

## (12) 发明专利

(10) 授权公告号 CN 101036330 B

(45) 授权公告日 2012. 02. 08

(21) 申请号 200580034136. 3

代理人 王怡

(22) 申请日 2005. 11. 30

(51) Int. Cl.

(30) 优先权数据

H04J 1/16(2006. 01)

11/001, 149 2004. 12. 01 US

(85) PCT申请进入国家阶段日

(56) 对比文件

2007. 04. 06

US 2003/0048754 A1, 2003. 03. 13, 全文 .

CN 1501644 A, 2004. 06. 02, 全文 .

W0 02/073902 A2, 2002. 09. 19, 全文 .

(86) PCT申请的申请数据

US 6813240 B1, 2004. 11. 02, 全文 .

PCT/US2005/043337 2005. 11. 30

US 2002/0118636 A1, 2002. 08. 29, 全文 .

(87) PCT申请的公布数据

US 6215765 B1, 2001. 04. 10, 全文 .

W02006/060491 EN 2006. 06. 08

CN 1452348 A, 2003. 10. 29, 全文 .

(73) 专利权人 思科技术公司

US 2002/0093954 A1, 2002. 07. 18, 全文 .

地址 美国加利福尼亚州

US 2004/0218542 A1, 2004. 11. 04, 全文 .

审查员 张德珍

(72) 发明人 詹姆斯·N·古伊查德

让-菲利普·瓦瑟尔

克拉伦斯·A·M·菲尔斯非斯

戴维·D·沃德

斯特凡诺·普雷维蒂

托马斯·D·纳德奥

(74) 专利代理机构 北京东方亿思知识产权代理

权利要求书 4 页 说明书 13 页 附图 7 页

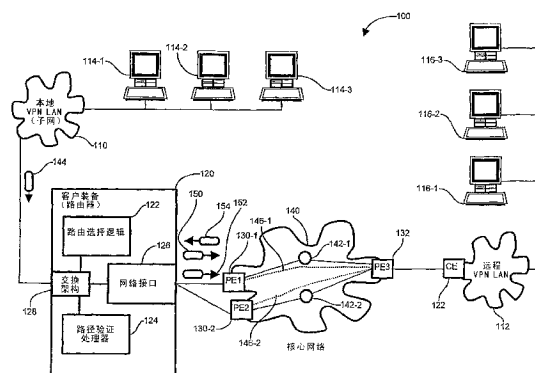
有限责任公司 11258

(54) 发明名称

用于检测网络故障的系统和方法

(57) 摘要

一种枚举发送到沿可疑路径的一组节点或路由器的一系列消息的路径验证协议 (PVP) 识别转发平面问题, 以实现控制平面级别的改变。这些消息包括这样一个命令, 该命令请求询问另一远程节点, 以便获得关于接收 PVP 消息的节点和另一远程节点之间的路径的信息 (202)。接收 PVP 消息的节点利用指示尝试达到另一远程节点的结果的命令响应作为答复 (206)。该系列消息总体上覆盖了沿着从发源者到接收者的路径的一组重要的预定路由选择点。对该系列 PVP 消息的聚集命令响应被分析, 以不仅识别整个路径是否可工作, 并且还识别问题的位置和性质 (308)。



1. 一种识别网络故障的方法,包括:

周期性地发送诊断消息到沿着去往目的地的路径的多个预定的路由选择点;

如果所述诊断消息指示沿着所述路径的中间节点存在问题,则发送一系列路径验证消息,每个所述路径验证消息包括一个命令,该命令可操作来引导中间节点发送另一消息到所述路径中的后继中间节点、接收来自所述另一消息的结果并且将所述结果作为命令响应进行报告,所述结果指示所述后继中间节点的可达性;

重复发送路径验证消息到指示所述问题的节点的沿着所述路径的后继节点;

分析接收到的来自发送到后继节点的路径验证消息的命令响应以识别问题;以及

基于所述分析确定替换路径,以绕过被识别为所指示问题的源的中间节点。

2. 如权利要求 1 所述的方法,其中分析还包括分析接收到的来自所述路径验证消息的命令响应,以识别所述命令响应指示有问题的第一中间节点和对所述诊断消息的命令响应指示正常操作的前面的中间节点。

3. 如权利要求 1 所述的方法,其中发送路径验证消息还包括:

根据诊断协议发送多个路径验证消息到多个预定的网络点;

接收与发送的路径验证消息相对应的命令响应,所述命令响应包括根据所述诊断协议的测试结果;以及

跟踪从沿着从源到目的地的路径发送的所述多个路径验证消息中的每一个接收的命令响应。

4. 如权利要求 3 所述的方法,其中所述中间节点表示网段,所述网段还包括客户设备和核心网络之间的本地网段,代表多个提供商设备的核心网段,以及所述核心网络和所述目的地之间的远程网段,跟踪还包括识别响应所源自的网段。

5. 如权利要求 4 所述的方法,其中所述网段还包括从客户边缘路由器到中间网络到远程边缘路由器的第一网段、提供商边缘路由器之间的第二网段和从提供商边缘路由器到远程客户边缘路由器的第三网段。

6. 如权利要求 3 所述的方法,其中分析还包括识别接收和未接收到命令响应,其中所述接收包括来自发送所述消息的预定网络点的可访问路径的指示,未接收表明根据诊断逻辑存在调解故障。

7. 如权利要求 6 所述的方法,还包括基于所述网络故障的位置和性质,识别在该处更改流量的网络点。

8. 如权利要求 7 所述的方法,还包括:

识别与具有实质性的逻辑路由选择价值的网络路由选择点相对应的重要前缀;以及

发送所述诊断消息到所述重要前缀。

9. 如权利要求 8 所述的方法,还包括确定活动前缀,所述活动前缀指示在前一阈值定时窗口期间的实质量的路由选择流量。

10. 如权利要求 9 所述的方法,其中所述实质量的路由选择流量是基于所传输的预定最小量字节的,并且重要路径对应于多个可用的替换路由选择路径。

11. 如权利要求 10 所述的方法,其中所述路径验证消息是根据预定协议的探测消息,所述探测消息包括测试指示符和目的地指示符,并且所述方法还包括根据所述预定协议应用诊断逻辑。

12. 如权利要求 11 所述的方法,还包括基于驱动端到端路径检查的可抖动且可配置的定时器来交错所述诊断消息。

13. 如权利要求 12 所述的方法,其中识别所述中间节点还对应于识别与测试发起者和目的地之间的网络跳相对应的网络前缀。

14. 如权利要求 1 所述的方法,其中:

分析还包括识别指示无法沿假定的最优路径传播消息的转发平面差错,并且确定还包括改变与假定可工作路径相对应的控制平面路由选择判决。

15. 如权利要求 14 所述的方法,其中发送所述路径验证消息还包括:

发送第一路径验证消息;

识别没有接收到与所述第一路径验证消息相对应的命令响应;

等待预定的阈值,该预定阈值对应于一个收敛时间,该收敛时间适合于允许自动路由选择表更新以补偿不稳定的路由;以及

发送第二路径验证消息,其中接收到对所述第二路径验证消息的命令响应指示绕开所述不稳定路由的路由选择表改变。

16. 一种用于定位网络故障的方法,包括:

根据诊断协议发送多个路径验证消息到多个预定的网络点,其中,每个路径验证消息包括一个命令,该命令可操作来引导中间节点发送另一消息到去往目的地的路径中的后继中间节点、接收来自所述另一消息的结果并且将所述结果作为命令响应进行报告,所述结果指示所述后继中间节点的可达性;

接收与发送的路径验证消息相对应的、包括根据所述诊断协议的测试结果的命令响应;

跟踪从沿着从源到目的地的路径发送的所述多个路径验证消息中的每一个接收的命令响应;以及

基于接收到来自所述预定网络点的响应,断定所述网络中消息流量的替换路由选择路径。

17. 如权利要求 16 所述的方法,其中基于接收消息而断定还包括:

基于在所述响应中接收的测试结果,确定是否在所述网络中重路由流量;以及

基于接收和未接收到来自特定网络点的响应,定位替换路径。

18. 一种用于定位有缺陷的网络互连的方法,包括:

识别从数据通信设备到远程网络目的地的路径,该路径还包括多个网段,每个网段由一跳来界定;

识别故障点包括识别由去往所述目的地的路径限定的网段顺序;

反复发送探测到沿着所述有序路径的每个后继跳;

如果针对特定跳有探测响应返回,则断定所述路径直到与返回的探测响应相对应的跳都无阻塞;

如果对于特定探测没有接收到探测响应,则断定在与所述特定探测相对应的跳和前一跳之间存在阻塞;

基于与断定出的阻塞相对应的跳,识别替换路径;以及

基于识别出的替换路径,确定是否将消息流量引导到所述识别出的替换路径。

19. 一种用于网络故障位置识别的方法,包括:

枚举一组重要路由,所述重要路由通过关键路径传送着实质性的流量负载;

基于最近传送的流量从所述重要路由中识别活动路由;

对于每个识别出的活动路由,确定是否存在无阻塞的网络路径;

对于被确定具有阻塞的每个活动路由,应用路径验证以识别与阻塞点相对应的路径网段,所述路径验证过程还包括:

查验多个中间跳中的每一个;

识别其查验响应有缺陷的跳;

如果所述查验响应有缺陷,则在等待收敛阈值延迟之后重查验所述跳;

如果接收到对重查验的响应,则断定已经重路由以绕开核心网络故障;以及

如果未接收到重查验响应,则确定在被重查验的跳和前一跳之间的某个点存在故障。

20. 一种识别网络故障的设备,包括:

用于周期性地发送诊断消息到沿着去往目的地的路径的多个预定的路由选择点的装置;

用于如果所述诊断消息指示沿着所述路径的中间节点存在问题则发送一系列路径验证消息的装置,每个所述路径验证消息包括一个命令,该命令可操作来引导中间节点发送另一消息到所述路径中的后继中间节点、接收来自所述另一消息的结果并且将所述结果作为命令响应进行报告,所述结果指示所述后继中间节点的可达性;

用于重复发送路径验证消息到指示所述问题的节点的沿着所述路径的后继节点的装置;

用于分析接收到的来自发送到后继节点的路径验证消息的命令响应以识别问题的装置;以及

用于基于所述分析确定替换路径,以绕过被识别为所指示问题的源的中间节点的装置。

21. 如权利要求 20 所述的设备,其中用于分析的装置还包括用于分析接收到的来自所述路径验证消息的命令响应,以识别所述命令响应指示有问题的第一中间节点和对所述诊断消息的命令响应指示正常操作的前面的中间节点的装置。

22. 如权利要求 20 所述的设备,其中用于发送路径验证消息的装置还包括:

用于根据诊断协议发送多个路径验证消息到多个预定的网络点的装置;

