类路由器—LSA 的过程相类似。然而，在这种情况下，路由器地址自身作为关联的路由器接口而被加入。这种类型路由器—LSA 条目的链路数据字段没有描述路由器的接口，而是描述了所连接的存根网络的网络掩码。因此，
15 不能确定路由器接口的 IP 地址接口。为了指定连接的目的，如下面所要描述的那样，这类链路的开始点被认为是路由器自身。

步骤 3.5（图 10）：所指定的路由器顶点被注释有他们的关联类型。根据路由器—LSA 中的 VEB 字段（图 4）中的 E 和 B 标志来标记路由器类型。如果设置了 B 标志，则路由器被标记为区域之间；如果设置了 E 标志，
20 则路由器被标记为 AS 之间或网络之间；否则顶点被标记为区域内。如果设置了 V 标志，则顶点是一个或多个虚拟链路的端点。

步骤 3.6（图 11）：对于每个 AS 外部—LSA，加入如 LSA 头部的通告路由器字段所标识的，且带有如“转发地址”字段（图 7）中的 IP 地址所标识的关联接口的路由器顶点，前提是还不存在这样的顶点。顶点被注释为“AS 之间的路由器”。类似地，对于每个总结—LSA，进行下述检查，
25 即检查存在用于由通告路由器字段中的 IP 地址标识的路由器的顶点以及该顶点被注释为“区域内的路由器”。该步骤具有两个目的：在还没有收集到完整拓扑之前，检查数据的完整性并加速探测器启动上的发现过程。

步骤 4.1 (图 12)：指定拓扑的区域之间的网络顶点。为此考虑第 3 类和第 4 类总结-LSA。这些 LSA 描述到包括网络 (第 3 类总结-LSA) 或区域之间的路由器 (第 4 类总结-LSA) 的区域之间目的地的连接。对于两个类型中的每个总结-LSA，由网络掩码字段和通告路由器字段的 IP 地址标识的网络顶点被加入到顶点列表中。这些顶点被注释有“总结网络”类型。

步骤 4.2 (图 12)：指定拓扑的 AS 之间的网络顶点。第 5 类 AS 外部-LSA 被用来指定用于表示到在含有探测器 40 的网络外部的网络的路由的一组外部顶点。这些路由的存在已经通过预配置的静态路由描述或诸如 BGP-4 的外部网关协议而得知。对于这些外部路由中的每个，OSPF 路由器将发布 AS 外部-LSA。对于每个 LSA，一个顶点被加入到顶点列表中，用于由网络掩码 (图 7) 和 LSA 头部的链路状态标识符字段中的 IP 地址标识的网络。该顶点被注释有“外部网络”类型。

步骤 5 (图 12 和 13)：指定网络中的边。

步骤 5.1：指定转接边。含有用于描述到转接网络的连接的条目的第 2 类路由器-LSA 被用于指定图形中的边，这些边将用来表示路由器的顶点和用来表示提供通过或转接服务的网络的任何顶点互连 (转接网络是具有两个或多个单独入口/出口点的网络)。对于含有用于描述到网络的转接连接的条目的每个路由器-LSA，在正逐渐形成的拓扑描述中从链路标识符字段 (图 4) 中定义的路由器接口到由链路数据字段定义的子网络，指定一条边 (根据 RFC 2328 “当连接到也产生 LSA 的目标 (即，另一路由器或转接网络) 时，链路标识符等于邻近 LSA 的链路状态标识符”)。因此，带有对应网络-LSA 的“链路状态标识符”的子网络被用于确定正在被指定的边的终点。所述边注释有遍历所述链路的成本，如度量/成本字段中所定义的那样。值得注意的是，可能存在不只两个连接到子网络顶点的边。

步骤 5.2：指定存根边。含有用于描述到存根网络的连接的条目的路由器-LSA 被用于指定相关路由器顶点和只有一个入口和出口点的网络顶点之间的边。对于含有用于描述到存根网络的连接的条目的每个第 3 类路

由器—LSA，加入一个起点在路由器接口处而终点在存根网络处的边；在本例中接口和路由器自身具有相同的地址（直接有效寻址路由器）。边起始点由通告路由器字段指示，而目的地是由链路标识符和链路数据字段定义的子网络。子网络地址由链路标识符字段中的 IP 地址和链路数据中的网络掩码来指示。所述边注释有在度量/成本字段中定义的遍历所述链路的成本。

步骤 5.3：指定点到点边。描述点到点和虚拟链路的路由器—LSA 被用于指定直接互连路由器顶点的边。虚拟链路在 RFC 2328 的 3.1 和 15 节中进行了描述，并且为了生成拓扑的目的，它们可以以和点到点链路相同的方式来操纵。对于含有用于描述到另一路由器的点到点连接的条目的每个路由器—LSA（第 1 类）或者含有用于描述虚拟链路的条目的每个路由器—LSA（第 4 类），将一条边加入到逐渐形成的拓扑中。所述边从由链路数据字段中的 IP 地址指示的路由器接口开始，且目的地路由器由链路标识符字段中的 IP 地址指示。所述边注释有度量/成本字段中定义的遍历所述链路的成本。

步骤 5.4：指定区域之间的边。总结—LSA 被用来指定将路由器顶点连接到用于描述任何区域之间目的地的顶点的边。存在两种类型，第 3 类描述的目的地是 IP 网络，第 4 类描述的目的地是其它区域之间的路由器。对于每个第 3 类总结—LSA，从路由器的接口（在本例中具有和路由器自身相同的地址，直接有效寻址路由器）到如链路状态标识符字段和网络掩码字段所定义的那样的区域之间的子网络加入一条边。对于第 4 类总结—LSA，网络掩码字段没有意义因而必须为零，而链路状态标识符是 AS 之间路由器的 IP 地址。在两种情形下，所述边都注释有度量/成本字段（图 6）中定义的遍历链路的成本。

