

基于软件定义网络的 MPLS TE 和 VPN 网络实现方法

张顺淼

(福建工程学院信息科学与工程学院, 福建 福州 350118)

摘要: 利用 SDN 网络可编程特性和全局网络资源抽象性, 提出了一种基于 OpenFlow 协议的软件定义 MPLS 流量工程 TE 和虚拟专用网 VPN 的实现方法. 该方法使用传统标准 MPLS 的数据面和基于 OpenFlow 协议的更加简单且可扩展的控制面, 且在原型系统中验证了 MPLS TE 和 VPN 功能. 原型验证结果也进一步说明了软件定义网络技术对于简化传统 TCP/IP 协议体系网络控制面和现有网络设备软件的复杂性均很有效果.

关键词: 软件定义网络; OpenFlow; 多协议标签交换; 流量工程; 虚拟专用网

中图分类号: TP393.1

文献标识码: A

MPLS TE and VPN implementation method based on software defined network

ZHANG Shunmiao

(College of Information Science and Engineering, Fujian University of Technology, Fuzhou, Fujian 350118, China)

Abstract: Leveraging the device programmability and the abstraction of global infrastructure, this paper propose an openflow protocol based method to implementing the traffic engineering (TE) and virtual private network (VPN), in which the control protocols (such as LDP, MP-BGP) is greatly simplified and the network data is also forwarded by MPLS labels. In the prototype, the paper verifies the method of proposed software defined MPLS control plane by the functionality of MPLS TE and VPN. The results of prototype verification further illustrates the method effectiveness to simplify the control plane in traditional TCP/IP protocol architecture and reducing the software complexity in implementing the device network control function.

Keywords: software defined networking; OpenFlow; MPLS; traffic engineering(TE); virtual private network (VPN)

0 引言

多协议标签交换(multiple protocol label switching, MPLS)用短而定长的标签来封装网络层分组数据,最初是为提高路由器的转发速度而提出的一个协议.随着硬件技术的进步,采用专用集成电路 ASIC 和网络处理器 NP 进行转发的高速路由器和三层交换机得到广泛应用, MPLS 提高转发速度的初衷已经没有意义.目前, MPLS 凭借其支持多层标签和面向连接的特点,主要用于实现 IP 网络的流量工程(traffic engineering, TE)和 QoS 方面,并提供 L2/L3 企业 VPN 服务.网络提供商基于 TE 以有效利用他们的网络基础设施,同时企业 VPN 也成为他们赢利最多的业务项.网络提供商对 MPLS 方案部署,发现:①尽管 MPLS 数据面简单,但厂商对 MPLS 的支持是作为他们非常复杂、能耗高和昂贵核心路由器的额外特征,如 Cisco 的 CRS-1 和 Juniper 的 T-640;② IP/MPLS 控制面也变得越来越复杂,导致成本和脆弱性的增加.

软件定义网络(software defined network, SDN)是斯坦福大学 Nick McKewon 在 OpenFlow 协议^[1-2]基础上提出的一种新型网络架构,开放网络联盟 ONF 组织于 2011 年发布了 SDN 新型网络范式技术白皮书^[3].总结起来,SDN 网络架构的主要特点是:①解耦网络控制面和转发面,重新抽象分组转发为基于流的数

收稿日期:2015-01-19

通讯作者:张顺淼(1974-),讲师,主要从事下一代网络、智能计算研究, zshunmiao@163.com

基金项目:国家自然科学基金资助项目(61101139);福建省杰出青年基金资助项目(2012J06015);福建省中青年教育科研基金资助项目(JA14217)

据转发, 流是所控制的基本单元; ② 通过逻辑上集中的控制面操控底层分布式转发设备, 控制面由网络操作系统实现, 构建和呈现整个网络的逻辑映射图给上层网络应用; ③ 对网络抽象重新定义, 通过开放标准 API 实现网络功能的可编程性和扩展性. 目前支撑 SDN 网络架构实现的标准化协议只有 ONF 定义的 OpenFlow 协议^[4]. 尽管该协议近期发展到了 1.4 版本, 从早期适用于局域网, 发展到如今适用于广域网和数据中心, 从支持 IPv4 网络到支持 IPv6 网络; 从支持单级流表到多级流表; 从支持以太网链路层以上的逻辑资源编排到支持光层波长动态分配的物理资源编排. 但若要构建全新 SDN 网络还有许多关键技术要研究. 例如以下四个方面: ① 支持大容量 TCAM、支持任意定义多级流表的 OpenFlow 芯片设计; ② 支持可伸缩和可靠控制面的集群技术及东西向接口协议; ③ 支持上层网络应用的北向开放 API 接口; ④ 甚至对网络重新抽象并对 TCP/IP 网络协议体系的重新定义. 之所以要对 TCP/IP 协议体系进行重新定义, 是因为现在的 TCP/IP 协议体系的核心是在假设分布式网络节点和物理链路不可靠前提下建立起来的, 数据转发是依赖于捆绑在各个网络节点的控制面进行路由路径的独立计算和链路状态的独立检测来完成. 而目前网络节点和链路状态的可靠性都得到了很大的提高, SDN 架构解耦了网络控制面和转发面, 提出对网络转发进行集中式管控的思想. 为此原先控制面的许多分布式交互协议可以得到一定程度的简化, 甚至被集中式算法替代. 文献[5]基于软件定义网络的集中管控平面, 提出了一种在线流量异常检测方法. 文献[6]利用控制器中的路由统计信息, 分析了从不同交换机获取流量统计数据网络负载问题, 从而构建整个网络的流量矩阵.

