



(12)发明专利

(10)授权公告号 CN 104025552 B

(45)授权公告日 2017.06.09

(21)申请号 201380004653.0

埃里克斯·塔斯特 杨一斌

(22)申请日 2013.03.26

(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 11258

(65)同一申请的已公布的文献号

申请公布号 CN 104025552 A

代理人 李晓冬

(43)申请公布日 2014.09.03

(51)Int.Cl.

H04L 29/06(2006.01)

(30)优先权数据

H04L 12/701(2006.01)

13/458,838 2012.04.27 US

(85)PCT国际申请进入国家阶段日

2014.06.30

(86)PCT国际申请的申请数据

PCT/US2013/033811 2013.03.26

(87)PCT国际申请的公布数据

W02013/162807 EN 2013.10.31

(73)专利权人 思科技术公司

地址 美国加利福尼亚州

(56)对比文件

CN 101529763 A, 2009.09.09,

CN 102404172 A, 2012.04.04,

US 2006/0072471 A1, 2006.04.06,

CN 101529763 A, 2009.09.09,

CN 102404172 A, 2012.04.04,

CN 101945042 A, 2011.01.12,

CN 101472344 A, 2009.07.01,

审查员 叶珊

(72)发明人 周宇晨 理查德·斯图尔特

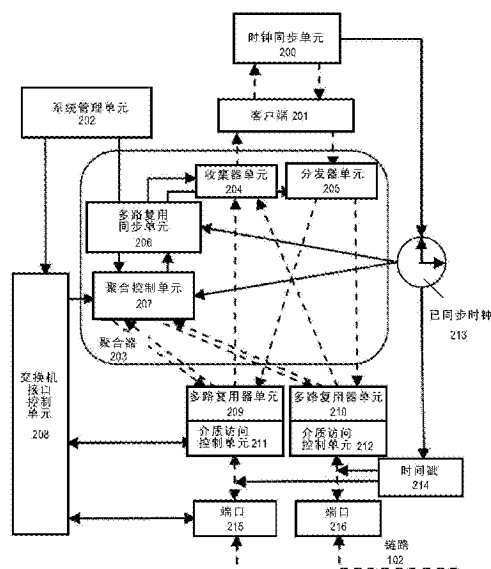
权利要求书2页 说明书10页 附图5页

(54)发明名称

链路聚合中流量多路复用的同步

(57)摘要

描述了链路聚合中多路复用的同步技术。在一实施例中，第一链路聚合器和第二链路聚合器与多个链路相关联。第一链路聚合器为每个链路保持一个标识符，至少指示启用或禁用状态。在第一链路聚合器和第二链路聚合器之间建立一个同步的时钟。多个链路中的特定链路被转变。其中，第一链路聚合器给第二链路聚合器发送第一消息，标识转变特定链路的特定时间。第一链路聚合器接收来自第二链路聚合器的第二消息，指示所述特定的时间是否可以接受。响应判定第二消息指示该特定时间可以接受并且同步的时钟达到特定时间，转变链路。



1. 一种用于链路转变同步的方法,包括:

将第一链路聚合器设备和第二链路聚合器设备与多个网络数据通信链路相关联,其中所述第一链路聚合器设备为每个链路保持指示至少启用或禁用状态的标识符;

在所述第一链路聚合器设备和所述第二链路聚合器设备之间建立同步的时钟;

通过以下步骤,转变多个链路中的特定链路到不同状态:

所述第一链路聚合器设备给所述第二链路聚合器设备发送标识转变特定链路的特定时间的第一消息;

所述第一链路聚合器设备从所述第二链路聚合器设备接受指示所述特定时间是否能够被接受的第二消息;

响应于判定所述同步时钟达到特定时间、并且所述第二消息指示该特定时间能够被接受,更新所述特定链路的所述标识符为所述不同状态;

响应于判定所述第二消息提出新时间并且所述同步时钟已达到所述新时间,更新所述特定链路的标识符为所述不同状态。

2. 如权利要求1所述的方法,还包括将所述第一链路聚合器设备和所述第二链路聚合器设备与新链路相关联,其中所述新链路处于禁用状态,所述特定链路是所述新链路,并且所述不同状态是启用状态。

3. 如权利要求1所述的方法,还包括移除所述特定链路与所述第一链路聚合器设备和所述第二链路聚合器设备的关联,以及将所述特定链路的标识符更新为指示禁用状态。

4. 如权利要求1所述的方法,其中所述第一链路聚合器设备包括收集到达多个链路上的数据帧的收集器单元,和使用所述多个链路发送来自客户端的数据帧的分发器单元;其中所述多个链路中每个单独链路的标识符被配置为控制收集器和分发器是否能够与各个链路进行交互。

5. 如权利要求1所述的方法,还包括将一个或多个设置值与所述第一链路聚合器设备和所述第二链路聚合器设备相关联,其中转变所述特定链路响应于所述设置值的改变而发生。

6. 如权利要求5所述的方法,其中设置值指示允许链路的最大数量、服务质量、或链路安全策略中的至少一项或多项。

7. 一个计算机系统,该系统包括:

被配置为将第一链路聚合器设备和第二链路聚合器设备与多个网络数据通信链路相关联的交换机接口控制单元,其中所述第一链路聚合器设备为每个链路保持指示至少启用或禁用状态的标识符;

被配置为在所述第一链路聚合器设备和所述第二链路聚合器设备之间建立同步的时钟的时钟同步单元;

被配置为通过以下步骤转变多个链路中的特定链路到不同状态的聚合控制单元:所述第一链路聚合器设备给所述第二链路聚合器设备发送标识转变特定链路的特定时间的第一消息;所述第一链路聚合器设备从所述第二链路聚合器设备接受指示所述特定时间是否能够被接受的第二消息;响应于判定所述同步时钟达到特定时间、并且所述第二消息指示该特定时间能够被接受,更新所述特定链路的所述标识符为所述不同状态;响应于判定所述第二消息提出新时间并且所述同步时钟已达到所述新时间,更新所述特定链路的标识符

为所述不同状态。

8. 如权利要求7所述的系统,其中聚合控制单元还被配置为将所述第一链路聚合器设备和所述第二链路聚合器设备与新链路相关联,其中所述新链路处于禁用状态,所述特定链路是所述新链路,并且所述不同状态是启用状态。

9. 如权利要求7所述的系统,其中聚合控制单元还被配置为移除所述特定链路与所述第一链路聚合器设备和所述第二链路聚合器设备的关联、并且将所述特定链路的标识符更新为指示禁用状态。

10. 如权利要求7所述的系统,其中所述第一链路聚合器设备包括收集到达多个链路上的数据帧的收集器单元,和使用所述多个链路发送来自客户端的数据帧的分发器单元;其中所述多个链路中每个单独链路的标识符被配置为控制收集器和分发器是否能够与各个链路进行交互。

11. 如权利要求7所述的系统,其中聚合控制单元被配置为将一个或多个设置值与所述第一链路聚合器设备和所述第二链路聚合器设备相关联,其中转变特定链路响应于所述设置值的改变而发生。