步骤 5.5：指定 AS 之间的边。AS 外部—LSA 被用来指定将路由器顶点连接到用于描述 AS 之外的任何外部目的地的顶点。对于每个 AS 外部—LSA，从由转发地址字段（图 7）指示的路由器接口到由 LSA 头部中的链路状态标识符字段和网络掩码字段（图 7）定义的外部网络，添加一条边。所述边注释有由 E 比特字段和度量/成本字段定义的遍历链路的成本。

如果 E-比特未被设置，则以和其它边的其它内部链路度量相同的单位来定义度量或成本。如果设置了 E-比特，则链路的成本被认为大于任何其它内部链路状态路径。

5 步骤 6 (图 13)：维护可行路径的图形。探测器 40 必须消除任何过时信息，从而确保只有可行网络路径被报告给使用拓扑信息的流量管理或其它应用。例如，存在下述的可能性，即探测器的链路状态数据库可能含有在网络断电或故障之前到达的 LSA，其中网络断电或故障引起网络的拆分。来自位于网络上远离拆分故障点一端的路由器的新 LSA 更新将不能到达探测器 40，其中在探测器 40 处它们将被用来去除过时信息。为了维护
10 准确的拓扑描述，探测器 40 必须消除代表受影响的路由器、网络和链路的顶点和边。

例如，在图 1 的情况下，路由器 16 和 18 可能由于电源故障而崩溃。图 14 示出了所得的网络配置。探测器 40 将继续从在其一侧的故障或“网络拆分”处的路由器 10 和 14 接收分组。然而，位于该拆分之外的路由器
15 12 不能与路由器 10 和 14 中的任何一个来传输改变。从而探测器 40 和路由器 10 与 14 中的链路状态数据库将继续包含从路由器 12 发送的 LSA。但是，这些 LSA 中的信息不再被认为是可靠的，因为所述信息来自探测器当前知道的路由区段之外。探测器 40 的目的是创建所有可能和可行网络路径的描述，所以这个描述不应该包括故障路由器 16 与 18 所处点之外的
20 的网络部分。

通过使用递归过程来系统地检查所有顶点，评价图形中的每个顶点的到达能力，其中递归过程开始于表示下述地点的顶点，所述地点为探测器 40 连接到网络的点。存在基于例如“宽度优先搜索”和“深度优先搜索”算法来确定图形中的到达能力的多种公知过程。在这些算法中，只有到每个
25 顶点的连通边才被考虑，因而图形中的未连通部分，在本例中为路由器 16 和 18 将被消除，并且结果是一个“连通图”，其消除了从路由器 12 到路由器 13 的路径以及之外的路径。

要注意递归过程的起始顶点可以是路由器或网络，这取决于哪个元件在逻辑上最接近探测器 40。例如，如果探测器通过抽头直接连接到点到点

连接上，则起始点是逻辑上最近的路由器，或者，如果探测器连接到转接子网络上，则应当使用该子网络。如果探测器连接到路由器上的 MIB，则应当使用该路由器。转接子网络通过每个与其连接的路由器提供到整个网络的多条可到达连接，并且在任何到达能力计算期间，必须考虑这些多条连接。所得的图形可以被输出到外部应用（例如，用于流量管理），并且仅包括探测器 40 当前知道的活动互连顶点。

通过适当形式的进程间通信，探测器 40 可以将图形输出到一个或多个外部应用。例如，可以使用公知的远程过程调用机制，或者 Java 远程方法调用（RMI）或公用对象请求代理体系结构（CORBA）的标准中描述的机制。用于探测器 40 的软件也可以直接嵌入到应用软件中，以创建简单、小型、轻便的便携式系统，其可以由操作员按需绕网络传输。

带注释的图形数据可以得到利用，从而在改变发生时，使应用认识到每个改变。这有时被称为“公布与预定”机制，由此在公布改变时，应用预定该改变。或者更为简单，新的拓扑可以在要求时被传送到应用。

输出的拓扑信息可以具有几种形式，但将包括活动顶点和活动边的某些类型的列表。活动顶点的列表通常包括：由 IP 地址和网络掩码或前缀长度指示的顶点身份；顶点表示的网络元件的类型，例如网络、AS 之间、区段之间与区段内、以及它们属于的区段和区段内。还可以包括由 IP 地址指示的接口的列表。边的列表通常包括它们连接的顶点（由 IP 地址和网络前缀指示）和使用该边所表示的链路的成本或度量。边的列表中还可以包括路由器上使用的接口，其也是由 IP 地址指示。

如果应用使用“公布与预定机制”在改变发生时请求该改变，则可以通过以 IP 地址和网络前缀来指定顶点，来去除变为不活动的边与顶点或网络改变。类似地，要被去除的边可以以两个连接的顶点来标识。当向应用通知边或顶点不再是活动的时，还可以提供诸如边度量或顶点类型的注释信息，但这不是必须的。

应用使用带注释的图形数据，通过例如提取出那些用于提供任何 AS 之间的连接性的顶点列表来确定当前活动逻辑路径的集合。这一类型的顶点包括网络的入口和出口点，并且是如上所述应用流量分类的最可能的地

点。该应用可以直接查询由那些顶点表示的路由器，来确定是否确实存在流量分类；一种可以用于这样的查询的公用机制是 SNMP（RFC 1157）。用于所选流量管理系统的关联管理信息库（MIB）和 SNMP 在绝大多数路由器上是可用的，并且提供一种用于访问这类网络管理数据的广泛接受的机制。如果存在下述可能，即在网络上存在任何内部流量分类，则还可以搜索内部路由器顶点。这较少发生，但在某些情形下可能发生。

