



SDN 技术在电信网络中应用的关键问题探讨

陈仲华¹, 沈成彬¹, 张坚平²

(1. 中国电信股份有限公司上海研究院 上海 200122;

2. 中国电信股份有限公司上海分公司 上海 200120)

摘要:通过对软件定义网络(SDN)技术总体架构的分析,提出了 SDN 技术在电信网络中规模应用面临的 4 个方面的问题,包括数据平面构建问题、控制平面构建问题、与应用的整合问题以及如何在电信网络内平滑演进的问题。并参考业内主流的解决技术,给出了 SDN 技术应用在电信网络中的解决方案。

关键词:SDN;控制器;OpenFlow;流表

doi: 10.3969/j.issn.1000-0801.2014.09.019

Discussion on the Key Problems in the Application of SDN Technology in Telecommunication Network

Chen Zhonghua¹, Shen Chengbin¹, Zhang Jianping²

(1. Shanghai Research Institute of China Telecom Co., Ltd., Shanghai 200122, China;

2. Shanghai Branch of China Telecom Co., Ltd., Shanghai 200120, China)

Abstract: Through the analysis on the architecture of SDN technology, the key problems in four aspects of deploying SDN in traditional communication network were provided, including data path constructing, control plane constructing, application integration and smooth evolution. Also, according to mainstream solution, a method of deploying SDN in communication network was given.

Key words: software defined networking, controller, OpenFlow, flow table

1 引言

近年来,随着互联网 OTT 业务的蓬勃发展,部分传统电信业务已经出现被 OTT 业务替代的趋势,传统电信网络被管道化已经是不争的事实。在未来电信网络管道化的大环境下,一方面,运营商参与互联网业务的竞争以维持自身业务发展是必然的选择;另一方面,如何提升网络的灵活性,满足丰富多变的互联网业务需要,也是电信运营商急需考虑的重要课题。

SDN 技术是互联网行业与电信行业融合的产物,它起源于大学校园,崛起于互联网数据中心,同时业内不少观

点也看好 SDN 技术能够应用于传统电信网络,并在提升传统网络价值方面获得成功。SDN 主流标准组织 ONF 给出的 SDN 技术的体系架构如图 1 所示。

由图 1 可知,SDN 技术的主要理念是控制层与转发层分离,将原来网络分散在各个物理设备上的控制单元集中到一起,形成总的网络控制平面,即 SDN 控制器。并通过标准化控制平面与数据平面之间的接口,实现控制层面对不同厂商提供的数据层面设备的统一控制,并提供与应用层的统一接口和数据模型。的确,SDN 控制与转发分离的理念在电信网络中的应用有助于解决长期以来困扰电信网络发展的几个问题。例如:全局优化的流量调度问题、精

