(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 105978817 A
(43)申请公布日 2016.09.28

(21)申请号 201610135770.6

(22)申请日 2016.03.10

(30)优先权数据
14/644,258 2015.03.11 US

(71)申请人 国际商业机器公司
地址 美国纽约

(72)发明人 B·G·巴纳瓦利卡

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 李晓芳

(51)Int.Cl.
H04L 12/761(2013.01)

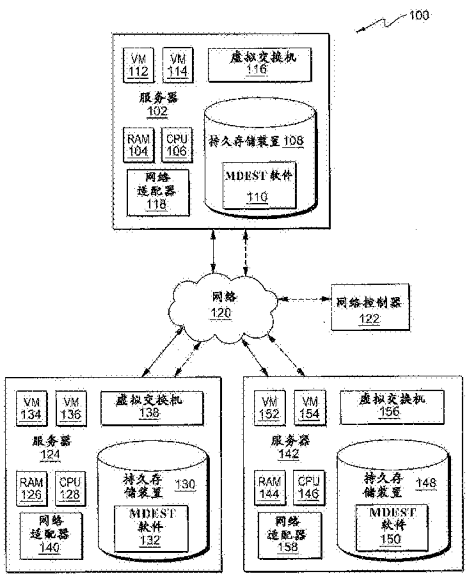
权利要求书2页 说明书9页 附图5页

(54)发明名称

用于传输数据的方法、存储器和网络适配器

(57)摘要

本申请涉及用于传输数据的方法、存储器和网络适配器。在实施例中，网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在网络适配器中存储的列表中标识出的叠加网络中的一个或多个虚拟交换机中的每一个的请求。针对在列表中标识出的一个或多个虚拟交换机中的每一个，网络适配器创建对多目的地分组的头端复制，获得针对所标识出的虚拟交换机的隧道端点信息，将所创建的对多目的地分组的头端复制与特定于所获得的隧道端点信息中标识的隧道协议的报头进行封装，以及将所封装的分组传输到在所标识的虚拟交换机上托管的接收器。



1.一种方法,包括:

网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在所述网络适配器中存储的列表中标识出的所述叠加网络的一个或多个虚拟交换机中的每一个的请求;以及

针对在所述列表中标识出的所述一个或多个虚拟交换机中的每一个:

所述网络适配器创建所述多目的地分组的头端复制;

所述网络适配器获得针对所标识出的虚拟交换机的隧道端点信息;

所述网络适配器将所创建的对多目的地分组的头端复制与特定于在所获得的隧道端点信息中标识的隧道协议的报头进行封装;以及

所述网络适配器将所封装的分组传输到在所标识的虚拟交换机上托管的接收器。

2.如权利要求1所述的方法,还包括:

响应于接收到由所述第一虚拟交换机做出的、对所述网络适配器的应用程序接口的调用,所述网络适配器将第二虚拟交换机的标识符添加到所述列表。

3.如权利要求1所述的方法,还包括:

响应于接收到由所述第一虚拟交换机做出的、对所述网络适配器的应用程序接口的调用,所述网络适配器从所述列表中删除第二虚拟交换机的标识符。

4.如权利要求1所述的方法,其中所述的获得包括:从软件定义的网络控制器中检索针对所标识出的虚拟交换机的隧道端点信息。

5.如权利要求1所述的方法,其中所述隧道端点信息包括与所标识的虚拟交换机所需要的隧道协议有关的信息。

6.如权利要求1所述的方法,其中所封装的分组标识多目的地接收器群组,所述多目的地群组包括所述多目的地分组的发送器和接收一个或多个被传输的封装的分组中的一个的每个虚拟交换机。

7.如权利要求6所述的方法,其中所述多目的地群组在所标识的虚拟交换机上的端口列表中被识别。

8.一种存储程序指令的存储器,该程序指令能够由网络适配器的处理器可执行以实施方法,所述方法包括:

所述网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在所述网络适配器中存储的列表中标识出的所述叠加网络中的一个或多个虚拟交换机中的每一个的请求;以及

针对在所述列表中标识出的所述一个或多个虚拟交换机中的每一个:

所述网络适配器创建所述多目的地分组的头端复制;

所述网络适配器获得针对所标识出的虚拟交换机的隧道端点信息;

所述网络适配器将所创建的对多目的地分组的头端复制与特定于在所获得的隧道端点信息中标识的隧道协议的报头进行封装;以及

所述网络适配器将所封装的分组传输到在所标识的虚拟交换机上托管的接收器。

9.如权利要求8所述的存储器,其中所述方法还包括:

响应于接收到由所述第一虚拟交换机做出的对所述网络适配器的应用程序接口的调用,所述网络适配器将第二虚拟交换机的标识符添加到所述列表。

10.如权利要8所述的存储器，其中所述方法还包括：

响应于接收到由所述第一虚拟交换机做出的、对所述网络适配器的应用程序接口的调用，所述网络适配器从所述列表中删除第二虚拟交换机的标识符。

11.如权利要求8所述的存储器，其中所述的获得包括：从软件定义的网络控制器中检索针对所标识出的虚拟交换机的隧道端点信息。

12.如权利要求8所述的存储器，其中所述隧道端点信息包括与所标识的虚拟交换机所需要的隧道协议有关的信息。

13.如权利要求8所述的存储器，其中所封装的分组识别多目的地接收器群组，所述多目的地群组包括所述多目的地分组的发送器和接收一个或多个被传输的封装的分组中的一个的每个虚拟交换机。

14.如权利要求13所述的存储器，其中所述多目的地群组在所标识的虚拟交换机上的端口列表中被识别。

15.一种网络适配器，包括处理器、存储器、以及被存储在所述存储器中以供所述处理器执行以实现方法的程序指令，所述方法包括：

所述网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在所述网络适配器中存储的列表中标识出的所述叠加网络中的一个或多个虚拟交换机中的每一个的请求；以及

针对在列表中标识出的所述一个或多个虚拟交换机中的每一个：

所述网络适配器创建所述多目的地分组的头端复制；

所述网络适配器获得针对所标识出的虚拟交换机的隧道端点信息；

所述网络适配器将所创建的对多目的地分组的头端复制与特定于所获得的隧道端点信息中标识的隧道协议的报头进行封装；以及

所述网络适配器将所封装的分组传输到在所标识的虚拟交换机上托管的接收器。

16.如权利要求15所述的网络适配器，其中所述方法还包括：

响应于接收到由所述第一虚拟交换机做出的、对所述网络适配器的应用程序接口的调用，所述网络适配器将第二虚拟交换机的标识符添加到所述列表。

17.如权利要15所述的网络适配器，其中所述方法还包括：

响应于接收到由所述第一虚拟交换机做出的、对所述网络适配器的应用程序接口的调用，所述网络适配器从所述列表中删除第二虚拟交换机的标识符。

18.如权利要求15所述的网络适配器，其中所述的获得包括：从软件定义的网络控制器中检索针对所标识出的虚拟交换机的隧道端点信息。

19.如权利要求15所述的网络适配器，其中所述隧道端点信息包括与所标识的虚拟交换机所需要的隧道协议有关的信息。

20.如权利要求15所述的网络适配器，其中所封装的分组识别多目的地接收器群组，所述多目的地群组包括所述多目的地分组的发送器和接收一个或多个被传输的封装的分组中的一个的每个虚拟交换机。

