一种互联网的稳定路由选择策略

李 琦 徐明伟 吴建平

(清华大学计算机科学与技术系 北京 100084)

摘 要 互联网中网络故障频繁,域间路由协议(BGP)并不能很好地适应网络故障.一般情况下,域间路由协议会经历一个比较漫长的路由搜索过程,导致互联网中大量的数据包丢包.虽然目前已提出了很多改进的BGP算法,但这些算法复杂度非常高,给路由器增加很多额外的计算代价.为了解决这个问题,作者提出了一个稳定的域间路由选择算法sBGP.在sBGP中,当路由器收到由故障触发的路由通告后,采用启发式的路由选择算法选择目前可选的最稳定路由为最佳路由.通过稳定路由选择,路由器可以选择有效的稳定路由,以避免无效的路由搜索以及路由不断更新引入的路由器处理开销.分析和模拟实验表明sBGP不仅能够有效提高BGP的收敛性能,而且可以减少收敛过程中的通信开销.

关键词 域间路由协议;BGP;稳定路由选择;sBGP;路由收敛

中图法分类号 TP393 **DOI号**: 10.3724/SP.J.1016.2012.02668

A Stable Routing Selection Scheme in the Internet

LI Qi XU Ming-Wei WU Jian-Ping

(*Department of Computer Science, Tsinghua University, Beijing* 100084)

Abstract Network failures always occur in Internet, however, Inter-domain routing, border gateway protocol (BGP), can not well adapt to these failures. In general, BGP may suffer slow path exploration and usually cause extensive packet loss. Although several improved schemes are proposed, they introduce high computation complexity. To address this problem, we propose a stable BGP (sBGP) scheme. In sBGP, a heuristic algorithm is designed to choose the most stable route as the best route. Through the stable routing selection, routers can quickly identify valid route and eliminate slow path exploration, and resource consumption is effectively reduced. The analysis and simulation study shows that sBGP not only effectively improves the BGP convergence performance but also reduces the communications overheads during convergence.

Keywords inter-domain routing; BGP; stable routing selection; sBGP; routing convergence

1 引 言

边界网关协议BGP[1]是互联网域间路由的标准协议.BGP路由的可用性和稳定性对整个互联网路由有很大的影响.Labovitz等人[2]发现实际网络中BGP路由会经历漫长的路由搜索过程.互联网上的测量研究表明,80%左右的网络故障延时不超过180 s[3-4].然而,这些故障导致的路由失效和无效路由切换导致的BGP路由收敛延时会长达几十分

收稿日期:2010-01-03;最终修改稿收到日期:2011-04-20.本课题得到国家自然科学基金(61073166,61133015,61161140454)、国家"九七三"重点基础研究发展规划项目基金(2009CB320502,2012CB315803)、国家"八六三"高技术研究发展计划项目基金(2011AA01A101)资助.
李 琦,男,1979年生,博士,主要研究方向为网络体系结构、网络和系统安全.E-mail: liqi@csnet1.cs.tsinghua.edu.cn.**徐明伟**,男,1971年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为计算机网络体系结构、高速路由器体系结构、互联网路由.**吴建平**,男,1953年生,博士,教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为计算机网络体系结构、计算机网络协议测试、形式化技术.

钟[5]. 无效的路由搜索不仅增加了路由器的处理开销，并且延长了互联网路由中的路由黑洞和回路. 由此进一步增加了数据包在网络中的传输延时和丢包率[6]，并直接影响现有各种互联网实时应用的性能[7].

在 BGP 中，链路或者设备故障会使网络中一些路由成为无效路由. 这些无效路由是同一事件引起的相关路由，不会立即从网络中消失. 路由器不断地在无效路由中搜索最优路由并传播，直到稳定的可靠路由被选择为止. 路径搜索过程大大增加了 BGP 路由收敛延时[8]. 为了解决这些路由的慢收敛问题，本文提出了一种稳定的 BGP 稳定路由选择算法(sBGP)，利用路由相关性使稳定路由得到更多的机会被选为最优路由，缩短路径搜索过程以改善 BGP 路由收敛性能. 稳定路由选择机制的基本思想是，记录 Adj-RIBs-IN 中每条路由的申明时间，根据多条路由在 Adj-RIBs-IN 的时间判断它们的稳定性. 一旦路由发生变化则 Adj-RIBs-IN 中路由计时将被清零. 当链路发生故障以后，由该故障链路触发的路由更新携带故障源信息. 当路由器收到路由更新，则 sBGP 将首先判断当前所选的路由是否受故障影响. 如果其中所选的路由受影响，则 BGP 重新执行路由选择，并选择 Adj-RIBs-IN 中最稳定的路由. 如果路由不受影响，则 sBGP 直接忽略这个由于故障触发的病态路由更新事件. 于是，sBGP 在路由选择算法的简单修改不仅改进了 BGP 的收敛性能，而且限制了故障信息传播范围并提高了 BGP 的稳定性. 目前我们已经在自主研发的虚拟路由器[9]中实现了 sBGP 方案，下一步将在 CERNET2 中进行大规模部署以研究和分析 sBGP 在真实互联网中的性能.

