域内自愈路由研究综述

徐明伟, 杨 芫, 李 琦
(清华大学计算机科学与技术系, 北京 100084)

摘 要: 路由的自愈性是指在网络故障发生后, 路由系统能够自动恢复或重建路由, 保证分组转发不受影响的能力. 虽然目前的域内路由协议都具备一定的自愈能力, 但是它们的自愈时间通常在几秒到上百秒之间, 难以完全满足用户的需求. 在自愈过程中, 网络路由可能是错误的, 甚至会出现"路由黑洞"或"路由环路", 导致分组丢失, 影响网络应用. 为了解决这个问题, 研究者设计了很多域内自愈路由方案. 本文在总结和分析了当前域内自愈路由的问题后, 提出了自愈路由模型, 依据模型将这些方案分成五类: 调整设置权值、限制洪泛、快速重路由、多路径和本地重路由. 在分析每类中典型算法的基础上, 总结对比了各类方案的特点, 详细分析了算法的有效性. 最后探讨了域内自愈路由研究中需要进一步考虑的问题.

关键词: 域内路由; 自愈路由; 网络故障
中图分类号: TP393 文献标识码: A 文章编号: 0372-2112 (2009) 12-2753-09

Survey of Intra-Domain Self-Healing Routing

XU Ming-wei, YANG Yuan, LI Qi
(*Department of Computer Science & Technology, Tsinghua University, Beijing 100084, China*)

Abstract: Self-healing is the ability with which a routing system can restore or rebuild routes automatically after network failures without impacting on packet forwarding. Although existing intra-domain routing protocols can adapt to network failures somewhat, they can not satisfy user requirements because their convergence time ranges from several seconds to more than one hundred seconds. During the period of self-healing, network routes may be incorrect and even have black hole or loop which will lead to packet loss and impact network applications. In order to solve this problem, several intra-domain self-healing routing solutions have been proposed. After analyzing the problems of current intra-domain self-healing routing, we propose intra-domain self-healing routing model, and divide the solutions into five categories according to the model: weight change, flood restriction, fast reroute, multi-path and local reroute. Based on the analysis of some typical solutions in each category, we compared the features of these solutions and studied the availability of the solutions in detail. Finally, we discussed several key issues for further study.

Key words: intra-domain routing; self-healing routing; network failure

1 引言

网络故障是不可避免的. 故障会导致网络连接中断, 危害网络服务, 轻则导致分组丢失, 重则使网络瘫痪, 例如 2006 年 12 月的台湾地震导致中国大陆通往台湾、北美、欧洲、东南亚等方向的互联网长时间大面积瘫痪. 互联网设计之初就非常重视网络故障对网络健壮性的影响, 设计并使用了自适应的动态路由协议. 但是随着互联网规模迅速扩大, 网络故障数量明显增加; 各种应用大量出现, VoIP、在线视频等实时应用广泛使用, 人们对网络端到端性能的要求越来越高. 目前路由协议的路由收敛时间难以完全满足用户的需求. 传统的内部网关协议 IGP(Interior Gateway Protocol) 中, RIP(Routing Information Protocol)[1,2] 在故障发生后自愈所需时间在 100 秒的数量级[3], 且存在"无穷计算"问题[4]; OSPF (Open Shortest Path First)[5]、ISIS (Intermediate System to Intermediate System)[6] 的路由收敛时间在几秒到几十秒. 在收敛过程中, 网络路由可能是错误的, 甚至会出现"路由黑洞", 导致分组丢失, 影响网络应用.

针对路由收敛慢、收敛过程中分组丢失或延迟等问题, 越来越多的研究者开始关注网络路由的自愈性研究. 路由的自愈性是指在网络故障发生后, 路由系统能够自动恢复或重建路由, 保证分组转发不受影响的能力.

互联网路由协议分为内部网关协议 IGP 和外部网关协议 EGP(Exterior Gateway Protocol). 解决自愈路由问题可以从域内和域间两方面入手. 本文重点关注域内自

收稿日期: 2008-09-25; 修回日期: 2009-06-14
基金项目: 国家重点基础研究发计划(973) (No. 2009CB320501); 国家技术研究发展计划(863) (No. 2007AA01Z2A2, 2009AA01Z205); "十一五"国家科技支撑计划重大项目(No. 2008BAB03)

愈路由技术, 研究网络层的域内自愈路由方法. 目前研究人员已经设计了很多域内自愈路由方案. 文献[7]将其中的一些方案分成响应处理和预处理两类. 这种分类方法过于粗糙, 实际上每一类方案中还存在很大的差别, 并且某些方案同时具有响应处理和预处理两种特性(例如故障迟钝路由[8]). 本文首先分析了自愈路由过程, 进而提出了自愈路由模型, 并依据模型对自愈路由方案进行了分类, 使得分类更加合理. 本文还分析了很多最近提出的新方案, 并对一些典型算法进行了模拟, 包括对单个故障和多个并发故障的模拟.

2 域内路由自愈形式化模型

2.1 域内路由的过程及问题

目前互联网中运行的典型域内路由协议有 RIP、OSPF 和 ISIS. 当网络发生故障时, 传统的域内路由协议的操作过程包括如下几个阶段:

(1) 故障检测. 相邻路由器通过互相发送消息确认链路是否正常工作, 若一定时间内没有收到消息则认为对方路由器或链路发生故障.

(2) 故障通知. 检测到故障后路由器需要将新的路由信息或链路信息通知到整个自治系统. 距离向量协议采用逐跳方式传播路由信息, 链路状态协议采用洪泛方式传播链路信息.

(3) 路由重计算. 路由器根据收到的路由信息或链路信息计算出新的路由, 并更新路由表.

(4) 转发信息更新. 路由器将路由表信息传送给转发表, 以便用来查表转发 IP 分组.

传统的域内路由协议在这四个过程中存在几个典型问题, 导致了路由的慢收敛甚至不收敛:

(1) 故障检测时间长. 故障检测时间受计时器的制约. 为了减少路由振荡, 计时器值往往较大, 不能及时检测出故障.

