



基于 SDN 架构的数据中心网络路由算法需求分析

樊自甫, 伍春玲, 王金红

(重庆邮电大学电子商务与现代物流重点实验室 重庆 400065)

摘要:随着 SDN 技术的发展,采用基于 SDN 架构的数据中心网络将越来越广泛。围绕 SDN 技术在数据中心网络中的应用问题,分析了现代数据中心网络及 SDN 的技术特征,设计了 SDN 在现代数据中心网络中的应用架构,并以路由为视角分析总结了基于 SDN 的数据中心网络路由性能需求,提出了相应的解决方案,以期对我国未来数据中心网络建设路由设计提供必要的参考。

关键词:软件定义网络;路由算法;数据中心网络;OpenFlow

doi: 10.11959/j.issn.1000-0801.2015038

Requirements Analysis of Data Center Network Routing Algorithm Based on SDN Architecture

Fan Zifu, Wu Chunling, Wang Jinhong

(Key Laboratory of Electronic Commerce and Modern Logistics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: With the development of SDN technology, SDN-based data center network will be more extensive. The applications of SDN technology in data center network were focused on; the technical characteristics of modern data center network and SDN technology were analyzed firstly, and a modern data center network architecture based on SDN was designed. Then the requirements of the data center network routing performance based on the SDN were analyzed and summarized and the solutions were proposed in order to provide the reference for the future construction and routing design of the data center network in China.

Key words: software defined networking, routing algorithm, data center network, OpenFlow

1 引言

云计算、大数据等应用促使互联网上的应用、计算和存储资源不断地向数据中心迁移^[1],使数据中心成为信息交互的中心,而数据中心网络(data center network, DCN)是一个连接了数万级、十万级甚至百万级的大规模服务器群,作为分布式存储和计算的桥梁,它的性能直接影响到云计算、大数据等业务的发展。为了满足数据中心网络性能

的新需求,对新型数据中心网络体系架构的研究成为了当前研究的热点。

数据中心网络的新需求促进了软件定义网络(software defined networking, SDN)技术的迅速发展,在 2012 年 Gartner 发布的未来 5 年全球 IT 十大关键技术中,SDN 位列第二。SDN 的核心在于数据转发与路由控制的分离,相较于传统网络,SDN 通过集中控制器取代了原来的路由协议自协商方式,极大地提升了网络的管控效率和开放程

度。在现代具有大规模服务器和带宽密集型数据中心网络中,SDN 技术可以提供大规模、高拓展性、灵活资源调度和高带宽等网络需求,表现出了天然适用于新型数据中心网络的特性。

在每个网络中,路由至关重要,对网络性能具有重要的影响,并随着新需求和新技术的产生而不断发展。在采用了 SDN 技术的数据中心网络中,新的体系架构和更高的网络服务质量需求给路由算法性能带来许多新的挑战,需要结合现代数据中心的网络特征和 SDN 技术特点加以改进,以适应多路径、大规模和节能等需求。本文以此为研究出发点,通过对现代数据中心网络和 SDN 技术特征的分析,提出了 SDN 在数据中心网络中的应用架构,并对面向数据中心网络的 SDN 路由性能需求进行了深入剖析,以便对未来在构建基于 SDN 的数据中心网络时如何设计路由提供借鉴。

2 现代数据中心网络特征

数据中心网络是数据中心内部通过高速链路和交换机连接大量服务器的网络^[2]。在传统数据中心网络中,普遍采用三层树型结构,即接入层交换机、汇聚层交换机和核心层交换机的三层交换机互联。在很长一段时间内,传统网络架构支撑了各种类型的数据中心网络,但其层次结构使得网络规模及性能受限于根节点,在很大程度上影响了网络的拓展性。随着虚拟化、云计算及相关新型业务的迅速发展,互联网上的应用、计算和存储资源不断向数据中心迁移,促使数据中心规模迅速膨胀和计算模式发生深刻变化^[3],仅仅使用传统的数据中心网络技术已经越来越无法适应新业务发展的需要,从而促进了对新型数据中心网

络体系架构的探索,使其成为近来学术界和工业界关注的热点。面向云计算等新型计算模式的数据中心呈现出许多新的特点见表 1。

3 SDN 及其在数据中心网络中的应用

3.1 SDN 网络架构

SDN 是一种新型的基于软件的网络架构及技术,其核心在于控制平面与数据平面相分离、灵活的编程接口开放^[6]。SDN 的典型架构共分 3 层,如图 1 所示,自上而下划分为:应用层、控制层和基础设施层。其中,基础设施层仅负责单纯的数据高速转发和网络状态收集,不具备任何智能性,只需关注硬件性能;中间控制层集中控制所有网元,根据应用需求计算数据转发路径,并将计算结果下发给基础设施层;控制器向上对应用层提供可编程接口,通过软件级的更新,新的协议和应用在控制器上可快速部署。SDN 大大降低了网络的复杂度,使其具有较大的灵活性,可以有效解决当前网络所面临的许多问题。

OpenFlow 是当前 SDN 控制器与转发设备的主流通信接口协议,包含了 OpenFlow controller (控制器)和 OpenFlow switch(交换机)两个网络元素^[9]。如图 2 所示,一个 OpenFlow 交换机内包含一至多张流表(flow table,即转发表),流表由流表项(flow entry)组成,每一流表项对应网络中传输的一条流,流表项决定了属于该条流的数据分组如何转发或处理。

在一条流的第一个数据分组刚进入 SDN 时,交换机将该数据分组或数据分组头上传到控制器,控制器通过路由计算及策略管理将该数据分组所属的流的处理方法下发给交换机存储在流表内。数据分组在交换机内的处理过

