

自治系统内 IP 子网和 SDN 子网的互联机制

史衍伟^{1,2}, 曹争^{1,2}

(1.东南大学 计算机科学与工程学院, 江苏 南京 211189; 2.计算机网络和信息集成教育部重点实验室, 江苏 南京 211189)

摘要: 软件定义网络 (SDN) 被认为是未来网络的发展趋势。SDN 网络与传统 IP 网络的互联机制成为当前学术界的研究热点, 但现有解决方案并不能适用于所有应用场景。为此本文提出了一个基于 OSPF 协议的 IMISA 架构, 在一个包含 SDN 子网 (基于 OpenFlow) 和 IP 子网的自治系统范围内, 通过给 SDN 控制器添加一个 OSPF 路由模块, 利用 OSPF 协议交换各自的网络信息, 最终实现了两种网络的互联。

关键词: 软件定义网络; OpenFlow; 自治系统; 互联

中图分类号: TP393

文献标识码: A

IMISA: Interconnection Mechanism for IP Subnet and SDN Subnet in Autonomous System

SHI Yan-wei^{1,2}, CAO Zhen^{1,2}

(1. School of Computer Science and Engineering, Southeast University, Nanjing 211189, China;

2. Key Laboratory of Computer Network and Information Integration, Ministry of Education, Nanjing, 211189, China)

Abstract: Software defined networking (SDN) is considered to be the trend of future network. The interconnection mechanism for SDN network and traditional IP network has been the focus in academia, but the current solution can't be used in every condition. To solve this problem, this paper proposes an architecture named IMISA which is based on the OSPF protocol. In the Autonomous System with both SDN subnet (based on OpenFlow) and IP subnet, an OSPF routing module was added to the SDN controller to help exchanging network informations and finally the networks can communicate with each other.

Key words: software defined networking; OpenFlow; autonomous system; interconnection

1 引言

软件定义网络 (Software Defined Networking, SDN) 将控制平面与转发平面分离, 采用集中控制的方式, 并提供可编程的网络和标准化的开放接口, 从而能够在了解整个网络信息的情况下做出全局优化和配置, 实现用软件的方式灵活控制整个网络, 能够很好地支持新协议和新技术, 有利于促进网络技术的发展与创新^[1,2]。

虽然 SDN 具有众多优点, 且被公认为是未来网络发展的必然趋势, 但由于其体系结构与传统网络体系结构具有非常大的差异, 目前 SDN 网络 and 传统网络并不能实现直接通信。对于 SDN 网络的

推广, 推翻现有网络从头构建一个纯 SDN 的网络是不现实的, 推广过程中必然出现 SDN 网络 and 传统 IP 网络共存的现象。因此, 研究 SDN 网络 and 传统 IP 网络的互联机制具有重要的现实意义。

2 相关工作

目前针对 SDN 网络与传统 IP 网络互联问题的解决方案主要有两个: RouteFlow 和控制器 BGP 组件。

2.1 RouteFlow

RouteFlow 由巴西最大的电信研究开发中心 CPqD 提出^[3], 是最早实现 SDN 网络与传统 IP 网络互联的机制之一。RouteFlow 主要包括三部分^[4]:

1) routeflowc。运行于 SDN 控制器 NOX 之上的

收稿日期:

修回日期:

基金项目:

Foundation Items:

应用，由 C++编写，负责控制器与 RouteFlow 服务器的交互。

2) RouteFlow 服务器。充当虚拟网络环境与控制器交互的中介，并维护一个存储网络信息的数据库。

3) 虚拟网络环境。RouteFlow 在虚拟网络环境中克隆了底层的物理网络，每个物理网络中的 OpenFlow 交换机分别对应一台虚拟网络环境中的虚拟机，并且两者的接口数目和连接情况完全相同。虚拟环境中的虚拟机运行开源路由引擎 Quagga^[5]。

当 SDN 网络中的 OpenFlow 交换机收到传统网络发来的路由协议报文时，该报文被发送到控制器，最终报文被送到虚拟网络环境中由 Quagga 按照传统网络的方式进行处理，处理结果最后经由控制器转发给 OpenFlow 交换机，然后发送给传统网络。这样，通过把传统网络与 SDN 网络的交互转换为传统网络与 RouteFlow 中虚拟网络的交互，间接实现了传统网络与 SDN 网络的互联。虽然思路很简单，但由于 RouteFlow 中每台 OpenFlow 交换机都必须对应一台虚拟机，当交换机数目较多时构建虚拟机的开销将变得很大，这大大限制了 RouteFlow 的应用范围。而且 RouteFlow 只是单纯得把传统网络路由信息转交给 Quagga 处理，真正处理路由信息的并不是 SDN 网络控制器，没有很好得体现 SDN 的特性。