12. 如权利要求11所述的系统,其中设置值指示允许链路的最大数量、服务质量、或链路安全策略中的至少一项或多项。

13. 如权利要求7中所述系统,包括任何分组数据交换机或分组数据路由器。

链路聚合中流量多路复用的同步

技术领域

[0001] 本公开总体上涉及一种由链路耦接的节点构成的数据通信网络中的链路聚合。

背景技术

[0002] 本节描述的方法可以被实行,但并不一定是先前已经构想出或者实行过的方法。因此,除非本文指出,本节描述的方法对本申请权利要求来说不是现有技术,并且不会因为包含在本节中就被认为是现有技术。

[0003] 一般地,链路聚合是组合多个并行网络连接(称为链路)以增加数据吞吐量,使之大于单个链路可以维持的吞吐量,或者为链路故障时进行保护引入冗余的过程。其结果是,多个被聚合的链路可以用作好像它们是网络拓扑结构中的单个链路。每一个被聚合的链路可以代表一个网络的物理链路,例如电力线或无线连接,或者网络的虚拟链路。

[0004] 因此,链路聚合可以发生在计算机网络层七OSI模型的多个不同层中。在第二层(也被称为数据链路层)中,聚合可以发生在将帧从一个网络传送到另一个网络设备的虚拟或物理端口之间。此外,在第四层(或称为传输层)中,链路传输部分,例如TCP连接,也可以被聚合。

[0005] 为建立网络的两个设备之间的链路聚合,这两个设备通常对待聚合的一组链路进行协商并达成一致。这组链路的协商过程由聚合协议控制管理。被称为链路聚合控制协议(LACP)的一个聚合协议,在数据链路层携带MAC帧的链路上执行链路聚合。在一个名为“局域网和城域网IEEE标准-链路聚合”的文档中描述了LACP协议,该文档的作者是IEEE计算机协会,该文档的全部内容通过引用被出于所有目的结合于此,其内容视为已充分阐述。被称为端口聚合协议(PAgP)的另一个聚合协议,在一组连接两个或多个网络设备的物理的或虚拟的端口上执行链路聚合。在一个标题为“端口聚合协议”的文档中对PAgP进行了描述,该文档作者为加利福尼亚州圣何塞的思科系统有限公司,该文档的全部内容通过引用被出于所有目的结合于此,其内容视为已充分阐述。

发明内容

[0006] 所附权利要求书可以作为本发明的概述。

附图说明

[0007] 在附图中:

[0008] 图1示出了实施例可以在其上执行链路聚合的两个网络设备。

[0009] 图2示出了用于可以同步链路转变的网络设备的链路聚合设计实施例。

[0010] 图3以状态图的形式示出了能发起链路传输同步的网络设备实施例。

[0011] 图4以状态图的形式示出了响应由另一个网络设备发起的链路传输同步的网络设备实施例。

[0012] 图5示出了在其上系统可以被实现的计算机系统。

[0013] 示例实施例说明

[0014] 描述了链路聚合中多路复用的同步技术。在下面的描述中,为了解释的目的,阐明许多具体的细节以提供对本发明的充分理解。然而,对本领域的技术人员来说,显然在没有这些具体细节的情况下也可以实现本发明。在其它示例中,以方框图的形式显示了公知的结构和设备以避免不必要地模糊了本发明。

[0015] 本文中根据以下提纲对实施例进行描述:

[0016] 1.0概述

[0017] 1.1示例网络

[0018] 1.2链路聚合

[0019] 2.0结构概述

[0020] 3.0功能概述

[0021] 3.1同步启动器

[0022] 3.2同步响应器

[0023] 4.0链路聚合控制协议示例

[0024] 5.0执行机制-硬件概述

[0025] 6.0扩展和替代

[0026] 1.0概述

[0027] 描述了链路聚合中多路复用的同步。在实施例中,第一链路聚合器和第二链路聚合器与多个链路相关联。第一链路聚合器为每个链路保持指示至少启用或禁用状态的识别符。在第一链路聚合器和第二链路聚合器之间建立同步时钟。多个链路中的特定链路被转变。转变的执行是通过所述第一链路聚合器给第二链路聚合器发送标识传输特定链路的特定时间的第一消息。第一链路聚合器接收来自第二链路聚合器的指示所述特定时间可以接受的第二消息。响应于判定第二消息指示该特定时间可以接受并且同步时钟到达特定时间,转换链路。

[0028] 在其它实施例中,本发明包括计算机装置和被配置为执行前述步骤的计算机可读介质。

[0029] 1.1示例网络

[0030] 图1示出了在其上可以实施例执行链路聚合的两个网络设备。图1中,网络设备100和网络设备101通过链路102连接,此外,网络设备100和网络设备101都是网络103的一部分。虽然图1中只描述了网络设备100和101,但是网络103可以包含未在图1中显示的许多其它链路和网络设备。

[0031] 在实施例中,网络设备100和网络设备101是负责为通过网络103的流量提供路由的交换机。在其它实施例中,网络设备100,101代表交换机、网关、防火墙、服务器、客户端端或任何其它类型的可以被包括在网络103内的设备。

[0032] 在实施例中,链路是物理链路,例如连接网络装置100和网络装置101的无线或以太网链路。在其它实施例中,链路是虚拟链路。意味着,在网络设备100和网络设备101之间可以存在着没有在图1中描述的物理设备。

[0033] 1.2链路聚合

[0034] 在一些实施例中,网络设备100、101使用聚合协议,如LACP或PAgP,协商和建立能

够聚合的多组链路。在一些实施例中,聚合协议把链路102分成一个或多个链路聚合组(LAG),其中每个LAG作为网络103的拓扑结构中的单个链路。

[0035] 然而,不是所有链路都能分在同一个LAG组中。在一些实施例中,链路本身是不相容的。例如,一个网络设备可能只能聚合端口位于相同电路板的链路或者以相同的速度传送流量的链路

[0036] 在其它实施例中,每个LAG与控制LAG内链路属性的设置相关联。例如,设置可以包括链路的最大数量,链路服务质量,或者链路安全策略。其结果是,如果参与将违反设置,则物理上能够聚合的链路仍可能被禁止加入同一个LAG。

[0037] 在一些实施例中,LAG初始建立后,网络设备100和网络设备101重新协商LAG并且据此重组链路102和LAG之间的关联。例如,与一个或多个LAG相关联的设置可能被修改。因此,聚合协议重新协商并使LAG适应新的设置。作为另一个例子,网络设备100和网络设备101之间的链路可能随时间而改变。某些链路可能会发生故障或者被移除,而其它链路可能会被修复或补充。其结果是,LAG被重新排列以补偿或优化新链路的增加或已有链路的移除。

[0038] 分发至LAG的流量在LAG内的链路中进行分发。判定如何在LAG内的链路中划分流量的策略可能基于许多不同的因素。在一些实施例中,流量划分基于源地址或目的地址。例如,帧的源地址或目的地址可以由哈希函数被散列,由结果判定哪个链路传送该帧。在其它实施例中,流量在LAG内的链路中以循环的方式被划分。然而,实施例选择要实施的确切策略不是本文描述的技术的关键。

[0039] 在一些实施例中,LAG内的链路中只有一个子集在传输流量时被活动地使用。称为备用链路的其它链路,被保持为备用以防活动链路发生故障。一旦链路故障被检测到,备用链路就替代故障电路以维持网络连接。