表 1 新型数据中心的特点

需求	说明
大规模	云计算数据中心具有通常可达上万乃至上百万台量级的大规模服务器 ^[4] ,且规模仍在迅速增长中,网络需要具有良好的可拓展性
低成本	在考虑成本的情况下,现代数据中心网络的构建应基于普通商业交换机
低配置开销	在大量服务器和交换机存在的前提下,数据中心网络不应该引入过多的交换机配置开销 ^[5]
自动化	多个租户共享使用数据中心网络资源且运行不同的企业应用,资源应以动态、弹性、按需的方式分配和共享,避免大量使用人工配置
高带宽	MapReduce 应用、虚拟机迁移以及其他带宽密集型应用触发的数据中心内部流量显著增加,网络带宽容易变成数据中心发展瓶颈
高利用率	数据中心流量分布不均,呈现局部、高动态和不可测的特性,从而使得传统流量工程方法无法有效工作 ^[6] ,网络内部存在严重的负载不均现象,大量交换机和链路闲置得不到有效利用,需提高网络利用率以增大网络吞吐量
可靠性	大量使用廉价的交换设备的同时,需要避免或减少服务组件的意外灾难或故障对业务的冲击
低能耗	数据中心属于能耗密集型产业,随着日益扩大的网络规模急需减少电源和制冷开销,削减运营成本 ^[7]

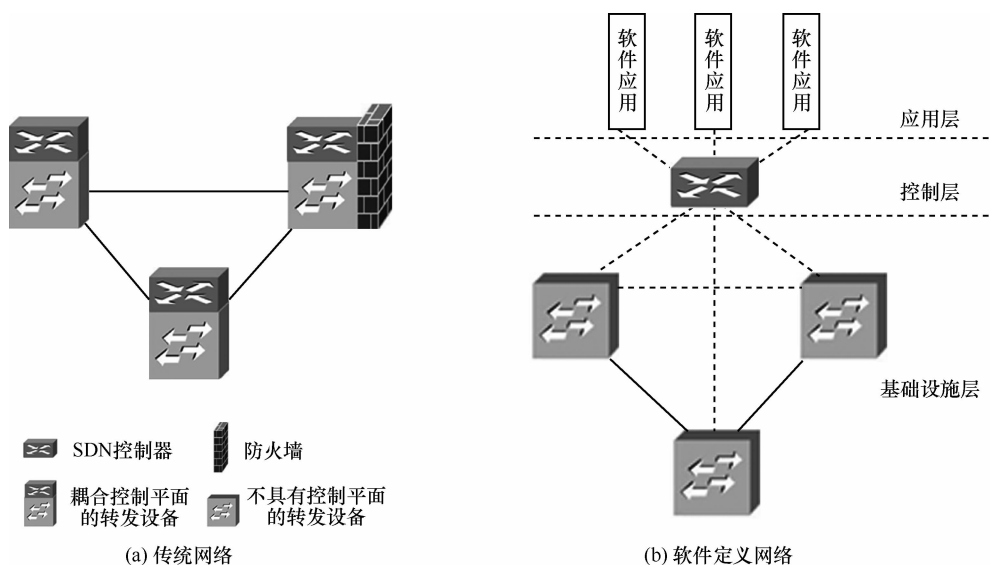


图1 传统网络架构和SDN网络架构

程为:数据分组到达交换机,首先进行分组头内容提取,然后与流表项匹配域进行匹配,若匹配到合适的流表项,则采用与该流表项相关联的指令进行处理;若没有匹配到相应的流表项,数据分组按默认流表项(table miss)进行处理,丢弃、转发到下一级流表或上传到控制器处理。

3.2 SDN在数据中心网络的应用

由于SDN集中控制的良好拓展性和灵活性天然符合数据中心网络发展的需求,在数据中心网络采用SDN技术的研究得到了广泛关注,而迫切需要改革的数据中心网络也成为了SDN的首要应用场景,推动SDN的发展。2010年

1月,Google开始采用SDN和OpenFlow;2012年初,Google全部数据中心骨干连接都已采用这种架构,网络利用率提升到95%,其他企业如Facebook、德国电信和NTT等企业和电信运营商也已经开始将SDN部署在数据中心。图2所示是采用SDN技术的数据中心网络可能的应用框架,网络主要由基础设施层、局部控制器和全局控制器构成:基础设施层由局部控制器进行集中管理,可适应任意拓扑结构,这里采用数据中心网络中较为流行的fat-tree^[10]结构;局部控制器除了管理本域内的交换机外,还需要将本域网络状态信息上传到全局控制器;全局控制器通过局部

