

## 域间流量工程体系结构的研究\*

刘亚萍, 龚正虎, 卢泽新

(国防科技大学 计算机学院, 湖南 长沙 410073)

**摘要** :随着 Internet 的快速发展,为了提高网络整体资源的利用率,出现了通过跨域网络性能的优化,达到优化网络资源使用的域间流量工程技术。在分析与评价域间流量工程体系结构研究状况的基础上,提出了面向域间流量工程的 RCP 体系结构以及今后需要研究的主要问题。

**关键词** :域间流量工程;体系结构;路由优化

**中图分类号** :TP393.4 **文献标识码** :A

## Research on the Architecture of Inter-domain Traffic Engineering

LIU Ya-ping, GONG Zheng-hu, LU Ze-xin

(College of Computer, National Univ. of Defense Technology, Changsha 410073, China)

**Abstract** :With the rapid development of the Internet, for the benefit of improving the using rate of the entire network resources, the technique of inter-domain traffic engineering was proposed for the performance optimization of traffic that originates in one administrative domain and terminates in a different one. Based on the analysis and evaluation of current researches on the architecture of inter-domain traffic engineering, the RCP architecture for inter-domain traffic engineering was suggested. Finally, some main key problems were discussed for future researches.

**Key words** :inter-domain traffic engineering; architecture; route optimization

随着 Internet 的发展,IP 网络技术仍面临着一系列极具挑战性的问题,其中一个很重要的问题就是网络性能和资源利用率低下。由于现有 IP 路由方法以静态链路度量为基础计算端到端的最短路径作为传输通路,因此,ISP(internet service provider)不得不依靠过度供给方法来实现对用户的服务承诺。其结果是,当网络的某一部分发生拥塞时,其它部分却有大量链路资源闲置。为解决上述问题,很多 ISP 通过流量工程技术来实现网络流量的均衡。

流量工程是研究如何评价 IP 网络的性能以及优化 IP 网络性能的技术,它主要包括如何对 Internet 流量进行测量、识别、建模以及控制的研究。根据流量工程作用范围的不同,可分为域内流量工程与域间流量工程。域内流量工程是指在一个管理域内实施的流量工程,域间流量工程是在多个管理域实施的流量工程。域内流量工程技术的研究较为成熟,如基于 MPLS<sup>[1]</sup>的流量工程<sup>[2-3]</sup>,基于域内路由协议的链路度量调整的流量工程<sup>[4]</sup>。域间流量工程与域内流量工程相比,具有分布式、拓扑信息的不完全性、网络状态信息的不完全性、控制管理的难协商性、流量模式的动态性等特点,因此其研究难度较大。当前 Internet 上运行的域间路由协议 BGP 协议,其设计没有考虑流量负载的平衡性,这样就存在域间流量的优化需求与域间路由协议不支持间的矛盾。在域间流量工程的研究中,应首先考虑如何建立一个良好的体系结构。

### 1 域间流量工程体系结构的研究现状分析

根据管理域间的相互关系,域间流量工程的体系结构可以分成 4 类:完全自私的结构、基于相邻域协商的结构、基于覆盖网络的结构和逻辑集中式控制的结构。

\* 收稿日期:2005-12-01

基金项目:国家 973 资助项目(2003CB314802);国家自然科学基金资助项目(90104001)

作者简介:刘亚萍(1973-),女,副研究员。

### 1.1 完全自私的结构

这种结构是指不考虑对其它域的影响,完全从本 ISP 利益出发的一种结构。该结构的基本原理是由流量统计信息预测流量需求矩阵,根据 BGP 路由信息、拓扑与配置信息进行路由优化与路由控制,将计算出的新路由或新策略的相关参数发布给相应的 BGP 路由器,从而达到出口流量的平衡。这种结构与当前 ISP 使用 BGP 协议的模式一致。

属于该类结构的研究有 Steve Uhlig 等人提出的针对 STUB AS 的出口流量平衡的域间流量工程的体系结构<sup>[5-6]</sup>、基于服务器的出口流量平衡的域间流量工程的体系结构<sup>[7]</sup>和应用于商业上的路由优化技术<sup>[8]</sup>。图 1 是文献[6]中体系结构的示意图, $R_1$ 、 $R_2$ 、 $R_3$  是 BGP 边界路由器,它向带优化工具的路由反射器 RR 提供 BGP 路由信息、配置信息、流量的统计测量信息;RR 将计算出的新路由策略的相关参数通过 iBGP 发布给相应的 BGP 边界路由器  $R_1$ 、 $R_2$ 、 $R_3$ (如修改 local preference),从而达到出口流量的平衡。

