(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 105579991 A
(43)申请公布日 2016.05.11

(21)申请号 201380079749.3
(22)申请日 2013.07.23
(85)PCT国际申请进入国家阶段日
2016.03.23
(86)PCT国际申请的申请数据
PCT/US2013/051674 2013.07.23
(87)PCT国际申请的公布数据
WO2015/012811 EN 2015.01.29
(71)申请人 慧与发展有限责任合伙企业
地址 美国德克萨斯州
(72)发明人 巴威·亚拉甘杜拉 卢奇安·波保 苏亚塔·班纳吉
(74)专利代理机构 北京德琦知识产权代理有限公司 11018
代理人 郭艳芳 康泉
(51)Int.Cl.
G06F 15/16(2006.01)
G06F 11/30(2006.01)

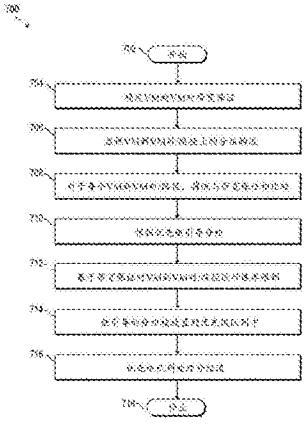
权利要求书2页 说明书12页 附图6页

(54)发明名称

使用优先级进行工作保持的带宽保证

(57)摘要

示例实施例涉及使用优先级进行工作保持的带宽保证。在一些示例中，方法可以包括确定源虚拟机(VM)和至少一个目的地VM之间的VM到VM带宽保证，包括源VM和特定目的地VM之间的特定VM到VM带宽保证。该方法可以包括监视从源VM到特定目的地VM的出站网络业务流。该方法可以包括将出站网络业务流与特定VM到VM带宽保证相比较。当出站网络业务流小于特定VM到VM带宽保证时，可以根据第一优先级来引导流的分组。当出站网络业务流大于特定VM到VM带宽保证时，可以根据第二优先级来引导流的分组。



CN 105579991 A

1.一种使用优先级进行工作保持的带宽保证的方法,所述方法包括:

确定源虚拟机(VM)和至少一个目的地VM之间的VM到VM带宽保证,所述VM到VM带宽保证包括所述源VM和特定目的地VM之间的特定VM到VM带宽保证;

监视从所述源VM到所述特定目的地VM的出站网络业务流;

将所述出站网络业务流与所述特定VM到VM带宽保证相比较;

当所述出站网络业务流小于所述特定VM到VM带宽保证时,根据第一优先级引导所述流的分组;以及

当所述出站网络业务流大于所述特定VM到VM带宽保证时,根据第二优先级引导所述流的分组。

2.根据权利要求1所述的方法,其中引导所述流的分组使所述第一优先级的分组以高达所述特定VM到VM带宽保证的速率行进,并且使所述第二优先级的分组仅仅当在用于连接所述源VM和所述至少一个目的地VM的网络中存在备用带宽时行进。

3.根据权利要求1所述的方法,其中所述源VM由第一计算设备托管,并且所述特定目的地VM由第二计算设备托管,并且其中硬件交换机连接所述第一计算设备与所述第二计算设备。

4.根据权利要求3所述的方法,其中根据所述第一优先级引导所述流的分组使所述分组被放置到所述硬件交换机的第一优先级队列中,并且其中根据所述第二优先级引导所述流的分组使所述分组被放置到所述硬件交换机的第二优先级队列中。

5.根据权利要求4所述的方法,其中所述源VM和所述至少一个目的地VM跨网络彼此进行通信,并且其中根据所述第一优先级或所述第二优先级引导所述流的分组允许使用超过所述VM到VM带宽保证的备用带宽容量。

6.根据权利要求3所述的方法,其中所述监视、所述比较和所述引导发生在所述第一计算设备的管理程序中。

7.根据权利要求3所述的方法,其中所述监视、所述比较和所述引导发生在所述源VM中,并且其中所述源VM使用多路径传送协议,其中第一虚拟路径与所述第一优先级相关联,第二虚拟路径与所述第二优先级相关联。

8.根据权利要求7所述的方法,其中所述源VM使用第一网络接口或地址来根据所述第一优先级引导所述出站网络业务流的分组,并且使用第二网络接口或地址来根据所述第二优先级引导所述出站网络业务流的分组。

9.根据权利要求2所述的方法,进一步包括:

分析小于所述特定VM到VM带宽保证的出站网络业务流,以确定所述流来自高级租户或应用;以及

进一步根据高级优先级引导所述流的分组,其中利用所述高级优先级引导所述流的分组使那些分组相比于与所述第一优先级相关联的而不是与所述高级优先级相关联的分组以较低的等待时间行进。

10.根据权利要求2所述的方法,进一步包括:

分析大于所述特定VM到VM带宽保证的所述出站网络业务流,以确定用于传送所述流的传送协议是否对网络拥塞作出响应;以及

当所述传送协议未对网络拥塞作出响应时,进一步根据第三优先级引导所述流的分

组，其中利用所述第三优先级引导所述流的分组使那些分组仅仅在仅与所述第二优先级相关联的分组之后行进。

11.一种使用优先级进行工作保持的带宽保证的主机计算设备，所述主机计算设备包括：

至少一个处理器，用于：

确定源虚拟机(VM)和至少一个目的地VM之间的VM对带宽保证，所述VM对带宽保证包括所述源VM和特定目的地VM之间的特定VM对带宽保证；

监视从所述源VM到所述特定目的地VM的出站网络业务流；

将所述出站网络业务流与所述特定VM对带宽保证相比较；以及

当所述出站网络业务流小于所述特定VM对带宽保证时，将所述流的分组作为有保证的业务来引导，否则将所述流的分组作为有工作保持的业务来引导。

12.根据权利要求11所述的主机计算设备，其中所述至少一个处理器进一步根据所述特定VM对带宽保证对所述有保证的业务进行速率限制。

13.根据权利要求11所述的主机计算设备，其中为了将所述流的分组作为有保证的业务来引导，所述源VM通过第一虚拟路径路由此类分组，否则通过第二虚拟路径路由所述流的分组，其中所述第一虚拟路径和所述第二虚拟路径使用多路径传送协议来实施。

14.一种机器可读存储介质，所述机器可读存储介质利用主机计算设备的至少一个处理器可执行的、以使用优先级进行工作保持的带宽保证的指令来编码，所述机器可读存储介质包括：

用于对于网络中的多个虚拟机(VM)之间的网络业务流分别确定VM对带宽保证的指令；

用于监视所述多个VM之间的网络业务流的指令；

用于将所述网络业务流与相应的VM对带宽保证相比较的指令；以及

用于在所述网络中的至少一个网络业务流低于其相应的带宽保证时、引导所述网络业务流的分组以使低于其相应的带宽保证的网络业务流的部分优先化并且允许网络业务流的剩余部分穿过所述网络的指令。

15.根据权利要求14所述的机器可读存储介质，其中引导所述网络业务流的分组使经优先化的部分的分组被放置到至少一个硬件交换机的至少一个高优先级队列中，并且使所述剩余部分的分组被放置到所述至少一个硬件交换机的至少一个低优先级队列中。

使用优先级进行工作保持的带宽保证

背景技术

[0001] 在云计算环境中，数据中心可以包括多个联网的计算设备。一些或所有计算设备(即，主机)可以均运行许多虚拟机(VM)。租户可以保留一个或多个VM，例如用于执行应用。租户可以是拥有(例如，支付)使用VM的权利的用户、机构，等等。特定租户的VM可以被合并在单个计算设备/主机上或者跨数据中心中的多个计算设备/主机分布。租户可以将其VM配置为在它们自己之间进行通信，以例如用于提供租户的统一的服务。租户的VM(多个)也可以与其他租户的VM进行通信。租户可能希望了解其VM之间的以及与其他租户的VM的最小通信规范(例如，带宽保证)，使得租户可以确定关于租户的应用和/或服务的性能的下界。

附图说明

[0002] 以下详细描述参考附图，其中：

[0003] 图1是使用优先级实施工作保持的带宽保证的示例性网络设定的框图；

[0004] 图2是使用优先级实施工作保持的带宽保证的示例性网络设定的框图；

[0005] 图3是用于使用优先级实施工作保持的带宽保证的示例性流管理器模块的框图；

[0006] 图4是流管理器模块可以根据优先级来引导信息的分组的示例性方案的流程图；

[0007] 图5是流管理器模块可以根据优先级来引导信息的分组的示例性方案的流程图；

[0008] 图6是流管理器模块可以根据优先级来引导信息的分组的示例性方案的流程图；

[0009] 图7是使用优先级进行工作保持的带宽保证的示例性方法的流程图；

[0010] 图8是使用优先级进行工作保持的带宽保证的主机计算设备的框图；以及

[0011] 图9是使用优先级进行工作保持的带宽保证的示例性方法的流程图。

具体实施方式

[0012] 如上所述，租户可能希望了解其VM之间的最小通信规范(例如，带宽保证)，使得租户可以确定关于租户的应用和/或服务的性能的下界。各个云提供商不能准确地提供此类带宽保证。此类带宽保证信息的缺乏阻止租户确定关于应用和/或服务性能的下界。此类带宽保证信息的缺乏也可能阻止租户将企业应用传递到云，这是因为许多企业应用要求可预测的性能保证。

