



综述

互联网新型服务质量保障技术研究^{*}

孟 飞, 兰巨龙, 胡宇翔

(国家数字交换系统工程技术研究中心 郑州 450002)

摘 要:网络流量规模高速增长且视频化趋势日益严重,使得传统网络为了数据传输而设计的体系架构在服务质量(QoS)保障、灵活性、可扩展性等方面经受着严峻的考验。工业界和学术界提出了很多新型网络体系结构,力求在体系结构层面增强网络的 QoS 保障能力。从“改良式”和“革命式”两种思路对互联网新型服务质量保障技术进行了详细分析,并介绍了可重构网络的宏电路技术且给出了部署考虑。可重构网络体系结构能以其对业务的适应性、功能的可扩展性等特点,适应未来网络业务发展,满足现有网络兼容融合演进的需求,为未来网络的 QoS 保障设计提供了一个解决途径。

关键词:网络体系结构;服务质量;可编程网络;可重构网络;宏电路

doi: 10.3969/j.issn.1000-0801.2014.09.017

Research on New Internet Quality of Service Guarantee Technology

Meng Fei, Lan Julong, Hu Yuxiang

(National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China)

Abstract: The scale of network traffic grows rapidly and the video trend of network traffic is becoming more and more serious, which make the traditional network architecture designed for data transmission can hardly satisfy current demands in quality of service (QoS) guarantee, flexibility, scalability, and other aspects. Industry and academia made a lot of new network architecture, aimed at enhancing the QoS capability at the architecture level. New internet QoS guarantee technology in both “incremental” and “clean slate” aspects was analyzed explicitly. In addition, wide-sense circuit and its deployment were introduced. Reconfigurable network architecture can adapt to future application development network and meet the compatible integration of evolution needs of existing network, with the characteristics of its adaptability to applications and functions scalability, providing a solution for the QoS guarantee of future network.

Key words: network architecture, quality of service, programmable network, reconfigurable network, wide-sense circuit

1 引言

随着网络规模的不断扩大及各种新业务、新应用的不断涌现,网络的 QoS 保障能力日益捉襟见肘^[1]。思科公司

预测,2012-2017 年,互联网流量综合年增长率(compound annual growth rate,CAGR)为 23%,总体规模将增加 3 倍,其中时延敏感视频类业务流量是增长的主要来源且占网络流量很大比重,P2P 技术的应用更加剧了这种现状^[2]。随

^{*} 国家重点基础研究发展计划(“973”计划)基金资助项目(No.2012CB315901, No.2013CB329104),国家自然科学基金资助项目(No.61372121, No.61309019),国家高技术研究发展计划(“863”计划)基金资助项目(No.2013AA013505)

着移动互联网的发展,视频类流量将占据更大比重。另外还有许多对实时性要求较高的应用,如IP电话、视频聊天等,对数据传送在时间域上提出了较高的时延和时延抖动保障要求。网络的QoS保障能力是网络使用者的应用需求。

现有网络体系结构僵化严重,网络控制平面与数据平面紧耦合,阻碍了网络参与者的有效沟通,使得网络各功能实体无法实现按需的资源调度。究其原因是有互联网在设计之初为了实现简洁、高效的数据传输及广泛的异构接入,形成了分层协作、端到端传输的设计原则。由于各层级之间的相对独立性,有利于上下层独立发展,促进了IP网络的成功,但各层之间的协调处理变得比较困难,使得传统QoS体系架构的实际部署困难重重。

相关研究机构对QoS的定义“异曲同工”^[1]:QoS是对各种服务性能参数,如业务可靠性、时延、抖动、吞吐量和可靠性等的描述,是服务效果的综合体现,这种服务效果决定了网络在多大程度上满足业务用户的要求,反映了用户对服务的满意程度。QoS的研究目标是提供一系列性能保障机制,有效地保障各业务的端到端传输性能,这要求系统具有端到端的接纳控制和资源预留过程以及QoS在系统范围内可配置、可维护。

由于互联网设计初期并未考虑QoS的问题,随着其日益成为社会的信息基础设施及商业利益的驱使,提高网络的QoS保障能力刻不容缓。相关研究者已经提出了许多解决方案,总结来看,主要有“改良式(Incremental)”和“革命式(clean slate)”两种。前者主要以现有互联网为基础,在不改变其体系架构的基础上进行“打补丁”式改进,继承和兼容现有网络,通过在各层插入新功能,在一定程度上提高网络QoS保障能力;后者着眼于对现有互联网体系结构进行全新设计,从而实现QoS质的提升,目前,研究者已提出了很多新型网络体系结构。

近年来,国内外已相继提出了许多相关研究计划,从不同侧面对新型网络体系结构QoS保障能力进行了积极的研究与探索。美国科学基金会(NSF)开展的FIND^[4]从路由机制和网络虚拟化等方面对网络多样化业务的QoS保障问题提出了解决方案,并开展了大量工作。GENI^[5]则旨在构建大规模的试验床框架,用于对未来网络体系架构的QoS保障能力进行试验验证。欧盟在FP7框架下开展了Challenge One^[6]项目和FIRE^[7]计划,前者研究网络虚拟化并提出全新的网络体系架构,包括4WORD^[8]和AUTOI^[9]两个子项目,对网络QoS保障方面的不足进行了研究;后者建

立了PII和OneLab两个实验床,用于为新提出的QoS保障技术提供大规模实验环境。日本的AKARI^[10]利用网络虚拟化支持多样化的数据接入和网络功能扩展的特点,针对新型网络体系结构及其关键技术对异构、复杂的多样化应用的QoS保障机制进行了深入研究。国家“973”计划“可重构信息通信基础网络研究”^[11]、“新一代互联网体系结构理论研究”^[12]、“智慧协同网络理论基础研究”^[13]等项目采用体系结构虚拟化、身份位置分离等机制构建支持多样化业务的QoS保障。

“改良式”和“革命式”两种思路是研究者从不同角度提出的QoS解决方案,目前这两种观点逐渐趋于一致,“创新”并不意味着一个全新互联网的出现,“改良”也可能导致协议和层次的变化。一种网络体系结构的QoS保障能力是由其整体功能结构形态所决定的,涉及网络的各个方面,只要是对网络应用有利的措施都可以认为提高了网络的QoS保障能力。对于解决未来网络体系结构QoS能力问题的研究还处于摸索阶段,还存在很多重大技术挑战。

2 改良式QoS保障技术研究

互联网在设计早期的主要目标是实现广泛的异构互联和有效的端到端传输,因此,形成了“核心简单,边缘智能”的网络现状。现在的互联网以IP技术为核心,采用了基于无连接的分组交换结构、存储转发的路由机制和“尽力而为”的服务模式,保证了异构网络之间的互连互通。Internet创始人之一Clark D将这种模式总结为“边缘论(end to end argument)”^[14]。“边缘论”设计模式保证了网络良好的可扩展性,促进了Internet的飞速发展,但由于其体系结构的封闭设计,控制平面与数据平面紧耦合,多样的业务资源需求无法得到按需满足。为了提升QoS,互联网服务提供商(Internet service provider,ISP)盲目地升级网络基础设施,为骨干网提供超量带宽,使得链路利用率仅为30%~40%^[15]。网络资源浪费严重,因此盲目地增加网络资源并不能从根本上解决现有网络QoS能力不足的问题。

