(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 107370676 A
(43)申请公布日 2017.11.21

(21)申请号 201710654604.1

(22)申请日 2017.08.03

(71)申请人 中山大学
地址 510275 广东省广州市海珠区新港西路135号

(72)发明人 王臻 谢逸 马海寿

(74)专利代理机构 广州粤高专利商标代理有限公司 44102
代理人 林丽明

(51)Int.Cl.
H04L 12/725(2013.01)
H04L 12/803(2013.01)

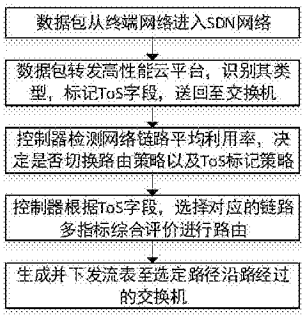
权利要求书4页 说明书15页 附图6页

(54)发明名称

一种融合QoS及负载均衡需求的路由选择方法

(57)摘要

本发明提供一种融合QoS及负载均衡需求的路由选择方法，它包括以下的步骤：首先，通过边界交换机和高性能云平台的协作实现流类型的分析与标记；其次，控制器通过网络测量得到链路实时状态信息，结合待转发流类型和网络负载均衡性计算链路综合评价指标；最后，控制器根据全局拓扑结构信息和链路综合评价指标，计算最优路由。本方法兼顾用户业务质量和网络负载均衡，使网络在实现动态负载均衡的同时，能够最大限度满足业务流的QoS需求。



CN 107370676 A

1.一种融合QoS及负载均衡需求的路由选择方法，其特征在于，包括以下步骤：

(1)终端网络将数据包发往SDN边界交换机；

(2)SDN边界交换机对已知的类型业务流进行ToS字段标注，将未知的业务流发送给高性能云平台，待高性能云平台分析识别该业务流后，SDN边界交换机根据分析结果对业务流进行ToS字段标注；

(3)SDN边界交换机把数据包转发给SDN控制器，以获得正确的转发策略；

(4)SDN控制器周期性测量网络，获得链路性能信息，并按预定的不同性能指标权重组合，计算不同性能组合模式下的链路综合评价指标，根据SDN边界交换机发来的数据包源和目的地址，及ToS字段，计算出最优路径及生成流表；

(5)SDN控制器把流表分发给特定路径上的交换节点；

(6)SDN交换机按照流表转发数据包，直到数据包到达目的端。

2.根据权利要求1所述的一种融合QoS及负载均衡需求的路由选择方法，其特征在于，上述高性能云平台的工作过程为：

(201)接收由SDN边界交换机发送的未知类型流，提取流的五元组信息；

(202)通过流量识别技术分析流，得到其业务类型；其中流量识别技术具体为使用深度包检测的技术分析，记录能够识别包的类型，无法识别的包再使用深度流检测技术基于统计信息进行分类；

(203)将分析结果发送至SDN边界交换机；其中分析结果指用于标识流的五元组和流类型的对应关系，流类型包括会话类、流类、交互类、背景类四种类型。

3.根据权利要求2所述的一种融合QoS及负载均衡需求的路由选择方法，其特征在于，SDN边界交换机是将网络功能网元以虚拟化的方式部属在通用的硬件服务器上，以硬件资源池分配的方法提供网络功能网元所需的计算、存储、网络资源，提高网络部署的灵活性；

上层建立有流标记存储虚拟模块、标记ToS虚拟模块及转发流量虚拟模块；

SDN边界交换机的工作过程为：

(301)流标记存储虚拟模块接收终端网络发送的流，然后向高性能云平台发送未知流量类型的数据包，接收由高性能云平台计算得到的分析结果，存储分析结果，存储方式为记录以流的五元组为键值、以流类型为数值的关联数组；

(302)标记ToS虚拟模块根据流标记存储虚拟模块存储的信息，判断数据包所属流类型，分析其QoS需求，标记ToS字段；

(303)转发流量虚拟模块将数据包送往SDN控制器用于寻路，SDN控制器根据ToS字段计算综合代价最低的路径，生成流表并将流表发送给交换机，交换机根据流表转发数据包。

4.根据权利要求3所述的一种融合QoS及负载均衡需求的路由选择方法，其特征在于，SDN边界交换机与高性能云平台通过一条专用链路实现相连，它们的信息交互过程如下：

(401)SDN边界交换机发送未知类型数据包至高性能云平台；SDN边界交换机将该数据包作为净荷，封装在以SDN边界交换机端口MAC地址为源物理地址、目的物理地址为广播地址的以太网数据帧中，然后从对应端口送出；

(402)高性能云平台接收该数据帧后，提取其净荷部分中的数据包，使用流五元组标识数据包所属业务流，并且识别业务流类型，得到五元组与流类型对应关系的分析结果，会话类、流类、交互类、背景类的类别编码依次为二进制码字00、01、10、11；

(403)分析结果由高性能云平台返回至SDN边界交换机;高性能云平台将五元组和流类型的数据以二进制表示,将此数据作为净荷,封装在以高性能云平台网卡MAC地址为源物理地址、目的地址为广播地址的以太网数据帧中,然后送回至SDN边界交换机。

5.根据权利要求4所述的一种融合QoS及负载均衡需求的路由选择方法,其特征在于,所述SDN控制器包含用于收集网络拓扑信息的拓扑发现模块、用于测量和收集网络链路信息的网络测量模块和用于综合评价链路,计算路由的路由功能模块;

SDN控制器周期性地与各个交换机交换信息,然后根据收集的信息为数据包生成流表,指导交换机转发数据包;SDN控制器的稳定性对整个网络的稳定性有决定性作用,合理的设置轮询周期时长能够保证稳定的网络性能,

SDN控制器的轮询周期时长的设置方法如下:根据SDN控制器计算能力设置最小轮询周期粒度τ,在不同CPU使用率的情况下,阶梯式地设置轮询周期T,具体如下:

$$T=\begin{cases}\tau, & CPU\ utilization\ rate\in[0,20\%]\\ 2\tau, & CPU\ utilization\ rate\in[20\%,40\%]\\ 3\tau, & CPU\ utilization\ rate\in[40\%,60\%]\\ 5\tau, & CPU\ utilization\ rate\in[60\%,100\%]\end{cases}。$$

6.根据权利要求5所述的一种融合QoS及负载均衡需求的路由选择方法,其特征在于,SDN控制器为业务流选择路由,需要掌握网络全局拓扑结构信息,采用所述拓扑发现模块拓扑信息收集网络拓扑信息,包括交换机间的连接关系、主机以及子网的位置信息;

(601)收集交换机间的连接关系:控制器周期性发送LLDP数据包,得到链路两端交换机以及对应端口的信息,以此获得整个SDN网络的交换机间的连接关系;

(2)收集主机及子网的位置信息:主机作为源端主动发起通信时,数据包到达交换机处,交换机没有能匹配该数据包的流表时,产生包含主机IP地址、MAC地址的PacketIn消息交由SDN控制器,此时SDN控制器记录主机IP地址、MAC地址、交换机编号、交换机端口号四元组的信息;若SDN控制器没有目的端的位置信息,由SDN控制器构造ARP请求数据包,通过PacketOut消息广播到全网交换机上非交换机互联的端口上,目的端收到此ARP请求后,发送ARP应答数据包,同样此应答包会被交换机发送至SDN控制器,据此SDN控制器能掌握网络中主机的位置信息;Openflow流表包头域中不仅能用IP地址作为匹配项,也能用子网号作为匹配项,因此通过将同一个子网下的主机聚集为子网号作为流表匹配项来减少流表项的数目。

7.根据权利要求6所述的一种融合QoS及负载均衡需求的路由选择方法,其特征在于,网络测量模块是基于SDN实现网络测量,测量的指标包括:链路时延、链路丢包率、链路最大可用带宽、链路负载;

链路时延测量的实现过程为:

(711)SDN控制器向交换机A下发一个Packetout报文;报文的数据段携带了一个约定的协议报文,其报文的数据段携带了控制器下发报文时的时间戳;Packetout报文的动作指示交换机将其泛洪或转发到某端口;

(712)与交换机A相邻的交换机B收到了交换机A发送过来的数据包,无流表项可匹配,从而将此数据包封装在Packetin消息并发送到SDN控制器;SDN控制器接收到这个数据包之后,和当下时间相减,得到时间差T_1;其时间差约等于数据包从控制器到交换机A、交换机A

到交换机B、交换机B到SDN控制器的时延总和；

(713)同理，SDN控制器向交换机B发送一个同样的报文；然后SDN控制器从交换机A收到Packet_in报文，记录下时间差T_2；所以T_1+T_2等于SDN控制器到交换机A的RTT、SDN控制器到交换机B的RTT和交换机A到交换机B的时延RTT三者之和；

(714)SDN控制器向交换机A和交换机B分别发送带有时间戳的echo request；交换机收到之后即刻回复携带echo request时间戳的echo reply消息；所以控制器可以通过echo reply的时间戳减去echo request携带的时间，从而得到对应交换机和控制器之间的RTT；通过这种方法测得SDN控制器到交换机A,B的RTT分别为T_a,T_b；

(715)交换机A到交换机B的RTT为$T_1+T_2-T_a-T_b$；假设往返时间一样，则交换机A到交换机B的链路时延为$(T_1+T_2-T_a-T_b)/2$；

链路丢包率测量方法如下：

(721)SDN控制器周期性向交换机发送ofp_port_status_request消息，收集交换机端口统计信息，得到ofp_port_status_reply消息中表示端口发送/接收比特数的计数器rx_packets和tx_packets；

