第 42 卷 第 6A 期
2015 年 6 月

计 算 机 科 学
Computer Science

Vol. 42 No. 6A
June 2015

iBGP 与集中式路由收敛时间分析

胡乔林 赵国林 刘剑豪 石子言

(空军预警学院 5 系 武汉 430019)

摘 要 iBGP 传播单条最佳路径的机制不能保证多样性路径及协议的正确性,且其路径探索会增加收敛时间,导致域间扰动。通过对分布式 iBGP 和集中式路由控制平台 RCP 的收敛时间进行详细分析,得出了 iBGP 路由协议收敛时间的理论上限值。通过实验证明了理论分析的正确性,集中式路由控制平台可有效降低收敛时间和域间扰动。

关键词 iBGP,瞬时失效,路由控制平台,域间扰动,收敛时间分析

中图法分类号 TP393 文献标识码 A

iBGP and RCP Routing Protocol Convergence Time Analysis

HU Qiao-lin ZHAO Guo-lin LIU Jian-hao SHI Zi-yan

(Department Five, Air Force Radar Academy, Wuhan 430019, China)

Abstract Mechanism of iBGP single best path propagation may prevent router-level path diversity and correctness, which resulting in long convergence time and inter-domain churn. Through the detailed route convergence time analysis for both iBGP and Route Control Platform(MP-RCP), We can get the theoretically upper limit convergence time for iBGP. The experiments prove the correctness of theoretically analysis, which also suggests that MP-RCP reduces convergence time and inter-domain churn.

Keywords iBGP, Transient failure, Route control platform, Inter-domain churn, Convergence time analysis

1 引言

Internet 承载了很多如 VoIP、VPN、远程医疗等对延迟及中断敏感的关键业务,测试发现域间链路失效与域内链路一样经常会出现短时间的失效。然而 Wolfgang 等人表明 AS 内路由器级的冗余路径更为广泛[1],当前 iBGP 协议中根据路径决策过程仅选择单条最佳路径传播,采用全互联、反射器结构,并不能保证各路由器获得充分的路径多样性,甚至在使用多冗余路由反射器的 Tier-1 AS 时,多数路由器仅能获得多条具有相同下一跳的路由。iBGP 隐藏了冗余路径,阻碍了前缀无关的收敛特征,无法将失效在本地进行隔离,这些为网络快速恢复、网络稳定性带来了不利影响。

当前对 iBGP 的研究主要集中于反射器结构下的 iBGP 多路径、路由震荡、环路、热土豆路由等问题。IETF 提出的 Add-Path[2] 可用于 iBGP 中以增强路径多样性,Virginie Van 等[3] 提出了在通告的路径中添加“PATH-DIVERSITY”属性,以指示 AS 内存在备份路径,避免了不必要的撤销消息,但对于快速路由恢复缺乏考虑。对此,部分研究工作提出了 RCP(Route Control Platform)[4,5] 的集中式路由管理机制。基于 RCP 的多路径路由控制平台 MP-RCP[6] 避免了复杂的配置,降低了维护会话的开销,可获得全局路由可视性并为每个客户端路由器分配多路径,从而降低了路由收敛时间以及域间扰动。

上述研究都没有从理论上分析 iBGP 收敛时间及收敛对网络性能的影响,而 Dan Pei 等仅分析了 eBGP 路由收敛[7]。本文侧重于对失效场景下的标准 iBGP、基于 RCP 的 MP-RCP 收敛时间进行理论分析,最后,通过模拟验证了理论的正确性,MP-RCP 在收敛时间、域间扰动方面有较大的改善。

2 iBGP 问题描述

BGP 协议具有 eBGP 和 iBGP 两种模式,iBGP 用于在同一个 AS 内部传递 eBGP 以获得外部可达信息。

本文以 AS 0 为研究对象,$G_I=(V,E_I)$ 表示 AS 0 中的 iBGP 信令图,目标前缀为 d,设 AS 0 中边界路由器 B_i 通过 eBGP 学习到的出口路径为 P_i,其中 P_i 为 B_i 通过 eBGP 学习到的外部路由,B_i 将向本 AS 其他路由器通告路由 P_i,B_i 称为出口点。则该 AS 出口点集合为 $E=\bigcup_{i=1}^{k}B_i$,AS 0 获得的外部可行路径集合为 $X=\bigcup_{i=1}^{k}P_i$。

AS 0 通常使用路由反射器取代全互联模式(本文仅考虑反射器结构),令从客户端到反射器的弧标签为 Up,而反射器之间的标签为 $Over$,从反射器到客户端的路由标签为 $Down$,则反射器结构下合法的信令路径表示为 $[Up]^m[Over]^t[Down]^n$,其中 m 和 n 为任意整数,$t\in\{0,1\}$。为了保证健壮性,通常将路由器与多反射器建立 iBGP 会话以获得冗余性,如图 1 中 B_1、B_2 与多个路由反射器连接,其中各路由器获得