改良式的互联网QoS保障能力提升方案中采用了“打补丁”方法,对现有的互联网体系结构以及网络运行体制进行相应的修改和增补,增强现有互联网的QoS保障能力,包括IntServ/RSVP、DiffServ、基于约束路由、MPLS与流量工程等。

2.1 IntServ/RSVP技术

传输层的综合服务/资源预留协议(Integrated service/



resource reservation protocol, IntServ^[16]/RSVP^[17]模型是在流传输之前,在传输路径上通过协商进行资源预留,对每一个流的需求和状态进行维护,是基于流的 QoS 保障。IntServ 可以分别向需要固定时延上界应用的保证型服务(guaranteed service)^[18]和可靠性要求较高应用的可控负载型服务(controlled-load service)^[19],提供无分组丢失保证和最小分组传送时延。RSVP 是面向接收端的资源预留协议,其工作流程如图 1 所示。

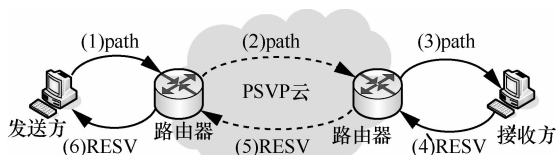


图 1 RSVP 信令过程

IntServ 模型由 4 个组件构成:信令协议(如 RSVP)、准入控制流程、分类器和分组调度器。信令协议用于在数据传输之前建立好路径并预留资源;准入控制流程是在预留资源请求时,用于判断被请求的资源是否能保证;分类器用于对数据分组进行分类,并根据分类结果把数据分组放到相应的队列中;调度器用于根据 QoS 需求对分组进行调度。

基于流量峰值预留资源具有很好的 QoS 保障效果,但却造成了资源的浪费。参考文献[20]通过分析多服务器多服务排队模型,为每个服务按照峰值预留部分资源,剩余的资源由所有服务共享,保证一定的 QoS,提高资源利用率。通过对建立的多速率流量排队模型进行分析,得到了每个服务有 QoS 保障的最小带宽资源。

为了增强 IntServ QoS 保障的可扩展性,参考文献[21]提出了一个可扩展 IntServ 架构,它将网络分为两层,每个域的边界节点保存本域内及域外相连节点间流的状态信息,减少保存的流状态信息量,在提供细粒度每流 QoS 保障的同时,解决了可扩展性问题。实际上,网络可以继续分为多层。

IntServ 架构存在几个缺陷:流状态信息量与流个数成正比,会耗费路由器大量的存储和计算资源,可扩展性很差;每个节点都要实现综合服务的 4 个组件,对于单个路由器节点的部署代价较高;保证型服务需要在整个域中部署,而经过网络瓶颈节点的可控负载型服务则需要增量部署。

2.2 DiffServ 技术

网络层的区分服务(differentiated service, DiffServ)^[22]的核心思想是“边缘分类,内部转发”。它重新定义了 IPv4 报

文头部的服务类型(type of service, ToS)域,称为 DS 域(differentiated service field)^[23],同时定义了一个分组转发的基本操作集合,称为每跳行为(per-hop behavior, PHB)。DS 域可以将数据分组划分为多个区分服务类,从而进行相应的分组处理。DS 域与原 IPv4 分组头的 ToS 字节关系如图 2 所示。其中 DS 域的前 6 bit 存储着流聚集的相关信息,被称为 DS 标记(differentiated service code point, DSCP)。不同的 DSCP 与不同的 PHB 相对应,这样才能提供区分服务。因此,PHB 实际上是一个 DS 节点调度转发处理分组头标有特定 DSCP 的 IP 分组流的外部行为特征。



图 2 ToS 与 DiffServ 域的关系

用户为了从 ISP 处获取区分服务,需要与 ISP 签订服务等级协议(service level agreement, SLA)。数据分组在 ISP 网络中的分类、监管和整形规则以及这些操作占用的存储资源都由 SLA 规定。

根据不同的分类、监管、整形和调度操作,数据分组接收的服务可以有很多种。典型的服务有奖赏服务(premium service)、保证服务(assured service)、奥林匹克服务(Olympic service)等。奖赏服务适用于需要低时延、低抖动的应用;保证服务适用于需要高可靠性的应用;奥林匹克服务提供了金、银、铜三层等级服务,其服务质量也是依次递减。

DiffServ 是一种基于资源共享策略的 QoS 保障技术,与基于资源分隔策略的绝对 QoS 保障技术——IntServ 是两种截然相反的 QoS 保障技术,二者各有千秋。参考文献[24]通过理论推导,证明了 IntServ 对最坏状况流的 QoS 保障效果较好,而 DiffServ 对全部流的平均 QoS 保障效果较好。

tsSDM^[25]是一个在 DiffServ 中融入 IntServ 的两阶段区分服务模型,如图 3 所示。第一阶段根据分组头是否包含 RSVP enabled DSCP 进行分类,若是,则数据分组被放入 RSVP 的动态队列中,否则按普通 DiffServ 程序处理;第二阶段,连接两队列池的优先级调度器分别给予 RSVP 队列、DiffServ 队列高优先级和低优先级,前者按 IntServ 模型进行细粒度预留资源保障。

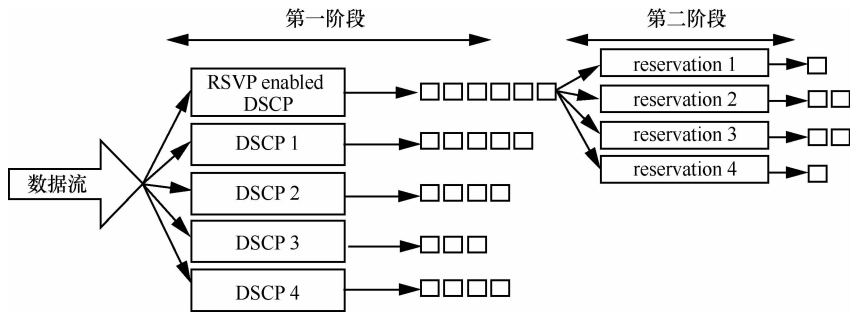


图 3 tsSDM 分类过程

DiffServ 架构中主动队列管理(AQM)是一个有效的控制机制,可实现分组丢失率与吞吐量之间的权衡,但它不能提供低复杂度且成本效益高的区分带宽分配。CHOKeR^[26]采用多优先级比例带宽分配策略,能提供 TCP 保护,确保 TCP 连接的带宽占用。CHOKeR 实现了不同优先级流的比例带宽分配及同等优先级流的公平分配,且只需保存优先级水平而不是每流状态信息,复杂度低。

实时带宽分配由于时延的存在不一定合理,且开销大,参考文献[27]基于马尔科夫链理论预测下一时段带宽需求,进行提前带宽分配,有效减小了开销和连接阻塞概率。

可以看出,目前的研究都是从结合 IntServ 和 DiffServ 二者的优点着手,在提供一定程度的细粒度 QoS 保障的同时,保证了体系的可扩展性。基于预测进行资源分配指明了一个新的研究方向,预测模型的准确性、复杂性是关键所在,未来还可以探讨实时分配与提前分配相结合的策略,根据网络状况,拥塞时基于预测提前分配,时延较小时实时分配。

