



(12) 发明专利

(10) 授权公告号 CN 111565142 B

(45) 授权公告日 2020.11.03

(21) 申请号 202010677756.5

(22) 申请日 2020.07.15

(65) 同一申请的已公布的文献号  
申请公布号 CN 111565142 A

(43) 申请公布日 2020.08.21

(73) 专利权人 鹏城实验室  
地址 518000 广东省深圳市南山区西丽街  
道留仙洞万科云城一期8栋

(72) 发明人 梁俊 付志鹏 郭长国 张晓梨

(74) 专利代理机构 深圳市世纪恒程知识产权代  
理事务所 44287

代理人 张婷

(51) Int. Cl.

H04L 12/46 (2006.01)

H04L 12/741 (2013.01)

(56) 对比文件

CN 110048963 A, 2019.07.23

CN 107454132 A, 2017.12.08

US 2020177629 A1, 2020.06.04

US 2018097734 A1, 2018.04.05

US 2018062923 A1, 2018.03.01

US 10673698 B2, 2020.06.02

CN 108600049 A, 2018.09.28

CN 109863720 A, 2019.06.07

郑杰.《基于DPDK的云AC系统设计与实现》.《中国优秀硕士学位论文全文数据库(信息科技辑)》.2017, (第02期), 第1138-387页.

严立宇 等.《云计算网络中多租户虚拟网络隔离的分布式实现研究》.《计算机应用与软件》.2016, 第33卷(第11期), 第93-98页.

审查员 匡仁炳

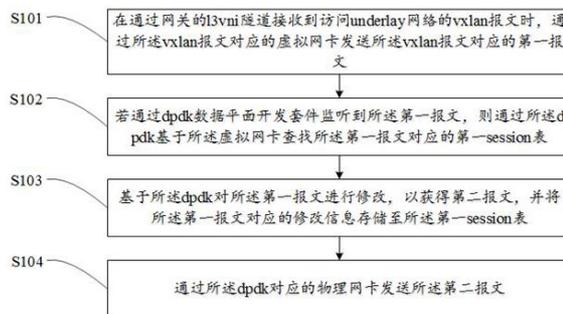
权利要求书2页 说明书12页 附图1页

(54) 发明名称

报文处理方法、装置及计算机可读存储介质

(57) 摘要

本发明公开了一种报文处理方法,包括:通过vxlan报文对应的虚拟网卡发送vxlan报文对应的第一报文;若通过dpdk数据平面开发套件监听到第一报文,则通过dpdk基于虚拟网卡查找第一报文对应的第一session表;基于dpdk对第一报文进行修改,以获得第二报文,并将第一报文对应的修改信息存储至第一session表;通过dpdk对应的物理网卡发送第二报文。本发明还公开了一种报文处理装置及计算机可读存储介质。本发明实现通过dpdk转发访问underlay网络的vxlan报文,控制平面采用MP-bgp多协议边界网关协议实现,数据平面充分利用dpdk数据转发的高性能,提高数据转发的效率,同时,由于每一个租户对应一个独立的session表以及虚拟网卡,进而能够通过session表以及虚拟网卡区分不同的租户网络,从而实现多租户的隔离。



1. 一种报文处理方法,其特征在于,所述报文处理方法包括以下步骤:

在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;

若通过dpdk数据平面开发套件监听到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;

基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;

通过所述dpdk对应的物理网卡发送所述第二报文。

2. 如权利要求1所述的报文处理方法,其特征在于,所述基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表的步骤包括:

基于所述dpdk更改所述第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得所述第二报文;

将所述第一报文中的目的ip、源ip、源端口以及目的端口,更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口存储至所述第一session表。

3. 如权利要求1所述的报文处理方法,其特征在于,所述通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文的步骤包括:

删除所述vxlan报文的报文头,以获得vxlan内层报文,并查询vrf虚拟路由转发设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;

若是,则基于所述vxlan内层报文生成所述第一报文,并通过所述虚拟网卡发送所述第一报文。

4. 如权利要求1所述的报文处理方法,其特征在于,所述报文处理方法还包括:

在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表;

基于第二session表,对所述mac报文进行转换,以获得第三报文,并将所述第三报文通过虚拟网卡发送至内核;

通过内核对所述第三报文进行封装,以获得目标报文,并转发所述目标报文。

5. 如权利要求4所述的报文处理方法,其特征在于,所述在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表的步骤包括:

在检测到所述物理网卡接收到mac报文时,通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表;

若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表。

6. 如权利要求4所述的报文处理方法,其特征在于,所述基于第二session表,对所述mac报文进行转换,以获得第三报文的步骤包括:

基于第二session表,对所述mac报文中的目的ip、源ip、源端口以及目的端口进行转换,并更新所述mac报文中的mac地址,以获得第三报文。

7. 如权利要求4所述的报文处理方法,其特征在于,所述通过内核对所述第三报文进行封装,以获得目标报文的步骤包括:

通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,以获得所述mac报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址;

基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文。

8. 如权利要求1至7任一项所述的报文处理方法,其特征在于,所述报文处理方法还包括:

创建vrf设备、bridge设备以及vxlan设备,将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;

通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备;

在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡;

运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点。

9. 一种报文处理装置,其特征在于,所述报文处理装置包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的报文处理程序,所述报文处理程序被所述处理器执行时实现如权利要求1至8中任一项所述的报文处理方法的步骤。

10. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质上存储有报文处理程序,所述报文处理程序被处理器执行时实现如权利要求1至8中任一项所述的报文处理方法的步骤。

## 报文处理方法、装置及计算机可读存储介质

### 技术领域

[0001] 本发明涉及数据处理技术领域,尤其涉及一种报文处理方法、装置及计算机可读存储介质。

### 背景技术

[0002] 目前,linux操作系统中,evpn(Ethernet Virtual Private Network,以太网虚拟专用网络)采用mp-bgp(Multiprotocol Border Gateway Protocol,多协议边界网关协议)作为vxlan控制平面,能够有效处理广播、未知单播、组播流量。

[0003] 但是,采用evpn进行vxlan网络(overlay network)和underlay网络之间的网络通信时,vxlan网络访问underlay网络需要进行nat网络地址转换,而nat实现依赖于linux内核的contrack连接跟踪机制,在高并发的情况下contrack的容量容易达到极限,造成新的连接会被拒绝,导致数据转发的效率低。并且,由于linux内核无法实现不同租户的contrack的隔离,当两个租户网络的网段重叠或者相同时,linux内核的nat转换无法保证这两个租户相同ip的虚拟机能同时通过网关成功访问underlay 网络。

[0004] 上述内容仅用于辅助理解本发明的技术方案,并不代表承认上述内容是现有技术。

### 发明内容

[0005] 本发明的主要目的在于提供一种报文处理方法、装置及计算机可读存储介质,旨在解决通过evpn进行vxlan网络(overlay network)和underlay之间的网络通信时数据转发效率低,以及无法实现租户之间的contrack的隔离的技术问题。

[0006] 为实现上述目的,本发明提供一种报文处理方法,所述报文处理方法包括以下步骤:

[0007] 在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;

[0008] 若通过dpdk数据平面开发套件监听到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;

[0009] 基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;

[0010] 通过所述dpdk对应的物理网卡发送所述第二报文。

[0011] 进一步地,所述基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表的步骤包括:

[0012] 基于所述dpdk更改所述第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得所述第二报文;

[0013] 将所述第一报文中的目的ip、源ip、源端口以及目的端口,更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口存储至所述第一session表。

[0014] 进一步地,所述通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文的步骤包括:

[0015] 删除所述vxlan报文的报文头,以获得vxlan内层报文,并查询vrf虚拟路由转发设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;

[0016] 若是,则基于所述vxlan内层报文生成所述第一报文,并通过所述虚拟网卡发送所述第一报文。

[0017] 进一步地,所述报文处理方法还包括:

[0018] 在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表;

[0019] 基于第二session表,对所述mac报文进行转换,以获得第三报文,并将所述第三报文通过虚拟网卡发送至内核;

[0020] 通过内核对所述第三报文进行封装,以获得目标报文,并转发所述目标报文。

[0021] 进一步地,所述在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表的步骤包括:

[0022] 在检测到所述物理网卡接收到mac报文时,通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表;

[0023] 若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表。

[0024] 进一步地,所述基于第二session表,对所述mac报文进行转换,以获得第三报文的步骤包括:

[0025] 基于第二session表,对所述mac报文中的目的ip、源ip、源端口以及目的端口进行转换,并更新所述mac报文中的mac地址,以获得第三报文。

[0026] 进一步地,所述通过内核对所述第三报文进行封装,以获得目标报文的步骤包括:

[0027] 通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,以获得所述mac报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址;

[0028] 基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文。

[0029] 进一步地,所述报文处理方法还包括:

[0030] 创建vrf设备、bridge设备以及vxlan设备,将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;

[0031] 通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备;

[0032] 在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡;

[0033] 运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点。

[0034] 此外,为实现上述目的,本发明还提供一种报文处理装置,所述报文处理装置包括:存储器、处理器及存储在所述存储器上并可在所述处理器上运行的报文处理方程序,所述报文处理方程序被所述处理器执行时实现前述的报文处理方法的步骤。

[0035] 此外,为实现上述目的,本发明还提供一种计算机可读存储介质,所述计算机可读

存储介质上存储有报文处理方程序,所述报文处理方程序被处理器执行时实现前述的报文处理方法的步骤。

[0036] 本发明通过在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;接着若通过dpdk数据平面开发套件监测到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;而后基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;然后通过所述dpdk对应的物理网卡发送所述第二报文,实现通过dpdk转发访问underlay网络的vxlan报文,控制平面采用MP-bgp多协议边界网关协议实现,数据平面充分利用dpdk数据转发的高性能,提高数据转发的效率,同时,由于每一个租户对应一个独立的session表以及虚拟网卡,进而能够通过session表以及虚拟网卡区分不同的租户网络,从而实现多租户的隔离。