本文受国家级社科基金网络电磁空间作战面临的威胁及对策(12GJ003-144)资助。

胡乔林(1979—),男,博士,主要研究方向为域间路由、虚拟网络、信息作战,E-mail:huqiaolin@gmail.com;赵国林 男,硕士,主要研究方向为信息作战;刘剑豪 男,博士,主要研究方向为信息作战;石子言 女,硕士,主要研究方向为计算机网络。

的路径如图 1 所示，其中 RR_2 按照 IETF 决策过程选择 P_2，这是由于其到达 B_2 的距离更短。

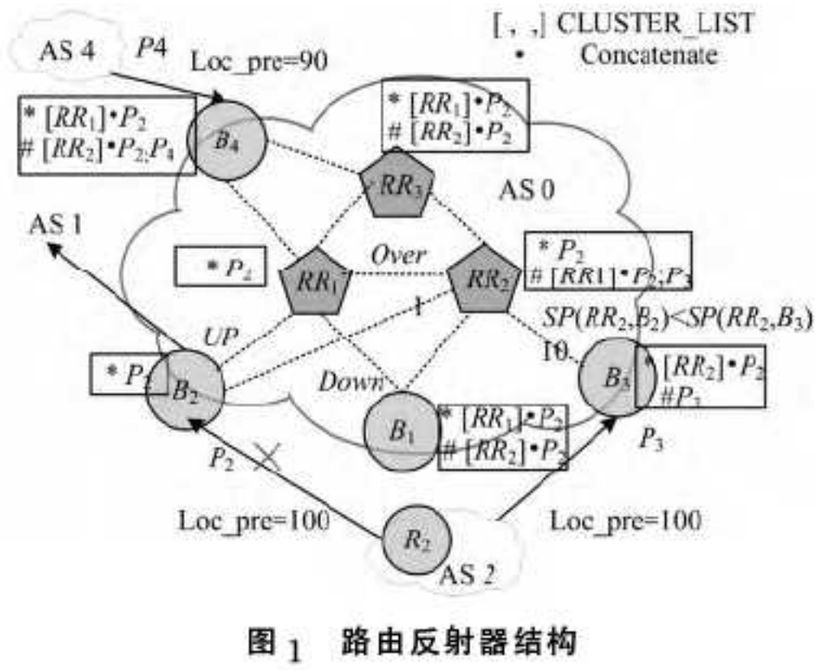

图 1　路由反射器结构

3　iBGP 与 MP-RCP 收敛时间分析

本文主要针对失效切换（T_{long}）场景下分布式 iBGP 与 MP-RCP 的收敛时间进行分析。

3.1　iBGP 收敛时间

如图 1 所示，本文假设 AS 0 中存在多条可行路径，且单链路失效后 AS 0 内部仍然存在合法路径，即失效可限制在 AS 内部处理，并不考虑与 eBGP 相关的更新。

在链路失效场景中，分布式 iBGP 协议收敛时间总结如下。如图 1 所示，当链路〈B_2，R_2〉失效后，节点 B_2 通过 Hello 消息或者 BFD 机制检测到失效，当 B_2 检测到失效后（失效信息也可通过 IGP 传播，这会影响收敛过程中的激活序列，但并不影响最终收敛）将删除 R_2 通告的路由，并启动路由更新发送过程；如果 B_2 最佳路由发生变化，B_2 将根据 rib-in 状态及导出策略发送撤销或者隐式替换路由消息到邻居节点；当 RR_1 接收到 B_2 发送的更新消息后也启动路由更新发送过程，直至最终路由收敛。在收敛过程中由于激活序列的不同，RR_1 可能首先接收到 B_4 通告的 P_4 并将其传递到邻居节点，很明显 P_4 并非最佳路径，这形成了“路径探索”；同样，如果 RR_3 先接收到 RR_1 的撤销消息，RR_3 甚至可能选择无效路由 $[RR_2]\cdot P_2$；直到 RR_2 接收到全局最佳路由 P_3 并通告给其它反射器，然后将最佳路由 P_3 发送到 AS 0 中其它路由器。最终，每个路由器根据 BGP 路由表和 IGP 表递归解析安装到 FIB 中。

令 $node\text{-}cvg(v)$ 表示节点 $v\in$ AS 0 处于稳定状态的时间，分布式 iBGP 协议需要 AS 内部所有节点都稳定才能达到收敛状态，其收敛时间可表示为：

$$iBGP\text{-}Cvg=\max_{v\in AS\,0}\{node\text{-}cvg(v)\} \tag{1}$$

因此，iBGP 收敛时间为从失效检测开始，直至 AS 0 内部最后一个节点进入稳定状态，并更新所有状态后能正常转发，其可以分为如下组成部分：

$$iBGP\text{-}Cvg=FD+UG+(Best\text{-}Path+PV+RU)+FIB+IGP\text{-}Cvg+CRR \tag{2}$$