DiffServ 架构存在几个缺陷:数据流根据 QoS 类别的不同而得到不同的处理,面向粗粒度汇聚流提供保障,QoS 保障效果不及基于每流的细粒度保障机制,灵活性差;由于 DiffServ 结构中网络和端系统之间缺乏信令通信,不能提供端到端的 QoS 保证;DiffServ 每跳处理,各转发节点独立配置策略,缺乏全局网络视图,无法做出最优 QoS 决策;只能提供相对的 QoS 承诺,不能向用户提供绝对的 QoS 保证。

2.3 基于约束的 QoS 路由技术

基于约束的路由(constraint-based routing,CBR)^[28]根据 QoS 参数的要求和策略选择路由。由 QoS 路由发展而来,QoS 路由在给定流或流聚集的 QoS 请求时,查找出最符合 QoS 需求的路由项。CBR 扩展了 QoS 路由,它不仅考虑流或流聚集的 QoS 需求,同时考虑网络的其他约束,如路由策略等。CBR 的目标包括两条:一是选出满足特定 QoS 需

求的路由;二是提高网络的利用率。CBR 在选择路由时,不仅考虑网络拓扑,还考虑流的传输需求、链路资源可用性以及一些由网络管理员指定的策略,因此 CBR 选出的路径可能不是最短,但是负载较轻,该路径要优于最短但负载很重的路径,网络流量的分布也因此更加均衡。

在链路状态更新的频率与路由开销之间需要做出权衡,参考文献[29]使用模糊路由算法,通过在链路开销更新过程中使用模糊逻辑关系函数修改 Dijkstra 算法,基于网络全局负载均衡限制计算新路径,以达到减小链路请求失败率和增加网络资源利用率的目的。CBR 路径计算有离线和在线两种^[30],在线路径计算对于网络简单、快速的路径计算,细致的资源分配及负载均衡等需求来说是必需的,可减小请求堵塞的可能性,在提高流的 QoS 保障效果的同时提高资源利用率。

CBR 的缺陷在于:较大的计算和通信开销;面向汇聚流而非微流,是粗粒度的 QoS 保障;路由表的存储开销大;CBR 选择的长路径路由带来资源开销和不稳定性。

2.4 MPLS 与流量工程

多协议标签交换(multi-protocol label switching,MPLS)^[31]是一种基于短标签的快速数据转发方式,MPLS 标签处于数据链路层(L2)和网络层(L3)之间。MPLS 数据分组头结构以及其与数据链路层数据帧头和网络层分组头的关系如图 4 所示。利用该标签,MPLS 将第二层交换的速度、性能和第三层路由选择的可扩展性、灵活性结合在一起,提供了一种高速高效的面向连接的传输技术。

一个支持 MPLS 的路由器称为标签交换路由器(LSR),LSR 在处理数据分组时仅考虑标签值。MPLS 通过

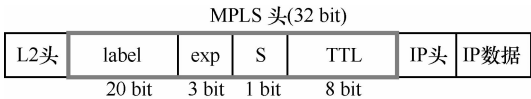


图 4 MPLS 头部结构及其与链路层和网络层间的关系



标签分发协议来建立标签交换路径(LSP)。LSP的建立可以采用控制驱动,即由控制流量(如路由更新)触发,也可以采用数据驱动,即由流或流量干线(traffic trunk)的请求触发。流量干线是具有相同服务类别的流的聚集。两个节点间的LSP可以与网络层逐跳路由相同,也可以是显式路由(explicit route)。在标签分发之后,LSR会建立一个由标签索引的转发表,转发表中的表项指明了对应标签的数据分组该如何处理。

数据分组在MPLS域的入口处根据分组头和一些其他信息,例如此数据分组到达的接口,确定其归属的转发等价类(forwarding equivalence class, FEC)。FEC会根据设置时考虑的信息,提供一个粗糙的或者精细的转发粒度,然后将一个与该FEC相关联的标签加在数据分组头。后面的LSR通过查看到达数据分组的标签并为其替换新标签,决定其转发出口,转发到下一个LSR。在数据分组离开MPLS域时,标签会被删除。

MPLS中LSP可以作为隧道使用,一旦LSP建立起来,数据分组的路径就完全取决于入口LSR分配给数据分组的标签。基于MPLS的QoS可行性体现在两个方面:一是它提供了更快的分组分类和转发机制;二是它提供了高效的隧道机制,由此MPLS在流量工程中也起了很大的作用。

流量工程^[32]是一种通过避免拥塞而实现网络流量QoS保障的机制,通过合理安排数据流流经网络,避免由于网络负载不均衡而导致的网络拥塞。基于MPLS实现流量工程的主要思想是将MPLS流量干线安排到合适的LSP中^[33]。目前,基于限制路由的标签分发协议(constraint-based routing label distributions protocol, CR-LDP)^[34]和隧道扩展的资源预留协议(RSVP-TE)^[35]是MPLS网络实现流量工程的信令协议。

MPLS网络在进行流量工程扩展之后,可以对流量干线按照显示路由的方式指定LSP,在避免拥塞的同时,也保证了传输的可靠性。虽然MPLS可以将数据流按照类型聚集成为流量干线,但是并不提供区分服务。MPLS-TE对LSP上的所有类型流量均采取相同的处理,不区分业务类型,在资源短缺时,无法提供对高优先级业务的QoS保障。区分服务感知的流量工程(DiffServ-aware traffic engineering, DS-TE)^[36]结合了MPLS高效的路由策略和DiffServ的区分服务特性,能够实现多类型业务QoS保障,同时提高网络资源利用率。DS-TE的实现需要将6 bit DSCP字段映射到3 bit exp字段,目前有E-LSP(exp-inferred-PSC LSP)和

L-LSP(label-inferred-PSC LSP)两种方案,前者表示的PHB太少,后者标签和PHB之间的映射管理过于复杂。参考文献[37]提出使用exp承载部分DSCP字段,把其余DSCP值编码到标签字段,经过解码之后便可恢复原始标签值和DSCP值,而且通过使用IPv6标签信令协议动态地在DiffServ域建立标签,在可用标签空间和区分服务的类别数量之间取得最优权衡。

备份路径通常提前部署在应用MPLS技术的核心网中,以避免单点链路失败造成的重大损失,提高对数据流服务的可靠性。Francois F^[38]等人提出了一个在线分布式的流量工程设计方案GBP,每个路由器周期性监测其本地流量状况并共同决定如何重路由流量,通过对长备份链路的限制使用,减少了传输时延,并且可以通过在线的休眠重配置来满足时延需求,提供QoS保障。

多限制条件最优网络负载均衡是一个NP难问题,且是MPLS-TE的重要组成部分,解决NP难问题通常使用智能算法。参考文献[39]提出了一个优化MPLS负载均衡问题的灵活基因算法(FGA),在解决流量均衡部署的同时,最小化了路由开销。

MPLS提供了简洁高效的面向连接转发策略,可对多样化业务流提供有效QoS保障。但在QoS保障方面仍然存在如下缺陷:MPLS的流分类基于特定长度标签,分类空间有限,是一种面向汇聚流的粗粒度QoS保障;缺乏动态自适应性,无法进行最优的服务配置;转发路径的选择仍基于分布式路由协议,没有对底层资源进行抽象,缺乏全部网络视图,无法进行全局的资源优化配置。