例如，如果流量分类正被用来通过 MPLS 标签交换路径（LSP）而将流量从给定的供应商沿着并非缺省最少成本路径的路径进行路由，则外部应用可以请求与当前使用的实际非最短路径有关的路由返回信息。SNMP 可以再次被用来获取包括已被分配给该逻辑路径的任何传输特性的这一路径信息，所述传输特性例如是预留带宽。

所发现的拓扑数据可以被用来确定网络范围内的路径集，包括用于该拓扑的缺省路径集。要注意，来自不同源路由器的多个逻辑路径可能潜在地遍历单个互连。结果，当确定替代性逻辑路径时，必须考虑网络范围内的路径集。没能考虑网络范围内的路径集可能导致路由器、子网络或互连上的阻塞和过多指定，其中所述路由器、子网络或互连服务于来自不同源/入口路由器的多条逻辑路径。需要网络范围内的路径集来确保这些计算的有效性。为了确定这个路径集，使用递归过程基于“宽度优先搜索”或“深度优先搜索”和 Dijkstra 算法（RFC 2328 中描述）来执行图形的遍历。这为入口/源和出口/目的地路由器的每个组合提供了网络范围内的路径集，包括最短路径。该算法的输入是有关路由器、子网络和互连的所发现的图形数据以及 IGP 成本度量。

例如，网络操作员可以使用组合信息（包括网络范围内的路径集和与它们关联的通过 AS 的成本/度量）结合覆盖请求的流量管理信息（有关 LSP），来监控逻辑非缺省路径部署。该组合信息为网络操作员提供下述方面的有用帮助，例如设计新路径、LSP 供应以及确保网络符合设计规范。

例如，通过比较网络范围内的路径、缺省路径、活动逻辑路径和与该活动路径相关联的路由目标，可以生成一组符合与该活动逻辑路径相关联

的路由目标的替代性逻辑路径。参考图 1（在本例中为简单起见假定 IGP 路径成本类似于所遍历的路由器数量，且所有链路具有相等的最大容量），应用将计算出从路由器 10 到路由器 12 的缺省路径是通过路由器 18 的。在路由器 10 和 18 之间发现活动 MPLS LSP，其需要等于链路最大容量的 75% 利用率的预留带宽。该 LSP 已经应 AS3 的管理者的请求而被建立，该管理者需要带宽的一个保障级别用于到 AS2 的连接。在路由器 10 和 12 之间发现了经由路由器 14 和 16 的第二路径，其被用于另一目的；该路径需要那些链路上的带宽的 20%。应用确定第一 LSP 在缺省路径上，并且第一和第二 LSP 的组合负荷不大于可用最大容量。因而应用可以推荐用于第一 LSP 的替代性路径可以经由路由器 14 和路由器 16，而不经由通过路由器 18 的缺省路径。

使用将 LSP 的任何改变（例如可以由因链路故障产生的内部转接网损失所引起）都通知给外部应用的机制，可以通过提供 LSP 改变的即时警告来帮助操作员减轻这种故障的影响。允许路由器提供反馈的这样一种机制是 SNMP 陷阱机制。SNMP 陷阱一旦被设置，将向外部应用通知目标 MIB 数据中的改变。只要 LSP 行为近实时反馈，新的 LSP 或 LSP 的特性的任何改变就可以被再次覆盖在改变的拓扑上。

因此，探测器 40 提供的带注释的拓扑能够在包括上述任务在内的多种网络管理任务方面来帮助操作员。所描述的过程还可以应用于但不限于其它形式的流量管理和其它技术，例如区分服务、虚拟个人网络（VPN）、IP 语音、SLA 和 QoS 机制。所述其它形式的流量管理和其它技术采用通过在入口和出口路由器处的分组分类在单独逻辑路径上的路由，作为一般最小成本路径路由的替代物。

