



(12) 发明专利

(10) 授权公告号 CN 102549972 B

(45) 授权公告日 2016. 05. 18

(21) 申请号 201080045182. 4

(51) Int. Cl.

(22) 申请日 2010. 10. 06

H04L 12/28(2006. 01)

(30) 优先权数据

61/250, 013 2009. 10. 09 US

12/723, 697 2010. 03. 15 US

(56) 对比文件

CN 1802837 A, 2006. 07. 12,

US 2007/0280124 A1, 2007. 12. 06,

US 6977416 B2, 2005. 12. 20,

(85) PCT国际申请进入国家阶段日

2012. 04. 06

审查员 吕淼

(86) PCT国际申请的申请数据

PCT/US2010/051698 2010. 10. 06

(87) PCT国际申请的公布数据

W02011/044288 EN 2011. 04. 14

(73) 专利权人 微软技术许可有限责任公司

地址 美国华盛顿州

(72) 发明人 J·D·帕德耶 S·坎杜拉 P·巴尔

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 高见

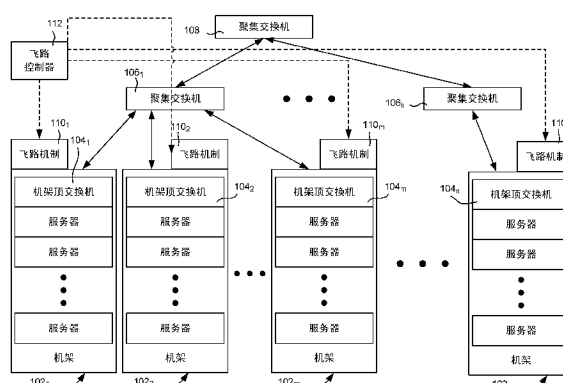
权利要求书1页 说明书7页 附图4页

(54) 发明名称

数据中心中的飞路

(57) 摘要

描述了通过使用动态预设的称为飞路的通信链路来需要时向超负荷基网络提供附加网络通信容量的技术。控制器检测对两个网络机器之间（例如，具有机架顶交换机的两个服务器机架之间）的附加网络通信容量的需求。控制器配置飞路机制（例如，每机架一个）来承载机架的机器之间的网络流量的至少一些，并由此提供附加网络通信容量。飞路机制可基于任何无线或有线技术，包括 60GHz 技术、光链路、802. 11n 或有线商用交换机。



1. 一种在计算机联网环境中的系统,所述系统包括一组飞路机构,所述一组飞路机构包括耦合至服务器机架的一个飞路机构,所述飞路机构与另一飞路机构通信,以向超负荷基网络提供附加网络通信容量,其中所述飞路机构包括能根据需要动态设置以及拆毁的附加网络通信路径。

2. 如权利要求 1 所述的系统,其特征在于,还包括检测对附加网络通信容量的需求的控制器,其中所述飞路机构由所述控制器来动态地预设以彼此进行通信。

3. 如权利要求 1 所述的系统,其特征在于,所述超负荷基网络按树型拓扑结构或网状拓扑结构来配置。

4. 如权利要求 1 所述的系统,其特征在于,所述飞路机构至少部分地基于所述网络的物理布局。

5. 如权利要求 1 所述的系统,其特征在于,所述飞路机构通过物理线路与所述另一飞路机构通信,或者其中所述飞路机构经由无线技术来与所述另一飞路机构通信。

6. 如权利要求 1 所述的系统,其特征在于,所述飞路机构经由无线技术来与所述另一飞路机构通信,并且其中所述无线技术包括 60GHz 频带技术,光链路,或基于 802.11 的 Wi-Fi 技术,或者 60GHz 频带技术、光链路、或基于 802.11 的 Wi-Fi 技术的组合。

7. 一种用于提供通信链路的方法,包括确定网络流量(402),以及基于所述网络流量,在所述网络中的机器之间预设用于承载这些机器之间的网络流量中的至少一些的通信链路,其中所述通信链路能根据需要被动态设置以及拆毁,其中所述通信链路是在耦合至所述机器的一组飞路机构之间提供的。

8. 如权利要求 7 所述的方法,其特征在于,还包括在机器之间预设另一通信链路,包括使用至少一个中间机器来用于多跳通信。

9. 如权利要求 7 所述的方法,其特征在于,确定所述网络流量包括与服务器机器进行通信或轮询交换机,或者既与服务器机器进行通信又轮询交换机。

10. 如权利要求 7 所述的方法,其特征在于,预设所述通信链路包括运行放置算法以选择在其间预设所述通信链路的机器。

11. 一种在计算环境中用于在至少一个处理器上执行的方法,包括:

确定各对网络机器之间的网络流量;

基于所述网络流量,选择所选机器对,其中该所选机器对之间尚未预设飞路;以及

在该所选机器对之间预设飞路,其中所述飞路包括能根据需要动态设置以及拆毁的附加网络通信路径,其中所述附加通信路径是在耦合至所述机器对的一对飞路机构之间提供的。

12. 如权利要求 11 所述的方法,其特征在于,选择所选机器对包括运行放置算法以计算最优路由。

13. 如权利要求 11 所述的方法,其特征在于,还包括预设至少两个飞路以启用不能通过单个飞路进行通信的两个机器之间的多跳网络通信。

数据中心中的飞路

背景技术

[0001] 大型网络数据中心提供规模经济、大型资源池、简化的 IT 管理以及运行大型数据挖掘作业的能力。在构建大型数据中心时,抑制网络成本是重要的考虑因素。网络成本是主要费用之一;如所知的,与提供服务器群集中的任意服务器对之间的线路速度通信带宽相关联的成本通常随着服务器群集的大小而超线性地增加。

