(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 104580124 B
(45)授权公告日 2019.04.05

(21)申请号 201310522423.5

(22)申请日 2013.10.29

(65)同一申请的已公布的文献号
申请公布号 CN 104580124 A

(43)申请公布日 2015.04.29

(73)专利权人 华为技术有限公司
地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 古强 文刘飞

(74)专利代理机构 北京同达信恒知识产权代理有限公司 11291
代理人 黄志华

(51)Int.Cl.
H04L 29/06(2006.01)

(56)对比文件
US 7424710 B1,2008.09.09,
US 2009323691 A1,2009.12.31,
US 2009083756 A1,2009.03.26,
CN 101667144 A,2010.03.10,
CN 101819564 A,2010.09.01,

审查员 李晨

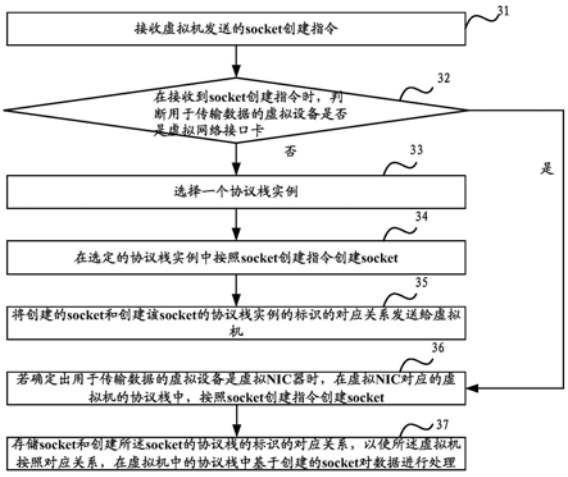
权利要求书3页 说明书19页 附图7页

(54)发明名称

协议栈选择方法、装置及设备

(57)摘要

本发明公开了一种协议栈选择方法、装置及设备，该方法包括：接收虚拟机发送的套接字socket创建指令；选择一个协议栈实例；在选定的所述协议栈实例中按照所述socket创建指令创建socket；将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机，以使所述虚拟机按照所述对应关系，在所述协议栈实例中基于创建的所述socket对数据进行处理。能够使得同一物理主机上设置的多个虚拟机之间可以共享网络协议处理能力，使得各虚拟机的协议栈负载均衡，提高系统的可靠性。



CN 104580124 B

1.一种协议栈选择方法,其特征在于,包括:

接收虚拟机发送的套接字socket创建指令;

选择一个协议栈实例;

在选定的所述协议栈实例中按照所述socket创建指令创建socket;

将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机,以使所述虚拟机按照所述对应关系,在所述协议栈实例中基于创建的所述socket对数据进行处理;

其中,所述选择一个协议栈实例,具体为:至少按照下述方式中的一种选择一个协议栈实例:

按照负载均衡原则,在主机中的至少一个协议栈实例中,选择一个协议栈实例;

确定待根据所述socket创建指令创建的socket对应的应用进程,按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征,按照所述应用进程的业务特征在所述主机中的至少一个协议栈实例中,选择一个协议栈实例;

按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系,在所述主机中的至少一个协议栈实例中,选择一个协议栈实例;

按照待根据所述socket创建指令创建的socket的属性信息,在所述主机中的至少一个协议栈实例中,选择一个协议栈实例;

所述选择一个协议栈实例之前,还包括:

确定待根据所述socket创建指令创建的socket对应的应用进程,按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征,按照所述应用进程的业务特征,确定用于传输数据的虚拟设备不是虚拟NIC。

2.如权利要求1所述的方法,其特征在于,在所述将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机之后,还包括:

接收所述虚拟机传递的对所述socket进行控制操作的控制指令,根据所述socket和所述协议栈实例的标识的对应关系,确定和所述控制指令中的所述socket对应的所述协议栈实例;

在所述协议栈实例上按照所述控制指令对所述socket进行操作;

将对所述socket进行操作的操作结果传输至所述虚拟机。

3.如权利要求2所述的方法,其特征在于,若所述控制指令是基于所述socket在虚拟机之间传输数据,则所述在所述协议栈实例上按照所述控制指令对所述socket进行操作,包括:

将基于所述socket在虚拟机之间传输的数据,通过所述协议栈实例,传输至待传输的虚拟机中的协议栈卸载装置,所述协议栈卸载装置在虚拟机中用于在协议栈实例和虚拟机之间进行数据传输。

4.如权利要求2所述的方法,其特征在于,在所述接收所述虚拟机传递的对所述socket进行控制操作的控制指令之后,还包括:

确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

5.如权利要求1所述的方法,其特征在于,在所述将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机之后,还包括:

接收将所述协议栈实例接收到的数据传输给所述socket对应的应用进程的数据传输指令，根据存储的所述socket和所述协议栈实例标识的对应关系，确定所述数据传输指令中的所述socket对应的所述协议栈实例；

将所述协议栈实例接收到的数据通过所述协议栈实例，传输至所述虚拟机中所述socket对应的应用进程。

6.如权利要求5所述的方法，其特征在于，所述协议栈实例接收到的数据至少是下述数据中的一种：

所述协议栈实例对接收到的数据处理后的处理结果；

所述协议栈实例接收到的同一主机中不同虚拟机之间传输的数据。

7.一种协议栈选择装置，其特征在于，包括：

接收模块，用于接收虚拟机发送的套接字socket创建指令，并将所述socket创建指令传输给选择模块；

所述选择模块，用于选择一个协议栈实例，具体用于按照下述方式中的至少一种选择一个协议栈实例：按照负载均衡原则，在主机中的至少一个协议栈实例中，选择一个协议栈实例；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照待根据所述socket创建指令创建的socket的属性信息，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；

创建模块，用于在选定的所述协议栈实例中按照所述socket创建指令创建socket；

发送模块，用于将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机，以使所述虚拟机按照所述对应关系，在所述协议栈实例中基于创建的所述socket对数据进行处理；

第一确定模块，用于确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征，确定用于传输数据的虚拟设备不是虚拟NIC。

8.如权利要求7所述的装置，其特征在于：

所述接收模块，还用于接收所述虚拟机传递的对所述socket进行控制操作的控制指令；

则，所述装置，还包括：第二确定模块，用于根据所述socket和所述协议栈实例的标识的对应关系，确定和所述控制指令中的所述socket对应的所述协议栈实例；

所述创建模块，还用于在所述第二确定模块确定的所述协议栈实例上按照所述控制指令对所述socket进行操作；

所述发送模块，还用于将对所述socket进行操作的操作结果传输至所述虚拟机。

9.如权利要求8所述的装置，其特征在于，若所述控制指令是基于所述socket在虚拟机之间传输数据，则：所述创建模块，用于在所述协议栈实例上按照所述控制指令对所述socket进行操作为：具体用于将基于所述socket在虚拟机之间传输的数据，通过所述协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置，所述协议栈卸载装置在虚拟机中用

于在协议栈实例和虚拟机之间进行数据传输。

10.如权利要求8所述的装置，其特征在于，所述创建模块，还用于确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

11.如权利要求7～10任一所述的装置，其特征在于，所述接收模块，还用于接收将所述协议栈实例接收到的数据传输给所述socket对应的应用进程的数据传输指令；所述选择模块，还用于根据存储的所述socket和所述协议栈实例标识的对应关系，确定所述数据传输指令中的所述socket对应的所述协议栈实例；

所述发送模块，还用于将所述协议栈实例接收到的数据通过所述协议栈实例，传输至所述虚拟机中所述socket对应的应用进程。

协议栈选择方法、装置及设备

技术领域

[0001] 本发明涉及计算机技术领域，尤其是涉及一种协议栈选择方法、装置及设备。

背景技术

[0002] 在计算机技术领域中，虚拟化技术是一种通过组合或分割现有的计算机资源，使得该些资源表现为一个或多个操作环境，从而提供优于原有资源配置的访问方式的技术。

[0003] 如图1所示，在虚拟化环境下，主要包括虚拟机(英文:Virtual Machine，缩写:VM)、虚拟机管理器(英文:Hypervisor)。虚拟化环境是以物理环境为基础的。也就是说，要实现虚拟化技术，需要物理主机提供运行虚拟化环境的基础。虚拟机是指通过软件模拟的具有完整硬件系统功能的、运行在一个虚拟化环境中的完整计算机系统。在虚拟化环境中，虚拟机管理器用于对虚拟机进行管理，转发虚拟机传输的数据，其中转发虚拟机传输的数据可以是通过虚拟机管理器中的虚拟交换机(Virtual Switch)来完成的，具体如图1中所示的实线传输路径。虚拟机在数据处理过程中，包括创建套接字(英文:socket)、协议处理、协议处理后的数据通过虚拟网络接口卡(英文:Virtual Network Interface Card，缩写:Virtual NIC)传输给虚拟交换机，通过虚拟交换机传输给物理网络接口卡。转发虚拟机传输的数据还可以是由虚拟网络接口卡将处理后的数据直接传输给物理网络接口卡，具体如图1中虚线所示的传输路径。

[0004] 在虚拟化环境下，协议处理功能一般设置在虚拟机内部，称之为虚拟机中的协议栈。由于虚拟机之间具有隔离性，因此，一个物理主机中虽然存在多个协议栈，但由于虚拟机之间的隔离性，虚拟机中的协议栈彼此之间也是隔离的，无法共享，即虚拟机A中的协议栈不能同时为虚拟机B服务。并且由于虚拟机之间的隔离性，虚拟机之间无法共享同一物理主机的网络处理能力。因此在该种情况下，由于虚拟机中的协议栈之间相互隔离，其负载不均衡导致协议处理能力较差，从而使得协议处理能力较差的协议栈可能成为对应的虚拟机的瓶颈。例如图1所示，数据处理过程中，虚拟机1#中的协议栈A的负载很高，已经达到100%满负荷情况，而虚拟机2#中的协议栈B的负载为50%，虚拟机3#中的协议栈C的负载为10%，但是协议处理过程中，协议栈A无法使用协议栈C或协议栈B提供的协议处理能力，当数据分配给协议栈A后，协议栈A只能超负荷运行进行协议处理或者不能进行协议处理。

[0005] 综上所述，在虚拟化环境下，在处理数据时，多个虚拟机之间不可以共享物理主机的网络协议处理能力，当部分虚拟机的协议栈负荷较高的情况下，系统可靠性较差。

发明内容

[0006] 本发明提供了一种协议栈选择方法、装置及设备，能够使得同一物理主机上设置的多个虚拟机之间可以共享网络协议处理能力，使得各虚拟机的协议栈负载均衡，提高系统的可靠性。

[0007] 第一方面，提供了一种协议栈选择方法，包括:接收虚拟机发送的套接字socket创建指令;选择一个协议栈实例;在选定的所述协议栈实例中按照所述socket创建指令创建

socket;将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机,以使所述虚拟机按照所述对应关系,在所述协议栈实例中基于创建的所述socket对数据进行处理。

[0008] 在第一方面的第一种可能的实现方式中,所述选择一个协议栈实例,具体为:至少按照下述方式中的一种选择一个协议栈实例:按照负载均衡原则,在主机中的至少一个协议栈实例中,选择一个协议栈实例;确定待根据所述socket创建指令创建的socket对应的应用进程,按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征,按照所述应用进程的业务特征在所述主机中的至少一个协议栈实例中,选择一个协议栈实例;按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系,在所述主机中的至少一个协议栈实例中,选择一个协议栈实例;按照待根据所述socket创建指令创建的socket的属性信息,在所述主机中的至少一个协议栈实例中,选择一个协议栈实例。

[0009] 在第一方面的第二种可能的实现方式中,所述选择一个协议栈实例之前,还包括:按照下述方式中的一种确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC:按照静态指定的方式,确定用于传输数据的虚拟设备不是虚拟NIC;确定待根据所述socket创建指令创建的socket对应的应用进程,按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征,按照所述应用进程的业务特征,确定用于传输数据的虚拟设备不是虚拟NIC;按照待根据所述socket创建指令创建的socket的属性信息,确定用于传输数据的虚拟设备不是虚拟NIC。