用于接收与发送的路径验证消息相对应的命令响应的装置,所述命令响应包括根据所述诊断协议的测试结果;以及

用于跟踪从沿着从源到目的地的路径发送的所述多个路径验证消息中的每一个接收的命令响应的装置。

23. 如权利要求 22 所述的设备,其中所述中间节点表示网段,所述网段还包括客户设备和核心网络之间的本地网段,代表多个提供商设备的核心网段,以及所述核心网络和所述目的地之间的远程网段,用于跟踪的装置还包括用于识别响应所源自的网段的装置。

24. 如权利要求 22 所述的设备,还包括:

用于识别与具有实质性的逻辑路由选择价值的网络路由选择点相对应的重要前缀的装置;以及

用于发送所述诊断消息到所述重要前缀的装置。

25. 如权利要求 24 所述的设备,还包括用于确定活动前缀的装置,所述活动前缀指示在前一阈值定时窗口期间的实质量的路由选择流量。

26. 如权利要求 25 所述的设备,其中所述路径验证消息是根据预定协议的探测消息,所述探测消息包括测试指示符和目的地指示符,并且所述设备还包括用于根据所述预定协议应用诊断逻辑的装置。

27. 如权利要求 20 所述的设备,其中:

用于分析的装置还包括用于识别指示无法沿假定的最优路径传播消息的转发平面差错的装置,并且

用于确定的装置还包括用于改变与假定可工作路径相对应的控制平面路由选择判决的装置。

28. 如权利要求 27 所述的设备,其中用于发送所述路径验证消息的装置还包括:

用于发送第一路径验证消息的装置;

用于识别没有接收到与所述第一路径验证消息相对应的命令响应的装置;

用于等待预定的阈值的装置,该预定阈值对应于一个收敛时间,该收敛时间适合于允许自动路由选择表更新以补偿不稳定的路由;以及

用于发送第二路径验证消息的装置,其中接收到对所述第二路径验证消息的命令响应指示绕开所述不稳定路由的路由选择表改变。

## 用于检测网络故障的系统和方法

### 技术领域

[0001] 本发明涉及用于检测网络故障的系统和方法。

### 背景技术

[0002] 计算机网络一般包括不同计算机之间的物理互连,以允许对程序和数据进行便利的交换。诸如交换机和路由器之类的多个连通性设备互连每个连接到网络的用户计算机。连通性设备维护关于计算机的路由选择信息,并且执行关于经由连通性设备在计算机之间传递的消息流量的路由选择判决。每个连通性设备或路由器对应于指示其可直接或间接访问的其他计算机的网络路由选择前缀(前缀)。因此,被从一个计算机路由到另一个计算机的数据沿着由两个计算机之间的路由器限定的经过网络的路径前行。以这种方式,网络中路由器的聚集限定了连接到网络的各种计算机之间的互连的图。

[0003] 因此,在图形表示中,这种网络可被设想为计算机之间的节点的图。该图限定了连接到网络的每个计算机之间的一个或多个路径。因此,路由器限定了网络中的节点,以及在网络上的一系列所谓的“跳”中在节点之间行进的数据。由于每个路由器一般连接到多个其他路由器,因此在给定计算机之间可能存在多个可能的路径。一般来说,路由选择信息被用于每个路由器中的路由选择表中,该路由选择表被用于确定去往目的地计算机或网络的路径。路由器利用路由选择表作出路由选择判决,以识别向其发送数据的下一“跳”或下一路由器,以便它最终到达目的地计算机。但是,可能出现网络问题,使得路由器和路由器之间的传输路径不可工作。这种故障在实际上消除了由网络限定的图中的节点或跳(这种故障本应当由控制平面检测到),因此干扰了本该通过受影响路径路由的数据流量。

[0004] 发明内容

[0005] 在典型的计算机网络中,可能发生阻止或延迟从一个节点到另一节点的传输的故障。这种故障可能在路由器本身处,例如不正常工作的端口或转发引擎,或者可以发生在去往下一跳的传输线路中,例如物理中断或线路缺口。传输线路故障一般可通过内部网关(路由选择)协议(IGP)来识别和绕过。但是,IGP却无法识别转发问题。因此,传统的方法通过手工地“查验(ping)”远程节点以识别可能发生问题的区域,来逼近这种事故。如本领域所知的,这种“查验”或者连通性检查包括发送请求确认的简单消息到远程节点。如果接收到确认(ack),则认为远程节点和中间的路径可工作。但是,这种传统的方法却有若干个缺陷。可能存在去往被“查验”的节点的多个路径,并且中间节点可能绕开故障路由查验和相应的ack。此外,否定结果仅仅是未接收到ack;而没有提供关于故障发生在何处以及为何发生的更多信息,或者关于故障是否将自己更正自身的更多信息,例如在传输线路故障的情形下就是如此。

[0006] 本发明的配置部分地基于以下观察到的事实:传统的网络诊断和故障排查机制一般识别不可达的目的地,但不识别问题的位置,例如断开的连接或发生故障的路由器。传统路由器的特定缺点在支持关于虚拟专用网(VPN)的因特网RFC 2547bis的设备中尤其明显。通常,这种影响沿着下一跳的数据传输的所谓“转发/数据平面”问题在“控制平面”或

者说判定路由选择路径的功能处是不明显的。因此,控制平面判决可能基于转发平面关于网络的不准确视图继续通过已废止的路径进行路由,而此时路由器或者使不可转发的分组排队,或者甚至丢弃不可转发的分组。后者有时被称为分组的“黑洞效应 (black holing)”,导致为了避免丢失数据而依赖于应用冗余性和重传机制,这两者都负面地影响着吞吐量。

[0007] 换言之,转发平面级别的问题或故障可能要到认识到丢失分组的积累拥堵或丢失分组变得显著时才会变得明显。直到有了这种认识并且操作者随后进行手工干预,控制平面判决才不会继续沿着不可工作的路径进行路由。因此,开发出这样一种路径验证机制将会是有益的,该机制可以探测特定路由选择路径,并且不仅识别端到端故障,比如常见的“查验”消息,还通过发送命令并接收指示对增长点 (incremental point) 来说可见的其他节点的响应来识别增长点或者说节点处的故障。以这种方式,一系列路径验证消息可识别发生这种转发平面问题的增长点,例如节点或路径,并且可能推翻数据平面路由选择判决,以便沿着绕开识别出的问题的替换路由选择路径前行。

[0008] 因此,本发明的配置通过提供一种枚举发送到沿可疑路径的一组节点或路由器的一系列消息的路径验证协议 (PVP),基本上克服了传统的网络故障检测和故障排查的缺点。这些消息包括这样一个命令,该命令请求询问另一远程节点,以便获得关于接收 PVP 消息的节点和另一远程节点之间的路径的信息。接收 PVP 消息的节点 (第一节点) 利用指示尝试达到另一远程节点 (第二节点) 的结果的命令响应作为答复。在特定的传统设备中,例如根据 RFC 2547bis 的设备中,某些客户装备 (CE) 边缘路由器不具有核心 (即中间公共网络) 内的可见性,因此依赖于另一节点例如提供商装备 (PE) 节点来执行这种验证。该系列消息总体上覆盖了沿着从发源者到接收者的路径的一组重要的预定路由选择点。路径验证处理器分析对该系列 PVP 消息的聚集命令响应,以尝试不仅识别整个路径是否可工作,并且还尝试识别问题的位置和性质 (端口、卡、传输线路等等)。以这种方式,下文进一步讨论的路径验证机制限定了用于从网络设备 (如路由器) 中的路径验证处理器枚举一组消息并且分析来自对路径验证消息作出响应的一组节点的命令响应以便定位故障的路径验证协议 (PVP)。

[0009] 在典型网络中,如上所述,数据采取消息的形式,消息在从源到目的地的一系列跳中在网络设备 (例如路由器) 之间传播。在适合用于这里讨论的方法和设备的示例性网络中,虚拟专用网 (VPN) 利用 VPN 服务互连两个或更多个本地网络 (例如 LAN),所述 VPN 服务可操作来向子网之间的网络流量提供安全性,以便每个子 LAN 的节点可与作为同一 VPN 的成员的其他子 LAN 的节点通信。在典型的 VPN 布置中,特定子网可以是具有分别有一个实质性的子网的多个不同站内的大型商业企业 (例如银行、零售商店或大公司) 的个体站点。这种环境中的传统 VPN 十分适合于向子网之间的通信提供透明保护。

[0010] 在典型的 VPN 中,每个子网具有一个或多个网关节点,或者说客户装备 (CE) 路由器,离开以去往其他子网或从其他子网进入的流量经过这些网关节点。网关节点连接到核心网络边缘处的可操作来提供去往 VPN 中的其他子网的传输的网络提供商路由器,或者说提供商装备 (PE)。CE 和 PE 路由器有时被称为“边缘”路由器,因为它们接近客户网络或提供商网络的边缘。核心网络可以是诸如因特网这样的公共接入网络、物理上分离的内联网或者其他互连,它提供去往远程 PE 路由器的传输。远程 PE 路由器耦合到代表远程子网或 LAN (其是 VPN 的一部分) 的入口的远程 CE 路由器。远程 CE 路由器在远程 VPN (LAN) 子网

内将消息流量继续转发到的目的地。

[0011] 在这种 VPN 布置中, VPN 源或者说发源者和 VPN 目的地或者说接收者之间特定的端到端路径代表多个网段。每个网段是沿着路径的某些节点之间的一个或多个跳的集合。多个网段代表一个路径, 并且包括从本地 CE 路由器到核心网络的本地 CE 网段, 核心网络的 PE 路由器之间的核心网段, 以及从远程 PE 路由器到远程 CE 路由器的远程 CE 网段, 下文将对此进行进一步讨论。也可以定义其他网段。

[0012] 具体而言, 在操作的一个级别, 这里讨论的配置执行一种用于通过以下步骤来定位网络故障的方法: 根据诊断协议发送多个路径验证消息到多个预定的网络点 (即节点), 并且接收来自这些节点的与发送的路径验证消息相对应的命令响应, 其中响应包括根据诊断协议的测试结果。该方法跟踪响应于沿着从源到目的地的特定路径发送的多个路径验证消息中的每一个而接收的命令响应, 并且基于接收到来自预定网络点的响应, 断定或计算一个路由选择判决, 该路由选择判决包括网络中消息流量的可能的替换路由选择路径。因此, 命令响应允许了发起路径验证消息的路由器或交换机基于在响应中接收的测试结果, 确定是否在网络中重路由流量, 以及如果是, 则基于接收和未接收到来自特定网络点的响应, 定位替换路径。

[0013] 更详细地说, 本发明的配置通过以下步骤执行网络故障识别: 周期性地发送诊断消息到沿着去往目的地的路径的多个预定的路由选择点, 并且如果诊断消息指示沿着路径的中间节点存在问题, 则发送一系列路径验证消息, 其中每个路径验证消息包括一个命令, 该命令可操作来引导中间节点 (第一节点) 发送另一消息到路径中的后继中间节点 (第二节点), 接收来自另一消息的结果, 并且将结果作为命令响应进行报告, 从而使得结果指示从第一节点到后继 (第二) 中间节点的可达性。该方法根据预定的跳 (即“重要”路由选择点) 以系统的方式重复发送路径验证消息到指示或报告问题的节点的沿着所述路径的后继节点。该方法分析接收到的来自后继路径验证消息的命令响应以识别问题或故障, 并且因此基于分析确定替换路径, 以绕过被识别为所指示问题的源的中间节点。

