OSPFv3 协议中平滑重启机制的实现①

胡建萍, 平 诞

(杭州电子科技大学 电子信息学院, 杭州 310018)

摘 要: 探讨了一种在 OSPFv3 协议中支持 GR 机制的方法来避免路由器设备重启过程中产生的路由振荡和数据转发中断, 从而提高了网络的可靠性. 在实现 OSPFv3 协议基本功能的基础上, 重点阐述了 IETF 标准 GR 机制的工作原理和具体实现, 详细分析了 OSPFv3 GR 的运行流程和报文交互过程. 测试结果表明, GR 机制能保证在协议重启过程中数据转发流量不中断.

关键词: OSPFv3 协议; 平滑重启; 协议重启; 主备倒换

Implementation of Graceful Restart Mechanism in OSPFv3 Protocol

HU Jian-Ping, PING Dan

(School of Electronic Information, Hangzhou Dianzi University, Hangzhou 310018, China)

Abstract: This paper discusses an approach that supporting GR mechanism in OSPFv3 protocol to avoid the routing oscillation and data forwarding interruption during router restarting, thereby improving the reliability of the network. Based on the realization of OSPFv3 basic functions, it focuses on the working principle and the concrete realization of IETF standard GR mechanism, analysis the running process and message interaction of OSPFv3 GR in detail. The test results show that the GR mechanism can ensure the data forwarding flow without interruption during protocol restarting.

Key words: OSPFv3; graceful restart; protocol restart; main/standby switch

近年来, 随着互联网技术的迅速发展, 网络规模不断扩大, 网络中的路由器数量不断增多. 与此同时, 互联网业务也越来越依赖于 IP 网络, 这也对 IP 网络运行的可靠性和持续性提出了更高的要求[1].

当今的主流路由器都由集中式设备向分布式设备过渡. 在分布式设备中其控制层面和转发层面相分离, 控制层面专注于业务的控制和管理, 而转发层面专注于数据的接收和转发. 在这种分布式处理情况下, 当控制层面进行协议重启时, 转发层面必须保证数据流量的不中断, 这对于运营商网络和大型企业网络来说至关重要[2].

GR(Graceful Restart, 平滑重启)是保证在设备重启时转发业务不中断的一种高可靠性技术, 能在协议重启或主备倒换过程中避免产生路由振荡和数据转发中断, 使整个系统保持不间断运行. 随着网络可靠性要求的不断提高, 各种主流路由协议对 GR 机制的支持和融合已成为一种必然趋势.

1 OSPFv3协议简介

OSPF(Open Shortest Path First, 开放最短路径优先)是 IETF 组织开发的一种基于链路状态的内部路由协议. OSPFv3 是 OSPF 版本 3 的简称, 主要提供对 IPv6 网络的支持.

OSPFv3 协议的工作机制是各设备之间通过周期性的交互 Hello 报文, 建立并维持邻居关系, 在形成邻接关系的设备之间互相扩散用于描述链路状态的 LSA(Link State Advertisement, 链路状态通告), 并且最终形成相同的 LSDB(Link State DataBase, 链路状态数据库), 在 LSDB 的基础上进行路由计算生成路由信息并下发到路由表中[3].

① 收稿时间:2012-04-11;收到修改稿时间:2012-05-21

1.1 邻居关系的建立

运行 OSPFv3 协议的路由器通过在接入网络的所有接口上周期性的发送 Hello 报文来宣告自身和发现邻居, 同时接收网络中的其它路由器发送的 Hello 报文. 达到双向邻居关系是指路由器收到的邻居发送的 Hello 报文中包含了自己.

形成双向邻居关系后, 两台路由器之间可能需要进一步形成邻接关系(根据网络类型和接口状态的不同情况而定). 为了保证自治系统内的所有路由器对网络拓扑结构形成一致的视图, OSPFv3 协议规定形成邻接关系的路由器之间必须保持 LSDB 的同步. LSDB 的同步过程中, 每台路由器通过 DD(Database Description, 数据库描述)报文向邻居描述自身的 LSDB 汇总情况. 邻居双方在获取对方的 LSDB 汇总情况后, 通过 LSR(Link State Request, 链路状态请求)报文向对方请求自身所需的 LSA, 通过 LSU(Link State Update, 链路状态更新)报文向对方发送其所需的 LSA, 通过 LSAck(Link State Acknowledgment, 链路状态确认)报文对收到的 LSA 进行确认, 直到双方的 LSDB 达到同步, 形成完全邻接关系.

1.2 协议重启

在正常情况下, 运行 OSPFv3 协议的路由器进行协议重启时, 会重新进行邻居关系的建立和 LSDB 的同步. 为了避免由于缺少 LSDB 的同步而造成路由环路或黑洞路由, 周边路由器认为该重启路由器在报文转发路径上已不可用, 会将其从邻居列表中删除, 断开与其的邻居关系, 重新生成相应的 LSA 来更新链路状态信息并通知其他路由器. 路由器协议重启结束后, 与周边路由器再次建立全邻接关系并同步 LSDB, 而周边路由器也会重新进行路由计算. 这样就会造成在路由器协议重启过程中, 网络中出现路由振荡和转发业务中断.