本文第 2 节介绍 BGP 路由选择策略和问题，包括 BGP 路由传播过程选路问题、由此形成的域间路由收敛问题和 BGP 路由的主要问题以及相关的研究工作；第 3 节介绍 sBGP 启发式路由选择机制的原理，并给出具体关闭算法以及实例分析；第 4 节通过模拟实验的结果评价了 sBGP 和相关 BGP 改进方案的性能；最后第 5 节总结全文.

2　BGP 路由选择策略和问题

本节首先分析 BGP 路由的基本思想以及存在的问题，接着分析和评价现有 BGP 路由选择的改进方案.

2.1　BGP 选路问题

互联网由不同的自治系统(AS)组成，各个自治系统通过 BGP 协议互联. BGP 路由信息封装在 BGP 的通告消息中，其中包括路由的声明和路由撤销消息. 当一个 BGP 路由器发现一条新路由，路由声明消息将这条路由传播给邻居路由器. 这条路由声明消息包括目前网络的 IP 地址前缀以及描述这条路由的属性信息；当一个 BGP 路由器发现到某个目的网络的路由不可用，则向邻居路由器发送路由撤销的通告消息；当 BGP 路由器发生路由改变(由于路由策略配置变化或者收到路由变化的通告消息)，就向邻居路由器发送路由声明消息，在这种情况下缺省取消前面已使用的路由.

BGP 会话分为内部的 BGP 会话(iBGP)以及外部的 BGP 会话(eBGP). iBGP 仅在一个自治系统中发送路由通告消息以达到自治系统内路由的一致性，而 eBGP 提供 BGP 路由器向其它自治性的邻居路由器发送路由通告消息. iBGP 和 eBGP 会话路由决策过程是相同的，唯一的区别是 BGP 收到的 iBGP 通告消息后决策得到的路由不会继续向自治系统内的其它 BGP 通告. BGP 路由器收到 BGP 通告消息以后将执行的路由选路决策过程如表 1 所示，即每一个路由器的路由决策结果由 BGP 通告消息中的属性来决定. 表 1 中的每一步的计算结果是从前一步的计算结果重新选择获得一个路由的子集合. 通过增加、修改和过滤通告消息中的属性值，管理员可以控制到达目的网络的路由选择决策过程.

表 1　BGP 选路决策过程

步骤	决策行为
1	Highest local preference
2	Lowest AS path length
3	Lowest origin type
4	Lowest MED
5	eBGP over iBGP-learned
6	Lowest IGP cost
7	Lowest router ID

2.2　BGP 和改进 BGP 方案的问题

由于域间路由协议 BGP 的设计原则路径向量路由的特点以及策略配置的灵活性[1]导致了整个互联网路由的慢收敛甚至不收敛. 我们通过采用同步网络模型分析 B-团(B-clique4)网络的路由收敛过程来分析这些问题. 如图 1 所示，由 6 个 AS 组成全连接 B-团拓扑，AS0 的网络设备发生故障后，到达目的结点的路由会在 AS1，AS2，AS3 和 AS4 中经历路径搜索过程. 在图 1 中，带下划线的路由表示

AS 当前所选择的最优路由，[ ]表示没有到达目的结点的路由，实线方框中的路由表示该 AS 本轮发送给邻居的路由，没有使用实线方框围起来的表示上一轮发送给邻居的路由，w 表示发送的是路由取消. 为了便于分析，本文采用同步网络模型描述到达目的结点的路由的收敛过程. 同步网络具有 3 个含义：(1) 任意 AS 间的传输延时都是固定值，本文的例子设为 1 s；(2) 网络中的 AS 同时收到邻居在上一轮发送来的路由消息，选择新的最优路由并传播出去；(3) AS 优先选择优先级高的路由. 图 1 所示 BGP Adj-RIBs-IN 中的路由已按优先级高低排列.

当 AS0 和 AS5 之间的链路失效以后，在 $t=1$ 时刻，AS1 和 AS2 删除来自 AS5 的所有路由，在 AS1，AS2，AS3，AS4 都存在到达目的地址的路由，路径分别为[205]，[105]，[105]，[3105]. 尽管这 3 条路由由于 AS5 和 AS0 的链路故障成为无效路由，但是 AS1 和 AS2 都无法知道这个信息. 于是分别从所选择的[205]和[105]作为新路由，并发送[1205]和[2105]给邻居. 在 $t=2$ 时刻，所有 AS 同时收到邻居 $t=1$ 时刻发送过来的路由声明，更新对应的 Adj-RIBs-IN 并重新选择路由. 由于收到的来自 AS1 和 AS2 的路由，收到 AS1 的[1205]，AS3 选择[1205]为最优路由，并通告给邻居. 类似地，当 $t=3$ 时刻，当由于收到来自 AS3 和 AS4 的路由形成了回路，AS1 不再有到达目的地址的路由，向邻居发送路由撤销通告. 同样 AS2 也检测到回路，撤销路由. 由于 AS3 先收到 AS1 的路由撤销通告，AS3 将选择[2105]. 但由于受 MARI 限制，AS3 不会马上发送[32105]给邻居. 当 $t=4$ 时刻，AS3 在 MARI 时钟超时前收到了 AS2 发送的路由撤销通告. 同时，AS4 收到 AS2 的撤销通告，从 Adj-RIBs-IN 中删除[2105]. 直到 $t=4+t'$ 时刻，AS3 收到 AS2 的撤销通告，删除 Adj-RIBs-IN 中的路由，并发送撤销通告给邻居. AS4 收到路由撤销通告，删除 Adj-RIBs-IN 中所有路由，路由收敛过程结束. 可见，在 BGP 路由收敛的过程中，每个 AS 会在所有邻居的 Adj-RIBs-IN 中不断搜索、选择和传播无效路由，直到所有 AS 的 Adj-RIBs-IN 中都不再存在任何路由.