用于传输数据的方法、存储器和网络适配器

技术领域

[0001] 本发明一般涉及数据传输的领域，并且更特别地涉及在叠加网络(overlay network)中传输多目的地(multi-destination)分组。

背景技术

[0002] 数据处理环境包括各种硬件、软件和固件网络组件。物理网络，也被称为底层(underlay)，是使用这样的组件来定义的网络。

[0003] 目前可以使用一些技术从这样的联网组件构造逻辑网络，该逻辑网络也被称为软件定义网络(SDN)叠加(以下称为“叠加”，“叠加网络”或“OVN”)。本质上，联网组件被抽象为对应的逻辑表示或虚拟表示，并且该抽象被用于定义叠加。也就是说，叠加是通过使用底层联网组件的逻辑表示来形成和操作的逻辑网络。

[0004] 物理网络通常存在于数据处理环境的划定边界内，该数据处理环境的联网组件在物理网络中被利用。与物理网络不同，叠加可以被设计为跨越一个或多个数据处理环境。例如，物理网络可以被包含在数据中心内，而叠加可以跨越一个或多个数据中心。

[0005] 作为示例，联网网关的逻辑表示可以加入叠加中，以使得归结于叠加中的联网网关的逻辑表示的功能实际上由叠加中的底层联网网关组件执行。

[0006] 在叠加中，因为执行联网功能的实际的联网组件被抽象为表示那些组件所提供的联网功能性的逻辑实体，而不是那些功能的实际实现，因此需要一些事物来将该联网功能性引入运行的逻辑网络中。SDN控制器是在叠加内管理和操作逻辑联网组件的组件。

[0007] 虚拟机器(VM)包括对数据处理系统中可用的真实硬件、软件和固件组件的虚拟化表示。数据处理系统可以具有配置在其上的任何数量的VM，并且利用在其中的任何数量的虚拟化组件。数据处理系统还被称为计算节点(computing node)、计算节点(compute node)、节点、或主机。

[0008] 在诸如数据中心中之类的大规模的数据处理环境中，数以千计的VM可以在任何给定时间在主机上运行，此时如果没有数以千计的这样的主机则数以百计的这样的主机可以在数据中心中是运行的。诸如所描述的数据中心之类的虚拟化数据处理环境经常被称为“云”，其基于需要向若干客户端提供计算资源和计算服务。

[0009] 虚拟交换机(在本文中有时被称为vSwitch)是允许VM之间的通信的软件应用。虚拟交换机是完全虚拟的，并且可以连接到网络接口卡(NIC)。虚拟交换机将物理交换机合并为单个逻辑交换机。这有助于增加带宽，并且在服务器和交换机之间创建活动网格(active mesh)。虚拟交换机可以被嵌入到服务器所安装的软件中，或被包括在服务器的硬件中作为其固件的一部分。

[0010] 通过定义叠加网络而进行网络虚拟化是在数据中心和云计算环境的管理和运行中新兴的趋势。网络虚拟化的一个目标是简化在多宿主数据处理环境以及专用的客户数据处理环境中的网络供应(provisioning)。

[0011] 单播(Unicasting)是点到点，即从单个发送器到单个接收器，发送数据的方法。广

播是将相同的数据发送给所有可能的目的地的方法。另一个多目的地分发方法(多播(multicasting))通过使用特别的地址分配将相同的数据仅仅发送给被称为接收器的感兴趣目的地。因特网协议(IP)多播是在IP分组的单个传输中将IP分组多播传送给若干接收器的过程。IP多播是用于帮助节约数据中心中的带宽并且减小服务器上的负荷的流行技术。

[0012] 在叠加网络中操作的IP多播被称为叠加多播。叠加多播可以以不同的方式实现,依赖于底层网络中提供的对多播的支持。基于多播的叠加多播要求底层网络提供对多播的支持。在底层网络中的多播目前在数据处理环境中并不普遍。基于多-单播(multi-unicast)的叠加多播是在叠加网络中传输多播分组的方法,在该叠加网络中底层支持单播,但不支持多播。

发明内容

[0013] 在一个实施例中,一种方法包括,网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在网络适配器所存储的列表中标识出的、叠加网络的一个或多个虚拟交换机中的每一个的请求。该方法还包括,针对在列表中标识出的一个或多个虚拟交换机中的每一个:网络适配器创建对多目的地分组的头端复制;网络适配器获得针对所标识出的虚拟交换机的隧道端点信息(tunneling endpoint information);网络适配器将所创建的对多目的地分组的头端复制与特定于所获得的隧道端点信息中标识的隧道协议的报头进行封装;以及网络适配器将所封装的分组传输到在所标识出的虚拟交换机上托管(host)的接收器。

[0014] 在另一个实施例中,存储器存储用于实施方法的、由网络适配器的处理器可执行的程序指令。该方法包括,网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在网络适配器所存储的列表中标识出的、叠加网络的一个或多个虚拟交换机中的每一个的请求。该方法还包括,针对在列表中标识出的一个或多个虚拟交换机中的每一个:网络适配器创建对多目的地分组的头端复制;网络适配器获得针对所标识出的虚拟交换机的隧道端点信息;网络适配器将所创建的对多目的地分组的头端复制与特定于所获得的隧道端点信息中标识的隧道协议的报头进行封装;以及网络适配器将所封装的分组传输到在所标识的虚拟交换机上托管的接收器。

[0015] 在另一个实施例中,网络适配器包括处理器、存储器和存储在存储器中以供处理器执行来实施方法的程序指令。该方法包括,网络适配器从叠加网络的第一虚拟交换机接收将多目的地分组传输到在网络适配器存储的列表中标识出的、叠加网络的一个或多个虚拟交换机中的每一个的请求。该方法还包括,针对在列表中标识出的一个或多个虚拟交换机中的每一个:网络适配器创建对多目的地分组的头端复制;网络适配器获得针对所标识出的虚拟交换机的隧道端点信息;网络适配器将所创建的对多目的地分组的头端复制与特定于所获得的隧道端点信息中标识的隧道协议的报头进行封装;以及网络适配器将所封装的分组传输到在所标识的虚拟交换机上托管的接收器。

附图说明

[0016] 图1是示出在根据本发明的实施例中的多目的地分组处理环境的功能性框图;

[0017] 图2是示出在根据本发明的实施例中的网络适配器的功能性框图;

