ISSN 1000-9825, CODEN RUXUEW
Journal of Software, Vol.21, No.7, July 2010, pp.1589−1604
doi: 10.3724/SP.J.1001.2010.03816
© by *Institute of Software, the Chinese Academy of Sciences*. All rights reserved.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

互联网无中断转发的生存性路由协议*

苏金树, 胡乔林+, 赵宝康

(国防科学技术大学 计算机学院,湖南 长沙 410073)

Disruption-Free Forwarding Survivable Routing Protocols on Internet

SU Jin-Shu, HU Qiao-Lin+, ZHAO Bao-Kang

(School of Computer, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: huqiaolin@nudt.edu.cn

Su JS, Hu QL, Zhao BK. Disruption-Free forwarding survivable routing protocols on Internet. *Journal of Software*, 2010,21(7):1589−1604. http://www.jos.org.cn/1000-9825/3816.htm

Abstract: Internet is becoming the infrastructure and starts to carry more critical mission traffic where even a short disruption can cause significant losses for certain applications. Nevertheless, traditional route protocols have the problem of long convergence delay, transient unreachability and loop upon the network topology changes due to links/nodes failure or various other reasons. Unfortunately, the transient routing failures are very common according to the experimental studies. Numerous routing protocols which can provide disruption-free forwarding and fast recovery have been proposed. This paper firstly studies the root cause of transient failures, and then presents classification standards for survivable routing protocols. Thereafter, it focuses on analyzing the fundamental mechanism of existing representative survivable routing protocols and comparing their characteristics, performance and overhead. Finally, the current research status and open research issues are concluded.

Key words: transient failure; route recovery; fast rerouting; multipath routing; survivable routing

摘 要: 互联网逐渐成为通信基础设施并承载了更多的关键业务流量,即使瞬时中断也会对某些应用造成巨大损失.然而,传统路由协议在出现链路/节点故障等拓扑变化时存在收敛时间长、瞬时不可达以及环路的问题.实际测量发现,路由瞬时失效相当普遍.因此,研究人员提出多种能够保证流量无中断转发和快速恢复的路由协议.在分析瞬时失效现象以后,提出了生存性路由协议的分类方法,重点对一些重要的路由协议的核心路由机制进行深入分析,并比较其特点、性能、开销等.最后,结合该领域研究现状以及存在的问题,指出未来生存性路由的研究重点.

关键词: 瞬时失效;路由恢复;快速重路由;多路径路由;生存性路由

中图法分类号: TP393 **文献标识码**: A

Internet 已发展成为重要的通信基础设施,承载了更多的关键业务流量.在重要的国防、经济场合,网络瞬时中断都会带来不可估量的损失.ISP 由于网络中断而引起的经济损失每小时以百万美元计算[1],致使网络生存性问题日益突出.为避免巨大损失,需要可用率达 99.999%的网络.然而遗憾的是,目前互联网连两个“9”的指标还

* Supported by the National High-Tech Research and Development Plan of China under Grant No.2008AA01A325 (国家高技术研究发展计划(863)); the National Basic Research Program of China under Grant No.2009CB320503 (国家重点基础研究发展计划(973))

Received 2009-09-08; Revised 2009-11-09; Accepted 2009-12-29

© 中国科学院软件研究所 http://www.jos.org.cn

不能达到[2,3].网络延迟和中断是两个重要的网络性能指标,直接影响着网络新兴数据业务的部署应用,如 VoIP、在线游戏、远程医疗、电子银行等对于持续端到端连接的要求非常高,Cisco 公司预测[4],到 2012 年,Internet 中客户 IP 流量将近 90%都是视频流量,这些流量对于延迟和中断都非常敏感,恢复时间低于 50ms 是寻常要求[5],而恢复时间在 200ms~2s 之间会使实时业务服务质量严重下降.

影响互联网生存性的因素众多,包括底层光网技术、拓扑、体系结构、路由协议等多方面因素,而在 IP 层进行故障处理具有低成本、灵活性等方面的优势[6].在给定拓扑、体系结构的情况下,路由协议是影响互联网生存性的最主要因素,直接影响到互联网承载新兴业务的能力.理想的路由协议应该保证只要底层拓扑连接就能够提供持续的端到端通信,然而,目前 Internet 中广泛采用的 OSPF、BGP 协议缺乏生存性保证,在面临失效时主要通过触发新的全局性、反应式收敛来应对拓扑变化,这种收敛过程对拓扑变化反应迟缓,难以满足 IP 实时业务的需求.比如,按照 RFC 建议设置 OSPF,需 40s~120s 检测失效,收敛时间达数十秒,而 BGP 收敛时间则以分钟计.在路由收敛转换过程中,当前路由协议由于隐藏冗余拓扑、使用单路径、失效全局可视化等问题,当出现链路/节点失效时,路由器不能快速、正确地切换工作路径,造成大量的丢包和服务中断.如何使互联网适应关键业务的无中断需求,成为工业界[7]、学术界共同关注的热点和难点问题,引起了 Sigcomm,NSDI,Infocom 等主流会议的关注.新型网络体系结构中明确提出 Internet 在可用性、可靠性方面存在不足,应针对路由生存性提出解决方案[8],保证网络可靠性以及应对失效的健壮性.“All over IP”的发展趋势更加凸显了生存性路由协议的重要性.

本文主要针对在失效条件下能够提供无中断流量转发、增强互联网生存性的路由协议进行研究,首先阐述了互联网的路由瞬时失效现象,并对现有生存性路由协议进行了分类,然后分别针对单失效、多失效场景深入分析、评价了这些典型路由协议的核心机制、特点、性能,最后讨论并指出未来的研究方向.

1 互联网生存性路由协议研究背景与分类

广泛使用的 OSPF,BGP 路由协议本质上都满足安全性、活性的需求,但是由于协议动态行为以及消息传播处理速度等物理限制,降低了路由更新速度,造成了收敛过程中各个路由器的拓扑视图、RIB、FIB 信息不能及时、同步,从而在网络状态发生改变时就有可能造成瞬时性不可达、路由与转发环路(将其统称为瞬时失效),导致了底层拓扑的连接并不意味着路由系统的可达.

1.1 互联网中瞬时失效现象

现有网络时常面临光缆割断、路由器崩溃、操作失误、攻击、iBGP/eBGP 会话失效等众多威胁,文献[9]通过对 Sprint 骨干网链路故障进行分析发现,失效几乎每天都会发生,且大多数失效属于短时间故障,这些威胁将频繁触发路由失效事件,造成网络的不稳定.

大量的研究表明,现有的域内、域间路由协议在面临失效事件时,包括 OSPF[10],BGP[11]在内的路由协议收敛时间相对较长,BGP 收敛甚至达到 30 分钟以上,而收敛期间的动态性也严重影响数据平面的转发性能[12]:大量路由器将会经历瞬时性不可达、路由环路等问题[13].近年来,通过对 ISP,VoIP 等应用的实际测量,发现 Internet 可用性相对较低,Wang 等人[14]通过对顶级 ISP 测量发现,即使在失效切换和链路恢复时,路由黑洞也会造成数十秒的报文丢失.Zhang 等人[15]指出,超过 39%的路由更新会导致可达性丢失,时间甚至长达 300s.Kushman 等人[16]通过对 VoIP 的分析发现,50%的 VoIP 中断与 BGP 更新高度相关.

路由收敛时间长、收敛期间的瞬时性不可达以及路由环路等问题严重降低了网络性能,难以支持关键、交互式以及实时应用.针对这些问题,学术界广泛开展了设计无环、降低收敛时间的路由协议的研究,文献[17]对 IGP 协议加速收敛的技术进行了总结,BGP 加速收敛的协议包括 BGP-RCN[18],EPIC[19]等.但是,由于大多数失效属于瞬时失效,这类降低收敛时间的协议却可能频繁收敛,造成消息更新风暴,导致出现很高的 CPU 和内存开销的现象,从而造成网络的不稳定[20].而对瞬时失效进行抑制却进一步恶化了路由的不稳定性[21],甚至可能恶化报文分发性能.目前,对 BGP 快速收敛的研究陷入了困境[22].此外,这些加速收敛的协议主要从控制平面上降低失效对于流量转发的影响,但是并不能彻底消除收敛期间对转发中断的影响,不能取得“无中断”的转发.

网络的迅速发展以及应用需求对路由协议提出了更高的要求,本文将生存性路由定义为,网络在遭受攻

© 中国科学院软件研究所 http://www.jos.org.cn

击、故障等各种路由失效事件的情况下,只要底层网络拓扑连接,忽略失效检测时间(双向转发检测 BFD 机制可将失效检测时间降至 15ms 以内),路由协议应该在网络状态改变的情况下快速、正确地恢复,保证流量无中断转发,以增强网络生存性.生存性路由协议需要降低重收敛时间长带来的负面影响,同时需要应对收敛期间中瞬时失效的挑战.如何在给定的环境下高效地利用资源并在性能和开销之间取得平衡,构成了生存性路由的研究核心.

1.2 生存性路由协议的分类

现有路由协议主要采用单路径、最短路径的方法,没有充分利用底层冗余拓扑[23],而缺乏多路径发现、转发能力;路由机制上缺乏故障诊断、隔离能力[24],失效发生时,路由协议难以确定导致路由更新的根源,不能及时作废无效路由.比如 BGP 难以识别非法路径,造成大量“路径探索”[25].以上问题也都影响了协议的生存性,因而需要突破传统路由协议的缺陷,比如利用 RCN[18]机制提供故障诊断能力等.而大多数生存性路由协议主要解决当前协议中存在的某些方面的问题,因此带来了路径建立方式、维护方式、报文转发方式等多个方面的改变.本文主要根据现有解决问题的场景、路径建立和维护时机、失效反应点等方面对生存性路由协议进行分类.

(1) 根据协议覆盖的故障类型,可以分为处理单链路/节点失效以及多个独立失效场景下的路由协议.现有的大多数协议主要集中于保证单链路/节点失效下流量的持续转发(单节点失效可能引起多链路失效,但这些失效是局部性且相关的);而对于独立多失效场景下的协议设计更为复杂,对网络体系结构、路由器实现都有更高的要求.

(2) 根据协议所采用的路径计算方法,可以分为集中式和分布式方法.集中式方法在计算效率、路径一致性等方面具有优势,但其可扩展性、路由收敛时间、通信开销却面临着挑战.RCP[26],Consensus[27]体系结构采用了集中式方法计算、分发路由,强制使各个路由器达到路由的一致性,以应对网络失效情景.