2.2 控制器 BGP 组件

该方案由清华大学、斯坦福与伯克利联合实验室 (Open Networking Lab) 和 NEC 联合提出^[6]，在网络操作系统 (也就是 SDN 控制器) 之上运行一个 SDN-IP 互联应用，该应用包括两部分：

- 1) BGP 路由模块。负责处理 BGP 报文，并将路由信息存储到路由信息库中。
- 2) 流表项预安装模块。负责根据从 BGP 学到的路由信息计算并安装流表项。

给控制器添加一个 BGP 组件实现 SDN 网络与传统网络互联的方案已经被 Google 采用，部署在惠灵顿 (新西兰首都) 的真实 SDN 网络中，取得了不错的效果。但这种方案一般用于规模较大的自治系统 AS (Autonomous System) 之间的互联，而 AS 内早期的 SDN 网络建设与实验规模都较小，并不需要采用该方案。

本文提出了一个基于 OSPF 协议的 IMISA (Interconnection Mechanism for IP Subnet and SDN

Subnet in AS) 架构，在一个包含 SDN 子网和 IP 子网的自治系统 AS 范围内，由 SDN 网络的控制器 POX 来解析和构造 OSPF 协议报文，并通过位于 SDN 网络边界的转换器与 IP 子网的区域边界路由器进行交互，交换彼此的网络信息，使得自治系统内的 IP 子网能够感知 SDN 子网连接的网络，SDN 子网能够感知 IP 子网连接的网络，从而实现自治系统内 SDN 子网与 IP 子网的互联。

3 IMISA 总体设计

IMISA 总体架构如图 1 所示，所选用的 SDN 控制器为开源的 POX^[7]。

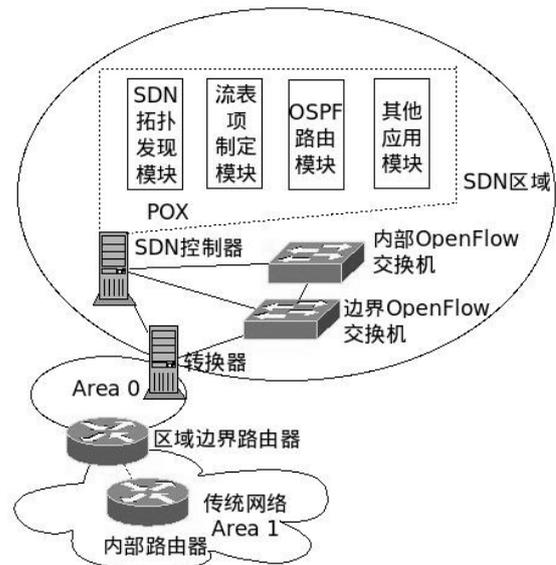


图 1 IMISA 总体架构图

在自治系统内，按照 OSPF 协议规范^[8]，将整个网络划分为三个基本区域：由传统网络设备构成的骨干区域 Area 0 和子网区域 Area 1；由转换器、OpenFlow 交换机^[9]及控制器构成的 SDN 子网区域。Area 1 通过区域边界路由器、SDN 区域通过转换器分别连接到 Area 0。

网络 Area 1 中，路由器按照 OSPF 协议规范进行交互，将本区域的传统网络连接信息添加到链路状态数据库中。SDN 网络区域中，SDN 控制器 POX 通过拓扑发现模块，获得 SDN 网络连接信息。

具体地，转换器收到 Area 0 区域边界路由器发来的 OSPF 协议报文，按照已设定的流表项将报文上传到控制器 POX。POX 调用 OSPF 报文解析模块对收到的 OSPF 协议报文进行解析，并将相应的网络信息添加到 OSPF 协议链路状态数据库中。然后，