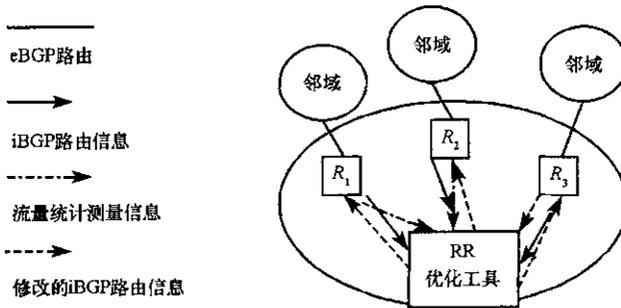


图 1 STUB AS 的出口流量平衡的域间流量工程的体系结构

Fig.1 Architecture of inter-domain traffic engineering for outgoing traffic of STUB AS

### 1.2 基于相邻域协商的体系结构

该结构的主要原理是,研究发现一个 ISP 如果仅根据本地信息进行路由优化,那么它通常会损坏其它域的利益,达到全局优化几乎是不可能的<sup>[9]</sup>,所以通过在相邻域间显示地协商流的路径来达到流量分布的均衡性以及 ISP 间利益双赢的局面<sup>[9-10]</sup>。以文献[9]提出的基于相邻域协商的结构为例,其基本思想是每个 ISP 都有一个路由优化器,它根据 BGP 边界路由提供的路由信息、配置信息、流量信息,给出希望调整流的路径推荐表,推荐表中包括对各条路径的喜好程度的评价。通过交换路径选择推荐表,提出优化的主动方在考虑对方利益的前提下给出路径选择并交给被动方,当双方均同意,才进行参数配置,调整路由。

### 1.3 基于覆盖网络的体系结构

该结构的主要原理是通过在 Internet 上建立应用层的一个小规模的网络,获取该覆盖网络的全部拓扑结构,将该覆盖网络可能涉及的域间链路的流量控制转换为传统的域内流量工程,从而达到负载平衡。例如弹性覆盖网络体系结构(RON)<sup>[11]</sup>、虚拟多宿主体系结构<sup>[12]</sup>均属于该类结构。以虚拟多宿主体系结构为例,其基本思想是每个 AS 有一个或多个多宿主服务器(MHS),它通过 iBGP 接收该域内所有 eBGP 路由器的路由信息,每个 AS 间的 MHS 构成一个多宿主覆盖网络,MHS 之间运行链路状态路由协议分发路由信息。作为服务器方的 MHS 要向客户方的 MHS 传送路由信息,但是客户方的 MHS 只接收路由,不广播路由信息。由源 MHS 根据目标地址选择目标 MHS。如图 2 所示,图中的 MHS 均是服务器方,实线表示域间链路,虚线表示 MHS 之间构成了一个覆盖网络。

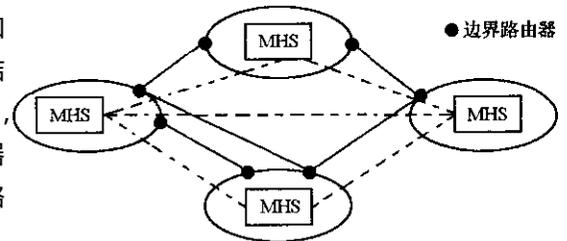


图 2 虚拟多宿主体系结构

Fig.2 Virtual multi-homing architecture

由于每个源 MHS 保存了到一个目的地址的跨域多路径以及相应的详细拓扑信息,因此,可以通过实施域内流量工程技术来实现端到端的服务保证,从而间接达到域间流量(该覆盖网络涉及的域间链路)和域内流量的

