在 BGP/MPLS VPNs 中用 BGP 实现域间流量工程*)

梁海英[1,2] 李 政[2] 高 远[1]

(东北大学信息科学与工程学院 沈阳 110004)[1] (吉林师范大学计算机学院 四平 136000)[2]

摘 要 在 BGP/MPLS VPNs 中,用 MPLS 实现的流量工程主要被限止在单个管理域内。然而,随着企业规模的不断扩大,VPN 跨越越来越多的管理域,急需管理域间流量的有效方法。以 BGP 属性、BGP 策略和 AS 关系为基础的,一方面通过配置 LOCAL-PREF 属性值,运用输入策略,控制 AS 的出界流量;另一方面,保证客户 AS 不在提供者间或对等体间过渡流量,或允许客户 AS 向它的部分提供者通告路由,或人为增长 AS_PATH,控制 AS 的入界流量。仿真表明此方法能有效地在 BGP/ MPLS VPNs 中用 BGP 实现域间流量工程。

关键词 边界网关协议,多协议标签交换,虚拟专用网,流量工程,自治系统

Implementing Inter-domain Traffic Engineering with BGP in BGP/MPLS VPNs

LIANG Hai-Ying[1,2] LI Zheng[2] GAO Yuan[1]

(School of Information Science and Engineering, Northeastern University, Shenyang 110004)[1]

(School of Computer, Jilin Normal University, Siping 136000)[2]

Abstract In BGP/MPLS VPNs, traffic engineering achieved by using MPLS was predominantly limited to intra-domain and single administrative domain. However with the rapid expansion of enterprise scale, a VPN has spanned a large number of administrative domains. Thus effective management of inter-domain traffic is urgently demanded. Presented the simplest method is on the basis of various BGP attributes, BGP import or export routing policy and AS relationship constructed according to bilateral economical agreements. To implement inter-domain traffic engineering, the best route in an AS depends on the routes coming from its neighboring ASes that apply export routing policies, as well as the import routing policies of the AS. The control of the outgoing traffic is often a requirement for providers that wish to optimize the distribution of their content. For this, they can rely on the LOCAL_PREF attribute to control the routes that will be chosen for the packets that leave each BGP router of the provider. A customer AS serving a large number of individual users or small corporate networks will typically have a very asymmetric inter-domain traffic pattern with several times more incoming than outgoing traffic. These ASes typically need to optimize their incoming traffic only. In order to balancing incoming traffic, ASes are allowed announce their prefixes to a selected subset of providers instead of all providers. According to BGP export policies, an AS cannot act as a transitive AS for its two providers or peers. Simulation shows that our approach can effectively implement inter-domain traffic engineering in BGP/MPLS VPNs.

Keywords BGP, MPLS, VPN, Traffic engineering, AS

1 引言

VPN(Virtual Private Network ,虚拟专用网)是指使用和专用网络一样的访问技术和安全策略,将客户通过公共设施连接在一起的网络。RFC2547[1]定义了建立 VPN 的一种方法(即 BGP/MPLS VPNs)。在这个方法中,BGP(Border Gateway Protocol,边界网关协议)[2]用于传送 VPN 路由信息,MPLS(Multiprotocol Label Switching,多协议标签交换)[3]用于传送 VPN 流量。

Internet 由大量的 AS (Autonomous System,自治系统)构成。AS 也被叫做管理域,它可以使用多种域内路由协议,如 OSPF(Open Shortest Path First,开放式最短路径优先),IS-IS(Intermediate System-Intermediate System,中间系统到中间系统)和 MPLS,但是,在 BGP/MPLS VPNs 中,用 MPLS 实现的流量工程主要被限止在管理域内和仅有单个管理域的情况[4]。然而,随着企业规模的不断扩大,VPN 跨越越来越多的管理域,急需对域间流量进行有效管理的方法。

本文介绍了一种用 BGP 实现域间流量管理的方法,该方法以各种各样的 BGP 属性、BGP 输入输出策略和通过商业合同建立的 AS 关系为基础,运用不同的路由策略选择来自其它 AS 的合适路由或向相应 AS 传播选择的路由可达信息。这些方法是通过配置 LOCAL-PREF 属性值,运用输入策略,控制 AS 的出界流量,或者保证客户 AS 不在提供者间或对等体间过渡流量,或允许客户 AS 向它的部分提供者通告路由,或人为增长 AS_PATH,控制 AS 的入界流量,从而实现流量工程。

本文组织如下:第 2 节对 BGP/MPLS VPNs 做一简要介绍;第 3 节介绍 BGP 属性和策略;第 4 节详细阐述用 BGP 在

*)国家自然科学基金资助(Internet 故障管理和路由稳定性研究,批准号:60073059;Internet 域间路由稳定性和可管理性的研究,批准号:60273078)。梁海英 副教授,博士研究生;李 政 教授,硕士生导师;高 远 教授,博士生导师。