2.5 改良式QoS保障技术缺陷

现有的QoS保障技术能够满足网络业务流量在可靠性和区分服务方面的需求,且不同技术有其自身的优势。然而,综合上述分析,现有技术仍存在如下缺陷。

- 缺乏支持新型业务的可扩展机制。现有技术在部署期间需要指定接受QoS保障的业务类型,在运行期间仅对这些指定的业务提供QoS保障服务,无法适应新业务的发展需求。
- 缺乏基于同类业务聚合的资源预留机制。由于现有技术在保障特定业务QoS需求时,并不考虑不同业务QoS需求的关系,无法将QoS需求相近的业务进行聚合,也无法对同类业务统一进行资源预留,导致网络在提供QoS保障服务时浪费资源且效率低下。

- 缺乏网络自适应性。由于网络中各类业务的流量以及流量的业务类型都在不断变化,而现有技术无法根据网络中的流量变化做出调整,使得网络不能合理分配资源,无法满足网络中业务流量的 QoS 需求。
- 缺乏面向连接特性。现有 QoS 保障机制大都着眼于寻找最优的转发路径、合理布局流量等,以优化网络资源利用率。但由于其无连接性,数据分组在传输过程中仍会存在乱序、分组丢失等问题,且数据分组本身的开销也大。
- 缺乏根本有效性。现有 QoS 保障技术均是对传统网络体系架构进行功能性修补,如图 5 所示,没有从根本上改变控制平面与数据平面、业务与服务之间复杂的紧耦合关系及数据的非连接交换方式。对各层添加一些辅助性措施,虽然一定程度上提高了网络的 QoS 保障能力,但是过多“额外”功能的加入使得网络承载的功能不断扩展,显得臃肿不堪,网络的结构变得十分复杂,管控功能日趋减弱。

应用层	
传输层	IntServ/RSVP DiffServ
网络层	CBR
	MPLS
链路层	

图 5 QoS 保障技术位置

当前互联网无法满足多样传输 QoS 需求的问题,体现了网络中存在两个方面的矛盾:一是网络服务模式的单一性与网络业务的多样性之间的矛盾;二是网络功能与体系结构的相对稳定性和网络业务需求的不确定性之间的矛盾。网络 QoS 难以保证在体系结构上存在深层次的原因,仅仅依靠更加先进的网络管理工具不能从根本上解决问题。一方面,互联网“边缘论”和面向非连接的设计思想保障了高效的互通及广泛的接入,但控制手段薄弱,难以满足 QoS 保证的需要;另一方面,互联网控制逻辑和转发逻辑紧耦合在一起,各种新功能的加入使得互联网已成为一个庞大的非线性复杂系统,不利于新协议、新技术的部署。

3 革命式 QoS 保障技术研究

当今信息网络依赖的基础互联传输能力是制约整个

信息网络总体功能的瓶颈,另一方面,网络内在的能力与结构对业务需求的适应性差,导致网络的 QoS 保障能力低下,使得业务需求与网络基础能力之间的差距愈发显著。而改良方案依然存在诸如网络层功能单一、业务与网络过度耦合、网络协同不力等问题,这些缺陷导致网络的 QoS 保障能力不足。

为了实现网络对数据传输的 QoS 保障,应该在宏观的全网范围和局部的转发节点两个层次上实现数据的按需转发。传统 QoS 体系架构是对 IP 网络的“修修补补”,没有从根本上改变网络层的“细腰”结构以增强网络的 QoS 保障能力。可编程网络的提出可以弥补现有 IP 网络的缺陷,它在顶层强调控制面与数据面的分离,底层强调控制面和数据面内的模块化和标准化。其思想为网络与业务解耦合,使网络服务具有普适性。国内外相关研究包括 IEEE P1520、IETF GSMP、MSF、IETF ForCES、NPF、SoftRouter、Intel 提出的 IXA、Stanford 提出的 SDN/OpenFlow、可重构网络等。

3.1 基于 ForCES 的 QoS 保障技术

ForCES^[40](forwarding and control element separation, 转发控制分离)将路由器控制件(control element, CE)和转发件(forwarding element, FE)分离,使其解耦为多个 FE、CE 及连接它们的 ForCES 协议^[41]。ForCES 网络件(network element, NE)由一个或多个 CE 与一个或多个 FE 构成,是对外隐藏内部细节的执行实体。NE 内 FE 与 CE 通过定义的标准化 ForCES 接口进行信息交互。CE 负责路由计算,执行控制和信令协议,并配置和管理 FE;FE 根据来自 CE 的控制命令对数据面分组进行转发,并通过上报机制产生事件消息。ForCES 整体架构^[42]如图 6 所示。

图 6 中, Fp (ForCES protocol, 转发控制分析协议)为

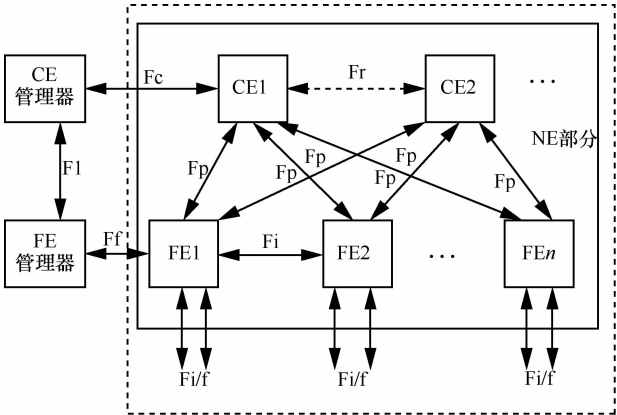


图 6 ForCES 整体架构



CE 和 FE 之间通信的协议。Fi 表示 FE 之间的接口,Fr 表示 CE 之间的接口,Fi/f 表示 FE 的外部接口,F1 为 CE 管理器与 FE 管理器之间的接口,Fc 为 CE 管理器与 CE 之间的接口,Ff 为 FE 管理器与 FE 之间的接口。

FE 逻辑上由多个细粒度定义的逻辑功能块(logical function block,LFB)组成,CE 对 FE 的控制操作本质上是对相关 LFB 的控制操作。细粒度定义的功能单一的 LFB 包括分类器、整形器、标记器、调度器和转发器等,多个 LFB 以一定的规则组合排序便可构成相应的复杂功能实现,如 IntServ、DiffServ 等,可实现控制平面要求的各种与 QoS 相关的流量处理。CE 对每个分组的处理实际上是其数据路径上一系列 LFB 对分组的操作,如图 7 所示。

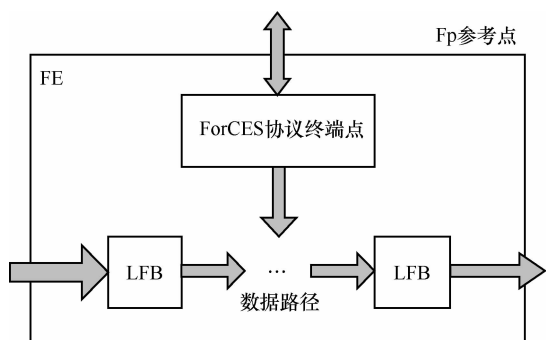


图 7 ForCES FE 基本结构

CE 通过 Fp 参考点和 ForCES 协议接口,可间接地对 LFB 的输入、输出和属性进行询问和操作。图 7 中,水平方向描述了一个 FE 内部整个转发平面各个 LFB 之间的输入和输出关系。垂直方向表示 CE 与 FE 之间通过 Fp 参考点形成的双向交互通道;CE 对 FE 执行一些配置、查询和控制工作,或向 FE 发送数据分组;FE 对 CE 的通信主要是向 CE 发送状态信息,如重定向数据分组、报错信息、监控或统计信息等。