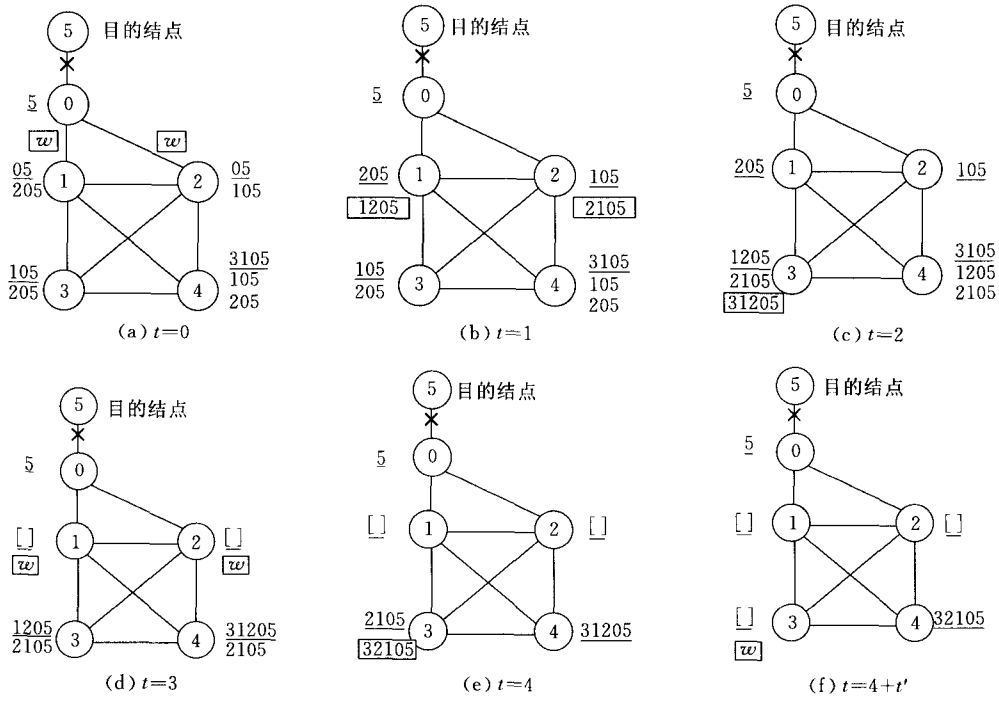


图 1 B-团拓扑的路由搜索过程

一般来说，域间路由协议 BGP 的以下几个问题导致了整个互联网路由的慢收敛甚至不收敛：

(1) 故障的错误检测. 在 BGP 协议中，路由器通过基于 TCP 可靠连接的 Keepalive 消息来检测邻居路由器，但是 TCP 会话的异常终止会引起 BGP 路由器不必要的路由重计算. 这种不必要的路由重计算会引起全局相关 BGP 路由的重计算，从而导致网络的不稳定.

(2) 路由计算时间过长. 不同于域内协议，作为一个路径向量协议，BGP 协议在路由被取消或路由发生切换时可能会经历时间较长的路径搜索过程，在路由重计算过程中，BGP 路由器经常选择的一些

路由可能已不再可用，这样的搜索过程导致了整个 Internet 中 BGP 路由收敛时间过长[8].

(3) 路由传播时延过长. BGP 协议中的一些改进机制也直接影响到 BGP 中的路由更新传播. 例如，BGP 协议中规定 BGP 路由器向其邻居发送到达同一目的网络路由的最小时间间隔值限定为一确定值 MRAI(Minimum Route Advertisement Interval). MARI 虽然可以在一定程度上减小通信开销，但同时也增加了路由收敛时间[8].

(4) BGP 路由不收敛. BGP 策略设置的灵活性和隐蔽性可能导致路由的不收敛. 不同的 AS 分别由不同的组织或机构管理，它们根据自身的需要设置不同的 BGP 路由策略，具有很大的灵活性. 此外，由于出于安全性等方面的考虑，各个 AS 的 BGP 路由策略通常是不对外公开的，具有很强的隐蔽性. BGP 策略的这种特性使得每个 AS 都无法获得全部路由策略信息，不能判断策略冲突，可能最终导致路由的不收敛[10].