[0010] 结合第一方面以及第一方面的第一种～第二种可能的实现方式,在第一方面的第三种可能的实现方式中,在所述将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机之后,还包括:接收所述虚拟机传递的对所述socket进行控制操作的控制指令,根据所述socket和所述协议栈实例的标识的对应关系,确定和所述控制指令中的所述socket对应的所述协议栈实例;在所述协议栈实例上按照所述控制指令对所述socket进行操作;将对所述socket进行操作的操作结果传输至所述虚拟机。

[0011] 结合第一方面的第三种可能的实现方式,在第一方面的第四种可能的实现方式中,若所述控制指令是基于所述socket在虚拟机之间传输数据,则所述在所述协议栈实例上按照所述控制指令对所述socket进行操作,包括:将基于所述socket在虚拟机之间传输的数据,通过所述协议栈实例,传输至待传输的虚拟机中的协议栈卸载装置,所述协议栈卸载装置在虚拟机中用于在协议栈实例和虚拟机之间进行数据传输。

[0012] 结合第一方面的第三种可能的实现方式,在第一方面的第五种可能的实现方式中,在所述接收所述虚拟机传递的对所述socket进行控制操作的控制指令之后,还包括:确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

[0013] 结合第一方面和第一方面的第一种～第二种可能的实现方式,在第一方面的第六种可能的实现方式中,在所述将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机之后,还包括:接收将所述协议栈实例接收到的数据传输给所述socket对应的应用进程的数据传输指令,根据存储的所述socket和所述协议栈实例标识的对应关系,确定所述数据传输指令中的所述socket对应的所述协议栈实例;将所述协议栈实例接收到的数据通过所述协议栈实例,传输至所述虚拟机中所述socket对应的应用进程。

[0014] 结合第一方面的第六种可能的实现方式,在第一方面的第七种可能的实现方式

中，所述协议栈实例接收到的数据至少是下述数据中的一种：所述协议栈实例对接收到的数据处理后的处理结果；所述协议栈实例接收到的同一主机中不同虚拟机之间传输的数据。

[0015] 第二方面，提供了一种协议栈选择装置，包括：接收模块，用于接收虚拟机发送的套接字socket创建指令，并将所述socket创建指令传输给选择模块；所述选择模块，用于选择一个协议栈实例；创建模块，用于在选定的所述协议栈实例中按照所述socket创建指令创建socket；发送模块，用于将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机，以使所述虚拟机按照所述对应关系，在所述协议栈实例中基于创建的所述socket对数据进行处理。

[0016] 结合第二方面，在第二方面的第一种可能的实现方式中，所述选择模块，具体用于按照下述方式中的至少一种选择一个协议栈实例：按照负载均衡原则，在主机中的至少一个协议栈实例中，选择一个协议栈实例；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照待根据所述socket创建指令创建的socket的属性信息，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例。

[0017] 结合第二方面，在第二方面的第二种可能的实现方式中，还包括：第一确定模块，用于按照下述方式中的一种确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC：按照静态指定的方式，确定用于传输数据的虚拟设备不是虚拟NIC；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征，确定用于传输数据的虚拟设备不是虚拟NIC；按照待根据所述socket创建指令创建的socket的属性信息，确定用于传输数据的虚拟设备不是虚拟NIC。

[0018] 结合第二方面或第二方面的第一种～第二种可能的实现方式，在第二方面的第三种可能的实现方式中，所述接收模块，还用于接收所述虚拟机传递的对所述socket进行控制操作的控制指令；则，所述装置，还包括：第二确定模块，用于根据所述socket和所述协议栈实例的标识的对应关系，确定和所述控制指令中的所述socket对应的所述协议栈实例；所述创建模块，还用于在所述第二确定模块确定的所述协议栈实例上按照所述控制指令对所述socket进行操作；所述发送模块，还用于将对所述socket进行操作的操作结果传输至所述虚拟机。

[0019] 结合第二方面的第三种可能的实现方式，在第二方面的第四种可能的实现方式中，若所述控制指令是基于所述socket在虚拟机之间传输数据，则：所述创建模块，用于在所述协议栈实例上按照所述控制指令对所述socket进行操作为：具体用于将基于所述socket在虚拟机之间传输的数据，通过所述协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置，所述协议栈卸载装置在虚拟机中用于在协议栈实例和虚拟机之间进行数据传输。

[0020] 结合第二方面的第三种可能的实现方式，在第二方面的第五种可能的实现方式中，所述创建模块，还用于确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

[0021] 结合第二方面，或第二方面的第一～第三种任一可能的实现方式，在第二方面的第六种可能的实现方式中，所述接收模块，还用于接收将所述协议栈实例接收到的数据传输给所述socket对应的应用进程的数据传输指令；所述选择模块，还用于根据存储的所述socket和所述协议栈实例标识的对应关系，确定所述数据传输指令中的所述socket对应的所述协议栈实例；所述发送模块，还用于将所述协议栈实例接收到的数据通过所述协议栈实例，传输至所述虚拟机中所述socket对应的应用进程。

[0022] 第三方面，提供了一种协议栈选择设备，包括：接口和信号处理器，接口和信号处理器之间通过总线连接传输数据，接口，用于接收虚拟机发送的套接字socket创建指令，并将所述socket创建指令传输给信号处理器；所述信号处理器，用于选择一个协议栈实例；在选定的所述协议栈实例中按照所述socket创建指令创建socket；接口还用于将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机，以使所述虚拟机按照所述对应关系，在所述协议栈实例中基于创建的所述socket对数据进行处理。

[0023] 结合第三方面，在第三方面的第一种可能的实现方式中，所述信号处理器，具体用于按照下述方式中的至少一种选择一个协议栈实例：按照负载均衡原则，在主机中的至少一个协议栈实例中，选择一个协议栈实例；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照待根据所述socket创建指令创建的socket的属性信息，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例。

[0024] 结合第三方面，在第三方面的第二种可能的实现方式中，所述信号处理器，还用于按照下述方式中的一种确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC：按照静态指定的方式，确定用于传输数据的虚拟设备不是虚拟NIC；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征，确定用于传输数据的虚拟设备不是虚拟NIC；按照待根据所述socket创建指令创建的socket的属性信息，确定用于传输数据的虚拟设备不是虚拟NIC。

[0025] 结合第三方面或第三方面的第一种～第二任一种可能的实现方式，在第三方面的第三种可能的实现方式中，所述接口，还用于接收所述虚拟机传递的对所述socket进行控制操作的控制指令；则，所述装置，所述信号处理器，还用于根据所述socket和所述协议栈实例的标识的对应关系，确定和所述控制指令中的所述socket对应的所述协议栈实例；在所述第二确定模块确定的所述协议栈实例上按照所述控制指令对所述socket进行操作；所述接口，还用于将对所述socket进行操作的操作结果传输至所述虚拟机。

[0026] 结合第三方面的第三种可能的实现方式，在第三方面的第四种可能的实现方式中，若所述控制指令是基于所述socket在虚拟机之间传输数据，则：所述信号处理器，用于在所述协议栈实例上按照所述控制指令对所述socket进行操作为：具体用于将基于所述socket在虚拟机之间传输的数据，通过所述协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置，所述协议栈卸载装置在虚拟机中用于在协议栈实例和虚拟机之间进行数据传输。

[0027] 结合第三方面的第三种可能的实现方式，在第三方面的第五种可能的实现方式中，所述信号处理器，还用于确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

[0028] 结合第三方面，或第三方面的第一～第三种任一可能的实现方式，在第三方面的第六种可能的实现方式中，所述接口，还用于接收将所述协议栈实例接收到的数据传输给所述socket对应的应用进程的数据传输指令；所述信号处理器，还用于根据存储的所述socket和所述协议栈实例标识的对应关系，确定所述数据传输指令中的所述socket对应的所述协议栈实例；所述接口，还用于将所述协议栈实例接收到的数据通过所述协议栈实例，传输至所述虚拟机中所述socket对应的应用进程。

[0029] 本发明提出的技术方案中，在主机中创建协议栈实例，在接收到socket创建指令时，在主机中的协议栈实例中，选定其中一个协议栈实例，然后在选定的协议栈实例上按照socket创建指令创建socket，并将创建结果传输给虚拟机。从而实现在虚拟化环境下，能够使得同一主机上设置的多个虚拟机之间可以共享网络协议处理能力，使得各虚拟机的协议栈负载均衡，提高系统的可靠性。

附图说明

[0030] 图1为通常情况中，提出的在虚拟化环境下协议栈选择系统结构组成示意图；

[0031] 图2a为本发明实施例一中，提出的在虚拟化环境下包含虚拟设备复用器的协议栈选择系统结构组成示意图；

[0032] 图2b为本发明实施例一中，提出的在虚拟化环境下未包含虚拟设备复用器的协议栈选择系统结构组成示意图；

[0033] 图2c为本发明实施例一中，提出的在虚拟化环境下包含虚拟设备复用器和虚拟交换机的协议栈选择系统结构组成示意图；

[0034] 图3为本发明实施例二中，提出的一种协议栈选择方法流程图；

[0035] 图4为本发明实施例二中，提出的一种协议栈选择方法流程图；

[0036] 图5为本发明实施例二中，提出的一种协议栈选择方法流程图；

[0037] 图6a为本发明实施例二中，提出的一种协议栈选择装置结构组成示意图；

[0038] 图6b为本发明实施例二中，提出的一种协议栈选择设备结构组成示意图。

具体实施方式

[0039] 针对一般情况下，在虚拟化环境中处理数据时，多个虚拟机之间不可以共享主机的网络协议处理能力，当部分虚拟机的协议栈负荷较高的情况下，系统可靠性较差的问题，本发明提出的技术方案中，在主机中创建协议栈实例，在接收到创建套接字(英文：socket)的socket创建指令时，在主机中的协议栈实例中，选定其中一个协议栈实例，然后在选定的协议栈实例上按照socket创建指令创建socket，并将创建结果传输给虚拟机。从而实现在虚拟化环境下，能够使得同一主机上设置的多个虚拟机之间可以共享网络协议处理能力，使得各虚拟机的协议栈负载均衡，提高系统的可靠性。

[0040] 下面将结合各个附图对本发明实施例技术方案的主要实现原理、具体实施方式及其对应能够达到的有益效果进行详细地阐述。

[0041] 实施例一

[0042] 本发明实施例一提出一种协议栈选择的系统，该系统结构组成可以包含三种不同的结构组成，分别如图2a，图2b，图2c所示，下面分别对三种不同的系统结构组成进行阐述。

[0043] 如图2a所示，是本发明实施例一提出的第一种虚拟化环境下协议栈选择的系统结构组成示意图，包含主机(图2a中未明确示出)、设置在主机上的至少一个虚拟机、设置在主机上的用于对虚拟机进行管理的虚拟机管理器、设置在主机中的用于提供协议处理功能的至少一个协议栈实例、设置在主机中的用于对NIC进行管理和分配的NIC复用(英文：multiplexer)模块、用于驱动NIC的NIC驱动(英文:driver)模块。

[0044] 其中，主机可以是个人计算机(英文:Person Computer，缩写:PC)、服务器(英文：server)、便携仪式计算机等。主机用于为虚拟机的创建及运行提供硬件环境。

[0045] 设置在主机中的虚拟机，可以是一个，也可以是多个，如果在一个主机中设置有多个虚拟机，设置的多个虚拟机之间可以通过虚拟交换机进行通信。如图1所示的系统架构，以及如图2c所示的系统架构组成图，均包含了虚拟交换机。本发明实施例一提出的三种不同的系统架构中，均以一个主机中设置有三个虚拟机为例来进行详细阐述。具体如图2a，图2b及图2c中所示的虚拟机#1，虚拟机#2和虚拟机#3。针对主机中包含的虚拟机中的任意一个虚拟机来说，在该虚拟机运行时，包含创建的套接字，具体如图2a，图2b，图2c所示的socket #a,socket #b,socket #c,socket #d,。每个虚拟机中可以包含至少一个创建的socket。创建的socket对应socket应用程序接口(英文:application programming interface，缩写:API)。在每个虚拟机中还包含至少一个用于提供协议处理功能的协议栈，以及用于传输数据的虚拟NIC。