其中，FD（Failure Detection）为失效检测时间，UG（Update Generate）表示更新报文产生时间。需要说明的是，由于分布式 iBGP 协议中每个节点路由依赖于邻居节点通告的路由，这里采用 $Best\text{-}propagate$ 表示 AS 0 内失效检测节点在产生第一个更新消息一直到最后一个节点 u 接收到失效后最佳路径的时间，本文将在后面具体分析 $Best\text{-}propagate$ 时间。PV（Path Vector）和 RU（Rib Update）分别表示 u 的最佳路径计算时间以及路由表更新时间，FU（FIB Update）表示更新转发表所需要的时间，而 $IGP\text{-}Cvg$（IGP Convergence）表示域内的 IGP 收敛时间，CRR（BGP Prefix Recursive Resolution）指路由器利用 IGP 对 BGP 前缀进行递归解析的时间。

在分布式 iBGP 收敛中，$Best\text{-}propagate$ 主要受两个因素的影响：即显式或隐式撤销消息作废非法路径以及传播新的最佳路径。这些因素受到 AS 中节点当前所处位置以及 rib-in 状态的影响。图 2 中，AS 收敛依赖于 B_3 接收到撤销消息并且将 P_3 通告到距离 B_3 最远的节点，此时收敛时间主要受到传播新的最佳路径因素影响。此外，AS 收敛也可能主要受到撤销消息因素的影响，如图 2 中，假设 RR_2 选择的最佳路径为 P_3，那么 RR_2 将会把 P_3 反射到 RR_1、RR_3，当 P_1 失效后，B_2 将会很快接收到 RR_1 或 RR_2 通告的路径 P_3，但是此时整个 AS 0 并未收敛，仍然需要撤销受影响节点的非法路径。

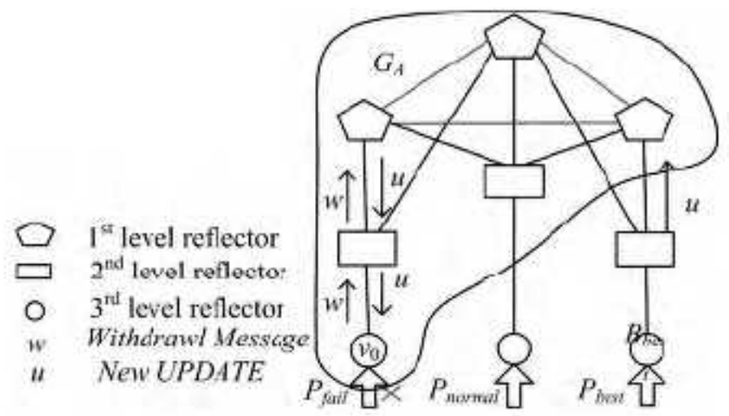

图 2　路径失效后 iBGP 收敛

如图 2 所示，设失效检测节点为 v_0，失效路径为 P_{fail}，$iBGP$ 信令图为 $G_I=(V,E_I)$，域间链路失效将 G_I 中的节点分为受到影响的集合 V_A，以及不受影响的集合 $V-V_A$，其中受到影响的节点是指 rib-in 中存在失效路径 P_{fail} 的节点。同时设 AS 0 中路径 P_{fail} 失效后的最佳路径为 P_{best}，P_{best} 对应的出口点为 B_{best}。

AS 内部最终收敛需要显式或隐式撤销消息传播到 V_A 中所有节点，且 V_A 中所有节点均接收到新的最佳路径。因此，$Best\text{-}Path$ 可以表示为：

$$Best\text{-}Path\leqslant\max_{v\in V_A}\{\max_{v\in V_A}\{withdraw(v\leftarrow v_0),receive(B_{best}\rightarrow v)\}\} \tag{3}$$

其中，$withdraw(v\leftarrow v_0)$ 表示 v_0 产生原始撤销消息以后，v 接收到其它节点传播的撤销消息时间。$receive(B_{best}\rightarrow v)$ 表示本 AS 中出口点 B_{best} 通告最佳路径 P_{best}，并且 v 接收到 P_{best} 的时间。需要注意的是，即使 AS 内部仍然存在失效后的最佳合法路径，也可能产生域间扰动。如图 2 中，P_2 失效后，B_2 发送了非安全撤销消息到 AS 1，但是当收敛之后 B_2 发送到 AS 2 的更新消息中的 AS-PATH 仍然为[AS0，AS2]，这与撤销之前的路由完全一样，这种域间扰动虽然并不影响 AS 0 的收敛，但这是 iBGP 设计应该避免的。因此，实际网络整体收敛时间 $T_{network\text{-}cvg}$ 为：

$$T_{network\text{-}cvg}=\max_{AS\,i\in Affected\text{-}AS}\{Best\text{-}Path_{AS\,i}\} \tag{4}$$

其中，Affected-AS 表示受到影响的 AS 集合。由于 $iBGP$ 与 $eBGP$ 之间的交互关系，如果 $iBGP$ 不能及时收敛，将可能导致失效信息传播到邻居 AS，从而实际上会造成更大的收敛时间。为比较的公平性，本文暂设失效信息不会影响到邻居 AS。