很多改进的 BGP 方案缓解或者解决了上面的 4 个问题，但是由于这些方案本身设计的缺点并不能实际部署，以有效提高 BGP 收敛性能. RCN[8] 和 EPIC[11] 通过在 BGP 通告消息中增加路由变化的根源信息以加快路由的收敛，但它们的改进性能在很大程度上依赖于网络的拓扑结构，而且这些方案并不支持增量部署. CA 方案[12] 通过检测路由通告中的无效路由加快路由收敛，但这个方案给路由器增加了很大的计算负载. Ghost Flushing 方案[13] 通过在通告新路由前撤销旧路由加快路由收敛，但无法避免路由的无效搜索过程. 此外，Ghost Flushing 方案增加了很多 BGP 通告消息，而且这个方案会恶化故障恢复时的收敛性能. 由于 BFD 方案[14] 已经解决了 BGP 协议的故障检测问题，所以，本文将重点解决问题 2、3 和 4.

3　稳定路由选择

本节将介绍稳定路由选择方案(sBGP). 路由稳定选择方案将有效避免由于不稳定路由或者无效路由引起的互联网路由的有效性和稳定性问题. 通过这个方案，路由的有效性和稳定性将大幅度提高. 首先改进域间路由协议(Stable BGP)的基本原则，其次介绍稳定路由选择路由协议的基本算法，最后通过实例分析介绍稳定路由选择的有效性.

3.1　基本思想和原则

域间路由协议的目标是确保互联网路由的有效和稳定性，特别在发生网络故障的情况下的路由有效性和稳定性. 由于域间路由本身的设计原则，路由选择的决策并不准确，短暂的网络故障会导致域间路由的很多问题，例如路由收敛过程中漫长的路由搜索过程以及域间路由的不稳定性问题. 为了解决这些问题，我们提出一个启发式的 BGP 路由选择算法，Stable BGP(sBGP).

有别于目前的 BGP 协议，sBGP 的目标是通过稳定路由选择本地化处理路由故障，尽量使路由变化再小范围内发生. 我们在 sBGP 决策过程中增加了额外的两个路由选择，即选择目前可用路由和选择运行时间最长的路由. 表 2 所示 sBGP 的路由选择决策过程. 表中步骤 1 和步骤 2 是 sBGP 路由的核心过程. 选择最佳的路由的第 1 个步骤是优先选择目前可用的路由避免了不必要的路由重计算过程. 由于收到一个新路由通告消息可能是由于旧路由无效的事件触发，所以在路由器通告中增加一个属性来标识路由变化的事件源. 第 2 个步骤选择运行时间最长的路由，确保 sBGP 选择最稳定的路由以减少由于不稳定路由导致的路由不稳定情况. 由于目前的主流路由器，例如 Cisco 路由器，已经在 Adj-RIBs-IN 中标识了已选的最佳路由以及记录了每条路由的达到时间，所以 sBGP 没有增加改进路由选择的实现复杂度.

表 2　BGP 启发式的路由决策过程

步骤	决策行为
1	Available route
2	Longest available time route
3	Highest local preference
4	Lowest AS path length
5	Lowest origin type
6	Lowest MED
7	eBGP over iBGP-learned
8	Lowest IGP cost
9	Lowest router ID

需要注意的是，在 sBGP 中，路由器缺省将采用初始 BGP 的路由决策过程，而启发式路由选择仅在路由发生变化时执行. 这种启发式的路由决策过程保留了原始 BGP 路由的所有特征，例如确保基于路由策略的路由优先选择. 通过这种方式可以在保证管理员策略配置有效的情况下提供域间路由的有效性.

3.2　sBGP 的路由选择算法

与传统的 BGP 协议不同，sBGP 不是仅仅在

BGP 通告消息标识路由变化的事件源，它提供的启发式的路由决策过程选择最稳定路由. 由于 sBGP 能够准确识别可选的最稳定路由，有效避免了 BGP 不断地选择无效路由，从而确保了域间路由选择的稳定性和有效性. 在 sBGP 协议的设计中，我们的目标是解决 BGP 的路径搜索时间长、路由的传播时间长以及路由策略导致的路由不收敛问题①.

表 2 中的步骤 1 和步骤 2 分别是实现 sBGP 稳定路由选择的核心，解决了目前 BGP 路由的问题. 具体来说，路由决策过程中的这两个步骤通过识别路由变化的故障源以及所有不稳定的路由以实现稳定路由选择的决策. sBGP 通过识别路由变化的故障源，避免了路由器收到撤销消息后搜索无效的路由，所以避免了漫长的路由搜索时间. sBGP 路由器只有在当前路由无效的时候才会进行路由的重新决策，因此 sBGP 路由器每次声明出来的路由总是当前路由中最稳定的，有效避免了路由的反复声明而引起的路由抑制. 类似地，由于 sBGP 路由器只有当当前路由不可用时进行路由重新决策，而决策过程中仅选择自己可选的最稳定路由，因此 sBGP 不会出现由于路由策略问题而导致的路由不收敛问题.