[0046] 在本发明实施例一提出的第一种系统架构中，虚拟机中还设置有在对数据进行协议处理时，用于对虚拟设备进行选择的虚拟设备复用器。在虚拟机中，对比图1可知，还增加了协议栈卸载装置(英文:stack offload device)及对应该协议栈卸载装置的驱动。协议栈卸载装置用于将数据不经过任何协议处理，直接发送至虚拟机管理器中，由虚拟机管理器中的协议栈实例对数据进行协议处理。

[0047] 本发明实施例一提出的系统架构中，还包括主机中设置的用于对主机中的虚拟机进行管理调度的虚拟机管理器(英文:hypervisor)。在虚拟机管理器中，设置有用于对设置在虚拟机管理器中的至少一个协议栈实例进行管理和调度的主机协议栈复用器。其中，设置在虚拟机管理器中的协议栈实例中的每一个协议栈实例，可以仅为其中一个虚拟机提供协议处理功能，也可以为多个虚拟机提供协议处理功能。即对于主机中设置的每个协议栈实例来说，可以仅为一个虚拟机提供服务，也可以由多个虚拟机共享使用。

[0048] 主机中的协议栈实例和虚拟机之间的对应关系可以由主机协议栈复用器设定，并维护设定的对应关系。设置在主机中的每个协议栈实例，可以根据需要，与一个或多个NIC中的队列相对应，对应关系可以通过NIC复用器配置并维护。

[0049] 图2b为本发明实施例一提出的第二种系统架构，对比图2a可知，图2b所示的系统架构组成中，在虚拟机中未设置虚拟设备复用器，虚拟设备复用器的功能可以通过socket API来实现。其它结构组成与图2a所示相同。这里不再赘述。

[0050] 图2c为本发明实施例一提出的第三种系统架构，对比图2a可知，图2c所示的系统架构组成中，在图2a所示的系统架构组成中，增加了虚拟交换机(英文:virtual switch)，可以实现同一个主机中的虚拟机之间相互通信。而其他结构组成均相同。这里不再赘述。

[0051] 由上述可知，图2a～图2c所示的系统架构，在结构组成上基本相同，因此本发明实施例一以图2a所示的系统架构，来进一步详细阐述本发明实施例提出的技术方案：

[0052] 可以预先在主机中的虚拟机管理器中创建协议栈实例。多个协议栈实例组成的集群可以称之为协议栈实例集群。在主机中创建协议栈实例之后，可以将创建的协议栈实例的标识发送给主机协议栈复用器中。主机协议栈复用器存储创建的协议栈实例的标识。其中，主机协议栈复用器存储的协议栈实例的标识是唯一能够标识该协议栈实例的编号。该编号可以是创建协议栈实例的顺序编号，也可以是虚拟机的标识和创建协议栈实例的顺序编号的组合形式。

[0053] 虚拟机对数据处理时，主要包含创建socket过程、对创建的socket控制操作过程，以及数据传输过程，下面分别进行阐述。

[0054] 虚拟机中的应用调用socket API，创建socket，虚拟机发送创建socket的socket创建指令。在socket创建指令中包含创建socket的相关信息。其中创建socket的相关信息包括本地互联网协议(英文：Internet Protocol，缩写：IP)地址、本地端口号、协议类型等。还可以包括远端IP地址，远端端口号等。当socket的相关信息中只包括本地IP地址、本地端口号以及协议类型三种信息时，可以称之为socket的三元组信息。当socket的相关信息中既包括本地IP地址、本地端口号以及协议类型三种信息，也包括远端IP地址，远端端口号时，可以称之为socket的五元组信息。

[0055] 虚拟机中的虚拟设备复用器接收socket创建指令，并将该socket创建指令发送给主机协议栈复用器。

[0056] 可选地，还可以在在接收到虚拟机发送的socket创建指令时，确定用于传输数据的虚拟设备是否是虚拟网络接口卡。若确定出用于传输数据的虚拟设备不是虚拟网络接口卡时，将socket创建指令发送给虚拟机中的stack offload device及其驱动。虚拟机中的stack offload device及其驱动接收socket创建指令，并将接收到的socket创建指令发送给主机中的主机协议栈复用器。

[0057] 可以按照下述方式中的至少一种，确定用于传输数据的虚拟设备是否是虚拟网络接口卡：

[0058] 第一种方式：按照静态指定的方式，确定用于传输数据的虚拟设备是否是虚拟NIC。

[0059] 该种方式下，由于是静态指定的方式，如果静态指定用于传输数据的虚拟设备不是虚拟NIC，即使系统中存在虚拟NIC，也始终不选择该虚拟NIC传输数据，而是选择stack offload device及驱动来传输数据。

[0060] 第二种方式：确定待根据socket创建指令创建的socket对应的应用进程，按照该应用进程创建socket的历史信息确定应用进程的业务特征，按照应用进程的业务特征，确定用于传输数据的虚拟设备是否是虚拟NIC。

[0061] 该种方式下，若来自于某socket的应用进程(例如在虚拟机中创建socket的主体)的其他之前创建的socket具有某种特征，例如连接持续时间长、业务吞吐量大等，可根据此类可表现出业务特征的因素，设定策略，并根据策略确定是否选择虚拟NIC。如来自于吞吐量大的应用进程1的新创建的socket，选定虚拟NIC。

[0062] 第三种方式：按照待根据socket创建指令创建的socket的属性信息，确定用于传

输数据的虚拟设备是否是虚拟NIC。

[0063] socket的属性信息包括socket三元组信息或socket五元组信息。该种方式下，可以按照待创建的socket中包含的三元组信息(本地IP地址，本地端口，协议类型)或五元组信息(本地IP地址，本地端口，远端IP地址，远端端口，协议类型)中的全部或部分信息，采用规则匹配的方式(如本地端口为8000～10000范围内，远端端口为80，协议类型为TCP的，由虚拟网络接口卡处理)，确定是否选择虚拟NIC。

[0064] 主机协议栈复用器在接收到socket创建指令时，在主机中的至少一个协议栈实例中，选定其中一个协议栈实例。

[0065] 主机中设置的协议栈实例具有协议处理功能。可以按照下述方式中的至少一种，在主机中的至少一个协议栈实例中，选定其中一个协议栈实例：

[0066] 第一种方式：按照负载均衡原则，在主机中的至少一个协议栈实例中，选择其中一个协议栈实例。

[0067] 该种方式下，需要获知主机中的每个协议栈实例的负荷情况。例如可以根据负荷情况，选择负荷较小的协议栈实例。

[0068] 第二种方式：确定待根据socket创建指令创建的socket对应的应用进程，按照该应用进程创建socket的历史信息确定应用进程的业务特征，按照确定出的应用进程的业务特征在主机中的至少一个协议栈实例中，选择其中一个协议栈实例。

[0069] 该种方式下，若来自于某虚拟机的其他之前创建的socket具有某种特征，如连接持续时间长、业务吞吐量大等，可根据此类可表现出业务特征的因素，设定策略(如来自于吞吐量大的虚拟机1的新创建的socket，选定协议栈实例2)，并根据策略选定其中一个协议栈实例。

[0070] 第三种方式：按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系，采用静态指定的方式，在主机中的至少一个协议栈实例中，选择其中一个协议栈实例。

[0071] 该种方式中，预先在主机中创建协议栈实例时，协议栈实例创建完成后，主机协议栈复用器中会预先存储该创建的协议栈实例的标识和虚拟机的标识之间的对应关系。同时，如果同一个协议栈实例为多个虚拟机共享，则主机协议栈复用器中会存储该协议栈实例的标识和共享的虚拟机的标识之间的对应关系。后续可以按照存储的对应关系，采用静态指定的方式，在主机中的协议栈实例中，选定其中一个协议栈实例。

[0072] 举一例来进行详细阐述，如图2a所示，假设主机中共有四个协议栈实例，分别是协议栈实例#A、协议栈实例#B、协议栈实例#C和协议栈实例#K。主机协议栈复用器中，存储的虚拟机的标识和协议栈实例的标识的对应关系如下：

[0073] 协议栈实例#A----------虚拟机1#，虚拟机2#。

[0074] 协议栈实例#B----------虚拟机1#。

[0075] 协议栈实例#C----------虚拟机2#，虚拟机3#。

[0076] 协议栈实例#K----------虚拟机1#，虚拟机2#，虚拟机3#。

[0077] 对于协议栈实例#A来说，同时为虚拟机1#，虚拟机2#提供服务。对于协议栈实例#B来说，只为虚拟机1#提供服务。协议栈实例#C同时为虚拟机2#，虚拟机3#提供服务。协议栈实例#K同时为虚拟机1#，虚拟机2#，虚拟机3#提供服务。按照上述第二种方式，静态指定的方式，主机协议栈复用器可以根据存储的对应关系分配协议栈实例。

[0078] 第四种方式:按照待根据socket创建指令创建的socket的属性信息,在主机中的至少一个协议栈实例中,选择其中一个协议栈实例。

[0079] 上述四种方式可以结合使用,例如可以根据套接字的三元组信息选定协议栈实例中的某几个协议栈实例,再通过负载均衡的方式,从某几个协议栈实例中选择其中一个协议栈实例。或者综合考虑套接字的属性信息和业务特征,将该些信息通过哈希等方式映射,根据映射结果选定协议栈实例中的一个协议栈实例。

[0080] 在选择出一个协议栈实例后,在选择出的协议栈实例中按照socket创建指令创建socket。并将创建的socket和创建socket的协议栈实例的标识的对应关系传输给发送socket创建指令的虚拟机,以使该虚拟机按照接收到的对应关系,在选定的协议栈实例中基于创建的socket对数据进行处理。

[0081] 例如,Socket创建完成后,将创建的socket和创建socket的协议栈实例的标识的对应关系发送给主机协议栈复用器。主机协议栈复用器维护接收到的创建的socket和创建socket的协议栈实例的标识的对应关系,并将创建的socket和创建socket的协议栈实例标识的对应关系传输给发送socket创建指令的stack offload device及其驱动,stack offload device及其驱动将创建socket的结果传输至虚拟机中的虚拟设备复用器。

[0082] 虚拟设备复用器接收创建socket的结果,并维护创建的socket和创建socket的协议栈实例的标识的对应关系,以及将创建结果反馈给虚拟机中的应用。

[0083] 上述创建socket的过程,是在确定出用于传输数据的虚拟设备不是虚拟网络接口卡的前提下来实施的,当确定出用于传输数据的虚拟设备是虚拟网络接口卡时,创建socket过程为:

[0084] 在接收到虚拟机发送的创建套接字socket的socket创建指令时,若确定出用于传输数据的虚拟设备是虚拟NIC时,在虚拟NIC对应的虚拟机的协议栈中,按照socket创建指令创建socket。存储socket和创建socket的协议栈的标识的对应关系,以使虚拟机按存储的对应关系,在协议栈中基于创建的socket对数据进行处理。

[0085] 同样地,虚拟机中的协议栈将socket创建完成后,将创建结果发送给虚拟机中的虚拟设备复用器。虚拟设备复用器接收创建socket的结果,并维护创建的socket和创建socket的协议栈的标识的对应关系,以及将创建结果反馈给虚拟机中的应用。

[0086] 其中,确定用户传输数据的虚拟设备是否是虚拟NIC的具体实施方式请参见上文中的详细阐述,这里不再赘述。

[0087] 在上述创建socket的过程中,在虚拟化环境下,当主机中的某一虚拟机中的协议栈负荷较大时,可以同时和多个主机中的协议栈实例相关联,创建socket的过程可以由主机中的协议栈实例来完成,多个虚拟机之间可以共享主机的网络协议处理能力,避免在虚拟机中的协议栈负荷较高的情况下,虚拟机中的协议栈成为虚拟机的瓶颈,使得系统可靠性较差的问题。并且,主机中的协议栈实例可以利用主机中的信号处理器直接进行协议处理,避免由于信号处理器仿真、设备仿真等带来的性能损失。