[0040] 当链路被添加到LAG组或从备用链路转变到活动链路,链路可以使用之前网络设备100和网络设备101都必须使链路在线。在实施例中,网络设备在使能链路上的收集和分发时使得链路在线。在这种实施例中,链路上的收集和分发可以被耦接在一起,这样使能其中一个意味着使能另一个,反之亦然。然而,在其它实施例中,网络设备能在链路上独立地使能收集和分发。结果,网络设备100和网络设备101可以分贝使得链路上的每个方向在线。例如,如果网络设备100在链路上使能了分发而网络设备101在同一个链路上使能了收集,从网络设备100到网络设备101的单向链路已由这两个设备在线。

[0041] 在一些情况下,一个网络设备,例如网络设备100,可以比其它的网络更快地使设备使链路在线。因此,在上述的示例中,网络设备101在链路上使能接收之前,网络设备100在链路上使能了分发并开始传送帧。结果,由网络装置100分发的帧被丢弃,直到网络设备101在链路上使能了收集。

[0042] 类似的问题也会发生在当链路从LAG组中被移除,或从一个活动状态转移到待机状态时。例如,如果在网络设备101禁用分发之前网络设备100在链路上禁用了收集,通过链路从网络设备100传输到网络设备101的流量将会被丢弃。

[0043] 为了示出清楚的示例实施例的目的,链路的“转变”包括引起链路从离线状态变为在线状态或者相反。

[0044] 在实施例中,网络设备100和网络设备101之间的链路转变被同步。这种同步确保

当一个网络设备在链路一端转变链路时,链路另一端的网络设备在几乎相同的时间也在进行同样的转变。其结果是,网络设备在另一个网络设备转变链路之前转变链路带来的延时造成的流量损失被消除或减至最小。

[0045] 2.0结构概述

[0046] 为了解释的目的,将参照数据链路层对技术进行描述。因此,链路将被描述为进行帧传输。然而,本文描述的技术不仅限于传送数据链路层流量的聚合链路。在其它实施例中,链路可以遵循任何数量的不同标准或协议传送消息,如因特网协议(IP)或者多协议标签交换(MPLS)。

[0047] 图2示出用于可以同步链路传输的网络设备的链路聚合设计实施例。为了解释的目的,将假设图2中的链路聚合设计已在网络设备100和网络设备101上实现。然而,在其它实施例中,网络设备100的链路聚合设计可以不同于网络设备101的链路聚合设计。本文中的“链路聚合设计”指的是可被配置为如图2所示并且能被安排完成本文描述的功能的计算设备、逻辑设备、软件元件、固件元件或硬件元件的任意组合。

[0048] 来自链路102的帧到达端口215或端口216,具体取决于传输帧的特定链路。在一些实施例中,端口215、216是物理端口。然而,在其它实施例中,端口215、216可以是虚拟端口。在这种实施例中,同一个链路可连接到多个虚拟端口。帧通过端口215、216到达相应的介质访问控制单元(MAC) 211、212,介质访问控制单元(MAC) 211、212被配置为给帧提供寻址和信道访问控制机制。一旦帧被MAC单元211或MAC单元212处理过,多路复用器单元209或多路复用器单元210中的相应的多路复用器单元就将该帧转发至作为帧的预定接收者的组件的聚合器203,例如收集器单元204或聚合控制单元207。在实施例中,帧的预定接收者由包含在帧的头部中字段判定,例如类型-长度-值(TLV) 字段或标记。

[0049] 聚合器203聚合来自多路复用器单元209、210的帧,并包含收集器单元204、分发器单元205、复用同步单元206、以及聚合控制单元207。

[0050] 收集器单元204收集由多路复用器单元209、210转发至它的帧,然后将那些帧提供给客户端201。在一些实施例中,客户端201是为进入和离开聚合器的帧提供寻址和信道访问控制机制的MAC单元。

[0051] 分发器单元205接收来自客户端201的帧并将那些帧分发到多路复用器单元209、210,用于沿链路102传输。每个被分发的帧被MAC单元211或MAC单元212处理并且通过和接收帧的多路复用单元对应的端口215、216。在一些实施例中,分发器单元205基于聚合控制单元207保持的链路分发策略判定哪个多路复用单元,以及相应的哪个链路用于传输帧。在实施例中,分发策略由一组存储在各单元共享存储区域的共享状态变量代表,所述共享状态变量存储在分发器单元205和聚合控制单元207之间共享。

[0052] 多路复用同步单元206控制哪些链路被使能用于收集器单元204和分发器单元205。在实施例中,多路复用同步单元206通过一组存储在各单元共享存储区域的共享状态变量代表。然而,在其它实施例中,多路复用同步单元可以通过发送指示使能链路或禁用链路的链路控制帧给收集器单元204和分发器单元205,来控制收集器单元204和分发器单元205。

[0053] 此外,多路复用同步单元206和位于链路102另一端的网络设备协同进行链路转变。因此,多路复用同步单元206可以引起分发器单元205通过链路102给另一个网络设备发

送同步控制帧。类似地,多路复用器单元209和多路复用器单元210接收的同步控制帧被转发到多路复用同步控制单元206进行处理。在一些实施例中,多路复用器单元209和多路复用器单元210直接将同步控制帧转发至多路复用同步单元206。然而,在其它实施例中,可以间接地进行转发,例如通过聚合控制单元207进行转发。

[0054] 聚合控制单元207管理链路102通过其被划分成LAG的策略以及帧通过其在LAG内的链路间被分发的策略。聚合控制单元207被配置成接收来自系统管理单元202和交换机接口控制单元208的指令,上述系统管理单元202和交换机接口控制单元208通过改变与聚合控制单元207相关的设置修改策略。例如,系统管理单元202可以是允许用户管理聚合控制单元207设置的图形用户界面或命令行。

[0055] 此外,聚合控制单元207和位于链路102另一端的网络设备协同进行LAG的建立和维护。因此,聚合控制单元207可以引起分发器单元205给其它的网络设备发送聚合控制帧。由于同样的原因,由多路复用器单元209或多路复用器单元210接收的聚合控制帧被转发到聚合控制单元207进行处理。

[0056] 此外,聚合控制单元207可以判定什么时候链路应当被转变。这种判定可以基于许多因素,包括与聚合控制单元207相关的设置、设置的修改、链路连接的变化、和/或与其它网络设备的通信。为了转变链路,聚合控制单元207使用多路复用同步单元206来同步与链路另一端的网络设备的转变。

[0057] 交换机接口控制单元208,是被配置为允许系统管理单元202控制聚合控制单元207、MAC单元211、212和端口215、216的接口。其结果是,系统管理单元202可以操纵与上述部件相关的设置,例如能够修改聚合控制单元207策略,或者MAC单元211、212的寻址和介质访问机制的设置。

[0058] 时钟同步单元200与链路102另一端的网络设备建立和维护同步时钟213。同步时钟213被用于产生可被插入在MAC单元211、212和端口215,216之间传输的帧中的时间戳214。此外,同步时钟213给多路复用转变器单元206和聚合控制单元207提供系统时钟。