[0002] 生产数据中心网络使用高带宽链路以及高端网络交换机来提供所需容量,但是它们仍是超负荷 (over-subscribed) 的 (有时缺乏容量),并由此遭受不定时的性能问题。超负荷通常是技术限制、需要昂贵的“大型”交换机的这些网络的 (例如,树型) 拓扑结构、以及网络管理者保持成本低的压力的组合的结果。其他网络拓扑结构具有类似问题。

发明内容

[0003] 提供本发明内容以便以简化形式介绍将在以下的具体实施例中进一步描述的一些代表性概念。本发明内容不旨在标识出所要求保护的主题的关键特征或必要特征,也不旨在以限制所要求保护的主题的范围的任何方式来使用。

[0004] 简言之,本文中描述的主题的各方面涉及通过使用称为飞路的通信链路来向超负荷基网络提供附加网络通信容量的技术。一般而言,每个飞路包括网络机器之间的根据需要可被动态设置以及按需拆毁的附加网络通信路径。

[0005] 在一个实现中,网络中的每个服务器机架具有相关联的飞路机制,该飞路机制可与其他飞路机制进行通信。飞路机制可基于无线或有线技术。控制器在检测到对两个机架之间的附加网络通信容量的需求时通过使飞路机制彼此通信来动态地预设 (provision) 飞路。一旦被预设,机器之间的附加通信链路就被用来承载这些机器之间的网络流量中的至少一些。

[0006] 结合附图阅读以下具体实施方式,本发明的其他优点会变得显而易见。

[0007] 附图简述

[0008] 作为示例而非限制,在附图中示出了本发明,附图中相同的附图标记指示相同或相似的元素,附图中:

[0009] 图 1 是示出包括了可建立飞路 (flyway) 的飞路机制的示例数据中心的框图。

[0010] 图 2 是在网络机器之间建立的飞路的表示。

[0011] 图 3 是示出各服务器机架以及安装在靠近一组机架的中心的机架的顶上的示例 60GHz 设备的范围的部分数据中心的表示。

[0012] 图 4 是用于预设一对网络机器之间的飞路的示例步骤的流程图。

具体实施方式

[0013] 本文中所描述的技术的各个方面一般地涉及保持现有网络的益处且还充分改进其性能的网络设计。为此,描述了混合式体系结构,其中预设基网络 (包括树型或网状) 来用于平均情形 (并且因此基网络是超负荷的),但是通过包括额外链路来处理热点,这些额

外链路按需被添加以在网络中需要附加容量的地方以及在需要附加容量之时提供附加容量。如本文中所述的,此类链路被称为飞路。注意,将来的网络设计(例如,Clos 网络)也可通过本文中描述的技术来得到增强。如将理解的,经增强的设计允许网络在服务器总数增大时维持其性能,而没有联网装置的数量和成本的显著增加。

[0014] 应当理解,此处描述的任何示例都是非限制性的示例。因此,本发明不限于在此描述的任何具体的实施例、方面、概念、结构、功能或示例。相反,此处所描述的实施例、方面、概念、结构、功能或示例中的任一个都是非限制性的,并且本发明一般能够以在计算和计算机网络方面提供好处和优点的各种方式来使用。

[0015] 图 1 示出了基于树型拓扑结构的网络。多个机架 102_1-102_n 各自具有诸服务器,这些服务器通过机架顶交换机 104_1-104_n 来通信。典型网络具有每机架二十至四十个服务器,其中沿着树向上其链路和交换机愈加强大。注意,飞路不限于树型拓扑结构,而是可被用在任一拓扑结构中,包括 clos 网络以及其他形式的网状拓扑结构,以及胖树 (FatTree) 拓扑结构。

[0016] 如图 1 中所表示的,机架顶交换机 104_1-104_n 中的每个通过一个或多个聚集交换机 106_1-106_k 来彼此耦合。以此方式,每个服务器可与任一其他服务器(包括不同机架中的服务器)通信。注意,在此示例中,更高层的聚集交换机 108 将机架层聚集交换机 106_1-106_k 耦合,并且可以有一个或多个附加聚集交换机耦合层。

[0017] 应用需求通常可由超负荷网络来满足,但是有时候网络不具有用于处理“热点”的足够容量。本文中描述的技术通过根据需要使用飞路来处理额外数据流量以提供附加容量。

[0018] 如图 1 和 2 中所表示的,飞路(图 2 中的弯曲箭头)是经由受飞路控制器 112 控制的飞路机制 110_1-110_n 来实现的。飞路可通过飞路控制器 112 根据需要来动态设置,并且在不需要(或其他地方需要)时拆毁。图 2 示出了可如何使用一个飞路 220 来链接机架 102_1 和 102_n 及其相应的机架顶交换机 104_1 和 104_n , 以及如何使用另一飞路 222 来链接机架 102_2 和 102_n 及其相应的机架顶交换机 104_2 和 104_n 。注意,机架/机架顶交换机在任何时刻可具有一个以上的飞路,如由机架 102_1 与 102_2 之间的飞路 221 所表示的。虽然示出了每机架单个飞路机制,但是可以理解,可以每机架可具有一个以上的飞路机制(或者单个飞路机制中多个设备)——可能使用不同的通信技术(例如,无线或光学)。

