ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.19, No.6, June 2008, pp.1465−1472
DOI: 10.3724/SP.J.1001.2008.01465

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

解决策略冲突导致 BGP 路由发散的自适应机制*

王立军+, 吴建平, 徐 恪

(清华大学 计算机科学与技术系,北京 100084)

An Adaptive Mechanism to Solve BGP Divergence Resulted from Policies Conflict

WANG Li-Jun+, WU Jian-Ping, XU Ke

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

+ Corresponding author: E-mail: wlj@csnet1.cs.tsinghua.du.cn

Wang LJ, Wu JP, Xu K. An adaptive mechanism to solve BGP divergence resulted from policies conflict. *Journal of Software*, 2008,19(6):1465−1472. http://www.jos.org.cn/1000-9825/19/1465.htm

Abstract: As a policy-based routing protocol, BGP (border gateway protocol) allows each AS to choose local routing policy independently. Possible policies conflict may result in BGP route persistent oscillation. This paper proposes an adaptive mechanism to guarantee BGP convergence with policies conflict, which neither impairs the flexibility of choosing routing policy, nor inserts additional information in BGP messages. With the mechanism, route stability is taken into BGP decision process so that unstable route is degraded to cause more stable route to be chosen to stop policies dispute. The new mechanism also can adapt to topology change and converge to new stable route.

Key words: border gateway protocol (BGP); inter-domain routing; policies conflict; routing convergence

摘 要: BGP(border gateway protocol)作为一种基于策略的协议,允许每个自治系统独立地选择本地路由策略.自治系统之间可能存在的路由策略冲突会引起 BGP 路由持续不稳定.当前提出的解决办法要么需要增加额外的通信开销,要么限制自治系统自由的选择路由策略.提出了一种解决策略冲突引起 BGP 路由不收敛的自适应方法,既不损害自治系统选择路由策略的灵活性,也不需要在 BGP 消息中增加额外信息.路由的稳定性被加入到 BGP 的判决过程中,不稳定路由的优先级被降低,使更加稳定的路由得以被选为最优路由,终止路由策略冲突引起的争执.在网络拓扑发生改变的情况下,这种新方法能够自适应地调整路由选择,重新收敛到新的稳定状态.

关键词: 边界网关协议;域间路由;路由策略冲突;路由收敛

中图法分类号: TP393 **文献标识码**: A

路由系统是互联网最核心的基础设施,它分为域内路由和域间路由两个层次.域间路由协议是自治系统之间的粘合剂,提供其他自治系统中网络的可达性信息,对端到端服务质量有非常重要的影响.域间路由协议从最初的 EGP(exterior gateway protocol)到 BGP(border gateway protocol),经过了几个阶段的发展,尽管目前使用的 BGP4 能够基本满足传递可达性信息的需求,但是仍然存在不尽如人意的地方.作为互联网上所有自治系统间

* Supported by the National Natural Science Foundation of China under Grant No.60473082 (国家自然科学基金); the National Basic Research Program of China under Grant No.2003CB314801 (国家重点基础研究发展计划(973))

Received 2007-03-06; Accepted 2007-03-19

的通信协议,BGP 需要具有良好的扩展性、稳定性和灵活性.如何在这几个方面提高 BGP 的性能,一直是计算机网络领域的讨论热点.

BGP 与域内路由协议(如 RIP 和 OSPF)最大的不同之处在于,它是一种基于策略的路由协议,并不以优化某种网络性能量度值为目标.每个自治系统根据自己的经济利益制定路由选择策略,尽管这些路由策略从本地看来是合理的,但是,不同自治系统路由策略间的冲突会引起 BGP 路由不收敛.在规模庞大的互联网中,定位和消除自治系统间可能存在的策略冲突是一项很大的挑战.

针对上述问题,本文提出了一种在存在路由策略冲突的情况下保证BGP路由收敛的自适应机制.新机制中,利用BGP的抖动抑制机制提供路由稳定性的依据,并将路由稳定性加入BGP路由选择标准中.不收敛路由被选择为最优路由的可能性不断降低,直到另外一条稳定的路由被选择.这样,自然消除了路由策略冲突造成的 BGP 路由不收敛的问题.

本文第 1 节简单介绍 BGP 协议.第 2 节介绍策略冲突导致 BGP 路由不收敛的问题以及相关研究.第 3 节讨论自适应机制的原理和设计.第 4 节通过模拟实验分析自适应机制解决 BGP 路由不收敛的效果.最后总结全文.

1 边界网关协议 BGP

互联网由超过 25 000 个自治系统(autonomous system,简称 AS)组成.每个 AS 由 16 比特自治系统号码标识.边界网关协议 BGP[1]是 AS 间的通信机制,负责传递网络层可达信息.BGP 路由封装在 Update 消息中传递,包括路由声明和路由取消两种类型.BGP 路由器发现一条新路由后,通过路由声明将这条路由传播给邻居,其中包括目的网络的 IP 地址前缀和一系列描述路由特征的路由属性.如果 BGP 路由器发现达到某个目的网络的路由不再可用,就向邻居发送路由取消,取消前面发送的到达该目的网络的路由.当 BGP 路由器的路由发生改变时,就向邻居发送新路由的路由声明,同时默认取消前面的发送路由.