为此, 本文讨论了 IP/MPLS 控制面的不足, 提出一种基于 SDN 网络架构实现 MPLS 网络的方法, 该方法通过传统 MPLS 数据面和网络操作系统(network operating system, NOS)上网络应用组件实现的控制面来提供 MPLS TE 和 VPN 业务, 以替代原来的 IP/MPLS 控制面. 为了验证技术可行性, 本文还阐述了一个通过可扩展编程控制面和 MPLS 数据面实现 MPLS TE 和三层 L3 VPN 网络应用的原型系统和结果展示, 并展现了原型系统的实现细节.

1 传统 IP/MPLS 控制面的局限性

在现在的分组交换网络中, 路由器既是对流量做路由决策的控制元素, 也是负责转发报文的转发元素. 而且这两种功能是紧密耦合在一起的, 如图 1 所示. 把控制面和转发面放在同一个盒子中必然使得网络设备更加复杂, 除了做路由决策和报文交换, 还需要具备收集和分发路由信息所需的智能, 如使用分布式路由协议 OSPF 或 IS-IS. 而在 IP/MPLS 网络中, 又还会增加如 RSVP-TE 和 LDP 的分布式信令和标签分发机制的复杂性. 在一个域内, IP/MPLS 网络可能还要另外支持别的如 iBGP、RIP、LMP、SNMP 和 MP-BGP 等协议, 甚至更多的组播和 Ipv6 协议. 所有这些叠加的网络特征导致盒子控制面过载, 进一步增加了脆弱性和硬件成本.

分布式链路状态路由协议达到稳定时是比较脆弱, 会出现处理器的路由负载和路由抖动. 若有网络状态在频繁地变化, 更新消息也被频繁地发送(有时也称为扰动), 那么路由器 CPU 将花费大量时间来进行路由计算, 导致入队列被阻塞, 同时进入的报文也开始被丢弃(包括 Hellos 报文). 丢弃 Hellos 报文反过来会引起超时, 并造成邻接信息丢弃, 又进一步导致更多的控制报文和更大的 CPU 负载. 这种连锁反应导致漫长的收敛时间, 路由循环才收敛, 更坏的情况, 甚至全部网络瘫痪. 为了避免路由协议的不稳定, 路由器厂商利用了一些阻尼机制, 它需要小心的调整以保证事情可控. 因此 IP 网络必须运维得非常小心. 在 MPLS-TE 网络中, 同样的路由协议被扩展为通告更多关于链路的信息, 甚至也给了更多的原因来产生控制流量, 如链路状态改变(通过预留带宽). 再者, 更多抑制的机制类如定时器和阈值也被用来保持事情稳定性, 比如以过时的链路状态信息为代价. 运行受限最短路径优先 CSPF 计算来产生流量标签交换路径 LSP 导

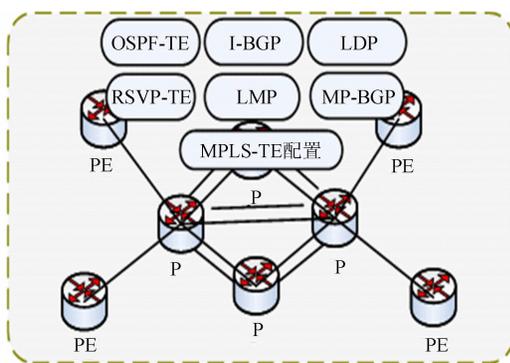


图 1 传统 IP/MPLS 网络

Fig. 1 Traditional network of IP/MPLS

致的产生过时信息的可能性,反过来也增加了使用分布式信令机制如 RSVP-TE 的必要性.