[0013] 可以提供带宽保证的各种解决方案也浪费带宽容量。例如，如果第一租户在特定时段期间未使用其全部带宽保证，那么备用带宽容量被浪费掉，例如，具有高需求的第二租户不能暂时地使用第一租户的备用带宽。可以尝试供应备用带宽容量的各种其他解决方案在完全地利用备用带宽方面是无效的。例如，当允许VM以比其最小带宽保证高的速率发送网络业务时，一些解决方案是保守的，例如以便避免例如在硬件交换机的由于拥塞的分组丢弃。该保守的行为导致网络带宽的浪费。

[0014] 各种其他解决方案可以通过利用多个队列来在网络中供应备用带宽。然而，这些解决方案的目的不在于提供工作保持的带宽保证，其中，通过利用网络中的备用带宽——例如，可用的备用带宽，用于两个特定VM之间的通信的带宽可能超过与这两个VM相关联的

最小带宽保证，这是因为网络中的其他VM不使用通过它们的带宽保证分配给它们的所有带宽。代之以，这些解决方案目的仅仅是对于不同的租户、应用、服务、不同类型的分组等等实施不同的服务水平或优先级。例如，高优先级队列可以处理用于第一租户、应用、服务、第一类型的分组等等的业务，并且较低的优先级队列可以处理用于第二租户、应用、服务等等的业务。对于这些解决方案，特定租户、应用、服务、特定类型的分组等等与特定硬件队列静态地相关联。另外地，对于这些解决方案，预先指定租户、应用、服务、分组的类型等等具有特定重要性、优先级或服务水平，并且该预先指定确定哪个队列用于租户、应用、服务、分组的类型，等等。

[0015] 本公开描述使用优先级进行工作保持的带宽保证。本公开描述一种解决方案，例如当其他租户和/或VM-VM对未使用它们的全部带宽保证时，允许(例如，特定VM-VM对的)租户以与特定带宽保证相比较高的速率来发送业务。该解决方案可以例如在连接多个主机计算设备的至少一个硬件交换机中、使用两个或更多优先级队列来实施。该解决方案也可以使用例如位于主机计算设备的管理程序中的或虚拟机中的至少一个流管理器模块。流管理器模块可以监视从源VM到目的地VM的出站网络业务流；并且对于特定VM到VM对，可以将流与用于VM到VM对的带宽保证相比较。流管理器模块可以基于流是否大于带宽保证来根据第一优先级或第二优先级对流的分组进行引导(例如，路由和/或加标签)。根据第一优先级或第二优先级引导流的分组可以使分组被放置到例如硬件交换机的第一优先级队列或第二优先级队列中。因而，可以将遵守带宽保证的业务指配给较高的优先级队列，而可以将超过带宽保证的“机会性业务”(即，工作保持的业务)指配给较低的优先级队列。以这种方式，当具有带宽保证的第一VM对上的业务未完全地利用保证时，第二VM对可以机会性地利用该备用带宽来发送超过第二对的带宽保证的业务。

[0016] 相比各种其他解决方案，本公开可以提供益处。首先，本公开为租户提供VM之间的最小带宽保证，这允许租户确定关于应用和/或服务性能的下界，并且也可以允许租户将企业应用传送到云。另外地，与浪费带宽容量的各种其他解决方案相比，本公开可以更高效地供应网络中的备用带宽。本公开也可以利用硬件支持(例如，硬件交换机中的优先级队列)，这可以简化在主机计算设备中实施的解决方案。因此，主机计算设备可以实施可以导致主机中的较低的开销以及较小的CPU利用率的较不复杂的算法。再者，本公开可以利用商品(例如，现成的)交换机，这使得解决方案实用并且廉价。

[0017] 贯穿本公开，术语“工作保持”、“工作保持的”等等可以指的是利用网络中的备用带宽容量以便尽可能少地浪费带宽容量的目标。术语“链路”可以指的是两个计算机组件之间的单个连接或单个通信组件(例如，电缆、接线、直接的无线连接，等等)。术语(例如，如在网络路径中的)“路径”可以指的是信息可以在网络中穿过的路线(例如，物理路线)，其中路径可以例如从源到目的地经过多个连接、网络接口、集线器、路由器、交换机等等。术语“虚拟路径”可以指的是穿过物理路径的逻辑路线，其中，其他逻辑路线穿过相同的物理路径。(例如，如在VM到VM对中的)术语“对”可以指的是两个计算模块(例如，两个虚拟机)之间的通信。例如，特定VM到VM对可以指的是从一个计算模块行进到另一个计算模块的所有数据。在一些实例中，(例如，从第一模块到第二模块)在一个方向上行进的数据可以被考虑为与(例如，从第二模块到第一模块)在相反方向上行进的数据不同的对。对的数据可以跨一个路径或多个路径行进，并且路径(多个)可以随时间而改变。因而，特定VM到VM对在网络中可

以是唯一的;然而,多个路径可以携带对的数据。同样地,特定路径可以携带多个VM到VM对的数据。在一些情形中,对可以用作指VM和VM的聚集之间的通信或者VM和租户之间的通信的通用术语。(如在网络流中的)术语“流”可以指的是在网络中在源与目的地之间的分组的聚合和/或序列。流的示例包括两个VM之间的传输控制协议(“TCP”)分组、两个VM之间的用户数据报协议(“UDP”)分组或者在源与目的地之间所传送的分组的任何其他聚合或聚集。在一些情形中,特定流可以与特定目的(例如,特定应用、服务、模块,等等)有关。术语“业务”可以指的是穿过网络中的点(例如,网络接口、集线器、路由器、交换机等等)的分组或流,其中分组或流可以从多于一个源行进和/或行进到多于一个目的地。

[0018] 图1是使用优先级实施工作保持的带宽保证的示例性网络设定100的框图。网络设定100可以包括许多主机计算设备(例如,102、104、106)以及至少一个硬件交换机(例如,交换机108)。主机计算设备102、104、106可以例如经由以太网接线或一些其他有线或无线网络链路与硬件交换机108进行通信。主机计算设备102、104、106可以通过将信息的分组传送到交换机108并且从交换机108接收信息的分组来与彼此进行通信。主机计算设备102、104、106可以均包括允许主机计算设备经由网络(例如,具有交换机108)进行通信的网络接口卡(NIC)(例如,118)。

[0019] 主机计算设备(例如,102、104、106)均可以是能够经由网络与其他计算设备进行通信并且能够运行虚拟机的任何计算系统或计算设备。应当理解的是,尽管一些主机计算设备(例如,106)可能未示出所有它们的内部组件,但它们可以包括与示出的其他主机计算设备(例如,102、104)类似的组件。主机计算设备(例如,102)均可以包括许多虚拟机(VM)(例如,110、112、114)。如上所述,租户可以保留一个或多个VM,例如用于执行应用。特定租户的VM可以被合并在单个主机计算设备上或者跨多个主机计算设备分布。主机计算设备(例如,102)均可以包括管理程序(例如,116)。术语“管理程序”可以指的是例示、运行和/或管理虚拟机的一款计算机软件、固件或硬件。管理程序(例如,116)可以为VM(例如,110、112、114)的操作系统呈现虚拟操作平台(例如,虚拟化硬件资源)。管理程序也可以管理VM操作系统的执行。术语“管理程序”通常也可以包含处于主机操作系统的控制下并且用于运行太复杂而不能在主机操作系统本身中运行的计算机软件的特殊VM。在图1的示例中,管理程序(例如,116)均可以包括保证确定和速率限制器模块122以及流管理器模块124。这些模块(例如,122、124)可以包括在机器可读存储介质上编码的并且可由主机计算设备(例如,102)的处理器执行的一系列指令。另外或作为替代,这些模块可以包括一个或多个硬件设备,该一个或多个硬件设备包括用于实施在本文描述的功能的电子电路。

[0020] 在本文的各种描述中,例如可以对至少一个源VM(例如,第一VM)和至少一个目的地VM(例如,第二VM)进行参考,其中,源VM和目的地VM形成特定VM到VM对。作为一个示例,并且参考图1,VM 110可以是源VM,并且VM 105可以是目的地VM。可以对同一租户保留VM到VM对的源VM和目的地VM两者,或者可以由不同的租户来保留VM到VM对的源VM和目的地VM。包括VM 110到VM 105的VM对可以与特定最小带宽保证相关联。此类带宽保证可以由保证确定和速率限制器模块122例如基于出自VM 110的最小发送保证带宽和到VM 105中的最小接收保证带宽来(至少部分地)确定。此类最小发送和接收保证带宽可以进一步基于与每个VM相关联的通用保证带宽。在下面可以关于如何确定各种带宽保证提供更多详情。