参考文献[43]基于网络演算理论推导了网络的端到端时延确定上界,对 ForCES 路由器系统的端到端通信性能进行了研究,为 QoS 保障提供了理论指导。参考文献[44]提出了一种提供 QoS 的 ForCES 路由器模型,通过在一个网络处理器上加入一系列支持 QoS 的可动态调整的 LFB,各 LFB 分别用来实现 DSCP 分类、六元分类、DSCP 标记、流量测量等,实现了一个转发与控制分离的 QoS 路由器。

ForCES 是编程网络思想的一个实现,实现了控制平面与转发平面的分离,有利于 QoS 保障的实现,但仍有些许缺陷,具体介绍如下。

- ForCES 框架协议还没有完善,许多标准正在制订中,如动态 LFB 的研究、ForCES 通道传输方案的研究及除 Fp 参考点外的其他参考点的细节问题和实现问题等。因此,其没有进行大规模的实际部署。
- ForCES 本质上没有改变网络的结构,仅在 NE 内实现了控制转发分离,对外仍表现为一个控制平面与转发平面耦合在一起的单个网络元素。
- 数据分组的传输仍为基于无连接的分组交换,数据传输 QoS 保障效果受限。

3.2 基于 SDN 的 QoS 保障技术

SDN (software defined networking, 软件定义网络)^[45]是编程网络的一种实现方法,它把网络的控制平面和转发平面解耦合,控制器把转发规则下发到转发平面,这样交换节点就可以去智能化,成为“哑的”简单的数据转发设备。这样做一方面减轻了网络的复杂、冗余,另一方面把网络的复杂性由对硬件设备的依赖转移至软控制器。

SDN 体系架构对网络进行两次抽象,使网络分为基础设施层、控制层和应用层 3 层,如图 8 所示,其中,流抽象屏蔽了转发层具体的传输数据细节,使各种转发方式表现为流表的形式,为控制平面提供统一灵活的转发操作。网络视图抽象能使应用层掌握全局网络细节描述,屏蔽网络应用状态分发细节,共同地、全局地优化网络功能。控制层中控制软件与基础设施层的底层转发设备经由控制数据平面接口(南向接口,如 OpenFlow)交互,并通过可扩展北向接口与应用层交互。OpenFlow 交换机由流表、安全通道和 OpenFlow 协议 3 部分组成,是整个 OpenFlow 网络的转发平面。转发规则以流表项的形式存在流表中,到达的数据分组通过匹配流表进行相应的操作处理,过程为:

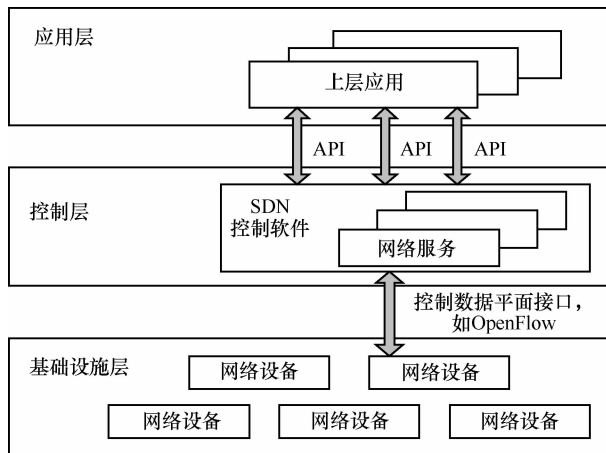


图 8 SDN 体系架构

OpenFlow 交换机接收到数据报文后, 首先进行流表匹配, 并执行相关动作; 若找不到匹配表项, 则根据此 table-miss 表项的配置, 把报文上交给控制层并由控制器决定转发行为——丢弃或转发至下一个流表。控制器通过安全通道按照 OpenFlow 标准协议更新 OpenFlow 交换机中的流表, 从而实现对整个网络流量的集中统一管控。

由于 SDN 可以通过控制器对全网流量进行统一操作, 对全网 QoS 部署自然有先天优势, 基于 SDN 解耦控制与转发和对网络集中式管控的思想提供 QoS 保障已得到广泛关注并取得很多研究成果。为了有效地部署 QoS 策略, 需要实时获取网络时延, 参考文献 [46] 描述了一种 OpenFlow/SDN 环境中测量时延的机制, 一个控制器监测全网所有交换机, 通过控制器与交换机之间按照 OpenFlow 协议设计的数据分组的交互, 实现网络时延的精确测量, 比 ping 更精确且不需额外的硬件等网络资源消耗。OpenFlow 支持的 QoS 能力有限, 在最新的 OpenFlow 1.3.0 版本中, 仅设计了每流的最大最小速率限制机制, 可在网络边缘用于流量整形, 但不能提供电信级 QoS 支持。参考文献 [47] 设计了 OpenFlow-MPLS 联合架构, 实现了具有相同 QoS 等级的多虚拟网隔离, OpenFlow 作为 MPLS 网络的接纳控制模块, 使得 MPLS 的全部功能得以利用, 从而支持电信级 QoS 保障。OpenFlow 控制器保存各虚拟网预留信息, 当来自某个虚拟网的具有特定 QoS 等级流量所需的网络资源可用时, 此流量便可进入网络, 本设计没有对 OpenFlow 协议和 MPLS 做任何修改, 实现了对每个虚拟网的 QoS 保障。为了对多媒体流量提供端到端 QoS 保障, OpenQoS^[48]首先把进入网络的流量分为数据流和多媒体流, 前者使用传统的最短路径转发, 后者动态分布于具有 QoS 保障能力的路径上, 这是一个约束最短路径 (constrained shortest path, CSP) 的 NP-complete (NP 完全) 问题, 路径管理功能更新 QoS 参数, 拓扑管理功能监测拓扑变化, 路径计算功能使用基于汇聚成本的拉格朗日松弛 (lagrangian relaxation based aggregated cost, LARAC) 算法解决上述 CSP 问题, 控制器相应地更新交换机流表实现动态 QoS 路由。利用 SDN 控制器具有全局网络视图的优势, VSDN^[49]为视频流动态选择最优转发路径并沿途预留资源, 从而提供端到端的 QoS 保障。

SDN 集中式的管控架构有利于流量的管理与操作, 控制、转发分离架构增强了 QoS 策略部署的可扩展性, 对 QoS 保障具有天然的体系架构优势。但 OpenFlow 仍在不断

发展完善中, 尚存在一些缺陷, 具体介绍如下。

- 目前, OpenFlow 协议支持的 QoS 特征有限, 只能设置队列最大最小速率, 不能进行队列策略配置, 无法进行更细粒度的流量管理。现有基于 OpenFlow/SDN 的 QoS 保障方案均是利用其体系架构优势, 通过应用层部署相应应用来实现的。
- 面向应用的可扩展北向接口仍处于需求讨论阶段, 尚未实现标准化。
- 底层资源的抽象粒度与管理开销之间要进行必要的权衡, 随着 OpenFlow 版本的不断升级, 匹配字段长度不断增加, 匹配时延及管理开销逐渐增大。
- 支持 QoS 的可扩展性问题。首先, 由于控制器存在性能瓶颈, 同时处理多个流时, 会导致流传输的端到端时延增加甚至无法及时应对, 造成链路拥塞。再者, 在流添加速度上存在硬件性能限制。