### 附图说明

[0037] 图1是本发明实施例方案涉及的硬件运行环境中报文处理装置的结构示意图;

[0038] 图2为本发明报文处理方法第一实施例的流程示意图。

[0039] 本发明目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

### 具体实施方式

[0040] 应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0041] 如图1所示,图1是本发明实施例方案涉及的硬件运行环境中报文处理装置的结构示意图。

[0042] 本发明实施例报文处理装置可以是PC,也可以是智能手机、平板电脑、电子书阅读器、MP3(Moving Picture Experts Group Audio Layer III,动态影像专家压缩标准音频层面3)播放器、MP4(Moving Picture Experts Group Audio Layer IV,动态影像专家压缩标准音频层面4)播放器、便携计算机等具有显示功能的可移动式终端设备。

[0043] 如图1所示,该报文处理装置可以包括:处理器1001,例如CPU,网络接口1004,用户接口1003,存储器1005,通信总线1002。其中,通信总线1002用于实现这些组件之间的连接通信。用户接口1003可以包括显示屏(Display)、输入单元比如键盘(Keyboard),可选用户接口1003还可以包括标准的有线接口、无线接口。网络接口1004可选的可以包括标准的有线接口、无线接口(如WI-FI接口)。存储器1005可以是高速RAM存储器,也可以是稳定的存储器(non-volatile memory),例如磁盘存储器。存储器1005可选的还可以是独立于前述处理器1001的存储装置。

[0044] 可选地,报文处理装置还可以包括摄像头、RF(Radio Frequency,射频)电路,传感器、音频电路、WiFi模块等等。其中,传感器比如光传感器、运动传感器以及其他传感器。具体地,光传感器可包括环境光传感器及接近传感器,其中,环境光传感器可根据环境光线的明暗来调节显示屏的亮度,接近传感器可在移动终端移动到耳边时,关闭显示屏和/或背光。作为运动传感器的一种,重力加速度传感器可检测各个方向上(一般为三轴)加速度的大小,静止时可检测出重力的大小及方向,可用于识别报文处理装置姿态的应用(比如横竖屏切换、相关游戏、磁力计姿态校准)、振动识别相关功能(比如计步器、敲击)等;当然,报文

处理装置还可配置陀螺仪、气压计、湿度计、温度计、红外线传感器等其他传感器,在此不再赘述。

[0045] 本领域技术人员可以理解,图1中示出的终端结构并不构成对报文处理装置的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件布置。

[0046] 如图1所示,作为一种计算机存储介质的存储器1005中可以包括操作系统、网络通信模块、用户接口模块以及报文处理方程序。

[0047] 在图1所示的报文处理装置中,网络接口1004主要用于连接后台服务器,与后台服务器进行数据通信;用户接口1003主要用于连接客户端(用户端),与客户端进行数据通信;而处理器1001可以用于调用存储器1005中存储的报文处理方程序。

[0048] 在本实施例中,报文处理方装置包括:存储器1005、处理器1001及存储在所述存储器1005上并可在所述处理器1001上运行的报文处理方程序,其中,处理器1001调用存储器1005中存储的报文处理方程序时,并执行以下操作:

[0049] 在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;

[0050] 若通过dpdk数据平面开发套件监听到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;

[0051] 基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;

[0052] 通过所述dpdk对应的物理网卡发送所述第二报文。

[0053] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以下操作:

[0054] 基于所述dpdk更改所述第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得所述第二报文;

[0055] 将所述第一报文中的目的ip、源ip、源端口以及目的端口,更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口存储至所述第一session表。

[0056] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以下操作:

[0057] 删除所述vxlan报文的报文头,以获得vxlan内层报文,并查询vrf虚拟路由转发设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;

[0058] 若是,则基于所述vxlan内层报文生成所述第一报文,并通过所述虚拟网卡发送所述第一报文。

[0059] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以下操作:

[0060] 在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表;

[0061] 基于第二session表,对所述mac报文进行转换,以获得第三报文,并将所述第三报文通过虚拟网卡发送至内核;

[0062] 通过内核对所述第三报文进行封装,以获得目标报文,并转发所述目标报文。

[0063] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以

下操作：

[0064] 在检测到所述物理网卡接收到mac报文时,通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表;

[0065] 若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表。

[0066] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以下操作：

[0067] 基于第二session表,对所述mac报文中的目的ip、源ip、源端口以及目的端口进行转换,并更新所述mac报文中的mac地址,以获得第三报文。

[0068] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以下操作：

[0069] 通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,以获得所述mac报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址;