[0021] 交换机108可以是能够连接多个计算设备、使得计算设备能够在网络中进行通信

的任何硬件交换机。交换机108可以是商品(例如,现成的)交换机或定制交换机。交换机108可以包括电路、固件、软件等等,以实施许多优先级队列120(例如,如在图1中描绘的Q0和Q1)。交换机108可以被配置为建立优先级队列(例如,一定数量的优先级队列及其优先级和优先级行为),并且可以被配置为基于各条信息(例如,分组的标签或报头、多路径TCP连接的特定路径或分组在其上行进的其他多路径传送连接,等等)将进入交换机的每个分组路由到优先级队列之一中。优先级队列及其优先级行为可以行动以调节业务(例如,分组)在业务进入交换机之后并且在其离开交换机之前的内部运动,这可以进而影响业务到下游的设备(例如,主机计算设备104和VM 105)的到达。在一些示例中,交换机108可以包括许多速率限制器,其可以调节业务可以进入和/或离开各种优先级队列120的速率。也可以配置速率限制器。在一些示例中,例如,如果主机计算设备中的模块(例如,122)执行足够的速率限制,则在交换机108中速率限制器可能不是必需的。

[0022] 在图1的示例中,交换机108中的优先级队列(例如,优先级队列120)包括Q0队列和Q1队列。Q0可以是低于(例如,没有超过)特定带宽保证(例如,用于从VM 110到VM 105的VM到VM对的带宽保证)的业务的较高的优先级队列。Q1可以是用于机会性(即,工作保持)业务的较低的优先级队列。在其他的示例中,交换机108可以包括多于一个高优先级队列和/或多于一个较低的优先级队列。可以在下面更详细地描述关于这两种类型的队列的更多详情以及可以如何将业务路由到这些队列。

[0023] 为了使主机计算设备(例如,102)能够向交换机108发送在交换机中可以被适当地路由到优先级队列的分组,可以配置主机计算设备中的NIC 118(例如,支持标记的分组的设置、多路径TCP,等等)。另外地,NIC 118也可以实施与交换机108中的优先级队列类似的优先级队列和/或优先级队列设置,例如用于保证有保证的业务通过从VM到VM的整个路径而不受较低优先级机会性业务的影响。在本文的各种描述可以主要描述硬件交换机中的优先级排队,但是应当理解,例如,作为对硬件交换机中的优先级排队的补充或者更换,可以将各种描述扩展为包括NIC中的类似的优先级排队。

[0024] 应当理解的是,图1仅仅示出一个示例性网络设定,其中经由单个交换机108将主机计算设备联网。在其他示例中,网络设定可以包括多个交换机,并且从一个主机计算设备中的一个VM到另一个主机计算设备中的VM的网络路径可以经过多个交换机——例如,级联的交换机,交换机的分级体系,等等。在此类网络设定中,路径中的一个或多个交换机可以实施与在本文描述的优先级排队类似的优先级排队。也应当理解的是,通过各种技术,租户可以接收用于同一主机计算设备内的VM到VM对以及不同的主机计算设备之间的VM到VM路径两者的VM到VM带宽保证。本公开主要地集中于不同的主机计算设备之间的VM到VM对,例如,经过至少一个硬件交换机(例如,108)的对。再者,在本文描述的技术和解决方案可以用于单个主机计算设备内的VM到VM对,例如被实施在虚拟网络接口、虚拟交换机,等等内。

[0025] 保证确定&速率限制器模块122可以确定网络设定(例如,100)中的各个VM(例如,110、112、114,主机计算设备104的VM,主机计算设备106的VM,等等)之间的VM到VM对带宽保证(例如,发送和接收)。可以假定,(例如,由一主机管理员、主机管理器,等等)为每个VM指配最小发送保证带宽和最小接收保证带宽。可以基于最大服务器占用率、最坏情况网络利用率和类似的度量来计算此类保证。模块122可以通过将用于第一VM(例如,110)的最小发送保证带宽划分或分割为用于与第一VM进行通信的每一个VM(包括第二VM(例如,105))的

初级VM到VM发送保证，来确定从第一VM到第二VM的发送VM到VM保证。可以对于第二VM关于第二VM的最小接收保证带宽来进行相同的划分或分割。那么，第一VM(例如，110)和第二VM(例如，105)之间的最小发送保证带宽可以是从第一VM到第二VM的第一VM的初级发送保证和从第一VM到第二VM的第二VM的初级接收保证中的较小的(即，最小的)。在一些示例中，一个主机计算设备(例如，104)中的保证确定和速率限制器模块(例如，125)可以与另一个主机计算设备(例如，102)中的类似模块(例如，122)进行通信，以共享最小带宽保证信息。可以在网络设定的任何数量的VM之间对于对(发送和接收两者)进行类似的最小带宽保证确定。

[0026] 如上所述的其中VM被指配最小发送和接收带宽保证并且其中将此类保证划分和分割为VM到VM保证的模型可以被称为用于带宽保证的软管模型。软管模型是保证确定&速率限制器模块122可以如何确定VM到VM带宽保证的一个示例。在其他的示例中，模块122可以使用其他带宽保证模型，例如管道模型。管道模型可以已经假定从源VM到各个目的地VM的以及从各个源VM到目的地VM的对保证。那么，模块122可以例如通过确定从第一VM到第二VM的第一VM的假定的发送保证和从第一VM到第二VM的第二VM的假定的接收保证中的较小的(即，最小的)，来确定在第一VM(例如，110)和第二VM(例如，105)之间的最小发送保证带宽。在其他的示例中，模块122可以使用提供带宽保证的任何其他类型的解决方案——例如不是工作保持的解决方案。在又一些其他示例中，模块122可以更换工作保持的解决方案中的负责工作保持(例如，效率低的工作保持)的模块或部分。

[0027] 系统管理员等等可以确保主机计算设备、交换机和其他网络/数据中心组件能够向所有租户、应用等等提供通告的最小带宽保证。例如，准入控制框架必须确保遍历任何一个路径的(例如，用于各个VM到VM对)的带宽保证的和小于该路径的容量。

[0028] 一旦保证确定&速率限制器模块122已经确定了VM到VM对带宽保证，模块122可以例如使用速率限制器来强制执行此类保证。通常，模块122可以确保用于特定VM到VM对或路径的对于高优先级被引导的(例如，路由和/或加标签的)分组的速率(以下被更详细地解释)不超过用于该对或路径的带宽保证(多个)(例如，在VM的对之间以及对于特定路径，业务被限制)。通过对VM中的每对之间的业务进行速率限制，可以满足(例如，管道模型中的)每个VM到VM对或(例如，软管模型中的)通常用于每个VM的(例如，如上所述确定的)最小带宽保证。在一些示例中，在管理程序(例如，116)中——例如在模块122中包括此类速率限制器。(例如，模块122中的)特定速率限制器可以对于特定VM到VM对(例如，在VM 110和VM 105之间)强制执行用于从VM 110到VM 105的通信的带宽保证，并且也可以强制执行用于从VM 105到VM 110的通信的带宽保证。在一些情形中，VM可以执行速率限制，如下所述。

[0029] 通过在主机计算设备中(例如，在模块122中)实施速率限制，在交换机108中可能不需要速率限制器。主机计算设备中的速率限制器可以确保到优先级队列120中的业务遵守带宽保证。例如，到高优先级队列(例如，Q0)中的业务可以受限于用于路由通过包含高优先级队列的交换机的特定VM到VM路径的(例如，用于VM到VM对)的带宽保证的和。那么，交换机中的优先级排队的行为可以允许当已经路由和发送所有高优先级业务时被引导到较低的优先级队列(例如，Q1)中的机会性业务行进或移动。

[0030] 图2是使用优先级实施工作保持的带宽保证的示例性网络设定200的框图。图2和网络设定200在众多方面与图1和网络设定100类似。例如，网络设定200可以包括许多主机

计算设备(例如,202、204、206)和至少一个硬件交换机(例如,交换机208)。交换机208可以包括许多优先级队列220。主机计算设备(例如,202)均可以包括许多虚拟机(VM)(例如,210、212、214)。主机计算设备(例如,202)均可以包括管理程序(例如,216)。管理程序(例如,216)均可以包括保证确定和速率限制器模块222。图2可以在各方面不同于图1。例如,管理程序(例如,216)可以不包括(例如,与124类似的)流管理器模块。代之以,类似的流管理器模块(例如,224、225)可以位于虚拟机(例如,210、212、214、205,等等)中的每一个中。在本文的各种描述通常可以指的是流管理器模块,并且应当理解,除非另外指出,此类描述可以适用于在(例如,如图1中所示的)管理程序中实施的流管理器模块或在(例如,如图2中所示的)VM中实施的流管理器模块。同样地,图1的组件的各种其他描述可以适用于图2的类似的组件,除非另外指示其他。