3.1.1　$Best\text{-}Path$ 中最大撤销时间

对于 $\forall v_k\in V_A$，设失效点 v_0 到 v_k 的传播路径为（v_0，v_1，

…,v_k),令 QP 表示邻居节点之间队列延迟 q、链路传播延迟 p 之和,即 $QP=p+q$。令邻居节点 MRAI 定时器等待时间为 M,邻居节点 v_j、v_{j-1} 之间的消息传播总延迟为 $d(v_j, v_{j-1})$,则有:

$$d(v_j, v_{j-1}) = QP_{v_j}^{v_{j-1}} + M_{v_j}^{v_{j-1}} \tag{5}$$

那么可得到 v_k 的原有路径被撤销的时间为:

$$withdraw(v_k \leftarrow v_0) = \sum_{j=k}^{1} d(v_j, v_{j-1}), v_j \in V_A \tag{6}$$

对于 iBGP 信令图而言,外部路径撤销时间为:

$$withdraw_{v \in V_A}(v \leftarrow v_0) \leqslant \max_{v_k \in V_A} \{withdraw(v_k \leftarrow v_0)\} \tag{7}$$

3.1.2 *Best-Path* 中传播新的最佳路径的最大时间

对于 $\forall v_k \in V_A$ 接收到失效后的最佳路径 P_{best},主要可以分为两步:即首先 B_{best} 接收到撤销消息(表示为 $withdraw(B_{best} \leftarrow v_0)$),然后 B_{best} 逐步将消息传递到 v_k(表示为 $ann(B_{best} \rightarrow v_k)$),其中 B_{best} 并不一定在 V_A 中。那么,最终 v_k 接收到最佳路径的时间为:

$$receive(B_{best} \rightarrow v_k) = withdraw(B_{best} \leftarrow v_0) + ann(B_{best} \rightarrow v_k) \tag{8}$$

类似地,$withdraw(B_{best} \leftarrow v_0)$ 也可表示为式(6)的形式,设撤销消息从 v_0 传播到 B_{best} 的路径为 $(v_0, u_1, \cdots, u_m, B_{best})$;而 $ann(B_{best} \rightarrow v_k)$ 表示从 B_{best} 到 v_k 的最佳路径传播过程,令 B_{best} 到 v_k 传播路径为 $(B_{best}, w_1, \cdots, w_n, v_k)$。那么,$withdraw(B_{best} \leftarrow v_0)$、$ann(B_{best} \rightarrow v_k)$ 分别为:

$$withdraw(B_{best} \leftarrow v_0) = \sum_{j=1}^{m-1} d(u_j, u_{j+1}) + d(v_0, u_1) + d(u_m, B_{best}) \tag{9}$$

$$ann(B_{best} \rightarrow v_k) = \sum_{j=1}^{n-1} d(w_j, w_{j+1}) + d(B_{best}, w_1) + d(w_n, v_k) \tag{10}$$

将式(7)、式(10)代入式(8),可得:

$$receive(B_{best} \rightarrow v_k) = (\sum_{j=1}^{m-1} d(u_j, u_{j+1}) + d(v_0, u_1) + d(u_m, B_{best})) + (\sum_{j=1}^{n-1} d(w_j, w_{j+1}) + d(B_{best}, w_1) + d(w_n, v_k)) \tag{11}$$

进而将式(7)、式(11)代入式(3),最终可以得到失效信息不传播到外部 AS 情况下的收敛时间。

3.2 MP-RCP 收敛时间

MP-RCP 结构以及设计机制参见文献[6]。MP-RCP 中失效检测机制与分布式 iBGP 类似,但是客户端需要将失效信息发送给 MP-RCP,由 MP-RCP 为每个客户端重新计算、分发路径;客户端将根据接收到的路径更新路由表。基于 RCP 平台协议的收敛时间类似,MP-RCP 的收敛时间主要包含如下主要部分。

$$MP\text{-}RCP\text{-}Cvg = FD + UG + |V| \cdot PV + \max_{v \in V}(QP_{MP\text{-}RCP}^{v} + M_{MP\text{-}RCP}^{v}) + FIB + IGP\text{-}Cvg + CRR \tag{12}$$

其中,$|V| \cdot PV$ 表示 MP-RCP 为所有客户端进行路径计算所需要的时间,$\max_{v \in V}(QP_{MP\text{-}RCP}^{v} + M_{MP\text{-}RCP}^{v})$ 表示将路由分发到客户的最大延迟。QP 表示发送队列延迟 q 与传播延迟 p 之和,即 $QP=p+q$。M 为 MRAI 定时器时间,其它符号意义与上面相同。

3.3 协议收敛时间对比分析

通过对分布式 iBGP 和 MP-RCP 结构收敛时间进行分析,在这些影响收敛的因素中,FD、UG、FU、PV、RU、CRR 等通常都比较小,且 IGP 收敛可以实现秒级的收敛,即 $IGP\text{-}Cvg$ 也可以忽略。依据上面的公式,分布式 iBGP 与 MP-RCP 收敛时间的主要差异在于 $Best\text{-}Path$ 与 $|V| \cdot PV + \max_{v \in V}(QPM^{v})$ 之间的比较。