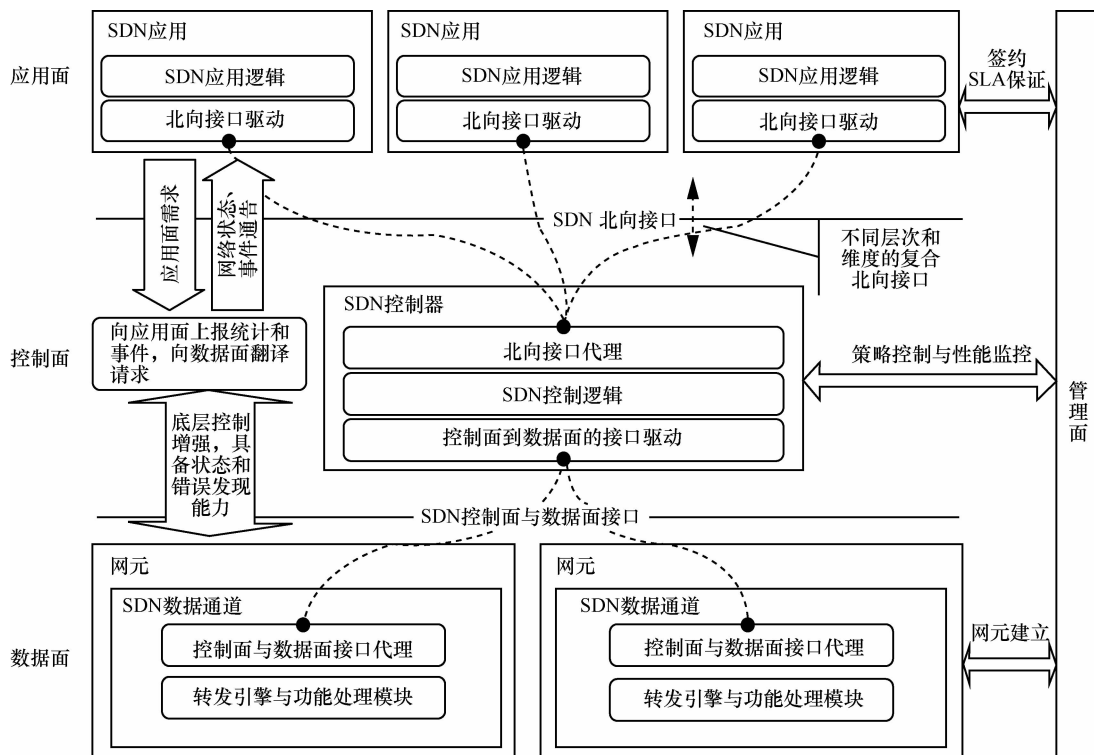


图1 SDN技术体系架构

细化的流量操控问题、与上层应用的深度结合问题以及网络功能多样化后带来的传统设备实现困难、成本高昂的问题。

但是,由于SDN技术对传统网络设备、组网方式、控制模型上都做了颠覆性的改变,因此要帮助传统电信网络实现这些功能还需要有众多的技术突破。从目前SDN技术发展的现状来看,至少需要解决数据平面的构建、控制平面的构建、与应用的深度整合、网络的平滑演进4方面的问题,才有可能在电信网络中实现预期的目标,获得广泛的应用。

2 数据平面的构建

SDN的水平分层的网络架构设计,导致传统网络设备的转发面需要采用更加开放的硬件实现。传统网络设备上用于报文转发查询的各种表项被统一成多级流表,通过流水线方式实现转发过程中的查询。流表的插入、修改、删除等操作及其数据结构被抽象成标准的数据模型和访问方法,开放给网络的控制面。要实现这样的功能开放,从目前SDN交换机实现的机制来看,大致可以分为3类。这3类SDN交换机的灵活性、扩展性和性能上存在较大的差异,未来应用在网络中的场景也可能不同。

第一类SDN交换机的实现采用ASIC架构,从硬件结构角度看,ASIC芯片实现SDN功能是比较困难的,传统ASIC实现的交换机将硬件查表、转发的功能直接通过硬件电路实现,几乎不能对查询表的结构进行扩展,因此很难适应SDN要求的由控制面来灵活定义转发面查询流表的结构和查表的方式。ASIC芯片方式实现的SDN交换机一般只能利用交换机本身的控制CPU来实现流表与ASIC硬件表的翻译,提取出流表中能够对应到ASIC表的字段,填充到原有的ASIC表中,以维持底层ASIC架构的正常工作。由此可见,第一类基于ASIC架构的SDN交换机会存在两个问题:其一,流表与ASIC表之间有一定的转换开销,在大流量转发时有可能成为系统的性能瓶颈点;其二,由于SDN流表非常灵活,有可能存在流表无法转换到ASIC表的情况,此时将发生部分SDN功能无法支持的情况。在实际使用中,第一类SDN交换机一般只能适应不需要频繁下发流表,同时流表功能比较简单的场景,例如只有二、三层转发需求的场景。但是由于第一类SDN交换机通过ASIC方式实现,其本身流量的转发性能是非常好的,可以应用在流量较大,但是转发流量的五元组会话信息变化较少的场景。