RSVP-TE 对流量工程协议来说是一个很差的选择,因为它对传统 RSVP 协议扩展了许多并不是用于流量工程的特征. 作为协议, RSVP 的软状态本质需要频繁的刷新以进行状态保活,因为它是直接在 IP 上跑的,所以也需要自己的可靠传输机制. 两种特征都导致控制面过多消息和交换上 CPU 的过重负载. 再者, RSVP 装载了过多的特征,如层次化的 LSP 建立,到多点的 LSP 建立, LSP 缝合, 帧中继建立和 GMPLS 扩展,使得已有负载协议更加膨胀. 简而言之,本研究认为 MPLS 控制面过于复杂,受 SDN 方法论的启发应该可受益于另一种设计.

2 OpenFlow/SDN 研究现状

受 SDN 思想的启发,本文提出一种开放可编程扩展网络(open programmable extensible network, OPEN)架构. 在该架构中,研究人员,或网络管理员或第三方可以在网络操作系统上通过编写软件程序来引入新的网络功能,以简化对网络分片逻辑映射的管理. 尽管 OPEN 的一些核心思想在过去被离散地提出过,但他们全部实现的意义更为深远. OPEN 希望减少各种网络基础设施的总体拥有成本,优化网络以达到他们想要的特征和服务;最后,允许更快的创新步伐,以软件速度帮忙网络更加快速的演进,也可能让更多开发人员参与以产生更多多样性/差异化解决方案.

在 OPEN 中,本文提出了使用 MPLS 转发元素作为数据面交换,他们通过 OpenFlow 协议连接到逻辑上集中的控制器上. 控制器运行管控整个网络的操作系统和一系列网络控制应用以实现所有上述提到的已存在的域内协议功能(如图 2 所示),消除了 OPEN 架构对传统 IP/MPLS 控制面协议的需求. 同时,本文认为一个新的网络特征/应用/服务提供所需要的,不等于一个新的协议或对现有协议的扩展. 一个协议应该可被使用 5~10 a,产生改变需求的时候应该是在最终的协议与原有协议思想非常不一致,或者许多厂商已经推出了多方无法互连的准标准实现的情况.

OpenFlow 协议允许控制器发现网络拓扑,并通过统计值、状态和错误消息更新来维护拓扑状态和网络状态.

OpenFlow 也提供了机制来操作每个数据面的流表. 基于网络操作系统的应用可以进行所有的网络控制决策,当提供了网络拓扑、状态和能力对流量进行流分类并控制交换转发. 控制器和应用可以决定每个流如何进行转发(无论是在新流建立时的被动下发流表,还是提前的主动下发流表),如何被路由,哪些被允许,哪儿被复制,以及接收的数据速率等等. 他们能够在每个数据面的流表中通过各种类型的动作缓存决策,因此 OpenFlow 允许对流的操作(如转发、多播、丢弃,压入、交换、弹出标记等). 因此,OPEN 的控制面网络应用决定了接入控制、路由、多播、负载均衡、隧道等等,减少了对全分布式路由和信令协议的需求;重要的是它同时也减少了每个新的服务或特征对新协议的需求——所有需要的只是一个新的应用.

控制逻辑上集中意味着决策以集中方式作出,而控制器自身可以以分布式方式分散在多个物理服务器以实现容错和性能伸缩. ONIX^[7]是一个分布式网络操作系统的例子,据称它能支持必要的规模、性能和可靠性;同时新成立的开源项目 OpenDaylight^[8]在其技术路线规划中也把控制器集群作为其基础需求之一. 相信再经过后续几年的发展,更多的网络操作系统针对不同的设计初衷将会出现用于各种研究或商业目的. 在下一个部分,本文描述了在既定的网络操作系统上,引入一种新的网络服务如 MPLS-TE、MPLS L3 VPN 到现有基于 IP 网络的开放网络环境中的简单性.

3 原型系统和实例

通过软件仿真了基于 OpenFlow 的广域网 MPLS 数据面,并在网络操作系统上构建了 MPLS 流量工程应用原型. 系统主要由下述几部分构成:

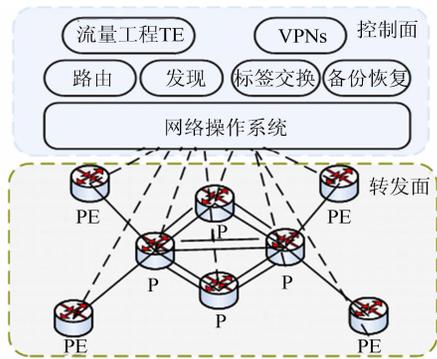


图 2 开放网络环境中的网络与 IP/MPLS 数据面
Fig. 2 Network and IP/MPLS data plate of open network environment

Open vSwitch(OVS^[9]): 一种开源的 OpenFlow 使能的软件交换, 修订后以执行 MPLS 数据面功能.