控制器 POX 调用 OSPF 协议构造模块,生成相应的 OSPF 协议报文,发送到转换器,由转换器发送给区域边界路由器。

传统网络与 SDN 网络以 OSPF 协议类型为 3 的 Network Summary LSA 为载体,互相告知对方自己的网络连接信息。POX 处理收到的 Network Summary LSA,得知传统网络连接有哪些网络,转换为 SDN 网络的流表项下发到各 OpenFlow 交换机,使得目的网络为传统网络的报文都被转发到边界 OpenFlow 交换机,然后发送给转换器并从相应端口发送到传统网络。同理,边界路由器收到 SDN 边界的转换器发来的 Network Summary LSA,得知 SDN 网络连接有哪些网络,并通过 OSPF 报文告知区域内的其他路由器。各路由器生成相应的路由表项,使得目的网络为 SDN 网络的报文都被转发到 SDN 边界的转换器,然后转发给边界 OpenFlow 交换机,再按照流表项进行转发,最终送达目的网络。

3.1 转换器

SDN 区域边界的转换器能够执行 OSPF 协议与传统路由器进行三层的交互。同时,转换器通过 Open vSwitch^[10],使得转换器对 SDN 控制器表现为一台 OpenFlow 交换机,SDN 控制器可以通过下发流表项控制转换器的工作,将不同的数据报文转发到不同方向。转换器连接图如图 2 所示,对应的主要流表项规则如表 1 所示。

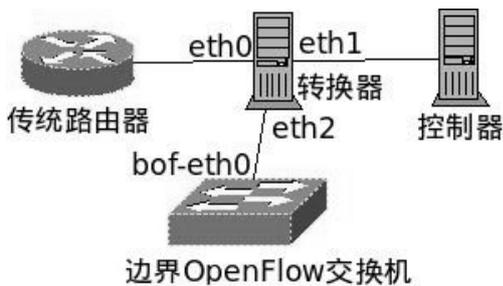


图 2 转换器连接图

3.2 SDN 拓扑发现模块

SDN 拓扑发现模块包括两个子模块:链路发现模块(用于获取 OpenFlow 交换机之间的连接情况)和网络发现模块。

POX 自带有链路发现模块。该模块控制每台交换机从其所有端口发出 LLDP(链路层发现协议)数据包,当邻居交换机收到该 LLDP 数据包时查找流表,没有找到相应流表项,便通过 Packet_In 消息上传给控制器,控制器进行分析后可获得各交

换机的连接情况^[11],并以(交换机 ID1,端口 1,交换机 ID2,端口 2)四元组的形式表示一条链路,同时获取当前时间,建立连接信息四元组到时间的映射并存储起来。注意,交换“交换机-端口”对的位置不改变链路,比如(交换机 ID2,端口 2,交换机 ID1,端口 1)被认为与上文所述链路为同一条链路。

表 1 转换器流表项主要规则

匹配域	动作
入端口=eth0 目的 IP=OSPF 组播地址	从 eth1 上报给控制器
入端口=eth0 目的 IP=SDN 所连网络地址 (由拓扑发现模块确定)	从 eth2 发到边界 OpenFlow 交换机
入端口=eth1 目的 IP=OSPF 组播地址	从 eth0 发到传统路由器
入端口=eth2	从 eth0 发到传统路由器

网络发现模块可以采用初始静态配置与动态管理相结合的方式。由于各 OpenFlow 交换机连接哪些网段基本是事先划定好的,因此可以写成配置文件的形式,以(交换机 ID,端口号,网络)三元组的形式作为一条网络连接的表示,并在系统启动时读入网络连接动态库。流表项制定模块会根据网络连接动态库计算并安装流表项,按照 OpenFlow 协议规范,流表项如果超过一定时间(保活时间)没有被命中过(说明该网络连接可能已经断掉)则会被删除,并向控制器发送 Flow-Removed 消息^[12]。控制器在收到 Flow-Removed 消息时,应将对应的可能已经断掉的网络连接信息从网络连接动态库中删除。当该网络连接再次启动时,该网络中主机第一次发送数据包到 OpenFlow 交换机时,由于交换机中没有相应流表项,就会通过一个 Packet_In 消息将数据包发送给控制器,控制器就可以从中获知该网络连接到的 OpenFlow 交换机的 ID 和相应端口,查询原来的配置文件便可得知原来的网络连接信息,将网络连接信息重新添加到网络连接动态库,并生成新的流表项。