第二类SDN交换机通过NP芯片方式实现,NP芯片

在编程能力上较 ASIC 芯片有一定增强,在不改变硬件设计结构的情况下,NP 芯片可以实现比较灵活的流表定义与查询。采用 NP 芯片设计的 SDN 交换机在性能上接近 ASIC 芯片 SDN 交换机,但是在流表设计的灵活性上仍旧有所欠缺,例如 NP 芯片的核通常采用流水线方式工作,核之间缺少数据共享和协同工作的能力。此外,采用 NP 芯片设计的流表也很难实现嵌套调用,分支处理这样的流表处理逻辑结构。此外,NP 芯片的成本也要高于 ASIC 芯片。因此,NP 芯片构建的 SDN 交换机通常可以考虑应用在网络的业务接入点,在保证转发性能的同时也可以提供一定的灵活控制功能。

第三类 SDN 交换机采用 x86 芯片来实现,x86 芯片在编程能力上是 3 种芯片中最强的,可以实现的流表功能也是最为丰富的。从原理上来看,基于 x86 芯片实现的 SDN 流表,其匹配字段可以是任意字段,而不仅仅局限在 IP 分组头或者 TCP/UDP 的报文头字段。但是从性能上来看,x86 芯片实现的 SDN 交换机在 3 种芯片实现中是最低的,即使只是实现二、三层报文转发功能,x86 芯片的交换机也只能实现 128 bit 长度报文的 10 Gbit/s 线速转发。同时,x86 芯片实现的 SDN 交换机在成本上也没有太多优势,大致与 NP 芯片的 SDN 交换机相当。因此,采用 x86 方案实现的 SDN 交换机通常只能用在网络边缘以及需要进行 4 层以上交换的场景中,以提升网络的智能功能。

SDN 数据平面的另一个重要问题是转发的性能及时延问题,数据平面的时延分为两个方面:一种是由于 SDN 控制器计算路由信息带来的首分组转发时延;另一种是由于控制器计算路由后下发的流表项的不合理,导致的 SDN 交换机数据转发查询流表的时延。

首分组转发时延源自 SDN 技术转发报文的原理,由于 SDN 采用了新建五元组流的首个报文上传 SDN 控制器进行路由计算,并将计算获得的路由信息通过流表项的方式下发给 SDN 交换机。这种机制在报文上传、路由计算、流表下发的过程中不可避免地将带来报文转发的时延。

SDN 交换机转发流表导致的时延,通常是由于控制器下发流表不合理造成的,可以考虑如图 2 所示的情况。

由于 SDN 控制器将二、三层转发信息下发到一张流表中,在实际应用中会导致在做特定转发(如完成二层报文转发)时,由于单一流表中包含了很多三层转发的表项,转发查询时大大增加了查表的规模,导致了报文转发的时延增大。如果采用多流表的设计方案,则如图 3 所示。

源IP地址	目的IP地址	源MAC地址	目的MAC地址	协议号	入端口	出端口

图 2 单流表实现 SDN 转发

源IP地址	目的IP地址	协议号	出端口	入端口

(a)

源MAC地址	目的MAC地址	协议号	出端口	入端口

(b)

图 3 多流表实现 SDN 转发

三层转发的流表信息与二层转发的流表信息被拆分到两个不同的流表中,在查询的时候,三层转发的信息只需要查询三层流表来获取,二层转发的信息也只需要查询二层流表来获取,这样流表查询的规模将大大缩小,设备查表的时间也能够大大缩短。因此控制器对多级流表的支持和流表拆分算法的优化可以减小流表规模及转发报文的时延。