(722)交换机n与m之间链路丢包率：$loss=\max(loss_{(m,n)},loss_{(n,m)})$，其中$loss_{(m,n)}$表示交换机m到交换机n方向链路丢包率，其值应为：

$$loss_{(m,n)}=\frac{(tx\_packets_{\{t,m\}}-tx\_packets_{\{t-1,m\}})-(rx\_packets_{\{t,n\}}-rx\_packets_{\{t-1,n\}})}{tx\_packets_{\{t,m\}}-tx\_packets_{\{t-1,m\}}}$$

下标{t,m}表示交换机m在第t轮次计数器的值，后文中下标代表相似含义；

链路最大可用带宽以及负载测量方法如下：

(731)在SDN控制器与交换机建立连接时，SDN控制器发送ofp_feature_request消息询问交换机支持功能，交换机通过ofp_port_features_reply消息，报告端口速率、双工模式、以及支持的链路介质，认为端口汇报速率即为最大可用带宽；

(732)SDN控制器周期性向交换机发送ofp_port_status_request消息，收集交换机端口统计信息，得到ofp_port_status_reply消息中表示端口发送/接收比特数的计数器rx_bytes和tx_bytes；交换机m与n之间链路已用带宽：

$Bandwidth=\max(Bandwidth_{(m,n)},Bandwidth_{(n,m)})$

$$Bandwidth_{(m,n)}=\frac{(tx\_bytes_{\{t,m\}}+rx\_bytes_{\{t,m\}}-tx\_bytes_{\{t-1,m\}}-rx\_bytes_{\{t-1,m\}})\times 8}{T}。$$

8.根据权利要求7所述的一种融合QoS及负载均衡需求的路由选择方法，其特征在于，路由功能模块用于接收SDN交换机发送的PacketIn消息，为之计算路由、生成流表、并且将流表下发至对应交换机；路由功能模块建立在拓扑发现模块和网络测量模块之上，路由功能模块需要拓扑发现模块提供的全局拓扑结构信息，需要网络测量模块提供的实时网络链路参数以优化路由；其工作过程为：

(801)SDN交换机收到无法匹配流表的数据包时，将此数据包封装在PacketIn消息中发送给SDN控制器，SDN控制器收到此消息后，解析出源地址和目的地址；

(802)根据拓扑发现模块提供的全局拓扑结构信息，构造存储拓扑信息的无向图；

(803)根据网络测量模块提供的实时网络链路参数，计算链路综合评价指标，更新无向图中边的代价值；

(804)使用路由算法为数据包选择最优路由；

其中步骤(803)根据网络测量模块提供的实时网络链路参数，计算链路的归一化时延代价、丢包率代价、负载代价，然后按照加权求和的方法计算链路综合评价指标，更新无向图中边的代价值；

链路综合指标$Cost_i=\sum_{p\in\{delay,loss,load\}}w_p\cdot c_{i,p}$，权重值$w_p$刻画客户端QoS需求和网络负载均衡性的重要程度，权重值$w_p=\frac{a_p}{\sum_{p\in\{delay,loss,load\}}a_p}$，其中$a_p$表示不同性能指标的权值因子；SDN控制器根据数据包IP包头中ToS字段，区分不同数据业务流的QoS需求；结合客户端QoS需求和网络负载均衡性，设置不同指标的权值因子，计算其权重值；同时根据网络负载均衡性设置a_{load}，网络负载越不均衡，则a_{load}值应越大，具体值需结合网络场景设计；将计算得到的链路综合评价指标作为无向图边的代价值；

最优路由是指从源端到目的端的最优路径，是所有从源端到目的端的可达路径当中逐段链路综合评价指标和最低的路径，即无向图中源顶点到目的顶点所有可达路径中逐段链路代价和最小的路径。

9.根据权利要求8所述的一种融合QoS及负载均衡需求的路由选择方法，其特征在于，SDN控制器采用以下方式计算最优路径：

(901)预先设置阈值，计算网络链路平均利用率；

(902)当网络链路平均利用率低于阈值，认为大多数网络链路较为空闲，SDN控制器以最短路径算法为业务流计算最优路由；

(903)当网络链路平均利用率高于阈值，认为大多数网络链路较为繁忙，SDN控制器以K最短路径Yen算法得到前K条最优路由，而后比较K条路由，选出路径负载均衡参数l_r最小的路径r作为最优路由。

一种融合QoS及负载均衡需求的路由选择方法

技术领域

[0001] 本发明涉及网络技术领域，更具体地，涉及一种融合QoS及负载均衡需求 的路由选择方法。

背景技术

[0002] 基于互联网的通信方式在不断扩张和普及的过程中，已经成为现代社会生活 中不可或缺的基本元素之一。随着网络承载数据和业务的不断演化，网络终端客 户和网络基础设施运营者对网络通信的要求也在不断提升，主要体现在两个方 面：(1)终端客户不再满足于简单的“发送/接收”，而是希望网络能够突破“数据 管道”的局限性，提供灵活的数据转发策略、保障不同类型的业务流服务质量 (Quality of Service，QoS)，从而提升用户体验质量(Quality of Experience，QoE)； (2)网络运营者不再停留在简单的“转发”工作，而是希望通过合理的优化方案有 效提高网络负载的均衡性，从而保证网络链路的可用性和稳定性，并以最优的网 络资源成本完成客户数据流的通信。

[0003] 针对上述要求，常见的解决方法主要有两种：以客户服务为中心的方案和以 网络性能为中心的方案。前者主要通过网络基础设施的升级和网络资源扩容的方 法保障网络客户业务流的QoS和终端用户的QoE。其主要问题在于：(1)升级与 扩容会导致网络运营者的成本增加；(2)以QoS/QoE为唯一目标，会导致大量低 配置链路的网络资源闲置，从而造成网络资源浪费，而一些优质链路则可能频繁 出现拥塞，并由此产生“呼损”。以网络性能为中心的方案则依据给定的调度策略， 通过负载均衡，避免链路过度的“空闲”和“拥塞”，提高网络的整体利用率、可用 性和稳定性。但其主要问题在于负载均衡的实现难以同时兼顾终端客户的 QoS/QoE需求。

[0004] 从现有的网络通信管理策略可见，非“0”即“1”的方法已经难以满足现代通信 的需要，设计一种能融合客户服务质量和网络负载均衡的路由选择方法是现代互 联网中亟待解决的关键问题之一，也是现代大型数据中心网络和大规模企业网所 面临的新挑战之一，因为这类网络的流量大、实时性要求高，而且业务流的类型 纷繁复杂。

[0005] 目前，面向SDN的负载均衡研究可分为以下几类：

[0006] (1)基于检测大流的方法：文献“Long H,Shen Y,Guo M,et al.LABERIO: Dynamic load-balanced routing in OpenFlow-enabled networks[C]//Advanced Information Networking and Applications(AINA),2013 IEEE 27th International Conference on.IEEE,2013:290-297.”和“Braun W,Menth M.Load-dependent flow splitting for traffic engineering in resilient OpenFlow networks[C]//Networked Systems (NetSys),2015 International Conference and Workshops on.IEEE,2015: 1-5”首先定义大流的概念，提出检测、定位大流的方法，动态检测网络中最繁忙 的链路，通过判别该链路上是否有大流，将该链路上大流分成若干小流，通过若 干条替代路径同时转发。文献“万晓榆，张丹，赵书宜，胡敏，樊自甫，王正强.一种基 于链路实时负载的SDN动态负载均衡调度方法[P].重庆：CN106411733A,2017-02-15.”提出的方法在超过预设的网络负载均衡度

阈值，检 测定位大流，若符合调度条件则对大流进行分流和调度。以上基于检测大流的方法重点关注负载均衡，未考虑业务流QoS需求。

[0007] (2)基于QoS的流量调度：在保证业务流服务质量的同时实现负载均衡，文 献“熊琦，郑亮，于治楼.一种SDN网络下基于业务特征的调度方法[P].山东：CN105656799A，2016-06-08.”为业务流设置不同的优先级，权衡业务优先级和链 路状态信息，为不同业务计算最优传输路径，但其业务优先级仅与业务最大容忍 时延相关，忽略其他影响服务质量的因素。文献“王英，李春，李云，吴广富，郑焕平. 网络效用最大化的SDN负载均衡方法[P].重庆：CN106656847A，2017-05-10”提 出利用IP层服务类型(Type of Service，ToS)字段区分业务类型，综合网络效用和 负载均衡为业务流计算最佳路由，但未明确如何标记ToS字段。

[0008] (3)基于算法设计的调度方法：文献“Li Y，Pan D.OpenFlow based load balancing for Fat-Tree networks with multipath support[C]//Proc.12th IEEE International Conference on Communications(ICC'13)，Budapest，Hungary.2013：1-5.”提出基于数据中心胖树拓扑网络的动态负载均衡算法，使用单跳贪婪的策 略，实现了局部路由最优决策，但是因未考虑全局网络状态，路由策略可能不是 全局最优，导致部分链路负载过重。文献“魏凯.基于蚁群算法SDN负载均衡的 研究[D].吉林大学，2015.”和“文强.SDN网络业务量工程技术研究[D].电子科技大 学，2016.”分别利用蚁群算法和遗传算法实现SDN的负载均衡和多QoS路由，但 它们均存在算法收敛速度慢、局部最优的问题。文献“Chen-xiao C，Ya-bin X. Research on load balance method in SDN[J].International Journal of Grid and Distributed Computing，2016，9(1)：25-36.”利用神经网络算法使用链路利用率、丢 包率、时延和跳数预测链路综合负载，从而选择链路综合负载最轻的路径，但是 由于调参是根据模型性能决定，学习训练得到的模型不具有普适性，即模型适用 于一种网络。文献“段元新，倪晓军，章韵.多指标综合评价的负载均衡路由策略 研究[J].小型微型计算机系统，2017，38(2)：209-212.”和“段元新，倪晓军，章韵.多指标综合评价的负载均衡路由策略研究[J].小型微型计算机系统，2017， 38(2)：209-212.”利用最短路径算法得到k条最短路径，而后对多条路径进行多指 标综合评价，从中选出最优转发路径，但是多指标评价归一化方法不够准确，单 一的加权评价模型未考虑业务流不同的QoS需求，同时为所有终端之间计算并 评价K条路径需占用较多的计算和存储资源。