MPLS 具有电信级的 QoS 保障能力, 但是由于其被封闭在僵化的体系结构中, 创新型部署困难。基于 OpenFlow 的 SDN 为网络的创新提供了必需的可编程性和灵活性, 目前, 已经支持 MPLS 的全部特性并可与现有 MPLS 设备共存, 如 Sharafat^[50]等人将 OpenFlow、MPLS 分别位于控制平面、数据平面, 执行 MPLS-TE 及 MPLS-VPN, 实验证明了其部署的易行性, 可在提供高质量传输效果的同时减少传统分布式控制的开销。对 OpenFlow 与 MPLS 结合的研究将是一个有价值的研究点。

3.3 可重构网络的 QoS 保障技术

可重构网络是一种通过网络结构及功能的自调整来动态匹配业务, 从而构建一个可以提供多样化和全方位的网络业务、满足多样性的业务要求、具备高质量通信效果的新型网络体系结构^[11], 其功能参考模型如图 9 所示。其主要理论有元能力理论、认知与网络重构机理及多态寻址路由机制。其本质是网络资源及其组合、使用方式等的动态改变, 以匹配时变的业务应用要求和特征, 是一种以“以变应变”方式进行网络服务, 提供高可靠 QoS 的结构形态。

可重构网络体系结构所提出的“业务—元服务—元能力”模型^[51]是一种细粒度组合服务思想的体现, 如图 10 所示, 通过网络的可分最小功能单元逻辑上的有机排列与组合, 形成的具有对应功能要素的实体, 可对特定业务实施有较强针对性和普适性的 QoS 保障。改变了传统网络服务与业务需求“一对一”的紧耦合关系, 使由抽象与聚类得来

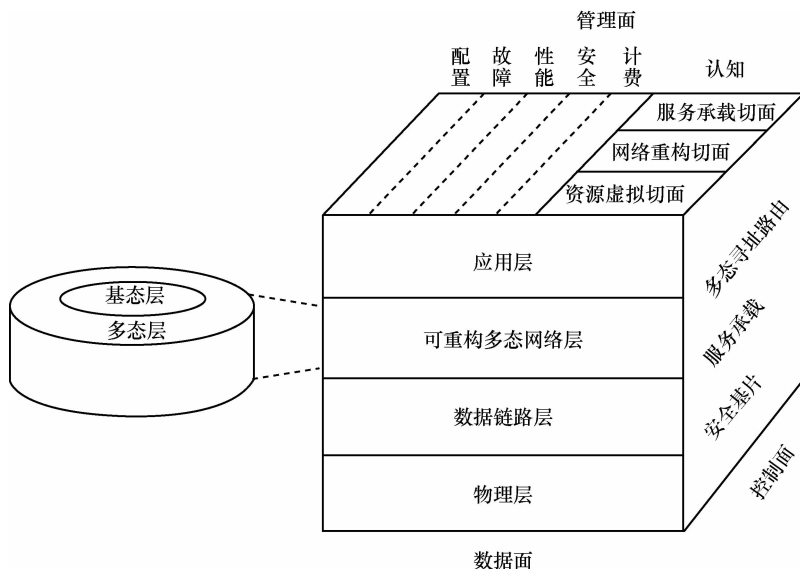


图9 可重构网络体系结构体系功能参考模型

的基本业务需求特征元素—元服务与网络基本承载功能组件—元能力进行动态匹配,使用普适的网络服务能力满足用户业务 QoS 保障需求。



图10 “业务—元服务—元能力”功能模型

在数据平面,有一个增强的网际互联传输层——可重构多态网络层,其功能设计目标是增强网络的基础互联传输能力,分为基态和多态两个子层。数据的传输模式有传统的具有网际互联传输能力的分组交换和宏电路两种,二者的内在逻辑关系如图 11 所示。宏电路是网络为一组具有共同传输路径的同类业务流动态建立的自适应型虚电路,是一种基于流的端到端的数据传输交换方式,增强和扩展了网络基础互联传输能力,对应于不同的数据流传输要求,宏电路有多种模式。

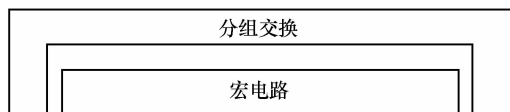


图11 可重构网络数据传递模式

宏电路中流的类别划分是根据特定 QoS 指标,如时延、抖动、分组丢失率等,将流量抽象聚合而成^[29]。借鉴了 ITU-T 对流量分类的方法,采用时延、抖动和分组丢失率这 3 个指标作为分类依据,根据网络认知功能获取网络的实时状况,动态调整流的分类粒度,以在保证流 QoS 保障的同时提高成功率。该方法的基本思想是将时延、抖动和分组丢失率划分成若干区间,即将 QoS 需求向量空间划分成若干子空间,每个子空间对应于一个类别。一条流的类别即该流 QoS 需求向量所处于子空间对应的流量类别。根据转发路径的链路带宽、节点 CPU 等资源利用状况及路径拥塞状况,实时调整流量类别划分粒度。

根据认知功能获得的底层流量视图,并由建立的流量预测模型,对多时间尺度上的流量行为进行预先估计和推断。据此,宏电路控制器控制底层转发节点动态地做出相应的构建动作,如建立、删除、规格调整。宏电路在边缘网络及不符合宏电路建立条件的链路上通过分组交换实现数据传输。在骨干网,按照网络业务流的性能要求,建立相应模式的宏电路。宏电路运行机制如图 12 所示。

可重构网络的核心特征是其内在结构的时变性,即由时变的结构驱动时变的服务能力,最终实现网络服务对应用要求和特征的动态匹配^[33]。由于可重构网络体系结构解耦了控制平面与数据平面,确保了可重构网络结构时变的可行性与易行性。