不同 AS 的路由器间建立的 BGP 会话称为 eBGP,而同一个 AS 内的路由器间建立的 BGP 会话称为 iBGP.同一 AS 内的边界路由器一般具有统一的路由策略,因此在本文中,我们忽略 AS 的内部细节,将一个 AS 抽象为一个点.丰富的 BGP 路由属性为制定灵活的路由策略提供支持.BGP 路由策略体现在路由处理的 3 个阶段上:输入过滤、路由判决和输出过滤.输入过滤阶段过滤掉某些来自邻居路由器的路由,并设定来自 eBGP 邻居的路由的 LOCAL_PREF 属性值.LOCAL_PREF 是 BGP 路由的重要属性,体现 AS 对这条路由的优选程度,路由的 LOCAL_PREF 属性值越大,越被 AS 优先选择.路由判决按照规则从多条路由的中选择一条最优路由.输出过滤阶段过滤某些路由并设定路由的属性,比如设定 NEXT_HOP 为发送接口的 IP 地址和将本 AS 的自治系统号码加入 AS_PATH 中,之后将路由发送给邻居路由器.路由判决规则由一系列不同优先级的规则组成,见算法 1,BGP 按照其中的顺序比较路由,如果按照第 1 条规则没有选出唯一的路由,那么按照下一条规则继续选择,直到选出的路由唯一为止.

算法 1. BGP 选择路由的顺序.

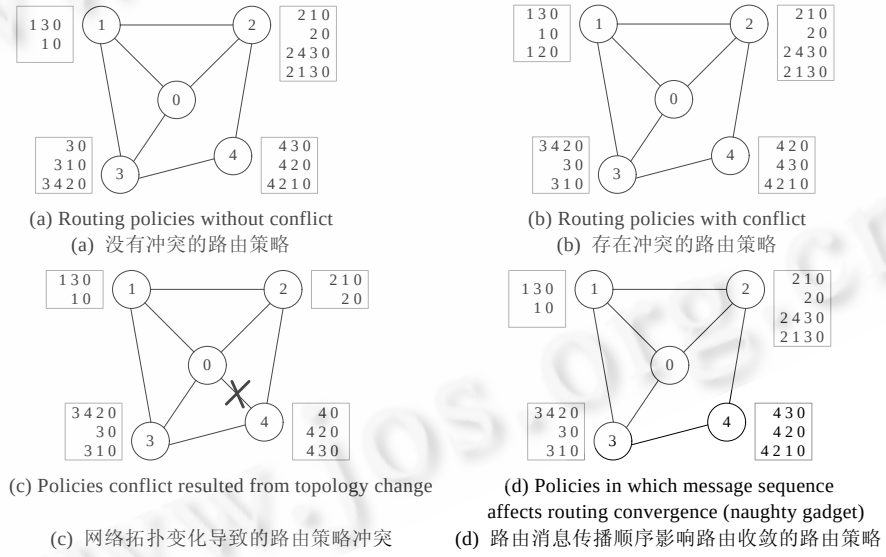
1. 优选值(LOCAL_PREF)最大;
2. AS_PATH 属性中路径最短;
3. ORIGIN 属性值最小;
4. 如果来自同一个 AS,则选择其中 MED 值最小的;
5. 选择来自 eBGP 的路由;
6. 到达下一跳路由器的 IGP 量度最小;
7. 路由的发送路由器 ROUTER ID 最小.

为了提高路由稳定性,BGP 引入一些控制机制以限制路由的变化频率.MRAI 规定了 BGP 路由器向一个邻居发送到达同一目的前缀的路由的最小时间间隔,默认值为 30s.除此以外,路由抖动抑制 RFD(route flap damping)[2]是提高 BGP 路由稳定性的重要机制.网络设备的软、硬件故障以及配置错误都会引起路由持续的反复变化,一般称为路由抖动(flapping).抖动路由会消耗网络设备大量的处理和通信能力,降低传输性能.RFD 为

每条路由维护一个惩罚值(penalty),记录路由过去变化的剧烈程度,以推测路由未来的稳定性.每次路由发生变化,惩罚值就增加一个固定的增量.路由取消导致的增量 P_W 和路由属性改变导致的增量 P_{AC} 往往被设定为不同值,一般情况下,$P_W<P_{AC}$.当 RFD 惩罚值超过预定的阈值 P_{cutoff} 时,BGP 抑制这条路由,也就是说,它不能再参加路由判决过程.路由被抑制的时间长短与路由的不稳定程度相关.路由稳定时,惩罚值按指数规律衰减.假设 t_0 时刻惩罚值为 $p(t_0)$,$t>t_0$ 时刻变为 $p(t)$,那么 $p(t)=p(t_0)e^{-\lambda(t-t_0)}$,其中,$\lambda=\ln 2/H$,$H$ 是衰减的半衰期.当惩罚值减小到小于预定阈值 P_{reuse} 时,解除路由抑制.RFD 被普遍认为是能够维护 BGP 路由表稳定的机制,已经被广泛实现在商用路由器中.尽管 MRAI 和 RFD 会限制路由变化的频率,但是本文第 3.1 节说明它们不能消除路由策略冲突导致的路由不收敛问题.