由于 AS 内部反射器结构特征,节点接收到最佳路径的最差情况为:失效检测点 v_0 发起的撤销消息依次传播到 AS 内部最远节点 B_{best},且 B_{best} 将最佳路径 P_{best} 再依次通告到 AS 中某个距 B_{best} 最远的节点。令 $G_A(G_A \subseteq G_I)$ 表示受影响节点集 V_A 构成的信令子图,$diameter(G_A)$ 表示信令子图 G_A 的直径。令邻居节点之间更新消息的队列、处理、iBGP MRAI 定时器等待时间之和为 $d(u,v)$,其中 $d(u,v)=QP+M$。同样,记邻居节点之间对于撤销消息的队列、传播、iBGP MRAI 定时器等待时间之和为 $d_w(u,v)$。根据式(8)、式(11),可以得出 $Best\text{-}Path$ 的近似上限:

$$\begin{aligned} Best\text{-}Path &\leqslant withdraw(B_{best} \leftarrow v_0) + ann(B_{best} \rightarrow v_k) \\ &\leqslant diameter(G_A) \cdot d_w(u,v) + diameter(G_A) \cdot d(u,v) \end{aligned} \tag{13}$$

在全互联结构中,iBGP 信令图中的网络直径为 1,此时根据公式可知全互联结构下最大收敛时间为:

$$Best\text{-}Path \leqslant d_w(u,v) + d(u,v) \tag{14}$$

在反射器结构中,iBGP 信令图具有结构化的特征,设反射器级数为 l(信令图中从最底层路由器到达最高层反射器的距离实际为 n,其中 $n=l-1$),则信令图中网络直径为 $\max\{[Up]^i[Over]^t[Down]^j\} \leqslant |n+1+n| = 2n+1$,其中 i、j、n 为正整数,$t \in \{0,1\}$。

(1)当设置撤销速率限制 WRATE 时,$d_w(u,v) \approx d(u,v)$,因此,由式(13)可以得到反射器结构下收敛延迟的上限:

$$Best\text{-}Path \leqslant 2 \cdot diameter(G_A) \cdot d(u,v) = 2 \cdot (2n+1) \cdot d(u,v) \tag{15}$$

(2)当不设置 WRATE 时,由于邻居节点之间撤销消息时间为 $d_w(u,v)$,忽略队列、传播延迟,故 $d_w(u,v) \approx 0$。此时,由式(13),收敛时间可为如下公式:

$$\begin{aligned} Best\text{-}Path &\leqslant diameter(G_A) \cdot d(u,v) \\ &\approx (2n+1) \cdot d(u,v) \approx (2n+1) \cdot iBGP\text{-}MRAI \end{aligned} \tag{16}$$

MP-RCP 结构下收敛时间的最主要部分为 $|V| \cdot PV + \max_{v \in V}(QP_{MP\text{-}RCP}^{v} + M_{MP\text{-}RCP}^{v})$,其收敛时间相对稳定,主要取决于 *MP-RCP* 的计算时间、*MP-RCP* 与客户端之间的 *MRAI* 定时器设置,且 *MP-RCP* 具有全局路由可视性,从而可以避免复杂的路径探索,不会增加消息数量。

4 实验评估

为验证收敛时间理论分析的正确性,本文对 SimBGP[8] 扩展实现了 RFC 4456 中的 iBGP 属性以及完整的路由决策过程,并实现了 MP-RCP 下的服务器端和客户端功能。由于 iBGP 会话建立在多跳 IGP 之上,实验中同时输入了网络物理拓扑,会话延迟设置为会话节点之间最短路径上的链路延迟之和。实验中主要采用的参数如表 1 所列,其中 MP-RCP 处理延迟根据文献[4]进行设置。MP-RCP 失效处理机制设置为失效后再次为每个节点分配路径(路径数为 2)。实验从收敛时间、消息数量方面对全互联、单反射器、冗余双反射器结构(简记为 RRs)、MP-RCP 结构进行对比分析。

表 1　实验参数设定值

参数	默认值
eBGP,iBGP/MP-RCP MRAI	30s,5s(Peer-based)
链路带宽	1000M
链路队列延迟	uniformly [0.01,0.1] ms
FIB 更新延迟	uniformly [0.001,0.01]ms
SSLD,WRATE	True,False
MRAI 抖动	[0.75,1]* MRAI
导入导出策略和优先级	Gao-Rexford Policy
MP-RCP 处理延迟	uniformly [0.001,0.1]s

4.1　Abilene 网络拓扑实验

根据 Abilene(AS 11537) 公开的连接关系，实验中从 68 个在最佳路由表中出现的活跃 AS 中，选择了与 Abilene 具有多连接的所有 27 个活跃邻居 AS 通告路径。实验中输入 iBGP 结构、eBGP 会话、Abilene 的 IGP 拓扑信息。

实验中设置了两种场景：邻居 AS 通告的所有路径具有完全相同的优先级，以及在邻居 AS 通告的路径中设置不同的 MED 值。

(1)对邻居 AS 通告的多路径设置完全相同的优先级。本实验中每次使邻居 AS 在多条链路上通告同一前缀，并将其中的某条链路断开。在分布式 iBGP 实验中设置了 3 种结构：即全互联、以 Kanas 作为单反射器、以 Kans 和 Chic 节点作为冗余双反射器的 iBGP 结构，反射器级数都为 2。容易得知，其全互联 iBGP 会话数量达 36 条，单反射器结构会话数量为 8 条，双冗余反射器 iBGP 会话数为 17 条(所有路由器与两个反射器建立会话)，MP-RCP 会话数为 9 条。