(3) 根据协议对备份路径的建立时间,可以分为先验式预计算和反应式按需计算路径方法,这两种使用备份路径的方法降低了转发对路由收敛的依赖程度.先验式预计算路径通过对所有可能出现的失效,预计算备份路径存储于 RIB,FIB,检测到失效后,立即切换到备份路径;而按需计算路径则主要根据即时的失效信息按需计算新的符合条件的恢复路径.

(4) 根据协议应对失效时反应点的位置,可以分为源端路径级恢复和本地链路级快速重路由(简称本地快速重路由).源端路径级恢复是指源节点在发现故障点以后,整体迅速切换到能够绕过失效点的备份路径;而本地链路级快速重路由是指当前节点发现故障后立即切换到可以绕过故障点的下一跳,而不用源节点切换路径.

此外,根据协议适用范围可以分为域内和域间生存性路由协议;根据报文转发实现方式可以分为基于报文封装和报文头标记修改、接口相关的转发等.生存性路由协议所覆盖的故障类型很大程度上影响着协议复杂性以及协议所采用的具体技术、机制,虽然上述(2)~(4)分类法也能对协议进行较好的归纳,但是为了提取出这些生存性协议的核心思想,特别是为了区分各种协议具体采用的技术、机制所能够适用的场景,本文主要从单/多失效场景的角度出发,根据各种协议的核心思想进行分类和总结,并详细分析了其中典型路由协议的研究成果.

2 单失效场景下生存性路由协议

该类路由协议主要应对目前失效比例较高的单链路/节点失效,当前节点检测到失效后,在无全局协调的情况下,根据自身路由信息库做出本地决策,而不会导致多节点之间的备份路径互相冲突.对失效位置可预知的情况,主要分析了顺序更新 FIB 相关的协议;对失效不可预知的情况,按照备份路径的结构化性质,分析了无结构化路径的域内/间本地快速重路由,以及基于结构化备份子图的本地重路由.

2.1 顺序更新FIB

文献[9]的数据显示,约 20%的故障是进行路由维护造成的,这种维护并非紧急的链路/节点事件,从而为解决瞬时失效问题创造了时间上的机会.Francois 等人针对链路状态协议提出了通过顺序更新 FIB 避免环路的机制 OFIB[28,29].OFIB 的主要思想是:要求将被关闭的链路 l 在收敛期间继续保持连接,确保其他节点在收敛期间

© 中国科学院软件研究所 http://www.jos.org.cn

仍能使用 l 而不会造成报文丢失;其次,强制要求所有路由器按照一定的次序更新 FIB,以破坏环路形成的条件;当收敛完成以后,所有的流量将不会经过 l,此时可以将需要维护的链路、路由器平稳关闭,而不会造成转发中断.

如图 1(a)所示,假设链路 $i\leftrightarrow j$ 将会被关闭,如果 s 到达目标节点 d 最短路径经过链路 $i\rightarrow j$,那么 s 可以将报文转发到其他节点 n.其中,节点 n 到达 d 的最短路径树 SPT 中不含有链路 $i\rightarrow j$.很明显,此时不会形成环路,即 $\forall s,d|P_{s\rightarrow d}=\{s,\ldots,i,j,\ldots,d\}$,且 $\forall s,d|P'_{s\rightarrow d}=\{s,n,\ldots,d\}$ and $i\rightarrow j\notin P^{init}_{n\rightarrow d}$,那么可以得到 $P'_{s\rightarrow d}=P^{final}_{s\rightarrow d}$.其中,$P_{s\rightarrow d}$, $P'_{s\rightarrow d}$, $P^{final}_{s\rightarrow d}$ 分别表示在链路关闭事件发生之前、使用顺序更新 FIB 以及链路关闭之后得到的从节点 s 到 d 的路径.从这个事实中可以得出如下结论:当链路 $i\rightarrow j$ 关闭引起路由收敛时,在收敛期间,如果所有路由器等待在 $rSPT_{i\rightarrow j}(j)$中的子节点更新 FIB 之后再更新 FIB,将不会引起环路.其中,$rSPT_{i\rightarrow j}(j)$是断开 $i\leftrightarrow j$ 后 j 的反向最短路径树.类似地可以得到,如果链路 $i\rightarrow j$ 恢复,路由器应该等待在 $rSPT^{final}_{i\rightarrow j}(j)$ 中的父节点路由器更新 FIB 以后再更新.根据这两个结论,通过将相关的信息编码端入 LSPs(链路状态报文),则很容易设计相应的协议避免环路[28].

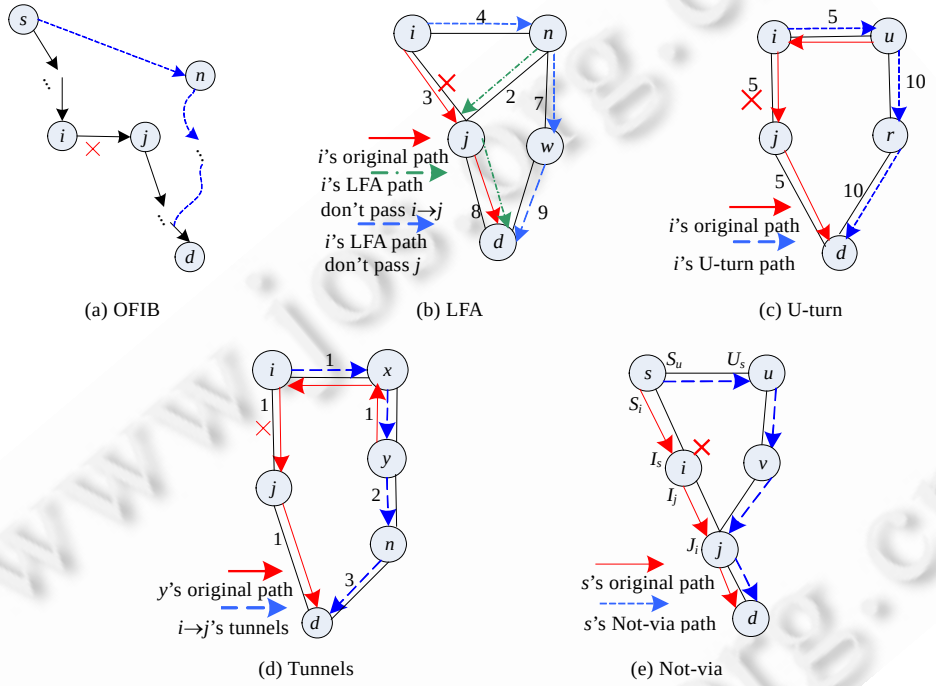


Fig.1 Example of fast local rerouting in IGP

图 1 IGP 中本地快速重路由示例

Francois 进一步针对 OSPF 提出使用步进设置链路权值的方法[29],这种方法不修改 OSPF(open shortest path first)协议,实际上是将预规划链路 $i\leftrightarrow j$ 关闭导致的路由收敛分解为多个子序列,通过步进式地增加链路 $i\leftrightarrow j$ 的权值Δ,每次权值的改变将会引起 OSPF 收敛并逐步更新所有路由器 FIB,且不会导致路由瞬时失效.可以证明,在运行 OSPF 协议的稳定网络中,通过逐步为 $i\leftrightarrow j$ 的权值加 1,必定会得到一个无环路收敛过程.为了减少对链路 $i\leftrightarrow j$ 调整的次数以提高效率,Francois 提出优化重路由权重序列的 ORMS(optimal reroute metric sequence)方法,以使得每次对 $i\leftrightarrow j$ 的调整步进足够大,减少调整次数,使得该方法实用可行.

OFIB 主要针对有规划的路由维护,由于强制所有路由器按照顺序更新 FIB,避免了转发环路,保证了无中断转发,但同时也会明显延长收敛时间.

2.2 域内本地快速重路由

链路/节点失效更多地是由意外故障引起的,IETF RTGWG 以及其他研究者致力于 IGP 协议的本地快速重

© 中国科学院软件研究所 http://www.jos.org.cn

路由[30],并提交了一些包括 ECMP 的草案,其应对失效场景主要包括单链路/节点、共享风险链路组 SRLG 等.这些研究主要采用预计算备份下一跳以应对当前下一跳的故障,其区别主要表现在对备份下一跳的计算方式上,比如,LFA[31]仅利用一跳内的节点信息计算备份下一跳,而 U-turn[32],Tunnels[33],Not-via[34],Deflections[35],FIR[36]等为了增加覆盖范围,考虑利用多跳外的节点信息预计算备份下一跳.而 SafeGuard[37]则采用报文显示携带信息的方法,根据不同携带信息,分别利用反应式或先验式预计算路径发起本地重路由.

(1) 无环备份(LFA)[31].LFA 为节点计算一个无环的备份下一跳.如图 1(b)所示,假设节点 i 经过 $i\to j$ 到达目标 d,当 LFA 对链路 $i\to j$ 进行保护时,将为默认下一跳 j 选择备份下一跳 n,其条件满足 $Cost(n\to\ldots d)<Cost(n\to\ldots i)+Cost(i\to\ldots D)$,表示 n 到达目标 d 的路径不会经过 $i\to j$,且 n 到达 d 的距离比 i 到达 d 的距离更短.由于域内路由协议采用最短路径,该条件也等价于$(i\to j)\notin SP(n,d)$.其中,$Cost(n\to\ldots d)$表示节点 n 到达 d 的最短距离,$SP(n,d)$表示节点 n 到达 d 最短路径集合.i 将流量转发到不受链路失效影响的备份节点 n 后,由于 n 到达 d 的距离更短,因此不会再经过 $i\to j$,从而避免环路.LFA 在保护节点时,其备份下一跳的选择条件类似于链路保护.

(2) U-turn[32].在某些拓扑中,LFA 不能计算得到备份下一跳节点,U-turn 对 LFA 的限制条件放松,可选择两跳之内节点作为备份下一跳,如图 1(c)所示.即,如果 i 的邻居节点 u 满足条件 $u\in N(i)$ and $r\in N(u)$ and $(i\to j)\notin SPT(r)$,那么节点 u 可用来保护链路 $i\to j$.其中,$N(i)$表示节点 i 的邻居节点集合,$SPT(r)$为以 r 为根节点的最短路径树.可理解为,如果 i 的邻居节点 u 具有 LFA 下一跳 r,则当 $i\to j$ 失效后,i 采用路径 $i\to u\to r$ 切换流量,以到达目标 d.但是在这种情况下,节点 i 和 u 之间可能会造成转发环路.因此,路由器 u 必须具有区分流量模式的机制.