[0019] 对来自数据中心网络的记录的分析示出在任何时刻,仅少数的机架顶交换机是“热的(hot)”,即它们正发送和/或接收大量流量。此外,当是热的时,机架顶交换机通常仅与少数的其他机架顶交换机交换其数据中的许多。这变成偏斜瓶颈,其中仅少数的机架顶交换机落后于其他机架顶交换机,并阻碍了整个网络。本文中描述的飞路向这些少数的机架顶交换机提供了额外能力,并由此显著改进整体性能。实际上,仅具有相对较低带宽的少数的飞路显著改进了超负荷数据中心网络的性能。

[0020] 能以相对小的附加成本将飞路添加到网络。这可通过以随机方式使用无线链路(例如,60GHz、光链路和/或 802.11n)和/或使用商用交换机增加容量来实现。通常,任何飞路机制可链接到任何其他飞路机制,只要它们满足耦合需求(例如,在无线范围内、具有光学视线等)。

[0021] 因此,可以各种方式来实现飞路,包括经由根据需要在飞路机制(例如,合适的无

线设备)之间建立的无线链路,和/或将机架顶交换机的子集进行互连的商用交换机。如以下所描述的,60GHz 无线技术是用于创建飞路的一个实现,因为其支持短程(1-10 米)、高带宽(1Gbps)无线链路。此外,60GHz 的高容量和有限干扰范围提供了诸益处。

[0022] 经飞路增强的超负荷网络的性能可接近或甚至等于非超负荷网络的性能。一个达到最大益处的方法是在适当的位置放置飞路。注意,网络流量需求在短时间范围上通常是可预测/可确定的,从而允许预设飞路来跟上改变的需求。如本文中所描述的,中央飞路控制器 112 收集需求数据,以动态方式适应飞路,以及切换路径以路由流量。

[0023] 无线飞路能按需形成链路,并由此可被用来将该可用容量分发给需要它(这由中央飞路控制器 112 来确定)的无论哪个机架顶交换机对。有线飞路提供等效益处;当廉价交换机被连接到机架顶交换机的子集时,可在被连接的许多机架顶交换机对中的需要带宽的那些机架顶交换机对之间划分这些廉价交换机的有限的背板带宽。有线飞路稍微更具限制性,因为仅当机架顶交换机对碰巧经由飞路交换机之一被连接时,它们才可从飞路直接获益(但如以下所描述的,多跳飞路是允许间接益处的一个替换方案)。然而,有线飞路更可能跟上有线速度(例如,由于 NIC 达 10Gbps,且链路达 40Gbps)。在任何情况下,相对于在所有机架上分布相同带宽的替换方案,飞路提供了益处,因为在不需要带宽的链路上这些带宽中的许多将不被使用。此外,现今可在生成数据中心的现有拓扑结构上部署飞路。

[0024] 为了添加飞路以在热的机架顶交换机与同其交换流量的某些其他机架顶交换机之间提供额外容量,可选择提供最高加速的飞路机制对。每条飞路还需要足够的容量。在具有最拥塞的机架顶交换机的机架与该交换机同其交换最多数据的另一机架之间建立飞路是简单直接的。然而,后续选择较不明朗,例如,需要考虑是否在同一机架处或者别处建立另一飞路。

[0025] 一般而言,过分稀疏地分布飞路,或者将它们集中在少数机架顶交换机上不能十分好地工作。例如,在排名前五十的机架顶交换机与其最大的通信者之间各放置一个飞路不会显著地降低热机架顶交换机的完成时间。相反地,在排名前五的机架顶交换机与它们的十个最大通信者的每一个之间放置飞路消除了排名前五的机架顶交换机处的拥塞,但是导致第六机架顶交换机成为瓶颈。一般而言,在帮助更多机架顶交换机与减少热机架顶交换机的每一个处的足够拥塞之间达成恰当的平衡会获得最高加速。以下描述了一个用于飞路放置的合适算法。

[0026] 关于每个飞路需要多少容量,考虑在排名前十的机架顶交换机与它们的同其交换最大数据的五个其他机架顶交换机中的每一个之间添加飞路(即,总共五十条飞路)的示例。大多数飞路要求少于机架顶交换机的上行链路带宽的百分之十就能有用;一个原因在于,机架顶交换机的上行链路承载去往所有其他机架顶交换机的流量,但飞路仅必须承载去往一个其他机架顶交换机的流量。

[0027] 飞路的有用性一般是由导致稀疏需求矩阵的应用特性来产生的。尽管飞路可能不为所有需求提供益处,但是一大组实际应用从飞路获益。例如,在支持 web 服务的数据中心中,请求流量被跨各服务器来进行负载平衡,这些服务器中的每一个进而通过可能要求少数其他服务器生成响应的诸部分(例如,广告)来组装响应页面。数据挖掘作业中的映射缩减的缩减部分可能是最坏情况的情景,其中每个缩减者(reducer)从所有映射者 mapper)拉取数据;该作业受阻,直至所有缩减者完成。甚至在那时,很少具有如此众多的映射者和