(2) 故障信息传播时间长. 域内路由协议需要将路由变化信息或链路变化信息传播到到整个自治系统. 距离向量算法采用逐跳方式传播路由信息, 传播时间长, 存在"无穷计算"问题; 链路状态算法采用洪泛方式传播链路信息, 传播时间较短, 但是传播时间受计时器的影响, 计时器值太小, 可能会引起路由振荡.

(3) 路由重计算时间长. 路由重计算时间受计时器的影响, 计时器值太小, 可能会引起路由振荡. 而且路由重计算的完成时间在自治系统内各路由器上存在差异, 可能会造成路由循环.

域内路由自愈时间是指从网络故障发生到自治系统中所有路由器的转发表达成一致所需要的时间. 在路由自愈时间内, 网络中可能会形成路由黑洞或路由环路, 导致一些 IP 分组被发往已失效的或错误的结点. 路由更新完成后, 失效链路原来承载的流量将被转移到其他链路上, 可能造成流量过载, 导致网络拥塞丢包.

2.2 域内路由自愈模型

网络结点故障可以被认为是与该结点相连的所有链路都发生了故障, 所以我们只需研究链路故障. 我们总结出了在单个链路发生故障的情况下的域内路由自愈模型, 如图 1 所示. 其中各状态及其持续时间, 以及状态变迁的含义如表 1 所示.

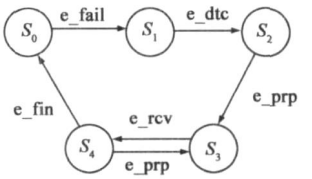


图1 域内自愈路由状态图

表 1 域内自愈路由状态转移表

状态	状态说明	状态持续时间	事件	事件说明	下一状态
S_0	网络路由稳定	--	e_fail	发生故障	S1
S_1	发生了故障但未被检测到	t_1	e_dtc	相邻结点检测到故障	S_2
S_2	故障被相邻结点检测到并进行处理	t_2	e_prp	故障开始向邻居传播	S_3
S_3	故障信息在网络中传播	t_3	e_rcv	非相邻结点接收到故障消息	S_4
S_4	非故障相邻结点处理故障	t_4	e_prp	故障开始向邻居传播	S_3
			e_fin	处理完成, 洪泛完成	S_0

假设故障信息需要经过 n 跳才能传播到整个网络, 那么从发生故障到网络重新稳定所需的时间, 即路由自愈时间为 $T=t_1+t_2+n(t_3+t_4)$. 在 t_1+t_2 时间内分组将被发送到故障链路上造成分组丢失, 如果相邻结点间出现路由循环, 即 S_3 与 S_4 之间发生无用循环, 则分组将在循环过程中由于 TTL(Time To Live) 减为 0 而被丢弃.

从模型的角度, 为了缩短路由收敛的时间及减小故障时网络中断的时间, 可以从以下几个方面进行优化: (1) 消除 S_3 与 S_4 之间的无用循环; (2) 减少 S_3 与 S_4 之间的循环; (3) 缩短 t_1、t_2、t_4 的长度, 优化 S_2、S_3、S_4 的结构; (4) 减小 t_1+t_2 时间内的丢包时间, 可保留或去除原有收敛过程.

故障链路重新恢复正常也是一次拓扑结构改变, 网络状态的变化过程与发生故障时类似. 多链路故障的域内路由自愈模型比较复杂. 文献[9]给出了基于 Petri Net 的多故障域内自愈路由模型.

3 域内自愈路由方案

3.1 方案分类

目前基于网络层的域内自愈路由研究已取得很多

成果. 依据上述域内路由自愈模型, 我们将这些方法分为如下五类:

(1) 调整权值设置. 对 OSPF 和 ISIS 协议中的链路权值(metric) 做更合理的设置. 这类方法在一定程度上减小了发生故障时的网络拥塞, 或者消除路由循环.

(2) 限制洪泛. 通过修改故障通知洪泛机制以限制故障通知的范围, 甚至取消故障通知, 减少故障对全网的影响. 这类方法使路由收敛时间 $T = t1 + t2 + n(t3 + t4)$ 中的 n 值减小, 并减小洪泛带来的额外开销.

(3) 快速重路由. 修改 OSPF 和 ISIS 协议中的计时器以及一些算法和实现细节. 这类方法的路由自愈模型简化为图 2(a), 一方面 $t1$、$t2$、$t3$、$t4$ 的时间被缩短; 另一方面故障信息传播比路由重计算优先级更高. 路由自愈时间缩短为 $T = t1 + n * t3 + t4$, 使得路由器完成路由更新的时间和整个网络的路由自愈时间显著缩短.

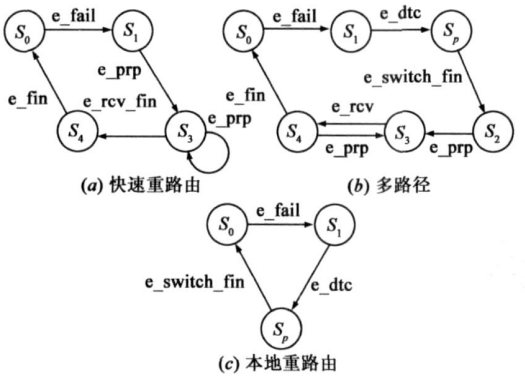


图2　几种改进的域内自愈路由状态图

(4) 多路径. 利用预先计算的多条路径进行传输. 多条路径提供并行的分组转发, 或者提供了主路径的备份. 在此类型的方法中, 网络状态模型可以用图 2(b) 来表示. 该模型在图 1 所示模型的基础上引入了新的状态 S_p, 通过事先规划好的有效的保护措施进行路径切换, S_p 持续的时间即切换时间 t' 通常为 10 毫秒级, 切换完成后分组通过保护路径转发, 此时控制平面进行通常的路由收敛过程. 当路由器完成了新的路由计算后, 即网络进入状态 S_3 时, 分组从保护路径切换到新的最短路径上, 与此同时邻近路由器开始计算新的保护路径. 通过引入保护状态, 可以在不增加抖动时间的基础上减少丢包时间.