[0018]　图3是在根据本发明的实施例中，描绘在图1中的多目的地分组处理环境内向多目的地群组接收器列表中添加虚拟交换机和移除虚拟交换机的操作步骤的流程图；

[0019]　图4是在根据本发明的实施例中，描绘用于在图1中的多目的地分组处理环境内将多目的地分组传输到多目的地群组的操作步骤的流程图；

[0020]　图5描绘了在根据本发明的实施例中的服务器计算机的框图。

具体实施方式

[0021]　根据本发明的实施例认识到网络适配器的显著优势：基于由基于软件的数据路径提供的卸载(offload)指令维持多目的地群组与隧道端点(TEP)的映射、提供应用程序接口(API)以操控这样的映射、以及处理针对多目的地分组的头端复制(head end replication)。本文所描述的实施例使虚拟交换机从创建一到许多的单播分组以用于多目的地分组，并且从针对新创建的分组中的每一个执行校验和以及封装中解脱。本文所描述的实施例消除了对用于在基于虚拟交换机软件的数据路径处处理多目的地分组的现有方法的限制。

[0022]　本文所描绘的实施例认识到在OVN TEP处的多目的地流量(traffic)处理是中央处理单元(CPU)昂贵的操作。一些在OVN TEP处的多目的地分组处理的实现对用于处理多目的地分组的底层网络的支持具有依赖性，这对这样的OVN解决方案的可部署性造成显著的限制，因为处理多目的地分组所需要的底层网络支持不总是存在于用户驻地(customer premises)中。用户驻地设备(CPE)是位于用户驻地(物理位置)而不是位于提供者的驻地或在之间的电话或其他服务提供者设备。电话手持设备、有线电视机顶盒、以及数字用户线路(DSL)路由器是CPE的示例。可替代地，OVN TEP使用重复的单播方法来处理多目的地分组递送。本文所描述的实施例认识到，这样的方法是CPU密集的，易于延迟的，并且具有显著的可缩放性问题，因为分组形成、校验和以及封装工作在基于软件的数据路径中被执行。本文所描述的实施例进一步认识到，这样的实现还遭受不一致的多目的地分组处理性能，并且影响对正常的单播分组的处理，因为对多目的地分组的重复单播的典型处理涉及在去往远程TEP的其他单播分组之前创建针对每个多目的地分组的一到许多的单播分组。这个问题在OVN TEP具有与底层网络的单个物理连接时变得甚至更严重。

[0023]　现在将参考附图详细描述根据本发明的实施例。图1是示出在根据本发明的实施例中的多目的地分组处理环境的功能性框图。

[0024]　多目的地分组处理环境100包括服务器120、网络控制器122、服务器124、和服务器142，这些全部通过网络120互联。服务器102、124和142各自可以是能够处理程序指令以及例如通过网络120来接收和发送数据的任何电子设备或电子设备的组合。在一些实施例中，服务器102、124和142可以是膝上型计算机、平板计算机、上网本计算机、个人计算机(PC)、台式计算机、服务器计算机、个人数字助手(PDA)、或智能电话中的任何一个。在其他实施例中，服务器102、124和142各自可以表示(诸如在分布式计算环境中)利用多个计算机作为服务器系统的服务器计算系统。

[0025]　服务器102包括RAM 104、CPU 106、持久存储装置108、VM 112、VM 114、虚拟交换机116以及网络适配器118。持久存储装置108可以例如是硬盘驱动器。可替代地，或除了硬磁盘驱动器以外，持久存储装置108可以包括固态硬驱动器、半导体存储设备、只读存储器

(ROM)、可擦除可编程只读存储器(EPROM)、闪速存储器、或能够存储程序指令或数字信息的任何其他计算机可读存储介质。持久存储装置108存储操作系统软件,以及使得服务器102能够通过网络120上的数据连接与网络控制器122、服务器124、和服务器142进行通信的软件。多目的地软件110(有时缩写为“mdest软件110”)也存储在持久存储装置108中。多目的地软件110是使得虚拟交换机116能够通过网络120与网络控制器122、服务器102上的VM 112和VM 114;服务器124上的VM 134、VM 136和虚拟交换机138;以及服务器142上的VM 152、VM 154和虚拟交换机156进行通信的软件。多目的地软件110还使得网络控制器122能够向虚拟交换机116、138和156供应针对给定的多目的地群组的多目的地群组接收器列表。

[0026] VM 112使用网络120上的叠加虚拟网络通过虚拟交换机116、虚拟交换机138和虚拟交换机156与VM 134和VM 152进行通信。VM 114使用网络120上的单独的叠加虚拟网络通过虚拟交换机116、虚拟交换机138和虚拟交换机156与VM 136和VM 154进行通信。虚拟交换机116被VM 112和VM 114使用以用于与网络120上它们各自的叠加网络上的节点进行通信。在一个实施例中,可以在多目的地分组处理环境100中存在许多虚拟机器和虚拟叠加网络。在其他实施例中,网络控制器122可以驻留在多目的地分组处理环境100中的服务器上,或作为连接到网络120的独立的计算机。

[0027] 网络适配器118是将服务器102连接到网络120的计算机硬件组件。网络适配器118允许服务器102与网络控制器122、服务器124和服务器142进行通信。网络适配器118还包括卸载能力,该卸载能力允许网络适配器118基于由基于软件的数据路径提供的卸载指令维持多目的地群组和TEP的映射、提供应用程序接口(API)以操纵这样的映射、以及处理针对多目的地分组的头端复制(head end replication)。网络适配器118还使得VM 112能够使用网络120上的叠加虚拟网络通过虚拟交换机116、虚拟交换机138和虚拟交换机156与VM 134和VM 152进行通信。网络适配器18还使得VM 114能够使用网络120上的单独的叠加虚拟网络通过虚拟交换机116、虚拟交换机138和虚拟交换机156与VM 136和VM 154进行通信。

[0028] 在图1中,网络120被示出为服务器102、网络控制器122、服务器124、和服务器142之间的互联结构。在实际中,网络120可以是任何可行的数据传输网络。网络120可以是例如,局域网(LAN)、诸如因特网之类的广域网(WAN)、或两者的组合,并且可以包括有线、无线、或光纤连接。一般来说,网络120可以是根据本发明的实施例支持服务器102、网络控制器122、服务器124、和服务器142之间的通信的连接和协议的任何组合。

[0029] 网络控制器122是具有虚拟网络、虚拟交换机、虚拟端点(多目的地的发送器和接收器)的端到端供应的可见性的SDN控制器。网络控制器122负责生成在给定的多目的地群组(例如VM 112、VM 134和VM 152)内的多目的地群组接收器列表,并且将其供应给具有针对给定的多目的地群组的多目的地发送器的、例如虚拟交换机156的虚拟交换机。在一个实施例中,网络控制器122可以是在多目的地分组处理环境100中的单独的计算机系统、服务器、或硬件。在另一个实施例中,网络控制器122可以是服务器102、服务器124或服务器142的一部分。一些实施例可以在多目的地分组处理环境100中包括多于一个的网络控制器,以便例如用作网络控制器122的备用控制器。