BGP/MPLS VPNs 中实现域间流量工程的方法；第 5 节介绍用 SSFnet 进行仿真实验，并对结果进行分析；最后对全文的一个简单总结。

2 BGP/MPLS VPNs 简介

2.1 BGP/MPLS VPNs 构成

BGP/MPLS VPNs(如图 1 所示)由 3 个部件组成：CE(Customer Edge，客户边界)设备、PE(Provider Edge，提供者边界)路由器和 P(Provider，提供者)路由器。

CE 路由器是客户边界设备，它通过数据链路，如帧中继、ATM 或租用线路，将客户结点连接到一个或多个 PE 设备。CE 设备可以是主机或者是路由器，典型地是路由器，它和它直接连接的 PE 设备建立邻接。在此情况下，CE 路由器与 PE 设备之间通过静态路由、IGP(Interior Gateway Protocol，内部网关协议)、或 EBGP(External BGP，外部边界网关协议)，在 AS 间交换客户 VPN 的网络可达性信息。

PE 路由器是骨干网络的边缘设备，它直接连接到 CE 设备。PE 路由器首先从 CE 设备获得当地 VPN 路由，然后与其它的 PE 路由器用 IBGP(Internal BGP，内部边界网关协议)在 AS 内交换 VPN 路由信息。

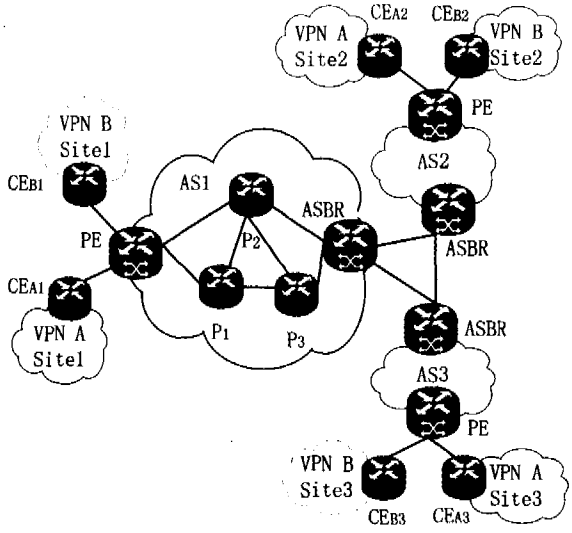

图 1 BGP/MPLS VPNs 的网络结构

P 路由器是服务提供者骨干网络的核心路由器，它不和 CE 设备相连。P 路由器作为 MPLS 的过渡 LSR(Label Switch Router，标签交换路由器)，基于 MPLS 标签，在 PE 路由器之间交换 VPN 的数据流量。P 路由器只需要维持到 PE 路由器的路由信息，而不需要维持到特定客户的 VPN 路由信息。

2.2 数据流和控制流

在 BGP/MPLS VPNS 中有两种流：数据流和控制流。控制流用于分发 VPN 路由和建立 LSP(Label Switched Path，标签交换路径)；数据流用于转发 VPN 客户的数据流量。

2.2.1 控制流

控制流由两个子流构成。

一个子控制流，负责在提供者骨干网边界交换路由信息以及在骨干网内部进行路由信息的分发。

在 BGP/MPLS VPN 中，分发 VPN 路由信息有 5 步：

第 1 步：从 CE 路由器到入口 PE 路由器分发。CE 路由器与入口 PE 路由器之间可以使用静态路由、IGP 或者 EBGP 来交换网络可达路由信息。

第 2 步：在入口 PE 路由器输出到提供者的 BGP。PE 路由器为每个直接连接的 VPN 维持一个 VRF(VPN Routing and Forwarding，VPN 路由转发)表，定义连接到这个 PE 路由器上的 VPN 客户成员关系。入口 PE 路由器从它直接连接的 CE 路由器获得路由后，将这些路由加上特定的 RT(Route Target，路由目标)属性，安装到相应的 VRF 表中，然后将此路由转换成 VPN-IPv4 路由，输出到提供者的 BGP。

第 3 步：穿过服务提供者骨干网络(即 VPN-IPv4 路由在入口 PE 路由器和出口 PE 路由器之间分发)。如果一个 VPN 的两个站连接到的 PE 路由器在同一 AS 内(如图 2 所示)，入口 PE 路由器能与出口 PE 路由器通过 IBGP 交换这些 VPN-IPv4 路由。另一方面，图 3 显示 VPN 站连接到的 PE 路由器在不同的 AS 中的情况。在这种情况下，入口 PE 路由器用 IBGP 向同一 AS 中的 ASBR(Autonomous System Border Routers，自治系统边界路由器)分发 VPN-IPv4 路由，然后，ASBR 用 EBGP 将这些路由分配到另一 AS 中的 ASBR，依此类推，直到最后一个 ASBR 用 IBGP 将这些路由分配给出口 PE 路由器为至。