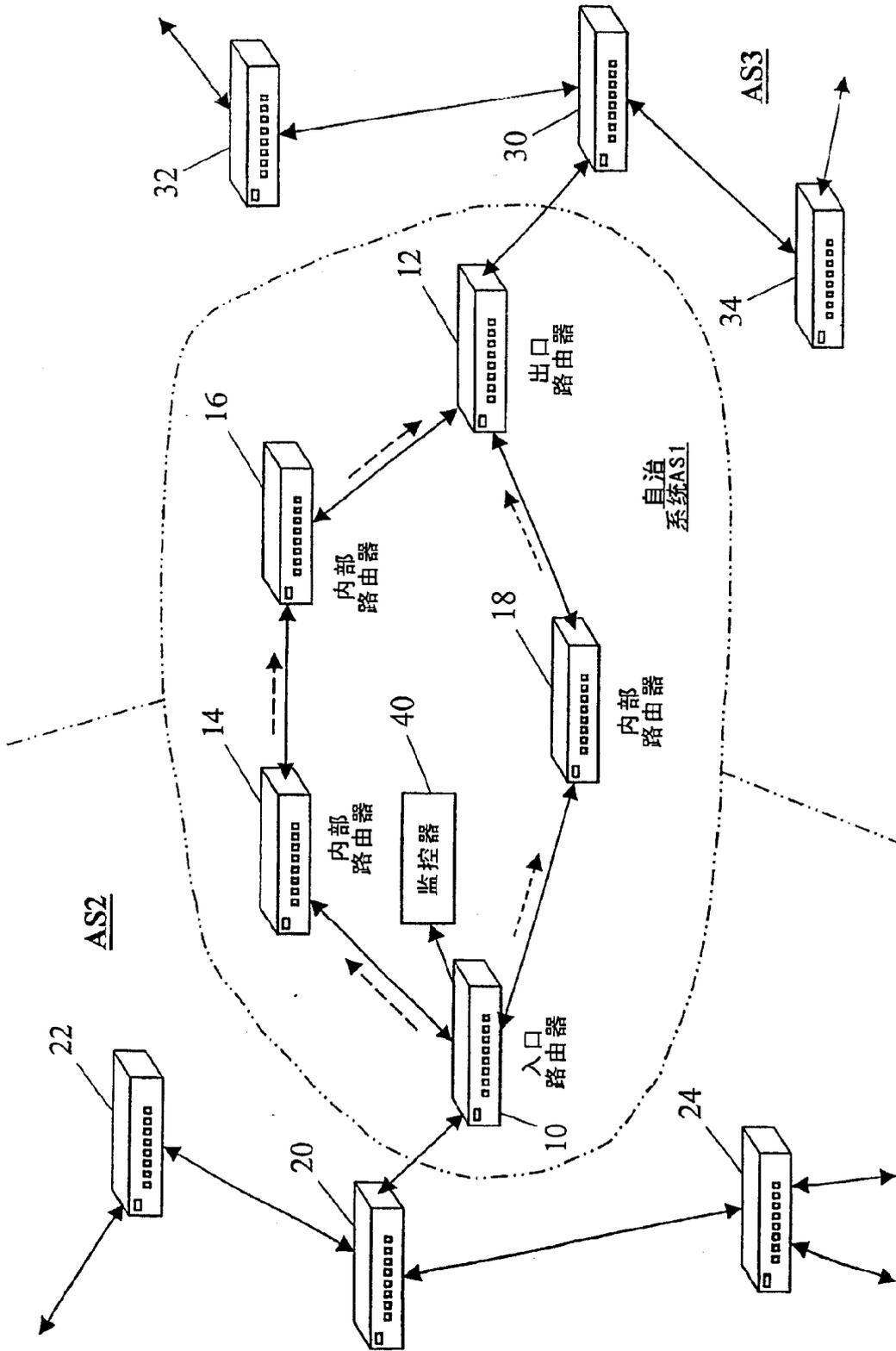


图1

LSA头部格式(所有LSA以此头部开始)

LSA年龄	选项	LS类型
链路状态标识符		
通告路由器		
LS序列号		
LS校验和	长度	

图3

路由器-LSA格式(LS类型=1)

头部(见图3)		
00000	VEB	00000000 #链路
链路标识符		
链路数据		
类型	# TOS 度量	度量/成本
TOS	00000000	TOS度量/成本
...
链路标识符		
...		

图4

网络-LSA格式 (LS类型=2)

头部(见图3)	
网络掩码	
附加路由器	
...	

图5

总结-LSA格式 (LS类型=3或4)

头部(见图3)	
网络掩码	
00000000	度量/成本
TOS	TOS度量/成本
...	...

图6

AS-外部-LSA格式 (LS类型=5)

头部(见图3)	
网络掩码	
E	00000000 度量/成本
转发地址	
外部路由标签	
E	TOS TOS度量/成本
转发地址	
...	