[0030] 服务器124包括RAM 126、CPU 128、持久存储装置130、VM 134、VM 136、虚拟交换机和网络适配器140。持久存储装置130包含类似于mdest软件110的mdest软件132。网络适配器140是将服务器124连接到网络120的计算机硬件组件。网络适配器140允许服务器124与

网络控制器122、服务器102和服务器142通信。网络适配器140还包括允许网络适配器140基于由基于软件的数据路径提供的卸载指令维持多目的地群组和TEP的映射、提供应用程序接口(API)以操纵这样的映射、以及处理针对多目的地分组的头端复制的卸载能力。网络适配器140还使得VM 134能够使用网络120上的叠加虚拟网络通过虚拟交换机138、虚拟交换机116和虚拟交换机156与VM 112和VM 152进行通信。网络适配器140还使得VM 136能够使用网络120上的单独的叠加虚拟网络通过虚拟交换机138、虚拟交换机116和虚拟交换机156与VM 114和VM 154进行通信。

[0031] 服务器142包括RAM 144、CPU 146、持久存储装置148、VM 152、VM 154、虚拟交换机156和网络适配器158。持久存储装置148包含类似于mdest软件110的mdest软件150。网络适配器158是将服务器142连接到网络120的计算机硬件组件。网络适配器158允许服务器142与网络控制器122、服务器102和服务器142进行通信。网络适配器158还包括允许网络适配器158基于由基于软件的数据路径提供的卸载指令维持多目的地群组和TEP的映射、提供应用程序接口(API)以操纵这样的映射、以及处理针对多目的地分组的头端复制的卸载能力。网络适配器158还使得VM 152使用网络120上的叠加虚拟网络通过虚拟交换机156、虚拟交换机116和虚拟交换机138与VM 112和VM 134进行通信。网络适配器158还使得VM 154使用网络120上的单独的叠加虚拟网络通过虚拟交换机156、虚拟交换机116和虚拟交换机138与VM 114和VM 136进行通信。

[0032] 在图1的示例实施例中，服务器142用作针对被称为“MDG-1”的、在虚拟交换机156上托管的多目的地群组的多目的地发送器，并且处理针对VM 112、VM 134、VM 152、虚拟交换机138和虚拟交换机116的网络数据。服务器142还用作针对被称为“MDG-2”(未示出)的、在虚拟交换机156上托管的多目的地群组的多目的地发送器，并且处理针对VM 114、VM 136、VM 154、虚拟交换机138和虚拟交换机116的网络数据。在其他实施例中，多目的地群组可以被托管在多目的地分组处理环境100中的一个或多个虚拟交换机上。

[0033] 图2是示出根据本发明的实施例中的网络适配器的功能性框图。网络适配器200是网络适配器118、140和158中的每一个的实施例，并且包含处理器202、存储器204、外围组件互连(PCI)接口206、介质访问控制208、和多目的地应用程序接口(API)210。应当理解的是，图2仅仅提供了对一个实现的说明，并且不意味着关于可以在其中实现不同的实施例的环境的任何限制。可以做出对所描绘的环境的许多修改。在一个实施例中，多目的地API 210可以驻留在存储器204中或驻留在单独的只读存储器(ROM)、可擦除可编程只读存储器(EPROM)、或闪速存储器中。在另一个实施例中，多目的地API 210可以驻留在于主机服务器，并且可以被下载到网络适配器118。

[0034] 多目的地API 210使得网络适配器200能够基于由例如虚拟交换机156的多目的地发送器托管的多目的地群组接收器列表来创建和存储多目的地群组列表。将多目的地操作从例如虚拟交换机156的虚拟交换机卸载到网络适配器200提供了用于多目的地分组处理操作的主机CPU利用的显著减少。将多目的地操作卸载到网络适配器200的另一个优势是改善分组发送延迟，因为处理在硬件级别进行。将多目的地操作卸载到网络适配器200还解决了随着增长的多目的地群组成员列表而产生的可缩放性问题。可缩放性是系统、网络或进程以可行的方式处理日益增长的工作量的能力，或其要被扩大以适应该增长的能力。对于针对每个多目的地群组成员的软件实施的分组复制和头端复制，当多目的地群组成员列表

增加时，针对被传输的每个分组，校验和计算必须被执行，从而在主机CPU上增加高负荷，该高负荷随着多目的地群组成员列表的增长而增长。通过这些操作被卸载到网络适配器200，诸如CPU 106、128和146之类的CPU能够向其他任务提供更多的处理周期。

[0035] 图3是描绘根据本发明的实施例中的向图1的多目的地分组处理环境内的多目的地群组接收器列表添加和移除虚拟交换机的操作步骤的、被总体表示为300的流程图。在这个示例实施例中，服务器142上的虚拟交换机156是MDG-1的多目的地发送器。网络控制器122向虚拟交换机156发送要将虚拟交换机116添加到网络适配器158上的MDG-1接收器的列表的请求，如步骤302所描绘的那样。MDG-1接收器的列表可以被存储在网络适配器158的存储器204中。在步骤304中，响应于接收到该请求，虚拟交换机156调用网络适配器158的多目的地API 210中的一个或多个，以通过将虚拟交换机116添加到网络适配器158上的MDG-1接收器的列表来更新MDG-1。响应于接收到对网络适配器158的多目的地API 210中的所述一个或多个的调用，网络适配器158通过将虚拟交换机116添加到网络适配器158上的MDG-1接收器的列表来更新MDG-1。

[0036] 网络控制器122向虚拟交换机156发送要将虚拟交换机138添加到虚拟交换机156上的MDG-1接收器的列表的请求，如步骤306所描绘的那样。在步骤308中，响应于接收到该请求，虚拟交换机156调用网络适配器158的多目的地API 210中的一个或多个，以通过将虚拟交换机138添加到网络适配器158上的MDG-1接收器的列表来更新MDG-1。响应于接收到对网络适配器158的多目的地API 210中的所述一个或多个的调用，网络适配器158通过将虚拟交换机138添加到网络适配器158上的MDG-1接收器的列表来更新MDG-1。

[0037] 虚拟交换机以类似的方式从多目的地群组被移除。在步骤310中，网络控制器1122向虚拟交换机156发送从MDG-1接收器的列表删除虚拟交换机116的请求。在步骤312中，响应于接收到该请求，虚拟交换机156调用网络适配器158的多目的地API 210中的一个或多个，以通过从网络适配器158上的MDG-1接收器的列表删除虚拟交换机116来更新MDG-1。响应于接收到对网络适配器158的多目的地API 210中的所述一个或多个的调用，网络适配器158通过从网络适配器158上的MDG-1接收器的列表删除虚拟交换机116来更新MDG-1。

