

IP 数据传输的快速故障恢复研究

Research on Fast Fault Recovery for IP Data Transmission

瞿辉¹ 祝坤洋¹ 王建兵¹ 康宗绪¹ 董彧焘²

Hui Qu¹ Kuniyang Zhu¹ Jianbing Wang¹ Zongxu Kang¹ Yutao Dong²

1. 重庆金美通信有限责任公司 中国·重庆 400030

2. 上海航天电子技术研究所 中国·上海 201199

1.Chongqing Jinmei Communications Co., Ltd., Chongqing, 400030, China

2.Shanghai Aerospace Electronic Technology Institute, Shanghai, 201199, China

摘要: 现有 SDN 数据传输网络中的链路故障恢复方案分为反应式故障恢复和先应式故障恢复。反应式故障恢复是控制器为受影响的数据流重新计算新的可用路径并配置转发规则。先应式故障恢复是 SDN 控制器预先计算保护路径, 当检测到故障发生后, 交换机自动切换并沿保护路径转发数据包。通过对既有技术的研究, 提出了分布式组网控制下, 依赖 OSPF 路由协议的同步技术, 同步全网的链路标签并计算中间节点到目的节点的标签转发路径, 从而在故障发生地采用标签转发路径到达目的。

Abstract: Existing solutions for link fault recovery in SDN data transmission network can be divided into reactive fault recovery and active fault recovery. Reactive fault recovery is where the controller recalculates a new available path for the affected data stream and configures forwarding rules. Active fault recovery means that the SDN controller calculates the protection path in advance. When detecting a fault, the switch automatically switches and forwards data packets along the protection path. Based on the research of the existing technologies, this paper proposes a synchronization technology which relies on the OSPF routing protocol under the control of distributed networking, synchronizes the link labels of the whole network, calculates the label forwarding path from the intermediate node to the destination node, and then adopts the label forwarding path to reach the destination where the fault occurs.

关键词: 软件定义网络; 标签转发; 开放式最短路径优先; 段路由; 链路标签

Keywords: SDN; lable forwarding; OSPF; SR; link label

DOI: 10.12346/csai.v1i3.7749

1 引言

在传统 SDN 数据传输网络中, 链路故障恢复方案分为先应式故障恢复和反应式故障恢复两种模式^[1]。现有链路故障处理策略的研究工作主要可以分为两大类: 故障恢复和故障保护, 亦有称被动恢复和主动恢复。

故障恢复 (被动恢复): 在恢复过程中需要 SDN 控制器的参与, 当检测到链路发生故障后, 控制器为受影响的数据流重新计算新的可用路径并配置转发规则, 以恢复中断的通信进程。

故障保护 (主动恢复): 在数据流到达前, SDN 控制

器预先为其计算保护路径, 当检测到故障发生后, 交换机会自动切换并沿保护路径转发数据包, 而无需控制器的参与。

二者的主要区别在于恢复过程中是否需要 SDN 控制器的参与。主动恢复方案由于没有控制器的参与, 链路恢复都是在本地进行的, 因此链路故障恢复的延迟低, 而由于其在业务到来时就在网络每个节点处存储主备两条路径, 相应的需要更多的存储空间^[2]。被动恢复方案与其相反, 网络节点无需事先额外存储备份路径, 需要的存储空间比主动恢复方案小, 而这种方案是当链路发生故障后再进行重路由计算, 因此耗时较长, 难以满足时延敏感业务的传输需求。

【作者简介】瞿辉 (1979-), 男, 中国重庆人, 本科, 高级工程师, 从事 IP 路由、IP 组播研究。

2 研究成果

链路故障主动恢复和备份路径有关,研究角度可以分为五类:故障检测方式、备份路径的计算方式、备份路径的切换机制、备份路径的存储方式、备份路径的传输方式。

故障检测技术分为接口检测、业务检测。接口检测技术较为简单,检测时间较短,相应的故障恢复时间也短。业务检测即端对端的检测,涉及本地接口,也涉及非本地接口,检测时间较长,故障恢复时间也长^[3]。

备份路径的计算方式主要有两种,第一种是在源节点和目的节点之间进行备份路径计算,即从源端到目的端除了一条主路径之外,另行设置一条不相交的备份路径,故障发生时流量按源路径返回源节点,通过备份路径进行转发。第二种是在中间节点到目的节点之间进行备份路径计算,即尽可能地在每个中间节点处都设置到目的节点的备份路径,相比于第一种可以节省流量返回的时间^[4]。