[0070] 基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文。

[0071] 进一步地,处理器1001可以调用存储器1005中存储的报文处理方程序,还执行以下操作：

[0072] 创建vrf设备、bridge设备以及vxlan设备,将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;

[0073] 通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备;

[0074] 在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡;

[0075] 运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点。

[0076] 本发明还提供一种报文处理方法,参照图2,图2为本发明报文处理方法第一实施例的流程示意图。

[0077] 本实施例中,该报文处理方法应用于设有linux操作系统的网关服务器(网关),在SDN(SoftwareDefinedNetwork,软件定义网络)中,该网关服务器对应的每一个租户的虚拟机的控制节点均已创建与网关服务器之间的l3vni隧道,其中,l3vni隧道是一个vxlan隧道,用来实现不同vni的vxlan间通信或者vxlan叠加网络(overlay network)和underlay之间的通信。

[0078] 本实施例中,该报文处理方法包括以下步骤：

[0079] 步骤S101,在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;

[0080] 本实施例中,在SDN中,各个租户均可通过其对应的l3vni隧道访问underlay网络,具体地,租户的虚拟机通过网关访问underlay网络,在访问时,虚拟机将访问underlay网络的报文通过l3vni隧道发送至网关,进而,网关可通过其多个l3vni隧道中的某一l3vni隧道接收访问underlay网络的vxlan报文。

[0081] 本实施例中,在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,

即通过其多个l3vni隧道中的某一l3vni隧道接收到访问underlay网络的vxlan报文时,网关通过vxlan报文对应的虚拟网卡发送vxlan报文对应的第一报文,网关先剥去vxlan报文的vxlan报文头,得到vxlan内层报文,而后在确定vxlan内层报文对应的目标网络的出口为虚拟网卡时,通过vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文。

[0082] 步骤S102,若通过dpdk数据平面开发套件监听到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;

[0083] 需要说明的是,在dpdk(Data Plane Development Kit,数据平面开发套件)完成各个虚拟网卡的创建后,实时监听各个虚拟网卡。其中,网关的控制器在接收到租户对应的虚拟机通过网关访问underlay网络的请求时,网关的控制器会下发通知信息至dpdk,通知dpdk创建虚拟网卡以及虚拟网卡关联的session表。

[0084] 本实施例中,若通过dpdk数据平面开发套件监听到第一报文,即虚拟网卡收到访问underlay网络的报文,则通过所述dpdk基于所述虚拟网卡查找第一报文对应的第一session表,具体地,dpdk设有监听虚拟网卡的组件,dpdk通过该组件监听虚拟网卡,若dpdk通过该组件监听到虚拟网卡收到的第一报文,则dpdk通过目标虚拟网卡查找第一报文对应的第一session表,该第一session表为该目标虚拟网卡所对应的session表,以实现每一个租户对应一个独立的session表以及虚拟网卡,以通过session表以及虚拟网卡区分不同的租户网络,从而实现多租户的隔离,其中,各个虚拟网卡分别对应唯一一个session表,目标虚拟网卡为收到访问underlay网络的报文的虚拟网卡。

[0085] 步骤S103,基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;

[0086] 本实施例中,在第一session表创建完成时,基于dpdk对第一报文进行修改,以获得第二报文,并将第一报文对应的修改信息存储至第一session表,具体地,对第一报文进行nat,即更改第一报文中的目的ip、源ip、源端口以及目的端口,以获得第二报文,并将更改前第一报文中的目的ip、源ip、源端口以及目的端口,以及更改后第二报文中的目的ip、源ip、源端口以及目的端口存储至所述第一session表,以通过第一session表实现vxlan报文所对应的nat转换。

[0087] 步骤S104,通过所述dpdk对应的物理网卡发送所述第二报文。

[0088] 在本实施例中,在获取到第二报文时,通过dpdk对应的物理网卡发送第二报文,具体地,通过把第二报文从dpdk绑定的物理网卡发出,进而实现通过dpdk转发访问underlay网络的报文(vxlan报文剥掉报文头之后的报文),以在数据平面充分利用dpdk数据转发的高性能,避免新的连接被拒绝的情况发生,提高数据转发的效率。

[0089] 本实施例提出的报文处理方法,通过在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;接着若通过dpdk数据平面开发套件监测到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;而后基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;然后通过所述dpdk对应的物理网卡发送所述第二报文,实现通过dpdk转发访问underlay网络的报文,在数据平面充分利用dpdk数据转发的高性能,提高数据转发的效率,同时,由于每一个租户对应一个独立的session表以及虚拟网卡,进而能够通过session表

以及虚拟网卡区分不同的租户网络,从而实现多租户的隔离。

[0090] 基于第一实施例,提出本发明报文处理方法的第二实施例,在本实施例中,步骤S103包括:

[0091] 步骤S201,基于所述dpdk更改所述第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得所述第二报文;

[0092] 步骤S202,将所述第一报文中的目的ip、源ip、源端口以及目的端口,更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口存储至所述第一session表。