负载均衡。

### 1.4 逻辑集中式控制的体系结构

逻辑集中式控制的体系结构的基本原理是将路由优化控制的计算、协商、甚至进行路由计算的过程等从传统的路由协议中分离出来,它形成了一个更为高级的控制平面与控制协议。域间路由由协议还是集中于完成其主要工作,传递可达信息。通过路由控制平面计算出的结果来控制、调整路由协议的参数,然后再将其调整的结果反映到转发平面上,从而达到控制调整路由的目的。例如 OPCA(overlay policy control architecture)体系结构<sup>[13]</sup>、路由控制剖面 RCP(route control platform)体系结构<sup>[14-15]</sup>等均属于该类结构。以 RCP 为例,每个 AS 都有一个逻辑集中控制的路由控制剖面点 RCP,RCP 代表整个域与相邻域的 RCP 交换路由策略信息,并拥有整个域的域内拓扑结构,如图 3 所示。其基本思想是 RCP 代表整个域的所有路由器进行路由选择,从而防止内部 BGP 路由协议的不一致性;RCP 在整个网络范围内进行路径选择和策略设置,能支持方便实施域间流量工程;RCP 可实现域间交换路由而不是路由器间交换路由;RCP 通过 iBGP 协议控制域内每个路由器的配置和路由表的修改。

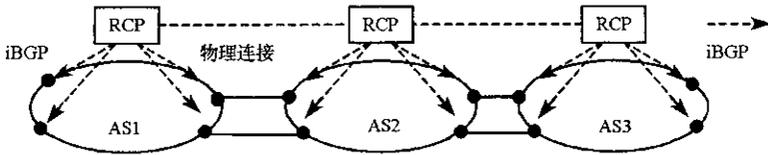


图 3 RCP 体系结构  
Fig.3 RCP Architecture

上述四类体系结构各有优缺点,其比较见表 1。其中,完全自私的结构会引发路由策略冲突、路由振荡等重大问题;基于相邻域协商结构会泄漏 AS 内部信息,覆盖网络结构易与 ISP 的利益相冲突,难以实施;逻辑集中式体系结构适合于流量工程技术的实施,但是 OPCA、RCP 结构都没有描述如何具体支持流量工程。

表 1 域间流量工程体系结构比较

Tab.1 Comparison between different inter-domain traffic engineering architectures

体系结构	路由策略的不一致性	协议振荡	配置错误	全局优化	安全性	域间流量工程	ISP 的利益	路由变化的实时性	BGP 协议的修改
完全自私结构	易导致	易导致	易导致	不支持	好	支持且有具体机制	吻合	较好	较少
基于相邻域协商结构	不易导致	不易导致	不易导致	支持	不好	支持且有具体机制	不易吻合	不好	较多
基于覆盖网络结构	不易导致	不易导致	易导致	支持小范围优化	不好	部分支持且有具体机制	不吻合	较好	较多
逻辑集中式控制结构	不易导致	不易导致	不易导致	支持	较好	支持但无具体机制	吻合	较好	较少

## 2 面向流量工程的 RCP

由于逻辑集中式体系结构较适合于流量工程技术的实施,因此我们提出面向流量工程的 RCP 结构。面向流量工程的 RCP 结构主要解决两个问题 (1)如何建立一种自主与协商平衡的架构,保证协商机制是激励兼容的、可扩展的、公平的、安全的和高性能的 (2)如何建立与 RCP 结构相容的面向流量工程的路由优化的工作模式。

### 2.1 面向流量工程的 RCP 体系结构描述

RCP 结构具体包括三部分:域内拓扑结构观察器、BGP 引擎以及路由控制服务器<sup>[15]</sup>。为支持流量工程,RCP 需要增加 4 个功能模块:流量信息模块、本地策略库模块、AS 拓扑关系推理模块、AS 协商模块,同时路由控制服务器模块需要增加基于流量平衡的路由优化功能。具体如图 4 所示。其中流量信息模块包括缺省配置的流量需求信息、上次策略调整的流量信息、当前的流量需求信息、瓶颈信息库,用

以预测流量需求的变化。本地策略库模块包含当前 AS 所实施的策略库(包括每个路由器相关的配置信息)与相邻域间的 SLA (service level agreement)。AS 拓扑关系推理模块包括分析、推理 AS 间关系,从而决定与哪些 AS 协商。AS 协商模块包括与一个 AS 间的协商和与多个 AS 间的协商。IGP viewer 通过与一个或多个路由器建立邻接关系,来获取 IGP 的拓扑信息。BGP 引擎通过运行 iBGP,学习 BGP 路由信息,并负责分发优化的路由。RCS 是控制中心,它代表 AS 进行路由选择的过程,包括 3 个子过程 (1)根据初始流量需求、SLA、AS 间关系等求出最优的路由策略设置,形成本地策略库 (2)根据当前策略,较实时地进行路由选择过程 (3)在较长时间范围内进行路由优化选择的调整。