[0038] 网络控制器122发送另一个要将虚拟交换机116添加到网络适配器158上的MDG-1接收器的列表的请求，如步骤314所描绘的那样。在步骤316中，响应于接收到该请求，虚拟交换机156调用网络适配器158的多目的地API 210中的一个或多个，以通过将虚拟交换机116添加到网络适配器158上的MDG-1接收器的列表来更新MDG-1。响应于接收到对网络适配器158的多目的地API 210中的所述一个或多个的调用，网络适配器158通过将虚拟交换机116添加到网络适配器158上的MDG-1接收器的列表来更新MDG-1。在网络适配器158中维持诸如MDG-1接收器的列表之类的多目的地群组列表，提供了在多目的地分组处理环境100内无缝地对该列表进行缩放的能力。

[0039] 图4是描绘根据本发明的实施例中的用于将多目的地分组传输给图1的多目的地分组处理环境内的多目的地群组的操作步骤的、被总体表示为400的流程图。在步骤402中，服务器142将用于MDG-1的多目的地分组传输给虚拟交换机156。在一个实施例中，多目的地分组可以来源于虚拟机器152或虚拟机器154。在另一个实施例中，多目的地分组可以来源于存储在持久存储装置148中的mdest软件150。虚拟交换机156调用网络适配器158的多目的地API 210中的一个或多个，以用于网络适配器158将多目的地分组传输给MDG-1接收器

的列表中的虚拟交换机，如步骤404所描述的那样。在步骤406中，响应于为网络适配器158将多目的地分组传输给MDG-1接收器的列表中的虚拟交换机、对网络适配器158的多目的地API 210中的所述一个或多个的调用的接收，网络适配器158通过创建针对MDG-1接收器列表中的每个虚拟交换机的分组的副件来执行针对多目的地分组的头端复制过程。对于MDG-1接收器的列表中的每个虚拟交换机，网络适配器158获得虚拟交换机隧道端点信息。该隧道端点信息可以包含对于目的地虚拟交换机所需要的隧道协议以及虚拟局域网(VLAN)端口群组。VLAN端口群组在初始的VLAN配置过程期间由管理员配置。VLAN端口群组包括物理网络接口卡(NIC)、VLAN信息和合作策略(teaming policy)。这些端口群组决定VLAN流量(traffic)如何通过物理NIC被运送进主机TEP和运送出主机TEP。网络适配器158将针对目的地虚拟交换机的所复制的多目的地分组与隧道协议所特定的报头封装起来。在计算机网络中，封装是设计模块化通信协议的方法，其中网络中的逻辑分离的功能通过包括在较高层的对象内隐藏的信息而从它们底层的结构被抽象。隧道协议允许网络用户访问或提供底层网络不直接提供或支持的网络服务。隧道协议的一个重要的用途是允许外来协议在不支持此特定协议的网络上运行，例如在IPv4上运行IPv6。

[0040] 网络适配器158将每个封装分组传输给其MDG-1接收器的列表中的目的地MDG-1接收器。在步骤408中，网络适配器158向虚拟交换机116发送第一封装分组。在步骤410中，虚拟交换机116接收并且解封分组，识别被包含在分组中的多目的地群组“MDG-1”，然后在端口列表中查找多目的地群组。在计算机联网中，解封分组是移除先前被封装的数据。虚拟交换机用作使用常用的网络适配器集合的端口配置的容器，该常用的网络适配器集合包括完全不包含网络适配器的集合。在一个实施例中，端口列表可以被存储在虚拟交换机中或本地服务器上。每个虚拟交换机提供有限数量的接口，通过这些接口，例如VM 152和VM 154的虚拟机器和网络服务可以到达一个或多个网络。虚拟交换机116将分组转发给服务器102上的网络适配器118上的目的地端口“X”，如步骤412所描绘的那样。

[0041] 在步骤414中，网络适配器158向虚拟交换机138发送第二封装分组。在步骤416中，虚拟交换机138接收并解封分组，并且在端口列表中查找多目的地群组“MDG-1”。虚拟交换机138将分组转发到服务器124上的网络适配器140上的目的地端口“Y”，如步骤418所描绘的那样。

[0042] 图5描绘了在根据本发明的实施例中的服务器计算机的组件的框图。应当理解的是，图5仅仅提供了对一个实现的说明，并且不意味着关于可以实现不同实施例的环境的任何限制。可以做出对所描绘的环境的许多修改。

[0043] 服务器计算机500是服务器计算机102、124和142中的每一个的实施例，并且包括通信结构502，该通信结构502提供一个或多个计算机处理器504、存储器506、持久存储装置508、通信单元510、以及一个或多个输入/输出(I/O)接口512之间的通信。可以用被设计用于在处理器(诸如微处理器、通信和网络处理器等等)、系统存储器、外围设备以及系统内的任何其他硬件组件之间传递数据和/或控制信息的任何体系架构来实施通信结构502。例如，通信结构502可以用一个或多个总线来实施。

[0044] 存储器506和外围存储装置508是计算机可读的存储介质。在这个实施例中，存储器506包括随机存取存储器(RAM)514和高速缓存存储器516。通常，存储器506可以包括任何适当的易失性或非易失性计算机可读存储介质。

[0045] 多目的地软件524(类似于多目的地软件110、132和150中的每一个)经由存储器506中的一个或多个存储器被存储在持久存储装置508中,以供相应的计算机处理器504中的一个或多个进行执行。在这个实施例中,持久存储装置508包括硬磁盘驱动器。可替代地,或除了硬磁盘驱动器以外,持久存储装置可以包括固态硬盘驱动器、半导体存储设备、只读存储器(ROM)、可擦除可编程只读存储器(EPROM)、闪速存储器、或能够存储程序指令或数字信息的任何其他计算机可读存储介质。

[0046] 由持久存储装置508使用的介质还可以是可移除的。例如,可移除硬盘驱动器可以被用于持久存储装置508。其他示例包括光盘和磁盘、拇指驱动器、以及被插入到驱动器以用于向也是持久存储装置508的一部分的另一个计算机可读存储介质进行递送的智能卡。

[0047] 在这些示例中,通信单元510提供与包括诸如网络控制器122之类的网络120的资源在内的其他数据处理系统或设备进行通信。在这些示例中,通信单元510包括一个或多个网络接口卡。通信单元510可以通过使用物理通信链路和无线通信链路中的任一个或两者来提供通信。通过通信单元510,多目的地软件524可以被下载到持久存储装置508。

[0048] 一个或多个I/O接口512允许用可以连接到服务器102的其他设备进行数据的输入和输出。例如,I/O接口512可以提供与外部设备518的连接,外部设备518诸如键盘、小型键盘、触摸屏、和/或一些其他合适的输入设备。外部设备518还可以包括便携式计算机可读介质,诸如拇指驱动器、便携式光盘或磁盘、以及存储器卡。例如多目的地软件524的、用于实行本发明的实施例的软件和数据,可以被存储在这样的便携式计算机可读存储介质上,并且可以经由一个或多个I/O接口512被加载到持久存储装置508。一个或多个I/O接口512还连接到显示器520。显示器520提供向用户显示数据的机构,并且可以是例如计算机监控器。