(5) 本地重路由. 故障发生时快速切换到预先计算的重路由路径, 然而与多路径方法不同的是它不进行通常的路由收敛过程, 而是将切换后的状态作为新的稳定状态, 如图 2(c) 所示. 当故障恢复时再次快速切换回原来的路由.

按照上面的分类, 各类下的典型算法如图 3 所示.

下面将分别对它们进行介绍和分析.

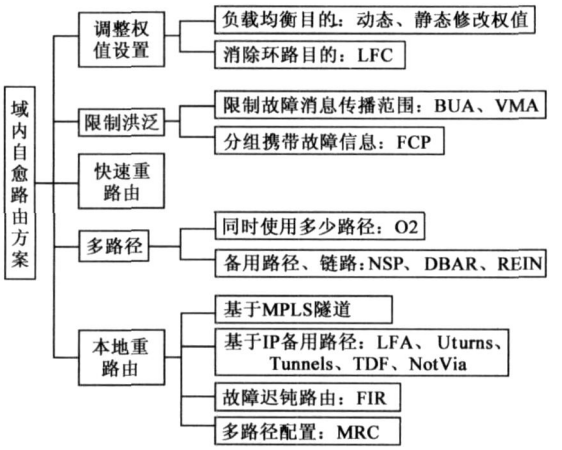


图3　域内自愈路由方案分类

3.2　调整权值设置

链路的权值是决定路由的重要因素, 通常情况下路由算法会选取链路权值之和最小的路径作为最佳路由. 一般情况下, 域内路由协议(例如 OSPF) 的权值是可以由管理员在允许范围内任意设置的, 于是调整权值设置成为一种提高路由自愈能力的方法.

3.2.1　以负载均衡为目的的权值设置

较早的研究将目标放在了网络负载均衡上. 文献[10] 提出的方法是, 通过故障发生后动态地改变一小部分链路的权值以达到负载均衡、避免拥塞的目标. 其中链路权值是根据链路容量和负载来决定的, 负载越大的链路权值越大, 即越不适合于继续增加负载. 该文献通过对美国国家级主干网进行的模拟实验指出, 在发生单个链路故障后, 仅改变 3 条链路的权值就能使流量的分布不超过最优解的 10%. 文献[11] 则使用了一种 Tabu 搜索的方法试图找出一种静态的链路权值分配方案, 其目标同样是负载均衡. 文献[10, 11] 提出的都是针对单个链路故障的方法.

此类方法对路由收敛过程没有改进, 它用于避免拥塞, 使分组转发少受故障的影响.

3.2.2　以消除环路为目的的权值设置

较新的研究不再考虑负载均衡, 而是提出了无环路收敛(Loop Free Convergence, LFC)[12]. 文献[12] 证明了: 在一个稳定的网络中, 将一条链路 A-B 的权值加 1 所触发的路由收敛过程中不会出现环路, 称这种收敛为无环路收敛. 进一步地, 如果要将一条链路 A-B 的权值从 m_1 增加到 m_t, 对于目的结点 D 而言, 定义关键权值为: 区间 $[m_1, m_t]$ 上所有满足条件的正整数 m_i, 使得至少一台路由器到目的地 D 具有不包含 A-B 的额外的等权值路径. 于是得到关键权值序列(Key Metric Sequence, KMS) $\{m_1, m_2, \cdots, m_i, \cdots, m_t\}$ 和重路由权值序列(Reroute Metric Sequence, RMS) $\{m_1, m_1+1, m_2, m_2+1,$

…, m_i, m_i+ 1, …, m_t}. 可以证明: 将 A-B 的权值按 RMS 逐步增大, 此过程中的路由收敛都是无环路收敛. 可以采用进一步的优化处理使 RMS 的长度最短. 利用这个结论, 当一台路由器需要暂时中断一条链路 A-B 时, 可以先依照 RMS 对 A-B 的权值进行修改, 直到 A-B 的权值足够大而不再承载流量, 整个过程中及之后的链路中断、恢复甚至抖动均无路由循环. 链路恢复时将权值减少的过程与权值增大过程类似.

文献[12]提出的方法可以在现有的路由器基础上很容易地实现, 并且它具备了处理多个链路故障的能力, 但它能处理的故障必须可预见, 对于突发故障没有效果.

在路由自愈模型中, 此方法对路由收敛过程的改进在于消除了状态 S_3 与 S_4 之间的无用的循环.

3.3 限制洪泛

传统的链路状态路由方法需要把故障信息洪泛到整个自治域, 但是实际上网络中有一些结点可能并不会受到故障的影响, 因而全域洪泛可能会造成一定的网络资源浪费. 限制洪泛也是域内自愈路由的一类方案.

3.3.1 限制故障消息传播范围

文献[13]设计了两种限制洪泛算法. 枝更新算法(Branch Update Algorithm, BUA)和向量权值算法(Vector-Metric Algorithm, VMA).

BUA 算法中, 当链路 A-B 发生故障时, 假设在目的结点 D 的汇集树(sink tree: 到某一目标结点的所有最短路径的集合)上 A 处在 B 的上游, 则故障信息仅仅洪泛到 A 的上游结点, 这些结点只需要重新计算目的结点位于结点 B 的下游的路由. BUA 只能处理单链路故障.

VMA 算法使用向量 V= (v_0, v_1, …, v_n)来表示一条链路的权值, 初始时 v_0= 0. 定义退化操作: v_0' = 0 且 v_i' = v_i - 1. 当链路 A-B 发生故障时, 结点 A 先计算出源结点到 B 的新的最短路径(或别的可用路径)P, 把 P 上的所有链路权值向量退化, 并以此来计算从 A 到目的结点的新路径, 接着按新路径转发; 故障信息只发送到 P 上的结点, 这些结点做与 A 同样的操作(需要保证各个结点计算的 P 一致, 这一点最短路径算法较易做到). VMA 算法可以处理多个故障, 但这些故障不能相互影响, 即一个故障不出现在由另一个故障所产生的新路径上.