可重构网络在 QoS 保障方面具有先天的体系结构优势,转发控制分离架构增强了网络的可扩展性,网络元能

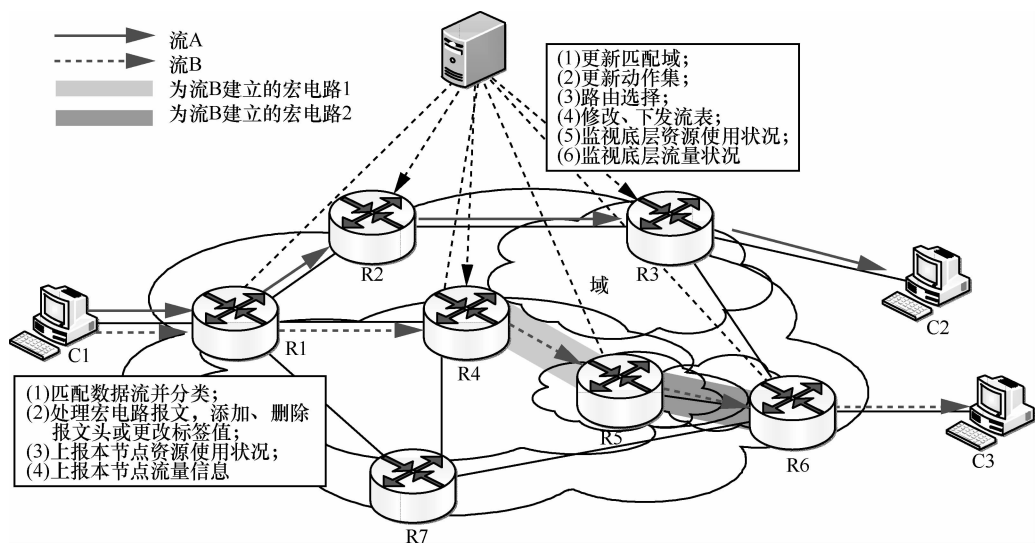


图 12 宏电路运行机制

力理论弥补了多样多变的业务特征需求与有限的网络服务能力之间的差异性,适应未来业务发展的延续性,面向连接的宏电路传输方式使得业务流可获得天然优秀的传输服务。集中式的控制架构使得流传输可在全网范围内进行优化,多级流表的配置使得流传输可在节点处进行调度,即可在宏观、微观两个层次上部署策略,实现按需的精确转发、传输。

4 可重构网络中宏电路的实验网部署研究

可重构网络以其柔性可变的结构形态和动态的功能组合,支持业务服务的灵活扩展、可管可控,对现有及未来业务具有优秀的 QoS 保障能力,获得了广泛的关注与研究,目前可重构实验网已搭建完毕^[1],由 5 个可重构路由交换平台和 5 个可重构光交换设备组成。宏电路控制平面位于可重构综合管理平台,控制服务器位于 Linux 系统上,基于新增宏电

路功能模块而扩展的 Floodlight 实现,宏电路的一切功能都运行于控制服务器之上。底层转发节点基于 NetFPGA 实现。

整个系统包括数据分组处理模块、资源监测与分配模块、流量监测与预测模块、路由选择模块、流表处理与下发模块。数据流在宏电路边节点处由数据分组处理模块首先判断是否为宏电路报文,并根据 QoS 参数分类,在链路层和网络层之间添加、删除标签;资源监测与分配模块实时监测底层资源使用状况,形成资源状态图,并在路径选择后沿途分配资源;流量监测与预测模块实时监测底层流量状况,并根据预测模型推测、估计未来一定时间内的流量行为趋势,生成流视图;路由选择模块根据流类别、资源状态图和流视图优化计算得到转发路径;流表处理与下发模块根据之前各个模块的处理结果生成或修改宏电路流表并下发至底层交换机。数据报文匹配流表并执行转发动作。系统设计结构如图 13 所示。

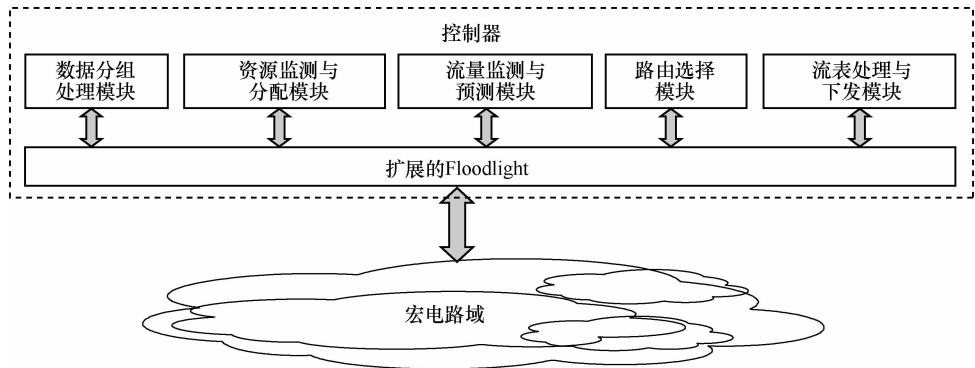


图 13 宏电路系统设计结构



5 结束语

结构决定功能,可重构网络体系架构改变了传统 IP 网络分层协作、端到端分组传输的设计原则,实现了控制平面与转发平面解耦合,使得复杂的控制逻辑加入不影响标准化的数据转发。数据流的传输使用宏电路与分组交换,在要求的路径使用宏电路可直接增强网络的基础传输能力,实现业务传输 QoS 保障。因此,可重构网络体系结构及宏电路是一种未来网络实现 QoS 自适应匹配、新型业务可扩展的有效解决方案。

参考文献

- 1 Feamster N, Gao L X, Rexford J. How to lease the internet in your spare time. ACM SIGCOMM Communication Review, 2007, 37(1): 61~64
- 2 Cisco System. Visual networking index(VNI) forecast. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.pdf, 2012
- 3 Aurecochea C, Campbell A T, Hauw L. A survey of QoS architectures. Multimedia Systems, 1998, 6(3):138~151
- 4 NSF NeTS FIND Initiative. Future internet design. <http://www.nets-find.net/>, 2005
- 5 Elliott C. GENI: opening up new classes of experiments in global networking. IEEE Internet Computing, 2010, 14(1): 39~42
- 6 EU FP7 ICT Challenge One. http://cordis.europa.eu/fp7/ict-programme/challenge_en.html, 2005
- 7 Gavras A, Karila A, Fdida S, *et al.* Future internet research and experimentation: the FIRE initiative. ACM SIGCOMM Computer Communication Review, 2007, 37(3): 89~92
- 8 The European Community's Seventh Programme. The FP7 4WARD project. <http://www.4ward-project.eu/>, 2007
- 9 The European Community's Seventh Programme. The FP7 autonomic internet project. <http://ist-autoi.eu/autoi/>, 2007
- 10 National Institute of Information (NICT). AKARI project. <http://akari-project.nict.go.jp>, 2007
- 11 兰巨龙, 程东年, 胡宇翔. 可重构信息通信基础网络体系研究. 通信学报, 2014, 35(1):128~139
- 12 吴建平, 林嵩, 徐格等. 可演进的新一代互联网体系结构研究进展. 计算机学报, 2012, 35(6):1094~1108
- 13 张宏科, 罗洪斌. 智慧协同网络体系基础研究进展. 电子学报, 2013, 41(7):1249~1254
- 14 Saltzer H, Reed D P, Clark D. End to end argument in system design. ACM Transactions on Computing System, 1984, 2 (4): 277~288
- 15 Sushant J, Alok K, Subhasree M, *et al.* B4: experience with a globally-deployed software defined WAN. Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM, Hong Kong, China, 2013: 3~14
- 16 Braden R, Clark D, Shenker S. Integrated services in the internet architecture: an overview. <http://www.rfc-editor.org/rfc/rfc1633.txt>, 1994
- 17 Flavius P, Ferdi P. A survey on the evolution of RSVP. IEEE Communications Surveys & Tutorials, 2013, 15(4):1859~1887
- 18 Shenker S, Partridge C, Guerin R. Specification of guaranteed quality of service. <http://www.rfc-editor.org/rfc/rfc2212.txt>, 1997
- 19 Wroclawski J. Specification of the controlled-load network element service. <http://www.rfc-editor.org/rfc/rfc2211.txt>, 1997
- 20 Lukasz B, Villy B I, Georgios K. Integrated service resource reservation using queueing networks theory. IET Networks, 2014, 3(1):16~21
- 21 Irena A, Charlie K, Radia P. Scalable per-flow resource management for large hierarchical networks. Proceedings of the IEEE 13th International Conference on High Performance Switching and Routing, Belgrade, Serbia, 2012: 160~167
- 22 Blake S, Black D, Carlson M, *et al.* An architecture for differentiated services. <http://www.rfc-editor.org/rfc/rfc2475.txt>, 1998
- 23 Nichols K, Black D L, Blake S, *et al.* Definition of the differentiated services field (DS field) in the IPv4 and IPv6 headers. <http://www.rfc-editor.org/rfc/rfc2474.txt>, 1998
- 24 Shigeo S. Fundamental trade-offs between resource separation and resource share for quality of service guarantees. IET Networks, 2014, 3(1):4~15
- 25 Flavius P, Ferdi P. A QoS enabled two-stage service differentiation model for the internet. Proceedings of the 6th International Conference on Signal Processing and Communication System, Gold Coast, QLD, 2012:1~10
- 26 Lu L Y, Du H F, Liu R P. CHOKeR: a novel AQM algorithm with proportional bandwidth allocation and TCP protection. IEEE Transactions on Industrial Informatics, 2014, 10(1):637~644
- 27 Geng X M, Luo A W, Sun Z J, *et al.* Markov chains based dynamic bandwidth allocation in DiffServ network. IEEE Communications Letters, 2012, 16(10):1711~1714
- 28 Xiao X P, Lionel M N. Internet QoS: a big picture. IEEE Network, 1999, 12(2):8~18
- 29 Khan J A, Alnuweiri H M. A fuzzy constraint-based routing algorithm for traffic engineering. Proceedings of the 2004 IEEE Global Telecommunications Conference(GLOBECOM'04), Dallas, TX, USA, 2004: 1366~1372
- 30 Hodzic H, Secerbegovic E. Online constraint-based routing as support for MPLS traffic engineering. Proceedings of the 51st International Symposium ELMAR-2009, Zadar, Yugoslavia, 2009: 127~130