(3) 隧道 Tunnels[33].在某些拓扑中,LFA 和 U-turn 都不能保证完全覆盖,而使用 IP 隧道为节点之间创建虚拟链路,可更进一步放松对备份下一跳节点的限制条件,隧道机制对报文封装使报文不受必须在最短路径上转发的限制.如图 1(d)所示,为了保护链路 $i\to j$,路由器 i 可以通过隧道将报文转发到路由器 n,路由器 n 满足条件 $(i\to j)\notin SPT(n)$ and $(i\to j)\notin SP(i,n)$.其中,$SP(i,n)$表示节点 i 到达 n 的最短路径集合.文献[33]也提出了通过计算集合 *F-space* 和 *G-space* 而得到可行路由器 n 的方法.其中,*F-space* 为 i 不经过 $i\to j$ 到达目标的节点集合,*G-space* 为不经过 $i\to j$ 到达 j 的节点集合,*F-space* 和 *G-space* 的交集即为可行隧道末端点,然而隧道也不能保证 100%的保护.

(4) Not-via 地址[34].Not-via 是对隧道机制的扩展,Not-via 的思想表现在:LFA,U-turn,Tunnels 都缺乏指示失效点的能力,不能显式地对需要重路由的报文进行操作以绕过失效点;而 Not-via 通过对报文封装,并显式地指示其他路由器在转发该报文时应该绕开失效点.

Not-via 要求网络中所有路由器合作,路由器将会在链路状态报文中为每个邻居通告一个 Not-via 地址,而其他路由器接收到具有 Not-via 计算请求的链路状态报文后,必须为每个特定的 Not-via 地址计算对应的 FIB,该 FIB 入口是通过移除 Not-via 对应的组件(链路/节点)后计算最短路径树所得到的.如图 1(e)所示,假设 s 经过 $i\to j$ 到达 d,s 检测到 i 失效后,i 将报文通过隧道封装转发到特定地址 J_i,实现无中断转发,其中 J_i 为特殊 Not-via 地址,其语义表示网络中所有节点假设移除 Not-via 地址所对应的组件后计算 FIB 所得到的备份下一跳.

Not-via 也可用于距离向量和路径向量协议,比如 RIP 和 BGP,此时需要将 Not-via 地址的通告方式进行修改,以压制不必要的 Not-via 地址传播.Not-via 的优点在于:只要底层拓扑保证连通,Not-via 就可以达到 100%的覆盖,并且支持组播和 SRLGs;但其缺点比较明显:全网需要通告大量 Not-via 地址,依赖于隧道机制,处理开销较大,容易带来次优路径的问题.Li 等人[38]对 Not-via 进行了改进,通过使用 Not-via 地址聚合、基于优先级的 Not-via 计算以及 rNotvia 算法降低计算和存储开销.

(5) 路由偏转(routing deflection)[35].路由偏转是一种允许节点通过设置标签选择非最短路径的机制.在路由器使用偏转机制时,如果出现故障链路/节点,路由器通过设置标签作为指示进行快速本地重路由,将报文偏转到绕开最短路径上的故障链路和节点.路由偏转主要包含两个部分:1) 偏转规则,决定了哪个邻居节点可用于偏转报文;2) 信令机制,使得端系统能够控制路径上的哪些路由器可以用于转发报文.其中,3 个逐步松弛的偏转规则主要用于计算不会导致环路的多个备份下一跳(称为偏转节点集合).令节点 n_i 表示当前路径上的节

© 中国科学院软件研究所 http://www.jos.org.cn

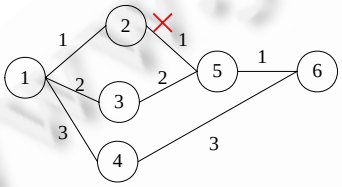
点,n_{i+1} 为 n_i 的偏转集合中的节点,$cost(n_i,d)$表示从当前节点 n_i 到达目的节点 d 的最短距离.

规则 1. n_i 的偏转节点集合中的节点 n_{i+1} 满足条件 $cost(n_{i+1},d)<cost(n_i,d)$.该规则等价于著名的无环条件 LFI,与 n_i 相比,偏转节点 n_{i+1} 到达目标节点的距离更短,以保证无环路.规则 1 与最短路相比产生了更多的可用路径.

规则 2. $cost(n_{i+1},d)<cost(n_i,d)$或 $cost(n_{i+1},d)<cost(n_{i-1},d)$,相对于规则 1,规则 2 放松了规则 1 的限制条件,允许当前节点 n_i 的上一跳 n_{i-1} 用作其下一跳,以临时允许增加开销,并且保证在下一跳能够充分地递减.因此,规则 2 产生的路径可能会使路径变长,但规则 2 能够获得足够多的多样性路径,提高了失效恢复能力.

规则 3. 对于任意的 $n_{i+1}\neq n_{i-1}$,$cost(G\backslash l_{i+1},n_{i+1},d)<cost(G\backslash l_i,n_i,d)$或 $cost(G\backslash l_{i+1},n_{i+1},d)<cost(G_i,n_{i-1},d)$.其中,$l_i$ 表示 n_{i-1} 和 n_i 之间的链路,$cost(G\backslash l_i,n_i,d)$表示在 $G\backslash l_i$(从拓扑 G 中删除 l_i)中 n_i 到达 d 的最短距离,其他符号语义类似.规则 3 比规则 2 更加灵活,进一步增加了路径的多样性.

(6) 故障不敏感路由 FIR(failure insensitive routing)[36]和失效推断快速路由 $FIFR_N$(failure inferencing based fast rerouting)[39].FIR 并不试图消除路由收敛中可能出现的环路,而是通过推断的方法判断故障链路.比如在图 2(a)中,节点 1 经过节点 2 发送报文到节点 6,如果链路 2→5 产生故障,在常规 OSPF 协议中,节点 1 和节点 2 之间的 FIB 不一致将导致环路.而在 FIR 中,节点 2 检测到失效后一段时间内将压制全局链路状态通告,并发起本地重路由,将报文发送到节点 1,节点 1 接收到从节点 2 发送的到达节点 6 的报文(在无故障发生时,节点 2 不会经由节点 1 到达节点 6),节点 1 推断必定是链路 2→5 或 5→6 失效,通过预计算的接口相关转发(interface-specific forwarding)技术,节点 1 将把报文转发到节点 4.FIR 中报文转发需要同时根据目标地址和报文入口决定,如图 2(b)中的转发表中斜体字部分.FIR 的关键算法在于识别关键链路集合(key link) $K^d_{j\to i}$,$K^d_{j\to i}$ 中的任意链路失效将会造成目标为 d 的报文从节点 j 传输到 i.而任意的边 $u\to v$ 属于关键链路集合,仅当该链路同时满足以下两个条件:$u\to v$ 未失效时,j 为 i 的下一跳;当 $u\to v$ 失效后,有向边 $j\to i$ 位于从 u 到 d 或者从 v 到 d 的最短路径上.通过求解关键链路集合即可构造接口相关转发表.



(a) Topology
(a) 拓扑示例

Routing table					
Dest	2	3	4	5	**6**
Next hop	2	3	4	2	***2***
Forwarding table					
2→1	—	3	4	3	**4**
3→1	2	—	4	2	2
4→1	2	3	—	2	2

(b) Routing table and forwardig table at node 1
(b) 节点 1 路由表和转发表

Fig.2 An example illustrating FIR

图 2 FIR 示例

$FIFR_N$ 对 FIR 进行了扩展,使其能够在节点失效的情况下保证无中断转发,$FIFR_N$ 需要求解关键节点集合.关键节点集合中的任意节点 u 满足以下两个条件:u 未失效时,j 为 i 的下一跳;u 失效以后,边 $j\to i$ 位于从 u 到 d 的反向最短路径上.通过计算关键节点集合即可构造接口相关的转发表.

FIR 和 $FIFR_N$ 的优点是,不修改现有路由协议以及报文格式,通过失效推断技术保证无中断转发.其缺点在于,不能保证 100%的覆盖,缺乏对 SRLG 的支持,可能采用次优路径,推断中需多次进行最短路径树的计算,实现较复杂.

(7) SafeGuard 路由[37].SafeGuard 报文中携带到达目的节点的剩余路径开销(remaining path cost)以辅助路由.在链路状态协议中,如果路由器 n_{i-1} 与其下行路由器的路由具有一致性,那么从 n_{i-1} 经过 n_i 到达目标节点 d 的开销满足条件 $cost(n_{i-1},d)=cost(n_i,d)+cost(n_{i-1},n_i)$.其中,$cost(n_{i-1},n_i)$表示从 n_{i-1} 到达 n_i 的最短路径开销.而任意的链路失效事件必定会导致网络拓扑变小,使得到达目标的开销将不低于未失效之前的开销.假定路由器 n_{i-1} 以 n_i 为下一跳发送到达目标节点 d 的报文 pkt,pkt 报文中剩余路径开销与路由器 n_i 到达目标 d 的 $cost(n_i,d)$将会出现 3 种情景:1) $pkt.cost=cost(n_i,d)$,表示 n_{i-1} 和 n_i 路由一致,路由器 n_i 将按照正常模式转发到 n_{i+1},并更新报文的

© 中国科学院软件研究所 http://www.jos.org.cn

剩余路径开销为 $pkt.cost=pkt.cost-cost(n_i,n_{i+1})$;2) $pkt.cost<cost(n_i,d)$,说明 n_{i-1} 与 n_i 之间路由不一致,网络此时处于收敛状态,路由器 n_i 应该具有更新的网络状态,n_i 将报文转发到默认下一跳 n_{i+1},并且更新报文剩余路径开销为 $pkt.cost=cost(n_i,d)-cost(n_i,n_{i+1})$;3) $pkt.cost>cost(n_i,d)$,说明 n_i 具有比 n_{i-1} 到达目标更低的开销,此时,n_i 不能继续按照默认下一跳进行转发,n_i 将按照备份路径数据库进行转发到 n'_{i+1},并且更新报文剩余路径开销为

$$pkt.cost=pkt.cost-cost(n_i,n'_{i+1}).$$

通过对这 3 种模式的处理,SafeGuard 能够在出现单链路失效的时候快速发起重路由,并加速网络收敛.