[0031] 流管理器模块(例如,224)可以使用多路径TCP协议或另一种类型的多路径传送协议。例如,在图2中,运行在VM(例如,210)中的多路径协议可以使用许多虚拟路径,其中每个虚拟路径可以与特定优先级(例如,高优先级和较低优先级)相关联。在该示例中,流管理器模块(例如,224)可以监视用于各个VM到VM对的业务,例如,发送给第二VM(例如,205)、出自第一VM(例如,210)的流。对于特定VM到VM对,流管理器模块224可以基于用于对的流是否已经超过用于该对的带宽保证来将分组路由到不同的虚拟路径中。在图2的示例中,附图标记215可以表示利用两个虚拟路径实施多路径TCP协议的网络连接。替换地,附图标记215可以表示通过管理程序暴露给VM的多个物理路径和/或多个虚拟网络接口,如以下更详细地描述的。应当理解的是,图1的示例(例如,流管理器模块124)也可以使用多路径协议来跨多个虚拟路径(例如,具有保证的一些和不具有保证的一些)分发分组。在该示例中,VM可以仅仅看见单个虚拟网络接口。

[0032] 在一些情形中,两个VM可以使用多个物理路径进行通信,在这种情况下,可以跨多个物理路径路由具有保证的业务(例如,高优先级业务)和工作保持的业务(例如,低优先级业务)两者。在该情形中,可以(例如,由分组分发器)跨所有物理路径散布分组,并且可以将多个物理路径看作单个较大的虚拟路径。然后,多路径传送协议可以看见或暴露一定数量的(例如,两个)虚拟路径。分组分发器模块可以在这些虚拟路径上(以及跨物理路径)扩散分组。替换地,可以(例如,通过多路径协议)将多个物理路径中的每一个看作分离的虚拟路径。在这种情况下,多路径协议可以看见或暴露2*N个虚拟路径,其中N是物理路径的数量。这些路径中的N个路径可以与有保证的业务以及具有工作保持的业务的N相关联。可以实施以上两个示例的任何组合(例如,一定数量的虚拟路径,以及用于每个物理路径的一个虚拟路径)。例如,可以通过多路径传送协议暴露N+1个虚拟路径,例如,用于具有保证的业务的一个虚拟路径以及用于工作保持的业务的N个虚拟路径。在这种情况下,分组分发器可以跨多个物理路径在具有保证的虚拟路径中散布分组。

[0033] 再次参考图2,在一些示例中,可以向VM提供多个虚拟网络接口或地址,以指示哪个业务将被速率限制(例如,低于保证的业务)以及哪个业务将是机会性的或工作保持的。例如,网络设定200的提供商或管理员可以对用于具有带宽保证的业务的VM暴露一个虚拟网络接口或地址,并且对于没有保证的业务,提供一个网络接口或地址。可以(例如,由模块222)将在两个VM的两个保证接口/地址之间行进的业务速率限制到保证带宽。然后,可以将此类业务路由到高优先级队列——例如,交换机208中的Q0中。可以将在两个VM的两个无保

证的接口/地址之间(或在一个有保证的接口/地址和一个无保证的接口/地址之间)行进的业务路由到至少一个较低的优先级队列——例如,交换机208中的Q1中。

[0034] 保证确定和速率限制器模块222可以类似于图1的模块122。模块222可以确保对于移出每个VM的高优先级所路由的总业务(例如,多个流)不超过与出自特定VM的VM到VM对相关联的带宽保证的总值。在图2的示例中,每个VM可以根据优先级将分组路由到不同的虚拟路径中,并且模块222可以对高优先级虚拟路径(例如,用于低于速率保证的业务的虚拟路径)上的业务进行速率限制。在一些情形中,保证确定和速率限制器模块222可以确定VM到VM对带宽保证并且然后可以向VM用信号通知用于每个虚拟路径的发送速率。VM然后可以对高优先级虚拟路径执行速率限制。替换地,模块222对高优先级虚拟路径执行速率限制。在该情形中,VM可以仍然有效地对它们的高优先级虚拟路径进行速率限制,这是因为VM可以通知丢弃的分组(例如,由于速率限制在模块222丢弃的)。在该情况下,VM可以发送对于高优先级路由的较少的分组(例如,低于速率保证的业务)。

[0035] 作为又一个示例性网络设定(在特定图中未示出),保证确定和速率限制器模块(例如,222)和流管理器模块(例如,124、224等等)可以部分地或完全地在NIC(例如,118)内部或交换机(例如,208)内部实施。因而,在本文提供的保证确定和速率限制器模块以及流管理器模块的各种描述可以适用于该示例。另外地,可以基于诸如IP地址或以太网地址之类的源地址和目的地地址在NIC或交换机应用优先级。

[0036] 图3是示例性流管理器模块300的框图。流管理器模块300可以类似于流管理器模块124、224和/或225。流管理器模块300可以接收由流管理器模块300监视的各个路径的信息的分组。可以在图3中将信息的此类分组指出为出站流308。出站流308可以表示出自至少一个VM并且发送给其他VM的分组的流。流管理器模块300可以输出信息的分组(例如,所接收的信息的相同的分组)。可以在图3中将信息的此类分组指出为路由的和/或加标签的流310。可以以特定方式(例如,根据优先级)来对由模块300输出的信息的每个分组进行引导(例如,路由和/或加标签)。流管理器模块300可以包括许多模块302、304、306。这些模块中的每一个可以被实施为在机器可读存储介质上编码的并且可由主机计算设备(例如,102)的处理器执行的一系列指令。另外或作为替代,这些模块可以包括一个或多个硬件设备,该一个或多个硬件设备包括用于实施在本文描述的功能的电子电路。图4是流管理器模块300可以根据优先级对信息的分组进行引导(例如,路由和/或加标签)的示例性方案的流程图400。

[0037] 流监视器模块302可以监视各个VM到VM路径上的业务(例如,出站流308),例如,发送给第二VM(例如,105或205)的、出自第一VM(例如,110或210)的流。流比较模块304可以对于特定VM到VM对确定通过该对的分组的流是否超过该路径的特定带宽保证。如图3中所示的,流比较模块304可以(例如,从模块122或222)接收由流管理器模块300监视的各个路径的最小带宽保证。只要业务流低于特定对的带宽保证,分组引导模块308可以根据第一优先级(例如,高优先级)对业务进行引导(例如,加标签和/或路由)。然后,当该路径上的业务流超过带宽保证(例如,机会性业务)时,模块308可以根据第二优先级(例如,较低优先级)引导业务。可以通过图3的流310来指示由模块300输出的所有引导的(例如,路由和/或加标签的)业务。

[0038] 参考图4,流程图400示出引导的(例如,路由和/或加标签的)流402(例如,与310类

似)可以如何使信息的各个分组被路由到不同的优先级队列(例如,404、408)中。到不同的优先级队列中的此类路由可以发生在网络交换机(例如,108或208)中和/或各个主机计算设备(例如,102)的NIC(例如,118)中。能够在图4中看出,可以使作为低于带宽保证被引导的信息的分组被放置到第一队列Q0(例如,高优先级队列)中,并且可以使作为超过带宽保证被引导的信息的分组(例如,机会性业务)被放置到第二队列Q2(例如,较低的优先级队列)中。通过执行此类优先级加标签和/或路由,如果网络中存在备用带宽,则可以允许机会性业务(例如,一旦其达到交换机108或208)行进或移动,这可以允许特定VM到VM对上的总体业务超过该对的带宽保证。

[0039] 流管理器模块300可以在例如用于相同的租户、应用、服务、分组的类型等等的相同的流内不同地(例如,利用不同的优先级)对信息的分组进行引导(例如,加标签和/或路由)。相比于使用优先级队列对于不同的租户、应用、服务、分组的类型等等简单地实施不同的服务水平的各种其他解决方案,这可以提供附加的好处。因而,流管理器模块300可以允许VM到VM对机会性地利用来自未完全地利用其带宽保证的VM的集合的备用带宽,以发送超过VM到VM对的带宽保证的业务。

[0040] 流管理器模块300(例如,经由模块308)可以以各种方式来引导信息的分组。术语“引导”、“引导的”、“正在引导”等等通常可以指的是基于引导使分组到达不同的目的地。加标签和路由可以是引导的两个特定示例。作为模块300可以如何引导分组的第一示例,并且如图1所示的,流管理器模块124可以位于主机计算设备(例如,102)的管理程序(例如,116)中。流管理器模块124可以监视各个VM到VM路径上的业务,例如,发送给第二VM(例如,105)的、出自第一VM(例如,110)的流。对于特定对,流管理器模块124可以利用优先级标签(例如,高优先级、低优先级,等等)对分组加标签。术语“标签”、“加标签的”、“加标签”等等可以指的是利用编码信息对分组(例如,分组中的数据比特)进行标记。可以以各种方式利用优先级对分组进行标记。例如,可以将优先级编码为分组报头,例如,IP报头中的适当的DiffServ比特。作为另一个示例,管理程序116中的流管理器模块124可以处理多路径TCP连接(或其他多路径传送连接),并且可以基于优先级将分组路由到连接的各个虚拟路径中。可以以与关于VM中的流管理器模块在本文描述的多路径TCP连接类似方式来实施此类多路径连接。