在路由自愈模型中, 此类方法对路由收敛过程的改进在于减少了状态 S_3 与 S_4 之间的循环次数.

3.3.2 分组携带故障信息

文献[14]提出的分组携带故障信息(Failure-Carrying Packet, FCP)方法能完全消除路由洪泛过程. 其基本思想是: 当一个结点检测到与它相连的链路或结点发生故障时, 先将要从该接口转发的分组缓存起来, 并计算新的 SPT(Shortest Path Tree), 接着将分组从新的出口转发出去, 并利用分组携带其经过路径上的所有故障链路信息. 用分组将故障信息传播到网络中其他结点, 因而不需要进行通常的故障信息洪泛. 当一个结点接收到携带着故障信息的分组后, 使用携带的故障信息实时地重新计算 SPT.

FCP 消除了路由洪泛过程, 且不丢弃分组. FCP 进行实时计算, 只要存在通向目的结点的路径, 就能够找到, 大大降低了故障发生时的分组丢失率. 但是 FCP 需要修改现有的链路状态协议和 IP 分组头结构(对于 IPv6 而言要容易一些), 此外部分结点可能要做最短路径树计算, 一段时间内的传输延迟可能会变得较长.

在路由自愈模型中, 此类方法同样减少了状态 S_3 与 S_4 之间的循环次数, 然而会使得状态 S_2 与 S_4 的持续时间增大.

3.4 快速重路由

文献[15]在 2000 年提出了减小目前 IGP 中各计时器的值以缩短路由收敛时间的想法, 近来的文献[16]中也有描述.

快速重路由对目前 IGP 的具体改动主要有以下几方面. 故障检测方面: 将 Hello 交换的周期缩短为亚秒级. 这样做能更快地检测出故障的发生, 但是有可能会引起路由抖动, 因为有时链路可能在失效后的短时间内恢复. 解决方法是让链路失效消息快速发送, 链路恢复消息延迟发送. 故障传播方面: 故障信息的传播相对最短路径的计算应该具有更高的优先级, 这样就能加快故障信息的传播. 最短路径计算方面: 对传统的 Dijkstra 算法进行改进, 采用增量算法, 只重新计算故障影响到的部分 SPT.

快速重路由的优点是不需要修改协议和基础设施就能很容易地实现, 并且能处理任何数量的故障. 它存在的问题包括: 多个故障的收敛时间尚未有研究结果, 快速 Hello 会消耗一定的带宽资源(虽然实际上很少), 此外因为网络中链路故障非常频繁[17], 所以较多的全域洪泛也会造成带宽资源的浪费.

在路由自愈模型中, 快速重路由方法对路由收敛过程的改进在于: 它优化了 S_2、S_3、S_4 之间的结构, 并缩短了状态持续时间, 将路由自愈时间由 T= t_1+ t_2+ n (t_3+ t_4)缩短为 T= t_1+ n* t_3+ t_4.

3.5 多路径

多路径方法是指在故障发生前预先准备好多条路径或多个下一跳(next hop)地址. OSPF 中的等权值路径

(equal cost path)也是多路径的一种形式,但是现实中出现等权值路径的情况不多. 可以同时用不同路径或下一跳来转发分组,例如 O2(Outdegree2)方法[18],而最常见的方法是: 平时使用通常的链路状态协议,当故障发生时使用备用的路径或下一跳,并同时进行正常的路由收敛过程,当后者完成后再切换回通常的链路状态协议,这样做是利用备用路径承担路由收敛过程中可能丢失的流量.

总的来说,多路径方法在路由自愈模型中增加了状态 Sp,使整个网络路由的收敛时间变长,但网络中断的时间可以得到缩短.

3.5.1　同时使用多条路径

O2[18]方法的目的主要是解决故障带来的突发流量冲击问题. 其基本思想是让网络中的每个结点尽可能同时向两个不同下一跳转发分组. 这样,在没有故障的时候,每一个结点的流量都被多条路径所分担. 当单一故障发生时,由于还有一个下一跳可用,分组不会丢失而会被继续转发. 由于故障链路的流量被分散到多条路径上,突发流量的冲击也会减轻. 当故障发生后,每个结点再重新计算O2路由,之后恢复正常的向两个不同下一跳转发分组的过程. 在目前连通度越来越高的网络拓扑结构下,绝大多数结点通常都能找到 2 个出度.

O2 算法形成多条连通路径,个别的故障不会影响到正常的数据传输,并且由于将网络流量分配给不同的链路,起到了一定的负载均衡作用. 虽然 O2 能处理多个故障,但对网络拓扑有一定的要求,绝大部分结点的出度应不小于 2,此外其主要缺点在于算法的复杂性高达 $O(n^3)$,以及它完全改变了目前的路由方式,因此较难得到实现.

3.5.2　备用路径或备用链路

文献[19, 20]采用了邻居最短路径(Neighbor's Shortest Path, NSP)的方法,利用计算出的NSP作为备用路径. NSP 指的是源结点到邻居结点的链路加上该邻居到目的结点的最短路径,并且要求邻居结点的最短路径不能通过源结点本身. 文献[20]进一步证明了以下结论: 假设 S 通过邻居结点 N 发送分组到目的地 D,令 $d(i,j)$ 表示从结点 i 到结点 j 的最小代价值,如果 $d(N,D)< d(S,D)$ 成立,则可断定利用 N 发送不会产生回路. 虽然并不是所有情况下都能找到NSP,但在多数情况下,当网络中流量均匀分布,且故障在各处发生的概率相等时,NSP具有较高的可用性. NSP 的另一个优点是与现有协议兼容,能实现增量配置.

有一些方法不是为每条路径计算一条备用路径,而是在每个结点为每个链路计算一个备用接口(链路). 因为一个结点的链路数量相比路由表项数量来说通常要少得多,因此备用链路表会比路由表要小得多. 这里重点介绍文献[21]的方法: 基于转移的迂回路由(Deflection Based Alternate Routing, DBAR),它利用汇集树 sink tree,将不在汇集树上且连接了汇集树的两棵子树的链路称为桥(bridge). 模拟结果表明,能计算出的作为备用链路的桥的数量与路由表重新计算后的平均下一跳数量很接近.