2.3 域间本地快速重路由

囿于 Internet 中 AS 的规模,由于顺序更新 FIB 的方法需要所有 AS 合作、计算量大、收敛时间延迟而使其在 BGP 中不可行.当前,应对域间链路失效主要使用本地快速重路由方法.本质上,域间快速重路由与域内快速重路由的原理相同且能够通用,主要区别在于备份路径的计算方法.BGP 协议中,链路无权值、无全局拓扑、单条最佳路径通告、典型的 Valley-free 策略[40]等阻碍了对更多冗余路径的发现,因此,IETF RTGWG 针对 IGP 的快速重路由也就不能直接应用于 BGP.考虑到当前 BGP 路由表、更新消息过多,已经影响 BGP 的稳定性,域间快速重路由协议必须限制开销,也进一步加剧了设计难度.

(1) 域间链路保护隧道[41].Bonaventure 等人[41]提出了为每条域间链路预先建立最佳 IP 保护隧道的方法,应对域间链路失效,该 IP 隧道的下一跳可以作为被保护链路的备份下一跳,主要使用主出口-次出口 pe-se(primary egress-secondary egress)隧道保护两个 AS 之间的并行链路,使用主出口-次入口 pe-se(primary egress-secondary ingress)隧道保护末端(stub)AS 与提供者(provider)AS 之间的域间链路.当域间链路失效时,路由器能够使用 IP 隧道发起本地快速重路由.这种在邻居 AS 之间预建立隧道的方法可以应对 AS 间的单链路失效,但不能应对跨 AS 域的失效,不具有通用性,而且需要设置带外的域间隧道.

(2) R-BGP[42].R-BGP 在 BGP 收敛过程中使用预计算的备份路径保证无中断转发,然而,在 BGP 协议中,使用合适的备份路径却需面临 3 个挑战:备份路径的选择分发、避免环路以及收敛后停止使用备份路径.对于第 1 个问题,R-BGP 使每个 AS 尽量为其最佳路径的下一跳通告一条到达目标前缀的最大不相交路径来加以解决;R-BGP 采用 BGP-RCN[18]的思想在每个路由更新消息中嵌入失效根源信息 RCN(root cause notification),以避免收敛过程中的环路并加速收敛;通过利用 AS 之间的层次关系,如果 AS 显式地接收到所有邻居的撤销消息,该 AS 可以判断自身将不会具有服从策略的路径,从而停止转发,这样就解决了第 3 个问题.R-BGP 降低了收敛延迟,但没有考虑 iBGP 相关的失效事件,且修改了 BGP 语义.

(3) BRAP[43].BRAP 保证当出现单条域间链路失效时,通过备份感知路由协议 BRAP 以执行非停止的路由,BRAP 的主要思想是:每个域间路由器 A 除通告最佳路径以外,在服从策略限制下必须具有以下通告能力:① 通告反向路径的能力,反向路径定义为不经过 A 的默认下一跳 B 到达目标的路径;② 通告无环备份路径的能力,无环备份路径定义为 A 到上行邻居的临时路径.假设路由器 B 具有最佳路径和反向路径,那么当 B 发现最佳路径中的下一跳域间链路失效时,通过使用反向路径即可进行快速重路由;而当出现 iBGP 链路失效以及恢复时,使用无环备份路径发起本地快速重路由.BRAP 放松了最大不相交路径的条件,对实际的部署进行了考虑,但其缺点是增加了一定数量的路由更新消息,修改了 BGP 路由协议.

(4) D-BGP[44].D-BGP 通过通告多路径的方法以增加域间路径多样性,在单链路失效后进行快速本地重路由.D-BGP 要求域间路由器除了最佳路径 $\mathcal{P}_{best}$ 以外,还通告一条与最佳路径最大不相交的备份路径 $\mathcal{Q}_{alt}$.

D-BGP 使用两个限制条件以减少通告消息以及存储开销.对于当前节点 u,假设 u 的所有邻居节点已经向 u 通告了其最佳路径以及备份路径,用 $\mathcal{P}_i$ 表示从邻居 i 学习到的最佳路由,$\mathcal{Q}_i$ 表示从邻居 i 学习到的备份路径,在这些学习到的路径中,u 选择最佳路径 $\mathcal{P}_{best}$ 以及备份路径 $\mathcal{Q}_{alt}$,使用 $|\mathcal{P}_a\cap\mathcal{P}_b|$ 表示路径 $\mathcal{P}_a$ 和 $\mathcal{P}_b$ 共享的边/节点数目.如果满足以下任意条件之一,则节点 u 将不会给邻居节点通告备份路径 $\mathcal{Q}_{alt}$:① 如果 u 具有 v 的最佳路径 $\mathcal{P}_v$ 或者备份路径 $\mathcal{Q}_v$;$|\mathcal{P}_{best}\cap\mathcal{P}_v|=0$ 或者 $|\mathcal{P}_{best}\cap\mathcal{Q}_v|=0$,则表示邻居节点 v 已经有了一条不相交路径;② 如果 $|\mathcal{P}_{best}\cap\mathcal{Q}_{alt}|<|\mathcal{Q}_{alt}\cap\mathcal{P}_v|$,则表示 u 通告的备份路径不能增加 v 的备份路径的不相交度.通过使用这两个条件可以减少通告消息数量,D-BGP 同时也使用 RCN[18]机制加速收敛以及避免环路.

© 中国科学院软件研究所 http://www.jos.org.cn

与域间链路保护隧道[41]的方法不同,R-BGP,BRAP,D-BGP 协议的核心思想较为相似,都是为 AS 通告一条与最佳路径不相交的 AS 级备份路径,这是一种无结构路径的方法,需要知道故障点和报文封装机制.

(5) 一致性路由(consensus routing)[27].Consensus 路由协议借鉴了分布式系统中利用快照达到全局一致性状态的经验,从而使得路由行为更可预测、更安全.Consensus 路由将域间路由协议的安全性和活性分离:安全性意味着路由器严格按照上行路由器采用的路径进行转发报文,除非遇到链路失效;而活性将保证快速响应失效、策略的变化.Consensus 路由通过要求每个时间段内 BGP 路由器达到全局稳定状态.与传统的 BGP 相比,Consensus 路由有意延迟了 BGP 的更新.Consensus 路由在报文转发时采用了两种逻辑上分离的模式:稳定模式和瞬时模式.处于稳定模式时,所有路由器具有全局一致性的路由视图;当由于链路失效等造成报文不能正常转发时,将转入瞬时转发模式,主要采用包括偏转(deflection)、绕道(detouring)、回退(backtracking)、备份(backup)路由机制发起本地重路由.

2.4 单失效下结构化备份子图

在域内快速重路由中,各个节点独立依据自身策略从多条可行路径中选择某条备份恢复路径.然而,该备份路径并不为其他节点所知,各节点的备份路径不具有全局一致性,在某些情况下甚至可能出现环路,本地重路由往往需要额外的信令机制.而结构化拓扑是指在特定的某个备份子图上所有节点具有一致性视图,其核心思想是,对原始图 G 抽取出多个符合条件的子图 G_i,独立在这些子图 G_i 上运行路由协议,其 G_i 中所有节点的路径、节点转发行为都是一致的.目前的结构化备份子图方法需要具有全局拓扑知识,而在 BGP 中,这一假设并不成立,BGP 中实现结构化备份子图较为困难,有待进一步的研究,目前,结构化备份子图方法主要适用于域内协议.

(1) 弹性路由层 RRL(resilient routing layer)[45].RRL 通过对需要保护的链路 l、节点 n 在原始拓扑 G 上删除 l、与 n 关联的链路后分别得到诱导拓扑子图,称为“层”,即每“层”中包含原始拓扑所有节点,但是仅包含链路子集.“节点 n 安全”表示该节点 n 在某“层”L_i 中仅有与该节点关联的一条链路包含在该“层”L_i 中,而 L_i 称为“节点 n 的安全层”.“层”L_i 中所有的节点依然保持连接,且 L_i 具有如下属性:在节点 n 的安全层中,n 不会用于传输流量;如果 n 故障,那么任何对于 n 安全的层中的节点对之间的路径仍然连通;如果 n 故障,任何以 n 为目标节点的流量都将会丢失.在 RRL 中,每个 L_i 都独立计算 RIB 和 FIB,当链路/节点故障时,当前路由器可以透明地将报文快速重路由到对应的“安全层”进行转发,且不会造成环路.

(2) 多路由配置 MRC(multiple routing configuration)[46].MRC 对 RRL 进一步加以改进,是一种自定义的结构化 RRL,与 RRL 的区别在于,MRC 通过限制链路权重以获得备份拓扑,图 3 表现了 MRC 的相关概念.MRC 通过在原始拓扑 $G=(V,E)$中对需要保护的节点 n 和链路 l 进行限制和隔离而得到对应的备份拓扑 C_p.链路 l 在备份拓扑 C_p 的被隔离是指其权值 $w_p(l)=\infty$,l 被限制是指 $w_p(l)=|E|\times w_{\max}$.其中,$w_{\max}$ 表示图 G 中的最大权值,通过对与节点 u 关联的链路进行隔离和限制,可以将节点 u 隔离,即如果节点 u 在 C_p 中满足条件:$\forall(u,v)\in E, w_p(u,v)\geqslant |E|\times w_{\max}$,且$\exists(u,v)\in E, w_p(u,v)<\infty$,此时,节点 u 即被隔离.可以理解为:边 l 被隔离,表示任何流量都不会经过 l 传输,链路 l 被限制,表示所有节点仅能使用 l 作为最后一跳到达被隔离的节点.除了在每个备份拓扑 C_p 中至少保留一条链路仅被限制以外,隔离节点 u 的所有相关的链路也将被隔离,使得其他节点能够通过限制链路到达节点 u.很明显,在 C_p 任意其他节点不会经过 u 到达目标.在 MRC 中,当检测到失效后,将流量重路由到对应的链路/节点被隔离的备份拓扑中,为避免环路,强制仅能够切换一次备份拓扑.MRC 可以较方便地在多拓扑路由 MTR[47] RFC 4915 中实现,易于控制备份路径.其缺点在于,需要计算大量的恢复拓扑.此外,Apostolopoulos[48]也对多拓扑路由进行了研究,并且对降低备份拓扑的数量问题进行了考虑.