[0041] 在图2的示例中,流管理器模块(例如,224、225)可以位于主机计算设备的VM(例如,210、212、214、205,等等)中。如上所述,可以通过多路径TCP连接或其他类型的多路径传送协议来部分地实施流管理器模块功能的一部分。每个多路径连接可以具有许多虚拟路径,其中每个虚拟路径可以与特定优先级(例如,高优先级和较低优先级)相关联。每个虚拟路径然后可以与硬件交换机208中的特定优先级队列(例如,用于高优先级的Q0和用于较低优先级的Q1)相关联。对于特定VM到VM对,流管理器模块224可以基于路径中的流是否已经超过用于该对的带宽保证来将分组路由到各种虚拟路径中。在一些示例中,高优先级虚拟路径与低于带宽保证的业务相关联,并且因此高优先级虚拟路径可以(例如,由模块222)进行速率限制,如上所述。

[0042] 不管对于各个VM到VM对如何对业务进行引导(例如,路由和/或加标签),业务分组可以利用特定标签或经由特定虚拟路径到达交换机(例如,108或208)。基于这些指定,交换机可以将分组放置在交换机的适当优先级队列中。然后,可以根据交换机的优先级方案来

处理分组。

[0043] 在一些示例中，可以将在本文描述的解决方案扩展为实施各种公平方案以将备用带宽公平地分配给各个VM到VM对。实施公平性例如可以阻止特定租户、VM或者VM到VM对消耗网络中的整个备用带宽。作为第一公平性示例，图5示出流管理器模块300可以根据优先级对信息的分组进行引导(例如，路由和/或加标签)的示例性方案的流程图500。在该示例中，可以使用多于两个优先级队列(例如，优先级队列Q0 504、Q1 506和Q2 508)。在该示例中并且参考图1和图2，交换机108和/或208可以包括三个优先级队列(例如，Q0、Q1和Q2)。返回到图5，能够看出，公平方案包括用于机会性业务(例如，已经超过带宽保证的业务)的两个队列(例如，Q1和Q2)。更具体地，用于机会性业务的第一队列(例如，Q1)可以用于对网络拥塞作出响应的业务，并且第二队列(例如，Q2)可以用于对网络拥塞无响应的业务。队列Q1和Q2两者都可以具有与Q0、用于遵守带宽保证的业务的队列相比较低的优先级。

[0044] 例如，如果业务根据被设计为适于网络拥塞的数据传送协议(例如，TCP)进行行进，则业务可以对网络拥塞作出响应。当网络拥塞(例如，由分组丢失、显式拥塞通知或ECN，等等所指示的)时，TCP协议可以限制更多业务被发送。例如，如果业务根据不考虑网络拥塞的数据传送协议(例如，UDP)行进，或者如果业务经由恶意流(例如，恶意TCP流)行进，则业务可以对网络拥塞无响应。甚至当网络拥塞时，UDP协议可以继续尽可能多地发送分组。因而，在一个特定示例中，Q0可以保持遵守带宽保证的业务，Q1可以保持超过带宽保证的TCP业务，并且Q2可以保持超过带宽保证的UDP业务。此类公平方案可以激励租户使用对拥塞作出响应的传送协议(例如，TCP)。

[0045] 作为第二公平性示例，备用带宽分配模块(例如，在管理程序中或在VM中被实施)可以确定如何公平地向各个其他VM到VM对分配(来自第一VM到VM路径)的备用带宽。备用带宽分配模块可以使用速率限制器来动态地增加各个其他VM到VM对上的超过(例如，由模块122)分配给那些对的保证的流的速率。当确定在网络中存在很少拥塞或没有拥塞时，备用带宽分配模块可以仅仅增加此类速率。备用带宽分配模块可以确定在其监视的VM到VM对上是否存在任何备用带宽，并且然后在各个VM到VM对之间公平地划分备用带宽。

[0046] 备用带宽分配模块可以基于各个VM到VM对的带宽保证(例如，与其成比例地)在各个VM到VM对之间划分备用带宽。例如，假定第一VM到VM对(例如，从VM X到VM Y)具有200Mbps的最小带宽，第二VM到VM对(例如，从VM Z到VM T)具有100Mbps的最小带宽，并且两个对都共享同一网络路径。备用带宽分配模块可以根据2:1比率来向第一对和第二对分配备用带宽。备用带宽分配模块因此可以使用加权方案来分配备用带宽，其可以实施公平性的水平并且可以阻止特定对消耗所有备用带宽。

[0047] 在一些示例中，可以将在本文描述的解决方案扩展为提供较低的等待时间保证，例如，扩展为“高级”租户、应用，等等。高级租户、应用等等可能仍然使它们的业务受限于特定带宽保证，但是与正常租户相比，他们可能遇到较低的等待时间。换句话说，主机计算设备可以(例如，经由模块122、222，等等)仍然确保：对于高级和正常租户、应用等等遵守保证的总业务量不超过每个路径上的带宽保证。正常租户、应用等等的业务可以仍然以最小带宽保证行进，但是可以仍然遇到一些延迟/等待时间。

[0048] 图6示出流管理器模块300可以根据优先级对信息的分组进行引导(例如，路由和/或加标签)的示例性方案的流程图600。在该示例中，可以使用超过两个优先级队列(例如，

优先级队列QP 604、Q0 606和Q1 608)。在该示例中并且参考图1和图2,交换机108和/或208可以包括三个或更多优先级队列(例如,QP、Q0和Q1)。返回到图6,能够看出,该解决方案包括用于遵守带宽保证的业务的两个队列(例如,QP和Q0)。更具体地,第一队列(例如,QP)可以用于针对高级水平的租户、应用等等(例如,业务的租户/应用承诺的低等待时间处理)遵守带宽保证的业务。第二队列(例如,Q0)可以用于针对正常水平的租户、应用等等遵守带宽保证的业务。替换地,对于特定VM到VM对,有保证的业务的一部分可以是低等待时间并且有保证的业务的一部分可以是正常的;以这种方式,用于VM到VM对的业务可以跨越在所有队列(例如,QP、Q0、Q1)上。交换机例如可以在其处理存储在Q0中的分组之前处理存储在QP中的分组,并且与Q1以及或许用于机会性业务的其他队列相比,队列QP和Q0两者都可以具有较高的优先级。

[0049] 在这些示例中,流管理器模块300可以基于流是否超过最小带宽保证并且基于租户、应用等等的水平(例如,正常或者高级)来对信息的分组进行引导(例如,路由和/或加标签)。能够在图6中看出,可以使作为低于带宽保证被引导(例如,加标签和/或路由)并且与高级租户、应用等等相关联的或者通常被确定为低等待时间业务的信息的分组被放置到第一队列QP(例如,最高的优先级队列)中,并且可以使作为低于带宽保证被引导(例如,加标签和/或路由)并且与正常租户、应用等等相关联的或者通常被确定为不是低等待时间的信息的分组被放置到第二队列Q0(例如,较高的优先级队列)中。可以使作为超过带宽保证被引导(例如,加标签和/或路由)的信息的分组(例如,机会性业务)被放置到另一个队列Q1(例如,较低的优先级队列)中。通过执行此类优先级引导(例如,加标签和/或路由),如果网络中存在备用带宽,则可以允许机会性业务(例如,一旦其到达交换机108或者208)行进或者移动,并且高级租户、应用等等可以体验最佳性能。

[0050] 图7是使用优先级进行工作保持的带宽保证的示例性方法700的流程图。以下参考可以与图1和/或图2的主机计算设备102、104、106、202、204和/或206类似的通用主机计算设备来描述方法700的执行。各种其它适当的计算设备——例如图8的主机计算设备800可以执行方法700。可以以存储在诸如存储介质820之类的机器可读存储介质上的可执行指令的形式,和/或电子电路的形式来实施方法700。在本公开的替代实施例中,可以基本上并行地或以与图7中示出的顺序相比不同的顺序来执行方法700的一个或多个步骤。在本公开的替代实施例中,方法700可以包括与在图7中示出的步骤相比更多或更少的步骤。在某些实施例中,方法700的步骤中的一个或多个可以在某些时候是持续的和/或可以重复。

[0051] 方法700可以开始于步骤702并且继续到步骤704,其中主机计算设备可以(例如,经由模块122或222)确定VM到VM对带宽保证。在步骤706,主机计算设备可以(例如,经由模块124或224)监视VM到VM对上的分组的流。在步骤708,对于每个VM到VM对,主机计算设备可以(例如,经由模块124或224)将流与带宽保证相比较。在步骤710,主机计算设备可以基于比较、根据优先级(例如,经由模块124或224)对分组进行引导(例如,路由和/或加标签)。在步骤712,主机计算设备可以基于带宽保证(例如,经由模块122或222)对VM到VM对进行速率限制。例如,可以对低于(多个)带宽保证的业务进行速率限制。在步骤714,分组的引导(例如,路由和/或加标签)可以(例如,在硬件交换机中)使各个分组被放置到优先级队列中。在步骤716,此类优先级队列和/或硬件交换机可以处理分组的流,例如,在硬件交换机中实施特定优先级方案。方法700可以最终继续到步骤718,其中方法700可以停止。