缩减者以致所有机架顶交换机同时拥塞。

[0028] 转到 60GHz 无线通信等,毫米波长无线通信已快速发展,并且是用于在数据中心中构建飞路的合适技术。60GHz 频带是 7GHz 宽的频带 (57-64GHz),其在 2001 年被 FCC 留出来供未经许可设备使用。由 ISO 和 IEEE 指定的信道带宽是 2.160GHz。60GHz 频带中实际上更大的带宽便于更高容量的链路。例如,达到少至 1bps/Hz 的简单编码方案使得可能构建具有标称带宽 7Gbps 的链路。在实现更好的调制和编码技术时,60GHz 设备很可能将变得在光谱上高效,并且七个正交 1Gbps 通道将是常见的,而 1Gbps 链路速度对于构建飞路是足够的。注意:通过使用 60GHz 频带中的多个信道,4-5Gbps 链路速度已是可用的。

[0029] 除了容量之外,60GHz 通信的其他特性使其对于实例化飞路而言尤其具有吸引力。作为其中的一个,60GHz 设备具有相对短的范围 (5 至 10 米)。在无线通信路径中,损耗与频率的平方直接成比例,因此与 2.4 和 5GHz Wi-Fi 信号相比,60GHz 信号随距离更快速地衰减。对于另一个,易于实现高定向 60GHz 通信。由于信号波长与频率成反比,因此 60GHz 信号具有非常短的波长 (5mm),并且这允许 RF 芯片设计者将天线阵列结合到收发机芯片中。将许多天线和具有适当定相的收发机整合到单个芯片中允许波束成形以及由此允许定向通信。另外,相控阵天线提升信号质量,从而造成在短距离上的具有低比特比特误差率的更干净的信号。例如,一平方英寸天线在 60GHz 下提供 25dBi 的信号提升。将天线整合到芯片上的附带益处在于,这避免了对用于将信号携带至芯片或从中携带出信号的电线的需求,从而降低封装的成本达两个数量级。

[0030] 低范围和高容量定向链路在其中服务器机架通常被紧密地打包的数据中心中是有益的,因为物理空间是价值高的资源且移动性以及每服务器机架大致十瓦的功率汲取不是问题。定向性允许网络设计者通过空间重用来提高整体频谱效率。例如,由于定向性和范围,四个机架顶交换机之间的两组通信可同时发生。

[0031] 在典型的数据中心服务器中,机架是大约 24 英寸宽,并且以密集且有规律的模式来布置,如图 3 中所示的。在图 3 中,每个框表示例如布置成十行的 24x48 英寸的机架。圆表示安装在中心里的机架的顶上的 60GHz 设备的 10 米范围,该圆包含大致 70 个其他机架。通过使用定向天线,安装在机架的顶部上的小型 (2-3 立方英寸) 60GHz 网络卡可根据需要实例化若干机架之间的多个非干扰吉比特飞路。具有大致 60 度的波束宽度的电可操纵相控阵天线可在毫秒等待时间下被操纵。

[0032] 注意,定向性伴随有后果。例如,发现不再是简单直接的,并且对于全向通信工作良好的流行的媒体接入控制协议必须被修改。尽管如此,这是研究团体中被很好理解的问题,并且存在许多解决方案。IEEE 和 ISO 60GHz 工作组正积极地工作以合并对定向通信和空间重用的支持。尽管他们的主要场景是室内多媒体通信,但是他们讨论的这些解决方案可被修改以应用于数据中心网络。特别地,由于网络被配置的方式以及网络的规模,问题可被解决,因为在所有可发现机架顶交换机之间存在可被转为控制信道的反向信道通信 (常常是有线的)。控制信道随后可被用来发现机架顶交换机,对准 60GHz 波束,分配信道以进行干扰减轻,以及仲裁无冲突媒体接入。

[0033] 关于 60GHz 的另一问题在于,在这些高频率下,信号被物体吸收,且非视线 (NLOS) 通信是不可能的。然而,在数据中心的,设备可被安装到机架的顶部,且不会挡人类操作者的道,由此视线不被挡住。此外,近来的技术进步已开始克服此限制。

[0034] 如以上所述的,频谱的巨大量(结合改进谱效率)允许构建并独立地操作许多飞路(对于 10 比特/Hz 谱效率,可构造达 70 条飞路,每条飞路在 1Gbps 下操作)。这加上归因于定向性的低干扰特性以及归因于路径损耗的低范围允许在数据中心中存在的数千服务器机架上进行显著的频率重用。

[0035] 此外,与其他高带宽无线技术不同,60GHz 频带是世界范围可用的。因而,芯片组和设备在全世界范围内可用,并提供规模经济以及拉低成本。

[0036] 转到实现具有飞路的网络,为了在机架顶交换机对之间形成飞路,放置在相对应的机架的顶上的一个或多个设备创建无线链路。对技术的选择影响可用带宽、可用于空间重用的信道的数目、干扰模式以及飞路范围。天线技术规定设置和摧毁飞路所需的时间。注意,每机架顶交换机添加少数无线设备仅稍微增加成本。

[0037] 有线飞路可通过使用附加交换机来构造,诸如作为将机架顶交换机随机子集的同时时代机架顶交换机的交换机,例如商用交换机可互联二十个机架顶交换机,每链路 1Gbps。为了使链路保持很短,飞路交换机可连接数据center中彼此靠近的机架。当部署有线飞路时,频谱分配或干扰不是问题。然而,其随机构造约束有线飞路;例如,交换大量流量且可从盈余容量获益的机架顶交换机对会无需有线飞路。

[0038] 在任何情况下,与跨所有机架顶交换机均等地划分相同的带宽量相比,有线或者无线飞路构造都是一种改进。与均等地分布带宽且使得许多带宽被浪费(如在需求矩阵稀疏时发生的)不同,飞路提供一种使用备用带宽来瞄准需求矩阵的可从盈余容量获益最多的各部分的方法。