[0088] 虚拟机在处理数据的过程中,还包括对创建的socket进行控制操作的过程。控制操作包括绑定(英文:bind)、监听(英文:listen)、连接(英文:connect)、关闭(英文:close)等操作,还可以包括数据接收、数据发送类操作,如接收(英文:receive)、发送(英文:send)等。

[0089] 由于创建socket的过程有两种形式，一种是本发明实施例一上述提出的创建socket的方法，还有一种方式是一般情况下创建的socket，即不采用本发明实施例一上述提出的创建socket的方法。基于此，对创建的socket进行控制操作的过程，也可以分为两种形式，一种方式是对本发明实施例一上述提出的创建的socket进行控制操作的过程，还有一种方式是对一般情况下创建的socket进行控制操作，下面分别进行阐述。

[0090] 第一种方式：对本发明实施例一上述提出的创建的socket进行控制操作的过程。

[0091] 若虚拟机接收到对socket进行控制操作的控制指令时，根据socket和创建socket的协议栈标识的对应关系，确定和控制指令中的socket对应的协议栈实例标识对应的协议栈实例，在确定出的协议栈实例上按照控制指令对所述socket进行操作；并将对socket进行操作的操作结果传输至所述虚拟机中。

[0092] 该种方式下，虚拟机中的应用调用socket API，操作创建的socket，发送对socket进行控制操作的控制指令。

[0093] 虚拟机中的虚拟设备复用器接收控制指令，并在接收到虚拟机传递的对socket进行控制操作的控制指令时，确定用于传输数据的虚拟设备是否是虚拟网络接口卡。

[0094] 其中，确定用于传输数据的虚拟设备是否是虚拟网络接口卡的具体实施方式请参见上文中的详细阐述，这里不再赘述。

[0095] 若确定出用于传输数据的虚拟设备不是虚拟NIC时，虚拟设备复用器将该控制指令发送给虚拟机中的stack offload device及其驱动。

[0096] 虚拟机中的stack offload device及其驱动将接收到的控制指令传输给主机协议栈复用器。

[0097] 主机协议栈复用器根据存储的socket和创建socket的协议栈实例的标识的对应关系，确定和控制指令中的socket对应的协议栈实例的标识对应的主机中的协议栈实例，在确定出的主机中的协议栈实例上按照接收到的控制指令对socket进行操作，并将对该socket进行操作的操作结果传输至虚拟机中的stack offload device及其驱动。

[0098] 虚拟机中的stack offload device及其驱动将接收到的对socket进行操作的操作结果传输至虚拟机中的虚拟设备复用器中。

[0099] 虚拟设备复用器将接收到的对socket进行操作的操作结果反馈给虚拟机中socket对应的应用进程。

[0100] 其中，若对socket进行控制操作的控制指令是基于socket在虚拟机之间传输数据，则在该种情况下，可以将基于socket在虚拟机之间传输的数据，通过确定出的协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置中，协议栈卸载装置是在虚拟机中的用于在协议栈实例和虚拟机之间进行数据传输。

[0101] 其中，如果需要将操作结果传输至同一主机中的其它虚拟机时，则主机协议栈复用器在获得对socket进行操作的操作结果之后，将获得的操作结果反馈给其他虚拟机中的协议栈卸载装置，由其他虚拟机中的协议栈卸载装置传输给虚拟机中对应的应用进程。

[0102] 举一例来进行说明，假设发送控制指令的虚拟机为源虚拟机，其他虚拟机为目的虚拟机。源虚拟机和目的虚拟机设置在同一主机中。如果需要将协议栈实例的操作结果传输至目的虚拟机，而不是传输给源虚拟机，则按照本发明实施例一提出的技术方案，主机协议栈复用器在获得对socket进行操作的操作结果之后，将获得的操作结果反馈给目的虚拟

机中的协议栈卸载装置，由目的虚拟机中的协议栈卸载装置传输给虚拟机中对应的应用进程。

[0103] 上述对创建的socket操作的过程，是在确定出用于传输数据的虚拟设备不是虚拟网络接口卡的前提下来实施的，当确定出用于传输数据的虚拟设备是虚拟网络接口卡时，对创建的socket过程为：

[0104] 若虚拟机接收到对socket进行控制操作的控制指令，且确定出用于传输数据的虚拟设备是虚拟NIC时，在确定出的虚拟NIC对应的虚拟机的协议栈中，按照控制指令对socket进行操作。并将操作结果传输是虚拟机中的虚拟设备复用器中。

[0105] 虚拟设备复用器将接收到的对socket进行操作的操作结果反馈给虚拟机中socket对应的应用进程。

[0106] 第二种方式：对一般情况下创建的socket进行控制操作。

[0107] 虚拟机中的应用调用socket API，操作创建的socket，发送对socket进行控制操作的控制指令。该种方式下，创建的socket并不是按照本发明实施例一上述提出的创建socket的过程来创建的，而是按照一般情况下进行处理创建得到的socket。

[0108] 该种方式下，除控制操作的socket的形式不一样之外，其它处理过程同一种方式相同，具体请参见上述第一种方式中的详细阐述，这里不再赘述。

[0109] 上述两种方式中，阐述了对创建的socket进行控制操作，以及将操作结果反馈给虚拟机中的应用进程。在虚拟化环境下，当主机中的某一虚拟机中的协议栈负荷较大时，可以同时和多个主机中的协议栈实例相关联，对socket控制操作的过程可以由主机中的协议栈实例来完成，多个虚拟机之间可以共享主机的网络协议处理能力，避免在虚拟机中的协议栈负荷较高的情况下，虚拟机中的协议栈成为虚拟机的瓶颈，使得系统可靠性较差的问题。并且，主机协议栈可以利用主机中的信号处理器直接进行协议处理，避免由于信号处理器仿真、设备仿真等带来的性能损失。

[0110] 虚拟机在处理数据的过程中，还包括对数据的传输。例如将网络事件或者控制操作的结果上报给应用进程。

[0111] 由于创建socket的过程有两种形式，一种是本发明实施例一上述提出的创建socket的方法，还有一种方式是一般情况下创建的socket，即不采用本发明实施例一上述提出的创建socket的方法。基于此，数据传输过程，也可以分为两种形式，一种方式是基于本发明实施例一上述提出的创建的socket，进行数据传输的过程，还有一种方式是基于一般情况下socket进行数据传输的过程，下面分别进行阐述。

[0112] 第一种方式：基于本发明实施例一上述提出的创建的socket，进行数据传输的过程。

[0113] 在将创建的socket和创建该socket的协议栈实例的标识的对应关系发送给虚拟机之后，在接收到将协议栈实例接收到的数据传输给socket对应的应用进程的数据传输指令时，根据存储的socket和协议栈实例标识的对应关系中，确定数据传输指令中的socket对应的协议栈实例标识；将协议栈实例接收到的数据通过确定出的协议栈实例标识对应的协议栈实例，传输至虚拟机中socket对应的应用进程。

[0114] 其中协议栈实例接收到的数据至少是下述中的一种：

[0115] 第一种：协议栈实例对接收到的数据处理后的处理结果。

[0116] 第二种:协议栈实例接收到的虚拟机中的协议栈对接收到的数据处理后的处理结果。

[0117] 第三种:协议栈实例接收到的同一主机中不同虚拟机之间传输的数据。

[0118] 其中,数据传输至主机中的协议栈实例。例如,网络事件或者其他数据传输至主机中的某一协议栈实例中。

[0119] 主机协议栈复用器获得主机中的协议栈实例中的数据,对获得的数据进行处理,需要将处理后的数据传输给虚拟机中的socket对应的应用进程。主机协议栈复用器在存储的socket和协议栈卸载装置标识的对应关系中,确定虚拟机中的socket对应的协议栈卸载装置标识对应的协议栈卸载装置,将接收到的数据传输给确定出的协议栈卸载装置。

[0120] 协议栈卸载装置将接收到的数据传输至该虚拟机中socket对应的应用进程。

[0121] 第二种方式:基于一般情况下提出的创建的socket,进行数据传输的过程。

[0122] 该种方式下,除socket的形式不一样之外,其它处理过程同一种方式相同,具体请参见上述第一种方式中的详细阐述,这里不再赘述。

[0123] 上述两种方式中,阐述了将数据传输给虚拟机中的应用进程。在虚拟化环境下,当主机中的某一虚拟机中的协议栈负荷较大时,可以同时和多个主机中的协议栈实例相关联,因此可以由主机中的协议栈实例来完成,多个虚拟机之间可以共享主机的网络协议处理能力,避免在虚拟机中的协议栈负荷较高的情况下,虚拟机中的协议栈成为虚拟机的瓶颈,使得系统可靠性较差的问题。并且,协议栈实例可以利用主机中的信号处理器直接进行协议处理,避免由于信号处理器仿真、设备仿真等带来的性能损失。

[0124] 采用上述技术方案,对于同一主机中的虚拟机之间进行通信时,通过主机中的协议栈实例共享的方式,首先减少了数据在同一主机中的虚拟机之间进行传输时,数据拷贝的次数和协议栈的处理过程,进一步提高了系统的性能。其主要原因为:一般情况下,数据在同一主机中的虚拟机之间进行传输时,如图1所示的系统架构中,数据在虚拟机#1和虚拟机#3之间传输时,需要将数据拷贝到虚拟机#1中的协议栈中,然后再将虚拟机#1中的协议栈中的数据拷贝至虚拟交换机中,再将虚拟交换机中的数据拷贝至虚拟机#3中的协议栈中,最后将虚拟机#3中的协议栈中数据拷贝给虚拟机#3对应的应用。共需要四次数据拷贝,以及两次需要两次协议处理过程。而采用本发明实施例一上述提出的技术方案,数据在同一主机中的虚拟机#1和虚拟机#3中之间进行传输时,如图2a所示的系统架构中,仅需要将虚拟机#1的数据传输给主机中的协议栈实例,然后主机中的协议栈实例中的数据传输给虚拟机#3对应的应用。即数据由虚拟机拷贝至主机,处理后,再拷贝至目标虚拟机。如果数据不需要进行协议处理,则可以不进行协议处理,直接拷贝至目标虚拟机。减少了数据拷贝的次数和协议栈的处理过程,进一步提高了系统的性能。

[0125] 其次在数据传输过程中,减少了虚拟中断的数量,降低了中断处理的负担。例如,对于基于传输控制协议(英文:Transmission Control Protocol,缩写:TCP)类的应用,由于基于TCP的应用,是面向流的应用,多个小的数据包可以合并为一个较大的数据包,通过本发明实施例一上述提出的主机中的协议栈实例,会将部分数据包合并后,再以中断的方式上报给虚拟机的协议栈卸载设备,这样可避免多个小数据包造成的多次中断,减轻中断处理带来的负载。

[0126] 本发明实施例一上述提出的技术方案中,是以图2a所示的系统架构进行详细阐述

的，具体实施中，还可以以图2b和图2c所示的系统架构实施本发明实施例一上述提出的技术方案。其中：

[0127] 若以图2b所示的系统架构实施本发明实施例一上述提出的技术方案，将图2a中提出的虚拟机中的虚拟设备复用器的功能，可以通过修改API接口来实现。例如可以修改标准的socket API接口，融入虚拟设备复用器的功能，感知虚拟机中虚拟网络和虚拟协议栈卸载设备，可为应用提供选择虚拟设备的接口，如图2b中虚拟机2#中的实线路径。通过修改标准接口，也可建立同一主机上不同虚拟机之间的虚拟直通链路，不经过协议栈实例，直接通过主机协议栈复用器，以内存拷贝/映射方式实现。如图2b中所示的虚线传输路径。也可不修改标准socket API接口，但在API实现时融合虚拟设备复用器的功能，应用不感知选择虚拟设备的操作。除此之外，虚拟机对数据的处理过程同图2a所示的系统架构基本相同，这里不再赘述。