[0009] 基于传统网络架构的QoS保障方法主要有综合服务(Integrated Service，IS) 模型和区分服务(Differentiated Service，DS)模型。IS使用资源预留协议，通 过预留网络资源保障高优先级数据流的QoS需求，要求所有网络节点监控网络 状态、存储数据流的信息，存在策略实现复杂、扩展性差的缺陷。DS以数据包 ToS字段标记流类型，优先处理转发ToS等级高的数据包，需要在边界交换机部 署ToS字段标记，但无法根据全局网络状态调整优先级标记的策略。

[0010] QoS路由“段元新.基于OpenFlow的负载均衡路由策略的研究[D].南京邮 电大学，2016.”是一种基于业务QoS需求和网络可用资源进行路由的机制，是网 络服务控制的核心技术问题。文献“崔勇，吴建平，徐恪，等.互联网络服务质量 路由算法研究综述[J].软件学报，2002，11.”提出一种分布式路由算法DRA，文献 “Chen S，Nahrstedt K.Distributed

quality-of-service routing in high-speed networks based on selective probing [C]//Local Computer Networks,1998.LCN'98. Proceedings.,23rd Annual Conference on.IEEE,1998:80-89."提出H_MCOP (Heuristic Multi-Constrained Optimal Path Selection)算法。上述路由算法能够较 好地解决多QoS路由问题,但由于其仅考虑QoS服务控制,忽略了网络负载均 衡。

[0011] 为实现QoS约束下的负载均衡,很多研究工作提出了不同的算法,文献 "Tekaya M,Tabbane N,Tabbane S.Multipath routing with load balancing and QoS in ad hoc network[J].IJCSNS international Journal of computer science and network security,2010,10(8):280-286."在无线自组织网络中根据QoS约束利用多径路由 实现负载均衡,文献"Casetti C,Cigno R L,Mellia M.QoS-aware routing schemes based on hierarchical load-balancing for integrated services packet networks[C]//Communications,1999.ICC'99.1999 IEEE International Conference on.IEEE,1999,1:489-493."提出一种基于负载均衡的静态路由算法,能够支持多 种QoS要求,文献"孙杰,李莉,沈苏彬.一种基于QoS和动态负载均衡的路由 策略[J].计算机技术与发展,2016,26(11):188-194."在SDN网络中设计SAS (Scheduling According to Stickiness)算法,通过调度实现负载均衡,提出用于 估计调度对流性能影响的链路粘值,重路由过程中兼顾流QoS需求,文献"刘军 伟.异构网络中基于QoS负载均衡的接纳控制算法研究[D].兰州交通大学, 2015."在异构网络中用户有多个供选择接入的无线网络的场景下,以用户QoS需 求和网络负载均衡综合效益最大化为优化目标,首先通过多属性评价确定最佳满 足QoS需求的无线网络,而后通过监控网络,寻找满足网络负载均衡要求下最 优QoS评价的无线网络。文献"Tekaya M,Tabbane N,Tabbane S.Multipath routing with load balancing and QoS in ad hoc network[J].IJCSNS international Journal of computer science and network security,2010,10(8):280-286."和"刘军伟.异构网络 中基于QoS负载均衡的接纳控制算法研究[D].兰州交通大学,2015."的方案适用 于低流量的无线网络环境,但对于大规模、高负载的复杂数据中心网络或园区网, 则难以发挥其优势。文献"Casetti C,Cigno R L,Mellia M.QoS-aware routing schemes based on hierarchical load-balancing for integrated services packet networks[C]//Communications,1999.ICC'99.1999IEEE International Conference on.IEEE,1999,1:489-493."提出的静态路由算法缺乏灵活性和扩展性,文献"孙 杰,李莉,沈苏彬.一种基于QoS和动态负载均衡的路由策略[J].计算机技术与 发展,2016,26(11):188-194."和"刘军伟.异构网络中基于QoS负载均衡的接纳控 制算法研究[D].兰州交通大学,2015."提出的算法复杂度高,而且未考虑端到端 QoS架构的实施方法。可见现有方法难以在保障网络性能的同时兼顾用户的体验 质量。

发明内容

[0012] 为克服现有技术存在的局限性,本发明提出一种融合QoS及负载均衡需求 的路由选择方法。该方法通过ToS字段标记网络流的QoS类型,从而确定不同 选路性能指标的权重,借助全局网络视图计算链路在特定指标权重方案下的综合 评价,依据各条链路的综合评价指标为网络流选择符合其QoS需求及有利于网 络负载均衡的最优路由,本方法综合考

虑了用户业务质量和网络负载均衡的需 求，有效解决现有技术的不足。

[0013] 为了实现上述目的，本发明的技术方案为：

[0014] 一种融合QoS及负载均衡需求的路由选择方法，包括以下步骤：

[0015] (1)终端网络将数据包发往SDN边界交换机；

[0016] (2)SDN边界交换机对已知的类型业务流进行ToS字段标注，将未知的业 务流发送给高性能云平台，待高性能云平台分析识别该业务流后，SDN边界交 换机根据分析结果对业务流进行ToS字段标注；

[0017] (3)SDN边界交换机把数据包转发给SDN控制器，以获得正确的转发策 略；

[0018] (4)SDN控制器周期性测量网络，获得链路性能信息，并按预定的不同性 能指标权重组合，计算不同性能组合模式下的链路综合评价指标，根据SDN边 界交换机发来的数据包源和目的地址，及ToS字段，计算出最优路径及生成流表；

[0019] (5)SDN控制器把流表分发给特定路径上的交换节点；

[0020] (6)SDN交换机按照流表转发数据包，直到数据包到达目的端。

[0021] 优选的，上述高性能云平台的工作过程为：

[0022] (201)接收由SDN边界交换机发送的未知类型流，提取流的五元组信息；

[0023] (202)通过流量识别技术分析流，得到其业务类型；其中流量识别技术具体 为使用深度包检测的技术分析，记录能够识别包的类型，无法识别的包再使用深 度流检测技术基于统计信息进行分类；

[0024] (203)将分析结果发送至SDN边界交换机；其中分析结果指用于标识流的五 元组和流类型的对应关系，流类型包括会话类、流类、交互类、背景类四种类型。

[0025] 优选的，SDN边界交换机是将网络功能网元以虚拟化的方式部属在通用的 硬件服务器上，以硬件资源池分配的方法提供网络功能网元所需的计算、存储、 网络资源，提高网络部署的灵活性；

[0026] 上层建立有流标记存储虚拟模块、标记ToS虚拟模块及转发流量虚拟模块；

[0027] SDN边界交换机的工作过程为：

[0028] (301)流标记存储虚拟模块接收终端网络发送的流，然后向高性能云平台发 送未知流量类型的数据包，接收由高性能云平台计算得到的分析结果，存储分析 结果，存储方式为记录以流的五元组为键值、以流类型为数值的关联数组；

[0029] (302)标记ToS虚拟模块根据流标记存储虚拟模块存储的信息，判断数据包 所属流类型，分析其QoS需求，标记ToS字段；

[0030] (303)转发流量虚拟模块将数据包送往SDN控制器用于寻路，SDN控制器 根据ToS字段计算综合代价最低的路径，生成流表并将流表发送给交换机，交换 机根据流表转发数据包。

[0031] 优选的，SDN边界交换机与高性能云平台通过一条专用链路实现相连，它 们的信息交互过程如下：

[0032] (401)SDN边界交换机发送未知类型数据包至高性能云平台；SDN边界交 换机将该数据包作为净荷，封装在以SDN边界交换机端口MAC地址为源物理 地址、目的物理地址为广播地址的以太网数据帧中，然后从对应端口送出；

[0033] (402)高性能云平台接收该数据帧后，提取其净荷部分中的数据包，使用 流五元

组标识数据包所属业务流，并且识别业务流类型，得到五元组与流类型对 应关系的分析结果，会话类、流类、交互类、背景类的类别编码依次为二进制码 字00、01、10、11；

[0034] (403)分析结果由高性能云平台返回至SDN边界交换机；高性能云平台将 五元组和流类型的数据以二进制表示，将此数据作为净荷，封装在以高性能云平 台网卡MAC地址为源物理地址、目的地址为广播地址的以太网数据帧中，然后 送回至SDN边界交换机。

[0035] 优选的，所述SDN控制器包含用于收集网络拓扑信息的拓扑发现模块、用 于测量和收集网络链路信息的网络测量模块和用于综合评价链路，计算路由的路 由功能模块；

[0036] SDN控制器周期性地与各个交换机交换信息，然后根据收集的信息为数据 包生成流表，指导交换机转发数据包；SDN控制器的稳定性对整个网络的稳定 性有决定性作用，合理的设置轮询周期时长能够保证稳定的网络性能，

[0037] SDN控制器的轮询周期时长的设置方法如下：根据SDN控制器计算能力设 置最小轮询周期粒度τ，在不同CPU使用率的情况下，阶梯式地设置轮询周期T，具体如下：

[0038] $$T=\begin{cases}\tau, & CPU\ utilization\ rate\in[0,20\%]\\ 2\tau, & CPU\ utilization\ rate\in[20\%,40\%]\\ 3\tau, & CPU\ utilization\ rate\in[40\%,60\%]\\ 5\tau, & CPU\ utilization\ rate\in[60\%,100\%]\end{cases}。$$

[0039] 优选的，SDN控制器为业务流选择路由，需要掌握网络全局拓扑结构信息，采用所述拓扑发现模块拓扑信息收集网络拓扑信息，包括交换机间的连接关系、主机以及子网的位置信息；