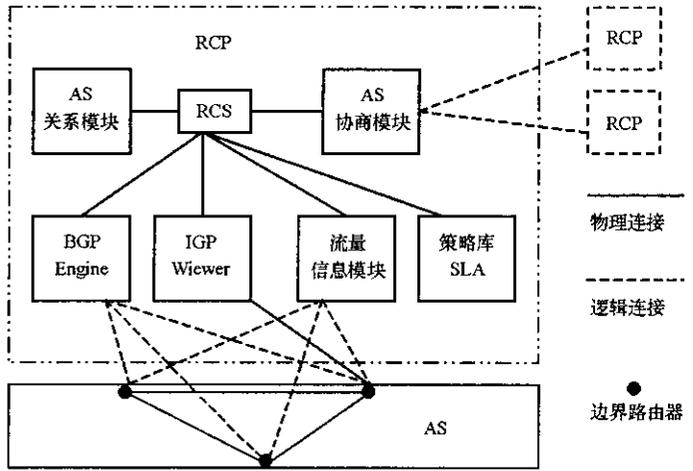


图 4 面向流量工程的 RCP 结构

Fig.4 Traffic engineering oriented RCP architecture

实施流量工程主要集中于子过程(1)和(3),子过程(1)是一种静态的行为,所以相对简单,难点主要集中于子过程(3)。在长时间的粒度<sup>[5-6]</sup>下,适合使用流量工程进行路由优化。有两种模式触发长时间粒度的优化路由:一是周期性触发,如以天、星期、月为周期;二是事件触发,如当出现较严重拥塞以及流量模式的显著变化等。然后 RCS 根据 BGP 引擎学到的路由、IGP viewer 学到的域内拓扑结构、当前的流量信息、SLA、AS 关系等进行最优路由的调整。为了降低路由优化调整所带来的副作用,一般先选择出口链路是同一个域的链路进行调整,启动相邻域间的协商,如果最终协商成功,就实施调整。如果需要多个域间的链路进行调整,则启动多个域的协商。

### 2.2 关键问题

在面向流量工程的 RCP 体系结构中,需要重点研究三类关键问题 (1)RCS 路由优化选择问题; (2)AS 协商模块问题 (3)实现及其性能分析。

#### 2.2.1 RCS 路由优化选择问题

RCS 路由优化选择实际是该体系结构的核心,因此需要研究如下问题:

(1)RCS 路由优化选择可以分为 3 个子过程,那么 3 个子过程间的关系是什么?如何使三者成为统一的整体?需要重点研究如何设置路由优化的频度和路由优化选择算法的运行条件。

(2)3 个子过程涉及的具体 BGP 多出口路径选择问题是 RCS 要解决的核心问题。RCS 路由优化问题是指如何优化路由控制,使得流量分布更加均衡。如何设计一个基于流量平衡的路由优化选择算法是研究的重点问题。

(3)BGP 路由选择会影响流量矩阵以及域内流量工程,而域内拓扑结构也可能影响路由优化选择,因此需要研究域内流量工程与域间流量工程的关系问题。

(4)由于引入了 AS 间的协商机制,因此,RCS 路由优化选择问题还需要研究如何解决自身利益、其它域的利益、与其它域的关系(如竞争关系)以及全局社会利益间的矛盾关系。

(5)为提高系统的鲁棒性,需要研究在一个 AS 中如何部署多个 RCS。

#### 2.2.2 AS 协商模块相关问题

引入 AS 协商模块的目的是用于实现社会利益最大化。为此,需要解决如下关键问题:

##### (1)AS 间协商采用的机制

如果采用中心控制仲裁机制,并且如果中心控制点足够安全,可以保证协商过程中 AS 间相互不泄漏策略和内部信息,从而易保证 AS 间收益的公平性,缺点是将安全信任转移到中心控制仲裁点,同时集中控制缺乏可扩展性。如果采用分布式机制,那么需要设计各 AS 间的通信协议。尽管理论上可以

采用加密手段来实现分布式多实体系统的私有性问题<sup>[16]</sup>,但是如何设计协议来保证不会泄漏 AS 本身策略以及 AS 间的公平性仍是研究的难点。