在 sBGP 路由决策步骤 2 中，如果路由的有效时间小于 τ，则采用收到的新申明路由. 我们通过 τ 来进行稳定路由选择的判断. 具体算法如算法 1 所示，我们首先选择路由运行时间最长的路由，而且该路由的有效时间超过了 τ. 如果目前所有路由的有效时间都没有超过 τ，则选择收到的新声明的路由，这时所选择的路由也是稳定的路由. 这个原则也是 BGP 的稳定路由的选择策略，因为 sBGP 有以下的定理.

算法 1. sBGP 路由决策算法.

avail (r): return the available time of route r

contains (u,v): return if the route contains the failed link

输入: Route r is withdrawn or replaced, which is caused by link failure between AS u and AS v

输出: the best route r'

```
1.  //AS determines the current route r* is impacted
      by the failed link or withdrawn
2.  If r*.contains(u,v)==FALSE Then
3.    return;
4.  Else  //AS needs to recomputed the best routers
5.    choose r in RIB-IN with the longest available
      time
6.    while (r!=null) {
7.      If r.contains(u,v)==FALSE Then
8.        r'=r;
9.        break;
10.     Else
11.       re-choose r in RIB-IN with the sub-longest
          available time;
12.     Endif
13.   }
14.   If (avail (r)<τ) Then
15.     //Available time of all routes in RIB-In is less
          than τ
16.     If exists announced route r Then
17.       select the announced route r';
18.     Endif
19.   Endif
20. Endif
```

定理 1. 在 sBGP 中，由于新申明的路由是邻居路由器采用的最长时间的可用路由，所以当本地无可用稳定路由时选择收到的申明路由也是一种稳定路由的选择策略.

在 sBGP 中当 AS 收到上层 AS 所选的稳定路由，则说明这条路由在网络中的稳定时间已经大于 τ. 所以，当路由发生改变而当本地无可选的稳定路由时(即路由有效时间大于 τ)，sBGP 可以选择上层 AS 所选择的稳定路由作为自己的稳定路由. 这个定理比较简单，详细的证明在本文中省略.

3.3 sBGP 实例分析

为了便于与传统的 BGP 比较和分析，我们同样采用同步网络模型来分析 sBGP 的性能. 如图 2 所

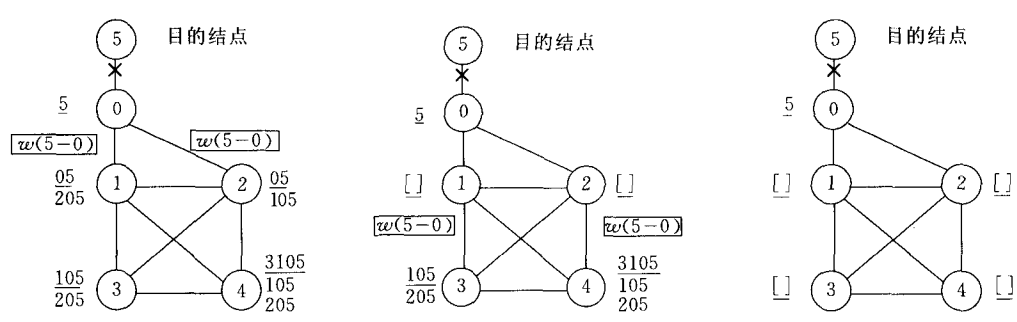


图 2 sBGP 在 B-团拓扑中的收敛过程

① 本文仅关注路由发生变化以后导致的路由不收敛问题.

示，AS5 和 AS0 的链路失效，AS0 发布给 AS1 和 AS2 路由撤销路由通告. 由于撤销通告包含了故障链路信息，AS1 和 AS2 会选择一个不受故障影响的路由. 但是，由于 AS1 和 AS2 中不存在可用路由，所以它们在 $t=1$ 时刻分别发布带有链路故障信息的路由撤销. $t=2$ 时刻，AS3 和 AS4 分别收到来自 AS1 和 AS2 的路由撤销. 同样，由于 AS3 和 AS4 均无可用的稳定路由. 到 $t=3$ 时刻，所有 AS 删除所有故障路由，达到稳定状态. 与传统的 BGP 相比（如图 1 所示），sBGP 在这个拓扑中缩短了 $2+t'$ 的收敛时间.

在发生故障后 sBGP 采用稳定路由选择策略，可以有效避免由于故障导致的路由不收敛问题. 我们采用文献[10]中的拓扑和策略配置实例进行分析和比较. 如图 3 所示，如果 AS0 和 AS4 链路发生故障，传统的 BGP 选择路由会导致各个 AS 将处在不断在 Adj-RIBs-IN 中选择路由而导致路由不收敛[10]. sBGP 很好地解决了这个问题. 当链路发生故障以后，$t=2$ 时刻 AS2 和 AS3 分别收到新的路由通告以后，由于当前路由不受故障影响. AS2 和 AS3 路由不受变化，此时路由收敛.