[0014] 发送路径验证消息还包括:

[0015] 1) 根据诊断协议发送一组路径验证消息到多个预定的网络点,

[0016] 2) 接收与发送的路径验证消息相对应的命令响应, 其中命令响应包括根据诊断协议的测试结果; 以及

[0017] 3) 跟踪从沿着从源到目的地的路径发送的多个路径验证消息中的每一个接收的命令响应, 其中跟踪还包括识别响应所源自的网段。

[0018] 分析接收到的命令响应包括分析路径验证消息, 以识别命令响应指示有问题的第一中间节点和对诊断消息的命令响应指示正常操作的前面的中间节点。如上所述, 中间节点表示网络中的网段, 其中网段还包括客户设备和核心网络之间的本地网段, 代表多个提供商设备的核心网段, 以及核心网络和目的地之间的远程网段。分析还包括识别接收和未接收, 其中接收包括从发送消息的预定网络点起的可访问路径的指示, 未接收根据诊断逻辑指示调解故障。分析可以识别指示无法沿假定的所选路径传播消息的转发平面差错, 从而使得确定替换路径包括改变与假定可工作路径相对应的控制平面路由选择判决。

[0019] 在特定布置中, 该方法基于网络故障的位置和性质, 识别在该处更改流量以重路由流量使之绕开故障的网络点。这种点是中间网络节点, 并且识别中间节点还对应于识别

与测试发起者和目的地之间的网络跳相对应的网络前缀。

[0020] 这里公开的配置通过以下步骤解决核心网络中的故障：发送第一路径验证消息，识别没有接收到来自核心网络中间路由器的与第一路径验证消息相对应的命令响应，以及等待预定的阈值，其中该预定阈值对应于一个收敛时间，该收敛时间适合于允许自动路由选择表更新以补偿不稳定的路由。该方法随后发送第二路径验证消息，其中接收到对第二路径验证消息的命令响应指示绕开不稳定路由的路由选择表改变，从而将所谓的核心网络的收敛属性用于重路由以利用冗余路径绕开故障。

[0021] 发送诊断消息包括识别与具有实质性的逻辑路由选择价值的网络路由选择点对应的重要前缀，以及发送针对重要前缀的诊断消息。在重要前缀中，该方法还可选地确定活动前缀，其中活动前缀指示在上一阈值定时窗口期间的实质量的路由选择流量。这种实质量的路由选择流量负载是基于所传输的预定最小量字节的，并且重要路径对应于若干个可用的替换路由选择路径，例如可能的瓶颈和周期性的突发门户。此外，该方法基于驱动覆盖端到端路径检查的一组消息的可抖动且可配置的定时器来交错诊断消息，从而避免 PE 路由器本身接收到诊断消息突发。

[0022] 在示例性的布置中，路径验证消息是根据预定协议的探测消息。探测消息包括出于确认特定路径的可用性或交换的目的而发送的消息和分组，而不是对数据有效载荷的传输。这里所使用的探测消息包括用于指定测试结果的测试指示符和用于指示所关注的节点的目的地址指示符，并且断定还包括根据预定协议应用诊断逻辑。该协议的诊断逻辑实现了指示或确定特定类型的故障（例如发生故障的转发引擎和灾难性的节点故障）的规则或状况。

[0023] 本发明的替换配置包括多重编程或多重处理的计算机化设备，例如工作站、手持式计算机或膝上型计算机，或者专用计算设备等等，其被配置以软件和/或电路（例如上面概括的处理器）以便处理这里作为本发明实施例公开的任何一种或全部方法操作。本发明的其他实施例包括软件程序，例如 Java 虚拟机和/或操作系统，其可单独工作、或者彼此结合工作、或者与多重处理计算机化设备相联合地工作，以执行上面概括的或在下面详细公开的方法实施例步骤和操作。一个这种实施例包括具有计算机可读介质的计算机程序产品，该计算机可读介质包括编码在其上的计算机程序逻辑，该计算机程序逻辑当在具有存储器和处理器的耦合的多重处理计算机化设备中被执行时，将处理器编程为执行这里作为本发明实施例公开的操作，以进行数据访问请求。本发明的这种布置一般是以布置或编码在计算机可读介质（所述计算机可读介质例如是光介质（例如 CD-ROM）、软盘或硬盘，或其他介质）上的软件、代码和/或其他数据（例如数据结构）、例如一个或多个 ROM 或 RAM 或 PROM 芯片或现场可编程门阵列（FPGA）中的固件或微代码，或者专用集成电路（ASIC）的形式提供的。软件或固件或其他这种配置可被安装在计算机化设备上（例如在用于执行环境的操作系统的安装期间），以使计算机化设备执行这里作为本发明实施例说明的技术。

## 附图说明

[0024] 本发明的前述和其他目的、特征和优点将从以下对附图中示出的本发明的优选实施例的更具体描述中显现出来，在附图中类似的标号在不同图中指代相同的部件。附图不一定是按比例绘制的，相反重点在于图示本发明的原理。

[0025] 图 1 是可用于本发明的包括经由多个提供商装备设备（路由器）限定路径的网络节点的网络通信环境的上下文图；

[0026] 图 2 是在图 1 的网络中使用路径验证机制的流程图；

[0027] 图 3 是图 1 的示例性网络中的路径验证设备的框图；以及

[0028] 图 4-7 是在图 4 的网络中利用图 3 的路径验证设备的路径验证机制的操作的流程图。

## 具体实施方式

[0029] 本发明的配置部分基于以下事实：即传统的网络诊断和故障排查机制一般识别不可达目的地，但不识别问题的位置，例如断开的连接或发生故障的路由器。通常，这种影响去往后继的下一跳的数据传输的所谓“转发平面”问题在“控制平面”级别或者说判定路由选择路径的功能（即路由选择逻辑）处是不明显的。因此，控制平面判决可能继续通过转发平面处已废止的路径进行路由，而此时路由器或者使不可转发的分组排队，或者甚至丢弃不可转发的分组。后者有时被称为分组的“黑洞效应”，导致为了避免丢失数据而依赖于应用冗余性和重传机制，这两者都负面地影响着吞吐量。

[0030] 换言之，转发平面级别的问题或故障可能要到认识到丢失分组的积累拥堵或丢失分组变得显著时才会变得明显，而在控制平面级别则永远不会明显。直到有了这种认识并且操作者随后进行手工干预，控制平面判决才不会继续沿着不可工作的路径进行路由。下面将进一步讨论的是一种路径验证机制，该机制可操作来探测特定路由选择路径，并且不仅识别端到端故障，比如常见的“查验”消息，还通过发送命令并接收指示对增长点来说可见的其他节点的响应来识别增长点或者说节点处的故障。以这种方式，一系列路径验证消息可识别发生这种转发平面问题的位置，并且推翻来自路由选择逻辑的数据平面路由选择判决，以便沿着绕开识别出的问题的替换路由选择路径前行。

[0031] 这里更详细描述的路径验证机制使用了一种路径验证协议，该协议可操作来根据协议发送路径验证消息，以便尝试诊断和识别网络故障的位置。因此，路径验证设备，例如具有这里公开的路径验证处理器的数据通信设备（即路由器）可操作来利用路径验证协议执行路径验证。

[0032] 因此，这里公开的系统包括在路由器中运行或执行的路径验证处理器，其具有用于执行这里详细公开的用于定位有缺陷网络互连的方法，包括识别从数据通信设备到远程网络目的地的路径，其中该路径还包括多个网段，其中每个网段由若干个跳界定。路径验证处理器通过识别去往目的地的路径所限定的网段顺序，并且反复地发送探测到沿着有序路径的相继的每一跳，来识别故障点。如果针对特定的跳的探测响应返回，则路径验证处理器断定该路径直到与返回的探测相对应的跳都无阻塞，而如果对于特定探测未接收到探测响应，则断定在与该特定探测相对应的跳和前一跳之间存在阻塞。路径验证处理器随后基于与断定的阻塞相对应的跳和 / 或先前的跳识别替换路径，并且基于识别出的替换路径确定是否将消息流量引导到识别出的替换路径。

[0033] 因此，本发明的配置通过提供一种枚举发送到沿可疑路径的一组节点或路由器的一系列消息的路径验证协议（PVP），基本上克服了传统的网络故障检测和故障排查（例如查验）的缺点。这些消息包括这样一个命令，该命令请求询问另一远程节点，以便获得关于

接收 PVP 消息的节点和 另一远程节点之间的路径的信息。接收 PVP 消息的节点利用指示尝试达到另一远程节点的结果的命令响应作为答复。该系列消息总体上覆盖了沿着从发源者到接收者的路径的一组重要路由选择点。对该系列 PVP 消息的聚集命令响应被分析,以不仅识别整个路径是否可工作,并且还尝试定位故障(端口、卡、交换架构等等)。以这种方式,路径验证机制限定了用于从网络设备(如路由器)中的路径验证处理器枚举一组消息并且分析来自对路径验证消息作出响应的一组节点的命令响应以便定位故障的路径验证协议(PVP)。

[0034] 图 1 是可用于本发明的包括限定路径的网络节点的网络通信环境的上下文图。参考图 1,网络通信环境 100 包括互连多个本地用户 114-1...114-3(总称为 114)的本地 VPN LAN 子集 110。本地 LAN 110 连接到网关客户装备 CE 路由器 120,该网关客户装备 CE 路由器 120 耦合到一台或多台提供商装备设备 130-1 和 130-2(总称为 130)。如下文将更详细讨论的,知晓多个 PE 路由器 130-1 和 130-2 的 CE 路由器 120 可基于这里讨论的考虑因素执行关于是否经由路由器 130-1 或 130-2 来路由流量的路由选择判决,所述考虑因素一般是位于核心网络 140 边缘的另一路由器。CE 路由器 120 或初始路径验证设备包括可操作以用于典型的控制平面路由选择判决的路由选择逻辑 122、可操作来定位故障并补充路由选择判决的路径验证处理器 124 以及用于转发和接收网络流量的网络接口 126。交换架构 128 可对路由选择逻辑 122 作出响应,用于经由设备上的物理端口(未具体示出)实现交换判决。核心网络 140 包括多个核心节点 142-1...142-2(总称为 142),例如各种路由器、集线器、交换机和其他连通性设备,它们互连由提供商服务的其他用户。远程提供商装备设备 132(即远程 PE 路由器)耦合到为比如 VPN LAN 112 这样的远程 VPN 子网服务的远程客户装备路由器 122。远程 VPN LAN 112 就像其对应子网 110 那样为多个远程用户 116-1...116-3(总称为 116)服务。