由于所有的路径具有同等优先级，在全互联结构中，所有路由器都具有多样性路径，在单链路失效场景中(见图 3 中收敛时间 CDF 图)，失效检测节点发送的主要是撤销消息，收敛速度最快；与单反射器相比，由于冗余双反射器结构中存在一定的"路径探索"，在一定程度上增加了双反射器结构收敛时间。在 MP-RCP 中，MP-RCP 与客户端会话 MRAI 定时器采用了标准 iBGP 设置，由于需要发送、撤销多路径，相对增加了收敛时间，而实际上由于 MP-RCP 具有全局可视路由，可将 MRAI 设置为更低，如图 3 中 MP-RCP(MRAI=0s，图 3 最左边)所示，在不增加消息数量的情况下，收敛时间大幅降低。

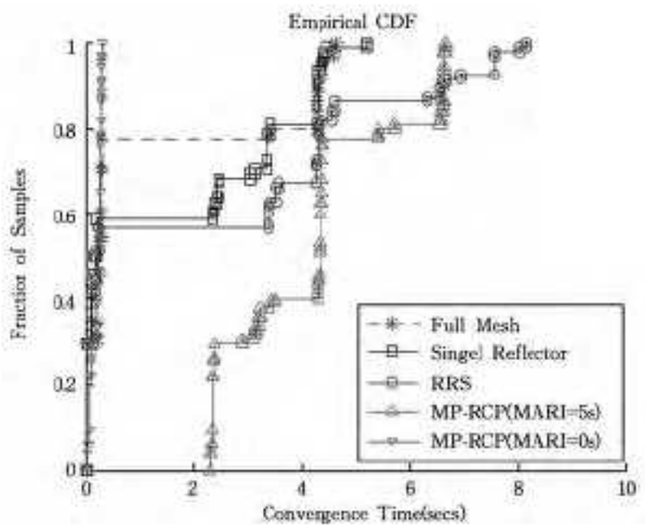

图 3　收敛时间 CDF 图

在单反射器和双反射器结构中，根据式(16)在不设置 WRATE 时的分析，收敛时间上限约为(2×1+1)×iBGP—MRAI=15s。实际上由于 Abilene 拓扑较小，Abilene 获得的多条路径具有同等优先级，使得受到影响的信令图直径缩小到 2 左右，从而降低了收敛时间，其绝大多数小于 15s，这也证明了理论上收敛时间的正确性。

(2)对邻居 AS 通告的多路径设置不同的 MED 值。实验中对同一 AS 通告的多条路径设置不同的 MED 值，这也比较符合实际情况。本实验中同样将发起通告的 AS 的某条路径断开，iBGP 结构设置同上。

当为邻居 AS 通告的多路径设置不同的 MED 值时，将会使全互联、反射器结构下更少的路由器获得多样性路径。如图 4 中收敛时间 CDF 图所示，MP-RCP 由于具有全局可视路由，可对失效进行稳定处理，其收敛时间显著降低，而全互联收敛时间仍然低于两种反射器结构，双反射器收敛时间比单反射器结构略有增加，但两种反射器结构下收敛时间上限都约为(2×1+1)×iBGP-MRAI=15s 左右，实验证明了理论分析上限值的正确性。

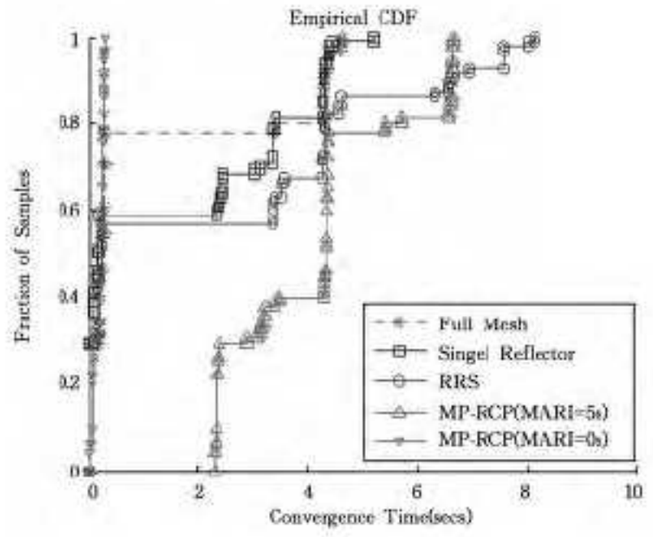

图 4　收敛时间 CDF 图

4.2　生成拓扑实验

实验采用 IGen 拓扑生成器产生的路由器级拓扑，该拓扑具有 AS 级拓扑及商业关系，然后利用 IGen 为每个 AS 生成路由器数量、IGP 拓扑、iBGP 结构以及 eBGP 连接关系，iBGP 拓扑采用了最常见的 ISP 冗余设计架构，遵循了反射器放置的指导原则。实验共采用了 9 个拓扑，每个拓扑 AS 数量约为 200 个左右，路由器数量为 5000 左右。实验从 9 个拓扑中选取 48 个 AS 部署 MP-RCP，为了比较的公平性，选取的 AS 中路由器数量分布在[20,27]之间，反射器级数为 2。实验中由测试节点的邻居 AS 通过多条链路发起路由通告，对于不同的邻居 AS 通告的路径按照"Gao-Rexford 原则"设置优先级，然后随机选择一条链路断开，共进行 250 次实验。