本地重路由方案及前面的几种备用路径(链路)方案会带来本地突发流量冲击问题. REIN(REliability as an INterdomain Service)方法[22]尝试在自治系统(Autonomous System, AS)之间建立备用通道,通过流量工程的算法建立备用路径,利用它们来处理 AS 内故障造成的网络分隔和突发流量冲击. REIN 还考虑了 ISP(Internet Services Provider)较为关心的 VPN(Virtual Private Network)流量,使得当故障发生时 VPN 流量仍然具有较高优先级. REIN 实现代价较小,能处理多个故障. 它较适合于持续较长时间的严重故障,对于瞬时且频繁发生的故障不合适.

3.6　本地重路由

本地重路由指的是检测到故障的结点采用路径快速切换的方法,将故障造成的影响限制在本地(即与故障相邻的路由器)的方法. 在路由自愈模型中,一般的本地重路由方法在状态 S_p 切换路由后回到稳态,路由收敛过程变得非常简单; 某些特殊的本地重路由方法在一定条件下仍需要进行一般的收敛过程,如故障迟钝路由.

3.6.1　基于 MPLS 隧道

文献[23]扩展了基于资源预留协议的流量工程(RSVP-TE),用于创建标签交换路径(Label Switch Path, LSP)隧道以修复路由. 用户的数据流能够在数十毫秒内切换到备用 LSP 隧道上,因此能满足一些实时应用的需求,例如 VoIP(Voice over Internet Protocol).

有两种基本的建立备用 LSP 的方式. 一种是 One-to-One 方式,每个路由器为经过它的每条路径都建立一条备用 LSP,备用 LSP 被用来绕过该结点在该路径上的下一跳结点或链路. 另一种是 facility 方式,每个路由器建立一条备用 LSP 用来绕过某一邻居结点或链路,它可以作为多条经过这一链路的路径的备用 LSP. 这两种方式都需要 n-1 条备用 LSP 来保护一条经过了 n 个结点的路径,然而 facility 方式的一条备用LSP能够保护一个路径集合.

MPLS 隧道方式能提供快速的路径切换,但它对基础设施有一定的要求,并且只有在支持 MPLS 以及 RSVP-TE 的网络中才能实现.

3.6.2 基于IP备用路径

由于简单修改计时器可能会导致路由振荡, Shand 等提出了基于 IP 重路由的方案[24,25]. 在这类方案中, 当检测到故障后不发布故障信息, 而是使用备份路由实现本地故障恢复. 通过此技术, 路由中断时间非常小, 因为只需要检测故障和激活备份线路的过程. 这套方案是有别于MPLS 重路由的纯 IP 网络的方案, 此外它也不同于多路径方案, 因为它没有通常的路由收敛过程, 而仅依靠切换路径来屏蔽故障的影响, 所以只能处理持续时间较短的故障.

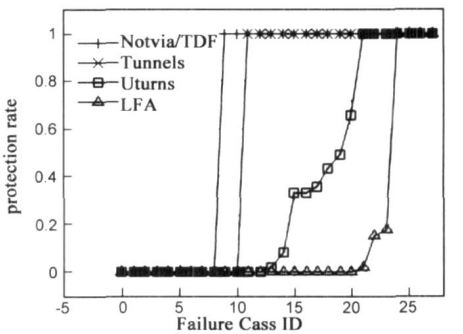
基于 IP 备用路径的本地重路由有多种方案(文献[26~31]), 包括 Equal Cost Path(ECP), Loop Free Alternates(LFA), Uturns, Tunnels, Tunnels with Directed Forwarding(TDF), NotVia addresses. 文献[28]采用隧道(Tunnels)的方法, 详细介绍了计算并选择无环路的备用隧道的方法. 文献[26]则为每个被保护的结点分配一个特殊的地址, 称为 not-via 地址, 一个 not-via 地址不仅指出了发生故障的结点 P, 还指出了目的结点或路径上的一个中间结点 B. 当发生故障时, 与故障相邻的结点将分组从无故障的接口转发给一个适当的 not-via 地址, 收到此分组的结点就知道 P 结点发生了故障, 并使用备用路径将分组转发到结点 B.

基于 IP的重路由方案是自愈性比较好的一类方法. 它可以在纯 IP 网络上实现, 对拓扑结构没有特殊的要求, 且路径切换速度快, 网络中断时间很短. 我们通过研究发现, 各算法在保护能力上有一定差距. 在单链路故障下, NotVia 和 TDF 算法对二连通网络具有完全的保护能力(可理论上证明), Tunnels 和 Uturns 算法能提供较高的保护率, 而 LFA 和 ECP 算法保护能力较差或不能提供保护. 在多个并发故障的情况下, 各算法的保护能力有所下降, 但只要故障不超过一定数量, NotVia 等算法仍然能提供一定的保护[32]. 图 4 给出了我们基于 CERNET2(China Education and Research Network 2) 拓扑对各算法的保护能力进行模拟的结果. 图中横坐标表示故障的情形, 纵坐标表示能够被保护的端到端路径所占的比例.

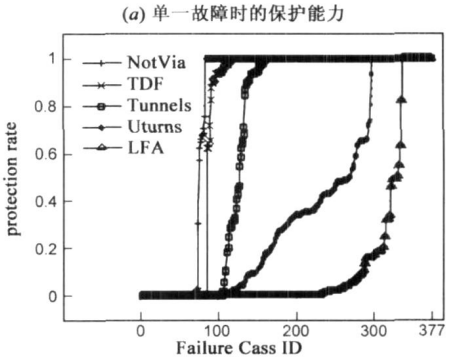
3.6.3 故障迟钝路由

文献[8]的方法称为故障迟钝路由(Failure Insensitive Routing, FIR). 当一个结点检测到与它相连的链路或结点发生故障, 不立即洪泛故障信息, 而是在一定的抑制时间(suppression interval)内尝试在本地消除故障的影响. 如果在抑制时间内链路恢复, 则抑制成功, 不需要洪泛故障信息, 否则进行故障信息洪泛等通常的 IGP 收敛过程. FIR 旨在消除链路短时间失效带来的洪泛等额外开销.