备份路径的切换机制主要有两种,第一种是组表方式^[5],SDN 控制器提前计算好备份路径,并在交换机中进行安装。第二种是优先级表,指为数据业务计算主备路径完成之后,采取高低优先级流表项的方式,为主路径的流表项设置高优先级,为备份路径的流表项设置低优先级,发生故障时触发主路径流表项失效机制,启动低优先级流表,即切换到备份路径^[6]。

备份路径的存储方式主要有两种,第一种是将备份路径存储在路由器软件上,切换时进行调用。第二种是将备份路径存储在交换芯片上,切换时直接使用。显然存储在交换芯片上效率更高,存储在路由器软件上更灵活^[7]。

备份路径的传输方式主要有两种:第一种是继续按照 IP 路由的方式转发,通过 SDN 控制器下发路由表时可以保证全网路由的同步性,如果是分布式路由机制可能导致全网路由未收敛造成路由环路。第二种是标签转发方式的源路由技术,从中间节点到目的节点通过标签规定了路由的中继跳,不会出现路由环路^[8]。

3 技术选择

经过上述分析并结合研究现状,在传统的路由交换设备(非 SDN 交换机,无 SDN 控制器)上实现链路故障主动恢复将采用的技术包括:接口故障检测技术;中间节点到目的节点的备份路径计算方式;组表方式的备份路径切换机制,考虑到工程实现的通用性,拟通过 FPGA 加载备份路径,并通过告警机制引导业务流到 FPGA 生成标签转发路径;备份路径存储在 FPGA 上,提高效率;SR 标签转发的源路由由数据转发技术,保证备份路径的到达率和无环性。

4 实现方法

4.1 标签同步

标签的格式:链路标签(SR-Link-Label)的标签值由 L

标志、SR-Link-SID 字段组成,遵循标准的 MPLS 标签 20bit 长度定义,L 标志值固定为 1,SR-Link-SID 由各节点为每条链路的邻居独立分配,格式如图 1 所示。

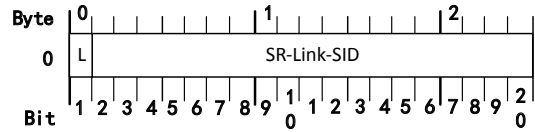


图 1 标签格式

标签的定义:链路标签由 OSPF 邻居握手后本地产生,只具有本地意义,不具有全局意义。

标签的同步:采用 OSPF 链路状态同步机制,增加 LSA (链路标签:本节点核心 IP、邻居节点核心 IP、出接口、链路标签值)以同步全网的链路标签,为标签路径计算做准备。

4.2 标签路径

采用两次运行 OSPF 的最短路径算法分别获取主路径和备份路径^[9]。基于全网拓扑数据进行第一次最短路径计算,得到下发给交换网络的转发表。剔除第一次计算结果的链路,再进行第二次最短路径计算,得到下发给 FPGA 的标签路径。如图 2 所示,C 节点计算目的地为 104.0.0.1,第一次最短路径计算后,得到 IP 转发路由表,出接口为 IF01。第二次最短路径计算得到标签路径:0x80001—0x80002—0x80004,出接口为 IF02。

4.3 故障检测

接口故障检测的时间是故障恢复时间性能指标的关键。接口故障检测依靠两种方法:控制程序业务级判断、接口板卡判断。工程实现时根据设备情况进行选择。

控制程序业务级判断方法是通过软件生成检测报文,在相邻的设备之间发送和接收,超过规定时间未收到检测报文即判断接口断开。假定检测报文每 10ms 发送一次,精度一般可到 50ms 级别。更快周期的检测报文不建议使用,容易造成 CPU 负担过重。

以太网板卡的快速检测技术是依托目前最新的网卡技术,可以毫秒级别检测到接口故障并上报 CPU。

4.4 标签栈生成

检测到接口故障时通知 CPU,由 CPU 形成控制报文给交换网络的驱动程序,驱动程序直接根据接口号将关联路由转发表的出接口转为 FPGA 接口。FPGA 根据报文目的 IP 地址进行查表,为数据包封装标签栈。如图 2 所示,C 节点 IF02 接口故障时封装的标签栈为:0x80001—0x80002—0x80004。

4.5 标签转发

交换芯片必须支持标签转发,且对于链路标签的执行规则是提前一跳弹出标签。如图 2 所示,C 节点交换芯片从 FPGA 接收到含标签栈的报文,即弹出顶层标签 0x80001,然后根据标签转发表(OSPF 邻居握手后下发)获取下一跳