(3) IP 冗余树 IPRT(IP redundant tree)[49].在传统的链路状态协议中,所有到达目标节点 D 的节点将形成最短路径树,IPRT 通过构造以目标节点 D 为根的两个冗余树 SPT^1 和 SPT^2,并对每个路由器的 FIB 进行扩展,使得每个路由器具有两个到达目标节点的下一跳.比如,当 SPT^1 下一跳的链路/节点失效以后,可以快速重路由到 SPT^2,以保证流量无中断转发.其中,冗余树的构造过程可以采用经典的图论算法.

(4) MARAs[50].与传统的单路径路由算法构造的以目标节点为根的有向无环图 DAG(即最短路径树)不同,由于 DAG 构造实际上等价于对网络图中的节点进行拓扑排序,而 MARAs 采用了 MA 排序法构造的 MARA

© 中国科学院软件研究所 http://www.jos.org.cn

DAG 包含网络图中的所有的符合约束的边,从而为所有节点到达目标节点提供了大量的备份路径.MARAs 引入了新的多路径概念,即多对一最大连接度、多对一最大流、多对一最大化最短备份路径树.MARAs 采用了集中式的方法对所有的节点进行排序,所有节点获得了大量的一致备份路径,从而可以进行快速的失效恢复.

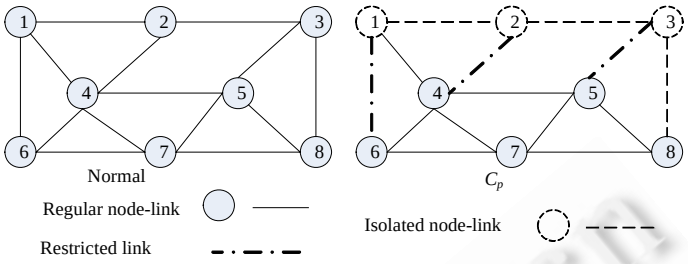


Fig.3 An example topology with one backup configuration in MRC

图 3 MRC 中具有单个备份拓扑配置示例

3 面向多失效场景的生存性路由协议

在独立多链路失效场景中,难以预测可能出现并发失效的链路,各失效检测节点也难以及时获取其他链路失效信息,这给协议设计带来了巨大挑战.如果采用顺序 FIB 更新,节点在不能得知其他失效链路位置时无法计算或仅计算出非法的 FIB 更新顺序;当采用非结构化的快速本地重路由时,由于发起重路由的节点仅能够根据本地局部知识,而无法感知到其他故障点以及其他节点的决策,报文可能传输到其他失效节点/链路而被丢失或陷入环路,仅使用本地快速重路由是难以保证无环、无中断转发的,必须有其他信息辅助转发,这是由协议的分布式特性的本质造成的.单失效场景下的生存性路由协议在多失效场景中不能完全保证协议的正确性和完整性.

目前,应对多独立失效场景的协议主要采用两种机制,即先验式计算多结构化子图覆盖各种可能的并发多链路失效,并在各逻辑子图上独立运行路径选择算法;或者在报文中嵌入“故障根源”等信息,通过报文隐式地传递控制平面失效信息,利用这些特殊信息进行反应式检测、纠正节点间的不一致决策,从而消除环路或黑洞.

3.1 多失效下的结构化备份子图

根据 Menger’s 定理,假定图 G 中 n 边/节点失效,网络保持连通的充要条件是 G 为 $k>n$ 的边/节点不相交连接图,该类协议通过构造能够容忍多失效的多个逻辑拓扑,以保证流量在多独立失效的场景下无中断转发.

(1) 2DMRC[51]. 2DMRC 对 MRC 扩展以应对两个独立并发失效,它必须满足以下 3 个条件:(1) 在隔离节点 u 的拓扑中,节点 u 必须不能作为中转节点,但是流量可以从一个隔离的入口节点转发到另外一个隔离的出口节点;(2) 在每个备份拓扑中,所有的节点必须都是连接的,但是不能使用隔离节点作为中转节点;(3) 所有的节点对之间必定至少存在一个备份拓扑将其隔离.条件(1)决定了限制链路的权重,而条件(2)和条件(3)决定了如何构造备份拓扑的哪些节点能够在同一个备份拓扑中同时隔离.以这 3 个条件作为输入限制产生能够容忍两个独立并发失效的备份拓扑,即任意两个组件失效的时候至少存在一个备份拓扑不使用这两个组件.当出现并发失效时,通过连续地对多个备份拓扑进行多次尝试以提供无中断转发.该协议效率不足,最差情况下算法复杂性将达到 $O(|V|^5)$.

(2) 独立双链路失效恢复[52].假定网络拓扑是 3 边连接图 $G=(V,E)$,该协议为每个节点 $u\in V$ 分配 4 个地址:1 个无失效下的常规地址 u_0,3 个备份保护地址 u_1,u_2 和 u_3.将与节点 u 连接的链路分成 3 个保护组,表示为 L_{u_1},L_{u_2} 和 L_{u_3}.节点 u 与 3 个保护图相关联,表示为 $G_{u_i}(V,E\setminus L_{u_i})$,$i$=1,2,3.其中,$G_{u_i}$ 是通过将 L_{u_i} 删除以后而所得到的,可以证明这种构造方法保证每个保护图 G_{u_i} 都是 2 边连接图,在每个保护图 G_{u_i} 中使用着色树(color tree)技术以路由报文.令 $S_{u_g}=\{v\mid(u,v)\in L_{u_g}\}$ 表示通过 L_{u_g} 的链路连接 u 的节点集合,当连接到节点 u 的链路失效以后,S_{u_g} 中的节点可以通过隧道将报文传输到保护地址 u_g.该协议将保护图和着色树技术结合起来以容忍并发两条链路

© 中国科学院软件研究所 http://www.jos.org.cn

失效,其协议的复杂度为 $O(|V||E|)$.

(3) 路径拼接(path splicing)[53].路径拼接通过将多个到达同一目的节点的最短路径树(单个以目的节点为根的最短路径树称为“片段”)组合形成路径,端系统通过报文头中的“拼接报文头(splicing header)”而切换报文转发路径进行全局恢复;而路由器可以通过改变“拼接报文头”发起本地快速重路由.路径拼接在给定的拓扑上随机扰动链路权重以获得多个拓扑,并运行多个路由协议实例为这些拓扑计算对应的最短路径树(即片段),随机扰动函数为 $L'(i,j)=L(i,j)+weight(a,b,i,j)\times random(0,L(i,j))$,$L(i,j)$表示原始链路的权重,$Weight(a,b,i,j)$表示节点 i,j 之间的属性函数.比如,节点的度数 $Random(0,L(i,j))$表示$[0,L(i,j)]$之间的随机数.其中,

$$\forall i,j,Weight(a,b,i,j)=f_{ab}(degree(i)+degree(j)),$$

f_{ab} 是位于区间$[a,b]$与 $degree(i)+degree(j)$相关的线性函数.在预计算多个片段以后,完整的路径即可由多个片段组合而成,如图 4 所示,其中的 01 字串表示“拼接报文头”.当链路失效后,节点通过设置报文中“拼接报文头”使得报文重路由到不同的分片上,实现无中断转发.

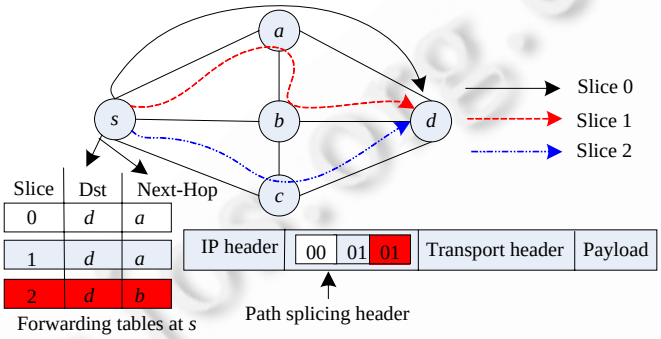


Fig.4 Path splicing

图 4 路径拼接

与 MRC 不同,路径拼接允许多次改变“拼接报文头”以应对多个失效,但是并不能保证 100%的覆盖,从而降低了实现复杂度.此外,当路径拼接通过随机扰动链路权重获得多个拓扑计算分片时,并没有将特定节点隔离,因此可能造成转发环路.路径拼接主要通过限制拼接位的数量来消除持久环路,通过偏转计数法消除瞬时环路.

将路径拼接扩展到域间路由时,报文转发时利用路由表中已经存在的多条路径,通过在报文头添加拼接位信息,从而可以选择 BGP 路由表中的备份路径.采用这种方法报文头开销过大,而且难以使用最大不相交路径.

3.2 信息携带辅助路由

一些研究工作采用了在报文中携带“故障根源”的方法,在不中断报文转发的同时,利用普通报文携带控制平面失效信息以辅助节点路由转发.采用这种方法应对多个失效改变了协议语义,也给路由器带来了额外的计算开销.

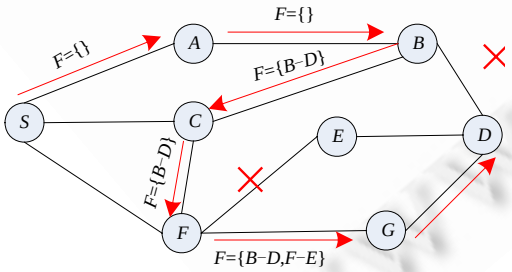


Fig.5 An example illustrating FCP

图 5 FCP 示例

(1) FCP 路由[54].与 OSPF 路由机制不同,FCP 并不致力于减少收敛时间,而是试图完全消除收敛过程;FCP 也不使用复杂的预计算备份路径,而是通过在报文中携带“故障根源”信息,使得报文自主发现可用路径,从而保证了流量的无中断转发.FCP 通过一个类似于 RCP[26]的平台分发全局一致性拓扑视图,当路由器路由失效后,FCP 在报文头中插入“故障根源”信息的黑名单,路由器接收到该报文后,只要底层网络拓扑连通,通过 FCP 算法就可以计算出避开故障链路/节点的新无环路径.图 5 展示了 FCP 的转发过程.