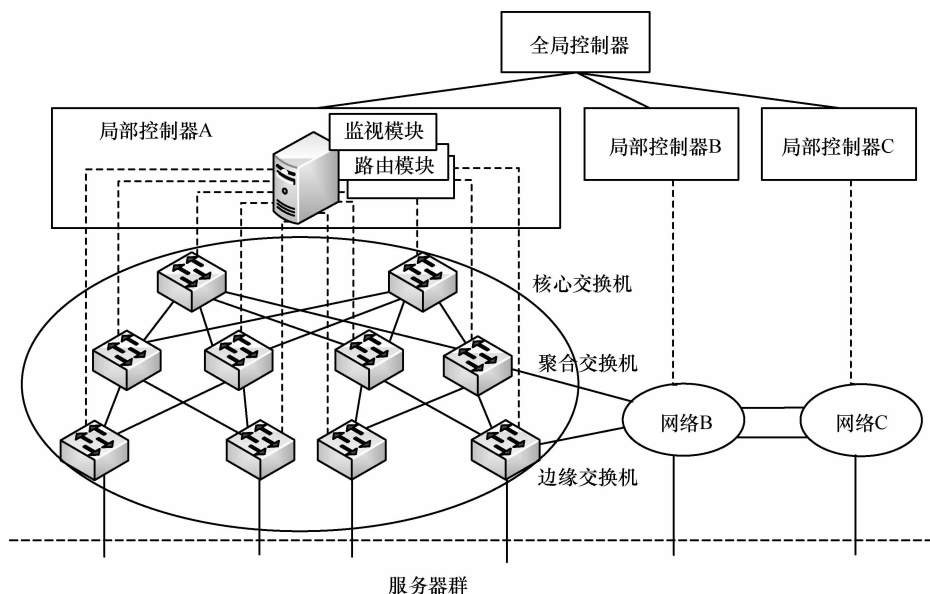


图2 基于SDN技术的数据中心网络应用框架

控制器对全网进行状态收集和管理,提供全网的网络状态信息视图。框架中的全局控制器不是必要的,局部控制器也可通过东西接口直接与其他控制域连接,进行信息交换。

基于SDN技术的数据中心网络有许多的优势。首先,SDN可管可控,网络运营者可及时获取全网状态信息,包括拓扑、链路、网络拥塞、服务质量和网络故障等,也可集中地进行设备配置、策略发放和新应用服务投入运行,从而获得大规模数据中心网络的高效运维和管理。其次,SDN技术可以有效提高网络资源利用率,在保障网络服务质量前提下,通过高效实施流量工程提高数据中心内链路带宽利用率和吞吐量,节约建网成本,同时满足快速频繁的资源调度和实时配置的需求。再者,SDN能够实现虚拟机管理自动化,可以有效管理云计算使用的海量虚拟机,同时通过与VM服务器的配合可以很好地实现自动化虚拟机部署及快速迁移。此外,SDN还有适用于组网灵活和多路径转发的环境,仅涉及软件更新的快速应用部署和网络升级等特性,这些特性都能很好地适应数据中心网络的建设开发。

4 基于SDN架构的数据中心网络路由性能需求

网络路由一直是网络的关键问题,路由算法层出不穷,目的都是为了寻找最佳路径(或者满足要求的路径)来传递信息,提高服务质量,同时提高网络整体资源利用率。在传统网络中,各个节点通过局部网络状态进行分布式路由计算,难以发挥算法的最佳性能。然而,具有全网视图的SDN可以在很大程度上提高路由算法性能,增加网络吞吐量,主要原因如下:

- 在集中控制的SDN中控制平面具有全网拓扑,可以优化路由算法性能而不需要考虑算法的收敛性;
- 控制平面维护了全局网络信息(包括流量信息、链路信息、路由限制等),根据网络能力,可按照不同流量的服务需求计算路由,有效地管理和调度流量;
- 转发平面实时上报网络信息到控制平面进行流量及网络状态分析,及时调整路由并统一下发流表更改路径,灵活、动态地调度流量,可以实现网络性能优化、提高网络链路利用率及负载均衡等目的。

结合现代数据中心网络的流量特征和业务模式,对基于SDN架构的数据中心网络,在面对云计算和大数据等

新型业务时的路由算法性能需求进行了全面分析,提出在未来数据中心网络建设与部署中应重点关注以下6个方面的路由性能。

(1) 负载均衡性能

数据中心网络属于高带宽网络,通过单纯增加网络链路来提高带宽,不仅增加了网络成本也使网络结构更加复杂,而最短路径路由算法总会选择代价最小的路径使得流量集中于相同链路,并没有有效利用所增加的带宽,使得实际上提高的吞吐量并不理想。由于使用SDN技术的数据中心网络具有全网视图,可获取链路利用率及剩余容量等网络信息,有效计算基于负载均衡的路由,对全网流量进行分配,从而提高链路利用率和网络吞吐量,满足高带宽需求。

基于负载均衡的路由算法主要利用了网络多路径特性,即源一目的对之间可能存在多条可用路径的特性,将网络流量分配到满足服务需求的空闲路径上,避免节点总是选择最短路径,提高链路利用率。实时的网络状态是确保路由有效性的关键,在传统网络中,一个节点需要保存所有网络节点信息,可能会导致路由的扩展性较差,而在SDN中全网视图可以很好地解决路由的可扩展性问题。

在路径的选择方式上主要有以下两种:一是在多条可用路径中,根据网络状态信息,在网络性能约束条件下,选择相对空闲的一条路径进行传输或者选择两条作为主备路径,备用路径仅在主路径失效后启用;二是选取源一目的对之间的多条路径按比例共同承载数据流。其中,第一种方式可以减少路径查找和安装的时延,路由更具稳定性,但因仅存一条路径,因此在网络状态发生变化时不能快速反应,虽然可以通过使用主备路径的方法来解决,但是备用路径存在可能已经过时的问题,同时也会增加交换机的维护路径开销;第二种方式由于在目的端接收的数据分组是乱序的,一定程度上影响了数据传输效率,但是由于多条路径的存在可以更有效地使用带宽,提高了端到端的可靠性,在网络状态发生改变的情况下可以有效地避免拥塞。

在多路径并行传输方式中,需要考虑选择多少条路径、如何选择这些路径、流量在每条路径上的分配问题,相对较为复杂,目前研究和关注较少,主要还是集中在单路径传输路由算法上。在单路径传输路由算法中,重点要考虑路径的查找和选择,目前主要有两种方式:一是基于负载均衡选择最优路径,例如,首先找出多条具有最小跳数