图7

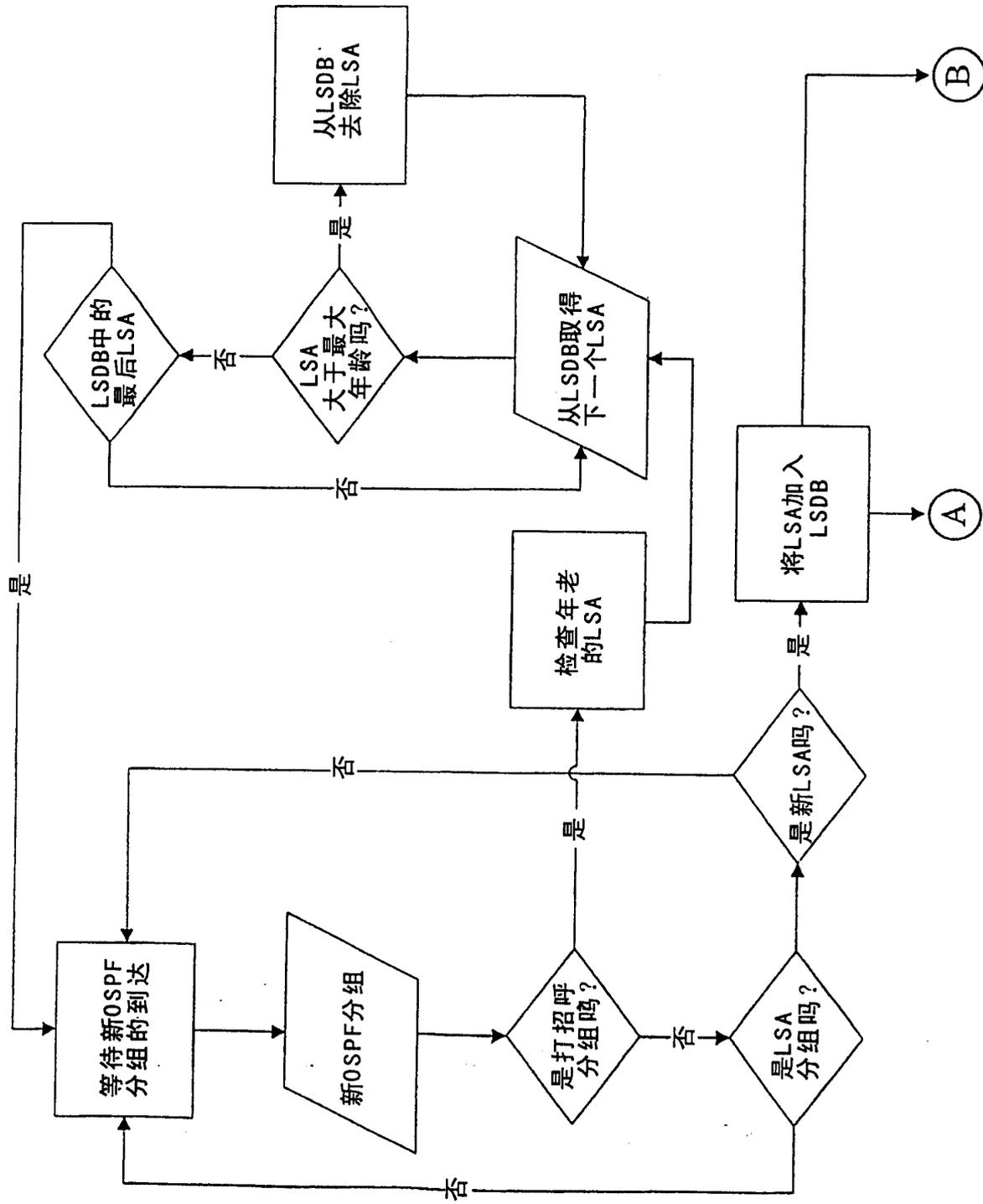


图8

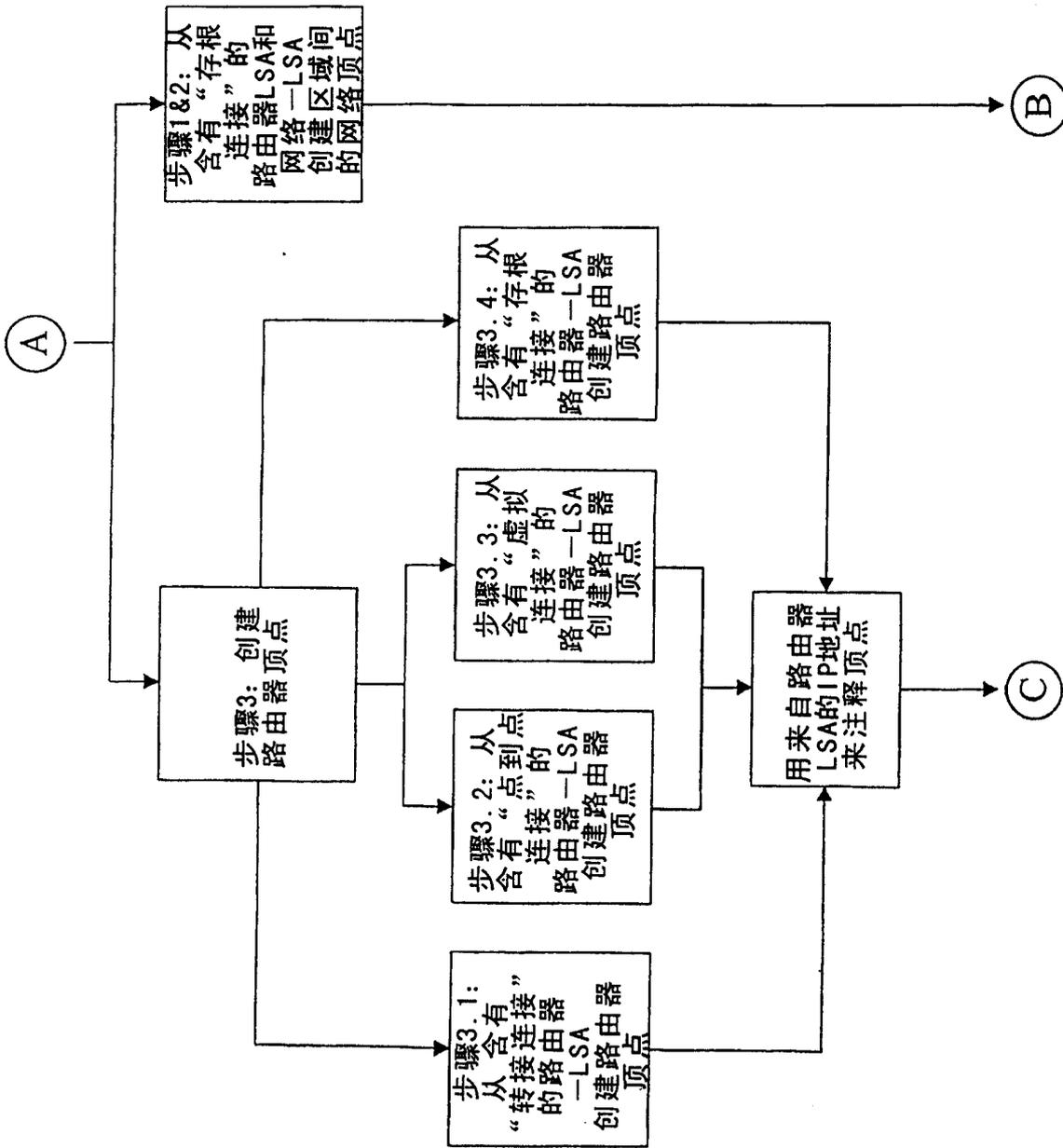