[0128] 若以图2c所示的系统架构实施本发明实施例一上述提出的技术方案，在图2a所示的系统架构上增加虚拟交换机，主机协议栈复用器可以选择是否和虚拟交换机连接，从而支持主机软件层和虚拟机之间的数据收发，主机中的协议栈实例与虚拟NIC可混合接入同一虚拟交换机的不同端口。除此之外，虚拟机对数据的处理过程同图2a所示的系统架构相同，这里不再赘述。

[0129] 实施例二

[0130] 本发明实施例提出一种协议栈选择方法，虚拟机对数据处理时，主要包含创建socket过程、对创建的socket控制操作过程，以及数据传输过程，下面分别进行阐述。在主机中设置至少一个虚拟机，以及在主机上创建至少一个具有协议处理功能的协议栈实例，虚拟机使用至少一个协议栈实例的协议处理功能，

[0131] 如图3所示，为本发明实施例提出的创建socket过程，其具体处理过程如下述：

[0132] 步骤31，接收虚拟机发送的socket创建指令。

[0133] 虚拟机中的应用调用socket API，创建socket，发送创建socket的socket创建指令。在socket创建指令中包含创建socket的相关信息。

[0134] 其中，创建socket的相关信息包括三、五元组信息，具体请参见上述实施例一中的详细阐述，本步骤中不再赘述。

[0135] 步骤32，在接收到socket创建指令时，判断用于传输数据的虚拟设备是否是虚拟网络接口卡。如果判断结果为否，则执行步骤33，反之，如果判断结果为是，则执行步骤36。

[0136] 具体地，确定用于传输数据的虚拟设备是否是虚拟网络接口卡的具体实施方式请参见上述实施例一中的详细阐述，本发明实施例二不再赘述。

[0137] 其中，步骤32是一种可选的执行过程，具体实施时，步骤32可以不执行，直接执行步骤33。

[0138] 步骤33，选择一个协议栈实例。

[0139] 在主机中的至少一个协议栈实例中，选定其中一个协议栈实例。

[0140] 该种情况下，确定出用于传输数据的虚拟设备不是虚拟NIC时，在主机中的至少一个协议栈实例中，选定其中一个协议栈实例。其中，选择协议栈实例的具体处理过程请参见上述实施例一中的详细阐述，本发明实施例二不再赘述。

[0141] 步骤34，在选定的协议栈实例中按照socket创建指令创建socket。

[0142] 步骤35，将创建的socket和创建该socket的协议栈实例的标识的对应关系发送给虚拟机，以使在选定的协议栈实例中基于创建的socket对数据进行处理。

[0143] 步骤36，若确定出用于传输数据的虚拟设备是虚拟NIC器时，在虚拟NIC对应的虚拟机的协议栈中，按照socket创建指令创建socket。

[0144] 步骤37，存储socket和创建所述socket的协议栈的标识的对应关系，以使所述虚拟机按照对应关系，在虚拟机中的协议栈中基于创建的socket对数据进行处理。

[0145] 在上述创建socket的过程中，在虚拟化环境下，当主机中的某一虚拟机中的协议栈负荷较大时，可以同时和多个主机中的协议栈实例相关联，创建socket的过程可以由主机中的协议栈来完成，多个虚拟机之间可以共享主机的网络协议处理能力，避免在虚拟机中的协议栈负荷较高的情况下，虚拟机中的协议栈成为虚拟机的瓶颈，使得系统可靠性较差的问题。并且，协议栈实例可以利用主机中的信号处理器直接进行协议处理，避免由于信号处理器仿真、设备仿真等带来的性能损失。

[0146] 虚拟机在处理数据的过程中，还包括对创建的socket进行控制操作的过程。控制操作包括bind、listen、connect、close等操作，还可以包括数据接收、数据发送类操作，如receive、send等。

[0147] 由于创建socket的过程有两种形式，一种是本发明实施例二上述提出的创建socket的方法，还有一种方式是一般情况下创建的socket，即不采用本发明实施例二上述提出的创建socket的方法。基于此，对创建的socket进行控制操作的过程，也可以分为两种形式，一种方式是对本发明实施例二上述提出的创建的socket进行控制操作的过程，还有一种方式是对一般情况下创建的socket进行控制操作，两种方式的区别仅在于socket的形式不同，但是对socket控制操作的处理流程相同，在将创建的socket和创建所述socket的协议栈实例的标识的对应关系传输给发送socket创建指令的虚拟机之后，如图4所示，其具体处理流程如下述：

[0148] 步骤41，接收虚拟机传递的对socket进行控制操作的控制指令。

[0149] 虚拟机中的应用调用socket API，操作创建的socket，发送对socket进行控制操作的控制指令。

[0150] 步骤42，确定用于传输数据的虚拟设备是否是虚拟网络接口卡。如果判断结果为否，执行步骤43，反之执行步骤46。

[0151] 其中，确定用于传输数据的虚拟设备是否是拟网络接口控制器的具体实施方式请参见上述实施例一中的详细阐述，本发明实施例二不再赘述。

[0152] 步骤42是一个可选地步骤，在具体实施时，可以直接执行步骤43，而不进行虚拟设备的判断。

[0153] 步骤43，若确定出用于传输数据的虚拟设备不是NIC时，根据socket和创建所述socket的协议栈实例的标识的对应关系，确定和控制指令中的socket对应的协议栈实例标识所对应的协议栈实例。

[0154] 步骤44，在确定出的主机中的协议栈实例上按照控制指令对socket进行操作。

[0155] 步骤45，将对socket进行操作的操作结果传输至所述虚拟机中。

[0156] 可以将对socket进行操作的操作结果传输至所述虚拟机中socket对应的应用进程。

[0157] 其中，对socket进行控制操作的控制指令是基于socket在虚拟机之间传输数据；将基于socket在虚拟机之间传输的数据，通过确定出的协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置中，协议栈卸载装置是虚拟机中的用于传输协议栈实例和虚拟机之间的数据的虚拟设备。

[0158] 举一例来进行说明，假设发送控制指令的虚拟机为源虚拟机，其他虚拟机为目的虚拟机。源虚拟机和目的虚拟机设置在同一主机中。如果需要将操作结果传输至目的虚拟机，而不是传输给源虚拟机，则按照本发明实施例一提出的技术方案，主机协议栈复用器在获得对socket进行操作的操作结果之后，将获得的操作结果反馈给目的虚拟机中的协议栈卸载装置，由目的虚拟机中的协议栈卸载装置传输给虚拟机中对应的应用进程。

[0159] 步骤46，若确定出用于传输数据的虚拟设备是虚拟NIC时，在确定出的虚拟NIC对应的虚拟机的协议栈中，按照控制指令对socket进行操作。

[0160] 步骤47，将操作结果反馈给虚拟机中与socket对应的应用进程。

[0161] 上述步骤41～步骤47的执行过程中，可以作为一个独立的对socket进行控制操作的流程来执行，也可以是在上述步骤35将创建的socket和创建socket的协议栈标识的对应关系传输给发送socket创建指令的设置在主机上的虚拟机之后，依次执行步骤41～步骤47。

[0162] 在上述步骤41～步骤47中，阐述了对创建的socket进行控制操作，以及将操作结果反馈给虚拟机中的应用进程。在虚拟化环境下，当主机中的某一虚拟机中的协议栈负荷较大时，可以同时和多个主机中的协议栈实例相关联，对socket控制操作的过程可以由主机中的协议栈来实例完成，多个虚拟机之间可以共享主机的网络协议处理能力，避免在虚拟机中的协议栈负荷较高的情况下，虚拟机中的协议栈成为虚拟机的瓶颈，使得系统可靠性较差的问题。并且，协议栈实例可以利用主机中的信号处理器直接进行协议处理，避免由于信号处理器仿真、设备仿真等带来的性能损失。

[0163] 虚拟机在处理数据的过程中，还包括对数据的传输。例如将网络事件或者控制操作的结果上报给应用进程。

[0164] 由于创建socket的过程有两种形式，一种是本发明实施例一上述提出的创建socket的方法，还有一种方式是一般情况下创建的socket，即不采用本发明实施例一上述提出的创建socket的方法。基于此，数据传输过程，也可以分为两种形式，一种方式是基于本发明实施例一上述提出的创建的socket，进行数据传输的过程，还有一种方式是基于一般情况下基于socket进行数据传输的过程，两种方式的区别仅在于socket的形式不同，但是基于socket进行数据传输的处理流程相同，如图5所示，其具体处理流程如下述：

[0165] 步骤51，接收将协议栈实例接收到的数据传输给socket对应的应用进程的数据传输指令。

[0166] 其中，接收到的数据至少是下述中的一种：

[0167] 第一种：协议栈实例对接收到的数据处理后的处理结果。

[0168] 第二种：协议栈实例接收到的虚拟机中的协议栈对接收到的数据处理后的处理结果。

[0169] 第三种：协议栈实例接收到的同一主机中不同虚拟机之间传输的数据。

[0170] 步骤52，根据存储的socket和协议栈实例标识的对应关系，确定数据传输指令中

的socket对应的协议栈实例。

[0171] 步骤53，将协议栈实例接收到的数据通过确定出的协议栈实例，传输至虚拟机中socket对应的应用进程。

[0172] 可以将接收到的数据通过协议栈实例传输给协议栈卸载装置，通过协议栈卸载装置传输至虚拟机中socket对应的应用进程。

[0173] 在虚拟化环境下，当主机中的某一虚拟机中的协议栈负荷较大时，可以同时和多个主机中的协议栈实例相关联，因此可以由主机中的协议栈实例来完成，多个虚拟机之间可以共享主机的网络协议处理能力，避免在虚拟机中的协议栈负荷较高的情况下，虚拟机中的协议栈成为虚拟机的瓶颈，使得系统可靠性较差的问题。并且，协议栈实例可以利用主机中的信号处理器直接进行协议处理，避免由于信号处理器仿真、设备仿真等带来的性能损失。

[0174] 采用上述技术方案，对于同一主机中的虚拟机之间进行通信时，通过主机中的协议栈共享的方式，首先减少了数据在同一主机中的虚拟机之间进行传输时，数据拷贝的次数和协议栈的处理过程，进一步提高了系统的性能。其主要原因为：一般情况下，数据在同一主机中的虚拟机之间进行传输时，如图1所示的系统架构中，数据在虚拟机#1和虚拟机#3之间传输时，需要将数据拷贝到虚拟机#1中的协议栈中，然后再将虚拟机#1中的协议栈中的数据拷贝至虚拟交换机中，再将虚拟交换机中的数据拷贝至虚拟机#3中的协议栈中，最后将虚拟机#3中的协议栈中数据拷贝给虚拟机#3对应的应用。共需要四次数据拷贝，以及两次需要两次协议处理过程。才用本发明实施例一上述提出的技术方案，数据在同一主机中的虚拟机#1和虚拟机#3中之间进行传输时，如图2a所示的系统架构中，仅需要将虚拟机#1的数据传输给主机中的协议栈实例，然后主机中的协议栈实例中的数据传输给虚拟机#3对应的应用。即数据由虚拟机拷贝至主机，处理后，再拷贝至目标虚拟机。如果数据不需要进行协议处理，则可以不进行协议处理，直接拷贝至目标虚拟机。减少了数据拷贝的次数和协议栈的处理过程，进一步提高了系统的性能。

[0175] 其次在数据传输过程中，减少了虚拟中断的数量，降低了中断处理的负担。例如，对于基于传输控制协议(英文：Transmission Control Protocol，缩写：TCP)类的应用，由于基于TCP的应用，是面向流的应用，多个小的数据包可以合并为一个较大的数据包，通过本发明实施例一上述提出的主机中的协议栈，会将部分数据包合并后，再以中断的方式上报给虚拟机的协议栈卸载设备，这样可避免多个小数据包造成的多次中断，减轻中断处理带来的负载。