图 5 为网络整体收敛时间 CDF。由于 MP-RCP 具有全局路由可视性，其收敛时间 iBGP-MRAI 相对稳定在 5s 左右。双反射器结构下平均收敛时间比单反射器结构下略微增加。

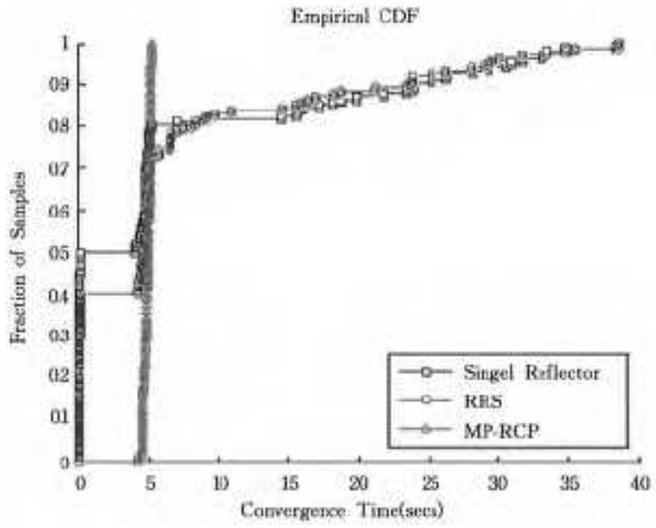

图 5　收敛时间 CDF 图

根据式(16)，生成拓扑的收敛时间上限为(2×2+1)×iBGP-MRAI=25s。实验中，在单、双反射器结构下，约 90% 的收敛时间在 25s 以内，而其中收敛时间较大的实验主要是由于发送了不必要的撤销，导致网络中其它 AS 扰动，这也从实验上证明了理论分析的正确性。通过分析发现，在单、冗余反射器结构下的非安全撤销中，50% 的受到影响的 AS 数量为 2 个，90% 的受到影响的 AS 数量小于 6 个，少数实验中受到影响的 AS 数量超过 10 个。MP-RCP 由于分配多路径以及失效检测、处理分离的机制，有效避免了域间扰动。

结束语　本文首次对 RFC 4456 中 iBGP 协议的收敛时

间进行分析，得出了理论上限值，对于 iBGP 定时器、WRATE 设置以及 iBGP 收敛的研究具有很好的借鉴意义。本文还对基于 RCP 的 MP-RCP 协议进行对比，验证了 MP-RCP 具有较好的性能。

参 考 文 献

[1] Mühlbauer W, Maennel O, Uhlig S. Building an as-topology model that captures route diversity[C]// Proc. of ACM Sigcomm'06. Pisa, Italy: ACM Press, 2006: 195-206

[2] Walton D, Retana A, Chen E, et al. Advertisement of Multiple Paths in BGP[EB/OL]. http://www. draft-walton-bgp-add-paths-06. txt, 2008

[3] van den Schrieck V, Francois P, Pelsser C, et al. Preventing the Unnecessary Propagation of BGP Withdraws[C]// Processing of Information IIFIP International Federation (NETWORKING 2009). LNCS, 2009: 495-508

[4] Caesar M, Caldwell D, Feamster N, et al. Design and Implementation of a Routing Control Platform[C]// Proc. of NSDI '05. Boston, MA Berkeley, CA, USA: USENIX Association, 2005: 15-28

[5] 赵丹. 基于逻辑集中控制的网络路由关键技术研究[D]. 长沙：国防科技大学, 2013

[6] 程柏林，胡乔林，陈新，等. MP-RCP：基于 RCP 的快速恢复 iBGP 协议[J]. 计算机应用与软件, 2014(1): 127-131, 147

[7] Pei D, Zhang Bei-chuan, et al. An analysis of convergence delay in path vector routing protocols[J]. Computer Networks, 2006, 50(3): 398-421

[8] Qiu J. simBGP: a lightweight event-driven BGP simulator[EB/OL]. http://www. bgpvista. com/simbgp. php, 2009

(上接第 265 页)

本文的结论一和结论二基本一致。

3) 混合搜索策略综合利用了宽度优先搜索和深度优先搜索的优势，10 分钟内其平均效率比 Blizzard 高 91.2%，比宽度优先搜索高 64.5%，比深度优先搜索高 27.4%。