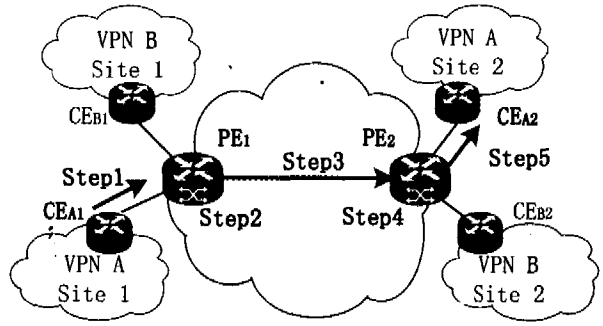

图 2 BGP/MPLs VPN 中路由分发(PE 路由器在同一 AS 中)

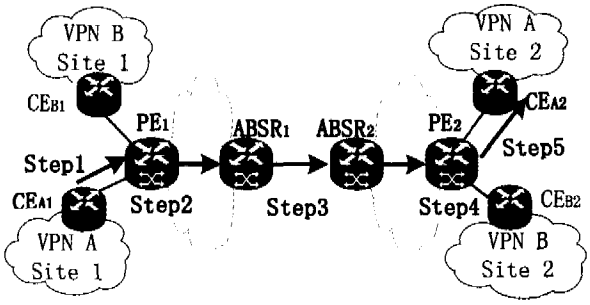

图 3 BGP/MPLs VPN 中路由分发(PE 路由器在不同 AS 中)

本文只讨论与域间流量相关的控制流在 ASBR 之间的分发。

第 4 步：在出口 PE 路由器从提供者 BGP 输出。一旦出口 PE 路由器从其它 PE 路由器或 ASBR 得到 VPN-IPv4 路由，它就为自己连接的 VPN 安装这些与 RT 属性相连的路由。

第 5 步：从出口 PE 路由器向 CE 路由器分发。最后，这些被选中的 VPN-IPv4 路由被转换回 IPv4 路由，并输入到相应的 VRF 表中。安装到 VRF 表中的路由均可以被分配给相应的 CE 路由器。

另一个子控制流是建立 LSP。为了使用 MPLS 转发 VPN 数据流量，使其经过骨干网络，必须在入口 PE 路由器和出口 PE 路由器之间建立 LSP。这些 LSP 可以通过 LDP

(Label Distribution Protocol,标签分发协议)、CR-LDP(Constraint-based Routing LDP 基于约束路由的 LDP)或 RSVP(Resource Reservation Protocol 资源预留协议)来建立,在 PE 和 PE 之间或 PE 和 ASBR 之间建立 LSP,以便使用 MLPS 沿此路径转发数据。提供者使用 LDP 在 PE 之间建立最努力的 LSP,也可以在需要给 LSP 分配带宽或使用流量工程时,使用 RSVP-TE 或 CR-LDP 来选择一条显示路径。

2.2.2 数据流

在 BGP/MPLS VPNs 中,使用两级标签。外层标签被分配给到出口 PE 路由器的路由,提供从入口 PE 路由器到出口 PE 路由器的转发。这些外层标签是由 LDP、CR-LDP 或 RSVP-TE 分发的。而内层标签控制在出口 PE 处的转发,它是由 BGP 和 VPN-IPv4 路由一起分发的。PE 路由器以及在多个 AS 之间对 VPN-IPv4 路由进行分配的 ASBR,需要把一个 4 字节地址前缀插入适用于骨干网络传输的 IGP 路由表中。这样,在骨干网络中的每一个结点,利用 MPLS 的 LDP,可以把一个与转发路径有关的标记分配到每一个 PE 路由器中。

入口 PE 路由器接收到来自 CE 的数据包时,它选择与这个 CE 相对应的 VRF,在此 VRF 中执行目的地址查找,用 MPLS 标签封装最佳匹配路由的 IP 包。两个标签被放到 VPN 包中,以便 VPN 包以隧道方式穿过骨干网络,到达相应的出口 PE 路由器。