Mininet^[10]: 一种开源工具, 在独立的 PC 上可创建由软件交换互连在一起的网络以仿真网络. 本文系统在 Mininet 上运行了 4 个 OVS 实例以形成 MPLS 数据面.

OpenDaylight^[8]: 一种开源的网络操作系统, 修订以满足 MPLS 带来对 OpenFlow 协议的改变, 同时在其服务抽象层提供的 Flow Programmer、Topology Manager、Statistic Manager、Inventory Service 等服务的基础上, 扩展了 MPLS 标签管理、隧道管理、LSP 路径管理、L3 VPN APP、流量工程 APP 及相应 Web UI 等组件. 该开源系统基于 Java 语言开发, 提供 REST API 以开放网络控制能力.

MPLS API: 基于 OpenDaylight 开发的 API, 以通过 OpenFlow 协议来分发标签、获取网络状态和统计数据以支持创建、修改和管理标签交换路径 LSP.

图形用户界面 GUI 和流量发生器: 扩展了 OpenDaylight 已存在的 Web UI 来实时展示网络状态, 使用软件流量发生器来产生 HTTP、VoIP、视频流量(在 UI 中以不同颜色编码).

在原型系统中, 本文通过数据面采用传统标签交换路径 LSP, 控制面采用类似隧道的特征从技术上验证了 MPLS L3 VPN 和 MPLS TE 实现的技术可行性.

3.1 MPLS L3 VPN 网络

如图 3 所示, 站点 A 与站点 D 之间建立了一条标签交换路径 LSP_1, 站点 B 与站点 C 之间建立另外一条标签交换路径 LSP_2. 不过 LSP_1 和 LSP_2 均不是由传统的标签分发协议 LDP 来请求分配和协商标签的, 而是通过控制器上的标签管理和 L3 VPN 隧道管理应用程序组件, 根据用户在 Web UI 界面上创建的 L3 VPN 触发产生自动分配, 然后控制器经由 OpenFlow 连接通道以流表形式下发到各个支持 MPLS 数据转发的交换设备. 从图 3 中可知, 转发面依然存在 CE、PE、P 设备, 只是它们与传统 IP/MPLS 的区别在于 MPLS 中的标签 Label 产生和 LSP 建立不再依赖于分布式协议 LDP, 同时各个 PE 之间维护的虚拟路由转发表 VRF 也无需依赖于扩展的 MP-BGP 等协议进行分发, 而统一由控制器维护的 VRF 路由表管理器统一维护, 通过为 PE 设备产生不同的内层标签来区分不同 CE 站点, 以允许各个 VPN 隧道所连接站点进行相同 IP 网段及地址的规划.

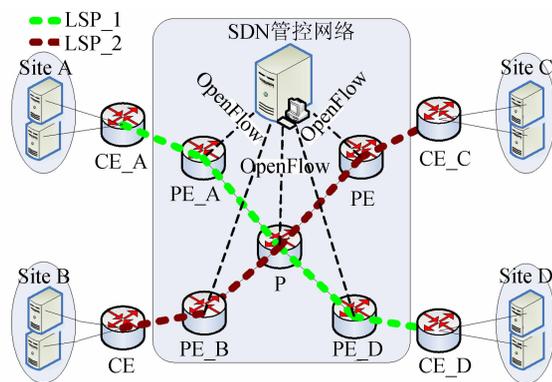


图 3 基于 SDN 的 MPLS L3 VPN 网络
Fig. 3 VPN network based on SDN's MPLS L3

3.2 MPLS 流量工程

如图 4 所示, 站点 A 与站点 B、站点 A 与站点 C、站点 B 与站点 C 之间均经过路由器 RouterC, RouterC 作为流量枢纽很容易成为性能瓶颈, 也容易造成流量路径的脆弱点.