2 BGP 路由策略冲突

域间路由策略冲突问题首先是由 Varadhan 等人[3]提出来的,他们发现,即使在一些非常简单的网络拓扑中,域间路由策略也会引起路由的持续抖动.如图 1 中的简单网络拓扑,每个节点代表一个 AS,旁边的方框表示根据节点路由策略形成的路由选择顺序,节点按照从上到下的顺序选择到达节点 0 的路由.在图 1(a)中,无论节点初始的路由是什么,节点按照图中策略选择路由,最后总能达到收敛状态:节点 1 的路由是⟨130⟩,节点 2 的路由是⟨20⟩,节点 3 的路由是⟨30⟩,节点 4 的路由是⟨430⟩.然而在图 1(b)中,如果节点根据设定的路由策略选择路由,则所有节点都不能得到到达节点 0 的稳定路由.假设节点 1 首先选择了⟨10⟩作为到达 0 的路由,将这条路由传递给邻居节点 2 和节点 3.节点 2 可选择的路由包括⟨20⟩和⟨210⟩,节点 2 选择⟨210⟩.节点 3 可选择的路由有⟨30⟩和⟨310⟩,因此,节点 3 的路由是⟨30⟩.节点 4 在收到来自节点 2 和节点 3 的路由后,可选的路由包括⟨4210⟩和⟨430⟩,节点 4 选择⟨430⟩.但是,节点 1 在收到节点 3 的路由⟨30⟩后,重新选择⟨130⟩作为到达 0 的路由,之后发送给节点 2.节点 2 的路由随之改变为⟨20⟩,并发送给节点 4.节点 4 的路由改变为⟨420⟩,并传递给节点 3,这样,节点 3 的路由变为⟨3420⟩,随之,节点 1 的路由再次变回⟨10⟩.如此,节点 1~节点 4 的路由反复变化,一直不能收敛.



(a) Routing policies without conflict
(a) 没有冲突的路由策略
(b) Routing policies with conflict
(b) 存在冲突的路由策略
(c) Policies conflict resulted from topology change
(c) 网络拓扑变化导致的路由策略冲突
(d) Policies in which message sequence affects routing convergence (naughty gadget)
(d) 路由消息传播顺序影响路由收敛的路由策略

Fig.1 A simple network topology and routing policies

图 1 一个简单的网络拓扑及路由策略

如果 AS 的路由策略不受限制,就很可能产生冲突;如果限制路由选择的条件,比如按最短路径长度选择,则又会降低 BGP 路由选择的灵活性.而且 BGP 的目标不是找到到达目的节点的最短路径.Griffin 等人[4]提出,把 BGP 看作是解决稳定路径问题(stable path problem,简称 SPP)的分布式算法.每个 AS 被抽象为一个点后,互联网可以表示为无向图 $G=(V,E)$,其中,节点集合 V 中的每个点表示一个 AS,边集合 E 中的每条边表示一个 eBGP 会

话.假设我们只研究到达某个特殊节点 o 的路由.如果路由 $R_u=\langle v,w,\ldots,o\rangle$是节点 u 选择的到达 o 的最优路由,那么根据节点 u 的路由策略,在来自所有邻居的到达 o 的路由中,R_u 具有最高的优选等级.因此,BGP 的解并不是全局最优的,而是使 V 中每个节点得到了各自认为是最优路由的均衡状态.BGP 负责找到稳定路径问题的解,如果无解,那么路由就会发散.路由策略存在冲突的节点选择出最优路由后,相互发送 BGP 消息,进而选出新的最优路由,如此不断反复,好像这些节点间在发生争吵,因此,策略冲突导致节点间无休止的交互过程也称为策略争执(policy dispute).Griffin 等人[4]应用争执环(dispute wheel)研究路由策略冲突问题,得到的结论是,稳定路径问题存在唯一解的充分条件是节点间没有形成争执环.

为了解决路由策略导致的路由不收敛问题,Griffin 和 Wilfong[5]提出了一种安全的路径向量协议.主要的改进是在路由消息中增加 route history 属性.路由变化信息记录在 route history 中,并根据检测其中是否包含环路判断是否存在冲突的路由策略.这种方法的缺陷是,在某些不存在策略冲突的情况下会作出错误的判断.Cobb 和 Musunuri[6]提出了一种利用 AS 间交换的代价量度(cost metric)来衡量路由发散可能性的方法.以上两种方法都需要在 BGP 消息中增加额外的信息,不仅增加了路由器的存储和通信开销,而且不利于在网络中的逐步部署.