[0059] 在所示出实施例中,为了示出清楚示例的目的,图2中只描述了特定数目的每个组件;然而,本文描述的技术不限于所描述的每个组件的数量。例如,实施例可能维持多个聚合器,其中每个聚合器代表不同的LAG。此外,该实施例仅描述两个链路;然而,本文描述的技术可以被应用于任何数量链路的链路聚合,包括非常大量的链路。

[0060] 3.0功能概述

[0061] 图3以状态图的形式示出发起链路传输同步的网络设备实施例。图4以状态图的形式示响应由另一个网络设备发起的链路传输同步的网络设备实施例。为了示出清楚示例的目的,图3将会由网络设备100代表,图4将会由网络设备101代表。

[0062] 以网络设备100代表图3,网络设备101代表图4是任意的选择,仅打算作为示出清楚示例的一种手段。在一些实施例中,网络设备100和网络设备101将既执行状态图3又执行状态图4。在这种实施例中,发起同步的网络设备遵循图3描述的状态图,而响应同步的网络设备遵循图4描述的状态图。由哪一个设备发起同步取决于许多因素,包括网络设备聚合控制单元的策略、用户定义的设置、和/或网络设备的相对处理能力,等等。然而,哪一个网络设备发起同步不是本文描述的技术的关键。

[0063] 正如前面提到的,这两个网络设备100和网络设备101被假定为实现图2所示的链

路聚合设计。其结果是,关于这两个网络设备100和网络设备101将会引用图2。为避免混淆,出现在3.1节中和图2相关的参考标号是关于网络设备100的,出现在3.2节中和图2相关的参考标号是关于网络设备101的。

[0064] 3.1同步启动器

[0065] 参照图3,网络设备100与网络设备101在方框3建立同步的时钟。在实施例中,时钟同步单元200与网络设备101建立同步的时钟213。在一些实施例中,时钟同步单元200实施精密时间协议(PTP),在由仪器与测量委员会起草的一份题为“IEEE标准网络测量和控制系统的精密时钟同步协议”的文件中对该协议进行了描述,该文档的全部内容通过引用被结合于此,其内容视为已充分阐述。在另一个实施例中,时钟同步单元200实施网络时间协议(NTP),在一份由互联网工程任务组起草的题为“网络时间协议版本四:协议和算法规范”的文件中对该协议标准进行了描述,该文档的全部内容通过引用被结合于此,其内容视为已充分阐述。然而,时钟同步单元200建立同步的时钟213所实施的确切时间同步协议并不是本文所述技术的关键。

[0066] 在方框301,网络设备100判定链路准备好被转变。在实施例中该判定由聚合控制单元207执行。聚合控制单元207基于各种不同因素判定转变链路,例如与聚合控制单元207相关的设置、设置的修改、链路连接的变化、和/或与网络设备101的通信。在实施例中,方框301执行判定后,聚合控制单元207调用多路复用同步单元206同步链路转变。

[0067] 在方框302,网络设备100给网络设备101发送指示链路转变的时间的消息。在实施例中,多路复用同步单元206通过生成同步控制帧来发送消息,并且使用分发器单元205将同步控制帧递送到网络设备101。同步控制帧指示的时间是至少部分地基于方框300中时钟同步单元200建立的同步的时钟213。在一些实施例中,同步控制帧指示的时间是在同步的时钟213指示的现有时间前的一组特定长度的时间。在实施例中,特定的时间长度由聚合控制单元207管理的一组设置控制。在另一些实施例中,基于一些因素例如收集到的关于网络设备101的速度和/或工作量的统计数据等,时间长度是可变的。在实施例中,分发器单元205在同一个将要被转变的链路上发送同步控制帧。然而,在其它实施例中,发送同步控制帧的链路由聚合控制单元207的策略判定。

[0068] 在方框303,网络设备100接收来自网络设备101的确认。在实施例中,该确认是同步控制帧。因此,当同步控制帧到达任一多路复用器单元209或210,该多路复用器单元直接或间接地将同步控制帧转发至多路复用同步单元206。

[0069] 在方框304,网络设备100判定确认是否提出了新的时间。在实施例中,该判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过检查同步控制帧头部中的标记执行判定。在其它实施例中,多路复用同步单元206通过检查帧内特定TLV字段的值执行判定。在其它实施例中,多路复用同步单元通过比较方框302提出的时间和同步控制帧内字段的值执行判定。

[0070] 如果确认没有提出新时间,则在方框305,网络设备100判定同步的时钟已达到方框302发送的消息中指示的时间。在实施例中,判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过周期性地检查同步时钟213是否达到方框302发送的同步控制帧提出的时间执行判定。

[0071] 如果确认确实提出新时间,则在方框306,网络设备100判定同步的时钟已达到方

框303接收的消息中指示的时间。在实施例中,判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过周期性地检查同步的时钟213是否达到方框303接收的同步控制帧提出的时间执行判定。

[0072] 在方框307,网络设备100转变链路。在实施例中,判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过操纵一个标识符来转变链路,该标识符被配置为控制收集器单元204和分发器单元205在一个给定链路上为使能或是禁用状态。在实施例中,网络设备100能够独立地控制收集器单元204和分发器单元205。在这种实施例中,收集器单元204和分发器单元205由不同的标识符来控制。在其它实施例中,收集器单元204和分发器单元205的控制是相互联系的。在这种实施例中,收集器单元204和分发器单元205由同一个标识符控制。

[0073] 3.2同步响应器

[0074] 参照图4,网络设备101与网络设备100在方框400建立同步的时钟。在实施例中,时钟同步单元200与网络设备100建立同步的时钟213。在一些实施例中,时钟同步单元实施PTP。在另一些实施例中,时钟同步单元200实施NTP。然而,时钟同步单元200建立同步时钟213所采用的确切时间同步协议不是本文所述技术的关键。

[0075] 在方框401,网络设备100接收指示转变链路的的时间的消息。在实施例中,该消息是同步控制帧。因此,当同步控制帧到达任一多路复用器单元209或210,该多路复用器单元将同步控制帧转发至多路复用同步单元206。在一些实施例中,转变链路的时间被储存在同步控制帧内的TLV中。

[0076] 在方框402,网络设备101判定到消息中指示的时间时链路是否准备好转变。在实施例中,判定由多路复用同步单元206基于方框401接收的同步控制帧指示的时间执行。这个判定可以基于许多因素,例如网络设备101的处理能力、网络设备101的当前工作负荷、基于过去性能的统计分析、和/或仍然有待被完成的工作等等,以使链路做好转变的准备。在一些实施例中,为网络设备402维持阈值时间值。在这种实施例中,多路复用同步单元206通过将同步控制消息提出的时间和同步的时钟213指示的当前时间之间的不同与阈值时间值的进行比较来执行判定。在实施例中,阈值时间值是由聚合控制单元207维持的设置,聚合控制单元207可以被系统管理单元202和/或交换机接口控制单元208控制。