数据转发平面的最后一个关键问题是关于混杂模式的设计问题。混杂模式是 SDN 技术应用在电信网络中必不可少的功能,所谓混杂模式就是指一台 SDN 交换机既可以工作在 SDN 流表转发的模式下,也可以在传统的二、三层转发模式下工作。混杂模式的设计有两种方式,其一是串行模式设计,其二是并行模式设计。串行模式设计如图 4 所示。

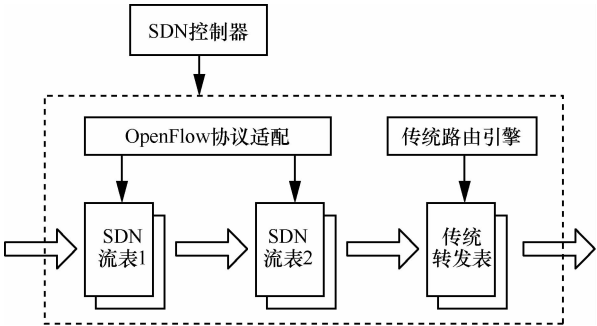


图 4 串行方式的混杂 SDN 路由器

基于串行方式实现的混杂模式 SDN 交换机,在流表转发的末端可以定义使部分流量的路由重新查询传统的路由转发表,而不是转发到 SDN 控制器进行路由计算,采用这种设计方式可以将一些不需要进行 SDN 优化转发的流量,使用传统的转发方式进行转发,有效降低 SDN 控制器的负载。在数据中心网络中,这样的设计模式可以应用在对大象流量和蚂蚁流量的差异化控制上,五元组变化不频繁的大象流量可以使用 SDN 流表进行转发优化,而五元组变化频繁的蚂蚁流量仍可以采用传统路由来转发。



并行模式设计如图 5 所示。

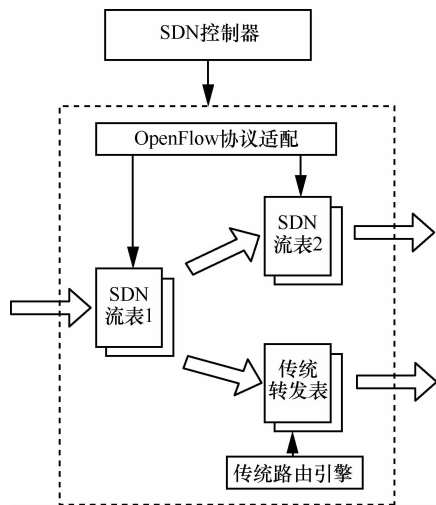


图 5 并行方式的混杂 SDN 路由器

基于并行方式的混杂模式 SDN 交换机，在一级 SDN 流表中进行协议字段的识别，例如进行 VLAN ID 的识别。通过不同的协议标记来选择采用 SDN 方式，还是采用传统路由方式对报文进行转发。这样的设计可以方便地实现叠加网络的功能，例如利用 VLAN ID 标识，甚至于 VxLAN VNI 标识可以将某些 VLAN/VxLAN 内的流量采用 SDN 流表方式进行转发，形成与传统路由表转发路径不同的虚拟、优化的转发路径。当然串行模式和并行模式也可以混合使用，实现灵活的 SDN 交换机，甚至可以同时使能这两种转发模式。

3 控制平面的构建

SDN 控制平面主要由 SDN 控制器构成，通常情况下 SDN 控制器都是构架在 x86 服务器上的一个软件系统。控制器软件的主要作用有获取数据转发面 SDN 交换机设备的硬件配置信息；网络的拓扑发现，并形成全局路由信息；根据 SDN 交换机上传的新建流的分组头信息计算转发路由，并以流表形式下发给 SDN 交换机执行。