[0049] 本文所描述的程序是基于在本发明的特定实施例中实施这些程序的应用而被识别的。但是应当理解的是,本文中的任何特定的程序术语仅仅是为了便利而被使用,并且因此本发明不应当被限制为仅仅在由这样的术语识别和/或暗示的任何特定应用中使用。

[0050] 本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质,其上载有用于使处理器实现本发明的各个方面的计算机可读程序指令。

[0051] 计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是——但不限于——电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身,诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如,通过光纤电缆的光脉冲)、或者通过电线传输的电信号。

[0052] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备,或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计

算机可读程序指令，并转发该计算机可读程序指令，以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0053] 用于执行本发明操作的计算机程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码，所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等，以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中，通过利用计算机可读程序指令的状态信息来个性化定制电子电路，例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA)，该电子电路可以执行计算机可读程序指令，从而实现本发明的各个方面。

[0054] 这里参照根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本发明的各个方面。应当理解，流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合，都可以由计算机可读程序指令实现。

[0055] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器，从而生产出一种机器，使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时，产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中，这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0056] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0057] 附图中的流程图和框图显示了根据本发明的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分，所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的是，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

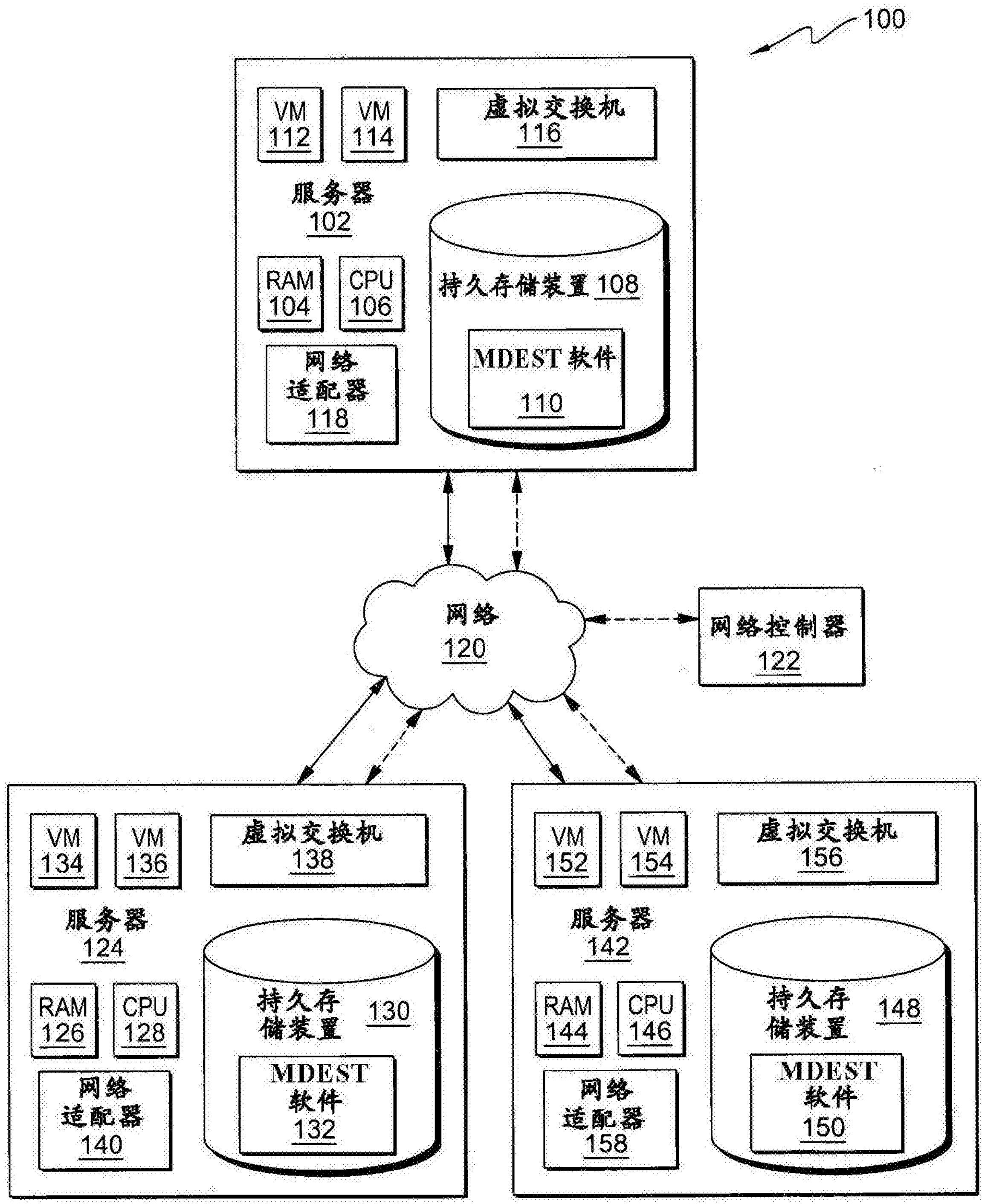

100
VM
112
VM
114
虚拟交换机
116
服务器
102
RAM
104
CPU
106
持久存储装置 108
MDEST 软件
110
网络
适配器
118
网络
120
网络控制器
122
VM
134
VM
136
虚拟交换机
138
服务器
124
RAM
126
CPU
128
持久存
储装置
130
MDEST
软件
132
网络
适配器
140
VM
152
VM
154
虚拟交换机
156
服务器
142
RAM
144
CPU
146
持久存
储装置
148
MDEST
软件
150
网络
适配器
158