[0077] 如果到消息中指示的时间时链路将要准备好,则在方框405网络设备101发送接受消息指示的时间的确认。在实施例中,多路复用同步单元206通过生成同步控制帧并使用分发器单元205将同步控制帧递送到网络设备100来发送确认。在实施例中,复用同步单元206通过在同步控制帧的头部中标记指示该时间被接受。在另一个实施例中,复用同步单元206通过包含字段,例如TLV字段来指示该时间被接受,同一时间被方框401接收。

[0078] 接下来,在方框406,网络设备101判定同步的时钟已达到方框401接收的时间。在实施例中,判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过周期性地检查同步的时钟213是否达到方框401接收的同步控制帧提出的时间执行判定。

[0079] 如果到消息中指示的时间时链路没有准备好,则在方框403网络设备101发送指示新的时间的确认。在实施例中,通过使用有关方框405中描述的同样的技术,确认被发送到网络设备100。然而,标记将指示一个新时间被提出和/或字段将包含一个与方框401接收到的时间不同的时间。在一些实施例中,确认指示的时间是基于有关方框402描述的阈值时间

值。

[0080] 接下来,在方框404,网络设备101判定同步的时钟已达到方框403发送的新时间。在实施例中,判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过周期性地检查同步时钟213是否达到方框403发送的同步控制帧指示的时间执行判定。

[0081] 在替代的实施例中,在方框403发送新提出的时间是可选的。例如,如果网络设备101不同意网络设备100提出的时间,该确认仅指示该提出的时间不可接受,而不会提出新的时间,相应地,链路转变同步可以被终止,网络设备100和网络设备101可以各自独立地转变链路。

[0082] 在方框407,网络设备101转变链路。在实施例中,判定由多路复用同步单元206执行。在一些实施例中,多路复用同步单元206通过控制标识符转变链路,该标识符被配置为控制收集器单元204和分发器单元205在一个给定链路上为使能或是禁用状态。在实施例中,网络设备100能够独立地控制收集器单元204和分发器单元205。在这种实施例中,收集器单元204和分发器单元205由不同的标识符控制。在其它实施例中,收集器单元204和分发器单元205的控制是相互联系的。在这种实施例中,收集器单元204和分发器单元205由同一个标识符控制。

[0083] 4.0链路聚合控制协议示例

[0084] 为了清楚示出示例的目的,将会参照LACP描述实施细节。然而,上述技术不限于LACP。因此,本节仅作为一个示例,展示上面讨论的技术如何被用于解决LACP面对的特定问题。

[0085] LACP,如上所述,是在携带数据链路层的MAC帧的链路中完成链路聚合的聚合协议。LACP的规范详细介绍了在链路上控制收集和分发的两种不同技术。

[0086] 第一种技术,被称为独立控制技术,维持收集状态和分发状态。当参与者进入收集状态,该参与者在链路上使能收集,而当参与者进入分发状态,该参与者在链路上使能分发。收集状态先被进入,一旦链路另一端的远程参与者发送指示远程参与者也已转变到收集状态的消息,(消息的形式为链路聚合控制协议数据单元(LACPDU)),随后就进入分发状态。因此,当一个参与者进入分发状态并开始通过链路传输,远程参与者也已进入收集状态并准备接受传输。

[0087] 然而,许多系统,通常包括旧有系统,没有被配置为通过链路提供精细粒度的控制。其结果是,LACP规范详细介绍了第二种技术,被称为耦合控制技术,它维持收集_分发状态。当一个参与者进入收集_分发状态,该参与者在链路上同时使能收集和分发。因此,不同于独立控制的情况,不存在保障措施确保两个参与者在链路上分发数据出现之前已准备好收集数据。

[0088] 此外,在耦合控制和独立控制这两种情况下,LACP都携带固有风险,也即当一个参与者离开耦合控制的收集_分发状态或独立控制的收集状态,由于缺乏链路同步,链路另一端的远程参与者仍然传输数据。

[0089] 在一种方法中,LACP的规范描述了一个可选的“标记议定书”以在携带转变发生之前清除链路。然而,标记议定书依赖于在链路上独立控制收集和分发的能力,不能解决参与者采用耦合控制的情况。其结果是,实施例可能实施LACP的修改版,在进入或离开收集_分发状态之前执行上述链路同步技术。注意到这种情况仅用于示出对LACP现有状态的可能的

改善,不限于两个参与者都实施耦合控制的情况。因此,其它的实施例可以采用修改版的LACP,在进入或离开独立控制的分发状态之前采用上述链路同步技术。

[0090] 5.0执行机制-硬件概述

[0091] 根据一个实施例,本文描述的技术通过一个或多个专用计算设备实施。专用计算设备可以是硬连线的以执行所述技术,或者可以包括数字电子设备例如一个或多个专用集成电路(ASIC),或者被永久编程为完成所述技术的现场可编程门阵列(FPGA),或者可以包括被编程为依照固件、存储器、其它存储设备或它们的组合上的程序指令执行该技术的一个或多个通用硬件处理器。这种专用计算设备也可以结合定制硬连线逻辑,ASIC,或自定义编程的FPGA来完成该技术。专用计算设备可以是台式计算机系统、便携式计算机系统、手持设备、网络设备或采用硬连线和/或程序逻辑来实现该技术的任何其它设备。

[0092] 例如,图5是示出了在其上本发明的实施例可以被实施的计算机系统500的方框图。计算机系统500包括总线502或者其它传达信息的通信机构,以及和耦接到总线502、用于处理信息的硬件处理器504。硬件处理器504可以是,例如,通用微处理器。

[0093] 计算机系统500还包括主存储器506,例如随机存取存储器(RAM)或者其它动态存储设备,主存储器与总线502相连,用于存储将由处理器504执行的信息和指令。主存储器506也可以被用于存储临时变量或者其它由处理器504执行的指令的执行过程中的中间信息。这些指令,当存储在可访问处理器504的非暂态存储媒介上,致使计算机系统500成为专用机器,被定制为执行指令中指定的操作。

[0094] 计算机系统500还包括只读存储器(ROM) 508或其它耦接到总线502的用于存储处理器504的静态信息和指令的静态存储设备。存储设备510,如磁盘或光盘,被提供并被耦接到总线502用于存储信息和指令。

[0095] 计算机系统500可以通过总线502耦接到显示屏512,例如阴极射线介质管(CRT),用于给计算机用户显示信息。包括字母数字和其它键的输入设备514,被耦接到总线502以传送信息和指令集到处理器504。另一种类型的用户输入设备是光标控制516,例如鼠标、轨迹球或光标方向键,用于传递方向信息和指令集给处理器504以及控制光标在显示屏512上的移动。这种输入设备通常在两个轴(例如第一轴(例如,x)和第二轴(例如,y))上具有允许设备指定平面中的位置的两个自由度。

[0096] 计算机设备500可以实施本文描述的技术,通过使用定制硬连线逻辑、一个或多个ASIC或FPGA、固件和/或可编程逻辑和计算机系统结合,导致或者对计算机系统编程使之成为一个专用机器。根据一个实施例,本文的技术是通过计算机系统500响应执行包含在主存储器506内的一个或多个指令的一个或多个序列的处理器504实现。这种指令可以从另一个存储介质,例如存储设备510,被读取到主存储器506。包含在主存储器506内的指令序列的执行导致处理器504执行本文描述的处理步骤。在另一个实施例中,硬连线电路可以代替软件指令或与软件指令结合使用。