[0093] 本实施例中,在第一session表创建完成时,基于dpdk更改第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得第二报文,具体地,先根据预设规则确定更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口、更改后的mac地址,而后,将第一报文中的目的ip转换为更改后的目的ip、将第一报文中的源ip转换为更改后的源ip、将第一报文中的源端口转换为更改后的源端口、将第一报文中的目的端口转换为更改后的目的端口,将第一报文中的mac地址转换为更改后的mac地址,进而得到第二报文,并将更改前第一报文中的目的ip、源ip、源端口以及目的端口,以及更改后第二报文中的目的ip、源ip、源端口以及目的端口(更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口)存储至所述第一session表。

[0094] 本实施例提出的报文处理方法,通过基于所述dpdk更改所述第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得所述第二报文;接着将所述第一报文中的目的ip、源ip、源端口以及目的端口,更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口存储至所述第一session表,能够准确获得第二报文,并存储第一报文对应nat的数据,便于后续调用。

[0095] 基于第一实施例,提出本发明报文处理方法的第三实施例,在本实施例中,步骤S101包括:

[0096] 步骤S301,删除所述vxlan报文的报文头,以获得vxlan内层报文,并查询vrf虚拟路由转发设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;

[0097] 步骤S302,若是,则基于所述vxlan内层报文生成所述第一报文,并通过所述虚拟网卡发送所述第一报文。

[0098] 需要说明的是,vrf设备能将一台物理路由设备虚拟成多个虚拟路由设备,一个vrf设备会关联一张策略路由表,可以通过mp-bgp进行ip路由学习和mac学习,学习得到的ip路由信息保在策略路由表,mac信息保存在vrf设备关联的neighbor表和系统fdb表。

[0099] 本实施例中,在获取到vxlan报文时,先删除vxlan报文的报文头(vxlan报文头),以得到vxlan内层报文,而后查询vrf虚拟路由转发设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;具体地,根据vxlan内层报文查询vrf虚拟路由转发设备关联的策略路由表,根据查询结果确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡,即策略路由表中vxlan内层报文对应的目标网络的出口是否为虚拟网卡。

[0100] 而后,若vxlan内层报文对应的目标网络的出口为虚拟网卡,则基于vxlan内层报文生成第一报文,并通过虚拟网卡发送所述第一报文,以实现通过虚拟网卡发送第一报文。

[0101] 本实施例提出的报文处理方法,通过删除所述vxlan报文的报文头,以获得vxlan内层报文,并查询vrf设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;接着若是,则基于所述vxlan内层报文生成所述第一报文,并通过所述虚拟网卡发送所述第一报文,能够通过虚拟网卡准确转发第一报文,由于每一个租户对应一个独立的session表、虚拟网卡、vrf设备,进而能够通过session表、vrf设备以及虚拟网卡区分不同的租户网络,从而实现多租户的隔离。

[0102] 基于第一实施例,提出本发明报文处理方法的第四实施例,在本实施例中,该报文处理方法还包括:

[0103] 步骤S401,在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表;

[0104] 步骤S402,基于第二session表,对所述mac报文进行转换,以获得第三报文,并将所述第三报文通过虚拟网卡发送至内核;

[0105] 步骤S403,通过对所述第三报文进行封装,以获得目标报文,并转发所述目标报文。

[0106] 需要说明的是,网关还可以接收网络反馈的mac报文,具体地,通过dpdk所绑定的物理网卡接收网络反馈的mac报文。

[0107] 本实施例中,在检测到物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表,具体地,dpdk根据mac报文中的mac报文中的目的ip、源ip、源端口以及目的端口,查找所有的session表,以确定mac报文对应的第二session表。

[0108] 接着,基于第二session表,对所述mac报文进行转换,以获得第三报文,具体地,对所述mac报文中的目的ip、源ip、源端口以及目的端口进行转换,并更新所述mac报文中的mac地址,以获得第三报文。

[0109] 而后,将第三报文通过虚拟网卡发送至内核(Linux系统的内核协议栈),并通过内核对所述第三报文进行封装,以获得目标报文,并转发所述目标报文,即将该目标报文转发至对应的目的节点。

[0110] 进一步地,在一实施例中,步骤S402包括:

[0111] 基于第二session表,对所述mac报文中的目的ip、源ip、源端口以及目的端口进行转换,并更新所述mac报文中的mac地址,以获得第三报文。

[0112] 具体地,将mac报文中的mac报文中的目的ip转换为第二session表中对应的目的ip、即mac报文中的目的ip在第二session表中所对应的目的ip,将mac报文中的源ip转换为第二session表中对应的源ip、即mac报文中的源ip在第二session表中所对应的源ip,将mac报文中的源端口转换为第二session表中对应的源端口、即mac报文中的源端口在第二session表中所对应的源端口,将mac报文中的目的端口转换为第二session表中对应的目的端口、即mac报文中的目的端口在第二session表中所对应的目的端口,并更新所述mac报文中的mac地址,以获得第三报文。