[0040] (601)收集交换机间的连接关系：控制器周期性发送LLDP数据包，得到 链路两端交换机以及对应端口的信息，以此获得整个SDN网络的交换机间的连 接关系；

[0041] (602)收集主机及子网的位置信息：主机作为源端主动发起通信时，数据 包到达交换机处，交换机没有能匹配该数据包的流表时，产生包含主机IP地址、MAC地址的PacketIn消息交由SDN控制器，此时SDN控制器记录主机IP地址、MAC地址、交换机编号、交换机端口号四元组的信息；若SDN控制器没有目的 端的位置信息，由SDN控制器构造ARP请求数据包，通过PacketOut消息广播 到全网交换机上非交换机互联的端口上，目的端收到此ARP请求后，发送ARP 应答数据包，同样此应答包会被交换机发送至SDN控制器，据此SDN控制器能 掌握网络中主机的位置信息；Openflow流表包头域中不仅能用IP地址作为匹配项，也能用子网号作为匹配项，因此通过将同一个子网下的主机聚集为子网号作 为流表匹配项来减少流表项的数目。

[0042] 优选的，网络测量模块是基于SDN实现网络测量，测量的指标包括：链路 时延、链路丢包率、链路最大可用带宽、链路负载；

[0043] 链路时延测量的实现过程为：

[0044] (711)SDN控制器向交换机A下发一个Packetout报文；报文的数据段携 带了一个约定的协议报文，其报文的数据段携带了控制器下发报文时的时间戳；Packetout报文的动作指示交换机将其泛洪或转发到某端口；

[0045] (712)与交换机A相邻的交换机B收到了交换机A发送过来的数据包，无流表项可匹配，从而将此数据包封装在Packetin消息并发送到SDN控制器；SDN控制器接收到这个数据包之后，和当下时间相减，得到时间差T_1；其时间差 约等于数据包从控制器到交换机A、

交换机A到交换机B、交换机B到SDN控 制器的时延总和；

[0046] (713)同理，SDN控制器向交换机B发送一个同样的报文；然后SDN控 制器从交换机A收到Packet_in报文，记录下时间差T_2；所以T_1+T_2等于SDN控 制器到交换机A的RTT、SDN控制器到交换机B的RTT和交换机A到交换机B 的时延RTT三者之和；

[0047] (714)SDN控制器向交换机A和交换机B分别发送带有时间戳的echo request；交换机收到之后即刻回复携带echo request时间戳的echo reply消息；所 以控制器可以通过echo reply的时间戳减去echo request携带的时间，从而得到 对应交换机和控制器之间的RTT；通过这种方法测得SDN控制器到交换机A，B 的RTT分别为T_a，T_b；

[0048] (715)交换机A到交换机B的RTT为$T_1+T_2-T_a-T_b$；假设往返时间一样， 则交换机A到交换机B的链路时延为$(T_1+T_2-T_a-T_b)/2$；

[0049] 链路丢包率测量方法如下：

[0050] (721)SDN控制器周期性向交换机发送ofp_port_status_request消息，收集 交换机端口统计信息，得到ofp_port_status_reply消息中表示端口发送/接收比特 数的计数器rx_packets和tx_packets；

[0051] (722)交换机n与m之间链路丢包率：$loss=\max(loss_{(m,n)},loss_{(n,m)})$，其中$loss_{(m,n)}$表示交换机m到交换机n方向链路丢包率，其值应为：

[0052] $$loss_{(m,n)}=\frac{(tx\_packets_{\{t,m\}}-tx\_packets_{\{t-1,m\}})-(rx\_packets_{\{t,n\}}-rx\_packets_{\{t-1,n\}})}{tx\_packets_{\{t,m\}}-tx\_packets_{\{t-1,m\}}}$$

[0053] 下标{t,m}表示交换机m在第t轮次计数器的值，后文中下标代表相似含义；

[0054] 链路最大可用带宽以及负载测量方法如下：

[0055] (731)在SDN控制器与交换机建立连接时，SDN控制器发送 ofp_feature_request消息询问交换机支持功能，交换机通过ofp_port_features_reply 消息，报告端口速率、双工模式、以及支持的链路介质，认为端口汇报速率即为 最大可用带宽；

[0056] (732)SDN控制器周期性向交换机发送ofp_port_status_request消息，收集 交换机端口统计信息，得到ofp_port_status_reply消息中表示端口发送/接收比特 数的计数器rx_bytes和tx_bytes；交换机m与n之间链路已用带宽：

[0057] $Bandwidth=\max(Bandwidth_{(m,n)},Bandwidth_{(n,m)})$

[0058] $$Bandwidth_{(m,n)}=\frac{(tx\_bytes_{\{t,m\}}+rx\_bytes_{\{t,m\}}-tx\_bytes_{\{t-1,m\}}-rx\_bytes_{\{t-1,m\}})\times 8}{T}$$。

[0059] 优选的，路由功能模块用于接收SDN交换机发送的PacketIn消息，为之计 算路由、生成流表、并且将流表下发至对应交换机；路由功能模块建立在拓扑发 现模块和网络测量模块之上，路由功能模块需要拓扑发现模块提供的全局拓扑结 构信息，需要网络测量模块提供的实时网络链路参数以优化路由；其工作过程为：

[0060] (801)SDN交换机收到无法匹配流表的数据包时，将此数据包封装在 PacketIn消息中发送给SDN控制器，SDN控制器收到此消息后，解析出源地址 和目的地址；

[0061] (802)根据拓扑发现模块提供的全局拓扑结构信息，构造存储拓扑信息的 无向图；

[0062] (803)根据网络测量模块提供的实时网络链路参数，计算链路综合评价指 标，更新无向图中边的代价值；

[0063] (804)使用路由算法为数据包选择最优路由；

[0064] 其中步骤(803)根据网络测量模块提供的实时网络链路参数，计算链路的 归一化时延代价、丢包率代价、负载代价，然后按照加权求和的方法计算链路综 合评价指标，更新无向图中边的代价值；

[0065] 链路综合指标$Cost_i=\sum_{p\in\{delay,loss,load\}}W_p \cdot c_{i,p}$，权重值$w_p$刻画客户端QoS需求 和网络负载均衡性的重要程度，权重值$w_p=\frac{a_p}{\sum_{p\in\{delay,loss,load\}}a_p}$，其中$a_p$表示不同 性能指标的权值因子；SDN控制器根据数据包IP包头中ToS字段，区分不同数 据业务流的QoS需求；结合客户端QoS需求和网络负载均衡性，设置不同指标 的权值因子，计算其权重值；同时根据网络负载均衡性设置a_{load}，网络负载越不 均衡，则a_{load}值应越大，具体值需结合网络场景设计；将计算得到的链路综合评 价指标作为无向图边的代价值；

[0066] 最优路由是指从源端到目的端的最优路径，是所有从源端到目的端的可达路 径当中逐段链路综合评价指标和最低的路径，即无向图中源顶点到目的顶点所有 可达路径中逐段链路代价和最小的路径。

[0067] 优选的，SDN控制器采用以下方式计算最优路径：

[0068] (901)预先设置阈值，计算网络链路平均利用率；

[0069] (902)当网络链路平均利用率低于阈值，认为大多数网络链路较为空闲， SDN控制器以最短路径算法为业务流计算最优路由；

[0070] (903)当网络链路平均利用率高于阈值，认为大多数网络链路较为繁忙， SDN控制器以K最短路径Yen算法得到前K条最优路由，而后比较K条路由， 选出路径负载均衡参数l_r最小的路径r作为最优路由。

[0071] 本发明考虑到SDN架构和网络功能虚拟化(Network Function Virtualization，NFV)技术已经被广泛应用于现代数据中心网络和大规模企业网，本发明将基于 SDN体系与NFV技术，设计一种面向大规模高速网络的新型负载均衡方法，该 方法融合业务流的QoS需求及网络负载的均衡性作为数据路由的双重约束条件， 从而使网络在实现动态负载均衡的同时，能够最大限度满足业务流的QoS需求。

[0072] 首先，通过边界交换机和高性能云平台的协作实现流类型的分析与标记；其 次，控制器通过网络测量得到链路实时状态信息，结合待转发流类型和网络负载 均衡性计算链路综合评价指标；最后，控制器根据全局拓扑结构信息和链路综合 评价指标，计算最优路由。

附图说明

[0073] 图1为ToS标记方式示意图。

[0074] 图2为系统总框架图。

[0075] 图3为数据包处理过程流程图。

[0076] 图4为业务识别分类流程图。

[0077] 图5为NFV架构示意图。

[0078] 图6为边界交换机标记ToS流程图。

[0079] 图7为以太网帧格式示意图。

[0080] 图8为五元组与业务类型编码示例图。

[0081] 图9为SDN控制器功能模块示意图。

[0082] 图10为拓扑发现过程示意图。

[0083] 图11为时延测量过程示意图。

[0084] 图12为ofp_port_status消息交互过程图。

[0085] 图13为ofp_port_status消息请求及答复报文格式图。

[0086] 图14为ofp_feature消息交互过程图。

[0087] 图15为ofp_feature消息请求及答复报文格式图。

[0088] 具体实施部分

[0089] 符号与定义说明：

[0090] IETF针对网络性能评估，提出了一系列性能指标，包括：时延、网络带宽、丢包率、路径利用率等。本发明采用“带宽利用率”衡量网络链路性能，利用“链 路时延”与“丢包率”刻画业务流的服务质量。