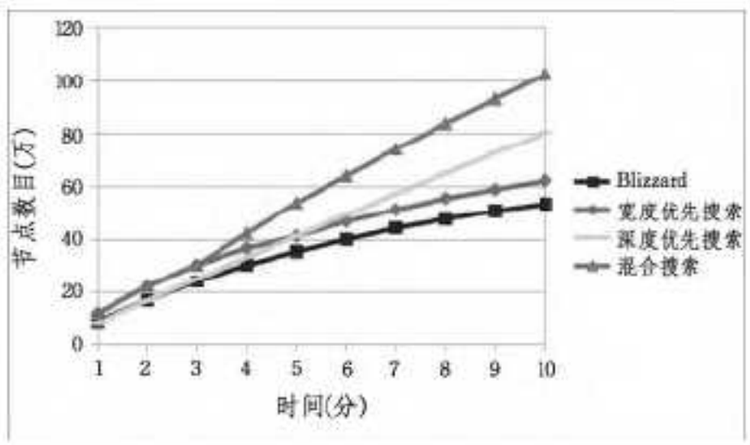

图 2 DHT 网络全局快照采集算法对比

3.3 路由表快照搜索策略对比

通过前期测量可知，Kad 节点的路由表中通常只有 40～120 个联系人。因此我们设置 g 分别为 3、5、7、9(将 g 设置为奇数，是为了将节点 ID 放在 $zone_g$ 的中间，提高检索效率)。为了对比，本文还实现一个随机算法，每次随机选择目标 ID 发送给接收者。

为了减小 DHT 动态变化对实验的影响，本文在每天不同时间共进行 10 组实验，每组实验随机选择 1000 个节点对其进行路由表快照获取。每种策略采集完每个节点路由表所用的路由请求报文平均数如表 2 所列。

表 2 不同搜索策略所使用的报文平均数

算法	随机	g=3	g=5	g=7	g=9
报文数	27.3	13.6	9.5	13.2	16.4

需要注明的是，$g=3$ 时平均使用了 13.6 个报文(上限是 2+4+8=14 个)，对于部分联系人较多的节点，该策略不能完全采集其路由表快照信息。

由表 2 可知：

1) 随机搜索策略的效率最低。这是由于它没有考虑到节点路由表的不均匀特性。

2) 对于 Kad 网络，选择 $g=5$ 具有最佳的路由表快照采集效率，比随机搜索策略高 187.4%，比 $g=7$ 时高 38.9%。

结束语 本文通过分析 DHT 网络的节点分布特性和路由表分布特性，提出了一个采集 DHT 网络全局路由表快照的混合双层模型。在 Kad 网络上的真实实验验证了本文提出算法的有效性。

下一步将使用现有算法测量更多 DHT 网络的全局路由表，分析各个路由表之间的关系、路由表存在的潜在风险，并深入研究各种路由表攻击的检测技术。

参 考 文 献

[1] Stoica I, Morris R, Karger D, et al. Chord: A scalable peer-to-peer lookup service for Internet applications[C]// Proceedings of SIGCOMM'01. 2001: 149-160

[2] Han J, Liu Y. Rumor Riding: Anonymizing Unstructured Peer-to-Peer Systems[M]. IEEE ICNP, Santa Barbara, California, USA, November, 2006

[3] 刘琼，徐鹏，杨海涛，等. Peer-to-Peer 文件共享系统的测量研究[J]. 软件学报, 2006, 17(10): 2131-2140

[4] Maymounkov P, Mazières D. Kademlia: A Peer-to-Peer Information System Based on the XOR Metric[C]// Proceedings of the 1st International Workshop on Peer-to-Peer Systems (IPTPS'02). 2002: 53-65

[5] Bhagwan R, Savage S, Voelker G. Understanding availability[C]// Proceedings of the 2nd International Workshop on Peer-to-Peer Systems (IPTPS'03). 2003: 256-267

[6] Brunner R. A performance evaluation of the Kad-protocol[M]. Master Thesis, 2006

[7] Steiner M, Biersack E W, Ennajjary T. Actively monitoring peers in KAD[C]// Proceedings of the 6th International Workshop on Peer-to-Peer Systems (IPTPS'07). 2007

[8] Guo L, Chen S, Xiao Z, et al. Measurement, analysis, and modeling of BitTorrent-like systems[C]// Proceedings of IMC'05. 2005

[9] Neglia G, Reina G, Zhang H. Availability in BitTorrent Systems [C]// Proceedings of INFOCOM'07. 2007

[10] Jimenez R, Osmani F, Knutsson B. Connectivity Properties of Mainline BitTorrent DHT Nodes[C]// Proceedings of P2P'09. 2009

[11] Yu J, Fang C, Xu J, et al. ID repetition in Kad[C]// Proceedings of IEEE P2P'09. 2009

[12] Yu J, Lu L, Li Z, et al. A simple effective scheme to enhance the capability of web servers using P2P networks[C]// Proceedings of ICPP'10. 2010

[13] 李强，李舟军，周长斌，等. Kad 网络中 Sybil 攻击团体检测技术研究[J]. 计算机研究与发展, 2014, 51(7): 1614-1624

[14] Steiner M, Carra D, Biersack E W. Long term study of peer behavior in the KAD DHT[C]// IEEE/ACM Trans. Networking. 2009