[0176] 相应地，本发明实施例二还提出一种下协议栈选择装置，如图6a所示，其具体结构组成如下述：

[0177] 接收模块601，用于接收虚拟机发送的套接字socket创建指令，并将所述socket创建指令传输给选择模块602。

[0178] 所述选择模块602，用于选择一个协议栈实例。

[0179] 具体地，上述选择模块602，具体用于按照下述方式中的至少一种选择一个协议栈实例：按照负载均衡原则，在主机中的至少一个协议栈实例中，选择一个协议栈实例；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征在所述主机中的

至少一个协议栈实例中，选择一个协议栈实例；按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照待根据所述socket创建指令创建的socket的属性信息，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例。

[0180] 创建模块603，用于在所述选择模块602选定的所述协议栈实例中按照所述socket创建指令创建socket，并将创建的socket传输给发送模块604。

[0181] 发送模块604，用于将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机，以使所述虚拟机按照所述对应关系，在所述协议栈实例中基于创建的所述socket对数据进行处理。

[0182] 可选地，上述装置还可以包括：

[0183] 第一确定模块，用于按照下述方式中的一种确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC：按照静态指定的方式，确定用于传输数据的虚拟设备不是虚拟NIC；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征，确定用于传输数据的虚拟设备不是虚拟NIC；按照待根据所述socket创建指令创建的socket的属性信息，确定用于传输数据的虚拟设备不是虚拟NIC。

[0184] 具体地，上述接收模块601，还用于接收所述虚拟机传递的对所述socket进行控制操作的控制指令。

[0185] 则，上述装置，还可以包括：

[0186] 第二确定模块，用于根据所述socket和所述协议栈实例的标识的对应关系，确定和所述控制指令中的所述socket对应的所述协议栈实例。

[0187] 上述创建模块603，还用于在所述第二确定模块确定的所述协议栈实例上按照所述控制指令对所述socket进行操作。

[0188] 上述发送模块604，还用于将对所述socket进行操作的操作结果传输至所述虚拟机。

[0189] 具体地，若所述控制指令是基于所述socket在虚拟机之间传输数据，则：所述创建模块603，用于在所述协议栈实例上按照所述控制指令对所述socket进行操作为：具体用于将基于所述socket在虚拟机之间传输的数据，通过所述协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置，所述协议栈卸载装置在虚拟机中用于在协议栈实例和虚拟机之间进行数据传输。

[0190] 可选地，上述创建模块603，还用于确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

[0191] 可选地，上述接收模块601，还用于接收将所述协议栈实例接收到的数据传输给所述socket对应的应用进程的数据传输指令；所述选择模块，还用于根据存储的所述socket和所述协议栈实例标识的对应关系，确定所述数据传输指令中的所述socket对应的所述协议栈实例。

[0192] 上述发送模块604，还用于将所述协议栈实例接收到的数据通过所述协议栈实例，传输至所述虚拟机中所述socket对应的应用进程。

[0193] 相应地，本发明实施例二还提出一种协议栈选择设备，如图6b所示，其具体结构组

成如下述，包括接口61和信号处理器62，接口61和信号处理器62之间通过总线连接，用于传输数据：

[0194] 接口61，用于接收虚拟机发送的套接字socket创建指令，并将所述socket创建指令传输给信号处理器62。

[0195] 所述信号处理器62，用于选择一个协议栈实例。

[0196] 具体地，上述信号处理器62，具体用于按照下述方式中的至少一种选择一个协议栈实例：按照负载均衡原则，在主机中的至少一个协议栈实例中，选择一个协议栈实例；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照预先存储的虚拟机标识和协议栈实例标识之间的对应关系，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例；按照待根据所述socket创建指令创建的socket的属性信息，在所述主机中的至少一个协议栈实例中，选择一个协议栈实例。

[0197] 上述信号处理器62，还用于在选定的所述协议栈实例中按照所述socket创建指令创建socket，并将创建的socket传输给接口61。

[0198] 上述接口61，用于将创建的所述socket和所述协议栈实例的标识的对应关系发送给所述虚拟机，以使所述虚拟机按照所述对应关系，在所述协议栈实例中基于创建的所述socket对数据进行处理。

[0199] 可选地，上述信号处理器62，具体用于按照下述方式中的一种确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC：按照静态指定的方式，确定用于传输数据的虚拟设备不是虚拟NIC；确定待根据所述socket创建指令创建的socket对应的应用进程，按照所述应用进程创建socket的历史信息确定所述应用进程的业务特征，按照所述应用进程的业务特征，确定用于传输数据的虚拟设备不是虚拟NIC；按照待根据所述socket创建指令创建的socket的属性信息，确定用于传输数据的虚拟设备不是虚拟NIC。

[0200] 上述接口61，还用于接收所述虚拟机传递的对所述socket进行控制操作的控制指令。

[0201] 上述信号处理器62，还用于根据所述socket和所述协议栈实例的标识的对应关系，确定和所述控制指令中的所述socket对应的所述协议栈实例。

[0202] 上述信号处理器62，还用于在所述第二确定模块确定的所述协议栈实例上按照所述控制指令对所述socket进行操作。

[0203] 上述接口61，还用于将对所述socket进行操作的操作结果传输至所述虚拟机。

[0204] 具体地，若所述控制指令是基于所述socket在虚拟机之间传输数据，则：所述信号处理器62，用于在所述协议栈实例上按照所述控制指令对所述socket进行操作为：具体用于将基于所述socket在虚拟机之间传输的数据，通过所述协议栈实例，传输至待传输的虚拟机中的协议栈卸载装置，所述协议栈卸载装置在虚拟机中用于在协议栈实例和虚拟机之间进行数据传输。

[0205] 可选地，上述信号处理器62，还用于确定用于传输数据的虚拟设备不是虚拟网络接口卡NIC。

[0206] 可选地，上述接口61，还用于接收将所述协议栈实例接收到的数据传输给所述

socket对应的应用进程的数据传输指令;所述选择模块,还用于根据存储的所述socket和所述协议栈实例标识的对应关系,确定所述数据传输指令中的所述socket对应的所述协议栈实例。

[0207] 上述接口61,还用于将所述协议栈实例接收到的数据通过所述协议栈实例,传输至所述虚拟机中所述socket对应的应用进程。

[0208] 本领域的技术人员应明白,本发明的实施例可提供为方法、装置(设备)、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、只读光盘、光学存储器等)上实施的计算机程序产品的形式。

[0209] 本发明是参照根据本发明实施例的方法、装置(设备)和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0210] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0211] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0212] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围的所有变更和修改。

[0213] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。

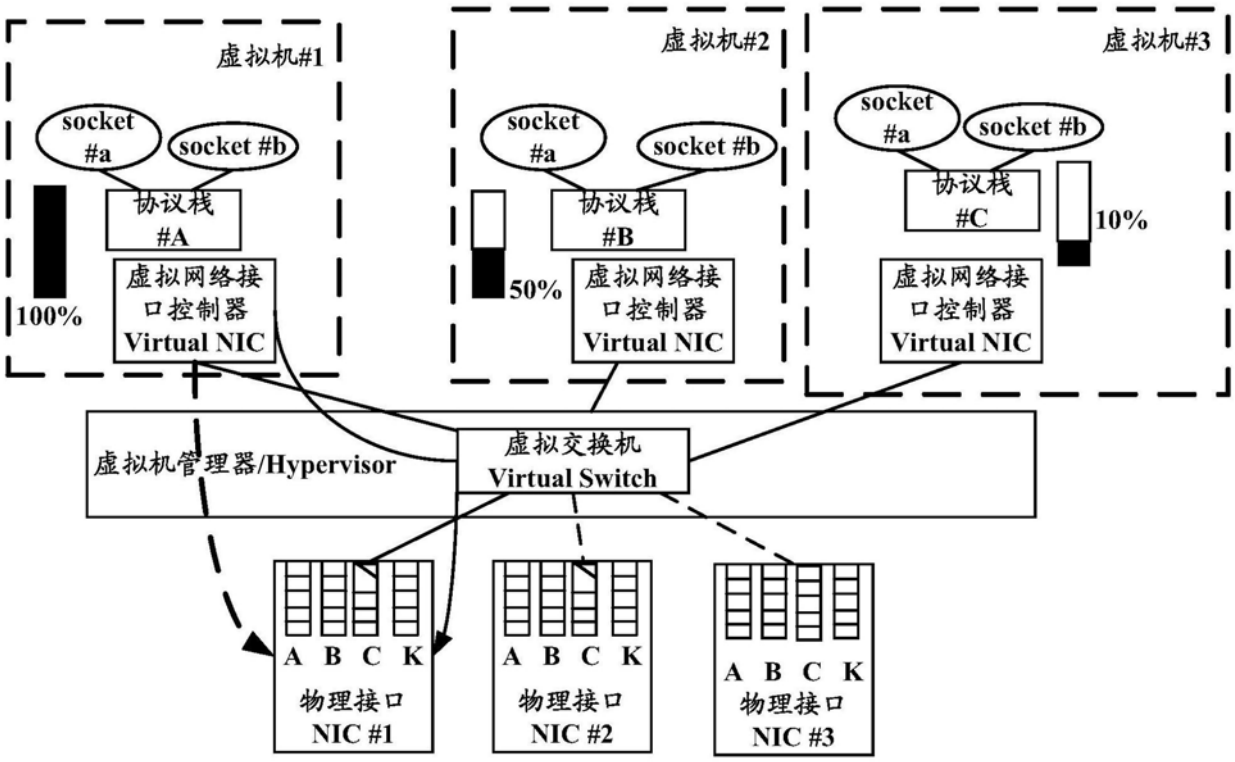

虚拟机#1
socket #a
socket #b
协议栈 #A
虚拟网络接口控制器 Virtual NIC
100%
虚拟机#2
socket #a
socket #b
协议栈 #B
50%
虚拟网络接口控制器 Virtual NIC
虚拟机#3
socket #a
socket #b
协议栈 #C
10%
虚拟网络接口控制器 Virtual NIC
虚拟机管理器/Hypervisor
虚拟交换机 Virtual Switch
A B C K
物理接口 NIC #1
A B C K
物理接口 NIC #2
A B C K
物理接口 NIC #3