当所有路由器无一致性拓扑视图时,采用基于源路由

© 中国科学院软件研究所 http://www.jos.org.cn

的 SR-FCP(source-routing FCP),第 1 跳路由器在报文中添加完整的到达目标的源路径,下行路由器按照报文头指定的路径转发直到检测到新失效链路.此时,路由器将失效链路加入报文头的失效信息域,并用新计算的路径替换报文中的源路径,SR-FCP 也可用于域间路由.

FCP 在多失效情况下可以保证无中断转发.但是,由于路由器几乎需要为每个失效报文都计算新的路径,即使通过路径缓存的方法也会造成路由器CPU的负担过重;其次,报文头开销大且长度不固定,在核心路由器中难以实现.

类似地,BISF[55]对 FIR[36]进行了扩展,报文中携带了失效链路的黑名单,当节点检测到失效链路后,通过发起类似于本地快速重路由的方法实现无中断转发.但是,该节点仍然必须利用黑名单信息以避免环路并辅助报文转发,使其容忍多失效,且 BISF 不能保证 100%的覆盖.

(2) ACF 路由[56].ACF 接受在路由收敛过程中全局路由视图不一致的现实,在路由不一致的情况下采用检测并恢复的方法,通过在报文头 p 中添加路径踪迹域(path trace)、黑名单域(black list)以及其他辅助功能域,进行动态按需路由计算.其中,路径踪迹域表示为 $p.pathTrace$,用以记录报文所经历过的 AS 列表;黑名单域表示为 $p.blackList$,黑名单域中的 AS 可能由于无服从策略的路径或该 AS 可能引起环路,从而不能将报文正确传输.

ACF 对报文处理过程如下:当在 AS_s 产生报文 p,且其目标域为 AS_d 时,当前 BGP 路由器检测 $p.pathTrace$,如果本地 AS 号不包含在 $p.pathTrace$ 中,报文将按照正常模式转发;如果本地 AS 号码包含在 $p.pathTrace$ 中,说明产生了环路,当前 BGP 路由器将 $p.pathTrace$ 中所有的 AS 号移除并添加到 $p.blackList$ 中,同时调用转发平面搜索 RIB 中不经过 $p.blackList$ 的备份路径.如果当前 BGP 路由器没有备份路径,可将报文传输到 AS_d 时,当前 BGP 路由器将会发起本地恢复转发(recovery forwarding),其核心思想类似于报文封装技术,将报文偏转到最有可能包含合法路径的 AS_r 中.比如 Top-10 AS, AS_r 同样可以继续发起本地恢复转发,以应对可能的多失效场景.

ACF 在报文头中增加了大量的开销,且对应的域长度可变,不利于高速核心路由器的实现.同时,由于采用恢复转发会造成路径长度较大,并有可能形成违反 AS 策略的转发路径.

4 生存性路由协议的比较与评价

基于不同应用场景、技术的生存性路由协议有着各自的特点,为了对当前的生存性路由协议有一个整体的了解,表 1 从多个方面对这些协议进行了比较,包括协议可应对失效类型、核心机制以及特点,着重比较了该协议通过何种方法解决收敛延迟大、瞬时失效的问题,另外,也列出该协议需要哪些假设条件、是否需要额外信令机制、报文转发方式、算法是否分布式、理论覆盖范围.表中的 None 表示无特别的机制改善该指标.

为了进一步揭示,表 2 对生存性路由协议的性能和开销进行了比较,从协议收敛延迟、路径伸展度(即备份路径长度相对于最优路径增加幅度)、各种开销、实现复杂度、协议兼容性等角度进行了综合比较.

生存性协议所能应对的失效场景影响了协议的复杂度、性能及开销,而路径计算是否采用集中式与特定应用、假设条件相关.通过表 1 和表 2 的比较,从生存性路由协议对路径建立和维护的时机、失效反应点两个角度分别对协议实现的复杂度、性能及开销进行了分析和总结.

(1) 先验式预计算与反应式按需计算路径的比较.对于先验式预计算路径的协议,由于其备份路径都是预先计算的,其算法的复杂度并不是影响性能的主要因素,当前节点检测到失效后立即切换到备份路径,先验式预计算路径与本地快速重路由结合几乎可以达到无缝持续转发.域内/域间快速重路由、结构化备份子图等都采用了先验式和本地链路重路由的方法,相关实验结果也表明了该方法的良好性能.这类方法的缺点是,需要在路由器中保存较多状态,且可能造成多次查找路由表,增加了开销.

反应式按需计算的协议主要包含信息携带辅助路由,这类协议算法复杂度、存储开销较低,但却需要利用每个报文头中的特殊信息进行即时的路径计算.比如 FCP[54],ACF[56]类的信息携带辅助协议其报文头开销长度可变,路由器转发平面实现困难.相比而言,结构化备份子图的方法其报文头开销固定,更易于实现.

(2) 源端路径级恢复与本地链路级快速重路由的比较.绝大多数生存性路由协议采用了本地链路级快速重路由方法,其转发中断程度相对更低.然而,本地链路级重路由需要复杂的路径切换机制,通常采用报文封装等

© 中国科学院软件研究所 http://www.jos.org.cn

开销较大的操作.首先,对于域内/域间快速重路由、结构化备份子图等协议,比如 Deflections[35],FIR[36]等,为了避开失效点,经常需要报文回退,报文可能会多次经过同一节点,造成网络资源浪费;其次,本地链路级重路由仅依据本地知识选择备份路径,可能发生环路、选择次优路径的问题.而源端路径级恢复可以避免类似问题,但却会增加中断时间.

生存性路由协议的选择涉及到网络拓扑、规模、协议性能、实现复杂度、兼容性、成本等多个关联因素,比如表 2 中的部分性能指标甚至会相互冲突.举例来说,MRC[46]使用更少的备份拓扑降低了存储开销,但必然会造成网络稀疏,从而导致备份拓扑中路径伸展度更高.如何达到效率和资源代价的合理折衷是需要着重研究的.

Table 1 Comparison of characteristics of survivable routing protocols

表 1 生存性路由协议特点比较

Protocol	Failure type	Proactive/ Reactive	Local/ Global	Multi-path	Domain	Solutions			Assump-tion	Signal	Distri-buted	Packet forwarding
						Long convergence	Transient unreachability	Loop				
OFIB[28]	Single	Proactive	Local	No	Intra	None	Ordered update	Ordered update	Planned failure	Yes	No	Normal
ORMS[29]	Single	Proactive	Local	No	Intra	None	None	Ordered update	Planned failure	No	No	Normal
LFA[31]	Single	Proactive	Local	Yes	Intra	None	Reroute, backup nexthop	SPT	GT	No	Yes	Normal
U-turn[32]	Single	Proactive	Local	Yes	Intra	None	Reroute, backup nexthop	SPT, encap-sulation	GT	No	Yes	Encap-sulation
Tunnels[33]	Single	Proactive	Local	Yes	Intra	None	Reroute, backup nexthop	SPT, encap-sulation	GT	No	Yes	Encap-sulation
Not-via[34]	Single	Proactive	Both	Yes	Both	None	Reroute, backup nexthop	SPT, encap-sulation	GT	Yes	Yes	Encap-sulation
Deflec-tions[35]	Single	Proactive	Both	No	Both	None	Deflection	Deflection rules	GT	Yes	Yes	Tag architecture
FIR[36], $FIFR_N$[39]	Single	Proactive	Local	No	Intra	None	Reroute	Failure inferencing	GT	No	Yes	Interface specific
Safe-Guard[54]	Single	Both	Local	No	Intra	None	Transient-mode	Remaining path cost	Carry cost, GT	Yes	Yes	Detect then rescue
Protec-tion[41]	Single	Proactive	Local	Yes	Inter	None	Reroute, protection tunnels	Encap-sulation	Policy, coll-aboration	No	Yes	Encap-sulation
R-BGP[42]	Single	Proactive	Local	Yes	Inter	RCN	Reroute, backup path	RCN	Policy, coll-aboration	Yes	Yes	Encap-sulation
BRAP[43]	Single	Proactive	Local	Yes	Inter	Replace withdraw	Reroute, backup path	Encap-sulation	Policy, coll-aboration	Yes	Yes	Encap-sulation
D-BGP[44]	Single	Proactive	Local	Yes	Inter	RCN	Reroute, backup path	RCN	Policy, coll-aboration	Yes	Yes	Encap-sulation
Consensus routing[27]	Single	Proactive	Local	No	Inter	Snapshot, consensus	Transient mode	Snapshot, consensus	Policy, coll-aboration	Yes	Yes	Encap-sulation
RRL[45]	Single	Proactive	Both	Yes	Intra	None	Backup layer	Layer identifier	GT	Yes	No	Packet mark
MRC[46]	Single	Proactive	Both	Yes	Intra	None	Backup topology	Topology identifier	GT	Yes	No	Packet mark
IPRT[49]	Single	Proactive	Both	Yes	Intra	None	Redundant trees	Tree identifier	GT	Yes	No	Packet mark
MARAs[50]	Single	Proactive	Both	Yes	Intra	None	Alternative routing	MA ordering	GT	No	No	Normal
2DMRC[51]	2	Proactive	Both	Yes	Intra	None	Backup topology	Topology identifier	GT	Yes	No	Packet mark
Dual link Failures[52]	2	Proactive	Both	Yes	Intra	None	Protection address	Protection graphs	GT	Yes	No	Encap-sulation
Splicing[53]	Multi-ple	Proactive	Both	Yes	Both	None	Multiple slices	Splicing header	GT	Yes	Yes	Packet mark
FCP[54]	Multi-ple	Reactive	Local	No	Both	None	Recompute route	Failure carrying	Network map	Yes	No	Failure carrying
BISF[55]	Multi-ple	Proactive	Local	No	Intra	None	Reroute	Blacklist	GT	Yes	Yes	Interface specific
ACF[56]	Multi-ple	Reactive	Local	No	Inter	None	Detecting and recovering	Path trace, black list	AS coll-aboration	Yes	Yes	Recovery forwarding

SPT: Shortest path tree; Intra: Intradomain; InterInterdomain; GT: Global topology; RCN: Root cause notification

© 中国科学院软件研究所 http://www.jos.org.cn

Table 2 Comparison of performances and overheads for survivable routing protocols