[0091] 本发明采用的符号与定义说明如下：

[0092] $\mathbb{G}=<\mathbb{N},\mathbb{E}>$表示网络拓扑，其中＝$\{\mathbb{N}=\{S_i\,|\,i=1,...,n\}\bigcup\{Subnet_i\,|\,i=1,...,m\}$表 示网络中的交换机节点和子网节点集合，$\mathbb{E}=\{(i,j)\,|\,S_i$与$S_j$相连，$S_j\in N,S_j\in N\}$ 表示节点间的链路集合，$|\mathbb{E}|$表示网络中的链路数目，其中(i，j)表示交换机S_i与 交换机S_j之间的连接关系，且(i，j)与(j，i)表示同一条链路，约定以(i，j)为下标 的参数指链路(i，j)的参数。链路(i，j)参数标记如下：其时延值用$delay_{(i,j)}$表示， 其丢包率用$loss_{(i,j)}$表示，其负载用$load_{(i,j)}$表示，其链路最大可用带宽用 $Band_{(i,j),max}$，链路代价用$C_{(i,j)}$表示。

[0093] 定义网络链路平均利用率为：$\bar{L}=\frac{1}{|\mathbb{E}|}\sum_{(i,j)\in\mathbb{E}}\frac{load_{(i,j)}}{Band_{(i,j),max}}$。

[0094] 定义网络负载均衡度：$LBD=\frac{1}{|\mathbb{E}|}\sum_{(i,j)\in\mathbb{E}}(load_{(i,j)}-\overline{LOAD})^2$，$\overline{LOAD}$表示网络 中所有链路的平均负载。网络负载均衡度越小则网络负载越平衡，可用于调整网 络负载，避免局部链路拥塞。

[0095] r＝(p_1，p_2，...，p_n)表示一条由源到达目的的路径，它包含n条链路，定义其路 径负载均衡参数为$l_r=e_r/f_r$，其中f_r表示路径r包含链路数目，e_r表示路径中负 载超过网络链路平均负载的链路数目，l_r值越小，路径r作为新业务流的路由造 成链路拥塞可能性越小。路径r端到端时延越小、丢包率越低，则认为路径r的 路由性能越优。路径r上第i条链路$p_{r,i}$的时延用$delay_{r,i}$表示，丢包率用$loss_{r,i}$表 示，负载用$load_{r,i}$表示。根据时延的可加性，即路径总时延等于逐段链路时延总 和，$D_r=\sum_{p_{r,i}\in r}delay_{r,i}$，以及丢包率的可乘性，即路径总丢包率等于逐段链路丢包 率的乘积，$Loss_r=1-\prod_{p_{r,i}\in r}(1-loss_{r,i})$。进一步为网络中链路(i，j)定义三个归一化 代价函数：

[0096] ●时延代价函数：$c_{(i,j),delay}=\frac{delay_{(i,j)}}{delay_{\max}}$，$delay_{max}$指所有链路时延的最大值；

[0097] ●丢包率代价函数：$c_{(i,j),loss}=\frac{\ln(1-loss_{(i,j)})}{\ln(1-loss_{\max})}$，$loss_{max}$指所有链路丢包率的最 大值；

[0098] ●负载代价函数：$c_{(i,j),load}=\begin{cases}\frac{load_{(i,j)}}{Band_{(i,j),\max}\cdot\overline{L}}, \frac{load_{(i,j)}}{Band_{(i,j),\max}}<\overline{L}\\ 1, otherwise\end{cases}$，通过优化路由 实现负载均衡，使新业务流倾向于分配到较为空闲的路径。

[0099] 其中，时延和丢包率体现了客户端QoS需求，网络负载刻画了当前网络侧 的负载均衡程度。为使选路能融合QoS和负载均衡性要求，根据上述三个代价 函数，定义以下链路综合评价指标：

[0100] $C(i,j)=\sum_{p\in\{delay,loss,load\}}W_p\cdot c(i,j),p$

[0101] w_p表示三个性能指标的权重，用于刻画不同类型网络业务对这三个性能的 需求程度。

[0102] 采用五元组方法(src_ip,dst_ip,proto,src_port,dst_por)t标识网络流，其 中src_ip表示源IP地址、dst_ip表示目的IP地址、proto表示传输层协议、src_port 表示源端口号、dst_port表示目的端口号。

[0103] 业务类型可根据具体的应用场景和需要灵活定义，以下通过一个简单的实例 进行说明：根据常见网络应用的特点，可以如所表1示将网络业务分为会话类、流类、交互类、背景类四类业务。会话类业务包括语音业务、视频会议等，它对 时延十分敏感，但容许一定的丢包率；流类业务包括音视频等流媒体业务，由于 流类业务的单向传输特点，其对时延要求比会话类业务要求低，容许一定的丢包 率；交互类业务指终端与远程服务器通信交互的业务，如Web服务、游戏服务 等。交互类业务的时延要求取决于用户对等待时间的容忍度，一般比会话类业务 要求的时延长。背景类业务包括SMTP、FTP等对时延无特殊要求但对丢包率要 求很高的业务。

[0104] 表1 网络业务QoS类型

[0105]

QoS 类别	会话类	流类	交互类	背景类
传输时延	严格限制（实时）	限制（实时）	宽松（非实时）	无限制（非实时）
时延抖动	严格限制	限制	宽松	无限制
低误码率	否	否	是	是

[0106] 由于SDN网络根据业务类型信息为数据包进行转发，因此需要对这些信息 进行编码，并在Open flow流表和数据包头部信息中进行定义，方法如下：

[0107] 如图1所示，Openflow的流表匹配域中包含十二个字段，其中“IP ToS”字段 包含6个比特，本发明使用其前4个比特，分为两组，每组两个比特。第一组包 括0和1位比特，用于表示业务对时延的要求程度；第二组包括2和3位比特， 表示业务对丢包率的要求程度。每组两位二进制比特的取值可以表示4种不同的 需求类型，包括：00表示“不要求”，01表“低要求”，10表示“中等要求”，11表 示“高要求”。因此利用“IP ToS”可以定义16种不同需求属性的业务流。

[0108] 数据包携带业务类型信息的方法是利用IP头部已有的ToS字段的前4个比 特，定义方法与图1一致。

[0109] 本发明是基于SDN架构提出一种融合QoS及负载均衡需求的路由方法；采 用NFV技术与云计算实现数据流的类型标记及转发；并设计链路综合评价指标 及路由选择的计算方法。

[0110] 一种融合QoS及负载均衡需求的路由选择系统如图2所示，包括SDN控制 器、高性能云平台、SDN网络以及终端网络。

[0111] SDN控制器是网络的逻辑控制部分，包含拓扑发现模块、网络测量模块、 路由功能模块。其中，拓扑发现模块用于收集网络拓扑信息，如交换机间连接关 系、主机和子网的位置信息等；网络测量模块用于测量和收集网络链路信息；路 由功能模块用于综合评价链路及计算路由。

[0112] 高性能云平台将流量分类的结果提供给边界交换机，作为对数据流的ToS 进行标记的依据。

[0113] SDN网络由众多互联的交换机组成，构成网络的基础设施部分，负责数据 包转发，边缘交换机负责标记数据包QoS类型，用于区分服务质量需求。

[0114] 终端网络指网络用户及接入网设备组成的网络边缘。

[0115] 如图2数据流所示，一种融合QoS及负载均衡需求的路由选择系统的工作 原理包括以下几个关键环节，其工作流程图如图3所示：

[0116] (1)终端网络数据包发往SDN边界交换机；

[0117] (2)边界交换机对已知类型业务流ToS字段标注，未知业务流发送给高性 能云平台，待其分析识别后，边界交换机根据分析结果对业务流ToS字段标注；

[0118] (3)边界交换机把数据包转发给SDN控制器，以获得正确的转发策略；

[0119] (4)SDN控制器周期性测量网络，获得链路性能信息，并按预定的不同性 能指标权重组合，计算不同性能组合模式下的链路综合评价指标。根据边界交换 机发来的数据包源和目的地址，及ToS字段，计算出最优路径及生成流表；

[0120] (5)SDN控制器把流表分发给特定路径上的交换节点；

[0121] (6)交换机按照流表转发数据包，直到数据包到达目的端。

[0122] 图2中各个功能模块的技术细节如下：

[0123] 高性能云平台。其工作原理包含以下几个重要环节：

[0124] (1)接收由边界交换机发送的未知类型流，提取流的五元组信息；

[0125] (2)通过流量识别技术分析流，得到其业务类型；

[0126] (3)将分析结果发送至边界交换机，分析结果指用于标识流的五元组和流类 型的对应关系，流类型包括会话类、流类、交互类、背景类四种类型。流量识别 可采用现有的技术，包括：基于端口的分类技术、基于深度包检测(Deep Packet Inspection，简称DPI)的分类技术、以及基于深度流检测(Deep Flow Inspection，简称DFI)的分类技术。

[0127] 不同的流量分类技术有不同的使用场景以及不同的准确性和效率。基于端口 的分类技术因为动态端口和端口复用的广泛使用，导致其分类结果准确性不高；DPI技术可以较为准确的区分流量，但耗时多，同时对加密的流量不能有效地区 分；DFI技术基于流量行为和流量统计特征区分流量，处理速度较快，可以处理 加密的流量，但是分类结果笼统、

不够精细。

[0128] 为兼顾流量分类的准确性和效率，本发明采用的分类方法是利用如图4所示 的方法区分流量，首先使用DPI深度包检测的技术分析，记录能够识别包的类 型，无法识别的包再使用DFI深度流检测技术基于统计信息进行分类。

[0129] 边界交换机。通过NFV技术使边缘交换机具有标记ToS字段的功能，并克 服了以往因匹配业务流而导致Openflow流表项随时间及流量的增加急速增长的 缺陷。边界交换机的结构如图5所示，将网络功能网元以虚拟化的方式部属在通 用的硬件服务器上，以硬件资源池分配的方法提供网络功能网元所需的计算、存 储、网络等资源，提高网络部署的灵活性。上层建立“流标记存储”、“标记ToS”、“转发流量”等虚拟模块。