SDN 控制器应用在电信网络时，由于单台 SDN 控制器的性能有限，引入多控制器组建控制平面的架构是唯一的选择，其控制平面在设计上需要面对以下 3 个问题。

- SDN 控制器需要获取全网的拓扑及路由信息，特别是在多控制器场景下，如何保持多控制器的全局路由表信息的一致是实现端到端转发的必要前提。

- SDN 控制器架构需要能够支持根据网络规模的扩展，进行水平性能扩展。在电信网络扩容时，也能够根据性能估算随时对控制平面进行性能的扩容。
- 在采用多控制器组建控制平面时，还必须关注单个控制器之间的负载均衡问题，如何保证多个控制器之间的负载尽可能平衡，以避免整个控制平面出现瓶颈点，是考量控制平面性能的重要指标。

目前业内控制平面的构建主要有 3 种架构，这 3 种架构各有优缺点，在实际使用中需要考虑综合因素进行选择。多控制器组建控制平面的第一种架构采用分布式文件系统进行设计，所有控制器都保存一份完整的全局路由信息，分布器之间采用发布/订阅模式进行信息的交互，网络事件以文件更新的方式反映到单台控制器的分布式文件系统中，保证信息的同步。多控制器组建 SDN 控制平面的第二种架构采用主从分布式架构，在启动时同属于一个控制器域的多个控制器将自主选择一个主控制器，这个主控制器维护域内的全局拓扑信息，域内的其他控制器作为从控制器监控主控制器的工作状态，一旦发现主控制器出现异常，立即选举合适的从控制器作为新的主控制器。这种架构通过分域选举主控制器的方式来达到控制器负载均衡的目的。多控制器组建 SDN 控制平面的第二种架构采用分层设计的控制器架构，这种架构将控制器分为两种：一种是本地控制器负责本地网络信息的管理、转发流表的生成以及网络设备的转发控制，在某些设计中本地控制器甚至可以是位于交换机设备（可以是软交换机，也可以是硬件交换机）上的一个代理；一种是根控制器，拥有全局的拓扑信息和路由信息，能够对流转发的全局路由进行优化和调度。通过两层控制器的架构，将 SDN 中的本地网络设备维护信息，例如设备发现、流发现等过程集中在本地控制器中，而全局路由优化的功能集中在根控制器中承担，减轻了不同层面控制器的性能压力。

多控制器组建控制平面的这 3 种架构各有优缺点。分布式文件系统架构，依赖成熟的分布式文件系统的设计，整体架构简单易行，但是由于文件系统的发布/订阅模式的性能不高，很难达到大规模网络拓扑管理的时延要求，因此比较适合节点数量较少、流量比较固定的网络。基于主从模式的分布式架构，将架构设计与具体的控制器实现进行了解耦，可以使用任意的控制器实现，如 NOX、

Floodlight, 在网络事件的处理性能上也有一定的保证。但是, 由于主从控制器之间需要进行实时的监控, 增加了网络带宽、时延的要求, 并且有可能要求建立带外的心跳机制, 导致系统的建设成本较高。基于两层控制器部署的架构, 对控制器的功能进行了分层, 这种设计的架构扩展性比较灵活, 但是层内的控制器仍需要有一定的架构设计。一种可行的方式是本地控制器通过分布式文件系统组建本地层面的控制平面, 而根控制器之间通过主从方式组建根控制平面。

4 与应用层面的整合

SDN 与应用层面的整合是 SDN 技术发展的一个重要方向, 可以说 SDN 的优势是通过与应用层面整合, 让应用层面能够直接调度网络设备的转发行为, 操控流量的转发而体现出来的。很多 SDN 的应用场景, 如业务链、优化流调度、负载均衡等都需要 SDN 对应用提供流操控的功能。SDN 与应用层的整合通过控制器的北向接口来实现。在 SDN 控制器的北向接口中, 需要将 SDN 控制器管理的网络设备、业务流、控制方法抽象成一定的数据结构, 并通过 XML 方式开放给上层应用调用。这种调用方式中, 应用可以直接访问到网络的具体功能, 从实现来看可能存在一定的网络安全风险, 特别是提供多个应用调用时, 如何控制权限, 是一个比较严重的问题。