表 2 生存性路由协议性能和开销比较

Protocol	Convergence delay	Path stretch	Coverage	Implementation complexity	Overhead				Compatibility
					Signal	Packet header	Computational complexity	RIB/FIB entry storage	
OFIB[28]	↑	Low	100%	Low	Low	None	High	Low	Yes
ORMS[29]	↑	Low	100%	Low	None	None	High	Low	Yes
LFA[31]	—	Low	Partial	Low	None	None	High	Low	Yes
U-turn[32]	—	Low	Partial	Medium	None	Low	High	Low	Yes
Tunnels[33]	—	Low	Partial	Medium	None	Low	High	Low	Yes
Not-via[34]	—	Low	Partial	High	High	Low	High	High, 2×e	No
Deflections[35]	—	Medium	100%	Medium	Low	Medium	High	Medium	Yes
FIR[36] FIFR$_N$[39]	—	Medium	100%	Medium	None	None	High	None	Yes
SafeGuard[54]	—	Low	100%	High	Low	Medium	Low	Medium	No
Protection[41]	—	Low	100%	Medium	None	Low	Low	Low, 2×*normal*	Yes
R-BGP[42]	↓	Medium	100%	Medium	Low	Low	Low	Low, 2×*normal*	No
BRAP[43]	↓	Medium	100%	High	High	Low	Low	Low, 2×*normal*	No
D-BGP[44]	↓	Medium	100%	Medium	Low	Low	Low	Low, 2×*normal*	No
Consensus[27]	↑	Medium	100%	High	High	Low	High	Low, 2×*normal*	No
RRL[45]	—	Medium	100%	Medium	Low	Medium	High	Medium, N×*normal*	No
MRC[46]	—	Medium	100%	Medium	Low	Medium	High	Medium, n×*normal*	No
IPRT[49]	—	Medium	100%	Medium	Low	Low	Medium	Low, 2×*normal*	No
MARAs[50]	—	Low	100%	Medium	None	None	High	High	Yes
2DMRC[51]	—	Medium	100%	High	Medium	Medium	High	Medium, N×*normal*	No
Dual link failures[52]	—	Medium	100%	High	Medium	Medium	Medium	Medium, 4×*normal*	No
Splicing[53]	—	Low	Partial	Low	Low	Medium	Low	Medium, k×*normal*	No
FCP[54]	None	High	100%	High	Medium	High	Low	Low	No
BISF[55]	—	Medium	Partial	High	Medium	High	High	Medium	No
ACF[56]	—	High	100%	High	Medium	High	Low	Low	No

e: Number of nodes in the network; n: Number of backup layers/topologies;
k: Number of slices; ↑: Increase; —: No change; ↓: Decrease

5 结论与未来展望

生存性路由协议可以显著增强互联网的生存性,符合网络作为基础设施、新兴应用发展的需求,ISP 也能够获得潜在利润.近年来,生存性路由协议引起了学术界和工业界的广泛关注,这些协议的核心思想也相互影响,众多路由器厂商参与制定相关协议草案.从目前研究的分布状况来看,针对域内协议的研究占多数,其原因在于,域内路由位于单管理域,可获得全局拓扑,开销在可承受范围内,出现了一些非常优秀的协议.相比而言,域间生存性路由协议的研究更为迫切,应成为未来研究的重点.我们认为,未来研究中需进一步解决的问题主要有:

(1) 可扩展性.它直接影响协议的可行性,目前,生存性路由协议计算需求较高,开销普遍较大.若能建立路由失效模型,或实施选择性保护,放松 100%覆盖要求,将降低算法以及实现复杂性,在性能和开销之间达到平衡.

(2) 生存性路由协议的实现与部署.首先,设计新协议时需要考虑其兼容性,尽量避免对 OSPF,BGP 协议的语义、消息格式进行较大改动,否则,即使 FCP[54]这样较好的技术也因为涉及到修改体系结构而导致目前难以实际使用.其次,域间路由协议扩展性、稳定性问题更加突出[57],生存性路由协议不能引入过多开销,甚至要降低开销.比如:降低更新消息数量以增强其稳定性、可扩展性;在路由器硬件体系结构中需要研究能够支持共享冗余的 RIB,FIB,并且不会降低查询、更新时间.目前,一些研究正在进行这方面的工作[58].

(3) 流量工程.由于生存性路由协议需要绕过失效组件快速重路由,必定导致大量流量转移,可能带来其他链路的拥塞,造成不必要的报文丢失.目前,考虑流量负载均衡分布、路由优化的工作偏少[48,59].

© 中国科学院软件研究所 http://www.jos.org.cn

(4) BGP 瞬时失效问题更加突出.Internet 中,AS 规模、策略等阻碍了 BGP 学习更多的路径.在不牺牲网络稳定性的前提下,进一步研究如何实现更灵活的 BGP 路由决策过程、如何通告或获取多样性路径[60]、如何改进转发过程等等.

(5) 多失效场景下的生存性路由协议更具挑战性.如何应对这种极端环境下的中断,需要进一步地从理论、机制上加以解决.同时,协议设计需要探索如何综合考虑域内和域间路由相互影响的问题.

(6) 生存性路由协议的模拟.对于域内协议的模拟相对较为真实,然而当前对于域间生存性路由协议的模拟还只是将每个 AS 看作单个路由器节点,而 AS 之间的连接仅看作单条边.这种模型缺乏合理性,与实际的 AS 拓扑并不相符.即使 CAIDA 组织对 Internet 中 AS 拓扑进行了标注,也未考虑 AS 之间多条边的存在这一问题.

致谢 衷心感谢对本文提出宝贵建议的匿名审稿专家以及对本文工作给予支持的老师和同学们.

References:

[1] Schilling WW, Alam M. Measuring the reliability of exsiting Web service. In: Proc. of the 2007 IEEE Intelector/Information Technology Conf. Los Alamitos: IEEE Computer Society, 2007. 356−361. http://dx.doi.org/10.1109/HICSS.2007.338

[2] Gummadi KP, Madhyastha HV, Gribble SD, Levy HM, Wetherall D. Improving the reliability of internet paths with one-hop source routing. In: Proc. of the OSDI 2004. San Francisco: USENIX Association, 2004. 183−198. http://portal.acm.org/citation.cfm?id=1251267

[3] Paxson V. End-to-End routing behavior in the internet. IEEE/ACM Trans. on Networking, 1997,5(5):601−615. [doi: 10.1109/90.649563]

[4] Cisco Visual Networking Index. Forecast and methodology. 2008−2013. 2009. http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[5] Cholda P, Mykkeltveit A, Helvik BE, Wittner OJ, Jajszczyk A. A survey of resilience differentiation frameworks in communication networks. IEEE Communications Surveys & Tutorials, 2007,9(4):32−55. [doi: 10.1109/COMST.2007.4444749]

[6] Iannaccone G, Chuah C, Bhattacharyya S, Diot C. Feasibility of IP restoration in a tier-1 backbone. IEEE Network Magazine, 2004, 18(2):13−19. [doi: 10.1109/MNET.2004.1276606]

[7] Wang H, Yang YR, Liu PH. Reliability as an Interdomain service. In: Proc. of the ACM SIGCOMM 2007. Kyoto: ACM Press, 2007. 229−240. http://doi.acm.org/10.1145/1282380.1282407

[8] Feldmann A. Internet clean-slate design: What and why ? ACM Computer Communication Review, 2007,37(3):59−64. [doi: 10.1145/1273445.1273453]

[9] Markopoulou A, Iannaccone G, Bhattacharyya S, Chuah CN, Ganjali Y, Diot C. Characterization of failures in an operational IP backbone network. IEEE/ACM Trans. on Networking, 2008,16(4):749−762. [doi: 10.1109/TNET.2007.902727]

[10] Basu A, Riecke JG. Stability issues in OSPF routing. In: Proc. of the ACM SIGCOMM 2001. San Diego: ACM Press, 2001. 225−236. http://doi.acm.org/10.1145/383059.383077

[11] Labovitz C, Ahuja A, Bose A, Jahanian F. Delayed Internet routing convergence. IEEE/ACM Trans. on Networking, 2001,9(3):293−306. [doi: 10.1109/90.929852]

[12] Pei D, Wang L, Massey D, Wu SF, Zhang L. A study of packet delivery performance during routing convergence. In: Proc. of the DSN 2003. San Francisco: IEEE Press, 2003. 183−192. http:// ieeexplore.ieee.org/iel5/8589/27228/01209929.pdf

[13] Wang F, Qiu J, Gao L, Wang J. On understanding transient interdomain routing failures. IEEE/ACM Trans. on Networking, 2009, 17(3):740−751. [doi: 10.1109/TNET.2008.2001952]

[14] Wang F, Mao ZM, Wang J, Gao L, Bush R. A measurement study on the impact of routing events on end-to-end Internet path performance. In: Proc. of the ACM SIGCOMM 2006. Pisa: ACM Press, 2006. 375−386. http://portal.acm.org/citation.cfm?id=1159956

[15] Zhang Y, Mao ZM, Wang J. A framework for measuring and predicting the impact of routing changes. In: Proc. of the INFOCOM 2007. Anchorage: IEEE Computer Society, 2007. 339−347. http://dx.doi.org/10.1109/INFCOM.2007.47

[16] Kushman N, Kandula S, Katabi D. Can you hear me now?! It must be BGP. ACM SIGCOMM Computer Communication Review, 2007,37(2):75−84. [doi: 10.1145/1232919.1232927]

[17] Rai S, Mukherjee B, Deshpande O. IP resilience within an autonomous system: Current approaches, challenges, and future directions. IEEE Communications Magazine, 2005,43(10):142−149. [doi: 10.1109/MCOM.2005.1522138]

[18] Pei D, Azuma M, Massey D, Zhang L. BGP-RCN: Improving BGP convergence through root cause notification. Computer Networks, 2005,48(2):175−194. [doi: 10.1016/j.comnet.2004.09.008]

© 中国科学院软件研究所 http://www.jos.org.cn

[19] Chandrashekar J, Duan ZH, Zhang ZL, Krasky J. Limiting path exploration in BGP. In: Proc. of the IEEE INFOCOM 2005. Miami: IEEE Computer Society, 2005. 2337−2348. http://dx.doi.org/10.1109/INFCOM.2005.1498520

[20] Choudhury G. Prioritized treatment of specific OSPF version 2 packets and congestion avoidance. RFC 4222, 2005.