[0097] 本文所用术语“存储介质”指的是存储数据和/或指令导致机器以特定方式操作的任何非暂态介质。这种存储介质可以包括非易失性介质和/或易失性介质。非易失性介质包括,例如光盘和磁盘,例如存储设备510。易失性介质包括动态存储器,例如主存储器506。存储介质的常见形式包括,例如,软盘,软磁盘,硬盘,固态硬盘,磁带,或任何其它磁性数据存储介质,CD-ROM,任何其它光学数据存储介质,具有孔的任何物理介质,RAM,PROM和EPROM,

FLASH-EPROM, NVRAM, 任何其它存储器芯片或盒式磁带。

[0098] 存储介质不同于传输介质但是可以和传输介质联合起来使用。传输介质参与存储介质之间的信息传输。例如, 传输介质包括同轴电缆, 铜线和光纤 (包括构成总线502的导线)。传输介质还可以采用声波或光波的形式, 例如在无线电波和红外数据通信期间生成的声波或光波。

[0099] 携带一个或多个指令的一个或多个序列给处理器504执行时涉及各种形式的介质。例如, 指令最初可能被携带在远程计算机的磁盘或固态硬盘上, 远程计算机可以将指令加载到它的动态存储器上并且使用调制解调器通过电话线发送指令。计算机系统500本地的调制解调器可以接收电话线上的数据并且使用红外发射机将数据转变成红外信号。红外检测器可以接收红外线信号携带的数据, 适当的电路将数据放置到总线502上。总线502将数据携带到主存储器506, 处理器504从主存储器获取并执行指令。主存储器506接收的指令被处理器504执行之前或之后可以可选地存储在存储设备510上。

[0100] 计算机系统500还包括耦接到总线502的通信接口518。通信接口518和连接到本地网络522的网络链路520耦接, 提供双向数据通信。例如, 通信接口518可以是综合业务数字网 (ISDN) 卡, 电缆调制解调器, 卫星调制解调器, 或给相应类型的电话线提供数据通信的调制解调器。另一个例子是, 通信接口可以是局域网 (LAN) 卡, 给兼容的LAN提供数据通信连接。还可以采用无线链路。在任何这样的实现中, 通信接口518发送和接收电, 电磁或光信号, 携带代表各种类型信息的数字数据流。

[0101] 网络链路520通常通过一个或多个网络给其它数据设备提供数据通信。例如, 网络链路520可以通过本地网络522提供与由互联网服务提供商 (ISP) 操作的主机524或数据设备的连接。ISP526反过来通过现在通常被称为“因特网”528的全球分组数据通信网络提供数据通信服务。本地网络522和因特网528都使用携带数字数据流的电, 电磁或光信号。携带来自和发至计算机系统500的数字数据的通过各种网络的信号和在网络链路520上以及通过通信接口518的信号是传输介质的示例形式。

[0102] 计算机系统500可以通过网络, 网络链路520和通信接口518发送消息和接收数据, 包括程序代码。在因特网示例中, 服务器530可以通过因特网528, ISP526, 局域网522和通信接口518给应用程序发送请求的代码。

[0103] 接收到的代码可以在接收到时由处理器504执行, 和/或被存储在存储设备510中, 或其它非易失性存储器上供以后执行。

[0104] 6.0扩展和备选

[0105] 在上述说明中, 参照实施方式不一样的多个具体细节对本发明的实施例进行了描述。因此, 关于什么是本发明的唯一且排他的指标, 以及申请人对什么是本发明的预期, 是本申请中的一组权利要求, 这些权利要求以特定的形式给出, 包括任何后续更正。对权利要求中包含的术语在本文中的任何详细的定义应规定有关术语在权利要求中使用时的涵义。因此, 没有在权利要去中明确叙述的限制、元件属性、特征、优势或属性不应以任何形式限制权利要求的范围。本说明书和附图, 相应的, 也可以被认为是说明性的而非限制性的。