P 路由器简单地基于外层标签交换 VPN 包。倒数第二个 LSR 在移交这个被标上标签的 VPN 包到目的 PE 路由器之前,弹出外层标签。

出口 PE 路由器接到带有内层标签的数据包,将数据包中的内层标签删除,然后将该数据包转发到相应的 CE 路由器中。

3 BGP 属性和策略

BGP 是在全球 Internet 范围内,在 AS 之间交换网络可达信息的域间路由协议。虽然 BGP 主要用在域间路由环境中,即 EBGP,但它也可以被 BGP 发言者用来在一个 AS 内交换外部网络可达信息,即 IBGP。

3.1 BGP 属性

每个 BGP 发言者路由器通过 BGP 会话,将更新消息发送给邻居。每个更新消息可以包括下列基本信息块。

(1)NLRI(Network Layer Reachability Information,网络层可达信息):以 IP 前缀路由的形式通告网络。NLRI 由一个或多个二元组〈长度,前缀〉的实例构成,长度是指一个特定前缀的掩码位数。

(2)路径属性:用来保持特定路由轨迹信息的参数的集合,包括路由优先级、下一跳以及聚合信息等。这些参数用于 BGP 过滤及路由决策处理。BGP 允许每一个 AS 按自己的路由策略接受和通告路由。这些策略是通过配置 BGP 属性实现的。BGP 路由属性描述如下:

NEXT_HOP:下一跳路由器地址(next hop);

AS_PATH:一个路由经历的 AS 路径的列表;AS PATH 属性是以{ASk, ASk-1,…, AS0}形式显示路由经过的 AS 序列,它能使 BGP 检查路由循环,加强当地的和全球的路由策略。在 AS_PATH 中最后一个 AS,即 AS0 是起源 AS,或者叫通告的起源。每个 AS 在接收到的路由的 AS_PATH 的前面加上自己的 AS 号后,再将此路由输出到它的邻居。

LOCAL_PREF:本地优先级(local preference);

MED:多出口鉴别(multi-exit discriminator)。

ORIGIN:路径信息的起源。取值为 0~2。

我们用 r. attribute 表示路由 r 的相应属性,如用 r. local pref 表示路由 r 的 LOCAL_PREF。用 r_{di} 表示 ASd 中的路由器从 ASi 中的路由器接收的路由。

(3)撤销路由:撤销路由是不可达路由的列表,它是不可用的或不再用的并需要从 BGP 路由表中撤销的路由。

BGP 路由信息库由下列三部分组成:

(1)Adj-RIBs-In:包含从其它 BGP 发言者接收的路由。

(2)Loc-RIB:包含当地 BGP 发言者将要使用的路由。

(3)Adj-RIBs-Out:包含通告给其它 BGP 发言者的路由。

3.2 BGP 输入路由策略

收到更新消息后,为了避免 AS_PATH 产生循环,路由器首先丢掉自己 AS 号出现在 AS_PATH 中的路由,然后依据基于 BGP 属性事先配置的输入策略,决定是使用这个路由还是拒绝这个路由,如果决定使用,还可以给这个路由分配 LOCAL_PREF 值,表示这个路由受欢迎的程度。

策略 1 对 $\forall i,j\in N$(N 是向 ASd 中路由器发送,到达同一目的地路由的,所有邻居 AS 中的路由器的集合),$\exists k\in N$,且 $i,j\neq d, i\neq j$。在 ASd 中的路由器根据 BGP 属性,从所有备选路由 r_{di} 中,计算出最佳路由 import(ASd←ASi)[{r_{di}}]的输入路由策略为[5,6]:

```
;选 LOCAL_PREF 最高的路由
if r_dk. loc_pref=max( r_di. loc_pref)and
  r_di. loc_pref≠r_dj. loc_pref then
    import(ASd←ASi)[{r_di}]=r_dk
;选择 AS_PATH 最短的路由
else if length(r_dk. as_path)=min(length(r_di. as_path))
      and r_di. as_path≠r_dj. as_path then
    I mport(ASd←ASi)[{r_di}]=r_dk
;选择 ORIGIN 最小的路由
else if r_dk. origin=min(r_di. origin)and
  r_di. next_hop≠r_dj. next_hop then
    import(ASd←ASi)[{r_di}]=r_dk
;选择 MED 最小的路由
else if r_dk. med =min(r_di. med)and
  r_di. med≠r_dj. med then
    import(ASd←ASi)[{r_di}]=r_dk
end if
```

注意:首先选择 LOCAL_PREF 值最高的路由,若 LOCAL_PREF 都相同,则选择 AS_PATH 最短的路由。因此,本文只讨论 LOCAL_PREF 和 AS_PATH 的设置对输入策略的影响。