[21] Mao Z, Govindan R, Varghese G, Katz RH. Route flap damping exacerbates Internet routing convergence. In: Proc. of the ACM SIGCOMM. ACM Press, 2002. http://doi.acm.org/10.1145/964725.633047

[22] Yannuzzi M, Masip-Bruin X, Bonaventure O. Open issues in interdomain routing: A survey. IEEE Network, 2005,19(6):49−56. [doi: 10.1109/MNET.2005.1541721]

[23] Mühlbauer W, Maennel O, Uhlig S. Building an as-topology model that captures route diversity. In: Proc. of the ACM SIGCOMM 2006. Pisa: ACM Press, 2006. 195−206. http://portal.acm.org/citation.cfm?id=1159937

[24] Feldmann A, Maennel O, Mao ZM. Locating Internet routing instabilities. In: Proc. of the SIGCOMM 2004. Portland: ACM Press, 2004. 205−218. http://doi.acm.org/10.1145/1030194.1015491

[25] Oliveira R, Zhang B, Pei D, Zhang L. Quantifying path exploration in the Internet. IEEE/ACM Trans. on Networking, 2009,17(2): 445−458. [doi: 10.1109/TNET.2009.2016390]

[26] Caesar M, Caldwell D, Feamster N, Rexford J, Shaikh A, van der Merwe J. Design and implementation of a routing control platform. In: Proc. of the NSDI 2005. Boston: USENIX Association, 2005. 15−28. http://portal.acm.org/citation.cfm?id=1251203.1251205

[27] John JP, Katz-Bassett E, Krishnamurthy A, Anderson T. Consensus routing: The Internet as a distributed system. In: Proc. of the NSDI 2008. Berkeley: USENIX Association, 2008. 351−364. http://portal.acm.org/citation.cfm?id=1387614

[28] Francois P, Bonaventure O. Avoiding transient loops during the convergence of link-state routing protocols. IEEE/ACM Trans. on Networking, 2007,15(6):1280−1292. [doi: 10.1109/TNET.2007.902686]

[29] Francois P, Shand M, Bonaventure O. Disruption free topology reconfiguration in OSPF networks. In: Proc. of the INFOCOM 2007. Anchorage: IEEE Computer Society, 2007. 89−97. http://dx.doi.org/10.1109/INFCOM.2007.19

[30] Francois PF, Bonaventure O. An evaluation of IP-based fast reroute techniques. In: Proc. of the ACM CoNEXT 2005. Toulouse: ACM Press, 2005. 244−245. http://doi.acm.org/10.1145/1095921.1095962

[31] Atlas A, Zinin A. Basic specification for IP fast reroute: Loop-free alternates. RFC 5286, 2008.

[32] Atlas A. U-turn alternates for IP/LDP local protection. IETF Internet Draft, Draft-Atlas-IP-Local-Protect-Uturn-03.Txt, Work in Progress, 2006.

[33] Bryant S, Filsfils C, Previdi S, Shand M. IP Fast reroute using tunnels. IETF Internet Draft, Draft-Bryant-Ipfrr-Tunnels-03.Txt, Work in Progress, 2007.

[34] Shand M, Bryant S, Previdi S. IP fast reroute using not-via addresses. IETF Internet Draft, Draft-Ietf-Rtgwg-Ipfrr-Notvia-Addresses-03.Txt, 2008.

[35] Xiaowei Y, David W. Source selectable path diversity via routing deflections. In: Proc. of the ACM SIGCOMM 2006. Pisa: ACM Press, 2006. 159−170. http://doi.acm.org/10.1145/1159913.1159933

[36] Nelakuditi S, Sanghwan L, Yinzhe Y, Zhang ZL, Chuah CN. Fast local rerouting for handling transient link failures. IEEE/ACM Trans. on Networking, 2007, 15(2):359−372. [doi: 10.1109/TNET.2007.892851]

[37] Li A, Yang X, Wetheral D. SafeGuard: Responsive routing with consistent forwarding. In: Proc. of the ACM CoNext 2009. Rome: ACM Press, 2009. 301−312. http://doi.acm.org/10.1145/1658939.1658974

[38] Li A, Francois P, Yang X. On improving the efficiency and manageability of NotVia. In: Proc. of the ACM CoNext 2007. New York: ACM Press, 2007. 1−12. http://doi.acm.org/10.1145/1364654.1364688

[39] Zifei Z, Nelakuditi S, Yinzhe Y, Lee S, Wang J, Chuah CN. Failure inferencing based fast rerouting for handling transient link and node failures. In: Proc. of the INFOCOM 2006. Barcelona: IEEE Computer Society, 2006. 1−5. http://dx.doi.org/10.1109/INFOCOM.2006.353

[40] Gao L, Rexford J. Stable Internet routing without global coordination. IEEE/ACM Trans. on Networking, 2001,9(6):681−692. [doi: 10.1109/90.974523]

[41] Bonaventure O, Filsfils C, Francois P. Achieving sub-50 milliseconds recovery upon BGP peering link failures. IEEE/ACM Trans. on Networking, 2007,15(5):1123−1135. [doi: 10.1109/TNET.2007.906045]

[42] Kushman N, Kandula S, Katabi D, Maggs BM. R-BGP: Staying connected in a connected world. In: Proc. of the NSDI 2007. Cambridge: USENIX Association, 2007. 341−354. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.86.4001

[43] Wang F, Gao L. A backup route aware routing protocol-fast recovery from transient routing failures. In: Proc. of the IEEE INFOCOM 2008. Phoenix: IEEE Computer Society, 2008. 2333−2341. http://dx.doi.org/10.1109/INFOCOM.2008.302

[44] Wang F, Gao L. Path diversity aware interdomain routing. In: Proc. of the IEEE Infocom 2009. Rio de Janeiro: IEEE Computer Society, 2009. 307−315. http://dx.doi.org/10.1109/INFCOM.2009.5061934

© 中国科学院软件研究所 http://www.jos.org.cn

[45] Fosselie HA, Kvalbein A, Cicic T, Gjessing S, Lysne O. Resilient routing layers for recovery in packet networks. In: Proc. of the Dependable Systems and Networks (DSN). Yokohama: IEEE computer Society, 2005. 238−247. http://dx.doi.org/10.1109/DSN.2005.81

[46] Kvalbein A, Hansen AF, Cicic T, Gjessing S, Lysne O. Fast IP network recovery using multiple routing configurations. In: Proc. of the INFOCOM 2006. Barcelona: IEEE Computer Society, 2006. 1−11. http://dx.doi.org/10.1109/INFOCOM.2006.227

[47] Psenak P, Mirtorabi S, Pillay-Esnault P. Multi-Topology (MT) routing in OSPF. IETF RFC 4915, 2007.

[48] Apostolopoulos G. Using multiple topologies for IP-only protection against network failures: A routing performance perspective. Technical Report, ICS-FORTH 377, Greece, 2006.

[49] Cicic T, Hansen AF, Apeland OK. Redundant trees for fast IP recovery. In: Proc. of the BROADNETS. Raleigh: IEEE Computer Society, 2007. 152−159. http://dx.doi.org/10.1109/BROADNETS.2007.4550419

[50] Ohara Y, Imahori S, van Meter R. MARA: Maximum alternative routing algorithm. In: Proc. of the Infocom 2009. Rio de Janeiro: IEEE Computer Society, 2009. 298−306. http://dx.doi.org/10.1109/INFCOM.2009.5061933

[51] Hansen AF, Lysne O, Cicic T, Gjessing S. Fast proactive recovery from concurrent failures. In: Proc. of the ICC 2007. Glasgow: IEEE Computer Society, 2007. 115−122. http://dx.doi.org/10.1109/ICC.2007.28

[52] Kini S, Ramasubramanian S, Kvalbein A, Hansen AF. Fast recovery from dual link failures in IP networks. In: Proc. of the Infocom 2009. Rio de Janeiro: IEEE Computer Society, 2009. 1368−1376. http://dx.doi.org/10.1109/INFCOM.2009.5062052

[53] Motiwala M, Elmore M, Feamster N, Nempala S. Path splicing. In: Proc. of the SIGCOMM 2008. Seattle: ACM Press, 2008. 27−38. http://doi.acm.org/10.1145/1402958.1402963

[54] Lakshminarayanan KK, Caesar MC, Rangan M, Anderson T, Shenker S, Stoica I. Achieving convergence-free routing using failure-carrying packets. In: Proc. of the ACM SIGCOMM 2007. Kyoto: ACM Press, 2007. 241−252. http://doi.acm.org/10.1145/1282380.1282408

[55] Wang J, Zhong Z, Nelakuditi S. Handling multiple network failures through interface specific forwarding. In: Proc. of the GLOBECOM 2006. San Francusco: IEEE Computer Society, 2006. 1−6. http://dx.doi.org/10.1109/GLOCOM.2006.33

[56] Ermolinskiy A, Shenker S. Reducing transient disconnectivity using anomaly-cognizant forwarding. In: Proc. of the ACM SIGCOMM Hotnet VII. Calgary: ACM Press, 2008. 91−96. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.141.3080

[57] Elmokash A, Kvalbein A, Dovrolis C. On the scalability of BGP: The roles of topology growth and update rate-limiting. In: Proc. of the CoNEXT 2008. Madrid: ACM Press, 2008. 1−12. http://doi.acm.org/10.1145/1544012.1544020

[58] de Carli L, Pan Y, Kumar A, Estan C, Sankaralingam K. PLUG: Flexible lookup modules for rapid deployment of new protocols in high-speed routers. In: Proc. of the ACM SIGCOMM 2009. Barcelona: ACM Press, 2009. 207−218. http://doi.acm.org/10.1145/1592568.1592593

[59] Kvalbein A, Cicic T, Gjessing S. Post-Failure routing performance with multiple routing configurations. In: Proc. of the IEEE INFOCOM 2007. Anchorage: IEEE Computer Society, 2007. 98−106. http://dx.doi.org/10.1109/INFCOM.2007.20

[60] Godfrey PB, Ganichev I, Shenkerz S, Stoica I. Pathlet routing. In: Proc. of the ACM SIGCOMM 2009. Barcelona: ACM Press, 2009. 111−122.

苏金树(1962－),男,福建莆田人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算机网络,信息安全.

赵宝康(1981－),男,博士,助理研究员,主要研究领域为无线网络安全,高性能路由器.

胡乔林(1979－),男,博士生,主要研究领域为网络生存性.

© 中国科学院软件研究所 http://www.jos.org.cn