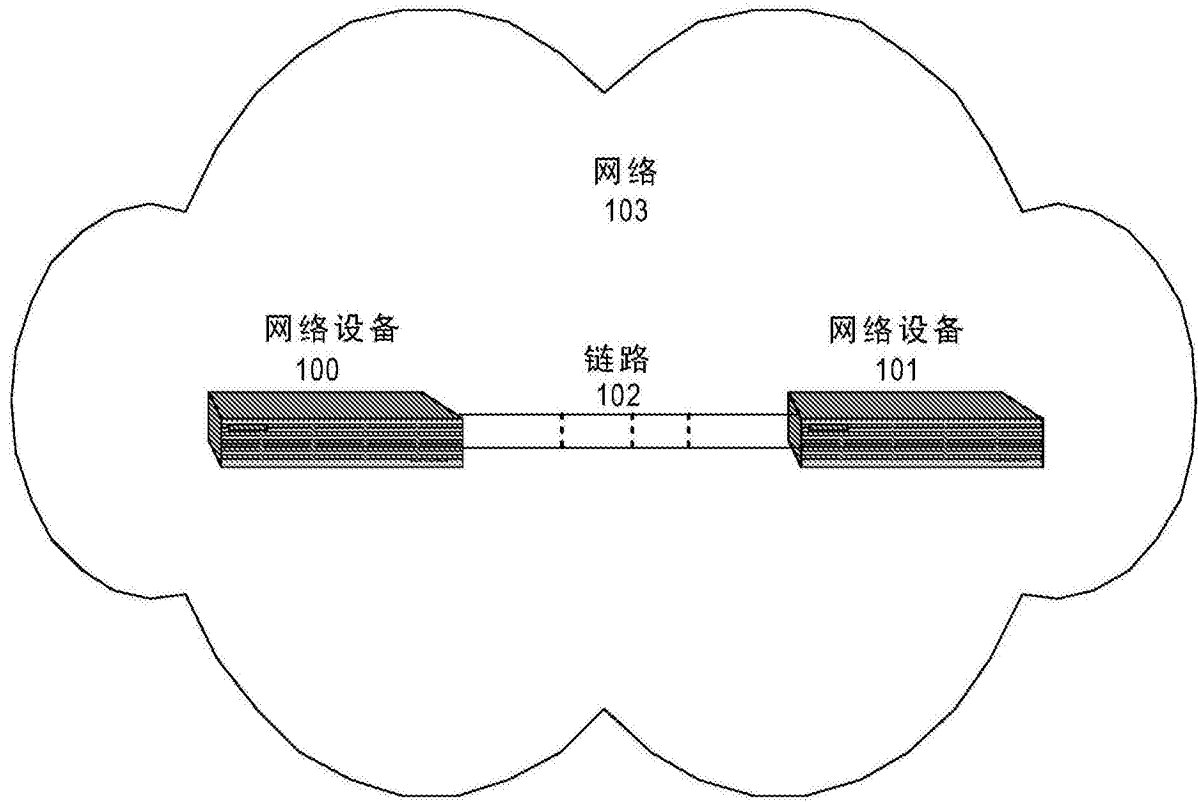


图1

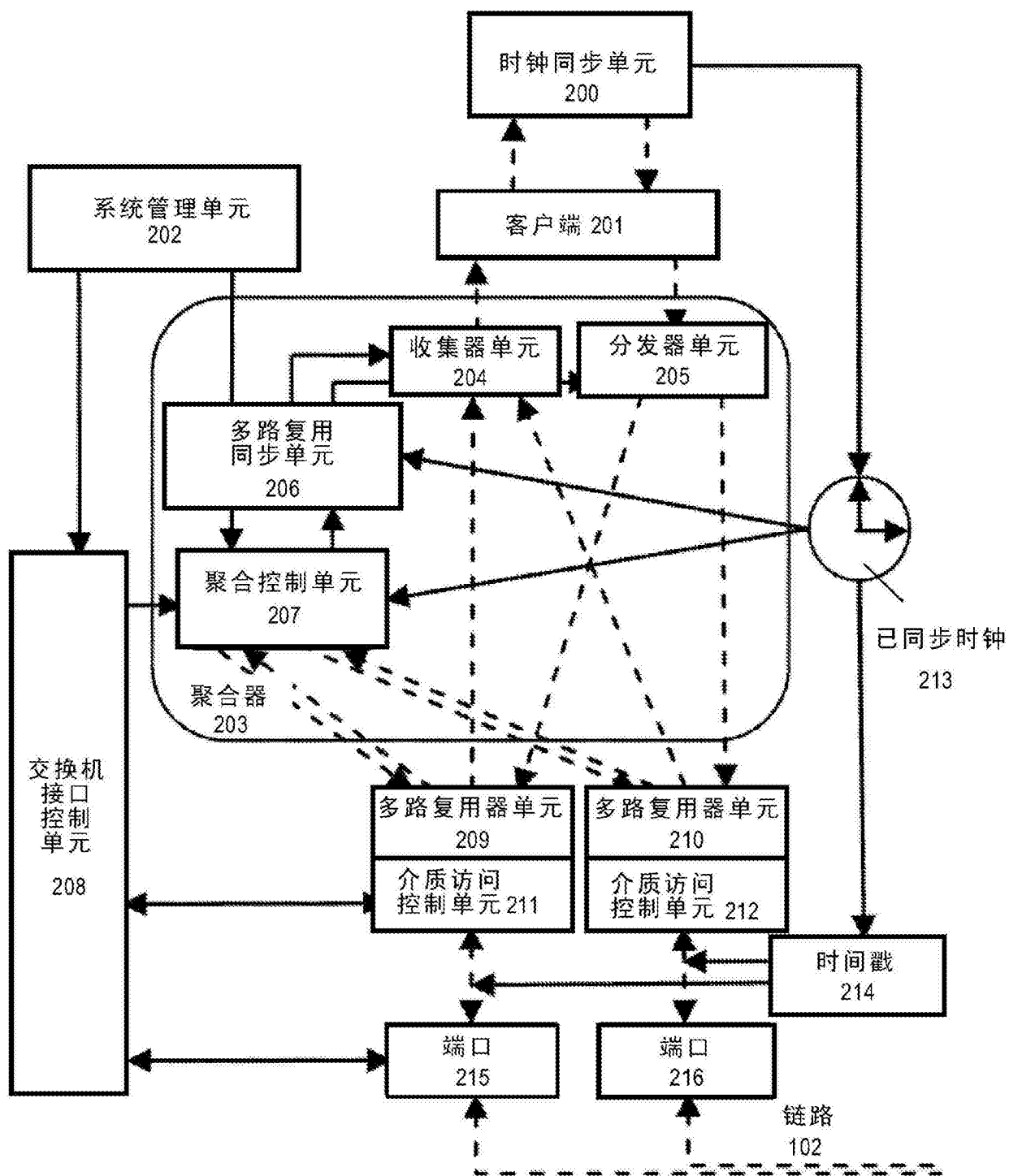


图2

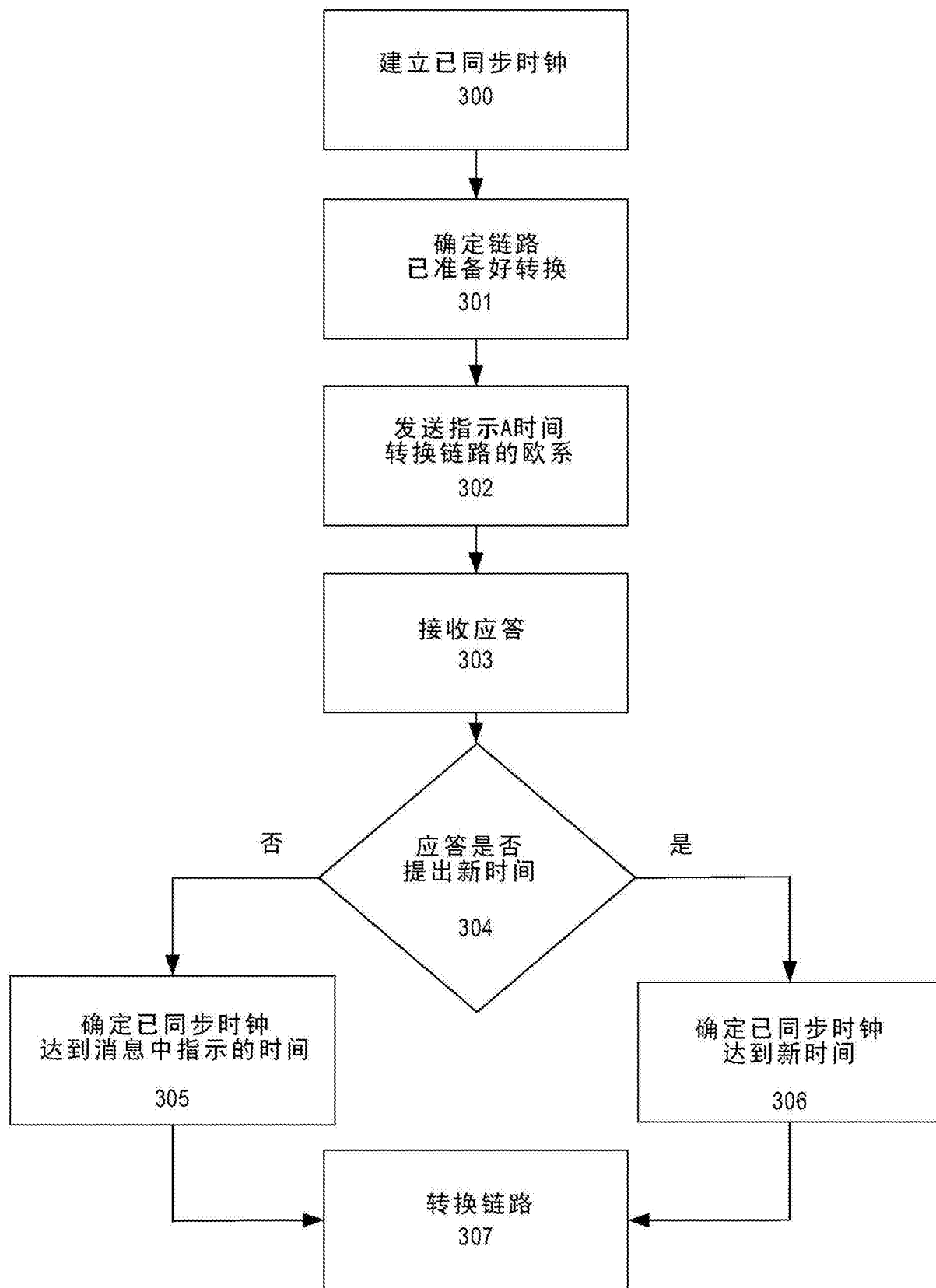