静态分析[3]是解决路由收敛问题的一种方法.它要求互联网上所有的 AS 提交它们的路由策略给 Internet Routing Registry(IRR).路由策略使用标准的路由策略描述语言[7,8]描述.根据登记的最新路由策略,分析其中可能存在的策略冲突.但是,这种全局协作方式的解决方案仍然面临难题:首先,出于经济和政治等方面的考虑,很多 AS 不愿登记和及时更新隐私的路由策略;其次,Griffin 等人[9]证明,通过静态分析发现路由策略冲突是 NP-完全问题;再次,即使一种路由策略在某种网络拓扑中不存在冲突,但是链路或者设备故障会引起网络拓扑变化,路由策略在新拓扑上就有可能生成冲突.因此,静态分析的方法并不能从根本上解决 BGP 路由收敛问题.

Gao 等人[10]提出了 AS 设置路由策略的指导规则,并证明只要 AS 按照这些规则配置路由策略,无须全局协作就能保证路由收敛.基本思想是,利用 AS 间的经济关系形成的互联网层次结构,为到达目的网络的路由设定一个偏序关系.在规则中,来自 Customer 的路由比来自 Peer 和 Provider 的路由被优先选择,来自 Peer 的路由比来自 Provider 的路由被优先选择.这与实际中域间流量控制的操作一致.但是,BGP 路由的收敛是由人为的配置保证,而不是 BGP 协议本身,不仅限制了配置 BGP 路由策略的灵活性,也增加了路由配置复杂度.错误的路由配置还可能引起路由不收敛.随着网络拓扑连接变得复杂化,AS 要求对流量有更强的控制能力,多路径[11]和多宿主[12,13]的使用都会对这种方法提出挑战.

3 自适应收敛机制 ACM(adaptive convergence mechanism)

针对现有策略冲突问题解决办法的缺陷,本文提出的路由自适应收敛机制 BGP-ACM 的设计目标包括:

(1) 不在 BGP 路由消息中增加额外的信息;

(2) 不限制设置 BGP 路由策略的灵活性;

(3) 网络拓扑发生变化后,路由能够自适应收敛到新的稳定状态.

设计的基本思想是,将路由稳定性动态地加入到路由选择中.如果路由策略规定某条路由具有较高的优先级,但是该路由不断发生变化,那么在路由选择中逐渐降低该路由的优选值,使其他稳定的路由有机会被选择.

3.1 路由策略冲突与RFD

RFD 会限制抖动路由的传播,但是,RFD 对路由策略冲突导致的路由变化有什么样的影响一直不清楚.因此,我们首先研究在 RFD 的作用下,BGP 路由会在策略争执中如何表现.模拟软件 SSFNet[14]是一种事件驱动的网络模拟器,在域间路由协议的研究中得到了广泛应用.我们使用 SSFNet 构建了图 1(b)中的网络拓扑,并在每个节点部署了相应的路由策略.模拟中,邻居节点间的消息传输延时设为 0.01s,MRAI 时钟设为默认值 30s.每个节点都使用了 RFD 功能,参数为:P_W=1.0,P_{AC}=0.5,P_{cutoff}=2.0,P_{reuse}=0.75.为了简化表述,我们使用 R_1,R_2,R_3,和 R_4 分别表示节点 1~节点 4 到达节点 0 的路由.R_2 的变化过程记录见表 1.根据节点 2 的路由策略,R_2 主要受到来自节点 1 的路由影响,因此,节点 2 收到的 R_1 及其 RFD 惩罚值也记录在表 1 中.

表 1 中第 1 行是事件发生的时间 t,由于节点 2 的路由主要受节点 1 路由的影响,因此,表 1 中的第 2 列是 t

时刻收到的来自节点 1 的路由 R_1,第 3 列表示 R_1 变化后经过更新的 RFD 惩罚值,第 4 列 R_2 表示节点 2 经过重新选择后得到的最优路由.在 t=5.08 时刻,节点 0 发源路由,经过 0.01s 的传输延时,节点 2 在 t=5.09 时刻得到直接到达节点 0 的路由⟨20⟩.在 t=5.10 收到来自节点 1 的路由⟨10⟩后,根据路由策略 R_2 变为⟨210⟩.之后,R_2 开始随着收到的 R_1 不断发生变化.在 t=123.61 时刻之前,节点 2 收到节点 1 路由的频率是由 MRAI 时钟控制的,R_1 的 RFD 惩罚值随着路由的变化不断增加.在 t=123.61,惩罚值超过了 P_{utoff},于是,节点 2 抑制了来自节点 1 的路由,模拟中,抑制直到 t=1635 才解除.需要注意的是,从 t=123.61s 之后,节点 3 的 RFD 功能开始发挥作用,并主导了其他节点路由的变化模式.