或满足服务需求性能的候选路径,通过SDN监测并计算这些路径上的链路利用率之和作为开销,选择开销最小的路径,如果有相同的最小开销路径,则优先选择剩余可用容量最大的路径^[1];另一种则并不总是选择最优路径,可以通过概率方式进行选择,如基于随机效用理论模型——MNL(multinomial logit,多项罗吉特)模型^[2]的轻量级路由算法就是出于大规模数据中心存在频繁的网络状态更新并不适合使用计算量太大的路由算法的考虑,通过概率进行路径选择的路由算法。个体从可选路径集 $N=\{1,2,\dots,n\}$ 中选择一个选项时,会选择对其效用最大的选项,MNL模型根据该理论,给出了路径 i 被选中的概率,如式(1)所示:

$$p_i = \frac{\exp[\gamma V_i]}{\sum_{j \in N} \exp[\gamma V_j]} \quad (1)$$

V_j 是选择 j 时的效用, γ 若用路径开销 C 作为选路效用,则 $V_j=C$,式(1)可化为:

$$p_r^{od} = \frac{\exp[-\gamma C_r^{od}]}{\sum_{j \in \Phi^{od}} \exp[-\gamma C_j^{od}]} \quad (2)$$

其中, o 为源节点, d 为目的节点, Φ^{od} 为源节点 o 到目的节点 d 的可选路径集, r 为其中一条路径, C_r^{od} 为路径 r 的开销, p_r^{od} 为选择路径 r 的概率^[13]。计算出各个路径的概率后,数据分组到达节点时即可根据 p 随机选择路径,同一条流的分组选择相同的路径,以减小TCP控制对流量吞吐量的影响,这种通过概率选取路径的方式可避免节点总是选择最短路径,提高链路利用率,同时控制器可周期更新开销重新计算 p ,进一步最大化利用网络。

(2)动态路由性能

由于数据中心网络中的流量具有高动态性,基于网络流初始化阶段的状态来进行的路由计算并不能保障在数据传输过程中动态网络的负载均衡。为此,需要对基于SDN架构的数据中心网络进行周期性实时监测以更新网络状态,将负载过重的链路上的流量进行路径转换或重新路由,从而使网络达到动态的负载均衡。检测负载是否过重的时间点可以使用周期性触发、门限触发和二者协同触发的方式。在判断负载是否过重时,既可以根据链路负载超过所占的链路容量比重(如75%)决定是否进行重新路由^[14],也可以通过对全网设置一个负载均衡参数进行判断,例如,负载均衡参数 $\delta(t)$ 如式(3)所示:

$$\delta(t) = \frac{\sum_{i,j} [\overline{load_{i,j}(t)} - load_{i,j}(t)]^2}{N} \quad (3)$$

式中, $load_{i,j}(t)$ 为在时刻 t 时节点 i 和 j 之间已占用的带宽, N 为网络中所有链路,当 $\delta(t) > \delta^*$ 时(δ^* 为针对该网预先测量的阈值),则启动动态负载路由算法,对网络中影响负载均衡的最大流进行调度,实现动态路由^[15]。

动态路由算法需要考虑的另一个问题是:链路过载发生后,应该对哪一条流进行重路由以使网络快速地重新恢复负载均衡状态,因此需要对流进行区分。在数据中心网络中混合了不同特征的流量,对大多数数据中心网络来说,网络流量表现出明显的大象流和老鼠流的特征^[3],即绝大部分的流均较小,而极少数的大流却占据了绝大部分的数据流量。因此,在进行流量调度时,路由协议可根据流量大小动态区分出吞吐量敏感的大流量和时延敏感的小流量。考虑到小流量因为持续时间相对较短,改变其路径极有可能会增加时延和开销,应通过迁移大流量实施流量调度,而在一般的路由协议如ECMP(等价多路径)无法对此进行区分处理,但由于SDN是基于流的转发,可以很好地对流进行统计区分,为数据流的判别奠定了基础。大象流的识别可以通过SDN数据统计机制来完成,每一条流都由其进入SDN时经过的第一个交换机进行流量监测,控制器每间隔一定时间(如5s)就向交换机进行数据采集,流的大小分类可采取式(4)进行计算:

$$\psi_t = (b_t - b_{t-p})/p \quad (4)$$

式中, ψ_t 是根据时间对流量大小进行的估计, b_t 是交换机在时刻 t 接收到的该流的所有字节数, p 是统计的时间间隔, b_{t-p} 即交换机在时刻 $t-p$ 接收到的该流字节数。通过对 ψ_t 进行界定即可区分大小流^[14]。在启用动态路由时,针对不同大小的流制订相应调度策略,如流量调度主要针对大流,可减少对小流的不必要干扰,有效实施负载均衡,提高动态路由性能。

(3)多播路由性能

数据中心网络中存在大量集群通信应用,如分布式文件存储系统、MapReduce^[16]集群计算和系统升级。集群通信模式是数据中心主要的通信模式,包括一对多、多对一和多对多^[17]。大量的集群数据影响了数据中心网络的性能,而传统的点到点通信方式浪费了大量的网络带宽且效率较低,多播技术因为可以节约网络带宽和提高应用的吞吐量,成为解决组通信的主流方法。IP多播是传统网络使用

最多的多播协议,但它不仅在可靠性、安全性和可拓展性上存在许多问题,且因不能利用网络多路径特性导致链路利用率低,分布式计算限制了最佳多播路由树的形成^[18]; SDN 具有可控可认证等优势,可以很好地解决多播的安全性和多播成员的出入组管理问题,同时全网拓扑为高效多播路由提供了有效信息。

在多播路由中,找出多播路由树是关键,初始状态的路由树应该是最佳路由树,但分布式计算显然较难获取最佳路由树,同时,多播因为存在大量的成员频繁出入多播组的情况^[19],所以多播路由树一直处于动态,很少能够达到稳定,也难以维持最佳的路由树,因此,如何优化动态的多播路由树也是多播路由计算的核心问题。传统 IP 多播主要使用的是 PIM-SM^[20]协议,其核心是构造并维护一颗单向共享树,如图 3(a)所示,共享树需要选择 PIM 中某一路由器作为公共根节点,即汇聚节点 (rendezvous point, RP),并根据 RP 形成多播共享树,该路由树并不是最佳路由树。多播数据由源节点通过 RP,沿着共享树向接收者转发,图 3(a)中带星号的为 RP,每当有新成员加入路由树时,都要连接到一个特殊节点(RP 或源节点),而不是连接到路由树中与它距离最短的那个节点。因此,PIM 多播效率并不高,且当通信量增大时,使用共享树将会出现流量集中到 RP 附近的瓶颈。