[0130] 边界交换机的工作原理包含以下几个重要环节：

[0131] (1)“流标记存储”模块接收终端网络发送的流，然后向高性能云平台发送未知 流量类型的数据包，接收由高性能云平台计算得到的分析结果，存储分析结果， 存储方式为记录以流的五元组为键值、以流类型为数值的关联数组，又称字典， 其对应关系示例如表2所示；

[0132] (2)“标记ToS”模块根据“流标记存储”模块存储的信息，判断数据包所属流类 型，分析其QoS需求，标记ToS字段；

[0133] (3)“转发流量”模块将数据包送往控制器用于寻路，控制器根据ToS字段计 算综合代价最低的路径，生成流表并将流表发送给交换机，交换机根据流表转发 数据包，其工作流程具体如图6所示。

[0134] 表2 五元组与流类型对应关系示例

[0135]

五元组	协议/应用类型	流类型
[59.34.148.244,172.18.217.201,UDP,6029,50851]	Skype	会话类
[192.168.31.131,192.168,31.1,UDP,50027,5060]	SIP	会话类
[216.58,205,66,192.168.1.7,UDP,443,54997]	Youtube	流类
[119.188.133.182,192.168.115.8,UDP,17788,22793]	PPStream	流类
[140.205.243.64,172.18.218.187,TCP,80,12943]	WEB	交互类
[24.105.29.24,172.18.216.87,TCP,3724,64668]	Game	交互类
[192.163.40.130,69.105.2.89,TCP,62927,25]	SMTP	背景类
[117.169.71.157,172.18.217.147,TCP,80,55393]	FTP_Data	背景类
……	…	

[0136] 边界交换机与高性能云平台通过一条专用链路实现相连，它们的信息交互过 程如下：

[0137] (1)边界交换机发送未知类型数据包至高性能云平台。边界交换机将该数 据包作为净荷，封装在以边界交换机端口MAC地址为源物理地址、目的物理地 址为广播地址(物理地址为FF-FF-FF-FF-FF-FF)的以太网数据帧中，格式如图 7所示，然后从对应端口送出；

[0138] (2)高性能云平台接收该数据帧后，提取其净荷部分中的数据包，使用流 五元组标识数据包所属业务流，并且按照图4所示的方法识别业务流类型，得到 五元组与流类型对应关系的分析结果，会话类、流类、交互类、背景类的类别编 码依次为二进制码字00、01、

10、11；

[0139] (3)分析结果由高性能云平台返回至边界交换机。高性能云平台将五元组 和流类型的数据以二进制表示，将此数据作为净荷，封装在以高性能云平台网卡 MAC地址为源物理地址、目的地址为广播地址的以太网数据帧中，格式如图7 所示，然后送回至边界交换机。为方便边界交换机解析出分类结果，五元组中每 个元素之间、五元组与流量类型之间需要用不同的特殊字符分隔，需要界定结束 符确认信息字段长度。在本例中五元组元素之间以空格字符分隔，五元组与流量 类型以分号字符分隔，以回车符表示结束，数据中字符通过Unicode统一码编码 为二进制数据。如图8所示，为方便表示，用十六进制数据表示编码的二进制数 据，0x20是五元组元素之间分隔符，0x59是五元组与流量类型的分隔符，0x10 为结束符。

[0140] SDN控制器。控制器功能模块如图9所示，包含拓扑发现模块、网络测量 模块、路由功能模块，拓扑发现模块用于收集网络拓扑信息，如交换机间连接关 系、主机和子网的位置信息等；网络测量模块用于测量和收集网络链路信息；路 由功能模块用于综合评价链路，计算路由。

[0141] 控制器周期性地与各个交换机交换信息，然后根据收集的信息为数据包生成 流表，指导交换机转发数据包。控制器的稳定性对整个网络的稳定性有决定性作 用。合理的设置轮询周期时长可以保证稳定的网络性能，轮询周期太长则网络测 量的结果不够准确，也不能及时发现因交换机宕机或接入网络导致拓扑变化，轮 询周期太短会造成控制器和交换机工作负荷过重，甚至可能影响到正常的网络通 信。

[0142] 因此，本发明提出以下方法以确定控制器的轮询周期：根据控制器工作负荷 选择增加或者缩短轮询周期，使用CPU使用率衡量控制器工作负荷，当CPU使 用率较高时，适当增加轮询周期时长，当CPU使用率较低时，可以为了得到更 准确的网络测量结果，可适当减少轮询周期时长。具体地，轮询周期时长的设置 方法如下：根据控制器计算能力设置最小轮询周期粒度τ，在不同CPU使用率 的情况下，阶梯式地设置轮询周期T，具体如下：

[0143] $$T=\begin{cases}\tau, & CPU\ utilization\ rate\in[0,20\%]\\ 2\tau, & CPU\ utilization\ rate\in[20\%,40\%]\\ 3\tau, & CPU\ utilization\ rate\in[40\%,60\%]\\ 5\tau, & CPU\ utilization\ rate\in[60\%,100\%]\end{cases}$$

[0144] 拓扑发现模块

[0145] 控制器为业务流选择路由，需要掌握网络全局拓扑结构信息，拓扑信息包括 交换机间的连接关系、主机以及子网的位置信息。

[0146] (1)本发明提出如图10所示链路拓扑发现过程，控制器周期性发送LLDP 数据包，得到链路两端交换机以及对应端口的信息，以此获得整个SDN网络的 交换机间的连接关系；

[0147] (2)收集主机及子网的位置信息：主机作为源端主动发起通信时，数据包 到达交换机处，交换机没有能匹配该数据包的流表时，产生包含主机IP地址、MAC地址的PacketIn消息交由控制器，此时控制器记录主机IP地址、MAC地 址、交换机编号、交换机端口号四元组的信息；若控制器没有目的端的位置信息， 由控制器构造ARP请求数据包，通过PacketOut消息广播到全网交换机上非交 换机互联的端口上，目的端收到此ARP请求后，发

送ARP应答数据包，同样此 应答包会被交换机发送至控制器，据此控制器可以掌握网络中主机的位置信息。Openflow流表包头域中不仅可以用IP地址作为匹配项，也可用子网号作为匹配 项，因此可以通过将同一个子网下的主机聚集为子网号作为流表匹配项来减少流表项的数目，减轻交换机的工作负担。

[0148] 网络测量模块

[0149] SDN由于集中控制的特点，可以通过控制器以主动测量的方式得到实时的 网络链路状态信息，目的是根据实时链路状态动态地进行流量调度。本发明提出 如下基于SDN的网络测量方法，测量的指标包括：链路时延、链路丢包率、链 路最大可用带宽、链路负载。

[0150] 链路时延测量方法如图11所示，具体步骤如下：

[0151] (1)控制器向交换机A下发一个Packetout报文。报文的数据段携带了一 个约定的协议报文(如链路层发现协议LLDP)，其报文的数据段携带了控制器 下发报文时的时间戳。Packetout报文的动作指示交换机将其泛洪或转发到某端 口。

[0152] (2)与交换机A相邻的交换机B收到了交换机A发送过来的数据包，无 流表项可匹配，从而将此数据包封装在Packetin消息并发送到控制器。控制器接 收到这个数据包之后，和当下时间相减，得到时间差T_1。其时间差约等于数据包 从控制器到交换机A、交换机A到交换机B、交换机B到控制器的时延总和。

[0153] (3)同理，控制器向交换机B发送一个同样的报文。然后控制器从交换机 A收到Packet_in报文，记录下时间差T_2。所以T_1+T_2等于控制器到交换机A的 RTT、控制器到交换机B的RTT和交换机A到交换机B的时延RTT三者之和。

[0154] (4)控制器向交换机A和交换机B分别发送带有时间戳的echo request。交换机收到之后即刻回复携带echo request时间戳的echo reply消息。所以控制 器可以通过echo reply的时间戳减去echo request携带的时间，从而得到对应交 换机和控制器之间的RTT。通过这种方法测得控制器到交换机A,B的RTT分别 为T_a,T_b。

[0155] (5)交换机A到交换机B的RTT为$T_1+T_2-T_a-T_b$。假设往返时间一样， 则交换机A到交换机B的链路时延为$(T_1+T_2-T_a-T_b)/2$。

[0156] 链路丢包率测量方法如下：

[0157] (1)控制器周期性向交换机发送ofp_port_status_request消息，收集交换机 端口统计信息，得到ofp_port_status_reply消息中表示端口发送/接收比特数的计 数器rx_packets和tx_packets，消息发送过程以及消息数据包格式如图12和图13 所示。

[0158] (2)交换机n与m之间链路丢包率：loss＝max(loss(m,n),loss(n,m))，其中 loss(m,n)表示交换机m到交换机n方向链路丢包率，其值应为：

[0159] $$loss_{(m,n)}=\frac{(tx\_packets_{\{t,m\}}-tx\_packets_{\{t-1,m\}})-(rx\_packets_{\{t,n\}}-rx\_packets_{\{t-1,n\}})}{tx\_packets_{\{t,m\}}-tx\_packets_{\{t-1,m\}}}$$

[0160] 下标{t,m}表示交换机m在第t轮次计数器的值，后文中下标代表相似含义。

[0161] 链路最大可用带宽以及负载测量方法如下：

[0162] (1)在控制器与交换机建立连接时，控制器发送ofp_feature_request消息询 问交换机支持功能，交换机通过ofp_port_features_reply消息，报告端口速率、双工模式、以及支持的链路介质，可认为端口汇报速率即为最大可用带宽。消息 发送过程以及消息数据包格式图14和图15所示。

[0163] (2)控制器周期性向交换机发送ofp_port_status_request消息,收集交换机 端口统计信息,得到ofp_port_status_reply消息中表示端口发送/接收比特数的计 数器rx_bytes和tx_bytes。交换机m与n之间链路已用带宽:

[0164] $Bandwidth = \max(Bandwidth_{(m,n)}, Bandwidth_{(n,m)})$

[0165] $$Bandwidth_{(m,n)} = \frac{(tx\_bytes_{\{t,m\}} + rx\_bytes_{\{t,m\}} - tx\_bytes_{\{t-1,m\}} - rx\_bytes_{\{t-1,m\}}) \times 8}{T}$$

[0166] 路由功能模块

[0167] 路由功能模块用于接收SDN交换机发送的PacketIn消息,为之计算路由、生成流表、并且将流表下发至对应交换机。路由功能模块建立在拓扑发现模块和 网络测量模块之上,路由功能模块需要拓扑发现模块提供的全局拓扑结构信息, 需要网络测量模块提供的实时网络链路参数以优化路由。其工作原理主要包含以 下几个环节:

[0168] (1)SDN交换机收到无法匹配流表的数据包时,将此数据包封装在PacketIn 消息中发送给控制器,控制器收到此消息后,解析出源地址和目的地址;

[0169] (2)根据拓扑发现模块提供的全局拓扑结构信息,构造存储拓扑信息的无 向图;

[0170] (3)根据网络测量模块提供的实时网络链路参数,计算链路综合评价指标, 更新无向图中边的代价值;

[0171] (4)使用路由算法为数据包选择最优路由。

[0172] 无向图包含顶点集和边集,顶点集中包含两类顶点,分别对应网络中的交换 机节点和子网节点;边集是顶点构成的无序二元组的集合,对应网络中链路,可 用邻接矩阵描述,如对于一个包含三个交换机节点S_1,S_2,S_3和两个子网节点 $Subnet_1$,$Subnet_2$的网络,其中(S_1,S_2),(S_2,S_3),($Subnet_1$,S_1),($Subnet_2$,S_3)节点对具有 连接关系,则邻接矩阵可以表示为:

[0173] $$\begin{vmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{vmatrix}$$

[0174] 矩阵中横轴和和纵轴均表示(S_1,S_2,S_3,$Subnet_1$,$Subnet_2$),邻接矩阵中1表示 有连接关系,0表示无连接关系。

[0175] 根据网络测量模块提供的实时网络链路参数,计算链路的归一化时延代价、丢包率代价、负载代价,然后按照加权求和的方法计算链路综合评价指标,更新 无向图中边的代价值。链路综合指标$Cost_i = \sum_{p\in\{delay,loss,load\}} w_p \cdot c_{i,p}$,权重值$w_p$刻 画客户端QoS需求和网络负载均衡性的重要程度,权重值 $w_p = \frac{a_p}{\sum_{p\in\{delay,loss,load\}} a_p}$,其中$a_p$表示不同性能指标的权值因子。控制器根据数据 包IP包头中ToS字段,可以区分不同数据业务流的QoS需求。结合客户端QoS 需求和网络负载均衡性,设置不同指标的权值因子,计算其权重值。前文中将对链路指标要求程度分为4个级别,为不同的级别设置不同的权重值,例如四个级 别的权值因子按要求程度从低到高依次为:0、2、4、8,同时根据网络负载均衡 性设置a_{load},网络负载越不均衡,则a_{load}值应越大,具体值需结合网络场景设计。将计算得到的链路综合评价指

标作为无向图边的代价值。

[0176] 从源端到目的端的最优路径是所有从源端到目的端的可达路径当中逐段链 路综合评价指标和最低的路径,即无向图中源顶点到目的顶点所有可达路径中逐 段链路代价和最小的路径。

[0177] 本发明提出以下的路由选择方法:

[0178] (1)预先设置阈值,计算网络链路平均利用率;

[0179] (2)当网络链路平均利用率低于阈值,可认为大多数网络链路较为空闲,控制器以文献“Dijkstra(Johnson D B.A note on Dijkstra's shortest path algorithm[J].Journal of the ACM(JACM),1973,20(3):385-388.)”或“Floyd(Hougardy S.The Floyd-Warshall algorithm on graphs with negative cycles[J].Information Processing Letters,2010,110(8-9):279-281.)”提出的最短路径算法为业务流计算最优路由;

[0180] (3)当网络链路平均利用率高于阈值,可认为大多数网络链路较为繁忙,控制器以K最短路径Yen算法(en J Y.Finding the k shortest loopless paths in a network[J].management Science,1971,17(11):712-716.)得到前K条最优路由,而后比较K条路由,选出路径负载均衡参数l_r最小的路径r作为最优路由。

[0181] (4)控制器根据计算出的最优路由,为沿路经过的交换机生成并下发流表,由入端口、源IP地址、目的IP地址、IP ToS位构成的四元组作为流表匹配项。

入端口	源MAC地址	目的MAC地址	以太网类型	VLAN ID	VLAN优先级	源IP地址	目的IP地址	IP 协议	IP ToS位	TCP/UDP源端口	TCP/UDP目的端口
Ingress Port	Ether Source	Ether Des	Ether Type	VLAN ID	VLAN Priority	IP Src	IP Des	IP Proto	IP ToS bits	TCP/UDP Src Port	TCP/UDP Dst Port

type	code
not required	00
low	01
medium	10
high	11

业务要求类型编码

第三组		第二组		第一组	
5	4	3	2	1	0
未定义		丢包率		时延	

图1

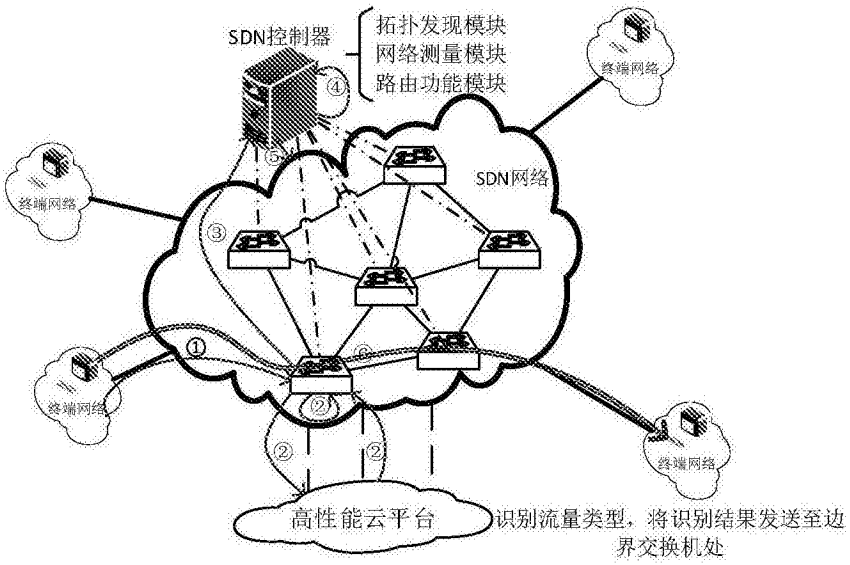


图2

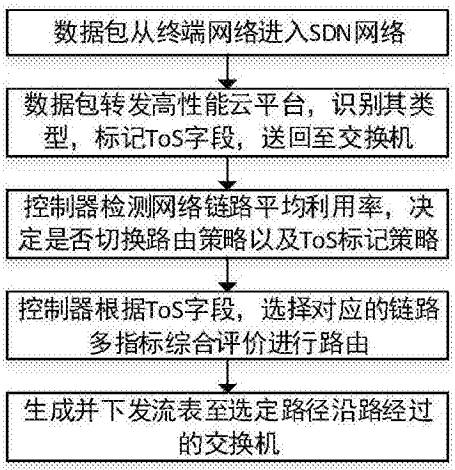


图3

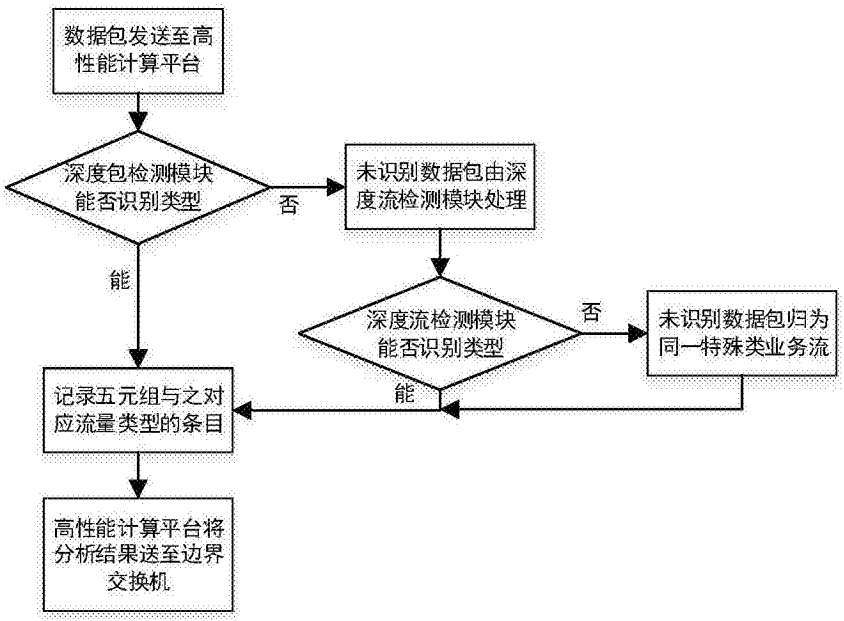

数据包发送至高性能计算平台
深度包检测模块能否识别类型
否
未识别数据包由深度流检测模块处理
能
深度流检测模块能否识别类型
否
未识别数据包归为同一特殊类业务流
能
记录五元组与之对应流量类型的条目
高性能计算平台将分析结果送至边界交换机