3.3 BGP 输出路由策略

选择最佳路由后,还需依据输出策略决定是否将这个更新消息传递给邻居 AS。

输出路由策略包括允许或拒绝一个路由;分配 MED,或在 AS_PATH 的前面添加多个相同的 AS 号,控制入界流量。

3.3.1 AS 关系

大多数 AS 之间通过商业合同确定了 AS 关系。通常有以下几种类型[7]:提供者—客户(Provider-Customer),对等关系(Peering-Peering)。下面给出相应的定义。

定义 1 对∀i,j∈N,设 ASi 和 ASj 是两个自治系统。如果 ASi 按商业合同为 ASj 提供有偿服务,ASj 可以通过 ASi 访问其它网络,这时称 ASi 与 ASj 为提供者－客户关系(Provider-Customer);记为:ASi∈ASj. providers,ASj∈ASi. customers。

客户－提供者关系是最常见的一类关系,在 Internet 中主干网往往是地区网的服务提供者,即它们是提供者(主干网)－客户(地区网)关系。显然,一个提供者可以有多个客户,一个客户又可有多个提供者。

定义 2 对∀i,j∈N,设 ASi 和 ASj 是两个自治系统,如果 ASi 与 ASj 相互提供无偿的内部的访问服务,这时称 ASi 与 ASj 为对等体-对等体关系(Peering-Peering);记为:ASi∈ASj. peerings,ASi∈ASj. peerings。

对等关系也是常见的一类关系,在同属一个 ISP 的两个区域网络经常建立对等关系。

上述 AS 关系在 Internet 路由体系中不但确定了 AS 之间的价格模式,而且在一定程度上决定 AS 之间的路由策略。

3.3.2 基于 AS 关系的路由输出策略

策略 2 令 export(ASi→ASj)[{r_{ji}}]表示从 ASi 可以分发到 ASj 的路由集;r_{ipr}、r_{ic}、r_{ipe} 分别表示 ASi 从它的提供者、客户及对等体获得的路由集;r_i 表示 ASi 自己的路由集。则基于 AS 关系的输出路由策略原则为[8]:

;客户 AS 可以将它自己的路由和它客户的路由输出给它的提供者。

if ASj∈ASi. providers then

export(ASi→ASj)[{r_{ji}}]=$r_i \cup r_{ic}$

;提供者 AS 可以将它自己的路由、从它客户、提供者或对等体获得的路由输出给它的客户。

else if ASj∈ASi. customers then

export(ASi→ASj)[{r_{ji}}]=$r_i \cup r_{iC} \cup r_{ipr} \cup r_{ipe}$

;对等体 AS 可以将自己的路由和它客户的路由,输出给它的对等体。

else if ASj∈ASi. peerings then

export(ASi→ASj)[{rji}]=ri∪ric

end if

策略 2 的另一种描述:一个 AS 不提供在它的任意两个提供者和对等体之间的传输服务。若用 first(r. as_path)表示出现在 r. as_path 中的第一个 AS(即路由 r 最后经过的 AS),则一个 ASu 应根据以下规则指定它的输出策略,以便过滤向 ASv 发送的路由 r_{vu}。

if ASv∈ASu. providers∪ASu. peerings and

first(r_{vu}. as_path)∈ASu. providers∪ASu. peerings

then export(ASu→ASv)[{r_{vu}}]={}。

4 在 BGP/MPLS VPNs 中用 BGP 实现域间流量工程的方法

AS 中的最佳路由的选取,既依赖于邻居 AS 应用输出策略发送来的路由,又依赖于自身 AS 应用输入策略过滤的路由。

4.1 控制出界流量的方法

如果提供者希望优化流量的分发,那么它常常需要控制出界流量。

方法 1 在 AS 路由器上配置 LOCAL_PREF 的相应值,控制离开每个路由器的流量所选择的路由,实现出界流量的控制分发。

图 4 表示由 6 个 AS 构成的 BGP/MPLS VPN 的简化的骨干网结构。其中,每个 AS 只有一个 ASBR。

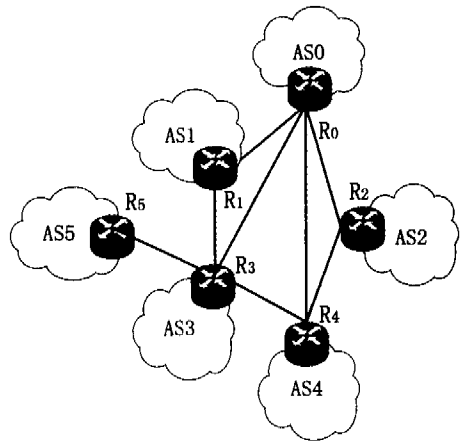

图 4 简化的 BGP/MPLS VPNss 的骨干网结构

下面,先看两种情况:

第一种情况是,对所有路由器均不改变任何 BGP 的属性。因此,各个路由器的各个链路上的 LOCAL_PREF 值都相同。按照输入路由策略,接下来,各个路由器应选择 AS_PATH 最短的路由作为到达目的地的最佳路由。考虑 AS0 中的路由器 R0,要将流量送到 AS4 中的 R4,它有 4 条可用路由,它们的 AS_PATH 分别为:AS_PATH 0 4,AS_PATH 0 2 4,AS_PATH 0 3 4 和 AS_PATH 0 1 3 4。依据策略 1,AS0 从中选取带有最短路径的 AS_PATH 0 4 的路由(命名为 AP1),从 AS0 的 R0 的一个相应的接口输出流量到目的地 AS4 中 R4。

第二个情况是,改变 LOCAL_PREF 的值。如果 R0 不想直接将流量输出到 AS4,而想经过 AS2 输出流量到 AS4,那么,它就可以将 r_{02}. local_pref 设置成四条链路中优先级最高的一个。根据策略 1,首先选取 LOCAL_PREF 值最高的带有 AS_PATH 0 2 4 的路由(命名为 AP2)而不选带有 AP1 的路由。

通过对比上述情况,可以看到,改变 LOCAL_PREF 可以实现出界流量的控制。

4.2 控制入界流量的方法

为大量的个人用户或小企业网络服务的客户 AS,在域间流量管理模式上非常不对称,处理的入界流量经常是出界流量的很多倍。这些 AS 就需要优化它们的入界流量。以下三种方法可以实现这一需求:

方法 2 依据基于 AS 关系的路由输出规则,保证客户 AS 不在提供者间或对等体间过渡流量。

回到图 4 所示的例子中,如果规定图中 6 个 AS 关系如下:

AS0. customers={AS1 AS2 AS4};

AS1. customers={AS3};

AS3. customers={AS4 AS5};

AS2. customers={AS4}。

那么,在配置路由输出策略时,应该考虑 AS4 的特殊地位,即 AS4 同时是 AS0,AS3 和 AS2 这三个 AS 的客户,所以它不能为它们过渡流量。例如:虽然 AS_PATH 2 0 3(命名为 AP3)和 AS_PATH 2 4 3(命名为 AP4)长度相等,但是依

据策略 2，AS2 将目的地属于 AS3 的数据包被迫经 AP3 发送而不经过 AP4 发送。

方法 3 为了平衡入界流量，允许 AS 将它们的地址前缀通告到部分提供者而不是全部提供者。这样，只有得到客户通告的提供者，才可以沿着通告的路径，将流量送入客户 AS 中；而没有得到客户通告的提供者就不能向这个客户 AS 直接注入流量，从而实现客户 AS 的入界流量控制。

还以图 4 为例，如果 AS3 为了平衡入界流量，将从 AS5 得到的路由只经过链路 R3-R0 通告给它的一个提供者 AS0。此时，AS3 的另一个提供者 AS1，没有得到 AS3 这个通告，所以，只能通过 AS_PATH 1 0 3 5 得到目的地属于 AS5 的路由通告。因此，AS1 不能从它与 AS3 的直接链路 R3-R1 向 AS1 发送目的地属于 AS5 的包。尽管 AS_PATH 1 0 3 5 比 AS_PATH 1 3 5 长，AS1 也不得不通过这个长的路径，向 AS1 发送目的地属于 AS5 的数据包。

方法 4 在不改变 LOCAL_PREF 值的情况下，允许 AS 通告各个路由时，通过人为增加某个路径上的 AS 号，使得这个路径增长，导致邻居 AS 不优选这个路径，从而控制这个路径上的入界流量。

对于上面的例子，AS1 将从 AS5 得到的路由从 R3-R0 和 R3-R1 两个链路分别通告给两个提供者 AS0 和 AS1，但是，在 R3-R1 链路上通告路由时，附带一个 AS_PATH 3 3 3 5 而不是 AS_PATH 3 5。因此，AS1 在分发目的地属于 AS5 的路由时，有两个路径 AS_PATH 1 0 3 5(命名为 AP5)和 AS_PATH 1 3 3 3 5(命名为 AP6)，依据策略 1，选择其中最短的路径 AP5。改变 AS PATH 属性的操作在实践中常被用到[9]。

5 仿真实验及结果分析

我们用 SSFnet[10] 来检验我们提出的用 BGP 在 BGP/MPLS VPNs 中实现域间流量工程的方法。

5.1 用 BGP 实现流量工程情况仿真

仿真拓扑结构如图 5 所示。图中只画出 BGP/MPLS VPNs 骨干网通过 ASBR 互连的情况。每个节点代表一个 AS，两个节点之间的边代表 BGP 会话，每个 AS 中只有一个 ASBR。