图1

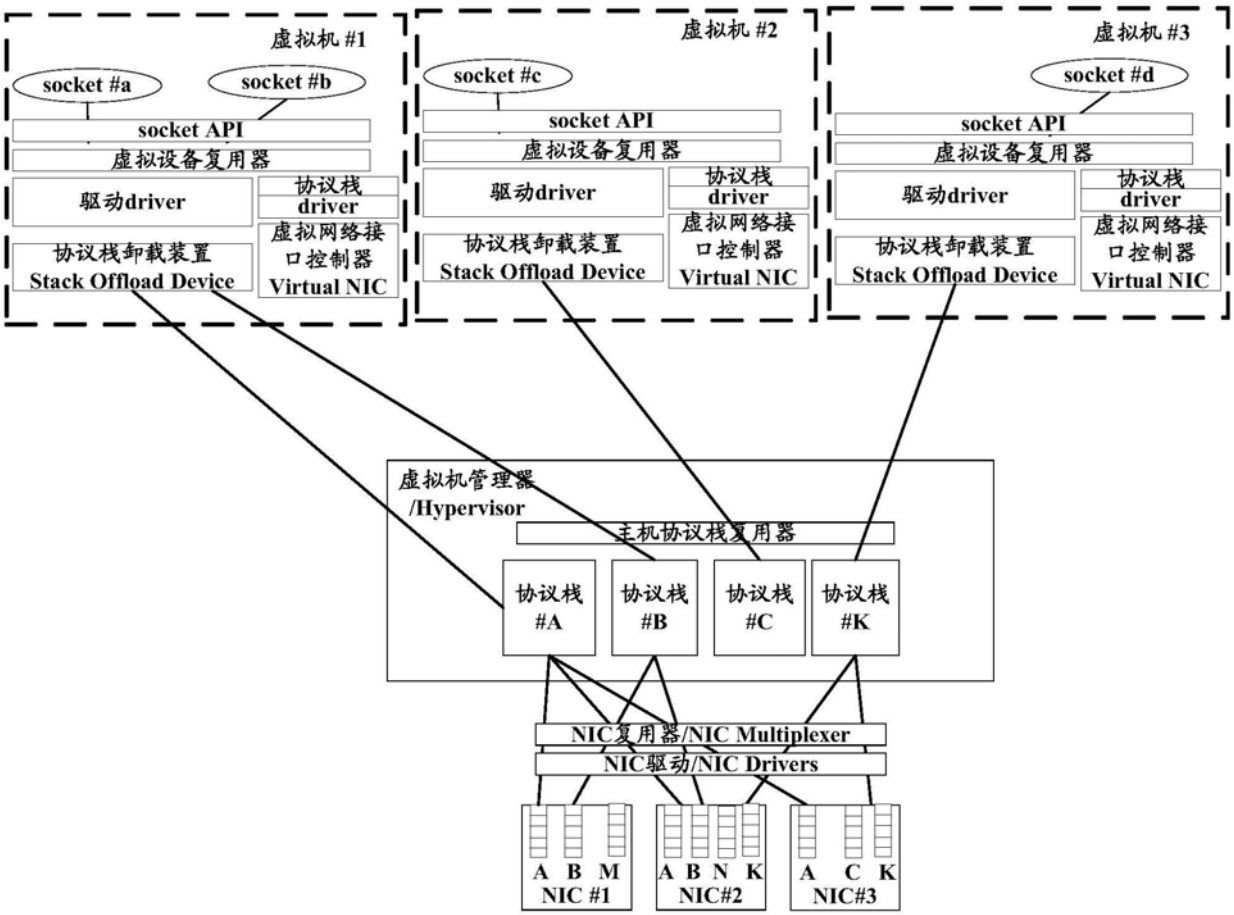

虚拟机 #1
socket #a
socket #b
socket API
虚拟设备复用器
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机 #2
socket #c
socket API
虚拟设备复用器
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机 #3
socket #d
socket API
虚拟设备复用器
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机管理器 /Hypervisor
主机协议栈复用器
协议栈 #A
协议栈 #B
协议栈 #C
协议栈 #K
NIC复用器/NIC Multiplexer
NIC驱动/NIC Drivers
A B M NIC #1
A B N K NIC#2
A C K NIC#3

图2a

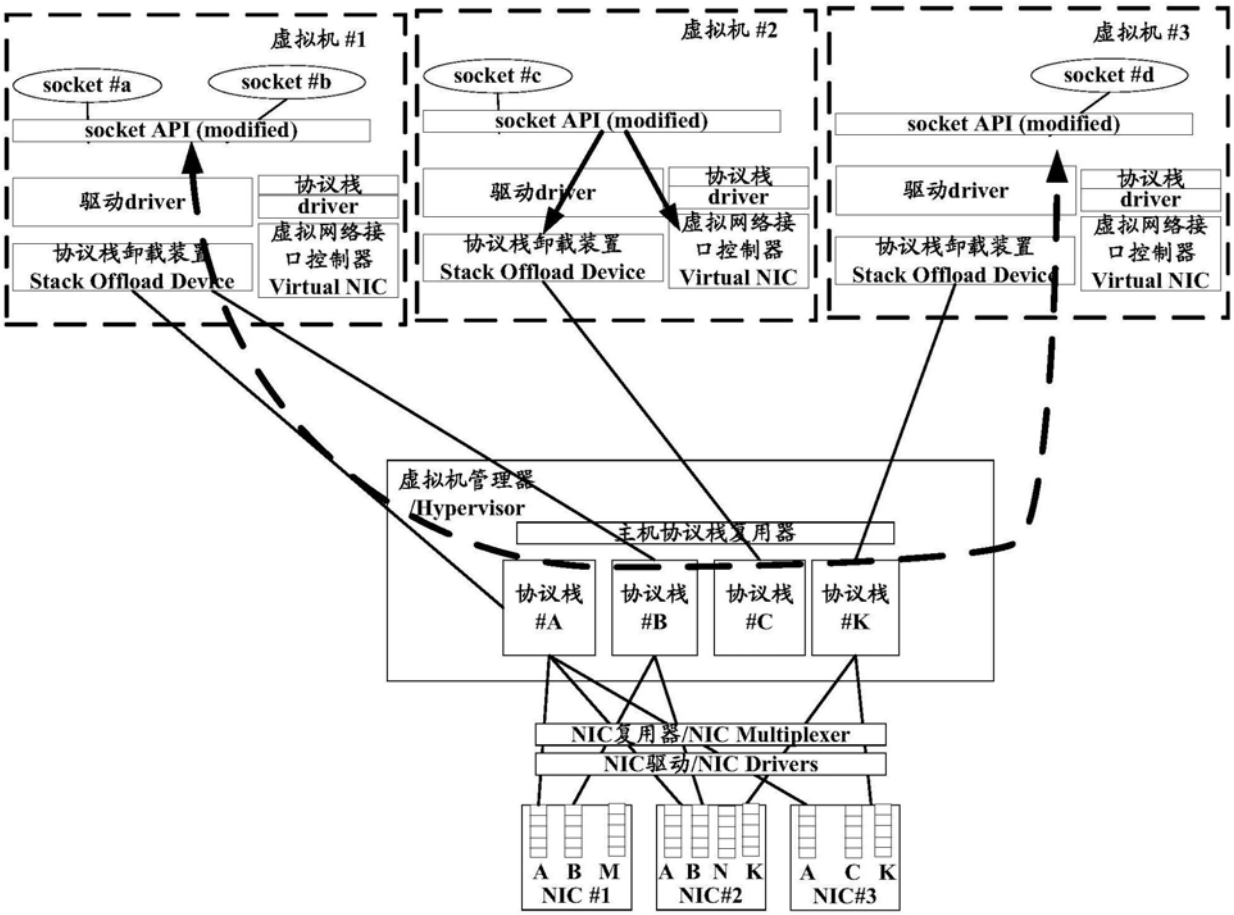

虚拟机 #1
socket #a
socket #b
socket API (modified)
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机 #2
socket #c
socket API (modified)
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机 #3
socket #d
socket API (modified)
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机管理器 Hypervisor
主机协议栈复用器
协议栈 #A
协议栈 #B
协议栈 #C
协议栈 #K
NIC复用器/NIC Multiplexer
NIC驱动/NIC Drivers
A B M NIC #1
A B N K NIC#2
A C K NIC#3

图2b

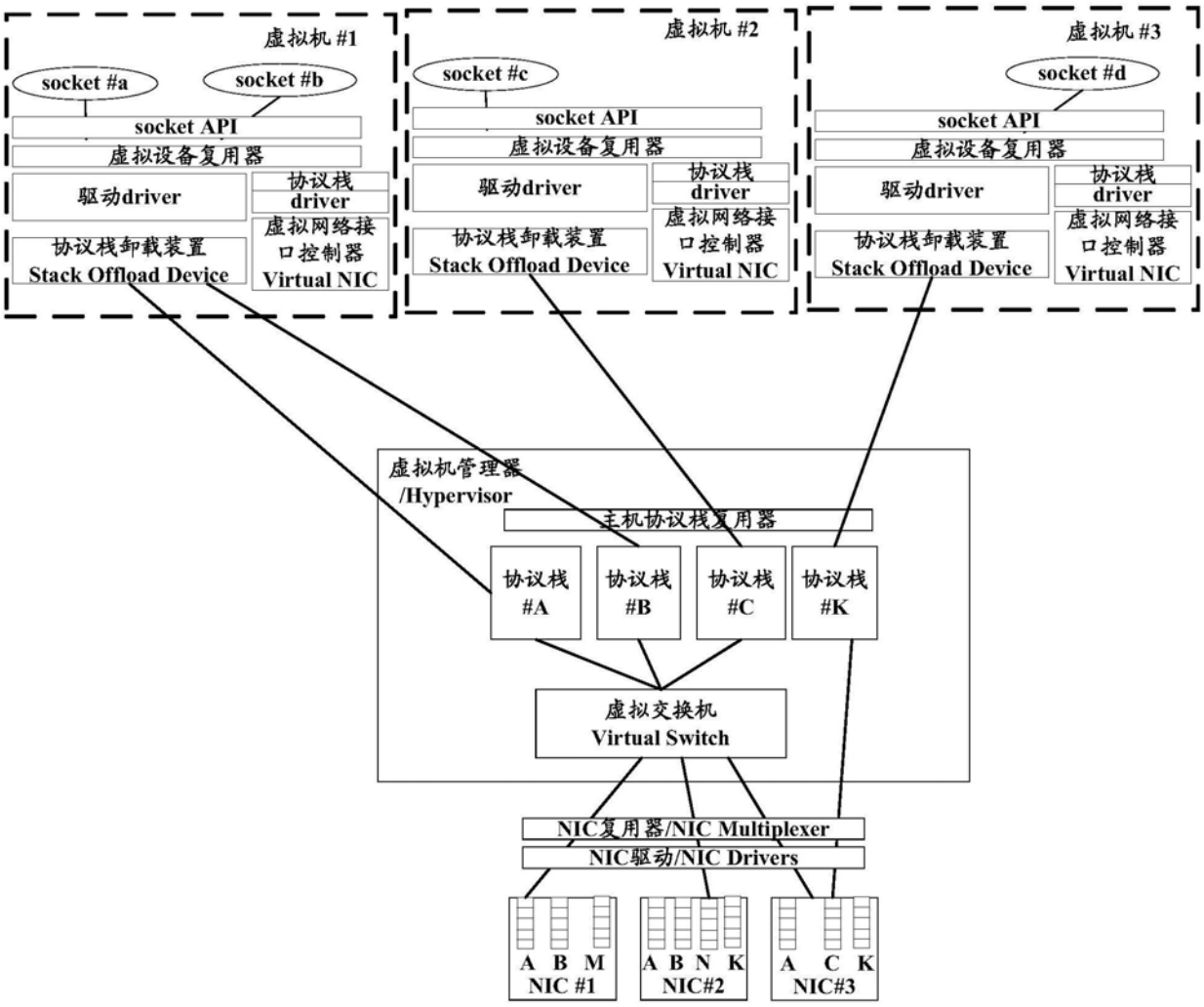

虚拟机 #1
socket #a
socket #b
socket API
虚拟设备复用器
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机 #2
socket #c
socket API
虚拟设备复用器
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机 #3
socket #d
socket API
虚拟设备复用器
驱动driver
协议栈 driver
协议栈卸载装置 Stack Offload Device
虚拟网络接口控制器 Virtual NIC
虚拟机管理器 /Hypervisor
主机协议栈复用器
协议栈 #A
协议栈 #B
协议栈 #C
协议栈 #K
虚拟交换机 Virtual Switch
NIC复用器/NIC Multiplexer
NIC驱动/NIC Drivers
A B M NIC #1
A B N K NIC#2
A C K NIC#3