Table 1 Route change of R_2, influenced by policies conflict and RFD

表 1 在路由策略冲突和 RFD 作用下,节点 2 的路由变化

Time (s)	5.09	5.10	28.79	52.20	76.20	99.91	123.61*
R_1		⟨10⟩	⟨130⟩	⟨10⟩	⟨130⟩	⟨10⟩	⟨130⟩
Penalty of R_1		0	0.5	0.99	1.47	1.94	**2.41***
The best route	⟨20⟩	⟨210⟩	⟨20⟩	⟨210⟩	⟨20⟩	⟨210⟩	⟨20⟩

Time (s)	1 883.98	1 910.22	3 310.48	3 336.72	4 736.98	4 763.23	6 163.48
R_1	⟨10⟩	⟨130⟩	⟨10⟩	⟨130⟩	⟨10⟩	⟨130⟩	⟨10⟩
Penalty of R_1	1.12	1.60	1.04	1.52	1.02	1.49	1.01
The best route	⟨210⟩	⟨20⟩	⟨210⟩	⟨20⟩	⟨210⟩	⟨20⟩	⟨210⟩

节点 3 发生的事件记录在图 2 中.根据节点 3 的路由策略,它的路由变化主要受到来自节点 4 路由的影响.在图 2 的上部,粗体实线表示 R_4 没有被节点 3 抑制,而虚线表示节点 3 抑制了 R_4.在 t=109.89 时刻,节点 3 抑制了 R_4,直到 t=1883.96 才解除抑制.因此在这段时间里,节点 3 的路由不再发生变化,网络中的策略争执也暂时停止.当节点 3 在 t=1883.96 时刻解除对 R_4 的抑制后,策略争执继续.发生两轮争执之后,节点 3 在 t=1884.02,再次抑制了来自节点 4 的路由.这样,网络中的节点 1、节点 2 都受到节点 3 路由的影响,发生类似模式的变化.表 1 中的时间间隔[1910.22,3310.48]和[3336.72,4736.98]都对应着节点 3 对 R_4 的抑制.

从模拟的结果来看,RFD 确实限制了策略冲突导致的路由变化的频率,但是,路由不收敛的问题并没有得到解决.每次被 RFD 抑制的路由被解除,策略争执就继续.

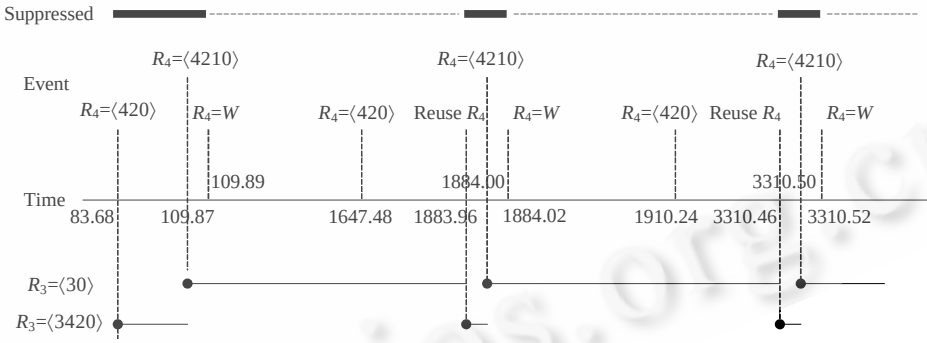


Fig.2 In the policies dispute resulted from policies conflict, node 3 suppresses route from node 4 periodically

图 2 在策略冲突导致的路由争执中,节点 3 周期性的抑制来自节点 4 的路由

3.2 设计原理

从第 3.1 节可以看出,每次解除路由抑制,策略争执都会继续.在 RFD 机制中,每条路由只有两个状态,像其他稳定路由一样参加路由判决过程和不能参加路由判决过程.两种状态之间没有过渡状态,在 BGP-ACM 的设计中,将路由稳定性加入到路由判决过程,给路由引入基于 RFD 惩罚值的中间状态;如果等级最高的路由在发生抖动,在该路由被 RFD 抑制之前,就使稳定的路由能够替换它成为最优路由;路由的稳定性越差,被选择为最优路由的可能性就越小.