[0113] 本实施例提出的报文处理方法,通过在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表;接着基于第二session表,对所述mac报文进行转换,以获得第三报文,并将所述第三报文通过虚拟网卡发送至内核;而后通过对所述第三报文进行封装,以获得目标报文,并转发所述目标报文,实现通过dpdk转发网

络发送的mac报文,充分利用dpdk数据转发的高性能,提高数据转发的效率,同时,由于每一个mac报文对应一个独立的session表以及虚拟网卡,进而能够通过session表以及虚拟网卡区分不同的租户网络,从而实现多租户的隔离。

[0114] 基于第四实施例,提出本发明报文处理方法的第五实施例,在本实施例中,步骤S401包括:

[0115] 步骤S501,在检测到所述物理网卡接收到mac报文时,通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表;

[0116] 步骤S502,若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表。

[0117] 本实施例中,在检测到物理网卡接收到mac报文时,dpdk获取mac报文中的目的ip、源ip、源端口以及目的端口,并通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表,即dpdk根据mac报文中的目的ip、源ip、源端口以及目的端口查找所有的session表,以确定是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表,若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表。

[0118] 需要说明的是,若第二session表中包括mac报文中的目的ip、源ip、源端口以及目的端口,则确定第二session表与mac报文中的目的ip、源ip、源端口以及目的端口匹配。

[0119] 本实施例提出的报文处理方法,通过在检测到所述物理网卡接收到mac报文时,通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表;若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表,能够根据mac报文中的目的ip、源ip、源端口以及目的端口准确查找到第二session表,进一步提高了数据转发的效率。

[0120] 基于第四实施例,提出本发明报文处理方法的第六实施例,在本实施例中,步骤S403包括:

[0121] 步骤S601,通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,以获得所述第三报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址;

[0122] 步骤S602,基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文。

[0123] 本实施例中,在获取到第三报文时,通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,即内核根据该第三报文查找这三张表,以得到该第三报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址。

[0124] 接着,基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文。

[0125] 本实施例提出的报文处理方法,通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,以获得所述mac报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址;接着基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文,实现了对第

三报文进行准确封装,进一步提升数据转发效率。

[0126] 基于上述各个实施例,提出本发明报文处理方法的第七实施例,在本实施例中,步骤S101之前,该报文处理方法还包括:

[0127] 步骤S701,创建vrf设备、bridge设备以及vxlan设备,将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;

[0128] 步骤S702,通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备;

[0129] 步骤S703,在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡;

[0130] 步骤S704,运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点。

[0131] 需要说明的时,为实现上述各个实施例,网关以及客户端均需要创建l3vni隧道。

[0132] 其中,网关创建l3vni隧道的方式包括:

[0133] 在linux系统上创建vrf设备、bridge设备以及vxlan设备;其中,vrf能将一台物理路由设备虚拟成多个虚拟路由设备,一个vrf设备会关联一张策略路由表,可以通过mp-bgp进行ip路由学习和mac学习,学习得到的ip路由信息保在策略路由表,mac信息保存在vrf设备关联的neighbor表和系统fdb表。bridge设备成为vrf的从设备后,bridge设备的mac地址会作为type-5路由的rmac。vxlan设备用于网关和控制节点(客户端)经过vxlan隧道进行双向通信,并且控制节点会创建一个vxlan vni相同的vxlan设备。

[0134] 将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备。vxlan设备的主设备设置为bridge后,当vxlan设备收到报文时,vxlan内层报文的mac地址是bridge设备的mac地址,然后报文会走bridge设备的内核协议栈。通过vxlan设备发送报文时,vxlan内层报文的源mac地址是bridge设备的mac地址。bridge设备的主设备设置为vrf设备后,网桥收到或者发送报文时,内核协议栈会查找vrf关联的策略路由表。

[0135] 通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备。虚拟网卡用于dpdk和内核之间的网络通信,dpdk和内核之间传递的数据报文是 mac报文。其中,在虚拟网卡创建完成后,创建独立的session表,dpdk可开始监听虚拟网卡。虚拟网卡的主设备设置为vrf设备后,同一个vrf设备下的vxlan从设备收到的报文的vxlan内层ip报文经过路由会从虚拟网卡发出,而dpdk会监听虚拟网卡,从而dpdk程序会收到vxlan内层ip报文。dpdk收到的报文也会通过虚拟网卡转发给内核,由内核完成vxlan报文的封装。

[0136] 在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡。

[0137] 运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点,具体地,运行mp-bgp将vxlan设备和目标网络通过 evpn type5路由通知到其他节点(其他控制节点例如客户端),其他控制节点收到vxlan设备和目标网络的路由通告后,后续将通过vxlan隧道访问目标网络。