由于 SDN 具有全网拓扑可以提供更多的可用路径和所需的链路状态,有效地提高了多播路由的效率。在多播路由树建立初始,根据带宽、时延、抖动等在 SDN 全网拓

扑上可构建最佳路由树。一种基于全网拓扑构建多播路由树的方法是通过 Steiner^[21]树问题来解决。Steiner 树问题是最短网络的一种,最小生成树是在给定的点集和边中寻求最短网络使所有点连通,而最小 Steiner 树允许在给定点外增加额外的点,使生成的最短网络开销最小,因此,使用 Steiner 树作为多播树时,在给定源节点和多播组员时可以在整个网络构造多播路由树,使得网络开销最小。在多播数据传送过程中,当有新节点加入时,同样可通过全网拓扑寻找该节点到原组播路由树的最短路径,一种简单的方法是通过迭代判断其相邻节点是否为多播树成员来进行接入^[22],如果是则接入相邻节点,否则继续判断该相邻节点的相邻节点是否为多播树成员,直到找到与新节点最近的多播树成员,如图 3(b)所示,R3 直接连接到与它距离最短的点,而不是直接连接源节点或公共根节点。通过优化多播路由树的动态变化过程,为多播路由提供更小开销,节约网络资源,也避免了性能瓶颈问题。

(4)多控制域间路由性能

数据中心网络规模巨大,并且存在多个数据中心互联的情况,若仅依赖于一个控制器来对整个网络进行管理,由于控制器处理能力和控制通道带宽的有限性,容易形成网络性能瓶颈。使用多控制器进行分域管理是解决 SDN 拓展性问题的有效方法^[23],即由一个独立控制器所控制的交换机组成一个控制域,控制域间无交叉,多个控制域协同管理整个网络。虽然多控制域可以实现大规模网络的有

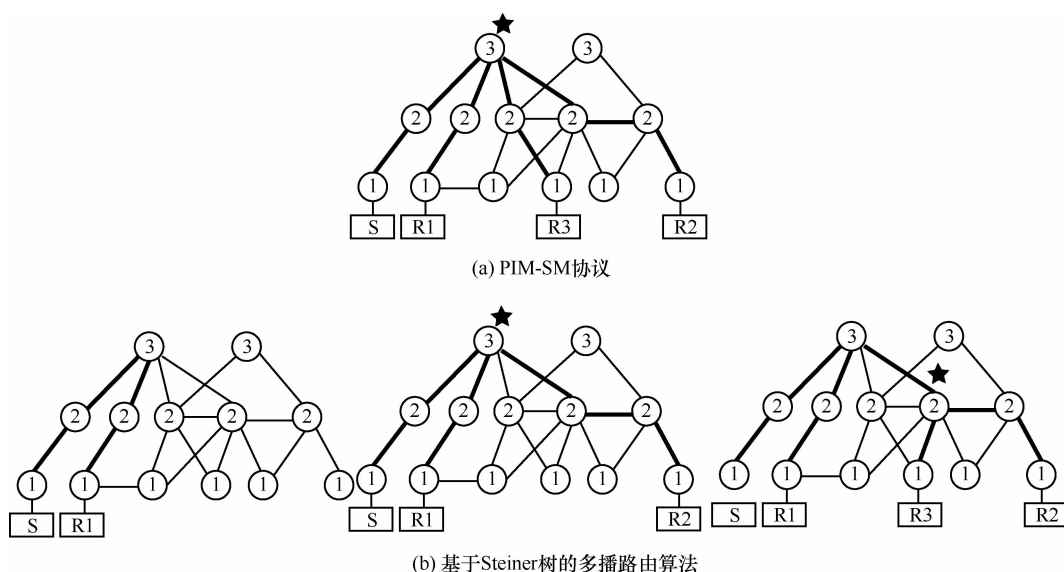


图3 多播路由树



效管理,但由于多个控制域的存在也增加了网络复杂度,而控制器只能计算域内的路由,也会大大降低路由效率,例如,域外数据分组到达后,控制器无法给出路由,只能通过泛洪的方式进行数据分组的传输,消耗大量的网络资源,因此要求控制域间能够共享各个域的网络状态信息,从而对域间路由算法提出了更高要求。多控制器的信息共享一般有两种方案:一种是控制器间直接通过专用接口(即南北向接口)进行连接和信息共享,这会产生大量交互信息并影响信息的实时性与准确性,所以可以采用另一种方案,使用控制器协同管理服务器(全局控制器)^[24]对每个控制器进行信息采集,形成全网网络状态视图,并为各个控制器提供全网路由信息。

通常,多控制域间路由算法可采用分级路由和无级路由两种路由方式,其中,在无级路由中,控制器对全网拓扑直接进行路由计算得到源一目的地的最短路径,在这个过程中,各个控制器需要提供大量的详细网络信息进行共享,在大型数据网络中心中,增加了网络控制信息开销和全局控制器的处理负担;而分级路由则区分控制器域内路由和域外路由,控制器间仅需交换域内路由计算结果,不仅减少大量的信息交互,而且降低了域间路由计算复杂度。

因此,为了提升多控制域间管理性能,建议采用以下分级路由方案^[25](如图4所示):首先计算出每个控制域内边界交换机(图中黑色圆点所示)间的最短路径和其开销,并将该路径转换为两个边界交换机直接相连的虚拟链路,如图4(b)所示,此时,整个网络拓扑得到简化,控制器接收到域外数据分组后就可以根据简化后的全网视图进行域间路由计算,实现多控制域环境下的数据中心网络路由。