[0035] 这里讨论的配置中实现的原理可由图 1 来概括,并且在下面参考图 3 和图 4-7 的流程图更详细讨论。本地 CE 路由器 120 将从本地 LAN 110 上的用户 114 发送来的分组路由到提供商装备路由器 130,该提供商装备路由器 130 表示进入核心网络 140 的入口。PE 路由器 130-1 和 130-2 可通过核心网络 140 分别经由特定路径 146-1 或 146-2 将分组转发向其预期目的地。为了说明方便,假设 PE1 例如通过调用 PE1 130-1 作为进入核心网络 140 的入口来将分组 114 转发到节点 142-1。

[0036] 如果例如在节点 142-1 处发生问题,则 CE 路由器 120 上的路径验证处理器 124 调用 PE 路由器 130 来经由一组周期性诊断消息 150 识别问题,并且 PE 路由器 130 经由一组路径验证消息 152 定位问题,两者都将在下文进一步讨论。因此,CE 路由器 120 上的路径验证处理器 124 引导路由选择逻辑 122 经由 PE2 路由器 130-2 路由分组 144。如上所述,控制平面和转发平面操作之间的差异由路径验证处理器 124 来有效地衔接。传统的路由选择逻辑为特定分组识别优选路由。在转发平面级别的问题在路由选择逻辑计算优选路由的控制平面可能是不明显的。因此,在上述示例中,PE 路由器 130-1 处的传统的路由选择逻辑继续使用路由 146-1,而不知晓节点 142-1 的问题。CE 路由器 120 上的使用这里讨论的路径验证协议的路径验证处理器 124 经由 PE 路由器 130-2 识别去往核心网络 140 的使用路径 146-2 上的节点 142-2 而不是已废止节点 142-1 的替换路径,并且推翻路由选择逻辑否则会使用的优选路由判决。注意实现路径验证处理器 124 执行的操作的软件、硬件和 / 或

固件可分布在 PE 和 CE 设备 130 和 132 上,并且为了简单在 CE 路由器 120 中被放大示出。PE 和 CE 设备(路由器)中的每一个可被实现以一个路径验证处理器,或者可对路径验证处理器 124 作出响应的其他机制以及因而实现的方法。由于经由 130-2 的替换路径使路径 142-2 至 146-2 从 PE2 延伸通过核心网络 140。因此,CE1 可执行路由选择判决以将其流量从 PE1 切换到 PE2,在所示出的示例性实施例中,这有效地绕过了核心中的节点 142-1,而改为选择节点 142-2。应当注意,这种路由选择判决应用来自网络的路径信息,例如其中路径 146-1 和 146-2 是分离的,只有 146-1 依赖于节点 142-1。

[0037] 图 2 是在图 1 的网络中使用路径验证机制的示例性客户装备设备(例如路由器 120)的流程图。参考图 1 和图 2,路由器 120 基于最近传送的流量从重要路由中识别活动路由,如步骤 200 所示。活动路由是当前传送实质量的流量的路径或交换选项,这可以根据观察交换架构的活动、嗅探或其他扫描机制来确定。路径验证处理器 124 为每个识别出的活动路由确定是否存在无阻塞的网络路径,如步骤 201 所示。路径验证处理器 124 可执行所谓的“查验”或其他机制,以确定每个活动路由的可用性。应当注意,活动路由是由负责沿活动路径的消息吞吐量的特定设备或路由器来表示的,并且在确定是否存在问题时是决定性的。这些路由器一般由指示它们所服务的 IP 地址的前缀来指定,这里本领域的技术人员已知的。

[0038] 因此,路径验证处理器 124 首先为每个活动路由发送一个诊断消息 150,例如“查验”或其他查询消息,以确定是否存在可能的阻塞,如步骤 202 所示。路径验证处理器 124 执行检查,以确定是否接收到指示非响应节点的否定回复,如步骤 203 所示。如果未从活动路由(即与活动路由相对应的路由器)接收到否定回复,则控制传递到步骤 204,以期待下一诊断间隔。

[0039] 在如步骤 202-204 中所公开的向每个活动路由发送周期性诊断消息之后,对于路径验证处理器 124 未接收到对其的响应的每个前缀,即活动路由,路径验证处理器 124 发送多个路径验证消息 152 到故障路径中当前使用的下一跳设备,如步骤 205 所示。路径验证消息 152 是以预定的顺序或模式发送到已被发现了问题的路径中的前缀(路由器)的。作为对路径验证消息 152 的响应,路径验证处理器 124 接收到与发送的路径验证消息相对应的命令响应 154,其中命令响应 154 包括根据诊断协议的测试结果,如步骤 206 所示。在步骤 207,执行检查以确定是否有更多命令响应 154 要取得,并且控制相应地回到步骤 206。注意,未接收到命令响应 154 也被视为否定响应,下文将更详细讨论这一点。路径验证处理器 124 跟踪从发送到本地 PE 路由器的多个路径验证消息 152 中每一个接收到的一组命令响应 154,如步骤 208 所示,并且基于从本地 PE 路由器接收到响应 154,断定网络 100 中消息流量 144 的替换路由选择路径,如步骤 209 所示。因此,从 CE 路由器 120 发送的该组路径验证消息 152 引发一组命令响应 154,每个命令响应指示路径验证信息。典型的路径验证信息例如是关于路由器是否能够与端到端网络的特定网段通信的指示。整个这组命令响应 154 指示哪些网段正确工作,因此使故障的发生本地化,现在将更详细讨论这一点。

[0040] 图 3 是图 1 的示例性网络的 CE 路由器 120 中的路径验证处理器 124 的框图。参考图 3,网络 100 包括客户装备 120-11..120-13(总称为 120),提供商装备 130-11、130-12 和 132-11,以及中间节点 142-11..142-15(总称为 142)。通过网络 100 的路径可被细分为网段 160,这些网段由客户装备 120 和提供商装备 130、132 定界,并且由点线 168 示出。本

地 VPN 网段 162 包括从本地 VPN 110 到提供商装备 130-11 和 130-12 的路径。

[0041] 核心网段 164 包括去往远程提供商装备 132-11 设备的核心网络 140, 远程 VPN 网段 166 覆盖从远程 PE 路由器 132-11 到远程 VPN 112 的路径。另外的多个主机 S1、S2 和 S3 位于远程 VPN LAN 子网 112 内, 例如本地 LAN 服务器节点, 下文将对其进行更详细讨论。

[0042] 在特定配置中, 路径验证处理器 124 利用 PE 节点 130 使用路径验证协议 (PVP), 以通知 CE 节点 120 路径可用性, 以识别发生故障的网段 160。由于对核心网段 164 的路由选择控制可能是有限的, 因此 CE 路由器 120 进行的路由选择判决的有效性可能是有限的。但是, 在本地网段 162 中, 可能有多个 PE 路由器 130-11、130-12 用于接入到核心网络 164 中。此外, 这些 PE 路由器 130-11、130-12 可连接到核心网络 140 中的不同节点 142, 例如分别连接到 142-11 和 142-14。因此, 使用不同的提供商装备路由器 130 的路由选择判决可有效地绕过了核心网络 140 中的故障。类似地, 多个 CE 路由器 120 可为特定的子网 VPN 服务。在所示的示例中, 远程 VPN LAN 112 耦合到 CE 路由器 120-12 和 120-13 (CE2 和 CE3)。因此, 如果 132-11 (PE3) 上的路径验证处理器 124 识别出 CE2 或 CE3 的问题, 则它可使用其他 CE 路由器来从提供商装备 132 接入远程子网 112。

[0043] 又例如, 继续参考图 3, 从 VPN 子网 110 到 VPN 子网 112 的优选路由包括节点 PE1、142-11、142-12、142-13, 在 PE3 处离开提供商网络并在 CE2 处进入远程 VPN 子网 112。进一步假设在节点 142-11 发生转发平面路由选择差错。因此, 120-11 (CE1) 上的路径验证处理器 124 经由周期性诊断消息 (下文进一步讨论) 识别存在问题, 并且调用路径验证协议如下。根据 CE1 的请求, PE1 上的路径验证处理器 124 发送路径验证 (PVP) 消息 152 到路由器 PE3, 该消息实际上询问“你能看见子网 112 吗?” PE3 可能接收到该 PVP 消息, 也可能未接收到。如果它接收到, 则 PE3 发送另一个 PVP 消息以测试去往子网 112 的远程网段 166, 并且确认连续性。因此, PE3 将命令响应 154 发送回 PE1, 指示从 PE3 到 CE2 的网段 166 正常工作。然而, 如果在 PE3 和 CE2 之间检测到问题, 则 PE3 上的路径验证处理器 124 可使用 CE3, 作为进入 VPN 子网 112 的重路由判决。如果接收到指示网段 164 和 166 完好 (例如在这种情况下是节点 142-11 没有问题) 的肯定响应, 则 CE1 上的路径验证处理器 124 可推断问题出在节点 130-11。

[0044] 如果 PE1 上的路径验证处理器 124 在设置的时间内未接收到对它发送到 PE3 的 PVP 消息的响应, 则它将假设在其自身和 PE3 之间存在问题, 因此通过核心网段 164 中的节点 142 前进, 最终尝试询问节点 142-11。PE1 上的路径验证处理器 124 发送 PVP 消息 152 到节点 142-11。由于 PE1 可工作, 并且它从节点 142-11 接收到对其 PVP 消息的肯定响应, 因此它可推断问题出在节点 142-11 和 142-12 之间, 并且 PE1 发送命令响应 154 到 CE1 120-11, 指示核心数据平面故障为故障源, 而不是正常的收敛事件。因此, 路由器 CE1 处的路径验证处理器 124 分析返回的命令响应 154, 并且确定不应当使用 PE1 路由器 130-11。此外, 路径验证处理器 124 识别出路由器 PE2 130-12 为进入核心网络 140 的替换入口, 它也提供了去往 PE3 的路径。因此, CE1 上的路径验证处理器 124 实现路由选择判决, 以推翻路由选择逻辑 122, 以经由提供商装备路由器 130-12 (PE2) 发送流量到核心网络 140。

[0045] 图 4-7 是在图 3 的网络中利用图 1 的路径验证设备 (即路由器) 120 的路径验证机制的操作的流程图。参考图 1 和图 3-7, 这里公开的使用路径验证处理器 124 识别网络故障的方法包括周期性地发送诊断消息 150 到多个预定的路由选择点, 例如目的地 116, 如步

骤 300 所示。路径验证处理器 124 为了识别中间节点,识别与测试发起者和目的地之间的网络跳相对应的网络前缀,如步骤 301 所示。该下一跳一般是本地附接的 PE 路由器。如上所述,典型的 TCP/IP(传输控制协议/因特网协议)路由选择配置为个体设备或路由器分配指示它可路由到或“看见”的 IP 地址的网络前缀。因此,CE1 上的路径验证处理器 124 只能看见其本地附接的 PE 路由器,因此必须依赖于来自这些 PE 的 PVP 处理结果。PE 路由器能够经由一组前缀识别活动路径 146,该组前缀限定了它和去往目的地 116 的出口点(即 132-11(PE3))之间的路径 146 中的路由器。