[0138] 需要说明的是,控制节点例如客户端创建l3vni隧道的方式包括:

[0139] 在linux系统上创建vrf设备、bridge设备以及vxlan设备;将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;

[0140] 同时把当虚拟机的连接的网桥的主设备设备为vrf设备,由于虚拟机的连接的网桥是配置了ip的,所以当vrf设备成为网桥的主设备的时候会在vrf关联的策略路由表中生成一条到网桥所在网段的路由,路由的出口是虚拟机连接的网桥。同时通过mp-bgp学习,会学习到可以通过网关访问到的网络。当 l3vni隧道收到网关发来的报文时,内核协议栈查找 vrf设备关联的策略路由表,如果目的网段是网桥所在网段会将报文发送到虚拟机连接的网桥,然后由网桥把报文转发给虚拟机。另一方面当虚拟机访问的目标网络和虚拟机所在的网络不一致时,会查找vrf关联的策略路由表进行三层转发,最后通过l3vni隧道将数据报文发送到网关。

[0141] 运行MP-bgp多协议边界网关协议,以发送vxlan设备和租户网络的路由通知至网关。

[0142] 本实施例提出的报文处理方法,通过创建vrf设备、bridge设备以及vxlan设备,将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;接着通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备;而后在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡;然后运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点,实现了为网关创建l3vni隧道,进一步提升数据转发的效率。

[0143] 此外,本发明实施例还提出一种计算机可读存储介质,所述计算机可读存储介质上存储有报文处理方程序,所述报文处理方程序被处理器执行时实现如下操作:

[0144] 在通过网关的l3vni隧道接收到访问underlay网络的vxlan报文时,通过所述vxlan报文对应的虚拟网卡发送所述vxlan报文对应的第一报文;

[0145] 若通过dpdk数据平面开发套件监听到所述第一报文,则通过所述dpdk基于所述虚拟网卡查找所述第一报文对应的第一session表;

[0146] 基于所述dpdk对所述第一报文进行修改,以获得第二报文,并将所述第一报文对应的修改信息存储至所述第一session表;

[0147] 通过所述dpdk对应的物理网卡发送所述第二报文。

[0148] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0149] 基于所述dpdk更改所述第一报文中的目的ip、源ip、源端口以及目的端口,并更改第一报文中的mac地址,以获得所述第二报文;

[0150] 将所述第一报文中的目的ip、源ip、源端口以及目的端口,更改后的目的ip、更改后的源ip、更改后的源端口以及更改后的目的端口存储至所述第一session表。

[0151] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0152] 删除所述vxlan报文的报文头,以获得vxlan内层报文,并查询vrf虚拟路由转发设备关联的策略路由表,以确定所述vxlan内层报文对应的目标网络的出口是否为虚拟网卡;

[0153] 若是,则基于所述vxlan内层报文生成所述第一报文,并通过所述虚拟网卡发送所述第一报文。

[0154] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0155] 在检测到所述物理网卡接收到mac报文时,基于所述dpdk查找所述mac报文对应的第二session表;

[0156] 基于第二session表,对所述mac报文进行转换,以获得第三报文,并将所述第三报

文通过虚拟网卡发送至内核；

[0157] 通过内核对所述第三报文进行封装,以获得目标报文,并转发所述目标报文。

[0158] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0159] 在检测到所述物理网卡接收到mac报文时,通过所述dpdk查找所有的session表中,是否存在与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表;

[0160] 若存在,则将与所述mac报文中的目的ip、源ip、源端口以及目的端口匹配的session表作为所述第二session表。

[0161] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0162] 基于第二session表,对所述mac报文中的目的ip、源ip、源端口以及目的端口进行转换,并更新所述mac报文中的mac地址,以获得第三报文。

[0163] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0164] 通过内核查找vrf设备关联的策略路由表、neighbor表和系统fdb表,以获得所述mac报文对应的外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址;

[0165] 基于所述外层目的ip地址、源ip地址、内层报文的mac地址以及源mac地址,通过内核对所述第三报文进行封装,以获得目标报文。

[0166] 进一步地,所述报文处理方程序被处理器执行时还实现如下操作:

[0167] 创建vrf设备、bridge设备以及vxlan设备,将vxlan设备的主设备设置为所述bridge设备,并将bridge设备的主设备设置为vrf设备;

[0168] 通过所述dpdk创建虚拟网卡,并将所述虚拟网卡的主设备设置为所述vrf设备;

[0169] 在vrf设备的路由域中添加目标网络对应的路由,以使所述目标网络的出口为所述虚拟网卡;

[0170] 运行MP-bgp多协议边界网关协议,以发送vxlan设备和目标网络的通知至其他节点。

[0171] 需要说明的是,在本文中,术语“包括”、“包含”或者其他任何其任何变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者系统不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者系统所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、方法、物品或者系统中还存在另外的相同要素。

[0172] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0173] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在如上所述的一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,空调器,或者网络设备)执行本发明各个实施例所述的方法。