图1

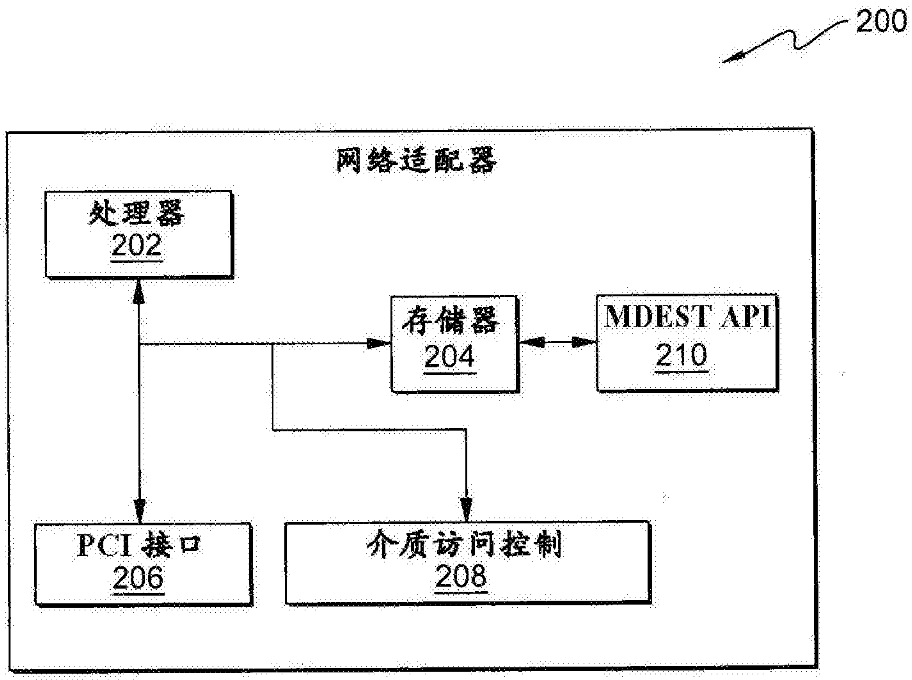


图2

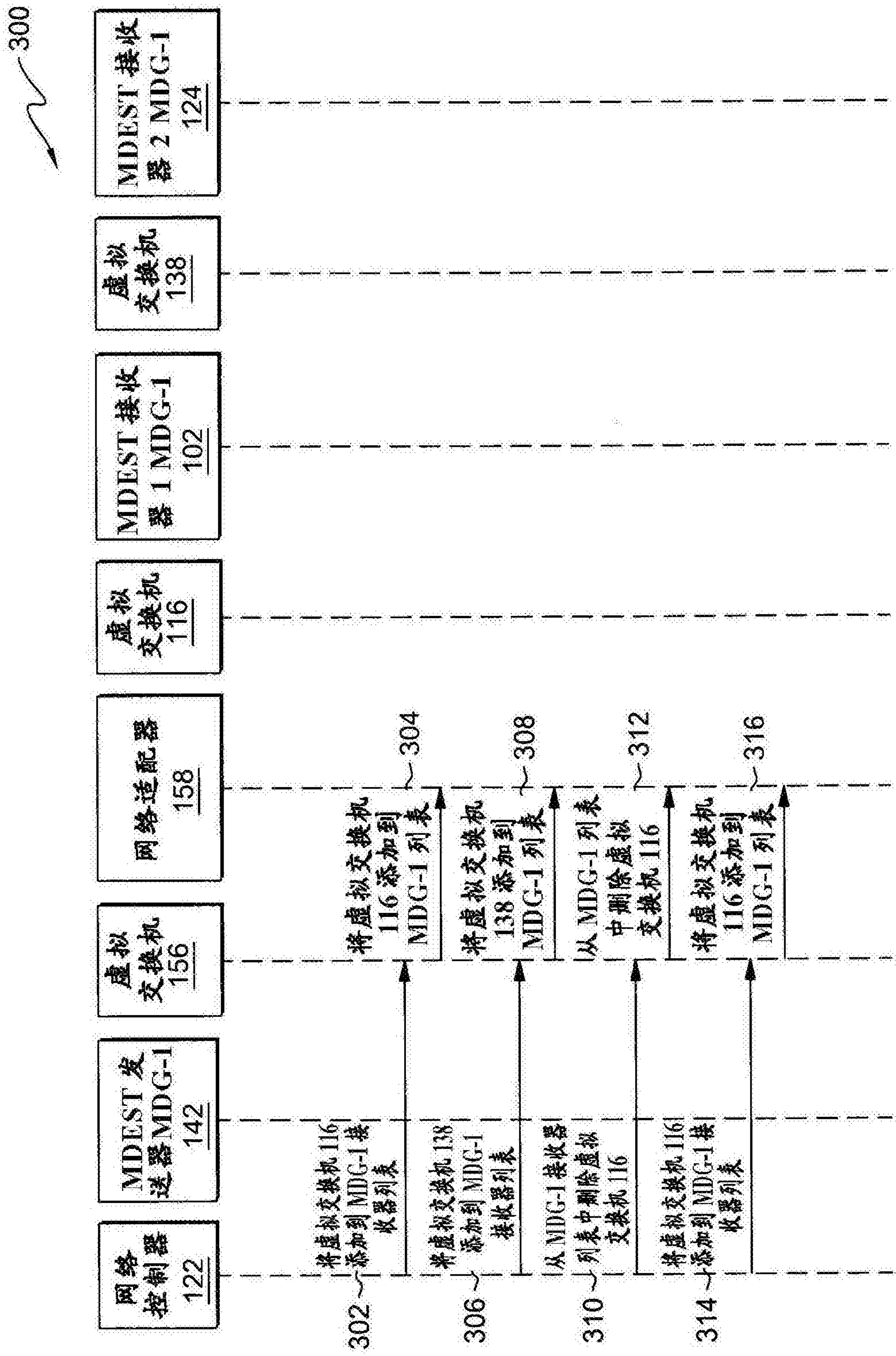

300
网络控制器 122
MDEST 发送器 MDG-1 142
虚拟交换机 156
网络适配器 158
虚拟交换机 116
MDEST 接收器 1 MDG-1 102
虚拟交换机 138
MDEST 接收器 2 MDG-1 124
302 将虚拟交换机 116 添加到 MDG-1 接收器列表
304 将虚拟交换机 116 添加到 MDG-1 列表
306 将虚拟交换机 138 添加到 MDG-1 接收器列表
308 将虚拟交换机 138 添加到 MDG-1 列表
310 从 MDG-1 接收器列表中删除虚拟交换机 116
312 从 MDG-1 列表中删除虚拟交换机 116
314 将虚拟交换机 116 添加到 MDG-1 接收器列表
316 将虚拟交换机 116 添加到 MDG-1 列表