图3

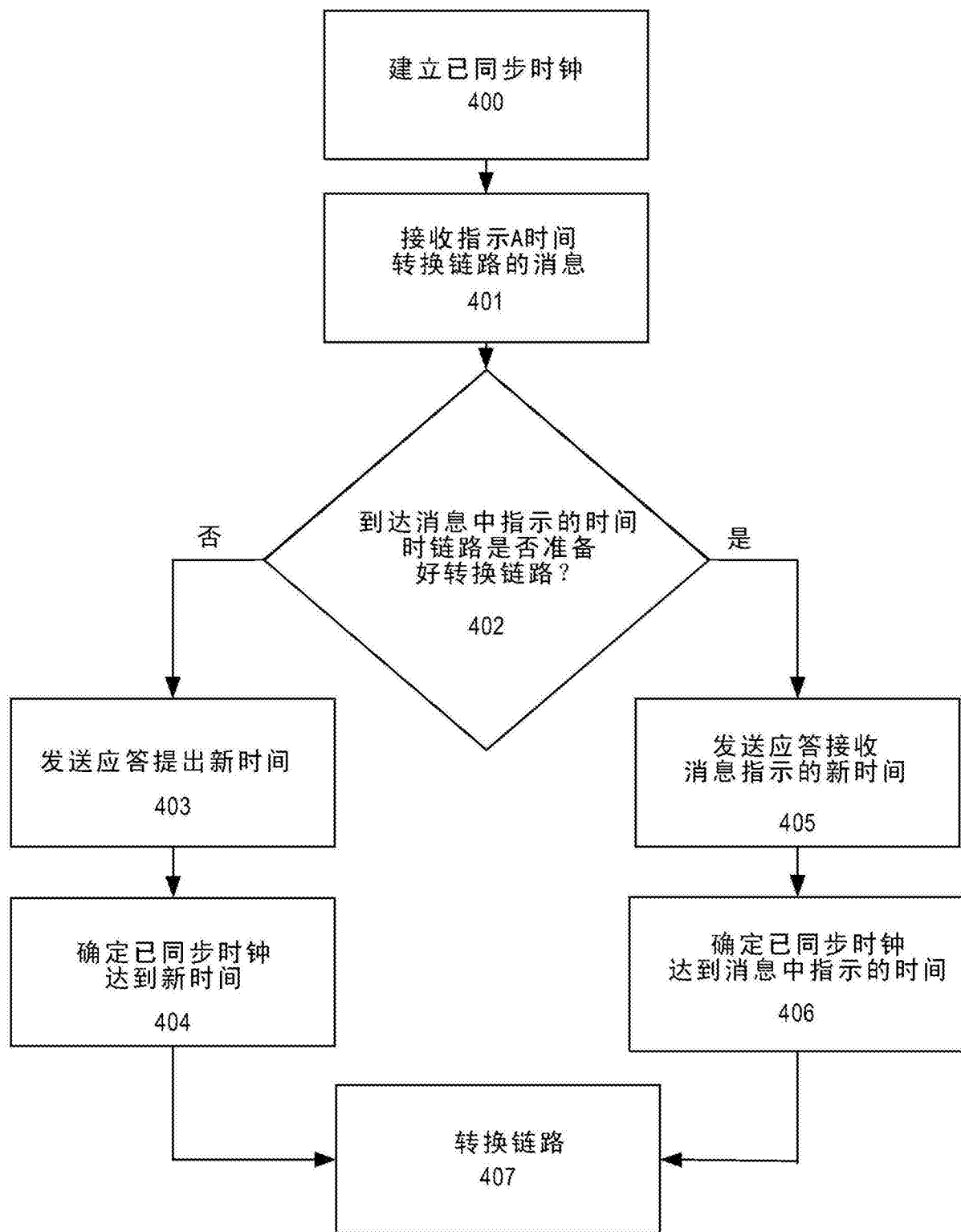


图4

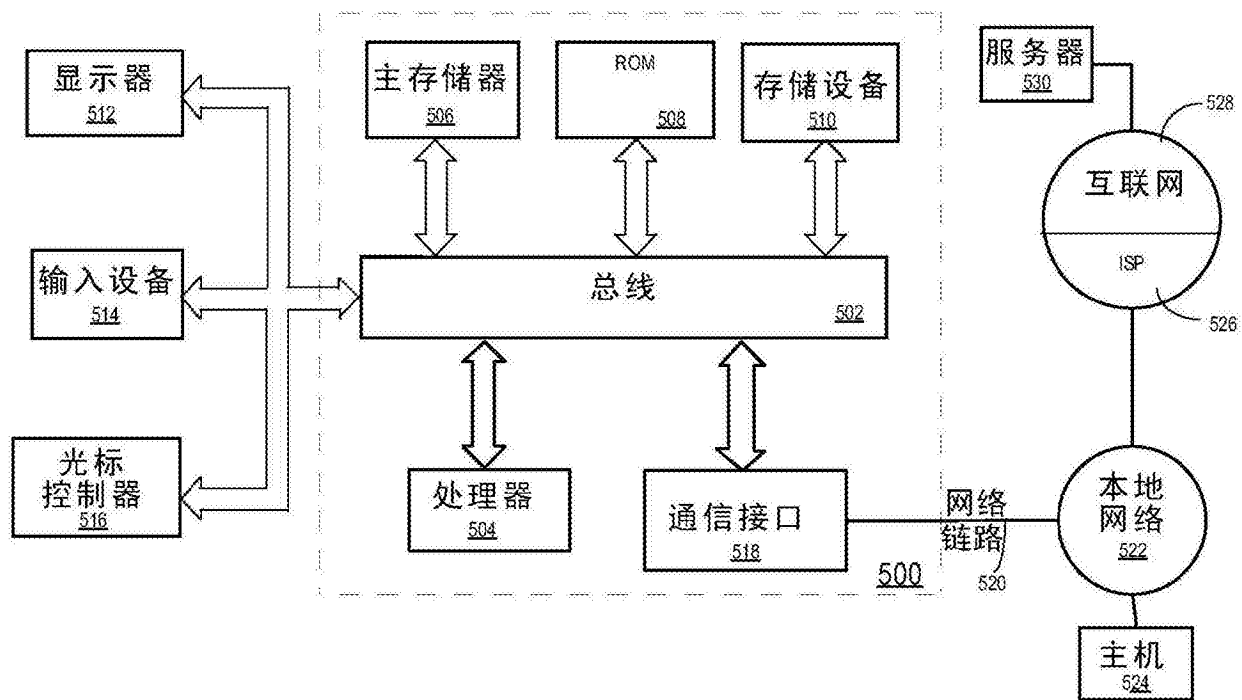


图5