[0174] 以上仅为本发明的优选实施例,并非因此限制本发明的专利范围,凡是利用本发明说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其他相关的技术领域,均同理包括在本发明的专利保护范围内。

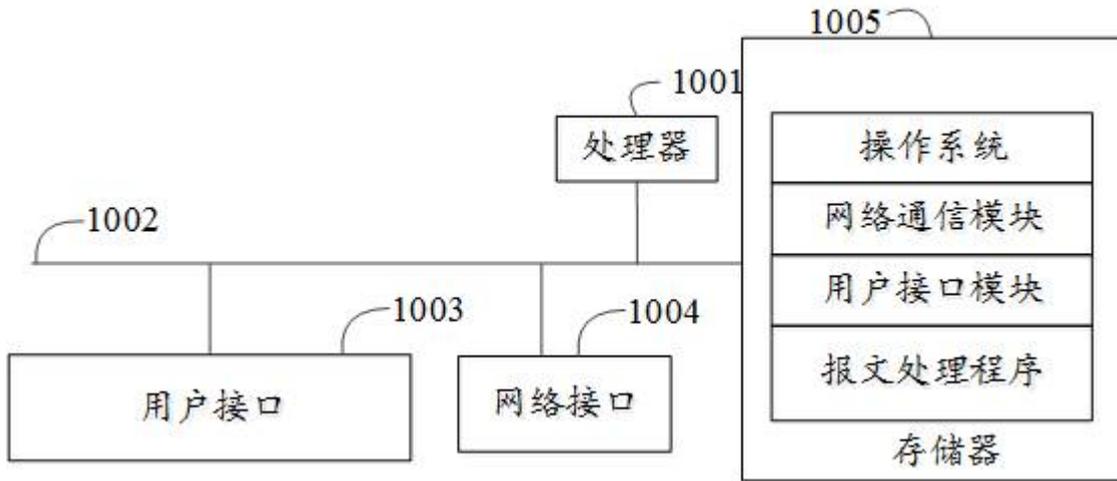


图1

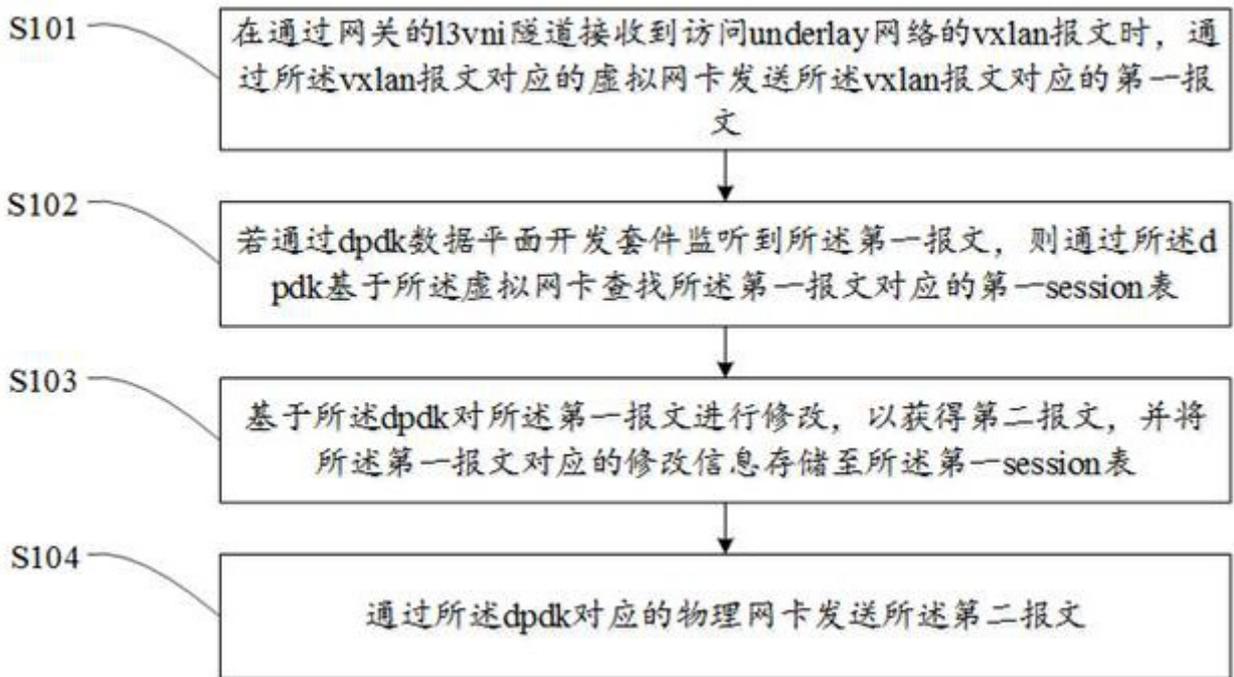


图2