[0046] 路径验证处理器 124 基于驱动端到端路径检查的可抖动且可配置的定时器来交错发送诊断消息 150 到每个前缀,如步骤 302 所示。路径验证处理器 124 以恒定的间隔发送或查询活动路由,如上所述。使消息 150 交错避免了在每个间隔诊断消息 150 的骤然突发。这种接收诊断消息 150 的前缀被标为重要前缀(其可利用访问列表来识别),并且对应于具有实质性的逻辑路由选择价值的网络路由选择点,如步骤 303 所示。此外,路径验证处理器 124 从重要前缀中确定活动前缀,其中活动前缀指示前一阈值定时窗口期间的实质量的路由选择流量,如步骤 304 所公开。此外,某些前缀可能经历休眠时段,或者可能主要在特定时间被利用,例如在每日或每周备份或下载时。因此,对实质量路由选择流量负载的确定是基于所传输的预定最小量字节的,并且重要路径对应于若干个可用的替换路由选择路径,如步骤 305 所示。例如,金融机构往往可能在业务周结束时(即星期五下午)进行许多事务。因此,某些前缀可被指明为仅在星期五下午活动,因为在这种时间,路由选择问题可能尤其会侵害业务操作。在确定活动前缀之后,路径验证处理器 124 发送诊断消息 150 到重要的活动前缀,如步骤 306 所示。

[0047] 如步骤 307 所示,路径验证处理器 124 执行检查以确定是否诊断消息 150 中的任何一个指示出问题,这一般是由未接收到确认引起的。如果没有诊断消息 150 指示了问题,则控制回到步骤 300 以等待下一间隔。但是,如果一个或多个目的地未确认诊断消息 150,则路径验证处理器 124 开始发送一系列路径验证消息,其中每个路径验证消息包括一个命令,该命令可操作来引导中间 PE 节点执行以下操作:a) 发送另一个消息到路径 中后继的中间节点,b) 接收来自该另一个消息的结果,以及 c) 报告该结果以作为命令响应,其中该结果指示后继的中间节点的可达性,如步骤 308 所示。

[0048] 在步骤 300-306 中公开的去往每个活动路由的周期性诊断消息 150(即去往活动前缀的消息)之后,路径验证处理器 124 应用路径验证以识别和定位未回复的前缀的问题。如步骤 309 所示,对于每个有问题的目的地,PE1 上的路径验证处理器 124 根据诊断协议发送多个路径验证消息 152 到多个预定的网络点(即活动前缀),如步骤 310 所示。在这里的示例性配置中,路径验证消息 152 是根据预定协议的探测消息,其中探测消息包括测试指示符和目的地指示符,以便探测消息 152 引发来自每个路径验证消息 152 的命令响应 154,从而允许 PE1 上的路径验证处理器 124 根据预定的协议应用诊断逻辑,如步骤 311 所示。路径验证消息 152 中的测试指示符和目的地指示符包括关于其他远程节点 142 及其可达性的信息。接收方节点 142 执行所请求的对目的地指示符中的节点的检查,并将测试结果写入测试指示符中。例如,节点 142-1 接收询问提供商边缘 PE 节点 132(图 1)的可达性的消息。此外,PE 节点 130 接收到类似的消息。如果 PE 节点 130 能够看见 PE 节点 132,但是节点 142-1 不能看见节点 132,则似乎在节点 142-1 处存在问题。可以假设,PE 节点 130

经由节点 142-2 访问 PE 节点 132, 去往 142-2 的后续路径验证消息可以确认。在这两种情况下, PE1 上的路径验证处理器 124 都请求和获得测试指示符字段中携带的关于通过不同的远程节点访问另一个不同的远程节点的信息, 而不仅仅是识别路径验证处理器 124 本身可达的节点。典型的传统查验和相关操作只识别针对发送方节点 (进行查验操作的节点) 的成功, 而不代表其他节点。

[0049] 继续上述示例, 路径验证处理器 124 接收与发送的路径验证消息 152 相对应的命令响应 154, 其中命令响应 154 包括根据诊断协议的关于目的地指示符中的节点的测试结果, 如步骤 312 所示。路径验证处理器 124 聚集命令响应 154, 以跟踪从沿着从源到目的地的特定可疑路径发送的多个路径验证消息 152 中每一个接收的命令响应, 如步骤 313 所示。在步骤 314, 执行检查以确定被跟踪的命令响应是否指示问题。如果不, 则路径验证处理器 124 继续重复发送路径验证消息到指示所述问题的节点的沿着所述路径的后继节点 (如步骤 315 所示), 因此遍历沿可疑路径的每个前缀以识别原因。

[0050] 如果特定命令响应 154 指示出问题, 则在步骤 314, PE1 上的路径验证处理器 124 尝试检测核心网段 164 中的核心网络问题。通常, 网络在路由器之间提供多个物理路径, 并且路由器自适应地改变路由以避开问题区域。这种作法被称为收敛, 并且可以在路径验证消息经由命令响应 154 指示故障之后立刻发生。

[0051] 因此, 路径验证处理器识别没有接收到与第一路径验证消息 152 相对应的命令响应或接收到与第一路径验证消息 152 相对应的否定命令响应, 如步骤 316 所示。路径验证处理器 124 随后等待预定的阈值, 其中该预定阈值对应于一个收敛时间, 该收敛时间适合于允许自动路由选择表更新以补偿不稳定的路由, 如步骤 317 所示。在收敛阈值时间之后, 路径验证处理器发送第二路径验证消息, 其中接收到对第二路径验证消息 152 的命令响应 154 指示绕开不稳定路由的路由选择表改变或其他收敛校正, 如步骤 318 所示。

[0052] 如果路径验证处理器接收到来自第二路径验证消息的肯定响应, 如步骤 319 所示, 则路径验证消息处理器断定核心网段 164 内的收敛问题, 并将此报告回 CE1, 并且控制回到步骤 300, 直到下一诊断间隔。如果收敛阈值检查没有解决故障, 则路径验证处理器分析接收到的来自后继的路径验证消息的命令响应, 以识别问题或故障, 如步骤 320 所示, 并将此报告给 CE1。

[0053] 路径验证处理器聚集并分析针对去往指示故障的前缀的路径所接收的响应 154。分析响应消息 154 还包括识别接收和未接收, 其中接收包括来自发送消息的预定网络点的可访问路径的指示, 未接收根据诊断逻辑指示调解故障, 如步骤 321 所示。在从源到目的地的特定路径中, 这种分析可包括分析接收到的来自路径验证消息的命令响应以识别命令响应指示有问题的第一中间节点以及对诊断消息的命令响应指示正常操作的前面的中间节点, 如步骤 322 所公开。换言之, 分析力求识别在该处可识别故障的第一网络跳。紧挨着的前一跳或者说路径上可达 (即响应 154 指示没有问题) 的最后一个成功前缀和第一个不成功跳, 从而标识发生故障的范围。这种分析可操作来识别指示无法沿假定的最优路径传播消息的转发平面差错, 如步骤 323 所示。如上所述, 转发平面差错, 例如关于转发引擎、端口或交换架构的故障, 可能不会立即在作出路由选择判决的控制平面 (即路由选择逻辑) 处变得明显。通过询问沿着已知有问题的路径的后继跳, 可以识别第一个发生错误的跳。

[0054] 在特定配置中, 收敛场景是基于整体流量的量来审查的。在拥塞网络中, 冒险丢弃

一些分组并等待滞后时间以便收敛阈值过去而不是通过已知的拥塞路由重新路由分组,可能是有益的。

[0055] 一旦分析指示出发生错误的位置、跳或节点,路径验证处理器 124 就基于网络故障的位置和性质识别在该处更改流量的网络点,如步骤 324 所示。例如,给定从本地 VPN LAN 110 到远程 VPN LAN 112(图 3)的路径,如果在路由器 PE1(130-11)中或在去往节点 142-11 的跳中发现问题,则替换路径是从 CE1 重新路由流量以使之在 PE2(130-12)而不是 PE1 处进入核心网络 140,以避免故障并且仍维持去往位于核心网络 140 远侧的 PE3 的路径。此外,中间节点表示网段 160,其中网段还包括客户设备和核心网络之间的本地网段 162、代表多个提供商设备的核心网段 164 和核心网络和目的地之间的远程网段 166,从而使得跟踪还包括识别响应所源自的网段,如步骤 325 所示。在特定配置中,网段可由与路径验证处理器或本地 CE 路由器 120 的距离来标识,其中网段还包括从客户边缘路由器到中间网络到远程边缘路由器的第一网段 162、提供商边缘路由器之间的第二网段 164 和从提供商边缘路由器到远程客户边缘路由器的第三网段 166,如步骤 326 所示。

[0056] 如上所述,对替换路由的确定可包括改变与假定的可工作路径相对应的控制平面路由选择判决,如步骤 327 所示。路径验证处理器 124 基于步骤 320 的分析确定替换路由,以绕过在步骤 324 被识别为所指示问题的源的中间节点,如步骤 328 所公开。基于响应消息的接收的路由选择判决的结论包括基于在响应中接收的测试结果确定是否在网络中重路由流量,如步骤 329 所公开的,并且如果是,则基于接收和未接收到来自特定网络点的响应来定位可操作来将流量传输到同一目的地或 VPN 子网的替换路径。

[0057] 参考图 3,控制平面路由选择判决可如下进行。从本地 VPN LAN 110 起的最优(最短)路径包括经由 142-11、142-12 和 142-13 从 PE1 到 PE3。参考上述示例,PE1 上的路径验证处理器 124 识别故障发生在去往远程 VPN LAN 112 的最优(最短)路径中包括的节点 142-11 中的转发引擎。传统的方法将会使控制平面继续沿最优路径进行路由,从而导致节点 142-11 处的黑洞效应和/或排队。还注意到,在一些情况下,核心网络可能是公共接入网络和/或外部提供商网络,因此不会直接对路径验证处理器作出响应(即不像 VPN 那样在用户的直接控制之下)。然而,CE1 上的路径验证处理器 124 观察到经由 PE2 通过节点 142-14 和 142-15 的替换路径在 142-13 处与最优(最短)路径汇合。CE1 上的路径验证处理器 124 推翻倾向于以 PE1 作为进入核心 140 的优选入口的路由选择逻辑 122,并且使用 PE2 作为替换路径。因此,路径验证处理器通过观察和确定路由选择逻辑 122 能够控制的新的 PE 设备而解决了核心网络中的问题(142-11),并且避免了会继续将流量引导到故障节点 142-11 的数据平面状况。类似地,如果问题被诊断为影响了 CE2,则从 PE3 进入远程 VPN LAN 112 的替换路径包括 CE3。