[0052] 图8是使用优先级进行工作保持的带宽保证的主机计算设备800的框图。主机计算设备800可以是能够经由网络与其他计算设备进行通信并且能够运行虚拟机的任何计算系统或计算设备。可以例如关于图1和图2的主机计算设备102、104、106、202、204、206来在以上描述关于各个主机计算设备的更多详情。在图8的实施例中，主机计算设备800包括处理器810和机器可读存储介质820。

[0053] 处理器810可以是适于调取和执行存储在机器可读存储介质820中的指令的一个或多个中央处理单元(CPU)、微处理器和/或其他硬件设备。处理器810可以运行主机计算设备的各个组件和/或模块，例如，管理程序(例如，116)和许多虚拟机(例如，110、112、114)。在图8中示出的特定实施例中，处理器810可以取出、解码并且执行指令822、824、826、828，以使用优先级来执行工作保持的带宽保证。作为对调取和执行指令的替代或补充，处理器810可以包括一个或多个电子电路，该一个或多个电子电路包括用于执行机器可读存储介质820中的一个或多个指令(例如，指令822、824、826、828)的功能的许多电子组件。关于在本文所描述和示出的可执行指令表示(例如，框)，应当理解，在替代实施例中，包括在一个框内的可执行指令和/或电子电路的部分或所有可以被包括在图中示出的不同的框中或者被包括在未示出的不同的框中。

[0054] 机器可读存储介质820可以是存储可执行指令的任何电子的、磁的、光学的或者其他物理存贮设备。因而，机器可读存储介质820例如可以是随机存取存储器(RAM)、电可擦可编程只读存储器(EEPRO)、存储驱动、光盘，等等。机器可读存储介质820可以被布置在主机计算设备800内，如图8中所示。在该情形中，可以将可执行指令“安装”在设备800上。替换地，机器可读存储介质820可以例如是便携式(例如，外部)存储介质，其允许阶段计算设备800远程地执行指令或从存储介质下载指令。在该情形中，可执行指令可以是“安装数据包”的一部分。如在本文所描述的，可以利用用于基于人工测试的基于上下文定位的可执行的指令来对机器可读存储介质820进行编码。

[0055] 带宽保证确定指令822可以确定在源VM和至少一个目的地VM之间的VM对带宽保证，例如，包括在源VM和第一目的地VM之间的第一VM对带宽保证。流监视指令824可以监视从源VM到第一目的地VM的出站网络业务流。流比较指令826可以将流与第一VM对带宽保证相比较。当流小于第一VM对带宽保证时，分组引导指令828可以将流的分组作为有保证的业务来引导(例如，路由和/或加标签)，否则可以将流的分组作为工作保持的业务来引导(例如，路由和/或加标签)。

[0056] 图9是使用优先级进行工作保持的带宽保证的示例性方法900的流程图。可以在以下将方法900描述为由主机计算设备800实行或执行；然而，也可以使用其它适当的主机计算设备，例如，图1和图2的主机计算设备102、104、106、202、204、206。可以以存储在诸如存储介质820之类的机器可读存储介质上的可执行指令的形式，和/或电子电路的形式来实施方法900。在本公开的替代实施例中，可以基本上并行地或以与图9中示出的顺序相比不同的顺序来执行方法900的一个或多个步骤。在本公开的替代实施例中，方法900可以包括与在图9中示出的步骤相比更多或更少的步骤。在某些实施例中，方法900的步骤中的一个或多个可以在某些时候是持续的和/或可以重复。

[0057] 方法900可以开始于步骤902并且继续到步骤904，其中主机计算设备800可以确定源VM和至少一个目的地VM之间的VM到VM带宽保证，例如，包括源VM和第一目的地VM之间的

第一VM到VM带宽保证。在步骤906，主机计算设备800可以监视从源VM到第一目的地VM的出站网络业务流。在步骤908，主机计算设备800可以将流与第一VM到VM带宽保证相比较。在步骤910，当出站网络业务流小于第一VM到VM带宽保证时，主机计算设备800可以根据第一优先级对流的分组进行引导(例如，路由和/或加标签)。当流大于第一VM到VM带宽保证时，主机计算设备800可以根据第二优先级对流的分组进行引导(例如，路由和/或加标签)。方法900可以最终继续到步骤912，在此方法900可以停止。

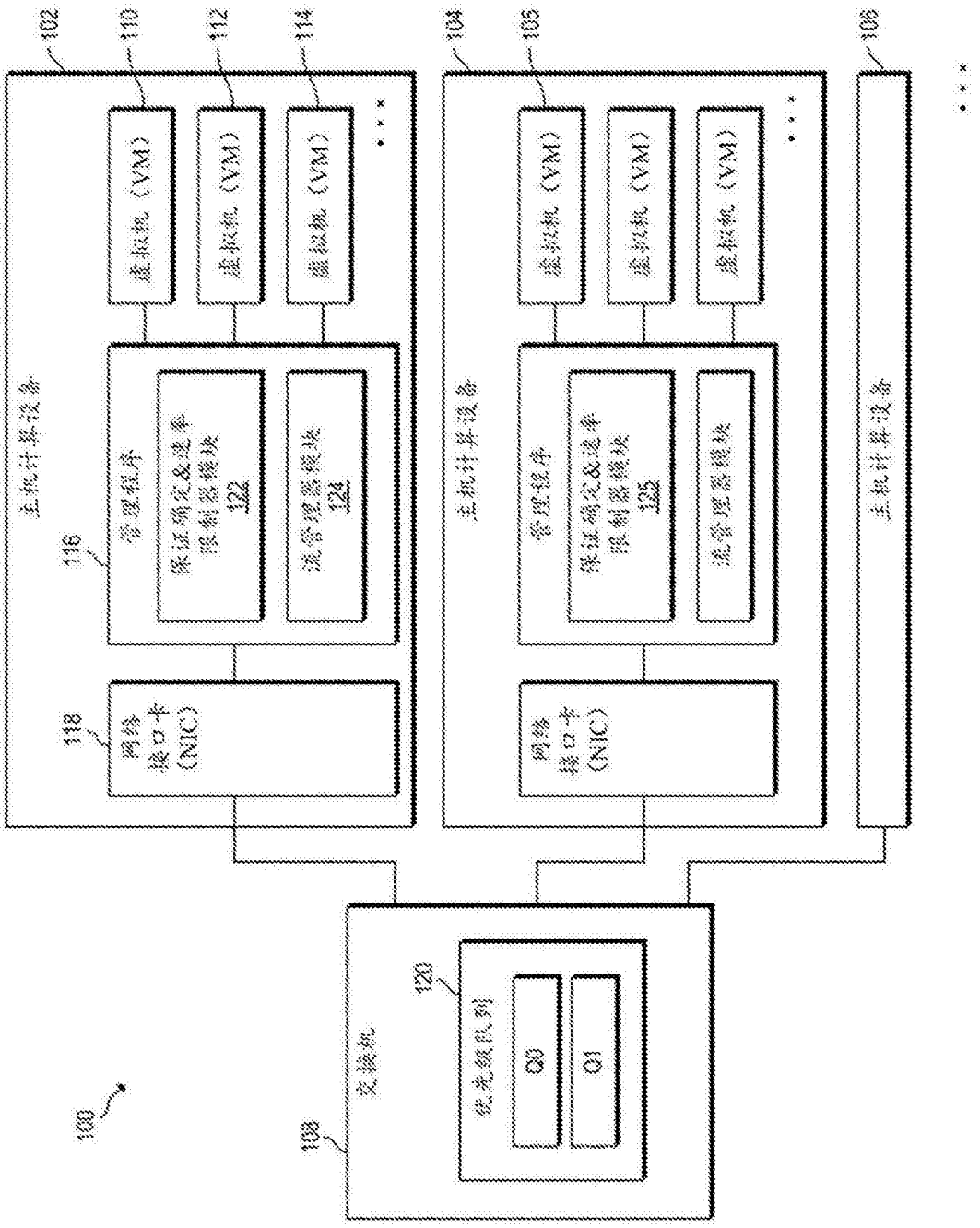

102
110
112
114
104
105
106
虚拟机（VM）
虚拟机（VM）
虚拟机（VM）
虚拟机（VM）
虚拟机（VM）
虚拟机（VM）
主机计算设备
管理程序
保证确定&速率限制器模块
122
流管理器模块
124
116
网络接口卡（NIC）
118
主机计算设备
管理程序
保证确定&速率限制器模块
125
流管理器模块
网络接口卡（NIC）
主机计算设备
100
交换机
108
优先级队列
120
Q0
Q1