#### (2) 协商的激励机制

如何保证 AS 进行域间协商的动力,是该机制是否可行的关键问题。

#### (3) 公平问题

如何保证 AS 的协商过程是公平的。

#### (4) 性能问题

协商会带来通信开销,需要研究如何降低协商的开销,使得系统的性能价格比可行。

#### (5) 协议的设计问题

如何设计协商协议,满足协商需求。

#### 2.2.3 实现及其性能分析

实现原型系统,并根据模拟器,或者直接采用当前 Internet 上的真实数据进行实验,分析与评价该体系结构下路由的收敛型、优化结果的性能以及可扩展性、可控制性等是验证该体系结构可行性的关键步骤。

### 3 总结与下一步的工作

本文从体系结构的角度,分析与评价了近年来域间流量工程的主要研究。分析表明,逻辑集中式的域间路由由体系结构适合进行域间流量工程技术的实施。因此本文在 RCP 的基础上,提出了基于域间流量工程的 RCP 结构,并描述了该结构的组成与基本工作原理。在此基础上,提出了需要研究与解决的关键问题。下一步的工作重点是(1)建立原型系统,为验证体系结构的思想打好基础(2)重点解决 AS 协商模块中协议与机制的设计(3)重点研究 BGP 多出口路径选择问题。

### 参考文献:

- [1] Awduche D. MPLS and Traffic Engineering in IP Networks[J]. IEEE Communications Magazine, 1999, 37(12): 42-47.
- [2] Elwalid A, Jin C, Low S, et al. MATE: MPLS Adaptive Traffic Engineering[A]. In: INFOCOM '01[C], April 2001.
- [3] Kar K, Kodialam M, Lakshman T V. Minimum Interference Routing of Bandwidth Guaranteed Tunnels with MPLS Traffic Engineering Applications[J]. IEEE JSAC 2000, 18(12).
- [4] Fortz B, Thorup M. Internet Traffic Engineering by Optimizing OSPF Weights[A]. In: INFOCOM '00[C], 2000: 519-528.
- [5] Uhlig S. Implications of Characteristics on Interdomain Traffic Engineering[D]. Phd Thesis, University Catholique de Louvain, 2004.
- [6] Bonaventure O, Uhlig S, Quoitin B. The Case for More Versatile BGP Route Reflectors[S]. Internet Draft draft-bonaventure-bgp-route-reflectors-00.txt, July 2004.
- [7] Ye T, Kalyanaraman S. A Recursive Random Search Algorithm for Black-box Optimization[J]. ACM SIGMETRICS Performance Evaluation Review, December 2004.
- [8] Allen D. NPN: Multihoming and Route Optimization: Finding the Best Way Home[J]. Network Magazine, 2000(6).
- [9] Mahajan R, Wetherall D, Anderson T. Interdomain Routing with Negotiation[R]. Tech. Rep. CSE-04-06-02, University of Washington, May 2004.
- [10] Winik J, Jamin S, Rexford J. Traffic Engineering between Neighboring Domains[EB/OL]. <http://www.research.att.com/~jrex/papers/interAS.pdf>, July 2002.
- [11] Andersen D G, Balakrishnan H, Kaashofk M F, et al. Resilient Overlay Networks[A]. In: 18<sup>th</sup> ACM SOSP[C], Banff, Canada, 2001: 131-145.
- [12] Li Z, Mohapatra P, Chuah C N. Virtual Multi-homing: On the Feasibility of Combining Overlay Routing with BGP Routing[R]. UC Davis ECS Technical Report, March 2005.
- [13] Agarwal S, Chuah C, Katz R. OPCA: Robust Interdomain Policy Routing and Traffic Control[A]. In: 6<sup>th</sup> International Conference on Open Architecture and Network Programming[C], IEEE OpenArch, April 2003.
- [14] Feamster N, Balakrishnan H, Rexford J, et al. The Case for Separating Routing from Routers[J]. In: ACM SIGCOMM Workshop on Future Directions in Network Architecture, August 2004.
- [15] Caesar M, Caldwell D, Feamster N, et al. Design and Implementation of a Routing Control Platform[A]. In: 2<sup>nd</sup> Symposium on Networked Systems Design and Implementation[C], Boston, MA, May 2005.
- [16] Machiraju S, Katz R H. Verifying Global Invariants in Multi-provider Distributed Systems[A]. In: Proceedings of Hotnets Workshop[C], 2004.