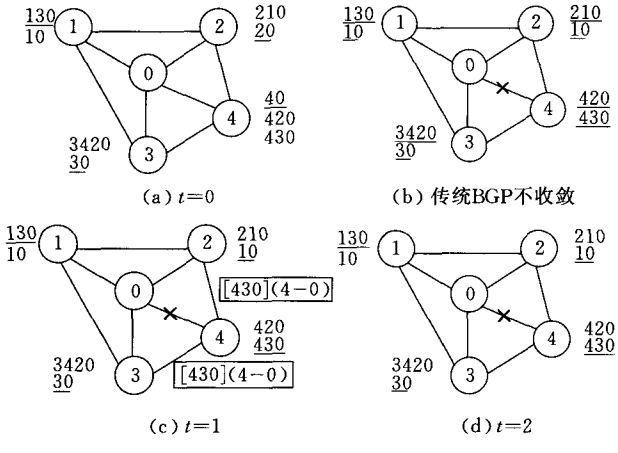


图 3　传统 BGP 和 sBGP 的收敛示意图

3.4　sBGP 讨论

sBGP 在选择稳定路由的过程中提供了路由的可用性. 由于所选择的是稳定路由，sBGP 极大地减少了路由变化次数，消除了路由收敛过程中的无效路由搜索，提高了路由的稳定性. 路由重计算仅在当前所选路由受故障影响的路由器中进行，所以路由故障信息只在有限结点中传播，解决了路由传播时间过长的问题；在 Adj-RIBs-IN 中不存在两个具有相同稳定性的路由，所以 sBGP 解决了路由故障发生后的不收敛问题. 此外，sBGP 可以灵活支持增量部署，解决了 BGP 改进方案的部署复杂性问题.

但是，sBGP 可能会引入路径长度延长以及选择非最优路由问题. 现有研究表明 80%左右的网络故障都是短暂的，一般故障延时都小于 180 s[3-4]. sBGP 会在发生这些故障时快速找到有效路径从而有效避免无效路由的搜索以及路由黑洞和回路. 然而，当发生网络故障以后，sBGP 可能会选择一些 ISP 中设置低优先级的路由. 我们认为从保证全局路由可用性和稳定性的角度来看，sBGP 取得了比较好的折中方案. 特别当故障发生后出现路由不收敛的时候，sBGP 可以选择稳定路由来中止路由不收敛. sBGP 中还实现了恢复正常路由选择机制. sBGP 在选择稳定路由后启动一个稳定选择计时器. 如果路由在计时器超时时间内不发生变化，则计时器超时后路由器将重新启用传统的路由选择算法. 如果计时器超时之前路由发生了变化，则计时器重新启动.

4　性能分析

4.1　实验环境和设置

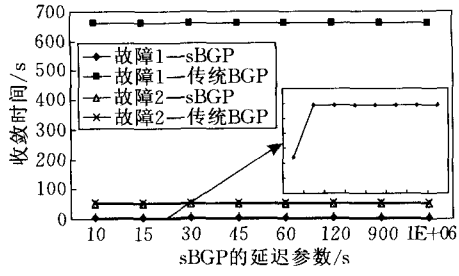
为了评价 sBGP 的性能，我们在模拟软件 SSFNet①的 BGP 代码里实现了 sBGP. 实验中使用了简单网络拓扑和由 SSFNet 提供的真实互联网拓扑. 简单拓扑分别使用了包含 15 个 AS 的二叉树拓扑，5 个 AS 的环拓扑，32 个 AS 的 B-团拓扑，5 个 AS 的线拓扑，6 个 AS 的团拓扑和 16 个 AS 的网格拓扑等几种典型的拓扑. 真实的互联网拓扑采用了包含 6 个 AS，29 个 AS，110 个 AS 以及 208 个 AS 的拓扑，其中后面 3 个拓扑是根据真实互联网中路由信息提出构造[9]. 在路由模拟中，每个 AS 都通告各自的前缀信息. 由于 Ghost Flushing 方案[13]和 Consistency Assertions 方案[12]无法有效避免无效路径的搜索，所以本文将不对这两类方案进行分析和评价. 在不同的拓扑上分别对传统 BGP，RCN 方案[8,11]以及 sBGP 进行了模拟，并分析和评价不同方案的收敛时间以及更新消息数量.

4.2　性能评价

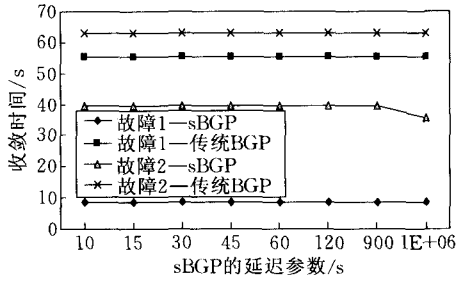
首先，为了分析稳定路由的性能我们分别用不同 τ 取值在 32 个 AS 的 B-团拓扑和 110 个 AS 的真实拓扑进行模拟. 由于模拟环境无法准确刻画互联网的故障特性，因而无法准确学习到 τ 的最优取值. 如图 4 所示，在简单拓扑和真实拓扑中分别只有一个故障的性能受 τ 的设置影响，但收敛时间

① The SSFnet project. http://www.ssfnet.org/homepage.html

的影响仅在 1% 以内. 这是由于发生故障以后，在 Adj-RIBs-IN 中所有的稳定路由比较多. 因此，根据模拟实验可知 τ 的选取对 sBGP 收敛性能影响不大. 在本文的实验中，我们取 τ 为 45 s.