[0058] 更详细地说,图 3 的系统中的示例性 PVP 场景如下。继续参考图 3,如果从同一 VPN 的不同的本地附接的客户端接收到针对同一远程目的地的多个请求,则 PE 路由器应当聚集路径验证检查。PE 在发起某些路由时充当下一跳本身。因此,接收来自请求验证去往 116-1 和 116-2 的路径的 CE 的 PVP 消息的 PE 将能够看到两个前缀都具有同一 BGP 下一跳(即远程 PE3)。利用这种信息,可按如下方式为路径的核心部分 164 聚集 PVP 过程:注意,对于下面的示例,如图 3 所示,CE1 连接到 PE1 和 PE2。还要注意 CE2 附接到 PE3,并且 CE3 也附接到 PE3。关于子网前缀 116-1、116-2 和 116-3,前缀 PE1 连接到 CE1、前缀 116-1、116-2

和 116-3 连接到 CE2 和 CE3, 如下面的序列所描述 :

[0059] -CE1 希望验证去往前缀 116-1、116-2 和 11-63 的路径。这假设了先前去往这些设备的查验已经失败

[0060] -CE1 读取前缀的共性 (community)/ 标签, 并找出哪些是前缀的下一跳, 如下 :

[0061] 116-1- > PE1

[0062] 116-2- > PE1

[0063] 116-3- > PE1

[0064] -CE1 准备 3 个 PVP 消息。这些消息可以是基本的或高级的, 详见下 :

[0065] o 目的地为 PE1 针对 116-1 的 PVP

[0066] o 目的地为 PE1 针对 116-2 的 PVP

[0067] o 目的地为 PE1 针对 116-3 的 PVP

[0068] -CE1 将 PVP 消息发送到 PE1

[0069] -PE1 检查这些 PVP 消息的目的地, 并找出前缀 116 实际上连接到同一 PE (PE3)。

[0070] -如果请求是基本检查, 则 PE1 将会向前缀 116 中的每一个发送一个查验。如果成功, 则它将会以肯定响应对 CE1 作出响应。如果查验失败, 则它将会以否定响应对 CE1 作出响应。

[0071] -如果请求是高级检查, 则 PE1 准备一个 PVP 消息 :

[0072] o 针对 PE3 的 PVP (包括针对 116-1、116-2 和 116-3 的 PVP)

[0073] -PE1 将该 PVP 消息发送到 PE3

[0074] -当 PE3 接收到 PVP 消息时, 它将会 :

[0075] o 检查谁是去往 116-1、116-2 和 116-3 的下一跳, 并发现使用了相同的下一跳

[0076] o 发起到适当 CE 的 PVP 消息。

[0077] -假设是对 BPV (基本路径验证) 的请求, 则客户端 (在本示例中即本地 VPN LAN 110 或 CE1 120-11) 将会接收来自 PE 130-11 的肯定或否定响应。如果接收到肯定响应, 则它将会假设问题存在于入口 PE 130-11 (PE1) 的交换路径内, 因为该 PE 能够达到远程目的地, 但是来自 CE1 的分组却不能, 从而指示在 PE1 上存在本地交换故障, 并且它因此将促成本地重路由 (其细节是特定于实现方式的, 取决于所使用的网络管理协议 / 机制)。如果响应是否定的, 则客户端将会假设收敛事件正在进行中并因而不采取进一步动作, 或者, 基于配置和诸如沿替换路径的路径成本增长这样的标准, 判定触发本地重路由。

[0078] -假设是对 APV (高级路径验证) 的请求, 则 PE 路由器 130 将会识别问题是存在于 (1) 核心网络 146 内, (2) 远程 PE 路由器 132 内, 还是 (3) 核心网络之外。

[0079] -如果问题在核心网络 164 内, 则 PE 130 将会向客户端 120-11 作出响应, 指示核心 146 问题。依赖于基于定时器的方法的技术可被用于此目的, 从而 PE 130 可启动一个定时器, 该定时器的值将会反映最坏 IGP 收敛时间。客户端应当将该信息理解为意味着正在发生收敛事件, 因此不采取动作。

[0080] -如果问题在于远程 PE 路由器 132, 则 PE 将会向客户端作出响应以指示这一点。远程 PE 路由器 132 的故障可以是真实故障 (例如路由处理器、电源、线路卡等), 在这种情况下收敛事件正在进行, 或者该故障可以是交换故障, 在这种情况下收敛事件并未进行。在任一种情况下, 如果经由另一 PE 的另一路径可用, 客户端 CE1 就将会发起本地重路由, 而不

论该路径的成本是否大于当前最佳路径（即，应当触发层间故障通知机制）。这将会增加到经由另一 PE 或经由同一 PE 的不同可工作接口到达目的地。

[0081] - 如果问题在核心网络 164 之外，则客户端 CE1 应当不采取动作。

[0082] - 在任何故障情况下，客户端应当记录从 PE 路由器接收的验证响应。

[0083] - 一旦本地重路由被发起，客户端就启动可配置的定时器 Y，在该定时器期满时新的验证被触发。这是为了确保一旦原始故障的原因已得以矫正并且倘若路由选择协议仍选择原始（故障前）路径作为最佳路径，更优的路径会被重新建立。

[0084] 本领域的技术人员将很容易认识到，这里定义的用于识别网络故障的程序和方法可以以许多形式递送到处理设备，所述形式包括但不限于：a) 永久存储在诸如 ROM 设备这样的不可写存储介质上的信息，b) 可改变地存储在诸如软盘、磁带、CD、RAM 设备和其他磁介质和光介质之类的可写存储介质上的信息，或者 c) 通过通信介质传输到计算机的信息，例如利用基带信令或宽带信令技术传输到计算机的信息，就像诸如因特网或电话调制解调器线路之类的电子网络中那样。这些操作和方法可实现在软件可执行对象中，或者实现为嵌入在载波中的一组指令。或者，这里公开的操作和方法可整体或部分地利用硬件组件或者硬件、软件和固件组件的组合实现，所述硬件组件例如是专用集成电路（ASIC）、现场可编程门阵列（FPGA）、状态机、控制器或其他硬件组件或设备。

[0085] 虽然已经参考其实施例具体地示出和描述了用于识别网络故障的系统和方法，但是本领域的技术人员将会理解，在不脱离由所附权利要求所限定的本发明的范围的情况下，可对其作出各种形式和细节上的改变。因此，本发明只打算由所附权利要求所限。