图1

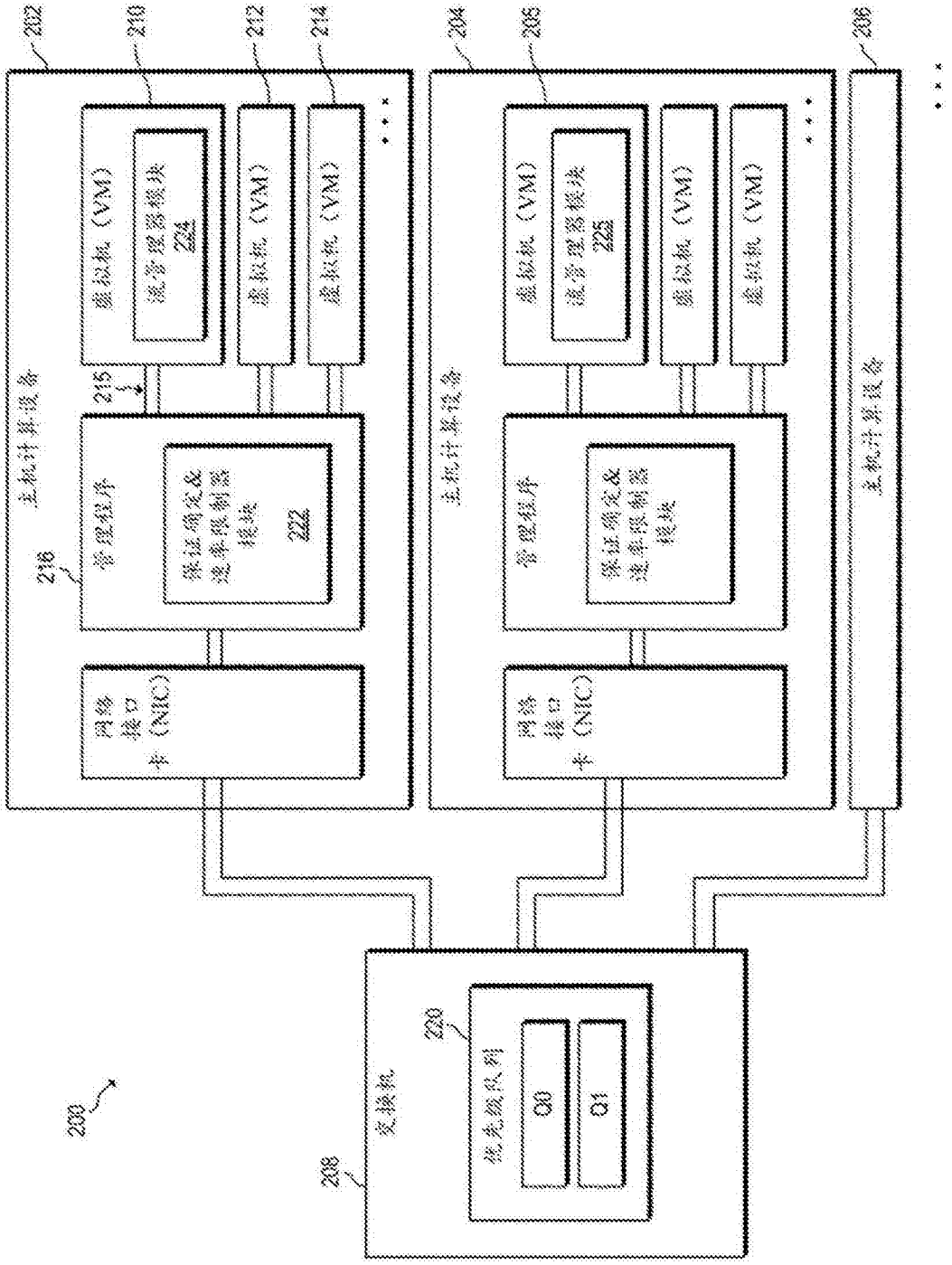

202
210
212
214
204
205
206
虚拟机（VM）
速率管理器模块
224
虚拟机（VM）
虚拟机（VM）
虚拟机（VM）
速率管理器模块
225
虚拟机（VM）
虚拟机（VM）
主机计算设备
215
管理程序
保证确定&
速率限制器
模块
222
216
网络
接口
卡（NIC）
管理程序
保证确定&
速率限制器
模块
网络
接口
卡（NIC）
主机计算设备
200
交换机
208
线光纤队列
220
Q0
Q1