(5) 节能性能

现代大规模数据中心中拥有数以百万计的服务器,设

备密度大,能耗提高,电力及制冷等运营成本也随之迅速增长^[26]。与此同时,为了适应服务需求,需提供高性能和高容错性的路由,新型数据中心大量采用 fat-tree、BCube、FiConn 等“富连结”的网络拓扑,引入了大量交换机参与路由,使得每对服务器间存在多条路径。这样虽然能提高网络性能,但也会造成大量的能耗。

对于数据中心的能耗问题,一般可以从设备级和网络级来考虑^[27],设备级节能方式仅关注单个硬件设备的能耗设计问题,网络级节能方式则需要对整个网络进行规划,需要解决的问题包括拓扑选取、网络路由和流量调度等。在网络路由方面,节能路由的核心思想是:在满足网络性能需求的条件下,根据网络状态的变化配置节能路由路径,将业务流量集中在主要网络链路上,并将空闲的网络设备关闭或置于休眠模式,减少不必要的交换机及链路的使用,从而达到节能的目的,然而,关闭部分网络设备后必然会对网络的服务质量造成一定影响,需要在网络性能和能耗两者之间做一定的权衡。节能路由机制涉及了网络状态的收集、流量调配和单点设备功耗模式切换等问题,这些在 SDN 集中控制的架构下都可以得到非常好的支持。

在节能路由的计算中,可以根据网络状态变化进行路由选择,也可根据网络拓扑直接计算节能路径。前者需要根据网络流量的实时变化动态调整路由,主要思想为首先为每条流计算初始路径,在路径集合中按照一定的策略对满足条件的节点和链路迭代地删除^[28],同时将流量调度到其他路径,调度策略可根据流量大小、链路利用率、能耗、连接度等进行制订。因为数据中心网络属于高动态网络,频繁地更改路由会影响网络的稳定性,而且大量的重路由增加了控制器的计算量、控制器与交换机间的信息开销和流量调度的复杂度,所以在采用 SDN 技术的数据中心网络,还需要考虑路由迭代计算周期。在根据网络拓扑节能

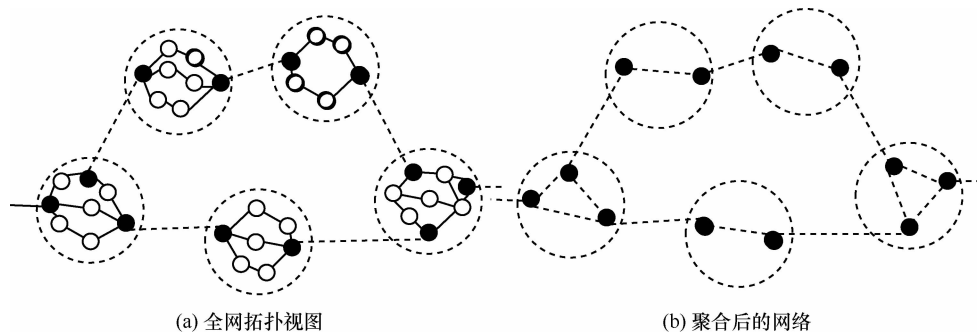


图4 一个简单的多控制域 SDN

路由的方法中,SDN 拥有全网拓扑可以很好地解决拓展性问题,算法根据全网拓扑直接计算路由,使用能耗作为路径开销,选取开销最小的路径,同时,可根据已存在的路径进行路由计算,使得新的路由最大程度地重复利用已有路径,从而集中流量,节省更多的网络资源。节能路由过于集中的流量容易造成网络性能的下降,需要根据实际的网络性能需求决定提供服务的节点和链路数。确保实时、准确的网络状态信息是节能路由选路的关键,SDN 架构的全局视图可以很好地解决该问题。同时,SDN 对全局的控制能力,使得节能路由算法可以以更小的代价获得更加复杂的路由算法和更有效地完成流量调度任务。例如,节能路由将流量集中到了网络主要链路,多条流会共享相同的链路和交换机,在这些节能路径中,负载并不均衡,多条流在主要链路带宽竞争激烈,而非主要链路相对空闲,浪费了大量传输能力,采用路径独占式路由方案^[20],让优先级较高的流独占一条链路进行传输,结束后再启动优先级较低的另一条流的传输,在同一时间、同一链路上仅有一条流传输,这样使得流不需要进行带宽竞争,且可将流量调度到较为空闲的链路上,提高链路利用率。

(6) 失效快速恢复性能

由于数据中心网络中广泛采用了低成本的低端设备,从而存在链路失效、服务器失效和交换机失效等多种故障和差错,而持续的可靠服务能力需要数据中心网络提供高效的失效恢复策略和容错机^[1]。

在非 SDN 架构的数据中心网络中,链路失效恢复采用分布式的方法实施,即检测到故障后,向各个交换机广播消息,在大规模数据中心网络中,一个小故障就会产生大量的控制消息,直接加大了管理开销,SDN 集中式控制能够有效减少故障处理的开销。

在基于 SDN 架构的数据中心网络中采用的失效恢复方式主要有两种:一种是控制器主导的失效恢复,即在链路失效消息发送到控制器后,暂停相关流的传输,对流重新进行路由计算,并在相应交换机上安装路径将流调度到其他链路,因为需要重新安装路由路径,延缓了恢复时间;另一种是交换机主导的失效恢复,在故障产生后,交换机立刻将流量切换到预先安装在交换机上的冗余路径^[20],恢复过程中不需要控制器的参与,减少了信息传递、重路由计算和路径安装的时间,达到快速恢复路由的目的,但是该方法无法处理备用路径也失效的情况,并且备用路径提前安装,增加了交换机流表项的存储负担,同时