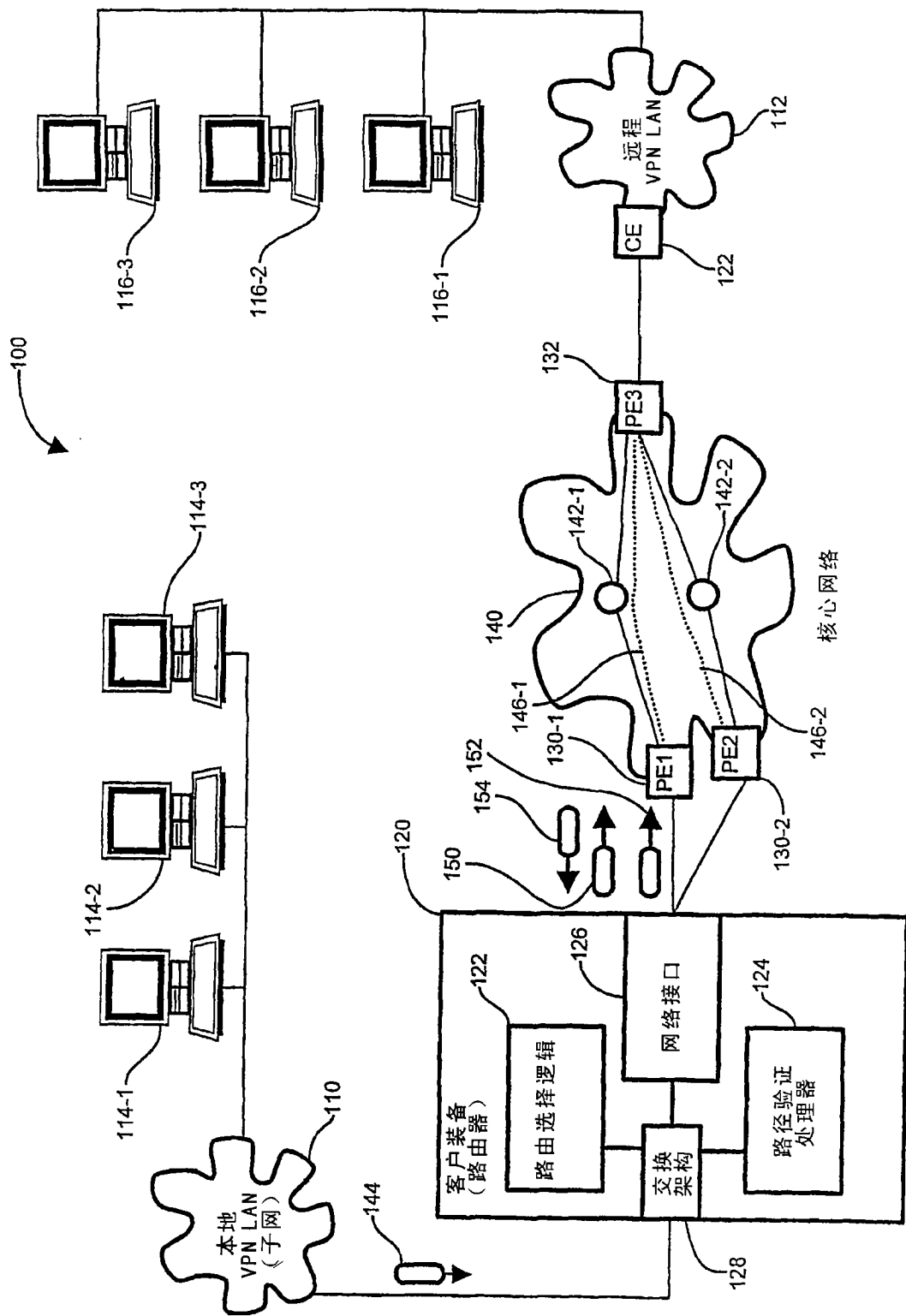


图1

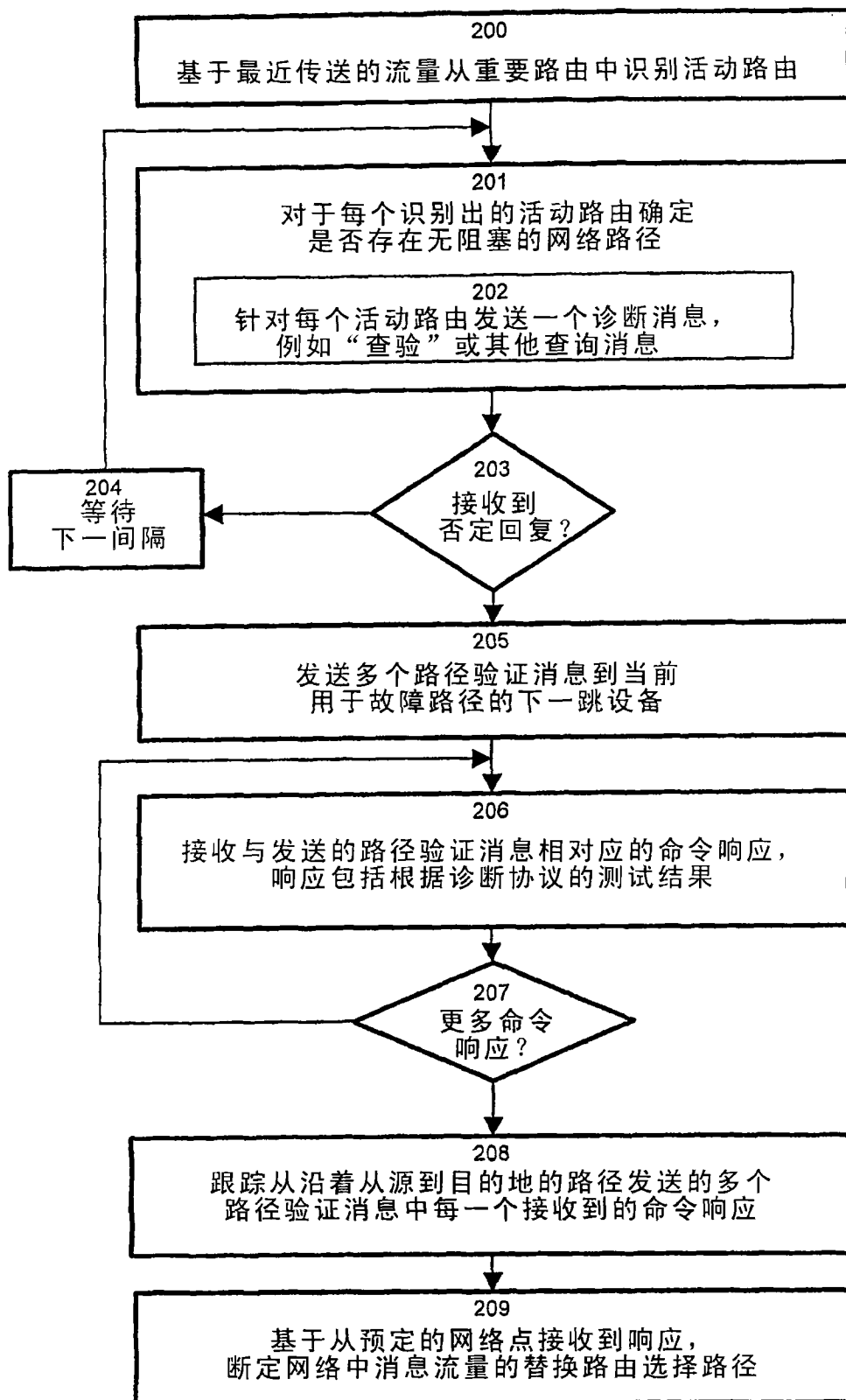


图 2

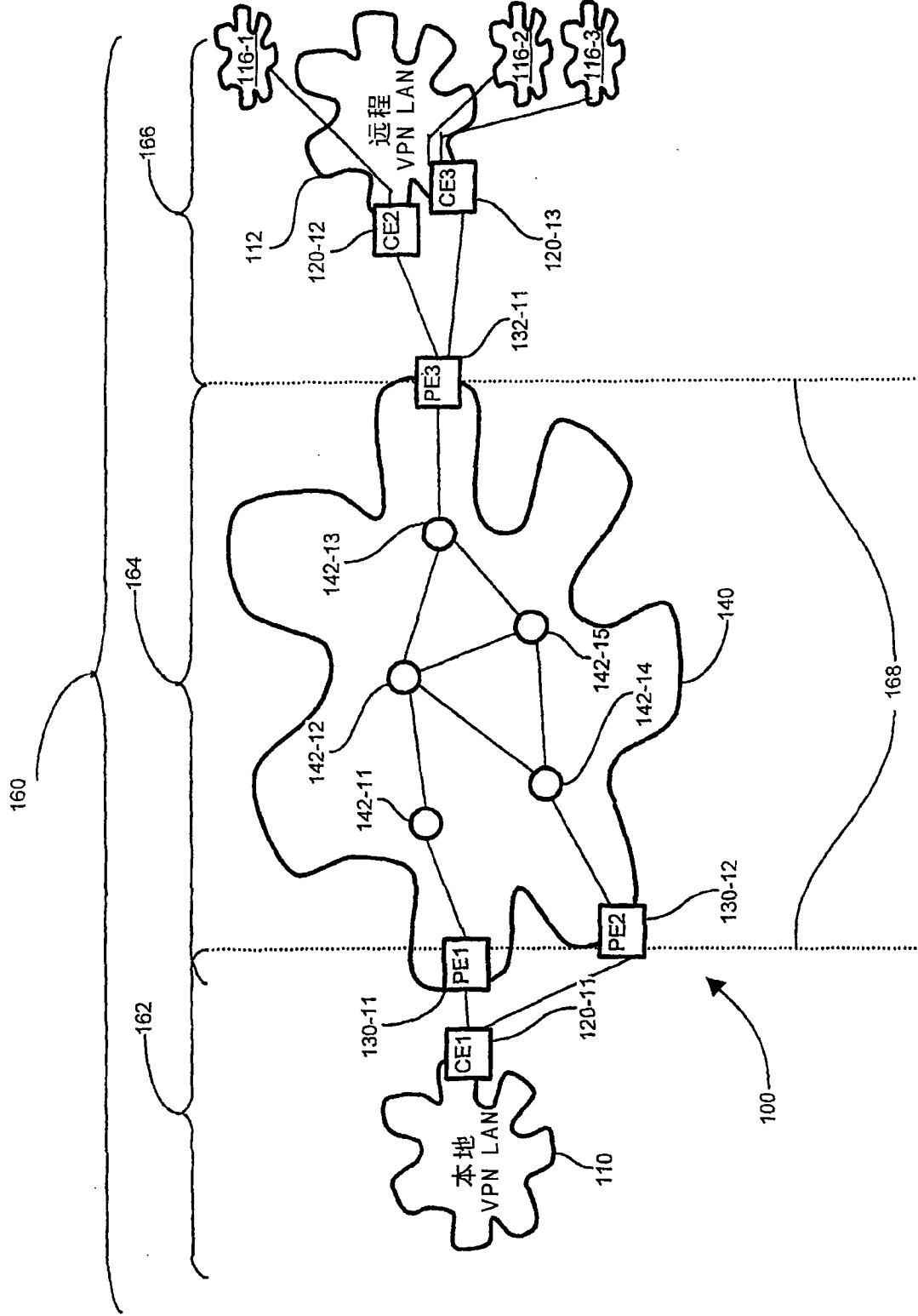


图3

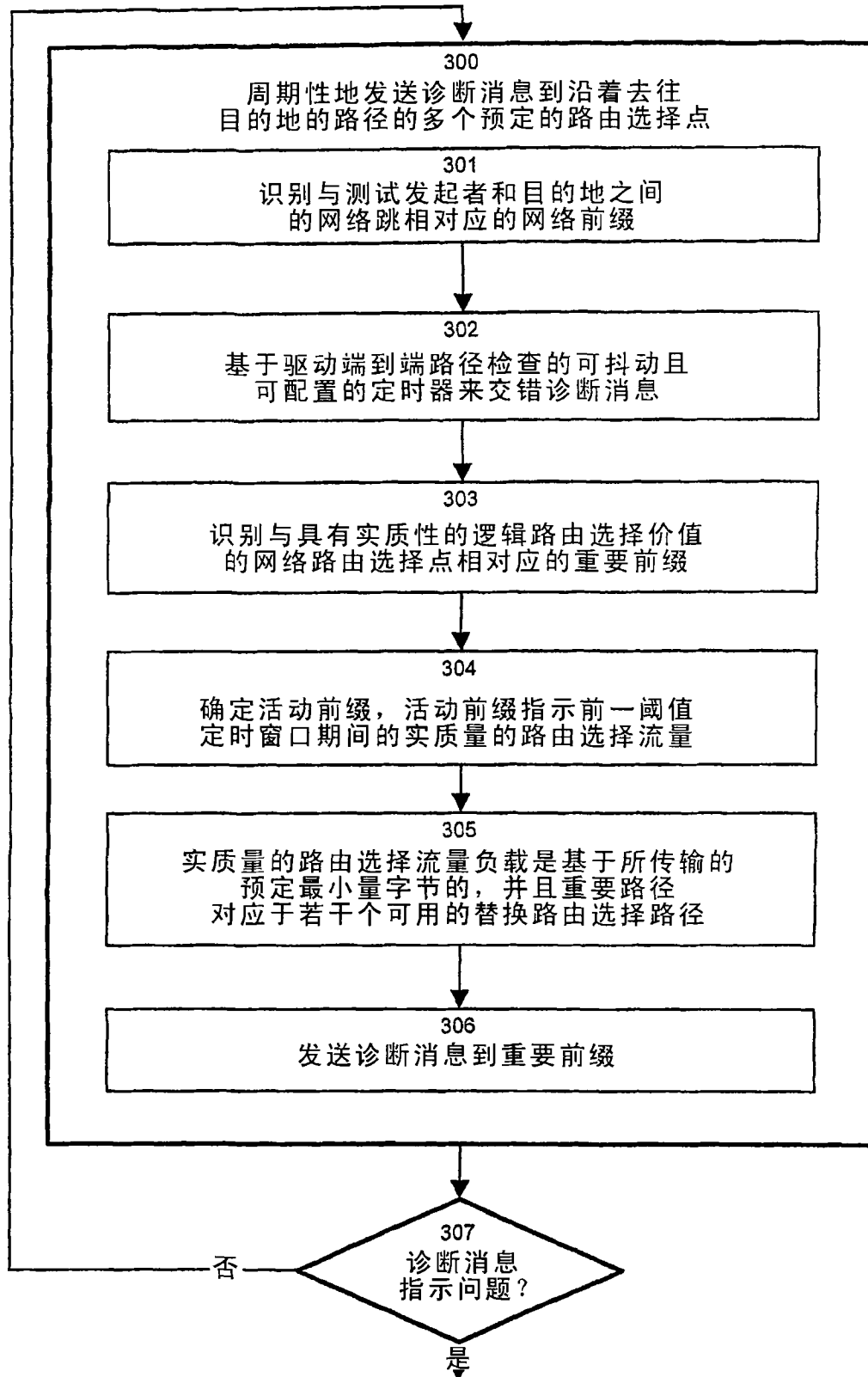