图9

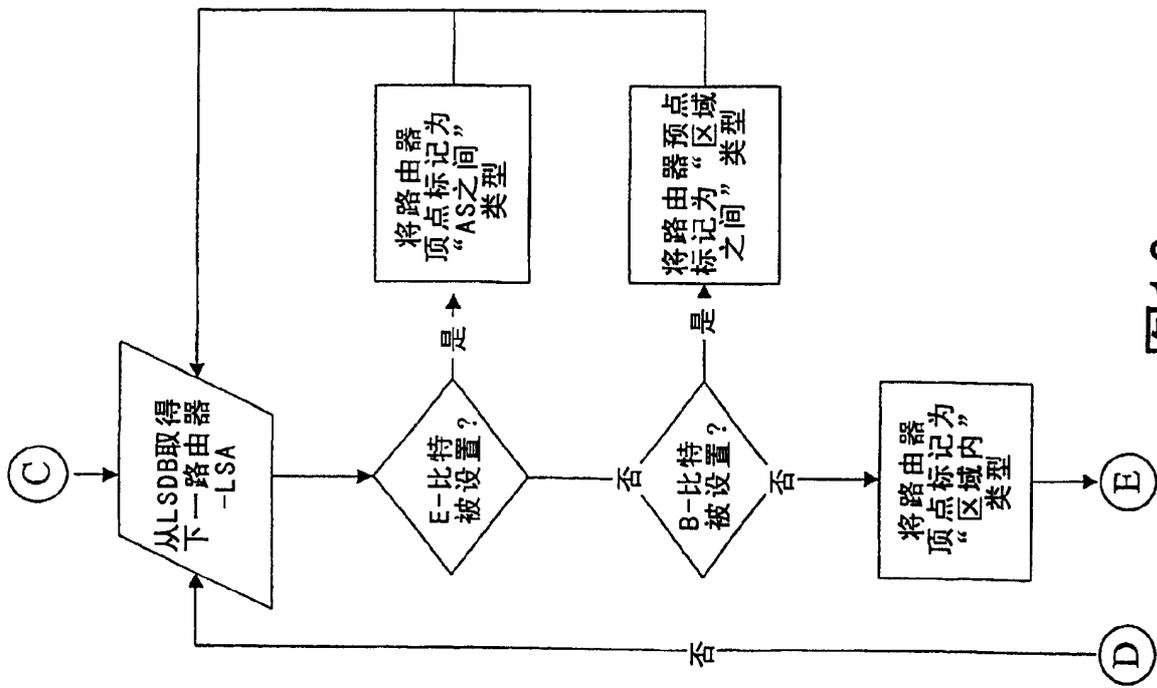


图10

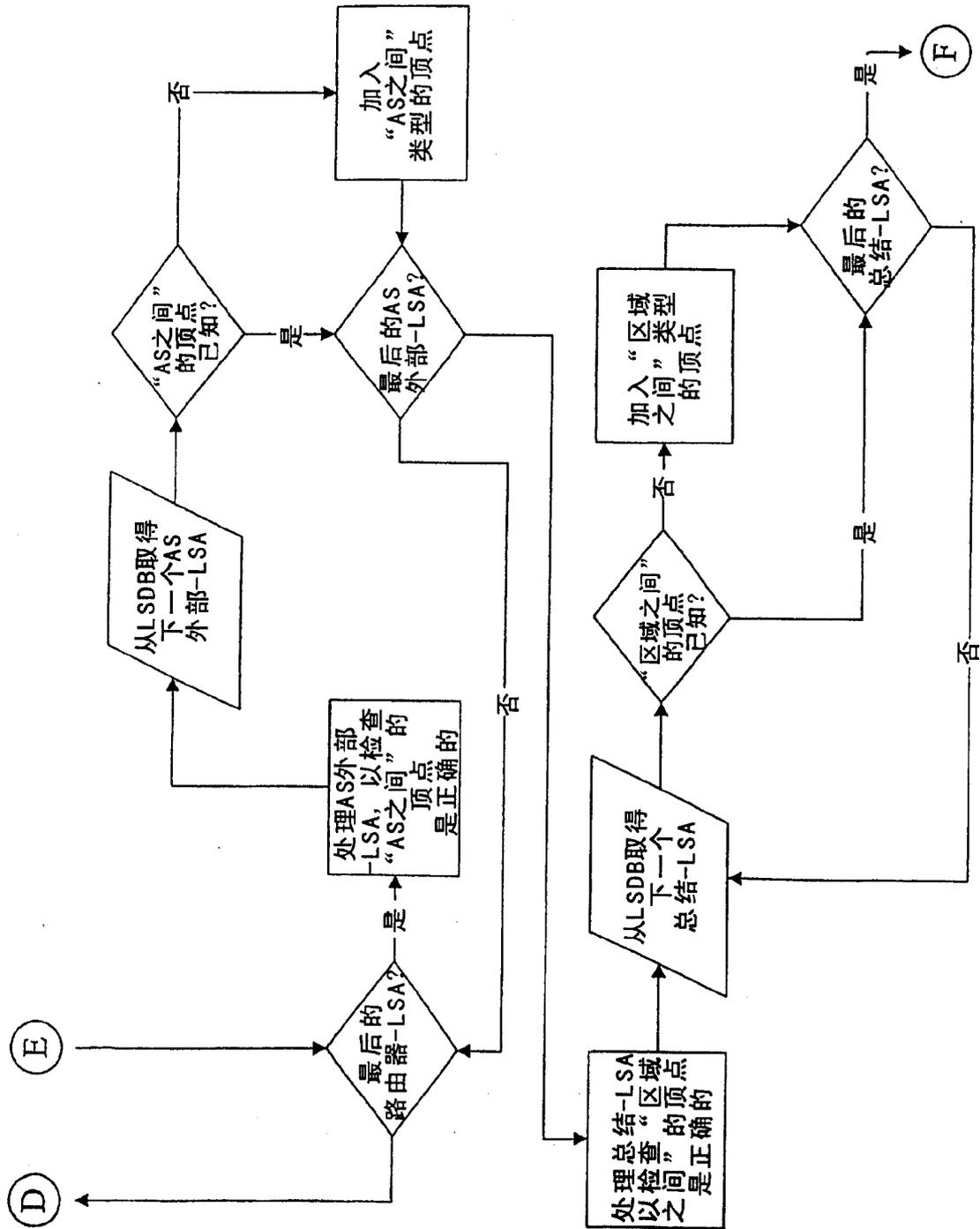