FIR 本地重路由的方法是通过接收到分组的接口来检测故障, 并推断出可能出现故障的链路, 根据推断出来的信息将分组转发到其他链路上, 以绕开可能发生了故障的链路. 此方法称为接口区别转发(Interface-Specific Forwarding). FIR 需要预先计算一张按接口区别的转发表, 故障发生时只需要查表就能简单地按备用链路转发. 定义关键链路的集合 K^d_{j-i}: 当 K^d_{j-i} 中的任意一条链路失效时, 发往目的地 d 的分组将通过结点 j 到达结点 i. 设网络中所有链路的集合为 E, 在 $E-K^d_{j-i}$ 中重新计算结点 i 到 d 的最短路径即可得到备用链路, 也就是抑制时间内的转发链路. 文献[8]给出了计算 K^d_{j-i} 的算法, 若使用增量算法则计算复杂度仅为一次 SPT 调用[33].



(a) 单一故障时的保护能力



(b) 2个并发故障时的保护能力

图4 基于CERNET2拓扑的本地重路由算法保护能力

3.6.4 多路由配置

文献[34,35]提出的方法称为多路由配置(Multiple Routing Configurations, MRC), 其基本思想为: 保存多种路由配置 C_0, C_1, C_2, …, C_n, 各配置的不同之处仅在于一些链路权值不同, 用这种方式隔离或限制一些结点和链路, 这样便能在发生故障后选用合适的备用配置绕开故障点; 另一方面可对不同的数据流采用不同配置, 以减小链路利用率, 起到负载均衡的作用. 图 5 所示是同一网络拓扑的四种不同配置, 实线代表使用的链路, 短虚线代表被隔离(不使用)的链路或结点, 长虚线代表被限制使用的链路. 可以看出, 只要有合适的算法和足够的空间, MRC 可以处理任意位置、任意数量的故障, 只要采用将故障隔离的配置就可以了.

MRC 具有比较强的路由保护功能, 同时还具有负载均衡的作用, 但存在一个较大的问题: 对于较大的网络拓扑, 备用配置的数量将急剧上升, 从而消耗大量存储空间, 而且初始化的时候计算很多的备用配置也会变得困难.

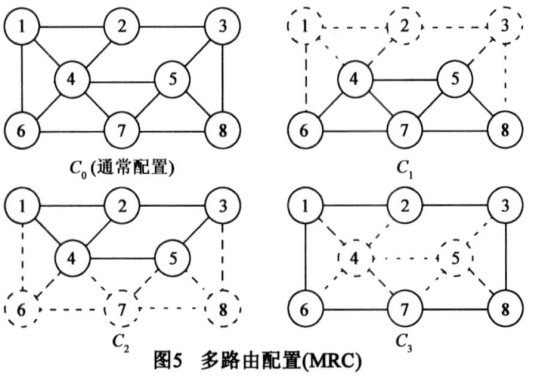


图5　多路由配置(MRC)

4　总结

表 2 对上述方法进行了比较.

表2　域内自愈路由方法比较

一级分类	优化的阶段	二级分类	路由收敛时间	网络中断时间	处理多故障	主要开销
调整权值设置	消除状态 S_3 与 S_4 之间的无用循环	负载均衡目的	10 秒级	10 秒级	不支持	权值分配的计算(少量)
		消除环路目的	10 秒级	10 秒级	不支持	权值序列的计算(少量)
限制洪泛	减少 S_3 与 S_4 之间的循环次数	限制故障消息传播范围	10 秒级	10 秒级	一定情况下支持	传播范围的计算(少量)
		分组携带故障信息	10 秒级	10 秒级	支持	实时 SPT 计算
快速重路由	优化 S_2、S_3、S_4 的结构, 缩短状态持续时间	– –	秒级	秒级	支持	少量的带宽消耗
多路径	增加了临时路由状态 S_p	同时使用多条路径	10 秒级	10 毫秒级	一定情况下支持	计算复杂度 $O(n^3)$, 多个下一跳存储
		备用路径或链路	10 秒级	10 毫秒级	一定情况下支持	备用路由表计算和存储
本地重路由	简化状态, 减少变迁	基于 MPLS 隧道	亚秒级	10 毫秒级	不支持	配置的备用路径的存储
		基于 IP 备用路径	亚秒级	10 毫秒级	部分支持	备用路由表计算和存储
		故障迟钝路由	亚秒级	10 毫秒级	不支持	备用路由表计算和存储
		多路径配置	亚秒级	10 毫秒级	支持	大量计算和存储

调整权值设置的方法能够均衡负载或消除路由循环, 它的实现不需要对现有协议进行任何修改, 但它能处理的故障情况较为有限.

限制洪泛方法旨在减小或取消故障消息洪泛带来的带宽消耗. *FCP* 方法不需要丢弃分组. 限制洪泛的代价在于不能处理较多故障或自愈时间较长, 并且它们都需要在不同程度上修改现有协议.

快速重路由方法修改了现有 *IGP* 协议参数, 对路由收敛过程进行了优化, 是目前可行且较为有效的自愈方案, 路由收敛速度很快, 实现起来也较为容易, 但需要一定的额外带宽作为代价. 它对多个故障的处理能力需要进一步研究.

多路径方法能快速地对故障进行处理, 其代价在于需要附加一定计算量和存储空间, 多数实现都需要修改现有协议.

本地重路由方法在本地快速对故障进行处理, 而避免进行通常的路由收敛过程, 但它不能很好地处理多个并发故障, 或者需要较高的计算和存储代价.

5　进一步的研究问题

尽管域内自愈路由研究已经取得了很多成果, 但是还没有完全解决现实中网络自愈能力不够的问题. 一方面需要在理论上解决一些根本的问题, 例如, 构造高自愈性、低成本的网络拓扑等; 另一方面还需要继续研究高性能、切实可行的自愈路由算法. 在研究过程中需要关注以下几个问题.