图2c

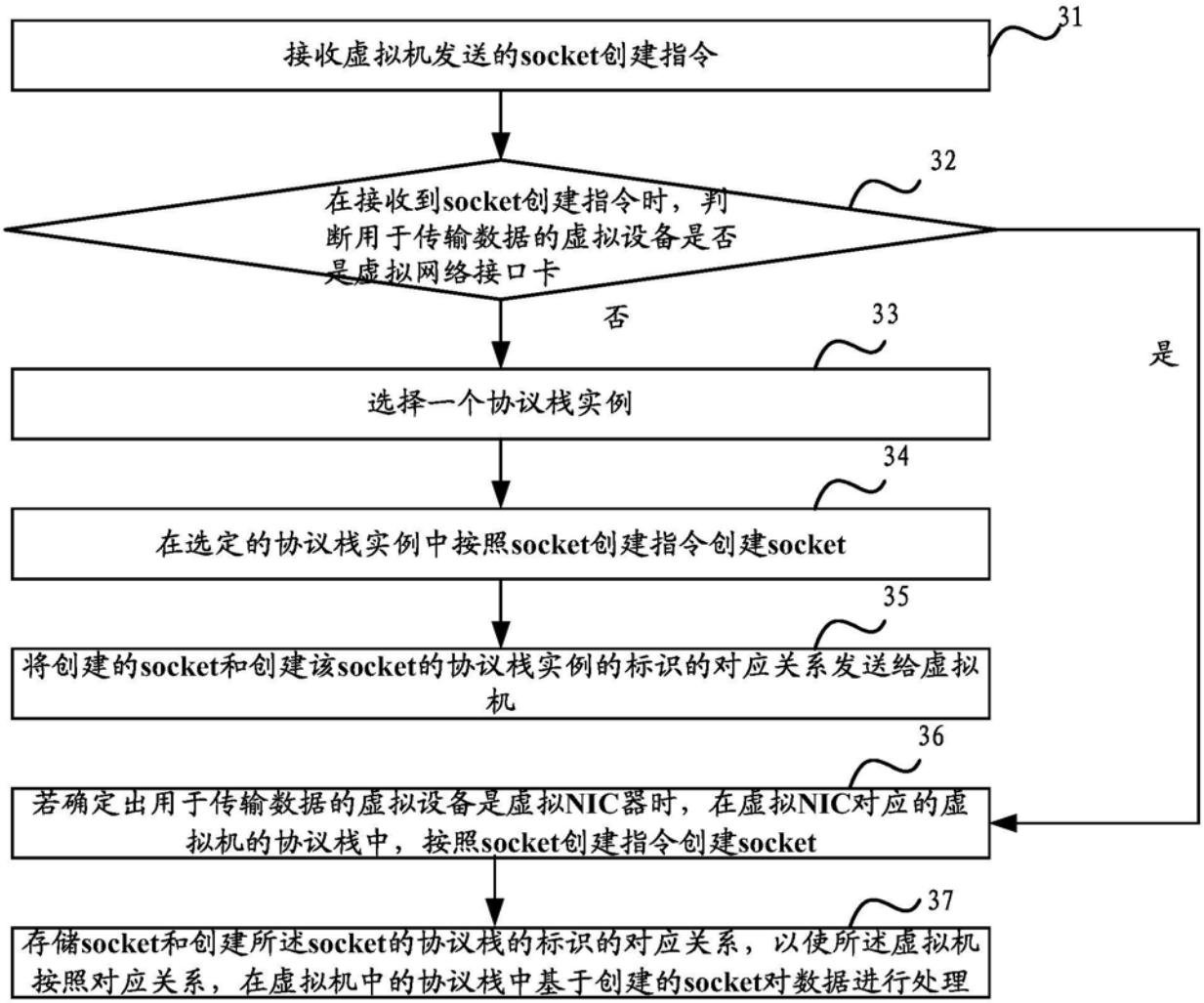

31
接收虚拟机发送的socket创建指令
32
在接收到socket创建指令时，判断用于传输数据的虚拟设备是否是虚拟网络接口卡
否
是
33
选择一个协议栈实例
34
在选定的协议栈实例中按照socket创建指令创建socket
35
将创建的socket和创建该socket的协议栈实例的标识的对应关系发送给虚拟机
36
若确定出用于传输数据的虚拟设备是虚拟NIC器时，在虚拟NIC对应的虚拟机的协议栈中，按照socket创建指令创建socket
37
存储socket和创建所述socket的协议栈的标识的对应关系，以使所述虚拟机按照对应关系，在虚拟机中的协议栈中基于创建的socket对数据进行处理

图3

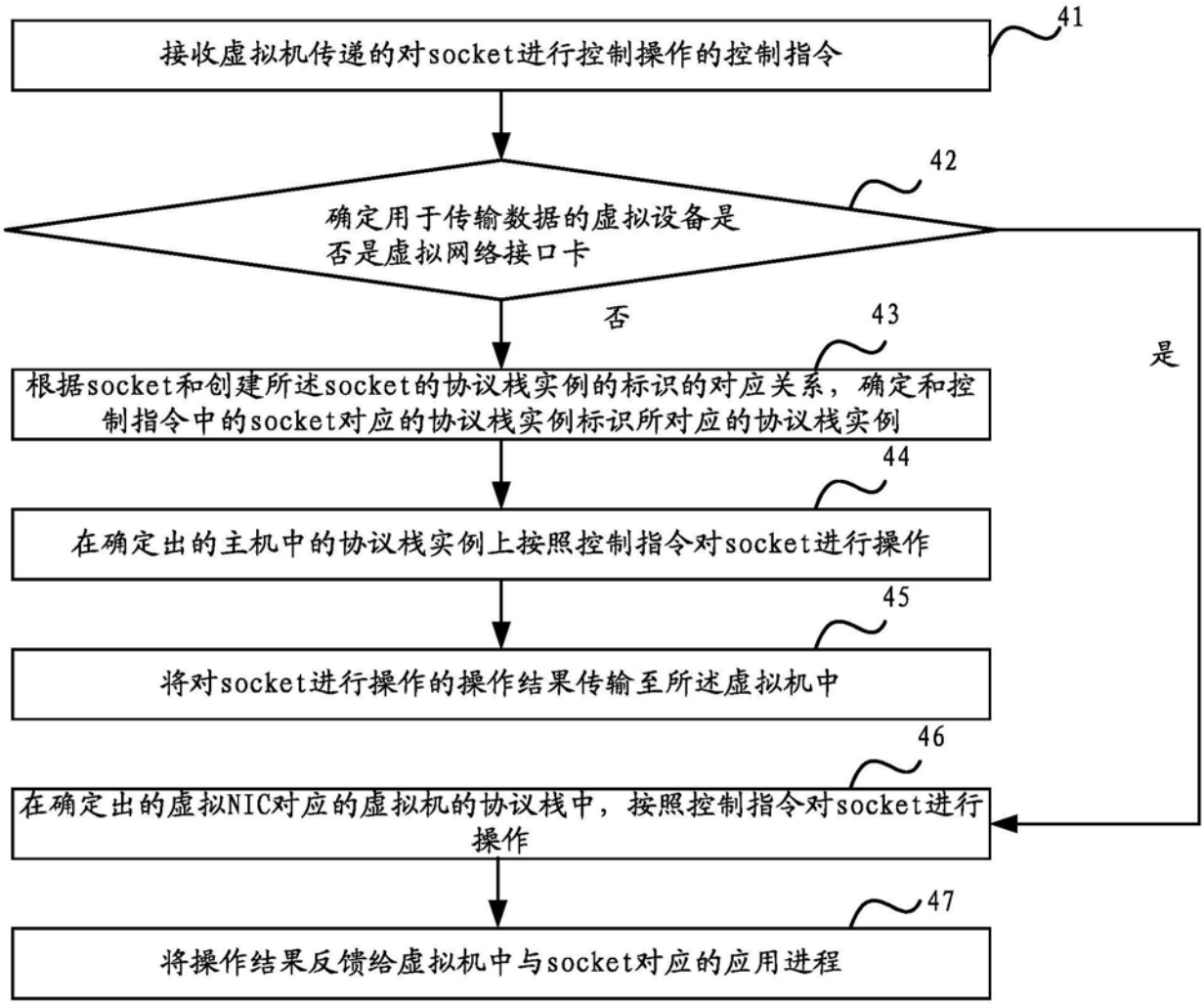

41
接收虚拟机传递的对socket进行控制操作的控制指令
42
确定用于传输数据的虚拟设备是否是虚拟网络接口卡
否
是
43
根据socket和创建所述socket的协议栈实例的标识的对应关系，确定和控制指令中的socket对应的协议栈实例标识所对应的协议栈实例
44
在确定出的主机中的协议栈实例上按照控制指令对socket进行操作
45
将对socket进行操作的操作结果传输至所述虚拟机中
46
在确定出的虚拟NIC对应的虚拟机的协议栈中，按照控制指令对socket进行操作
47
将操作结果反馈给虚拟机中与socket对应的应用进程

图4

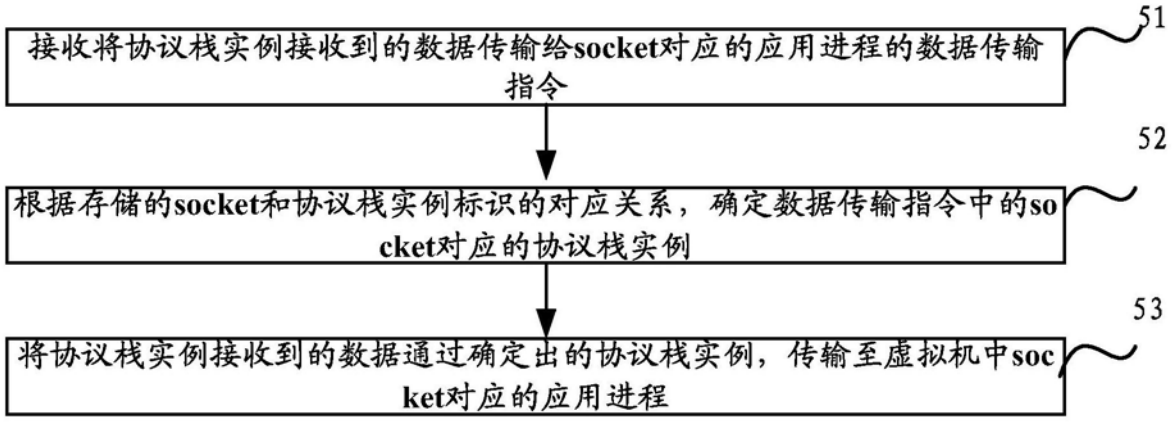

51
接收将协议栈实例接收到的数据传输给socket对应的应用进程的数据传输指令
52
根据存储的socket和协议栈实例标识的对应关系，确定数据传输指令中的socket对应的协议栈实例
53
将协议栈实例接收到的数据通过确定出的协议栈实例，传输至虚拟机中socket对应的应用进程

图5

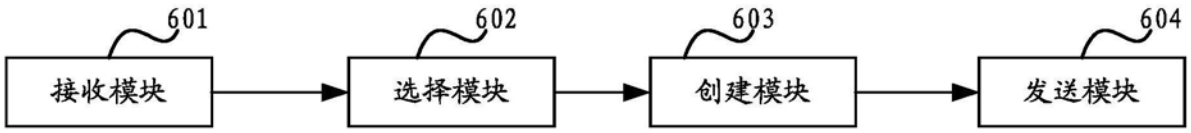


图6a

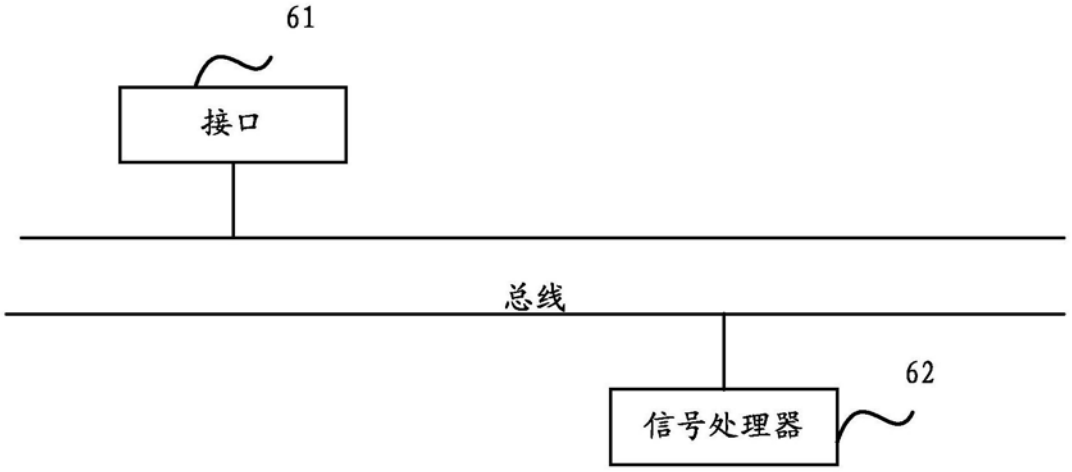


图6b