AS 路由策略对路由选择的影响主要体现在路由优选值的设置上.比如,到达同一目的网络的来自 Peer 和

Customer 的两条路由,AS 会给来自 customer 的路由设定更高的优选值,使来自 customer 的路由被选择.为了将路由稳定性加入到路由选择中,我们采用的方法是把路由的 RFD 惩罚值映射为优选值的负增量,使路由从优选值中减去负增量后,它被选择为最优路由的概率减小.假设在时刻 t,节点 u 从邻居得到的到达目标节点 d 的路由是 $R_1,R_2,\ldots,R_n$,根据 u 的路由策略,这些路由的优选值为 $L(R_i),i=1,2,\ldots,n$,并满足 $L(R_i)>L(R_j),i<j$,RFD,惩罚值分别为 $p(R_i),i=1,2,\ldots,n$.在 BGP-ACM 的作用下,每次路由发生变化后,节点按照如下步骤重新计算路由的优选值:

(1) 得到路由优选值的差值 $D_i=L(R_i)-L(R_{i+1}),i=1,2,\ldots,n-1,D_n=0$;

(2) 计算负增量,

$$\delta_i=\sum_{j=i}^{n-1}p(R_i)\times D_i \tag{1}$$

(3) 在优选值 $L(Pi)$中减去负增量,得到更新的优选值,记为 $L'(R_i)=L(R_i)-\delta_i$.

最终,加入路由稳定性影响的路由优选值表示为

$$L'(R_i)=L(R_i)-\sum_{j=i}^{n-1}p(R_i)\times(L(R_i)-L(R_{i+1})) \tag{2}$$

假设 R_1 由于受策略冲突的影响不断发生变化,而 R_2 是一条稳定路由.当 $0<p(R_1)<1.0$ 时,R_1 经过更新的优选值 $L'(R_1)$会减小,但是仍然有 $L'(R_1)>L'(R_2)$,因此在路由判决过程中,R_1 仍然会被选择为最优路由.当 $p(R_1)$超过 1.0 时,$L'(R_1)<L'(R_2)$,于是,R_1 在路由判决过程中被降低到 R_2 以下,尽管这个时候 R_1 没有被 RFD 抑制.假设 R_2 也受到了策略冲突的影响,而 R_3 是一条稳定路由,那么当 $p(R_1)>1.0,p(R_2)>1.0$ 时,更新后的优选值 $L'(R_1)<L'(R_3)$,$L'(R_2)<L'(R_3)$,于是,R_3 被选择为最优路由.与 RFD 抑制路由的过程相比,在不稳定路由被稳定路由代替之后,没有解除抑制时重新选择路由的过程.因此,策略争执一旦被 BGP-ACM 停止下来,它就不会再继续.即使由于网络拓扑发生改变导致产生临时的策略冲突,BGP-ACM 也能够自适应地调整路由优选值.例如图 1(b)中,在 $t=99.91$ 时刻,$R_1=\langle 210\rangle$的惩罚值是 1.94,$\langle 210\rangle$更新后的优选值小于$\langle 20\rangle$更新后的优选值,于是,在路由判决中节点 2 选择的路由就不再是$\langle 210\rangle$而是$\langle 20\rangle$,终止了策略争执.

4 评 价

为了评价 BGP-ACM 的有效性,我们在模拟软件 SSFNet 的 BGP 代码里实现了 BGP-ACM.实验中使用了图 1 中的简单网络拓扑和策略冲突场景.根据文献[4],策略冲突起源于网络中某些形成“争执环”的节点,而网络中不在争执环中的节点不影响策略冲突的形成和消除,因此,在模拟评价中可以忽略不参与策略冲突的节点.这样,尽管实验中使用的拓扑简单,但是同样能够有效评价 BGP-ACM 对解决策略冲突引起路由不收敛问题的效果.

4.1 有效性

第 1 个模拟实验构建了图 1(b)中的拓扑,并在每个节点实现了相应的路由策略.模拟中用到的参数设置与第 3.1 节中的设置完全一样.实验结果是,BGP-ACM 在 3 个 MRAI 时间后少于 90s,终止了策略争执.稳定状态时,各个节点的路由分别是 $R_1=\langle 130\rangle,R_2=\langle 20\rangle,R_3=\langle 30\rangle,R_4=\langle 420\rangle$.显然,节点 3 将来自节点 4 的路由$\langle 3420\rangle$降级到了$\langle 30\rangle$以下,选择了稳定的$\langle 30\rangle$,终止了策略争执.

第 2 个实验评价 BGP-ACM 在网络拓扑暂时发生改变情况下的表现,我们模拟了图 1(c)中的场景.在最开始的时候,节点 4 直接与节点 0 连接,所有节点的路由达到稳定状态:$R_1=\langle 130\rangle,R_2=\langle 20\rangle,R_3=\langle 30\rangle,R_4=\langle 40\rangle$.在 $t=t_1$ 时刻,节点 4 与节点 0 之间的连接中断,这样在节点 1~节点 4 之间就形成了图 1(b)中的策略冲突,节点间开始了策略争执.但是,由于节点都实现了 BGP-ACM,在争执持续了少于 90s 之后就恢复了稳定,再次稳定后的路由是:$R_1=\langle 130\rangle,R_2=\langle 20\rangle,R_3=\langle 30\rangle,R_4=\langle 420\rangle$.节点 3 降低了路由$\langle 3420\rangle$的等级,使网络中的路由到达稳定状态.在 $t=t_2$ 时刻,节点 0 和节点 4 之间的连接恢复,那么,节点的路由立即恢复到了连接中断前的稳定状态.这个实验结果说明,BGP-ACM 不仅能够终止策略争执,而且在暂时的策略冲突消除后,能够使路由恢复到策略冲突前的状态.这表明,BGP-ACM 能够适应网络拓扑的动态变化,使节点得到稳定的最优路由.