由于没有参考实时网络状态,极可能将流路由到负载较重的链路上,而且数据中心网络属于高动态网络,因此目前大多数研究建议主要集中在控制器主导的失效恢复。

在控制器主导的失效恢复中,新路径的安装是影响路由恢复时间的主要因素,因此修改尽量少的交换机路由表项是缩短该步骤使用时间的核心。一般通过改善路径安装方式和重路由使得需要修改的交换机最少。通过重路由有两种典型方案:一种是需找失效路由到目的地的最短路径,另一种则仅进行局部的路由^[30]。如图 5 所示,当交换机 C 和 B 之间的链路断开后,只找 C 和 B 之间的迂回线路(C→A→D→B),虽然该路径不是最短路径,但大大减少了需要修改流表项的交换机数量。

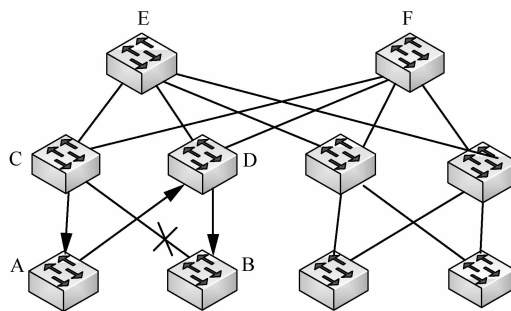


图 5 局部重路由示意图

5 结束语

随着信息社会的进一步发展,大量不同类型的数据在数据中心内进行交换,数据中心逐步成为整个信息网络的中心,其爆炸式增长的信息处理使其对网络性能的要求越来越高,而 SDN 技术是一种革新性的网络架构,它对整个网络的可靠可控性非常适合数据中心网络的新需求。与此同时,当 SDN 应用于数据中心网络时,由于需求的多样性给路由算法带来了更多的挑战。本文根据现代数据中心网络的需求和 SDN 技术特点,对面向 SDN 架构的数据中心网络路由算法的主要性能需求进行了分析总结,可以为未来数据中心网络建设提供必要的借鉴和参考。当然,本文所做的分析与建议不能代表所有基于 SDN 的数据中心网络路由算法需求,在具体网络建设中还需要根据实际需求进行针对性的选择。

参考文献

- 1 魏祥麟,陈鸣,范建华等. 数据中心网络的体系结构. 软件学报, 2013, 24(2): 295~316



- Wei X L, Chen M, Fan J H, *et al.* Architecture of the data center network. *Journal of Software*, 2013, 24(2): 295~316
- 2 Guo C X, Wu H T, Tan K, *et al.* Dcell: a scalable and fault-tolerant network structure for data centers. *ACM SIGCOMM Computer Communication Review*, 2008, 38(4): 75~86
- 3 邓罡, 龚正虎, 王宏. 现代数据中心网络特征研究. *计算机研究与发展*, 2014(2): 395~407
- Deng G, Gong Z H, Wang H. Characteristics research on modern data center network. *Journal of Computer Research and Development*, 2014(2): 395~407
- 4 Chen K, Hu C C, Zhang X, *et al.* Survey on routing in data centers: insights and future directions. *IEEE Network*, 2011, 25(4): 6~10
- 5 Benson T, Anand A, Akella A, *et al.* Understanding data center traffic characteristics. *Proceedings of the 1st ACM SIGCOMM 2009 Workshop on Research on Enterprise Networking*, Barcelona, Spain, 2009: 65~72
- 6 Ho Trong Viet, Deville Y, Bonaventure O, *et al.* Traffic engineering for multiple spanning tree protocol in large data centers. *Proceedings of the 23rd International Teletraffic Congress (ITC23)*, San Francisco, USA, 2011: 23~30
- 7 Heller B, Seetharaman S, Mahadevan P, *et al.* ElasticTree: saving energy in data center networks. *Proceedings of the 7th USENIX Symposium on Networked Systems Design and Implementation*, NSDI 2010, San Jose, USA, 2010: 2~17
- 8 Hu F, Hao Q, Bao K. A survey on software-defined network and OpenFlow: from concept to implementation. *IEEE Communications Surveys & Tutorials*, 2014, 16(4): 2181~2206
- 9 McKeown N, Anderson T, Balakrishnan H, *et al.* OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 2008, 38(2): 69~74
- 10 Al-Fares M, Loukissas A, Vahdat A. A scalable commodity data center network architecture. *ACM SIGCOMM Computer Communication Review*, 2008, 38(4): 63~74
- 11 Kanagavelu R, Lee B S, Felipe Miguel R, *et al.* Software defined network based adaptive routing for data replication in data centers. *Proceedings of the 19th IEEE International Conference on Networks (ICON)*, Singapore, 2013: 1~6
- 12 Honma Y, Aida M, Shimonishi H, *et al.* A new multi-path routing methodology based on logit type assignment. *Proceedings of the 2009 IEEE GLOBECOM Workshops*, Honolulu, USA, 2009: 1~6
- 13 Shinohara Y, Chiba Y, Shimonishi H. An adaptive multipath routing algorithm for maximizing flow throughputs. *Proceedings of the World Telecommunications Congress (WTC)*, Miyazaki, Japan, 2012: 1~6
- 14 Kanagavelu R, Mingjie L N, Khin MiMi, *et al.* OpenFlow based control for re-routing with differentiated flows in data center networks. *Proceedings of the 18th IEEE International Conference on Networks (ICON)*, Singapore, 2012: 228~233
- 15 Long H, Shen Y, Guo M Y, *et al.* LABERIO: dynamic load-balanced routing in OpenFlow-enabled networks. *Proceedings of the IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, Barcelona, Spain, 2013: 290~297
- 16 Dean J, Ghemawat S. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 2008, 51(1): 107~113
- 17 Guo C X, Lu G H, Li D, *et al.* BCube: a high performance, server-centric network architecture for modular data centers. *Proceedings of the ACM SIGCOMM 2009 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, Barcelona, Spain, 2009: 63~74
- 18 Sun S M, Li H, Cho Sungchol, *et al.* Performance optimization of media distribution in overlay networks using OpenFlow. *Proceedings of the International Conference on Information Networking (ICOIN)*, Phuket, Thailand, 2014: 276~281
- 19 Marcondes C A C, Santos T P C, Godoy A P, *et al.* CastFlow: clean-slate multicast approach using in-advance path processing in programmable networks. *Proceedings of IEEE Symposium on Computers and Communications (ISCC)*, Cappadocia, Turkey, 2012: 94~101
- 20 IETF RFC 4601. Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification (Revised), 2006
- 21 Ramanathan S. Multicast tree generation in networks with asymmetric links. *IEEE/ACM Transactions on Networking*, 1996, 4(4): 558~568
- 22 Iyer A, Kumar P, Mann V. Avalanche: data center Multicast using software defined networking. *Proceedings of the Sixth International Conference on Communication Systems and Networks (COMSNETS)*, Bangalore, India, 2014: 1~8
- 23 Shimonishi H, Ochiai H, Enomoto N, *et al.* Building hierarchical switch network using OpenFlow. *Proceedings of the Intelligent Networking and Collaborative Systems (INCOS'09)*, Barcelona, Spain, 2009: 391~394
- 24 Phan Xuan Thien, Thoai Nam, Kuonen Pierre. A collaborative model for routing in multi-domains OpenFlow networks. *Proceedings of 2013 International Conference on Computing, Management and Telecommunications (ComManTel)*, Ho Chi Minh City, Vietnam, 2013: 278~283
- 25 Egilmez H E, Civanlar S, Tekalp A M. A distributed QoS routing architecture for scalable video streaming over multi-domain OpenFlow networks. *Proceedings of the 19th IEEE*