[0039] 中央飞路控制器 112(图 1)收集诸机架交换机对之间的需求的估计。例如,可从端服务器本身处的轻量装置收集信息,或者通过轮询交换机处的 SNMP 计数器来收集信息。通过使用这些估计,控制器 112 可周期性或以其他方式运行放置算法(以下所描述的)来放置可用飞路。

[0040] 基于飞路的网络的拓扑结构因此是动态的,并且要求多路径路由。针对类似问题的解决方案是已知的。控制器 112 确定一对机架顶交换机之间的流量中有多少沿着基网络进行,或改为采取从发送机架顶交换机至接收机架顶交换机的飞路(若存在一个)。机架顶交换机通过将不同的流分配给不同的 MPLS(多协议标签交换)标签交换路径来根据此比例来划分流量。注意:仅少数的飞路(若有的话)在每个机架顶交换机处是可用的。因而,每个交换机处需要的 LSP(标签交换路径)的数目很小,且跨基网络和各一跳飞路划分流量的问题比其他流量划分技术简单得多。

[0041] 创建最优飞路的问题可被认为是最优化问题。给定机架顶交换机 i, j 之间的需求 D_{ij} ,且如果 C_l 是链路 l 的容量,则最优路由是使最大完成时间最小化的路由:

[0042]

$$\text{使得} \quad \min \max \frac{D_{ij}}{r_{ij}} \quad (1)$$

[0043]

$$\sum_{l \in \text{传入}} r_{ij}^l - \sum_{l \in \text{传出}} r_{ij}^l = \begin{cases} D_{ij} & \text{在 ToR } j \text{ 处} \\ -D_{ij} & \text{在 ToR } i \text{ 处} \\ 0 & \text{在所有其他 ToR 处} \end{cases}$$

[0044]

$$\sum_{ij} r_{ij}^l \leq C^l \quad \forall \text{ 链路 } l$$

[0045] 其中 r_{ij} 是机架顶交换机对 i, j 所达到成的速率, 而 r_{ij}^l 是该对在链路 l 上的流量部分。

[0046] 计算最优 / 接近最优飞路放置涉及适当地改变拓扑结构以及重新求解以上最优化问题。例如, 可添加所有可能的飞路, 且具有这样的约束: 不超过特定数目的飞路可同时活动, 或者没有飞路可具有大于特定量的容量。并非由于技术约束而导致的以上最优化问题的所有变体都是易处理的。替代地, 可使用贪婪过程, 其通过求解以上最优化问题来每次增加一条飞路以找到减少完成时间最多的飞路。其他过程可能是更好的逼近。

[0047] 图 4 的示例流程图中一般地表示了一示例, 其始于表示收集正被考虑的的网络流量数据的步骤 402。一般而言, 基于流量数据, 在步骤 404 处选择在两个所选机器对 (机架顶交换机) 之间所需的最优飞路 (即, 交换机是热的, 但是尚未预设飞路)。

[0048] 注意, 飞路还被用来提供至其到其余机架的主连接已故障的机架的连接。此类故障补偿飞路 (或诸飞路) 例如可被认为是步骤 404 处所需的最优飞路, 或者可被分开考虑, 诸如在基于容量选择其他之前。

[0049] 步骤 406 表示评估最优飞路是否可被预设; 例如, 如果飞路的总数是个约束, 则在此时可能没有另一飞路可用。然而, 如果没有飞路可用, 但是比起当前预设的另一飞路, 更需要最优非预设飞路, 则步骤 408 可拆毁该另一飞路 (或在使用多跳场景的情况下拆毁多条飞路, 如以下描述的), 以使得飞路变得可用。例如, 可拆毁最不需要的飞路。

[0050] 然而, 注意, 拆毁飞路可能是不恰当的。例如, 可能的情形是: 所有已预设的飞路正处理甚至更多流量, 且由此与在步骤 404 处的所选对的飞路相比是更需要的, 在此情形中, 不采取动作 (例如, 始自步骤 408 的虚线)。

[0051] 步骤 410 预设最优飞路 (如果合适的话)。一旦预设, 飞路就可被用来处理所选对之间的网络流量中的至少一些, 如步骤 412 所表示的。该过程随后重复。注意: 图 4 仅是一个示例; 作为替换, 在步骤 402 处对于对的每次轮询之间, 尚未被预设的多达排名前 N 的最优飞路可被添加。

[0052] 通过使用飞路, 需要考虑网络的物理拓扑结构, 以及归因于该技术的约束。无线飞路受范围的约束, 而通过将机架顶交换机的随机子集进行互连来构造的有线飞路仅可在这些随机子集之间提供盈余容量。60GHz 飞路横跨十米的距离, 并且使用来自生产数据中心的数据中心布局 (例如, 图 3)。对于有线飞路, 可使用 24 端口、1Gbps 交换机, 且这些端口中的一些用于飞路信道。虽然此类约束影响了飞路的益处, 但是获利仍是显著的。

[0053] 注意: 需要添加多得多的有线飞路才能获得与无线飞路相同的益处。例如, 当添加五十个 24 端口交换机时, 1,200 个双工链路被添加到网络。无线飞路通过仅 50 个半双工链路就能提供等效性能。这是因为无线飞路以有目标的 (targeted) 方式被添加: 它们帮助加速需要附加容量的那些机架顶交换机对。有线飞路被随机地添加, 且仅当它们碰巧在所选对之间时才使该对获益。