该模块还应该提供接口,允许用户手动向网络连接动态库中添加新的网络连接信息,新添的网络连接信息也会添加到原来的配置文件当中。

除了网络发现模块的动态管理,链路发现模块也会定期发送 LLDP 报文。当检测到某台 OpenFlow 交换机连接中断时,应把相应的网络连接信息从网

络连接动态库中删除；当检测到该交换机重新连接时，可以读取配置文件重新把相应的网络连接信息添加到网络连接动态库中。

SDN 网络拓扑结构与存储信息的映射关系如图 3 所示：

3.3 OSPF 路由模块

该模块为 SDN 控制器提供像传统路由器处理 OSPF 协议报文那样的功能，是传统 IP 网络和 SDN 网络交换网络可达性信息的关键，包括 OSPF 报文解析模块和 OSPF 报文构造模块两个子模块。

3.3.1 OSPF 报文解析模块

当转换器收到传统网络边界路由器发来的 OSPF 协议报文时，按照流表项规则将报文上传到控制器。控制器的 OSPF 报文解析模块按照 OSPF 协议规范对 OSPF 报文进行解析，并根据解析出来的结果维护本地的 OSPF 协议链路状态数据库。

因为 IMISA 架构进行的是自治系统内的网络互联，所以 OSPF 报文解析模块需要进行解析的是 OSPF 协议的 Hello 报文、DD 报文（数据库描述报文）、LSR 报文（链路状态请求报文）、LSU 报文（链路状态更新报文）和 LSAck 报文（链路状态应答报文），以及型号为 1 的 Router-LSA、型号为 2 的 Network-LSA 和型号为 3 的 Network Summary LSA。如果收到传统网络路由器发来的型号为 4 和 5 的 LSA，直接丢弃即可。

3.3.2 OSPF 报文构造模块

该模块用于按照 OSPF 协议规范构造 OSPF 协议报文，并通过转换器发送到传统网络的边界路由器以进行交互。在边界路由器看来，和它进行交互的转换器仿佛就是一台传统路由器。

OSPF 报文构造模块需要构造的报文种类与 OSPF 报文解析模块需要处理的报文种类相同。特别地，该模块会根据拓扑发现模块中网络连接动态库存储的 SDN 网络连接信息，生成型号为 3 的

Network Summary LSA，其中网络地址填入 Link State ID 字段，掩码填入 Network Mask 字段。传统网络边界路由器收到转换器发来的 Network Summary LSA，便可以得知 SDN 网络连接有哪些网络，进而形成相应的路由表项。

3.4 流表项制定模块

系统启动之初，流表项制定模块应主动生成上文表 1 中所列规则并下发到转换器，确保转换器能正常工作。其中目的 IP 是 SDN 所连网络地址的流表项还要根据拓扑发现模块网络连接动态库存储的网络连接信息动态更新。

OSPF 报文解析模块维护的 OSPF 协议链路状态数据库存储了传统网络通过 OSPF 协议报文发来的有关传统网络的所有信息。在为目的地址是传统网络的报文生成转发路径控制的流表项前，为方便计算，可以先根据链路状态数据库中的 Summary Net Link States 部分提取出传统网络摘要信息，即传统网络所连的网络地址和掩码。当数据包的目的地址为摘要信息中包含的网络时，该数据包就是需要发往传统网络的数据，需要先转发到指定的边界 OpenFlow 交换机（与转换器相连的 OpenFlow 交换机，在控制器上通过配置文件静态设置）。流表项制定模块根据 SDN 网络拓扑信息计算各 OpenFlow 交换机到边界 OpenFlow 交换机的路径（该路径的逆向路径便是相反方向的数据转发路径），制定相应的流表项并下发到 OpenFlow 交换机。数据到达边界 OpenFlow 交换机后，从指定端口（同样通过配置文件静态设置）发送到转换器。