图 4

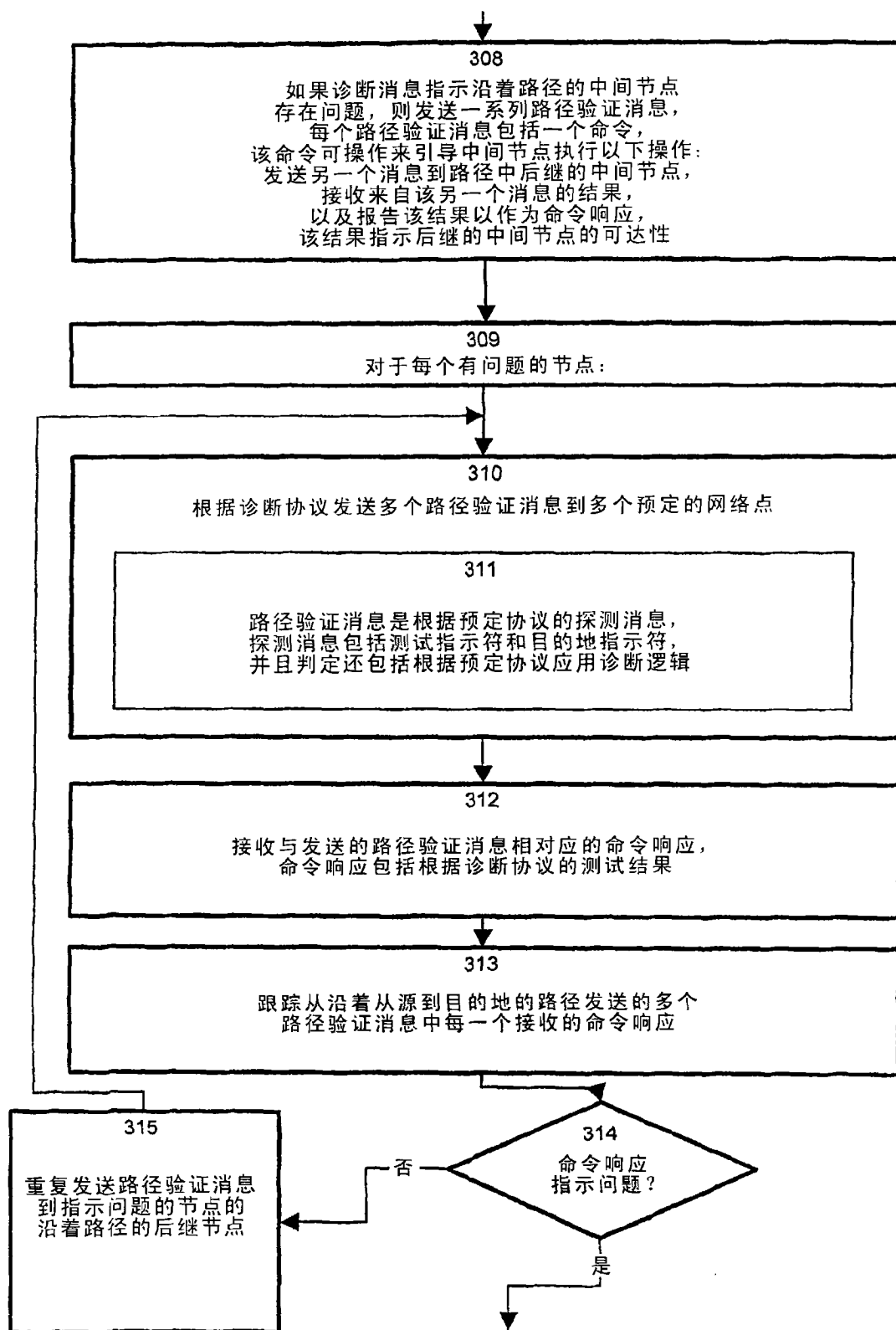


图 5

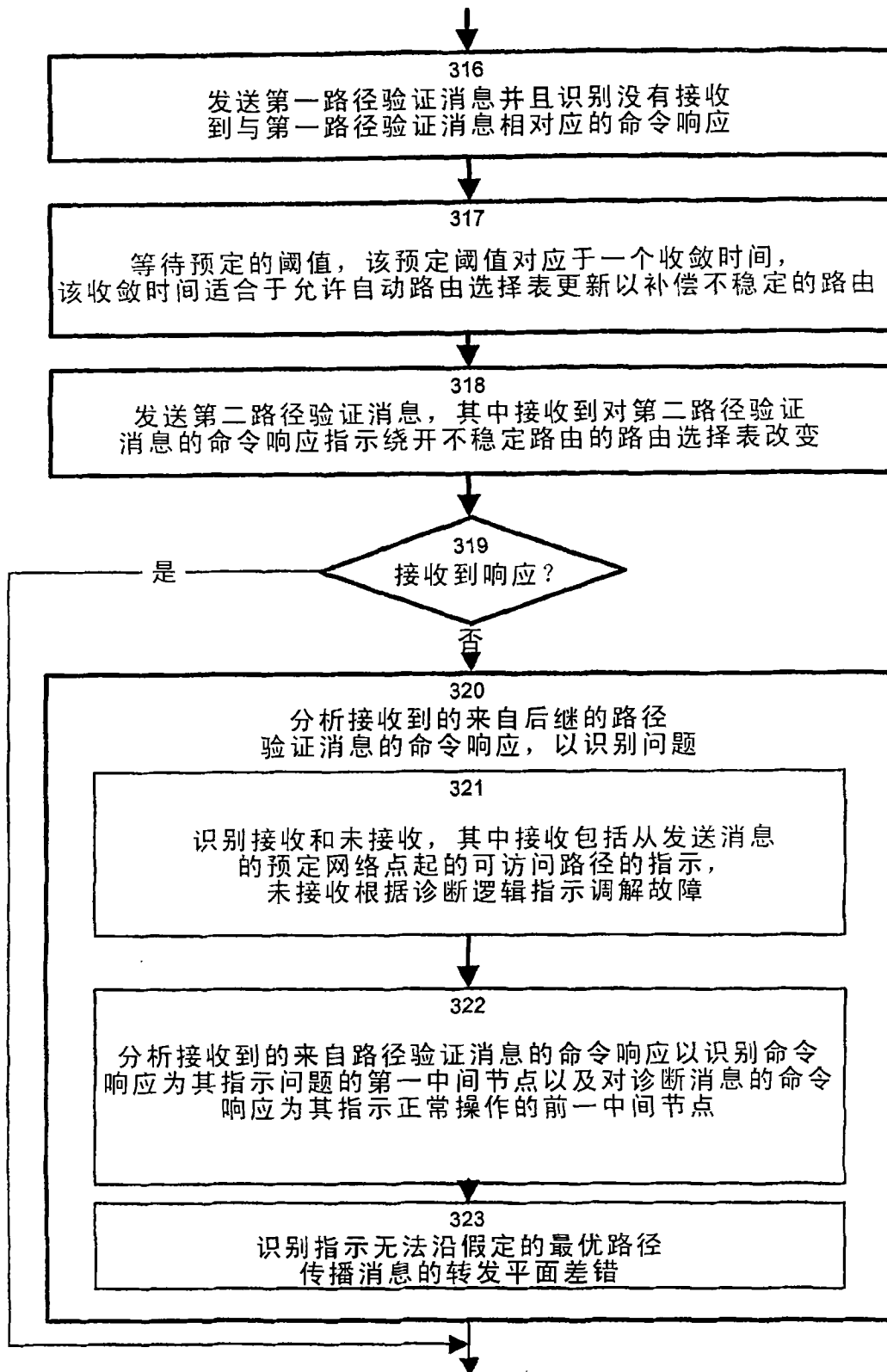


图 6

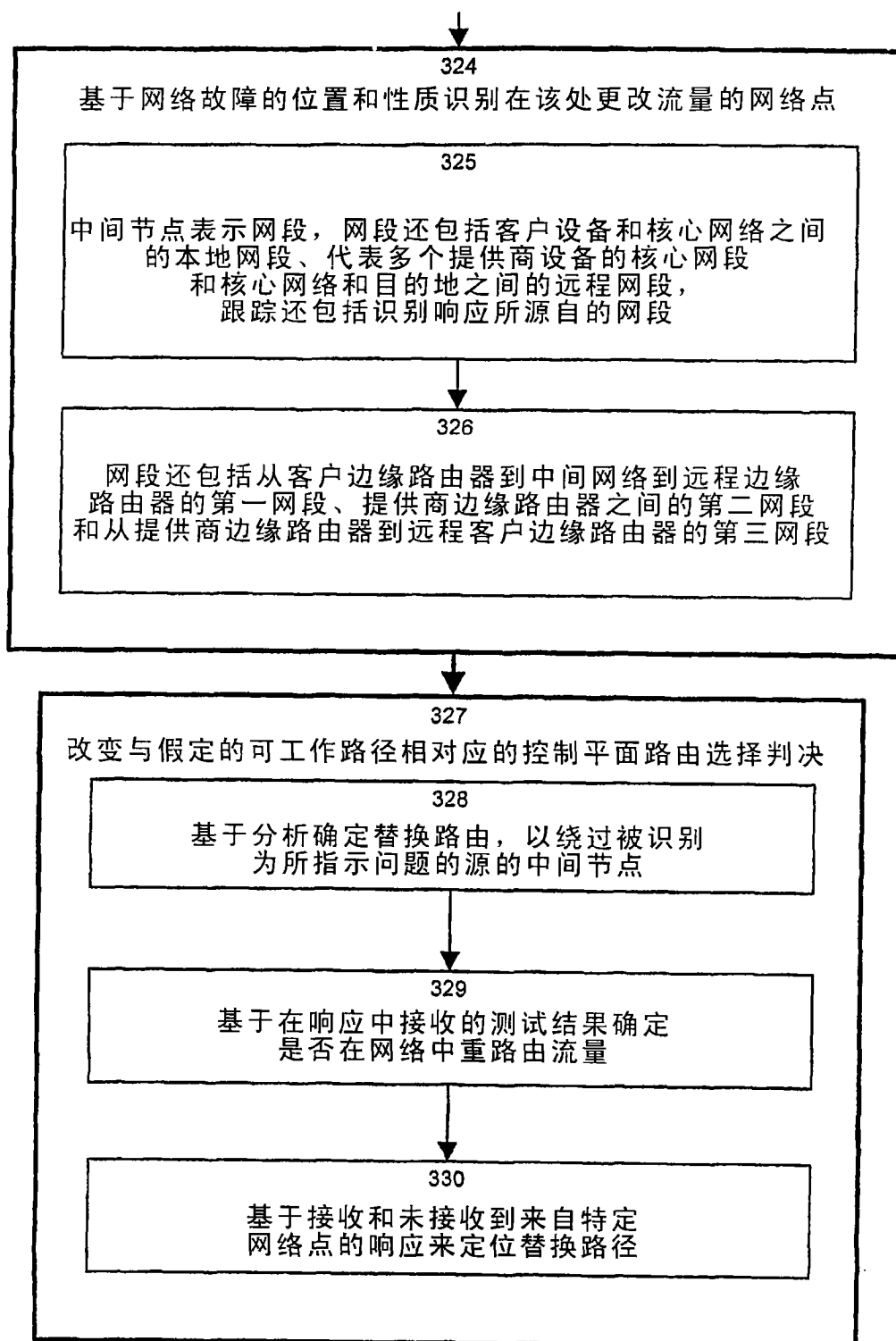


图 7