[0054] 多跳飞路是可行的。考虑图 3 中的圆内的机器想要与圆外的机器通信。然而, 来自圆内部的足够靠近另一机器的机器的第二跳可被用来完成通信。

[0055] 此外, 控制器 112 (图 1) 可考虑两个机器 (例如, A 和 Z) 是否在单跳范围的外部。

若是,则可在靠近但经由机架顶交换机通信的两个其他机器(例如,C和M)之间构建飞路,其中它们的通信被切换至该飞路。为C和M新构建的飞路所提供的带宽节省随后可被用来释放机架顶交换机,以允许机器A和Z使用机架顶交换机来通信。

[0056] 尽管本发明易于作出各种修改和替换构造,但其某些说明性实施例在附图中示出并在上面被详细地描述。然而应当了解,这不旨在将本发明限于所公开的具体形式,而是相反地,旨在覆盖落入本发明的精神和范围之内内的所有修改、替换构造和等效方案。

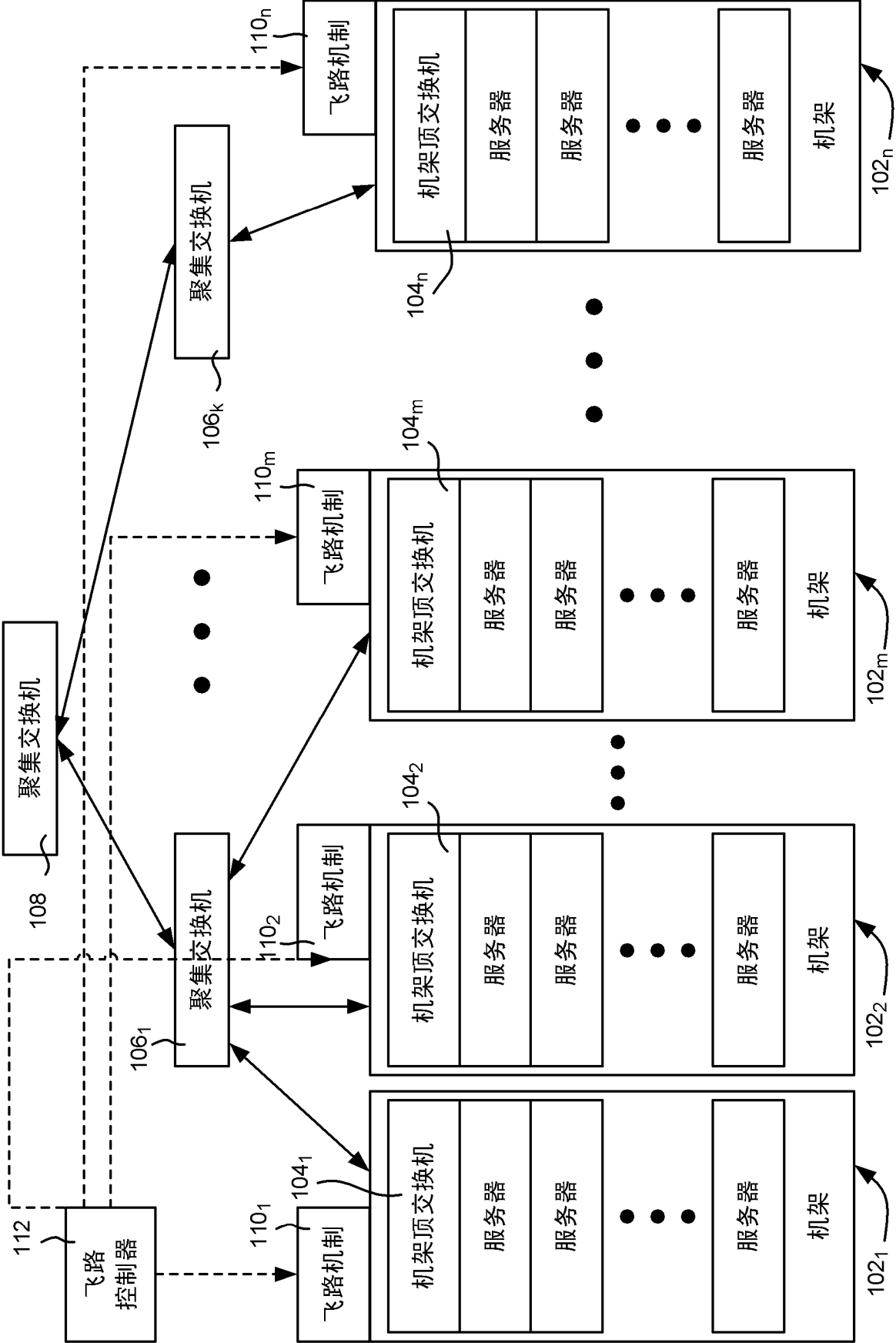


图 1

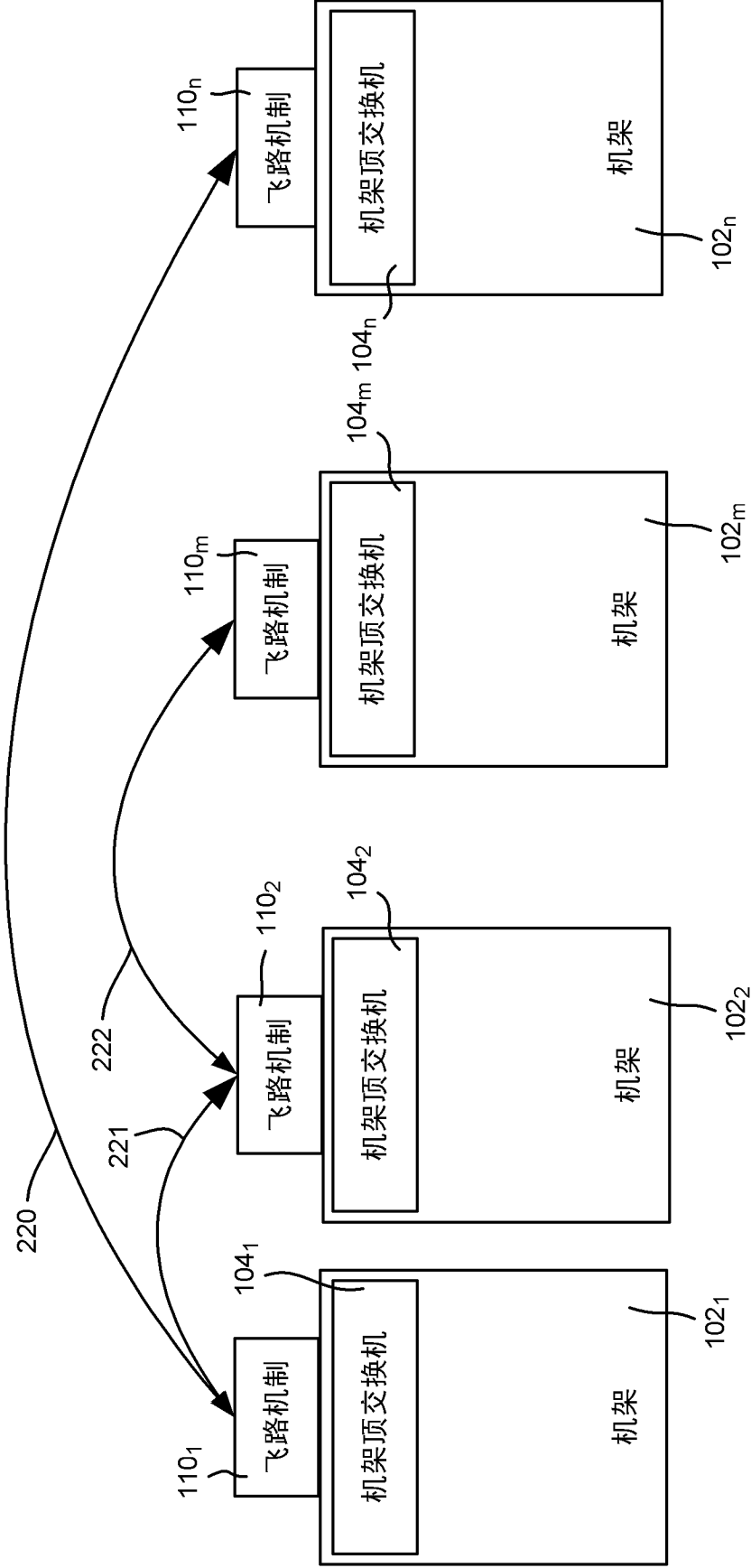


图 2

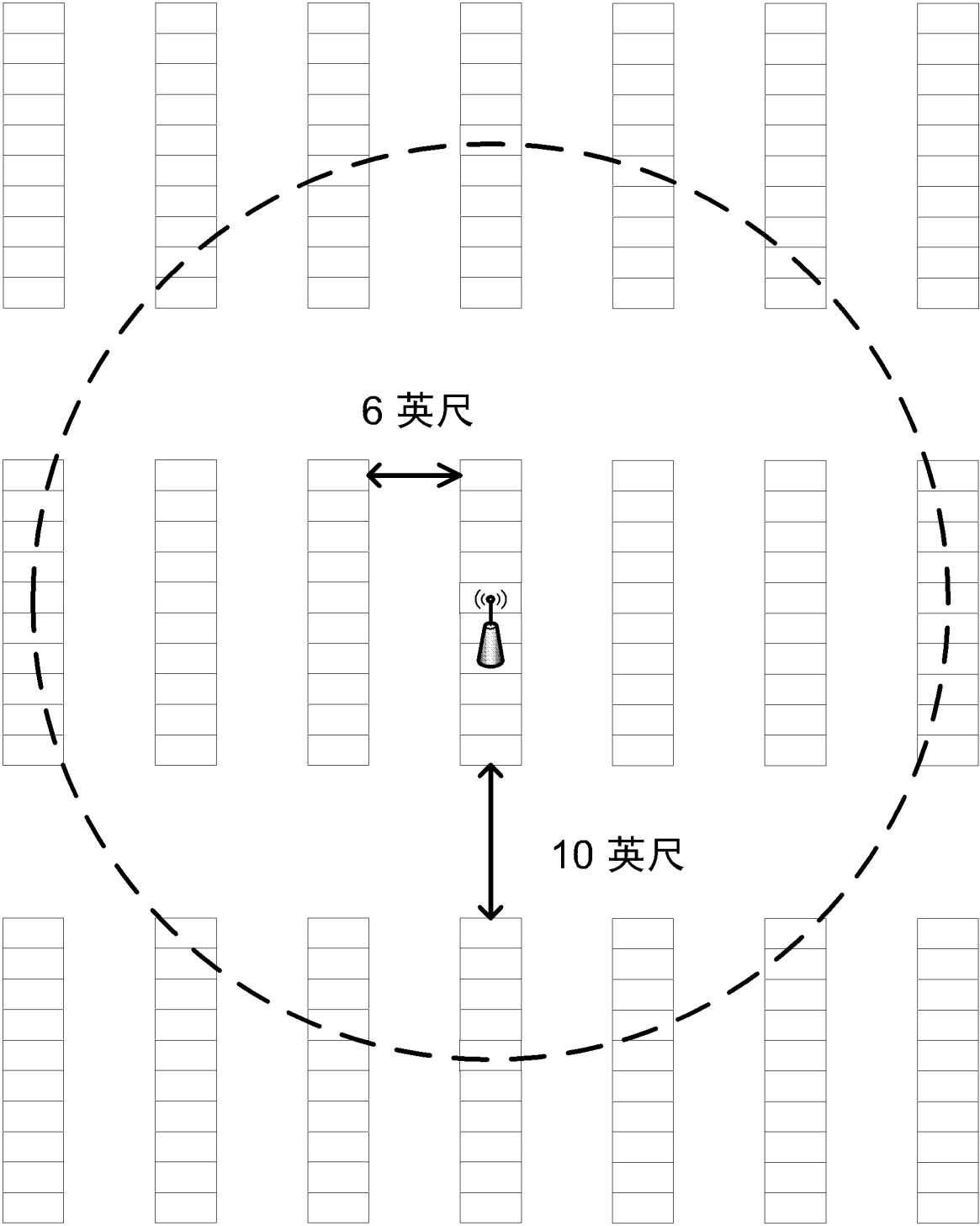


图 3

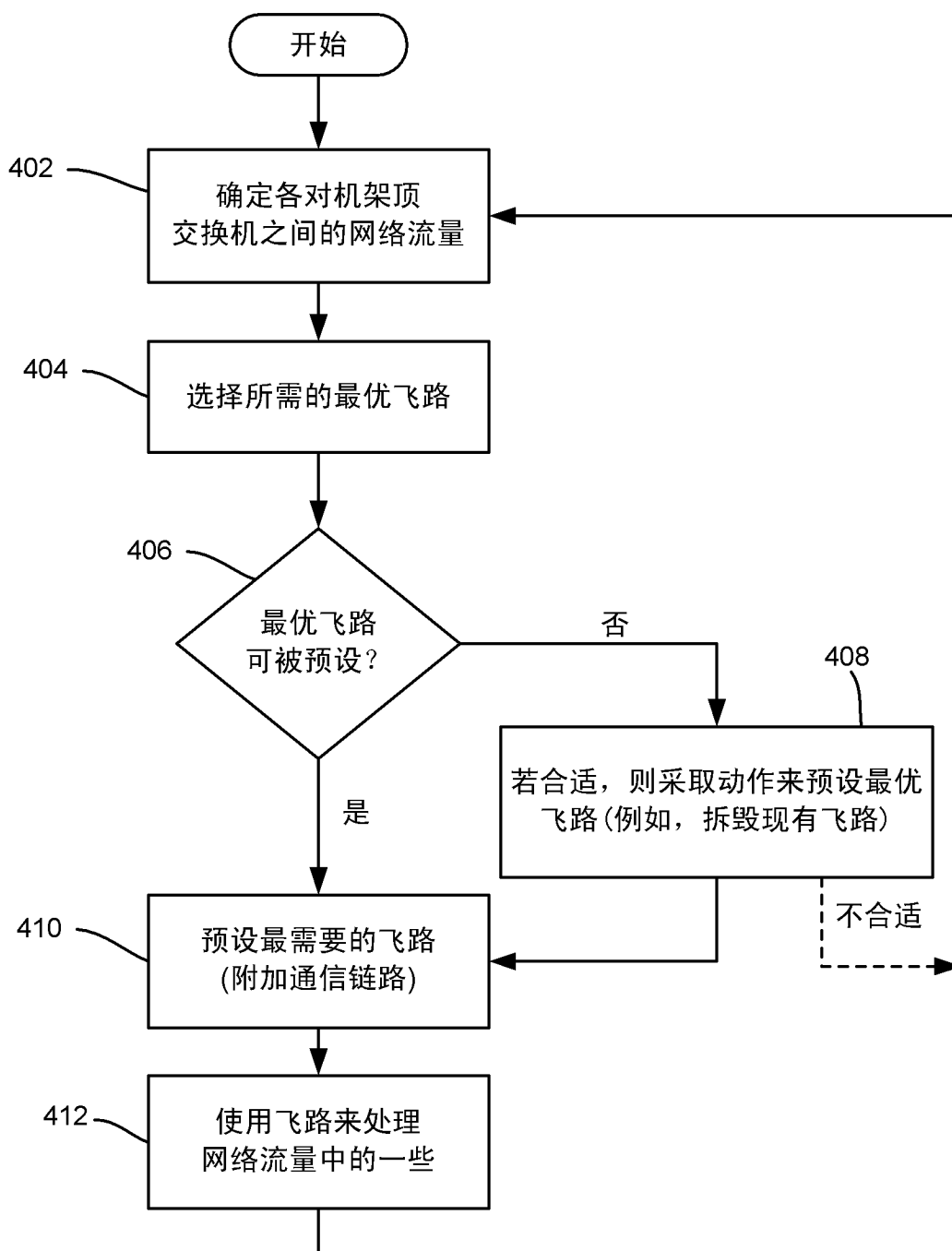


图 4