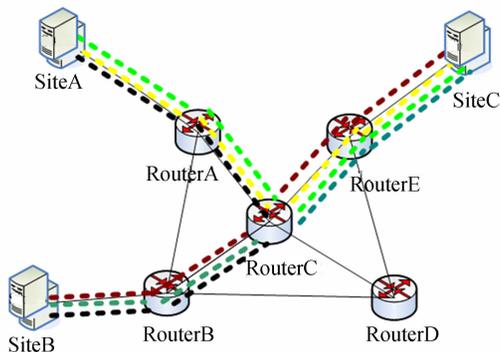


图 4 传统网络基于 CSPF 算法的 MPLS 流量工程
Fig. 4 MPLS traffic engineering based on traditional network CSPF algorithm

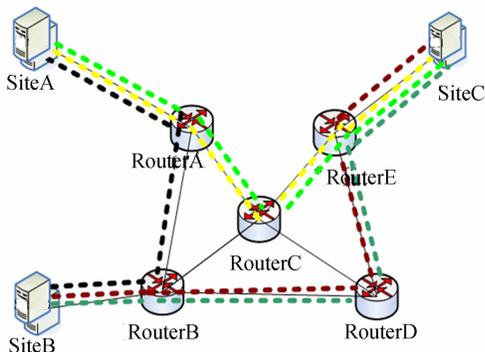


图 5 SDN 网络中基于 CSPF 算法的 MPLS 流量工程
Fig. 5 MPLS traffic engineering based on SDN CSPF algorithm

通过流量工程,在链路上预留不同优先级的带宽资源,执行条件受限最短路径 CSPF (constrained shortest path first, 受限最短路径优先) 算法功能以寻找到满足所有约束条件的隧道最短路径,并展示了管理控制的新的流量工程标签交换. 如图 5 所示,站点 A 和站点 B, 站点 B 和站点 C 之间均在路由路径上预留了足够的带宽,最终,站点 A 和站点 C 之间走 RouterA -> RouterC, 站点 A 和站点 B 走 RouterA -> RouterB 之间的直连路径, 站点 B 和站点 C 之间走 RouterB -> RouterD -> RouterE 之间的路径. 同时,本文也验证了进入的流量既可以路由到隧道,也可以路由到 IP 连接,之前使用常规 IP 连接的流量可以自动路由到新创建的流量工程标签交换路径. 用户可以在受限最短路径优先 CSPF 中指定隧道优先级以抢占低优先级的隧道,同时隧道的流量类型也可以指定(如站点 B 和站点 C 之间的流量只允许是视频流量).

4 结论

论述了传统 IP/MPLS 控制面对网络设备计算单元所带来的过多计算负载和为支持各种分布式链路状态扩展及 LDP 等协议的软件复杂度. 基于 SDN 新型网络架构实现 MPLS 网络的方法,提出使用 IP/MPLS 数据面和 SDN 开放可编程网络环境的控制面,在原型软件系统中验证了 MPLS L3 VPN 和流量工程原型系统. 目标在于支持 MPLS 数据面的 OpenFlow,使能在交换硬件中复制这种试验. 进一步简化传统 TCP/IP 协议体系控制面协议的复杂性,同时验证其他的 MPLS 控制能力,实现帧中继和 L2 VPNs 等功能.

参考文献:

- [1] 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.
- [2] Open Networking Foundation. OpenFlow protocol[EB/OL]. [2011-02-28]. <http://www.openflowswitch.org/wp/documents/>.
- [3] Open Networking Foundation. SDN[EB/OL]. [2013-08-03]. <https://www.opennetworking.org/~sdn-resources/sdn-library/whitepapers>.
- [4] Open Networking Foundation. OpenFlow switch specification 1.4.0[EB/OL]. [2013-10-14]. <https://www.opennetworking.org/images/stories/downloads/sdn-resources/onf-specifications/openflow/openflow-spec-v1.4.0.pdf>.
- [5] 左青云, 陈鸣, 王秀磊, 等. 一种基于 SDN 的在线流量异常检测方法[J]. 西安电子科技大学学报: 自然科学版, 2015, 42(1): 155-160.
- [6] Tootoonchian A, Ghobadi M, Ganjali Y. OpenTM: traffic matrix estimator for openflow networks[C]//Proceedings of the 11th International Conference on Passive and Active Measurement. Heidelberg: Springer, 2010: 201-210.
- [7] Koponen T, Casado M, Gude N, *et al.* Onix: a distributed control platform for large-scale production networks[C]//Proceeding of Operating Systems Design and Implementation. Vancouver: [s. n.], 2010: 1-6.
- [8] The Linux Foundation. OpenDaylight open source project[EB/OL]. [2012-11-07]. https://wiki.opendaylight.org/view/Main_Page.
- [9] Nicira Networks. Open vSwitch information[EB/OL]. [2013-1-22]. <http://openvswitch.org>.
- [10] Lantz B, Heller B, McKeown N. A network in a laptop: rapid prototyping for software-defined networks[C]//Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks. [s. l.]: ACM, 2010: 19.

(责任编辑: 林晓)