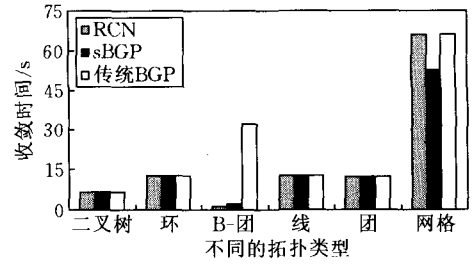
(a) 32个AS网络的收敛性能



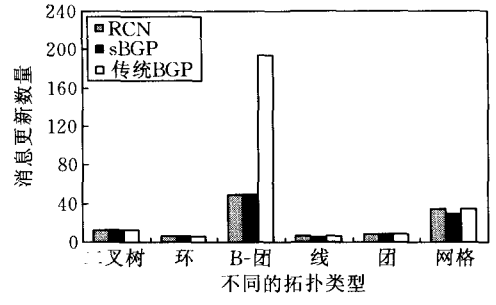
(b)110个AS网络的收敛性能

图 4 不同延迟参数 τ 对 BGP 收敛性能的影响

图 5 分别对不同的简单拓扑的收敛性和更新消息数量进行分模拟. 如图 5(a)所示，RCN 和 sBGP 在 B-团的拓扑中消除了无效路径搜索，收敛时间分别改进了 96% 和 94%. 在网格拓扑中，由于故障链路没有引入无效搜索，RCN 获得了和传统 BGP 同样的性能. 但 sBGP 有效消除了无效的路由计算，因此改进了 21% 的收敛性能. 其它拓扑中不存在无效路由，而且拓扑中不存在很多有效的路由，因此 RCN、sBGP 以及传统 BGP 的收敛性能类似. 图 5(b)为各个拓扑的更新数量，B-团拓扑中，RCN 和 sBGP 都减少了 75% 和的消息数量. 其它拓扑中，3 种方案的更新数量类似. 图 6 显示了不同网络规模的性能. 在 6 个 AS 和 29 个 AS 的拓扑中没有很多可选的稳定路由，RCN、sBGP 和传统 BGP 的性能是类似的. 但随着网络规模的变大，sBGP 的收敛性能改进越多. 例如，在 110 个 AS 和 208 个 AS 的网络拓扑中，sBGP 获得了比较好的性能改进. RCN 获得了传统 BGP 类似的性能，这是由于故障的链路并部没有引入大量的无效路由搜索. 在 110 个 AS 的拓扑中，与 RCN 和传统 BGP 相比，sBGP 的收敛性能分别改进了 77% 和 79%，更新消息数分别减少了 1% 和 31%. 在 208 个 AS 的拓扑中，sBGP 改进了大约 64% 的收敛时间以及减少了 34% 的更新消息数量. 由此可以看出，sBGP 随着网络规模的增加，性能改进的效果越好. 因此，sBGP 能够提高现有互联网路由性能应对网络故障并提高收敛性能.

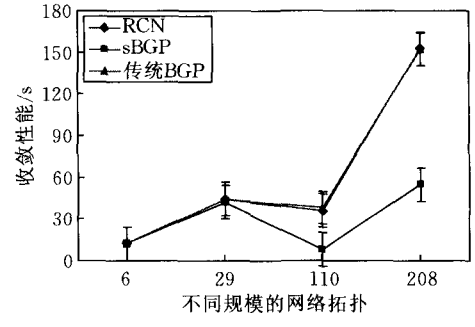


(a) 收敛性能

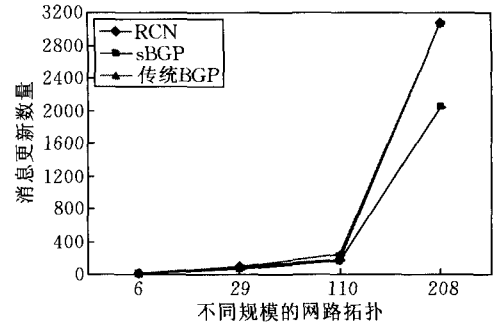


(b) 更新消息数量

图 5 简单拓扑的 BGP 性能比较



(a) 收敛性能



(b) 更新消息数量

图 6 互联网拓扑的 BGP 性能比较

5 结束语

BGP 路由慢收敛问题直接影响互联网路由的性能. 本文提出了稳定路由选择(sBGP)的算法以改

善 BGP 路由收敛. 首先检测了算法路由的有效性. 只有当目前所选路由受故障影响，sBGP 才会重新选择路由. 不同于传统的 BGP，sBGP 选择可选的最稳定路由作为最佳选择. 模型分析和模拟实验结果表明 sBGP 能够大大缩短 BGP 路由收敛时间并且有效减少收敛过程中的 BGP 路由更新消息数量. 目前我们已经在自主研发的虚拟路由器中实现了 sBGP 方案，下一步将在 CERNET2 中进行大规模部署以研究和分析 sBGP 在真实互联网中的性能.