在第 3 个实验中,我们评价了 BGP-ACM 部分部署情况下的有效性.在每次模拟中,图 1(b)中只有 1 个节点部署 BGP-ACM,实验结果见表 2.当节点 1 部署 BGP-ACM 时,路由收敛时间是 109.89s.稳定状态的路由⟨10⟩*表示节点 1 降低了⟨130⟩的等级,选择了稳定的⟨10⟩.从表 2 可以看出,节点 2 部署 BGP-ACM 与节点 1 部署的情况相似,路由都很快收敛.这说明 BGP-ACM 的部分部署对 BGP 路由的稳定性有很大提高.但是,节点 3 部署 BGP-ACM 时的情况有所不同,收敛时间达到了 1697.58s.

Table 2 Convergence with partial deployment of BGP-ACM

表 2 BGP-ACM 部分部署时的路由收敛情况

The nodes with BGP-ACM	The convergence time (s)	R_1	R_2	R_3	R_4
1	109.89	⟨10⟩*	⟨210⟩	⟨30⟩	⟨430⟩
2	83.83	⟨10⟩	⟨20⟩	⟨3420⟩	⟨420⟩
3	1697.58	⟨130⟩	⟨20⟩	⟨30⟩	⟨420⟩
4	∞	φ	φ	φ	φ

图 3 描述了 R_3 在模拟中的变化情况,可以看出,R_3 相对于 R_1 和 R_2 有较长的收敛时间,是由于在策略争执停止之前,节点 2 和节点 4 的 RFD 抑制了路由.但是,当抑制被解除后,在 BGP-ACM 的作用下,路由还是很快达到了收敛状态.从此也可以看出,BGP-ACM 对网络中的复杂变化有较强具有适应能力.

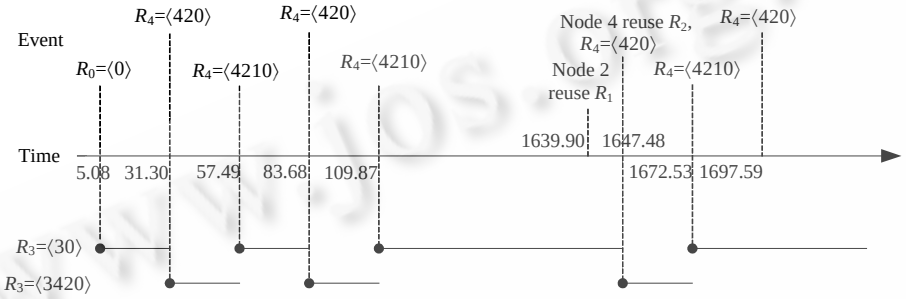


Fig.3 Changing behavior of R_3 with BGP-ACM deployed only on node 3

图 3 仅有节点 3 部署 BGP-ACM 时,R_3 的变化情况

最有意思的是,在仅有节点 4 部署 BGP-ACM 的情况下,策略争执并没有像前 3 个节点部署 BGP-ACM 那样被终止,而是一直持续了下去.表 2 中使用收敛时间为∞表示路由不能达到收敛状态,使用φ表示路由处于不断变化的状态.在节点 4 将⟨420⟩的等级降低后,路由判决过程中⟨430⟩比⟨420⟩被优先选择.但是这时,节点 4 的路由选择顺序形成了另外一种策略冲突的形式,Griffin 称其为“Naughty Gadget”[5].在这种特殊的策略关系中,网络路由是否收敛取决于节点收到路由消息的顺序.图 1(d)是一个典型的例子,在初始状态下,如果来自节点 3 的⟨30⟩先于来自节点 2 的⟨20⟩到达节点 4,节点就会达到稳定状态:R_1=⟨130⟩,R_2=⟨20⟩,R_3=⟨30⟩,R_4=⟨430⟩,否则,路由就不会收敛.上述实验中的结果,就是因为节点 4 在 BGP-ACM 的作用下形成了一个“Naughty Gadget”.节点 4 收到路由消息的顺序导致新的路由选择顺序仍然不能使路由收敛.由此可见,部署 BGP-ACM 的节点最好有到达目的节点的稳定不变的路由,以免形成“Naughty Gadget”.