5.1　路由可用性与稳定性

可用性与稳定性是自愈路由首要解决的问题. 一方面, 在发生故障时要尽可能找到可用的路径继续数据的传输. 目前的研究中, 备用路径的覆盖率是衡量域内自愈路由算法的一个重要指标, 然而目前很少有某种单独的算法可以提供较高的可用性, 而结合多种方法的代价过于昂贵. 寻找能提供更高可用性的自愈路由算法是非常必要的. 当前的研究多在故障发生的邻近路由器处理故障, 如果能在稍大的范围内加以考虑或许能找到更多有效的方法. Uturns 方案[29] 就在寻找重路由路径时考虑了距离邻近路由器 2 跳的路由器.

另一方面, 当故障频繁发生时要能抵御故障带来的振荡. 文献[36] 对 OSPF 的稳定性做了详细的研究. 文献[37] 也指出应当防止网络振荡的发生. 快速与稳定在一定程度上是相互制约的. 不同的故障具有不同的性质, 如果能对故障发生的原因进行区分, 并进行不同处理, 便能够在实现快速路由自愈或实施有效保护的同时提供较高稳定性.

5.2　对多故障的适应性

目前能够处理单一链路故障的域内自愈路由方法比较多, 理论模型也较为完善. 但是实际网络中的故障是随机的、突发的, 需要有一个很好的模型来描述网络在发生多个故障时的路由情况, 更需要能够处理多故障的域内自愈路由器方法. 可以对原来只能处理单一故障的方法加以改进, 使之对多故障的适应性提升, 或者设计新的方法. 目前的一些方案也具有了处理多个

并发故障的能力,例如快速重路由,我们还需要对其性能进行进一步评估,并研究在多故障模型下协议参数选取的问题.若能结合稳定性研究的成果,快速重路由将成为比较好的域内自愈路由方案之一.

5.3 其他考虑

"简单实用"是互联网的一条准则.思想简单、易于实现、能与目前的网络兼容,并能提高路由自愈性的方法就是好的方法.在这方面,采用隧道等方式的自愈方法较被看好.此外,以往的研究多注重路由收敛,而很少考虑到用户分组丢弃的问题.应用层的发展趋势使人们开始从用户的角度出发,提供面向用户的自愈路由,不仅仅关注路由的情况,更重视如何保证数据的转发.

6 结束语

本文对域内自愈路由问题进行了深入的分析和讨论,给出了自愈路由模型,并对现有的研究成果进行了综述,分析了各种方案的优点及其存在的问题,然后探讨了需要进一步研究的问题.

参考文献:

[1] C Hedrick. Routing Information Protocol[EB/OL]. http://www.ietf.org/rfc/rfc1058.txt, 1988.

[2] G Malkin. RIP Version 2[EB/OL]. http://www.ietf.org/rfc/rfc2453.txt, 1998.

[3] G Iannaccone, C Chuah, S Bhattacharyya, C Diot. Feasibility of IP restoration in a tier 1 backbone[J]. IEEE Network Magazine, 2004, 18(2): 13-19.

[4] Floyd S, Jacobson V. The synchronization of periodic routing messages[J]. IEEE/ACM transactions on networking, 1994, 2(2): 122-136.

[5] J Moy. OSPF Version 2[EB/OL]. http://www.ietf.org/rfc/rfc2328.txt, 1998.

[6] D Oran, OSI IS-IS Intra-domain Routing Protocol[EB/OL]. http://www.ietf.org/rfc/rfc1142.txt, 1990.

[7] S Rai, B Mukherjee, O Deshpande. IP resilience within an autonomous system: current approaches, challenges, and future directions[J]. IEEE Communications Magazine, 2005, 43(10): 142-149.

[8] Sanghwan Lee, Yinzhe Yu, Srihari Nelakuditi, Zhi-Li Zhang, Chen-Nee Chuah. Proactive vs reactive approaches to failure resilient routing[A]. Proceedings of INFOCOM 2004[C]. Hong Kong: IEEE Press, 2004. 176-186.

[9] Lingtao Pan, Mingwei Xu, Qi Li, Dan Wang. The Self-Healing Ability of Link State Routing Systems: Modeling and Analysis [EB/OL]. http://network.cs.tsinghua.edu.cn/teacher/xumingwei/pub/0801.pdf, 2008.

[10] B Fortz, M Thorup. Optimizing OSPF/IS-IS weights in a changing world[J]. IEEE Journal on Selected Areas in Communications, 2002, 20(4): 756-767.

[11] A Nucci, B Schroeder, S Bhattacharyya, N Taft, C Diot. IGP link weight assignment for transient link failures[A]. Proceedings of 18th International Teletraffic Congress (ITC)[C]. Berlin, Germany: Elsevier Press, 2003, 321-330.

[12] Pierre Francois, Mike Shand, Olivier Bonaventure. Disruption free topology reconfiguration in OSPF networks[A]. Proceedings of INFOCOM 2007[C]. Anchorage, AK: IEEE Press, 2007. 89-97.

[13] P Narvaez. Routing Reconfiguration in IP Networks[D]. Massachusetts, USA: Massachusetts Institute of Technology, 2000.

[14] Karthik Lakshminarayanan, Matthew Caesar, Murali Rangan, Tom Anderson, Scott Shenker, Ion Stoica. Achieving convergence-free routing using failure-carrying packets[A]. Proceedings of ACM SIGCOMM 2007[C]. Kyoto, Japan: ACM Press, 2007, 37(4): 241-252.

[15] C Alaettinoglu, V Jacobson, H Yu. Towards Millisecond IGP Convergence[EB/OL]. http://www.nanog.org/meetings/nanog20/abstracts.php?pt=MTA3MiZuYW5vZzIw&nm=nanog20, 2000.

[16] Pierre Francois, Clarence Filsfils, John Evans, Olivier Bonaventure. Achieving subsecond IGP convergence in large IP networks[J]. ACM SIGCOMM Computer Communication Review, 2005, 35(2): 35-44.