图11

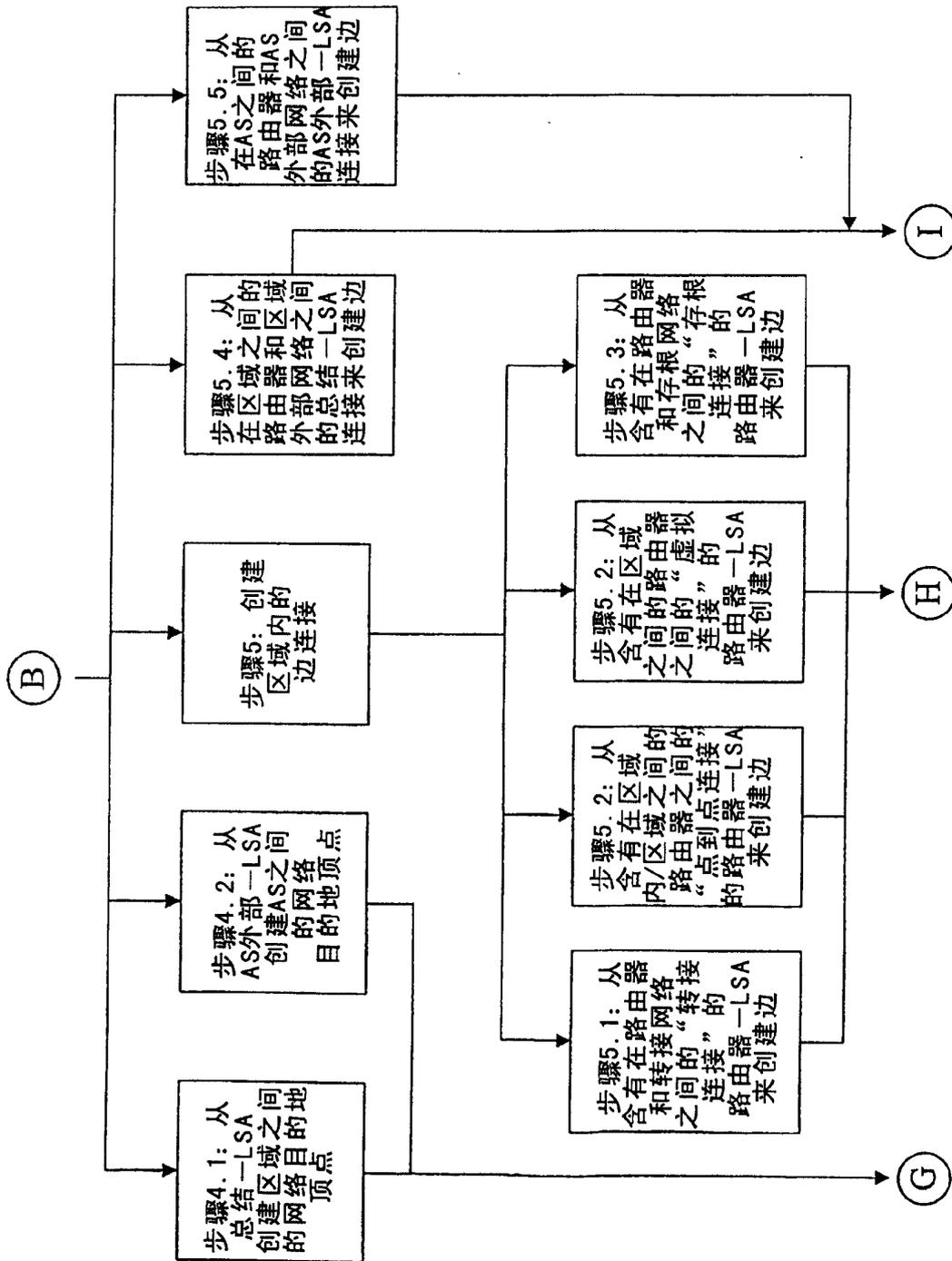


图12