- International Conference on Image Processing (ICIP), Orlando, USA, 2012: 2237~2240
- 26 Bruschi R, Lombardo A, Morabito G, *et al.* Green extension of OpenFlow. Proceedings of the 26th International Teletraffic Congress (ITC), Karlskrona, Sweden, 2014:1~6
 - 27 Shang Y F. Research on power-aware routing in data center networks. Dissertation of Tsinghua University, 2013
 - 28 Chiaraviglio L, Mellia M, Neri F. Reducing power consumption in backbone networks. Proceedings of the IEEE International Conference on Communications (ICC'09), Dresden, Germany, 2009: 1~6
 - 29 Li D, Shang Y F, Chen C J. Software defined green data center network with exclusive routing. Proceedings of the INFOCOM, Toronto, Canada, 2014: 1743~1751
 - 30 Ramos R M, Martinello M, Esteve Rothenberg C. SlickFlow: resilient source routing in data center networks unlocked by OpenFlow. Proceedings of IEEE the 38th Conference on Local Computer Networks (LCN), Sydney, Australia, 2013: 606~613
 - 31 Li J, JongHwan Hyun, Jae-Hyoung Yoo, *et al.* Scalable failover method for data center networks using OpenFlow. Proceedings of Network Operations and Management Symposium (NOMS), Krakow, Poland, 2014: 1~6

[作者简介]



樊自甫,男,重庆邮电大学经济管理学院应用经济学系主任、副教授,主要研究方向为通信运营管理、下一代网络技术。



伍春玲,女,重庆邮电大学通信与信息工程学院硕士生,主要研究方向为下一代网络技术、SDN。



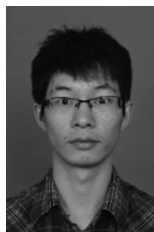
王金红,女,重庆邮电大学通信与信息工程学院硕士生,主要研究方向为下一代网络技术、SDN。

2015038-10

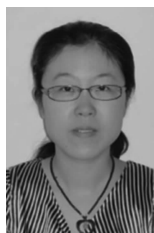
(上接 2015019-7)

- 13 胡紫巍. MIMO 和协作中继系统中的空时编码技术研究(博士学位论文). 北京邮电大学, 2013
- Hu Z W. Research of space-time coding in MIMO and cooperative relaying system (doctoral dissertation). Beijing University of Posts and Telecommunications, 2013
- 14 Muralidharan V T. Distributed space time coding for wireless two-way relaying. IEEE Transactions on Signal Processing, 2013, 61(4): 980~991
- 15 Scutari G, Barbarossa S. Distributed space-time coding for regenerative relay networks. IEEE Transactions on Wireless Communications, 2005, 4(5): 2387~2399
- 16 Sirkeci-Mergen B, Scaglione A. Randomized space-time coding for distributed cooperative communication. IEEE Transactions on Signal Processing, 2007, 55(10): 5003~5017
- 17 Jing Y D, Hamid J. Distributed differential space-time coding for wireless relay networks. IEEE Transactions on Communications, 2008, 56(7): 1092~1100
- 18 Seddik K G, Sadek A K, Ibrahim, *et al.* Design criteria and performance analysis for distributed space-time coding. IEEE Transactions on Vehicular Technology, 2007, 57 (4): 2280~2292

[作者简介]



蒋涉权,男,北京理工大学信息与电子学院硕士生,主要研究方向为无线通信、阵列信号处理。



王晶,女,北京理工大学信息与电子学院副教授,主要研究方向为音频信号处理、移动通信。



彭超,女,北京 66019 部队工程师,主要研究方向为通信信号处理、视频调度。

2015019-8