同时，该模块提供流表项定义接口，方便用户人为控制一条数据流的转发。

3.5 其它应用模块

除了上文提到的模块，还可以根据特定需要为控制器 POX 添加其它应用模块。IMISA 架构默认提供一个流量统计模块。该模块定期向 OpenFlow

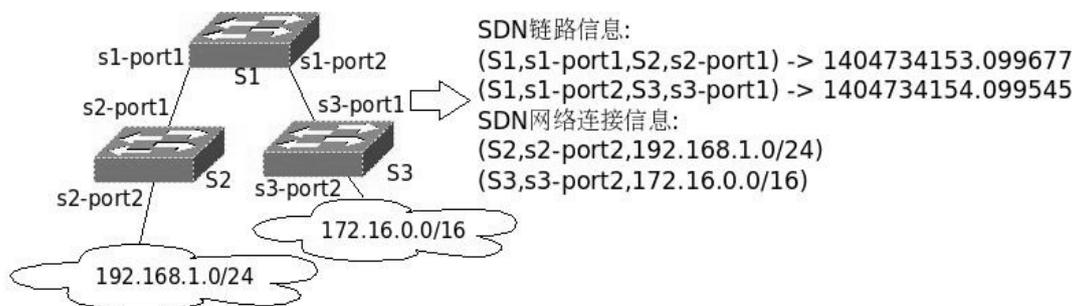


图 3 网络拓扑与存储信息映射图

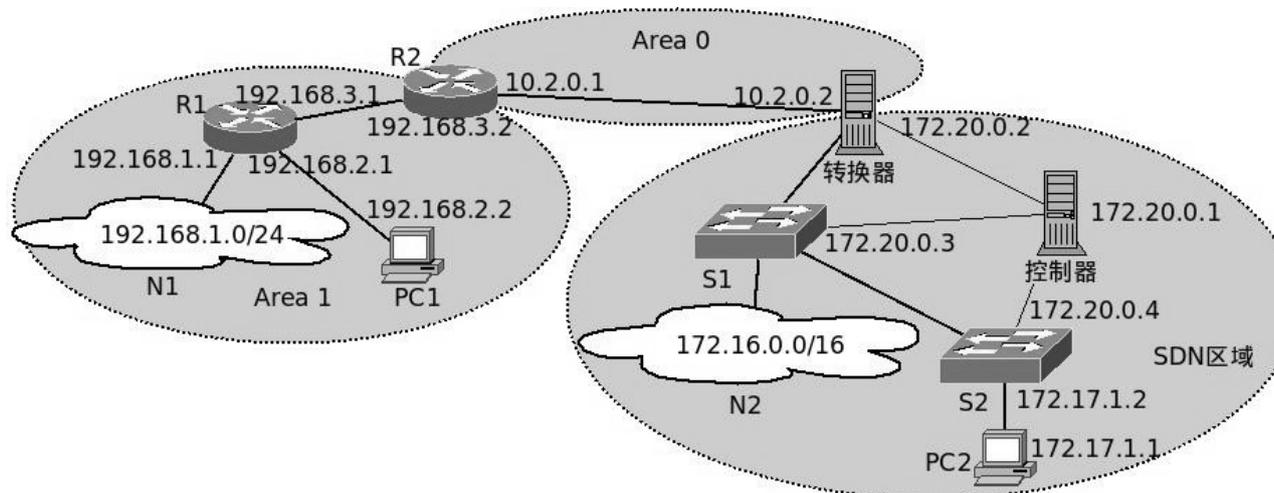


图 4 IMISA 架构部署网络拓扑图

交换机发送状态请求报文, 获得各数据流的流量统计信息, 并保存一定时间内流量最大的前 N 个流的信息 (N 根据应用需求确定)。同时, 该模块提供流定义接口, 可以按照需求对指定数据流的流量信息进行统计和汇总。

4 部署方案与验证测试

IMISA 架构部署的网络拓扑如图 4 所示:

为验证 SDN 网络与传统 IP 网络的互通性, 首先进行非常直观的测试:

1) 由 SDN 网络中的主机 PC2 对传统 IP 网络中的主机 PC1 执行 ping 命令 (ping 192.168.2.2), 发现能收到正确的返回结果。

2) 在传统 IP 网络中的主机 PC1 上架设一个简单的 WWW 服务器, 由 SDN 网络中的主机 PC2 用浏览器访问该服务器, 发现能正常访问显示网页。

然后, 登陆到传统网络路由器上执行查看路由表的命令, 可以发现到有到 SDN 网络 (比如 172.16.0.0/16) 的路由; 而 SDN 网络中的内部 OpenFlow 交换机有这样的流表项: 匹配域中目的网络地址为传统网络 (比如 192.168.1.0/24), 动作域为转发到能到达边界 OpenFlow 交换机 S1 的下一交换机。

为验证网络可达性信息能否动态更新, 首先人为将传统 IP 网络中路由器 R1 所连的 192.168.1.0/24 网络断开连接, 30 秒后 (当网络规模不大时, OSPF 的收敛速度为 10 秒左右^[13], 考虑到控制器计算并下发流表的时延, 适当延后) 查看 SDN 网络中的流表项, 发现原来与 192.168.1.0/24 相关的流表项