[17] Athina Markopoulou, Gianluca Iannaccone, Supratik Bhattacharyya, Chen-Nee Chuah, Christophe Diot. Characterization of failures in an IP backbone[A]. Proceedings of INFOCOM 2004[C]. Hong Kong: IEEE Press, 2004. 2307-2317.

[18] Gero Schollmeiers, Joachim Charzinski, Andreas Kirstadter, Christoph Reichertt, Karl J Schrodi, Yuri Glickmant, Chris Wider. Improving the resilience in IP networks[A]. IEEE High Performance Switching and Routing 2003 (HPSR 2003)[C]. Torino, Italy: IEEE Press, 2003. 91-96.

[19] S Kini, Yibin Yang. Traffic restoration in link state protocols using neighbor's shortest path[EB/OL]. http://www.potaroo.net/ietf/old-ids/draft-kini-traf-restore-nsp-00.txt, 2002.

[20] V Naidu. IP Fast Reroute using Multiple Path Algorithm (MPA)[EB/OL]. https://datatracker.ietf.org/drafts/draft-venkatar-ipfrr-mpa/, 2004.

[21] S Vellanki, A L N Reddy. Improving Service Availability During Link Failure Transients through Alternate Routing[EB/OL]. http://dropzone.tamu.edu/techpubs/2003/TAMU-ECE-2003-02.pdf, 2003.

[22] Hao Wang, Yang Richard Yang, Paul H Liu, Jia Wang, Alexandre Gerber, Albert Greenberg. Reliability as an interdomain service[A]. Proceedings of SIGCOMM 2007[C]. Kyoto, Japan: ACM Press, 2007, 37(4): 229-240.

[23] P Pan, G Swallow, A Atlas. Fast Reroute Extensions to RSVP-TE for LSP Tunnels[EB/OL]. http://www.ietf.org/rfc/rfc4090.txt, 2005.

[24] M Shand, S Bryant. IP Fast Reroute Framework[EB/OL]. http://tools.ietf.org/html/draft-ietf-rtgwg-ipfrr-framework-10, 2009.

[25] M Shand, S Bryant. A Framework for Loop-free Convergence [EB/OL]. http://tools.ietf.org/html/draft-bryant-shand-lf-conv-frmwk-03, 2006.

[26] S Bryant, Shand M, S Previdi. IP Fast Reroute Using Notvia Addresses[EB/OL]. http://tools.ietf.org/html/draft-ietf-rtgwg-ipfrr-notvia-addresses-03, 2008.

[27] S Nelakuditi, S Lee, Y Yu, Z L Zhang, C N Chuah. Fast local rerouting for handling transient link failures[A]. Proceedings of IEEE/ACM Transactions on Networking (TON) 2007 [C]. Piscataway, NJ, USA: IEEE Press, 2007, 15(2): 359-372.

[28] S Bryant, C Filsfils, S Previdi, M Shand. IP Fast Reroute using tunnels[EB/OL]. http://tools.ietf.org/html/draft-bryant-ipfrr-tunnels-03, 2007.

[29] A Atlas. U-turn Alternate for IP/LDP Fast-Reroute [EB/OL]. http://tools.ietf.org/id/draft-atlas-ip-local-protect-uturn-03.txt, 2006.

[30] Albert J Tian, Naiming Shen. Fast Reroute using Alternative Shortest Paths[EB/OL]. http://tools.ietf.org/html/draft-tian-rtgwg-fr-alt-shortest-pathr-01, 2004.

[31] A Atlas, A Zinin. Basic Specification for IP Fast-Reroute: Loop-free Alternates [EB/OL]. http://tools.ietf.org/id/draft-ietf-rtgwg-ipfrr-spec-base-12.txt, 2008.

[32] Yuan Yang. Simulation and Research of Intra-domain Self-Healing Routing [EB/OL]. http://network.cs.tsinghua.edu.cn/teacher/xumingwei/pub/0802.pdf, 2008.

[33] Srihari Nelakuditi, Sanghwan Lee, Yinzhe Yu, Zhi-Li Zhang. Failure insensitive routing for ensuring service availability [A]. Proceedings of International Workshop on Quality of Service 2003 (IWQoS 2003) [C]. Monterey, California: Springer Press, 2003. 287-304.

[34] A Kvalbein, A F Hansen, T Cicic, S Gjessing, O Lysne. Fast IP network recovery using multiple routing configurations [A]. Proceedings of INFOCOM 2006[C]. Barcelona, Spain: IEEE Press, 2006. 23-29.

[35] A Kvalbein, T Cicic, S Gjessing. Post-failure routing performance with multiple routing configurations[A]. Proceedings of INFOCOM 2007[C]. Anchorage, AK: IEEE Press, 2007. 98-106.

[36] Anindya Basu, Jon G Riecke. Stability issues in OSPF routing [A]. Proceedings of SIGCOMM 2001[C]. New York: ACM Press, 2001, 31(4): 225-236.

[37] 张民贵, 刘斌. IP 网络的快速故障恢复[J]. 电子学报, 2008, 36(8): 1595-1602.
Zhang Min-gui, Liu Bin. Fast failure recovery of IP networks [J]. Acta Electronica Sinica, 2008, 36(8): 1595-1602. (in Chinese)

作者简介:

徐明伟　男, 1971 年生于辽宁朝阳. 1994 年和 1998 年在清华大学计算机科学与技术系分别获学士与博士学位, 1998 年留校任教. 现任网络所所长, 教授, 博导. IEEE 会员, 中国计算机学会会员. 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.
E-mail: xmw@csnet1.cs.tsinghua.edu.cn

杨　芫　男, 1984 年生于山东淄博. 2006 年毕业于清华大学计算机系, 获工学学士学位, 现为清华大学硕士研究生. 主要研究领域为互联网路由.
E-mail: yyang@csnet1.cs.tsinghua.edu.cn

李　琦　男, 1979 年生于浙江长兴, 2003 年和 2007 年分别在清华大学和中国科学院获工学学士学位和硕士学位. 现为清华大学博士研究生. 主要研究领域为网络体系结构与网络安全.
E-mail: liqi@csnet1.cs.tsinghua.edu.cn