图2

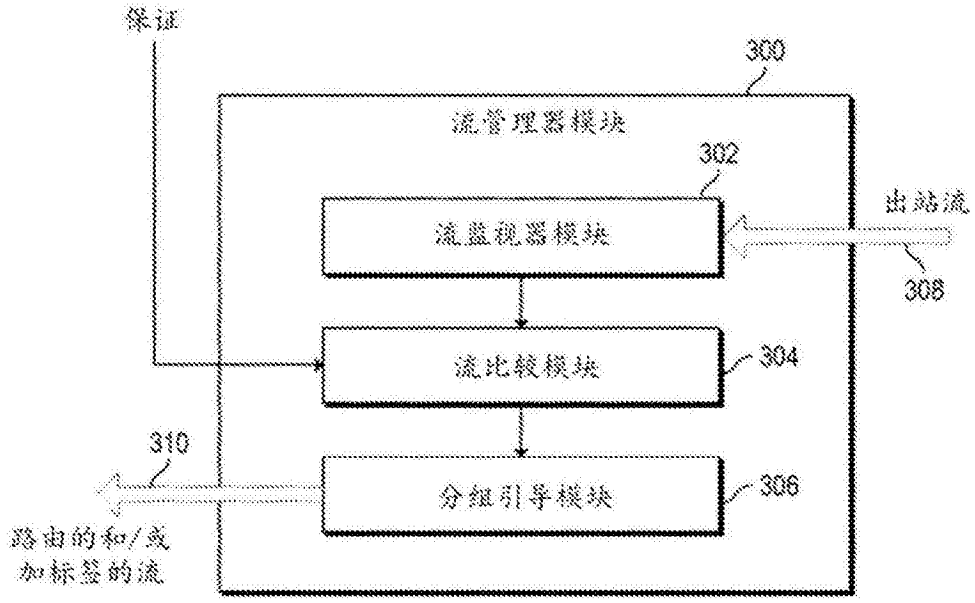

保证
300
流管理器模块
302
流监视器模块
出站流
308
流比较模块
304
310
分组引导模块
306
路由的和/或
加标签的流

图3

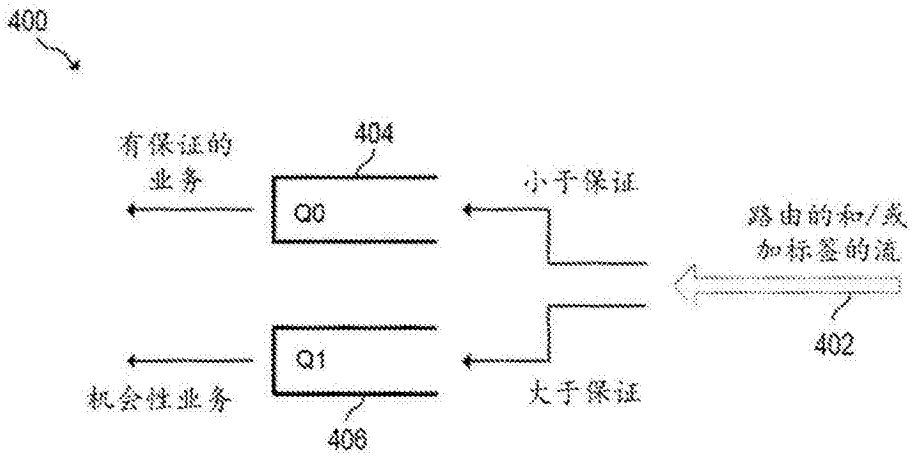

400
有保证的
业务
404
Q0
小于保证
路由的和/或
加标签的流
402
Q1
机会性业务
大于保证
406

图4

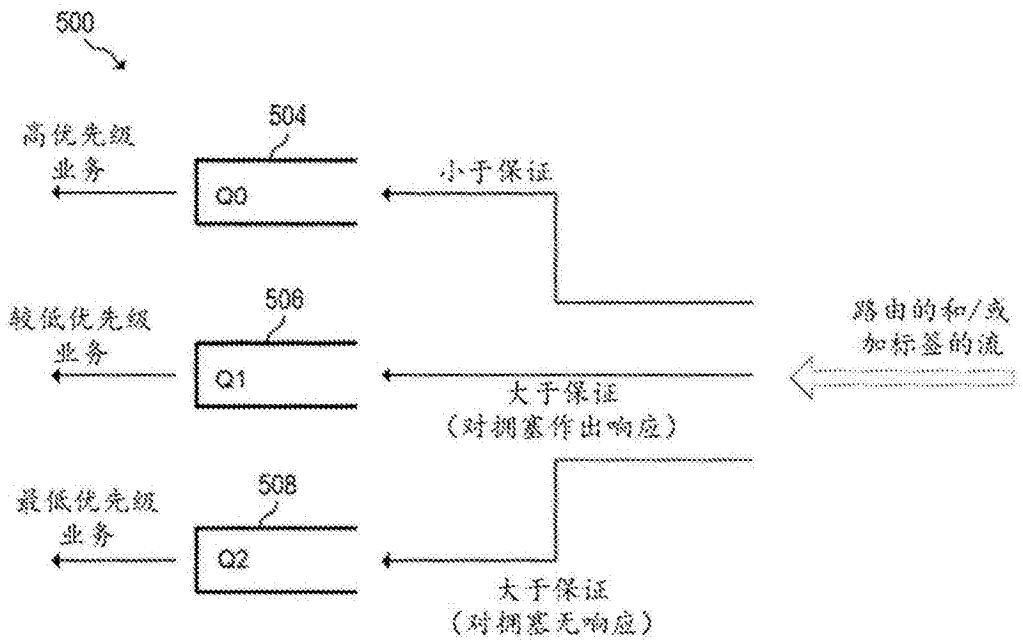

500
504
高优先级
业务
Q0
小于保证
506
较低优先级
业务
Q1
路由的和/或
加标签的流
大于保证
(对拥塞作出响应)
508
最低优先级
业务
Q2
大于保证
(对拥塞无响应)

图5

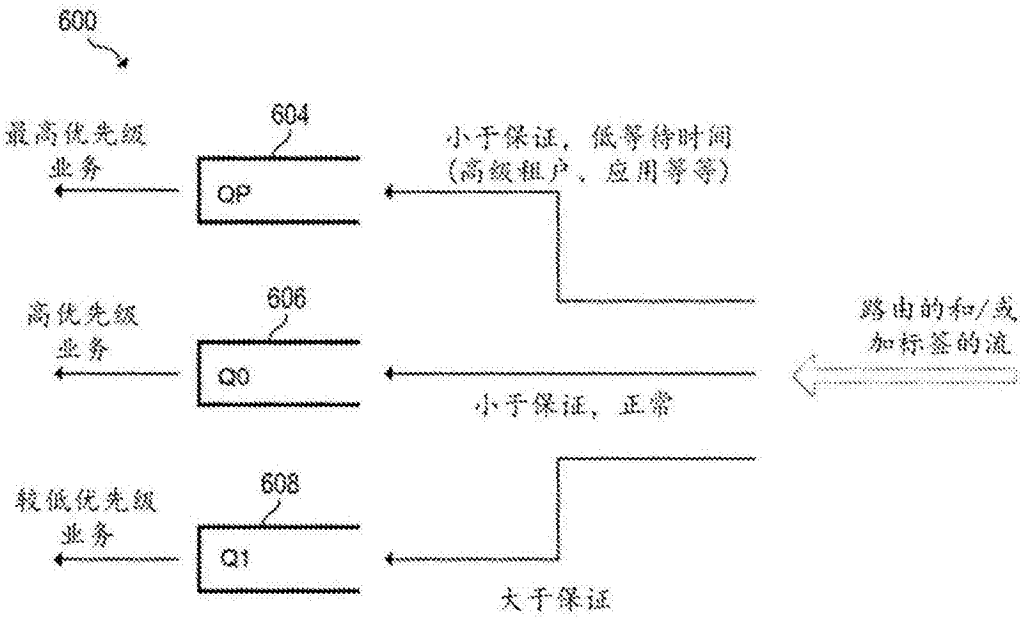

600
604
最高优先级
业务
QP
小于保证，低等待时间
(高级租户、应用等等)
606
高优先级
业务
Q0
路由的和/或
加标签的流
小于保证，正常
608
较低优先级
业务
Q1
大于保证

图6


700
702
开始
704
确定VM到VM对带宽保证
706
监视VM到VM对/路径上的分组的流
708
对于每个VM到VM对/路径，将流与带宽保证相比较
710
根据优先级引导分组
712
基于带宽保证对VM到VM对/路径进行速率限制
714
使引导的分组被放置到优先级队列中
716
优先级队列处理分组流
718
停止

图7

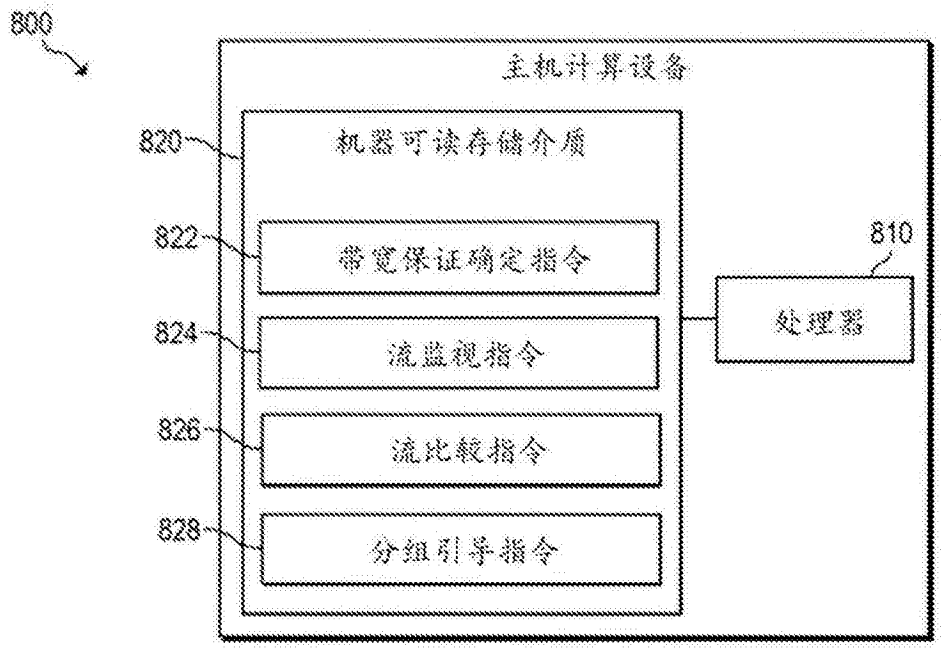

800
主机计算设备
820
机器可读存储介质
822
带宽保证确定指令
824
流监视指令
826
流比较指令
828
分组引导指令
810
处理器

图8

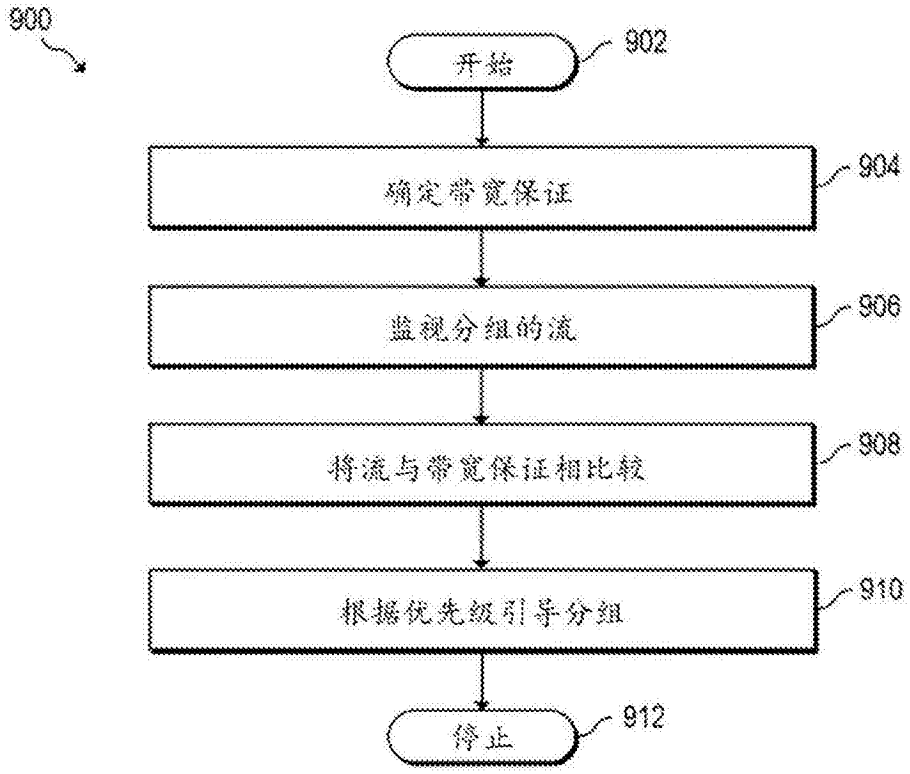

900
开始
902
确定带宽保证
904
监视分组的流
906
将流与带宽保证相比较
908
根据优先级引导分组
910
停止
912

图9