已经被删除; 然后人为将 SDN 网络中 OpenFlow 交换机 S1 所连接的 172.16.0.0/16 网络断开连接, 60 秒后 (IMISA 架构中设置流表项的保活时间为 40 秒, 考虑到控制器重新生成和发送 OSPF 报文以及路由器重新计算路由表的时延, 适当延后) 查看传统网络中路由器的路由表, 发现原来到 172.16.0.0/16 网络的路由表项已经被删除。

5 下一步工作

针对传统网络与 SDN 网络互联的问题, 本文首次提出基于 OSPF 协议的 IMISA 架构, 通过 OSPF 协议规范中类型为 3 的 Network Summary LSA 交换各自网络的连接信息, 实现自治系统内 SDN 子网与 IP 子网的互联。接下来的工作主要是将 IMISA 架构部署到真实网络环境中, 测试其实际性能 (比如每秒能处理的流数目、网络互通的时延等); 同时着重探索该机制能否用于以下两种场景:

1) 不同域的 OpenFlow 交换机之间交换各自网络的连接信息, 实现控制器的跨域交互 (即东西向交互)。

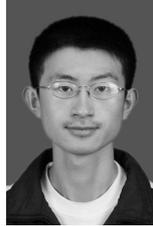
2) SDN 网络 and 传统 IP 网络的连接不再是单一出口, 存在多出口的情况。

参考文献:

- [1] Fundation O N. Software-defined networking: The new norm for networks[J]. ONF White Paper, 2012.
- [2] 左青云, 陈鸣, 赵广松, 等. 基于 Open Flow 的 SDN 技术研究[J]. 软件学报, 2013, 24(5): 1078-1097.
ZUO Q Y, CHEN M, ZHAO G S, et al. Research on OpenFlow-Based SDN Technologies[J]. Journal of Software, 2013, 24(5): 1078-1097.

- [3] 王楠. OpenFlow 网络中路由机制的研究与实现[D]. 北京: 北京邮电大学, 2012.
WANG N. Research and Implementation of Routing Mechanism in OpenFlow Networks[D]. Beijing: Beijing University of Posts and Telecommunications, 2012.
- [4] Nascimento M R, Rothenberg C E, Salvador M R, et al. Virtual routers as a service: the routeflow approach leveraging software-defined networks[C]//Proceedings of the 6th International Conference on Future Internet Technologies. ACM, 2011: 34-37.
- [5] Quagga Routing Suite[EB/OL]. <http://www.nongnu.org/quagga/>, 2013.
- [6] Lin P, Hart J, Krishnaswamy U, et al. Seamless Interworking of SDN and IP[C]//Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM. ACM, 2013: 475-476.
- [7] POX[EB/OL]. <http://www.noxrepo.org/pox/about-pox/>, 2013.
STEVENS W. TCP/IP Illustrated Volume 1 : The Protocols[M]. Beijing: China Machine Press, 2000. 226-227.
- [8] OSPF [RFC 2328][EB/OL]. <http://www.ietf.org/rfc/rfc2328.txt/>, 1998.
- [9] McKeown N, Anderson T, Balakrishnan H, et al. OpenFlow: enabling innovation in campus networks[J]. ACM SIGCOMM Computer Communication Review, 2008, 38(2): 69-74.
- [10] Open vSwitch[EB/OL]. <http://openvswitch.org/>, 2014. STEVENS W. TCP/IP Illustrated Volume 1 : The Protocols[M]. Beijing: China Machine Press, 2000. 226-227.
- [11] 雷葆华, 王峰, 王茜, 等. SDN 核心技术剖析和实战指南[M]. 电子工业出版社, 2013.
LEI B H, WANG F, WANG Q, et al. Deciphering SDN: Core Techniques and Practical Guide[M]. Publishing House Of Electronics Industry, 2013.
- [12] OpenFlow Switch Consortium. OpenFlow Switch Specification Version 1.0. 0[J]. 2009.
- [13] 王之梁, 尹霞, 范伦挺, 等. 网络路由收敛性能测试研究[J]. 厦门大学学报 (自然科学版), 2007, 2.
WANG Z D, YIN X, FAN L T, et al. Towards Network Routing Convergence Performance Testing[J]. Journal of Amoy University (Natural Science), 2007, 2.

作者简介:



史衍伟, (1989-), 男, 山东青岛人, 东南大学硕士生, 主要研究方向为网络体系结构、网络管理、下一代网络技术。



曹争, (1958-), 男, 江苏武进人, 硕士, 东南大学副教授, 主要研究方向为网络体系结构、网络管理、下一代网络技术。