图4

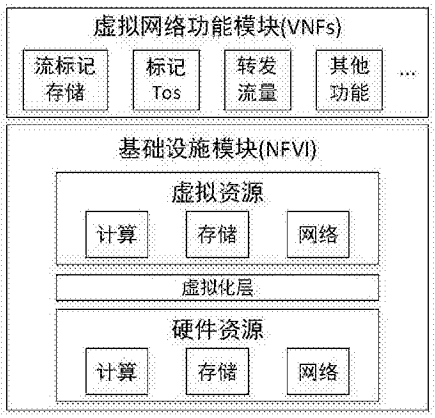

虚拟网络功能模块(VNFs)
流标记存储
标记Tos
转发流量
其他功能
...
基础设施模块(NFVI)
虚拟资源
计算
存储
网络
虚拟化层
硬件资源
计算
存储
网络

图5

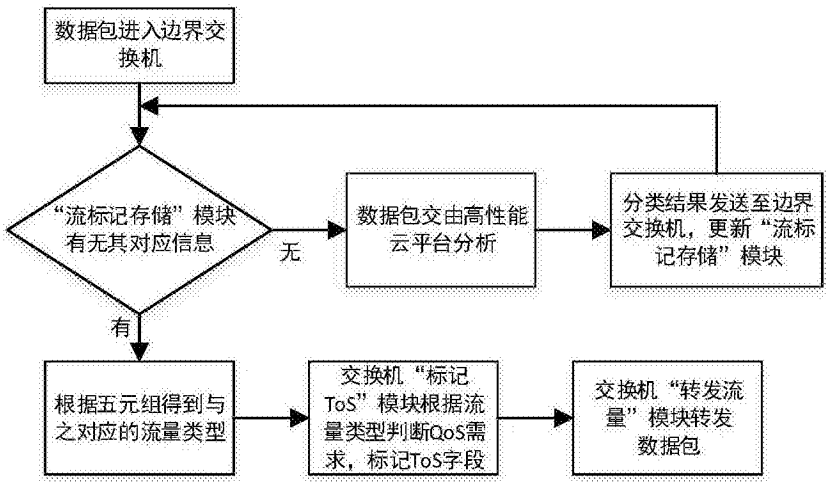

数据包进入边界交换机
“流标记存储”模块有无其对应信息
无
数据包交由高性能云平台分析
分类结果发送至边界交换机，更新“流标记存储”模块
有
根据五元组得到与之对应的流量类型
交换机“标记ToS”模块根据流量类型判断QoS需求，标记ToS字段
交换机“转发流量”模块转发数据包

图6


目的地址
源地址
类型
净荷
CRC

图7

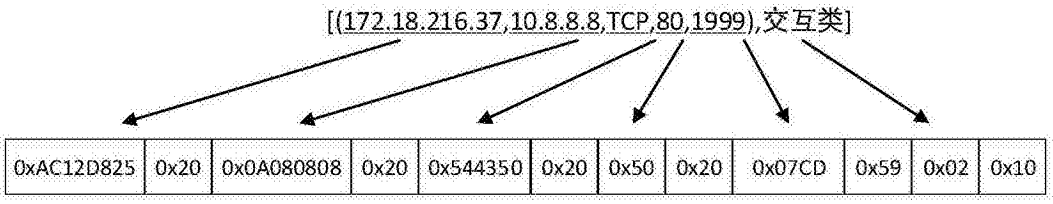

[(172.18.216.37,10.8.8.8,TCP,80,1999),交互类]
0xAC12D825
0x20
0x0A080808
0x20
0x544350
0x20
0x50
0x20
0x07CD
0x59
0x02
0x10

图8

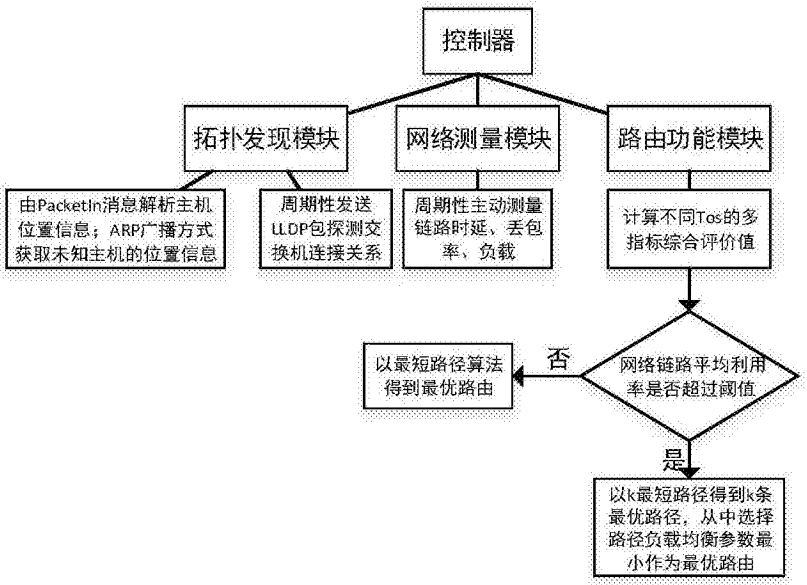

控制器
拓扑发现模块
网络测量模块
路由功能模块
由PacketIn消息解析主机位置信息；ARP广播方式获取未知主机的位置信息
周期性发送LLDP包探测交换机连接关系
周期性主动测量链路时延、丢包率、负载
计算不同Tos的多指标综合评价值
网络链路平均利用率是否超过阈值
否
以最短路径算法得到最优路由
是
以k最短路径得到k条最优路径，从中选择路径负载均衡参数最小作为最优路由

图9

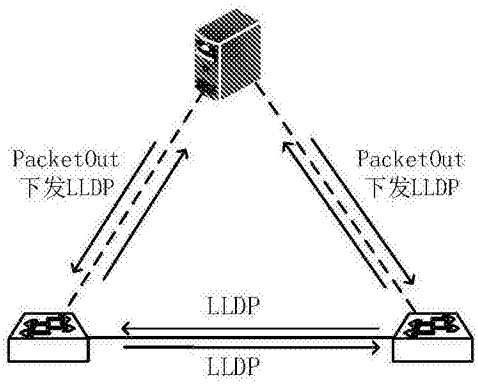

PacketOut
下发LLDP
PacketOut
下发LLDP
LLDP
LLDP

图10

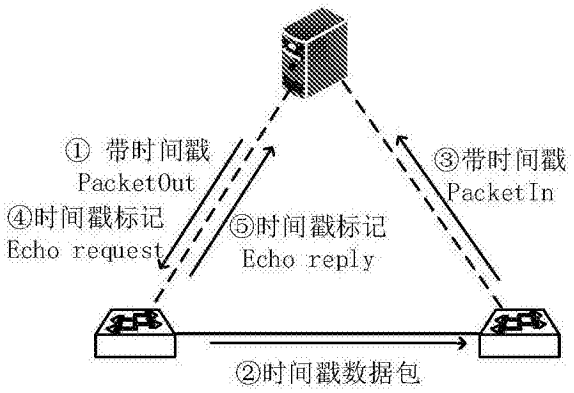

① 带时间戳
PacketOut
④时间戳标记
Echo request
⑤时间戳标记
Echo reply
③带时间戳
PacketIn
②时间戳数据包

图11

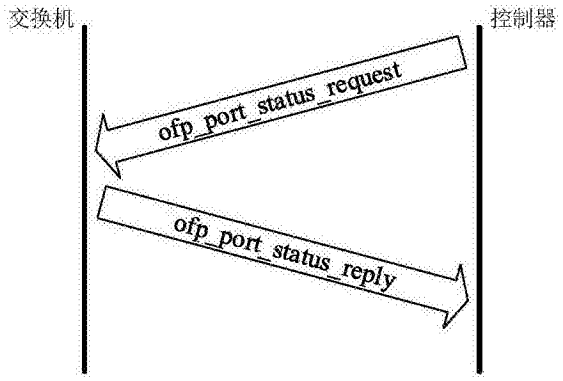


图12

ofp_port_status_request:

version	type	length	xid	stats_type	flag	port_no	pad	body
1B	1B	2B	4B	2B	2B	4B	4B	request

ofp_port_status_reply:

version	type	length	xid	stats_type	flag	port_no	pad	body
1B	1B	2B	4B	2B	2B	4B	4B	reply

port_num	pad	rx_packets	tx_packets	rx_bytes	tx_bytes	...
4B	4B	8B	8B	8B	8B	...
⋮						
port_num	pad	rx_packets	tx_packets	rx_bytes	tx_bytes	...
4B	4B	8B	8B	8B	8B	...

图13

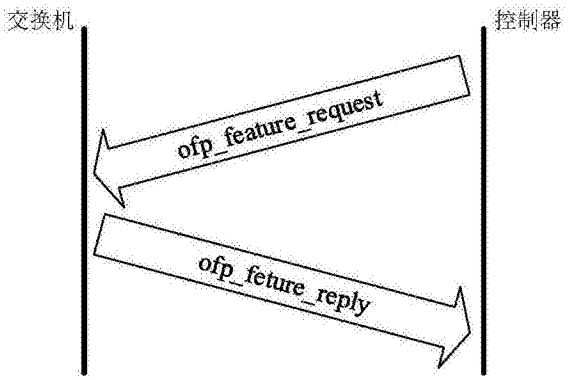


图14

ofp_feature_request:

version	type	length	xid
1B	1B	2B	4B

ofp_feature_request:

version	type	length	xid	datapath_id	n_buffers	n_tables	capabilities	reserved	body
1B	1B	2B	4B	8B	4B	1B	4B	4B	of_port_desc_list

port_num	pad	hw_addr	name	config	state	curr	advertised	supported	peer	curr_speed	max_speed
4B	4B	6B	string	4B	4B	4B	4B	4B	4B	4B	4B
⋮											
port_num	pad	hw_addr	name	config	state	curr	advertised	supported	peer	curr_speed	max_speed
4B	4B	6B	string	4B	4B	4B	4B	4B	4B	4B	4B

图15