这 5 个 AS 之间的关系定义如下：

AS1. providers={AS0 AS3}；

AS2. providers={AS0 AS1}；

AS3. providers={AS0}；

AS4. providers={AS2 AS3}。

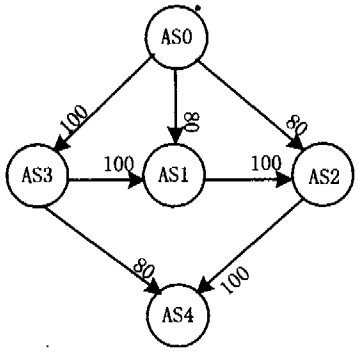

图 5 模拟拓扑结构(有策略)

实验 1 我们仿真的目标是，让 AS4 从众多路由中选择最长路径 AS_PATH 4 2 1 3 0 的路由，将其流量送往目的地 AS0，且不允许向 AS4 注入任何流量。

为了达到这个流量控制的目标，我们综合考虑输入策略和输出策略。

首先，配置各路由的 LOCAL_PREF 值如下：

r_{42}. local_pref=100；r_{43}. local_pref=80；

r_{21}. local_pref=100；r_{20}. local_pref=80；

r_{13}. local_pref=100；r_{10}. local_pref=80；

r_{30}. local_pref=100。

其次，保证一个客户 AS 不为任何两个提供者或两个对等体过渡流量。

最后，让 AS2、AS1 和 AS3 只向它们部分提供者 AS0 通告各自的前缀，而 AS4 不向任何提供者通告它的前缀。

通过上述配置后，用 SSFnet 仿真，得到的仿真结果如表 1 所示，该结果是用各 AS 中路由器的 Loc_RIB(包含当地路由器将要使用的路由)表示的。

表 1 仿真结果(有策略)

(a)Loc-RIB at bgp@0:1:

NetworkNHI	NextHopNHI	LocPrf	ASPathNHI
*>0	self	-	i
*>1	1:1(1)	-	1
*>3	3:1(1)	-	3
*>2	2:1(1)	-	2

(b)Loc-RIB at bgp@1:1:

NetworkNHI	NextHopNHI	LocPrf	ASPathNHI
*>0	3:1(2)	100	3 0
*>1	self	-	i
*>3	3:1(2)	100	3
*>2	3:1(2)	100	3 0 2

(c)Loc-RIB at bgp@2:1:

NetworkNHI	NextHopNHI	LocPrf	ASPathNHI
*>0	1:1(2)	100	1 3 0
*>1	1:1(2)	100	1
*>3	1:1(2)	100	1 3
*>2	self	-	i

(d)Loc-RIB at bgp@3:1:

NetworkNHI	NextHopNHI	LocPrf	ASPathNHI
*>0	0:1(3)	100	0
*>1	0:1(3)	100	0 1
*>3	self	-	i
*>2	0:1(3)	100	0 2

(e)Loc-RIB at bgp@4:1:

NetworkNHI	NextHopNHI	LocPrf	ASPathNHI
*>0	2:1(3)	100	2 1 3 0
*>1	2:1(3)	100	2 1
*>3	2:1(3)	100	2 1 3
*>2	2:1(3)	100	2
*>4	self	-	i

从表 1 中可以看出，AS4 确实选择 ASPathNHI 2 1 3 0 的路由到目的地 AS0，同时，没有任何路由可以到达 AS4。达到了我们预期的目标。

通过不同的配置，可以实现不同路由路径选择的要求，进而实现各路径上的流量分配。

实验 2 我们仿真的目标是，让 AS4 从众多路由中选择路径 AS_PATH 4 2 1 0 的路由，将其流量送往目的地 AS0，且不允许向 AS4 注入任何流量。

为此，在图5中只更改 r_{10}. local_pref 的值，由原来的80，改成100，就可以让 AS4 从众多的路由中选择路径 AS_PATH 4 2 1 0 的路由，将其流量送往目的地 AS0。

5.2 没实现流量工程的情况仿真

实验3 我们仿真的目标是，让 AS4 从众多路由中选择最短路径 AS_PATH 4 3 0 的路由，将其流量送往目的地 AS0，且允许向 AS4 注入任何流量。

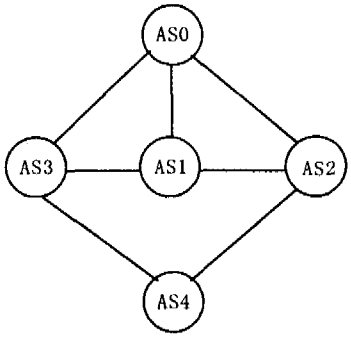

图6 模拟拓扑结构(无策略)

表2 仿真结果(无策略)

(a)Loc-RIB at bgp@0:1:

NetworkNHI	NextHopNHI	ASPathNHI
*>0	self	i
*>1	1:1(1)	1
*>3	3:1(1)	3
*>2	2:1(1)	2
*>4	3:1(1)	3 4

(b)Loc-RIB at bgp@1:1:

NetworkNHI	NextHopNHI	ASPathNHI
*>0	0:1(1)	0
*>1	self	i
*>3	3:1(2)	3
*>2	2:1(2)	2
*>4	3:1(2)	3 4

(c)Loc-RIB at bgp@2:1:

NetworkNHI	NextHopNHI	ASPathNHI
*>0	0:1(2)	0
*>1	1:1(2)	1
*>3	4:1(2)	4 3
*>2	self	i
*>4	4:1(2)	4

(d)Loc-RIB at bgp@3:1:

NetworkNHI	NextHopNHI	ASPathNHI
*>0	0:1(3)	0
*>1	1:1(3)	1
*>3	self	i
*>2	4:1(1)	4 2
*>4	4:1(1)	4

(e)Loc-RIB at bgp@4:1:

NetworkNHI	NextHopNHI	ASPathNHI
*>0	3:1(3)	3 0
*>1	3:1(3)	3 1
*>3	3:1(3)	3
*>2	2:1(3)	2
*>4	self	i

为此，我们对图5所示的仿真拓扑结构图，不采用任何流量工程控制策略，只保证各 AS 间的连通关系，如图6所示。

用 SSFnet 对图6所示情况进行仿真，得到的仿真结果如表2所示。

从表2可以看出，AS4 不再选择 ASPathNHI 4 2 1 3 0 或 ASPathNHI 4 2 1 0 的路由到目的地 AS0，而选择最短路径 ASPathNHI 4 3 0 到达目的地 AS0；同时，所有路由都可以到达 AS4。

通过对表1和表2的比较可以发现，不配置 BGP 的属性，路由通路是默认选取最短路径的路由，这样会导致这个路径上的流量出现拥塞现象。而利用本文提出的方法，针对不同要求，适当配置 BGP 的相应属性，合理利用 BGP 的输入策略或输出策略，使流量分配到不同的路径上，实现入界流量或出界流量的控制，进而实现域间流量工程。

结论 由于域间流量很难控制，所以域间流量工程是个很重要的研究课题。很好地控制流量能更好地利用资源，优化性能。

本文提出的用 BGP 实现的域间流量工程的方法，是以 BGP 属性、BGP 策略和 AS 关系为基础，通过配置 BGP 的 LOCAL-PREF 属性，运用输入路由策略，控制 AS 的出界流量；另外，也可以在保证一个客户 AS 不在提供者间或对等体间过渡流量的情况下，允许客户 AS 只向它的部分提供者而不是全部提供者通告路由，或人为增加 AS_PATH，控制出界流量。通过应用不同策略，对入界流量和出界流量进行协调控制，实现域间流量工程。

仿真表明此方法有效地在 BGP/MPLS VPNS 中用 BGP 实现了域间流量工程的管理。

参 考 文 献

1 Rosen E, Rekhter Y. RFC2547: BGP/MPLS VPNs [S]. Reston, Virginia: Internet Engineering Task Force, March 1999

2 Rekhter Y, Li T. RFC 1771: A Border Gateway Protocol 4 (BGP-4)[S]. Reston, Virginia: Internet Engineering Task Force, Mar 1995

3 Rosen E, Viswanathan A, Callon R. Multiprotocol Label Switching Architecture [S]. Reston, Virginia: Internet Engineering Task Force, January 2001

4 Cristallo G, Jacquenet C. An approach to inter-domain traffic engineering. [EB/OL]. http://www.ist-tequila.org/publications/wtc2002-idte.pdf

5 Wang F, Gao L. Inferring and Characterizing Internet Routing Policies [J]. In: ACM SIGCOMM Internet Measurement Conference 2003. Karlsruhe: ACM SIGCOMM, 2003. 317~433

6 Halabi S, McPherson D. Internet routing architectures [M]. 2nd ed. Indianapolis: Cisco Press, 1997. 111~202

7 Gao Lixin. On Inferring Autonomous System Relationships in the Internet [J]. IEEE/ACM Transactions on Networking, 2001, 9 (6): 733~745

8 Huston G. Interconnection, peering, and settlements[EB/OL]. http://www.potaroo.net/papers/books/peerdocs/peering.pdf. 2004

9 Nemeth B E, Claffy K. Internet expansion, refinement and churn [EB/OL]. http:// www.caida.org/outreach/papers/2002/EGR, 2002

10 The SSFNET Project[EB/OL]. http://www.ssfnet.org. 2004