图3

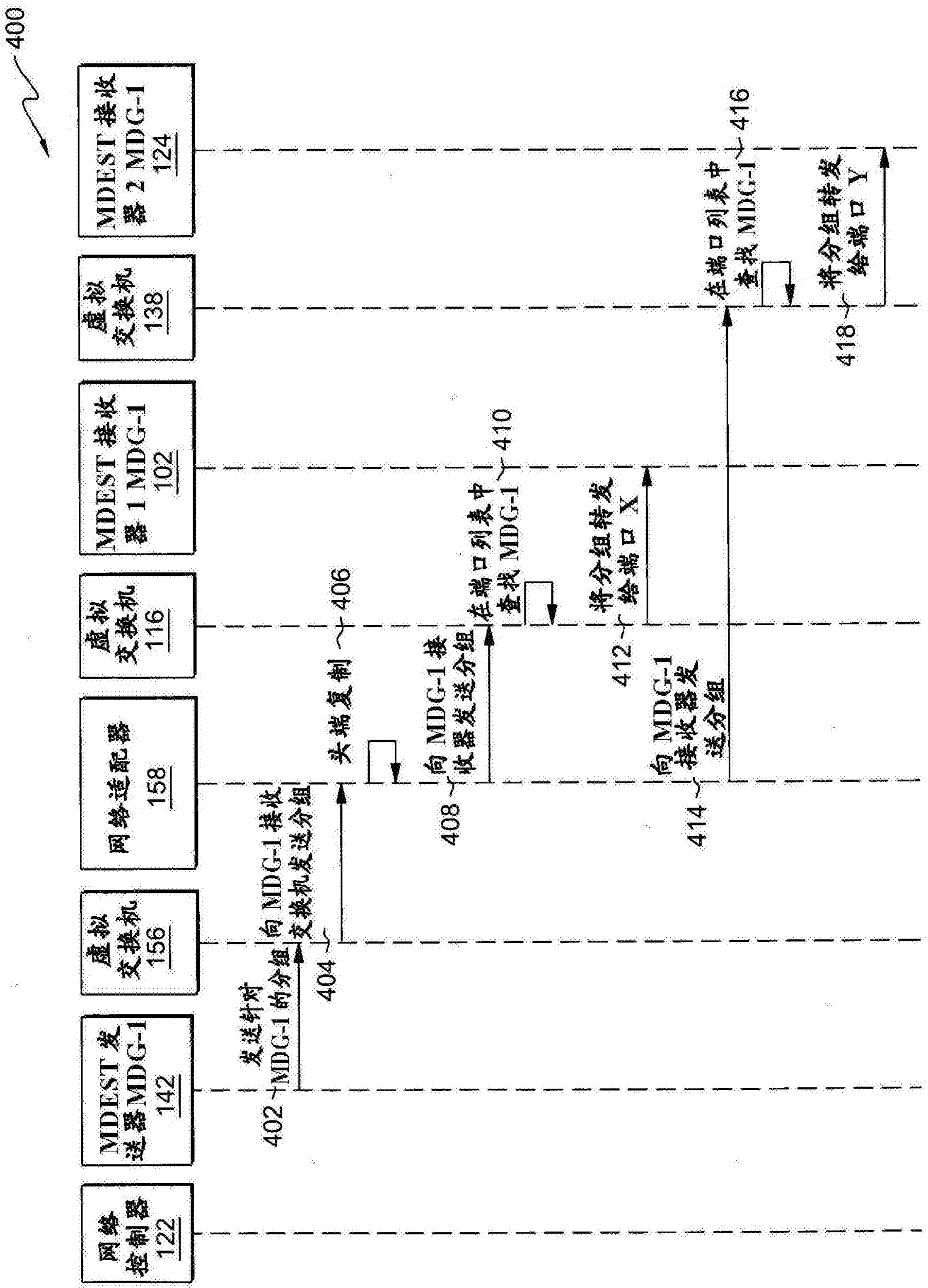

400
网络控制器 122
MDEST发送器MDG-1 142
虚拟交换机 156
网络适配器 158
虚拟交换机 116
MDEST接收器1 MDG-1 102
虚拟交换机 138
MDEST接收器2 MDG-1 124
402 发送针对MDG-1的分组
404 向MDG-1接收交换机发送分组
406 头端复制
408 向MDG-1接收器发送分组
410 在端口列表中查找MDG-1
412 将分组转发给端口X
414 向MDG-1接收器发送分组
416 在端口列表中查找MDG-1
418 将分组转发给端口Y

图4

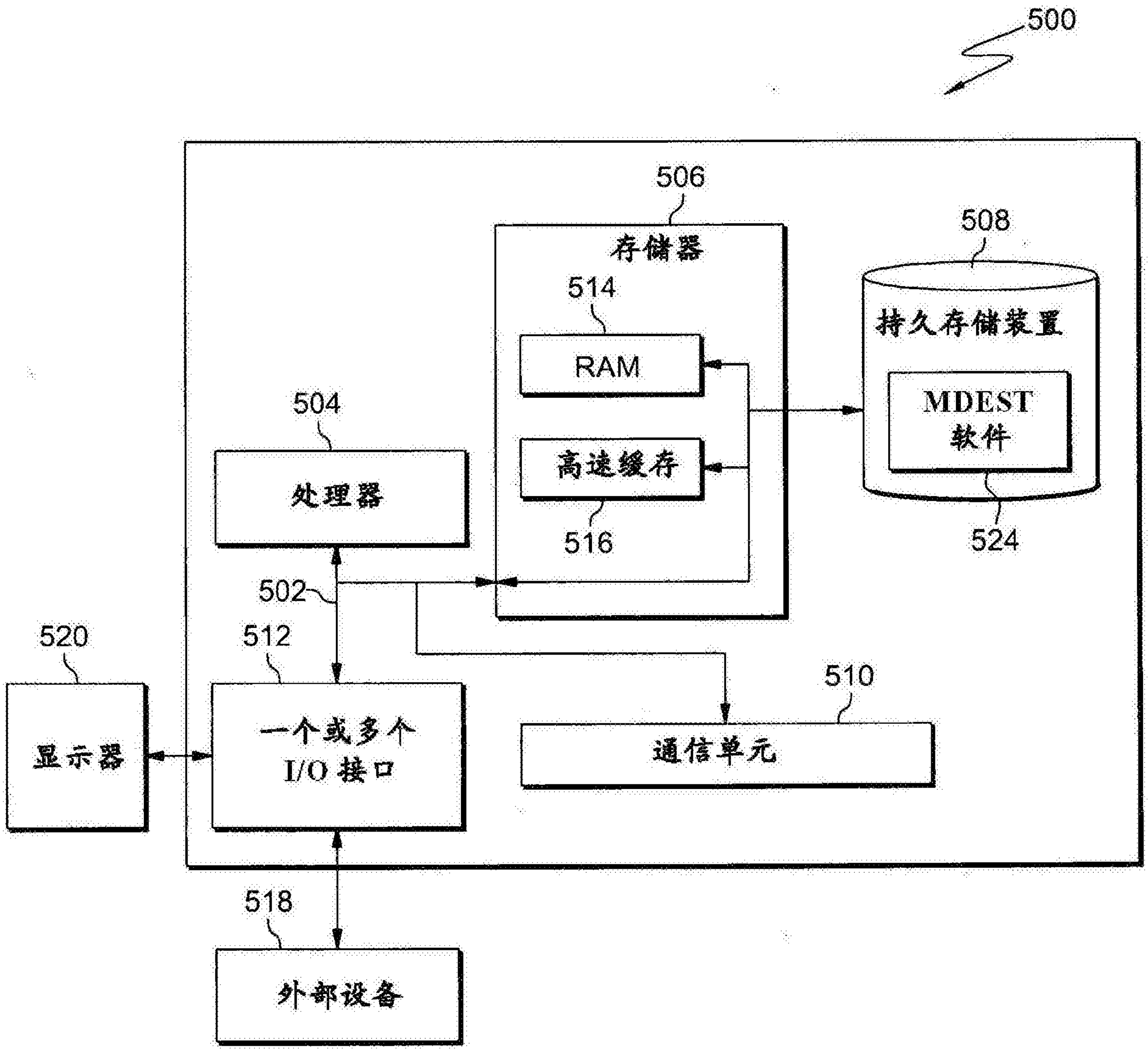

500
506
存储器
514
RAM
高速缓存
516
508
持久存储装置
MDEST
软件
524
504
处理器
502
520
512
显示器
一个或多个
I/O 接口
510
通信单元
518
外部设备

图5