4.2 对RFD的影响

BGP-ACM 不仅改变策略冲突中路由的优选值,而且也对 RFD 的作用产生影响.在没有路由策略冲突的情况下,当相对稳定路由的 RFD 惩罚值超过一定值的时候,尽管没有达到 RFD 的阈值,仍然可能被降低路由的优先等级.但是我们认为,BGP-ACM 对 RFD 的影响是可以接受的.假设 R 是节点 u 到达节点 d 的唯一路由.受链路断开/恢复的影响,R 不断抖动,那么,BGP-ACM 对节点 u 抑制 R 的行为没有任何影响,因为根据公式(2),路由 R 的负增量一直是 0.如果节点 u 到达 d 的路由是集合$\{R_1,R_2,\ldots,R_n\}$,$L(R_i)>L(R_{i+1})$,i=1,2,…,n−1.在 BGP-ACM 的作用下,节点 u 在 R_i 没有被 RFD 抑制前可能会选择 R_{i+1} 为最优路由.假设路由 R_f 处于抖动状态,而且 R_f 不是节点 u 最优先选择的路由,那么,节点 u 的最优路由选择不会受到 R_f 变化的影响.因为根据公式(1),优选值大于 R_f 的所

有路由的优选值都减小 $p(R_f)\times(L(R_f)-L(R'))$,其中,$R'$是优选值仅次于 R_f 的路由.综上,在没有备用路由的情况下,BGP-ACM 不会对 RFD 路由选择产生影响;如果节点有备用路由,节点就会选择稳定的备用路由.这与节点使用严格的路由抖动抑制的不同之处在于,不稳定路由没有被抑制,而是其优选值随着 RFD 惩罚值的增加而降低.

5 结束语

随着越来越复杂的路由策略在 BGP 中的实施,为了维护整个互联网路由稳定,策略冲突导致路由不收敛的问题必须找到一种更好的解决办法.我们提出了一种自适应的路由收敛机制——BGP-ACM,保证在存在路由策略冲突的情况下,BGP 能够自动地收敛到稳定路由,并且能够根据网络拓扑变化作自适应调整.在 BGP-ACM 的作用下,路由在路由判决过程中的角色,由参加和不参加两种状态转变为一种渐变的过程.也就是随着稳定性变差,路由在判决过程中的优先等级不断降低,直到最后被 RFD 抑制.与现有的方法相比,BGP-ACM 的好处是对 BGP 协议的修改很小,不需要在节点间增加传输开销和对路由策略的设置作任何限制.在部分部署的情况下,BGP-ACM 也能取得很好的效果,这使它成为一种易于在实际网络中部署的可行方法.

致谢 我们向对本文工作给予支持的老师、对论文提出批评和建议的审稿老师和编辑同志表示感谢.

References:

[1] Rekhter Y, Li T, Hares S. A border gateway protocol 4 (BGP-4). RFC 4271, 2006.
[2] Villamizar C, Chandra R, Govindan R. BGP route flap damping. RFC 2439, 1998.
[3] Varadhan K, Govindan R, Estrin D. Persistent route oscillations in inter-domain routing. Computer Networks, 2000,(32):1−16.
[4] Griffin TG, Shepherd FB, Wilfong G. The stable paths problem and interdomain routing. IEEE/ACM Trans. on Networking, 2002, 10(2):3104−3107.
[5] Griffin T, Wilfong G. A safe path vector protocol. In: Proc. of the IEEE INFOCOM, Vol.2. Tel Aviv: IEEE, 2000. 490−499.
[6] Cobb JA, Musunuri R. Enforcing convergence in inter-domain routing. In: Proc. of the GLOBECOM, Vol.3. Dallas: IEEE, 2004. 1353−1358.
[7] Alaettinoglu C, Villamizar C, Gerich E, Kessens D, Meyer D, Bates T, Karrenberg D, Terpstra M. Routing policy specification language (RPSL). IETF RFC 2622, 1999.
[8] Meyer D, Schmitz J, Orange C, Prior M, Alaettinoglu C. Using RPSL in practice. IETF RFC 2650, 1999.
[9] Griffin T, Wilfong G. An analysis of BGP convergence properties. In: Proc. of the ACM SIGCOMM, Vol. 29. New York: ACM, 1999. 277−288.
[10] Gao L, Rexford J. Stable Internet routing without global coordination. IEEE/ACM Trans. on Networking, 2001,9(6):681−692.
[11] Xu W, Rexford J. MIRO: Multi-Path interdomain routing. In: Proc. of the ACM SIGCOMM. New York: ACM, 2006. 171−182.
[12] Goldenberg DK, Qiu L, Ye H, Yang YR, Zhang Y. Optimizing cost and performance for multihoming. In: Proc. of the ACM SIGCOMM. New York: ACM, 2003. 79−92.
[13] Akella A, Pang J, Maggs B, Seshan S, Shaikh A. A comparison of overlay routing and multihoming route control. In: Proc. of the ACM SIGCOMM. New York: ACM, 2003. 93−106.
[14] The SSFNet project. 2007. http://www.ssfnet.org/

王立军(1978－),男,河北唐山人,博士生,主要研究领域为互联网域间路由协议.

徐恪(1974－),男,博士,副教授,CCF 高级会员,主要研究领域为路由器软件体系结构.

吴建平(1953－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为下一代互联网体系结构,网络协议测试.