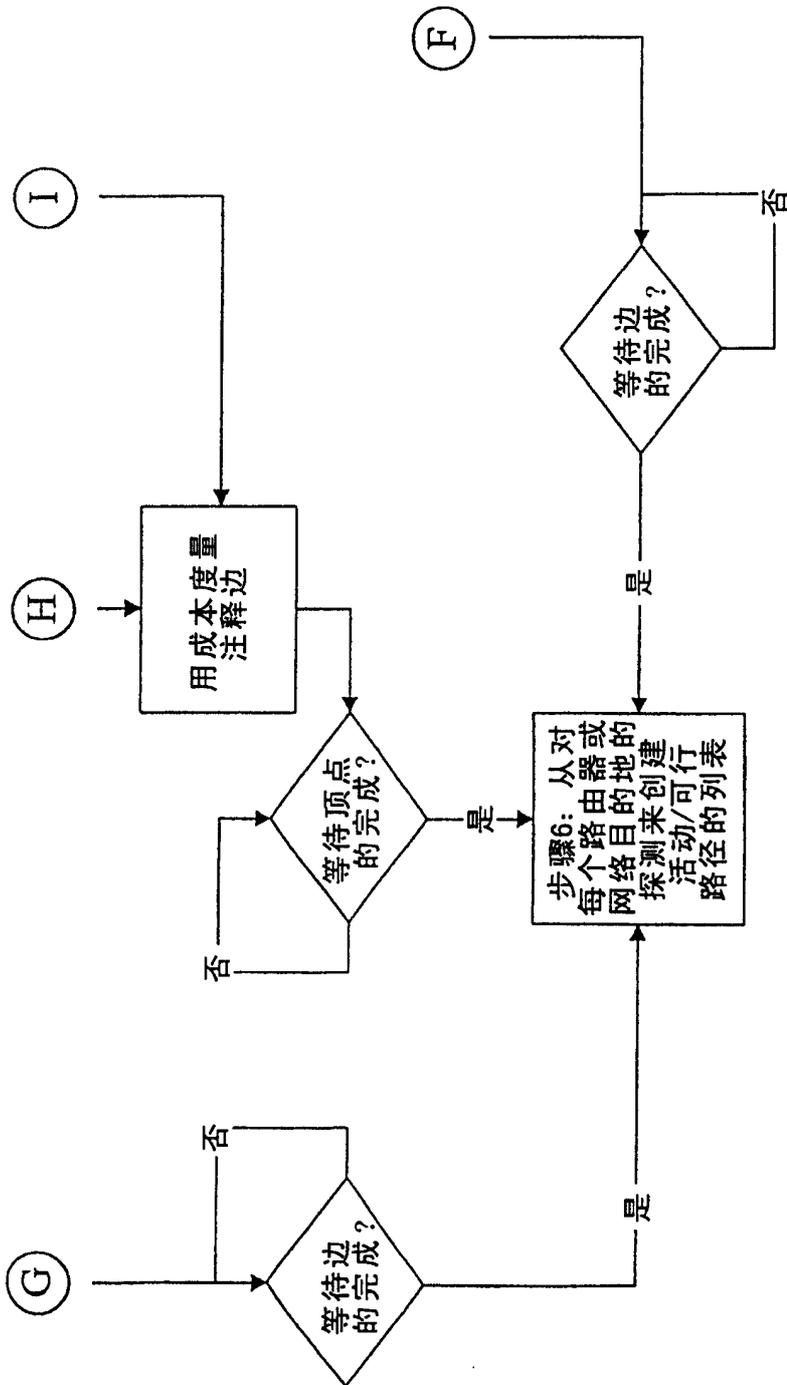


图13

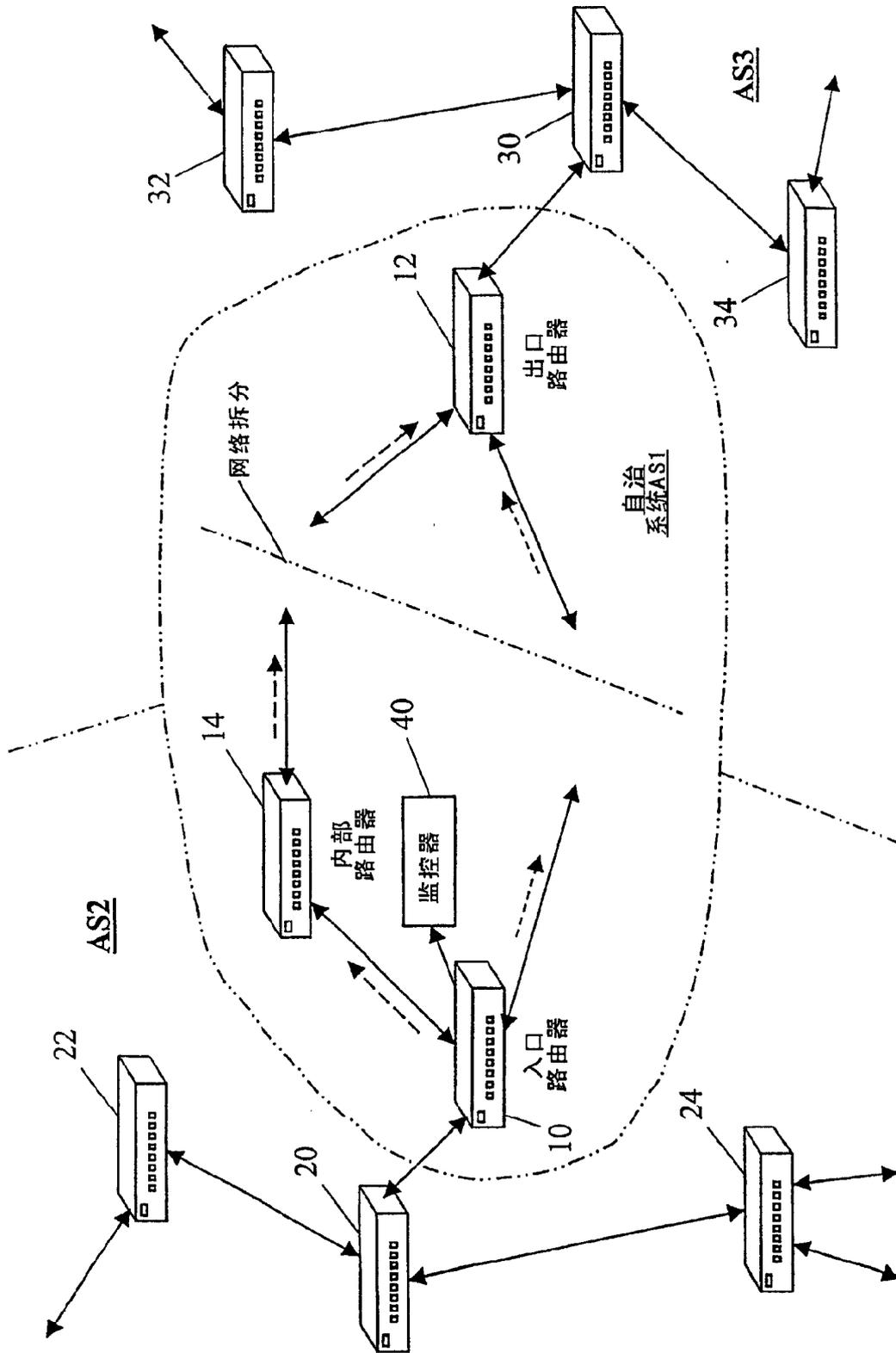


图14