- 31 Rosen E, Viswanathan A, Callon R. Multiprotocol label switching architecture. <http://www.rfc-editor.org/rfc/rfc3031.txt>, 2001
- 32 Awduche D, Chiu A, Elwalid A, *et al.* Overview and principles of internet traffic engineering. <http://www.rfc-editor.org/rfc/rfc3272.txt>, 2002
- 33 Awduche D, Malcolm J, Agogbua J, *et al.* Requirements for traffic engineering over MPLS. <http://www.rfc-editor.org/rfc/rfc2702.txt>, 1999
- 34 Asif M, Farid Z, Lal M, *et al.* Analysis of the similarities and differences between MPLS label distribution protocols RSVP and CR-LDP. *International Journal of Computer Science*, 2012, 54(9): 48~55
- 35 Swallow G, Aggarwal R, Ali Z. Non-penultimate hop popping behavior and out-of-band mapping for RSVP-TE label switched paths. <http://www.rfc-editor.org/rfc/rfc6511.txt>, 2012
- 36 Faueheur F L, Lai W. Requirement for Support of Differentiated Services-Aware MPLS Traffic Engineering. IETF RFC3564, 2003
- 37 Chen C L. A proposal of next generation network: QoS mapping for MPLS-DiffServ and label forwarding. *Proceedings of the 5th International Conference on BioMedical Engineering and Informatics (BMEI 2012)*, Chongqing, China, 2012:1416~1419
- 38 Francois F, Wang N, Moessner K, *et al.* Leveraging MPLS backup paths for distributed energy-aware traffic engineering. *IEEE Transactions on Network and Science Management*, 2014, 11(2): 235~249
- 39 Jain A, Chaudhari N S. Genetic algorithm for optimizing network load balance in MPLS network. *Proceedings of the Fourth International Conference on Computational Intelligence and Communication Networks*, Mathura, India, 2012: 122~126
- 40 Haas R. Forwarding and control element separation (ForCES) MIB. <http://www.rfc-editor.org/rfc/rfc5813.txt>, 2010
- 41 Doria A, Haas R, Salim J H, *et al.* Forwarding and control element separation (ForCES) protocol specification. <http://www.rfc-editor.org/rfc/rfc5810.txt>, 2010
- 42 Yang L, Dantu R, Anderson T, *et al.* Forwarding and control element separation (ForCES) framework. <http://www.rfc-editor.org/rfc/rfc3746.txt>, 2004
- 43 柯辽原. ForCES 通道传输方案及性能模型研究. 浙江工商大学博士学位论文, 2011
- 44 郑安枕. 基于网络处理器的 ForCES 路由器中 QoS 技术实现研究. 浙江工商大学博士学位论文, 2006
- 45 Nick M, Tom A, Hari B, *et al.* OpenFlow: enabling innovation in campus networks. *ACM SIGCOMM Computer Communication Review*, 2008, 38(2):69~74
- 46 Phemius K, Bouet M. Monitoring latency with OpenFlow. *Proceedings of the 9th International Conference on Network and Service Management*, Zurich, Switzerland, 2013: 122~125
- 47 Khan A, Kiess W, Perez-Caparras D, *et al.* Quality-of-service (QoS) for virtual networks in OpenFlow MPLS transport networks. *Proceedings of the IEEE 2nd International Conference on Cloud Networking*, San Francisco, CA, USA, 2013:10~17
- 48 Egilmez H E, Dane S T, Bagci K T, *et al.* OpenQoS: an OpenFlow controller design for multimedia delivery with end-to-end quality of service over software-defined networks. *Proceedings of the 2012 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*, Hollywood, CA, USA, 2012:1~8
- 49 Owens H, Durresi A. Video over software-defined networking (VSDN). *Proceedings of the 16th International Conference on Network-Based Information Systems*, Gwangju, Korea, 2013:44~51
- 50 Sharafat A R, Das S, Parulkar G, *et al.* MPLS-TE and MPLS VPNS with OpenFlow. *Proceedings of ACM SIGCOMM 2011 Conference*, Toronto, ON, Canada, 2011: 452~453
- 51 程东年, 汪斌强, 王保进等. 网络结构自调整的柔性内涵初探. *通信学报*, 2012, 33(8):214~222
- 52 Sun L, Lan J L. A flow-view based dynamic QoS guarantee mechanism. *Proceedings of the 4th IEEE International Conference on Software Engineering and Service Science*, Beijing, China, 2013:882~885
- 53 兰巨龙, 邢池强, 胡宇翔等. 可重构技术与未来网络体系架构. *电信科学*, 2013, 29(8):16~23

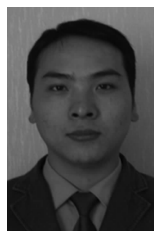
[作者简介]



孟飞,男,国家数字交换系统工程技术研究中心硕士生,主要研究方向为宽带信息网络。



兰巨龙,男,国家数字交换系统工程技术研究中心教授、博士生导师,国家重点基础研究发展计划(“973”计划)首席科学家,下一代广播电视网专项总体组专家,主要研究方向为宽带信息网络、网络与信息安全。



胡宇翔,男,国家数字交换系统工程技术研究中心讲师,主要研究方向为新型网络体系结构、网络安全等。

(收稿日期:2014-06-30)