一般一个大型的电信级 SDN 会构建一个编排层, 编排层将负责汇总应用的访问, 对应用的权限进行认证, 同时保证对网络功能的合理调用。目前, 比较通用的实现方式是基于开源的 OpenStack 等软件, 构建 SDN 的编排层。通过这些开源软件的部署, 不仅能够实现对网络资源的合理调用, 还能够实现网络资源与存储资源、计算资源(如虚拟机)的协同操作, 对云应用提供支持。

5 如何平滑演进

SDN 技术在电信网络中的应用必须考虑网络的平滑演进, 从演进的技术上来看可以分为以下 3 个阶段。

阶段一属于发展初期, SDN 技术会在局部网络区域内进行部署的尝试, 同时进行一些业务流控制的验证。其组网如图 6 所示。

在 SDN 区域与传统网络互通时可以采取两种技术方

案, 其一, 采用控制器或者网关设备与传统网络建立 iBGP 邻居方式与传统网络建立 peer 关系, 将 SDN 的路由信息动态发布到传统网络; 其二, 由于 SDN 区域规模相对较小, 网络配置也较简单。在网络规划时期可以为 SDN 区域单独规划地址, 网关设备与传统网络之间采用静态路由与传统网络互通。

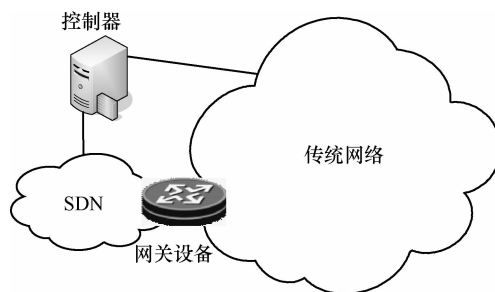


图 6 SDN 演进阶段一网络示意图

阶段二中, 网络中已出现一定数量的 SDN 区域网络, 这些网络内部通过流表转发, 与传统网络采用路由或 MPLS 方式互通。SDN 边缘地址数量增加, 很难通过事先规划方式进行管理。在这个阶段可以采用如图 7 所示的分区域路由方式进行网络的演进。

将网络划分为 SDN 区域和传统网络区域, SDN 区域内通过流表方式控制数据的路由转发, 传统网络内仍旧维持传统的路由表转发。在两个 SDN 内的 IP 地址采用 IP 隧道方式进行封装, 实现两个 SDN 的连接。在这样的架构中, SDN 控制器需要包含全程全网的拓扑结构, 并计算 SDN 的路由。因此 SDN 控制器可以执行传统路由协议, 与传统网络的路由器设备建立邻居关系, 以获得传统网络的路由信息。

在最后的演进阶段, SDN 技术已经在网络中得到了充分的应用。但是, SDN 技术并不能完全取代传统网络, 它与传统网络将会是一种共存关系, 并可能通过叠加网络的方式构建在传统网络之上。其架构如图 8 所示。

未来在传统网络上可能通过 SDN 技术构建多个虚拟的网络, 以实现业务的灵活部署。构建虚拟网络可以采用两种方式。

- 方式一, SDN 控制器本身可以在流表逻辑层面实现虚拟网络, 但在实体物理设备上不存在虚拟网络的概念, 这种方式组建的虚拟网络流表与转发之间对资源存在竞争关系。