参 考 文 献

[1] Rekhter Y, Li T, Hares S. A border gateway protocol 4 (BGP-4). IETF, RFC 4271, 2006

[2] Labovitz C, Malan G R, Jahanian F. Internet routing instability. IEEE/ACM Transactions on Networking, 1998, 6 (5): 515-527

[3] Markopoulou A, Iannaccone G, Bhattacharyya S, Chuah C-N, Ganjali Y, Diot C. Characterization of failures in an operational IP backbone network. IEEE/ACM Transactions on Networking, 2008, 16(4): 749-762

[4] Turner D, Levchenko K, Snoeren A C, Savage S. California fault lines: Understanding the causes and impact of network failures//Proceedings of the ACM SIGCOMM. New Delhi, India, 2010: 315-326

[5] Labovitz C, Ahuja A, Bose A, Jahanian F. Delayed Internet routing convergence. IEEE/ACM Transactions on Networking, 2001, 9(3): 293-306

[6] Zhang B C, Massey D, Zhang L X. Destination reachability and BGP convergence time//Proceedings of the IEEE Global Telecommunications Conference. Los Angeles: IEEE, 2004: 1383-1389

[7] Kushman N, Kandula S, Katabi D. Can you hear me now?! It must be BGP. ACM SIGCOMM Computer Communication Review, 2007, 27(2): 75-84

[8] Pei D, Azuma M, Massey D, Zhang L X. BGP-RCN: Improving BGP convergence through root cause notification. Computer Networks, 2005, 48(2): 175-194

[9] Chen Wen-Long, Xu Ming-Wei, Yang Yang, Li Qi, Ma Dong-Chao. Virtual network with high performance: VegaNet. Chinese Journal of Computers, 2010, 33(1): 63-73(in Chinese)
(陈文龙，徐明伟，杨扬，李琦，马东超. 高性能虚拟网络 VegaNet. 计算机学报，2010，33(1)：63-73)

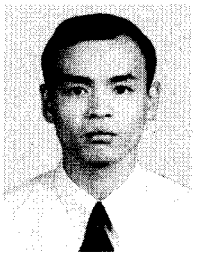
[10] Griffin T G, Shepherd F B, Wilfong G. The stable paths problem and interdomain routing. IEEE/ACM Transactions on Networking, 2002, 10(2): 232-243

[11] Chandrashekar J, Duan Z H, Zhang Z L, Krasky J. Limiting path exploration in BGP//Proceedings of the IEEE Annual Joint Conference of the IEEE Computer and Communications Societies. Miami: IEEE, 2005: 2337-2348

[12] Pei D, Zhao X L, Wang L, Massey D, Mankin A, Wu S F, Zhang LX. Improving BGP convergence through Consistency Assertions//Proceedings of the IEEE Annual Joint Conference of the IEEE Computer and Communications Societies. New York: IEEE, 2002: 902-911

[13] Afek Y, Bremler-Barr A, Schwarz S. Improved BGP convergence via Ghost Flushing. IEEE Journal on Selected Areas in Communications, 2004, 22(10): 1933-1948

[14] Katz D, Ward D. Bidirectional forwarding detection, IETF draft. http: //tools. ietf. org/html/draft-ietf-bfd-base-09



LI Qi, born in 1979, Ph. D.. His research interest includes network architecture and protocol design, system and network security.

XU Ming-Wei, born in 1971, Ph. D. , professor, Ph. D. supervisor. His research interests include computer network architecture, high-speed router architecture and Internet routing.

WU Jian-Ping, born in 1953, Ph. D. , professor, Ph. D. supervisor. His current research interests include computer network architecture, next generation Internet, and protocol testing and formal methods.

Background

Internet routing stability is a key problem in networking research community. In particular, inter-domain routing stability directly impacts availability of Internet. The current inter-domain routing protocol, i. e. , BGP, will experience sever path explorations and incurs many invalid route changes under network failures. During the path explorations, there exist many routing blackholes and loops, which induce lots of packet loss. Although several improved BGP proposals are proposed to mitigate or partly solve the problem, these proposals are unable to deploy in real networks because of their complexity. In this paper, we proposed a stable routing selection scheme (sBGP) which aims to select available routes to ensure successful packet delivery after network failures. The performance evaluation shows that sBGP not only improves the BGP convergence performance, but also reduces computations and communication overheads. The proposed sBGP scheme casts light on designing and improving inter-domain routing in the Internet.

The research is supported by the National Natural Science Foundation of China under Grant Nos. 61073166, 61133015, and 61161140454, the National Basic Research Program of China (973 Program) under Grant Nos. 2009CB320502 and 2012CB315803, and the National High Technology Research and Development Program (863 Program) under Grant No. 2011AA01A101.