出接口和网关地址 (IF: 01-101.0.0.1), 然后转发到节点 A (101.0.0.1)。节点 A 继续执行标签转发, 弹出顶层标签 0x80002, 然后根据标签转发表获取下一跳出接口和网关地址 (IF: 02-102.0.0.1), 然后转发到节点 B (102.0.0.1)。节点 B 继续执行标签转发, 弹出顶层标签 0x80004, 然后根据标签转发表获取下一跳出接口和网关地址 (IF: 04-104.0.0.1), 然后转发到节点 D (104.0.0.1)。数据报文到达节点 D 时已经还原为标准的 IP 报文, 可以继续执行标准的 IP 转发。

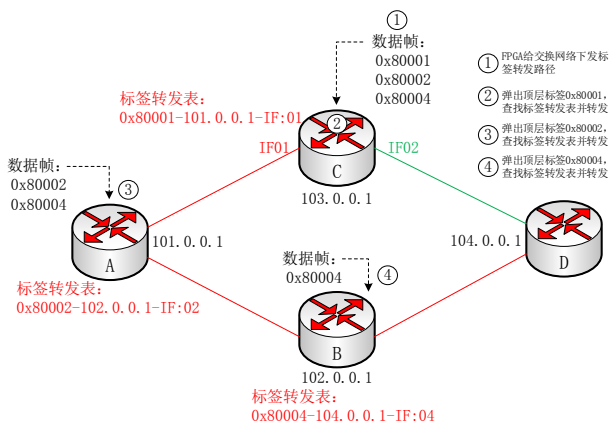


图 2 标签转发数据的流程

5 算力分析

获取全网拓扑数据后, 运行一次最短路径计算可以算出本节点到全网其他节点的路径。但是在备用路径的计算上, 计算的次数是 $N-1$ (N 代表网络节点数)^[10]。

如图 2 所示, 计算 C—D 的备用路径时, 需要剔除 CD 之间的链路; 计算 C—A 的备用路径时, 需要剔除 CA 之间的链路; 计算 C—B 的备用路径时, 需要剔除 CA 或者 CD 之间的链路。

假如存在多系统互联或者接入大型网络的情况, 建议仅

就系统内部进行备份路径的计算, 避免该算法向系统外无限延伸, 增加算力负担。可以增加系统编号 (非 OSPF 区域号) 加入 LSA 消息中, 用以隔绝系统外的标签。

6 结语

通过 OSPF 的同步机制和最短路径算法, 能够计算从原始节点到目的节点的多条路径, 然后分别通过 IP 转发和标签转发让数据流按照指定路径传输, 结合快速故障检测方法即能实现业务快速切换到保护路径, 增强了系统的抗毁性。

参考文献

- [1] 张淑娟, 王颖, 邱雪松, 等. 基于负载均衡的 SDN 数据平面故障恢复方法[C]//2016 年全国通信软件学术会议程序册与交流文集, 2016.
- [2] 冯斯详. 基于共享环路与流量负载均衡的 SDN 数据平面故障恢复方法[D]. 北京: 北京邮电大学, 2019.
- [3] 侯乐, 汪硕, 林毅凯, 等. 基于 SDN 的链路故障恢复[J]. 电信科学, 2015, 31(6): 6.
- [4] 张淑娟. 基于备份资源的 SDN 数据平面故障恢复方法[D]. 北京: 北京邮电大学, 2017.
- [5] 段洁, 高江明, 程克非, 等. 基于流类型的 SDN 数据平面故障恢复算法[J]. 重庆邮电大学学报, 2018, 30(1): 135-136.
- [6] 许延伟, 黄志林, 徐晔, 等. 一种基于标签交换的 OpenStack SDN 网络高效流表方案[J]. 电脑知识与技术, 2018, 14(9): 54-55.
- [7] 孙宇, 梁毅娟. 针对有限 TCAM 的 SDN 网络灵活局部路由故障恢复[J]. 现代电子技术, 2018, 41(8): 14-15.
- [8] 黄建洋, 兰巨龙, 胡翔宇, 等. 一种 SDN 中基于 SR 的多故障恢复与规避机制[J]. 电子学报, 2017, 45(11): 2762-2763.
- [9] 崔文岩, 孟相如, 杨欢欢, 等. QoS 约束的链路故障多备份路径恢复算法[J]. 电子与信息学报, 2016, 38(8): 1851-1852.
- [10] 金勇, 刘亦星, 王欣欣. 基于 SDN 的数据中心网络多路径流量调度算法[J]. 计算机科学, 2019, 46(6): 91-92.