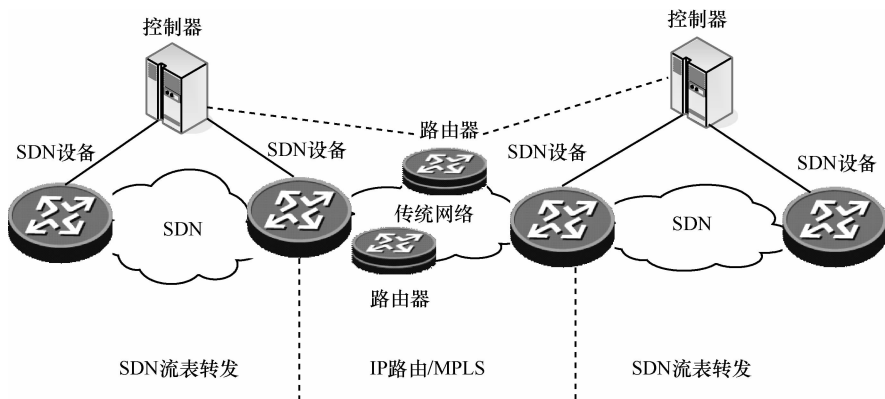


图7 演进中期组网示意

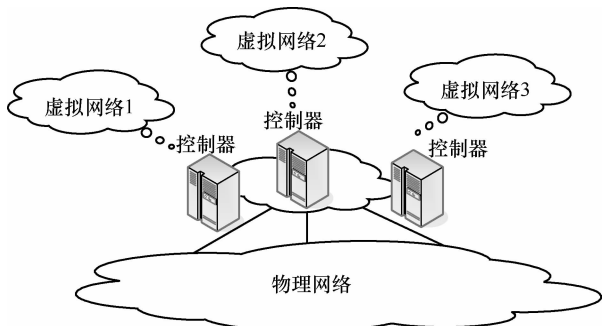


图8 演进后目标组网示意

- 方式二, 采用 VxLAN、NVGRE 等封装技术构成叠加网络, 使用 SDN 技术在物理设备上可以基于封装来实现资源的预留, 虚拟网络之间可以保证资源的使用。

6 结束语

SDN 技术对于传统网络来说, 由于其改变了网络的转发机制, 引入转发与控制分离的概念, 使网络能够更加彻底地向应用层面开放, 更加灵活地满足应用层面对网络的需求, 因此 SDN 技术很有希望成为未来网络技术的一个主要方向。但是 SDN 技术由于其具有颠覆性, 目前只能够在部分网络场景中使用, 要在传统电信网络内应用还需要解决较多的技术问题。同时, 从技术原理可以看出 SDN 技术并不是万能的, 它要想更好地在电信网络中发挥作用, 必然还需要与其他技术实现协同。

参考文献

- 1 Naous J, Erickson D, Covington A G, *et al.* Implementing an OpenFlow switch on the NetFPGA platform. Proceedings of the 2008 ACM/IEEE Symposium on Architecture for Networking and

Communications Systems (ANCS 2008), San Jose, California, USA, 2008

- 2 江国龙, 付斌章, 陈明宇等. SDN 控制器的调研和量化分析. 计算机科学与探索, 2014(2)
- 3 Alizadehzy M, Greenberg A, Maltzy A D, *et al.* Data center TCP(DCTCP). Proceedings of ACM SIGCOMM 2010, New Delhi, India, 2010
- 4 ONF. OpenFlow Switch Specification Version 1.3.1, 2012
- 5 ONF. SDN Architecture Overview Version 1.0, 2013

[作者简介]



陈仲华, 男, 中国电信股份有限公司上海研究院高级研发工程师, 多年从事 IP 网络技术研究和开发工作, 主要研究方向为 IPv6、物联网、SDN/NFV、IP 安全等。



沈成彬, 男, 博士, 中国电信股份有限公司上海研究院教授级高级工程师、技术总监, 多年从事宽带光接入与传输技术及其应用研究工作, 主导制订了多项 BBF、IEEE、ITU-T、CCSA 等国际国内标准, 先后发表了 20 多篇学术论文, 申请近 20 项专利并有多项专利转化为标准, 多次获得国家级和省部级奖励。

张坚平, 男, 中国电信股份有限公司上海分公司总工程师助理, 多年从事 IP 网络技术研究和工程规划工作, 主要研究方向为 IP 城域网、软件定义网络、IPv6 等。

(收稿日期: 2014-07-05)