然而, 如果在协议重启过程中, 网络拓扑仍然保持稳定不变, 重启路由器仍然能保持它的转发表正常进行报文转发, 则该重启路由器在报文转发路径上仍然是安全可用的, 周边路由器可继续维持与其的邻居关系, 无需删除邻居信息以及重新产生相应的 LSA.

2 GR机制实现

GR 是一种在协议重启或主备倒换时保证转发业务不中断的机制. 当设备进行协议重启或主备倒换时, 能够通知其周边设备, 使周边设备到该设备的邻居关系和路由信息在一定时间内保持稳定. 在设备协议重启或主备倒换结束后, 周边设备协助其进行数据(包括各种拓扑、路由和会话信息)同步, 在较短时间内恢复到重启前的状态. 在协议重启或主备倒换过程中不会产生路由振荡, 报文转发路径也没有任何改变, 整个系统可以实现不间断运行.

GR 过程中有两个角色: GRRestarter 和 GRHelper.

GRRestarter: 发生协议重启或主备倒换事件且具有 GR 能力的设备.

GRHelper: 和 GRRestarter 具有邻居关系, 协助完成 GR 的设备[4].

2.1 Grace LSA

Grace LSA 是由进行协议重启或主备倒换的设备生成, 用于通告邻居设备自身进入 GR 重启流程的一种 LSA, 其中携带了 GR 的相关信息. OSPFv3 中 Grace LSA 的格式如图 1 所示.

<table>
<tr><td colspan="2">LS Age</td><td>U</td><td>S1</td><td>S2</td><td colspan="2">LS Type</td></tr>
<tr><td colspan="7">Link State ID</td></tr>
<tr><td colspan="7">Advertising Router</td></tr>
<tr><td colspan="7">LS Sequence</td></tr>
<tr><td colspan="3">LS Checksum</td><td colspan="4">LS Length</td></tr>
<tr><td colspan="3">Type</td><td colspan="4">Length</td></tr>
<tr><td colspan="7">Value</td></tr>
<tr><td colspan="3">Type</td><td colspan="4">Length</td></tr>
<tr><td colspan="7">Value</td></tr>
</table>

图 1 Grace LSA

其中, LS Age 为该 Grace LSA 自生成后所经过的时间. LS Type 为 LSA 所属的类型, Grace LSA 的 LS Type 为 0x000b. Link State ID 和 Advertising Router 分别为发送该 Grace LSA 的接口 ID 和路由器 ID. LS Sequence、LS Checksum 和 LS Length 分别为该 Grace LSA 对应的序列号、校验和及长度. TLV 部分为该 Grace LSA 的数据部分, 其中携带了 GR 的相关信息. OSPFv3 协议中, 每个 Grace LSA 都带有两个 TLV 三元组:

① Type = 1, Length = 4 的 TLV, 此时 Value 为 GR 周期时间间隔. GR 周期时间长度由重启设备在进入 GR 重启流程时设置, 用于通告周边设备在该时间段内继续维持到重启设备的邻居关系, 并协助其进行信息同步;

② Type = 2, Length = 1 的 TLV, 此时 Value 为设备进入 GR 的原因. 取值为 0 表示未知原因, 为 1 表示

软件重启，为 2 表示软件重载/升级，为 3 表示主备倒换[5].

2.2 GRRestarter 实现

GRRestarter 进入 GR 重启流程后，会重新进行邻居关系的建立和 LSDB 的同步. 在 GR 重启过程中，GRRestarter 不会生成各类 LSA，收到重启前自己生成的 LSA 也不进行刷新处理，而是直接接收并打上老化标记，等GR重启流程结束后统一进行刷新处理. 周边设备仍使用 GRRestarter 重启前生成的 LSA 进行路由计算，以保证 GR 过程中路由表项的稳定和各类转发业务的不中断.

当 GRRestarter 在 GR 周期内与 GRHelper 重新建立了邻接关系并同步了 LSDB，则 GRRestarter 完成GR重启，退出GR流程，进入OSPFv3正常流程. 在退出GR状态后，GRRestarter会Flush Grace LSA，即生成并向 GRHelper 发送 LS Age 为 3600s 的 Grace LSA，以此通告 GRHelper 自身已完成 GR 重启. 同时，GRRestarter需要向接入的所有区域重新产生各类LSA，在 GR 过程中收到的自己产生并打上老化标记的 LSA 也需要统一进行老化和刷新处理. 此外，GRRestarter 还需要重新触发全部路由计算并下发路由表，刷新相应表项.

2.3 GRHelper 实现

周边设备在收到GRRestarter发送的Grace LSA后，从中提取出 GR 周期和 GR 重启原因等 GR 相关信息，进行报文格式、邻居状态和网络拓扑等一系列判定[6]. 若判定通过，则进入 GR 流程，成为 GRHelper.

设备在成为 GRHelper 后，需要在 GR 周期时间间隔内协助 GRRestarter 完成 GR 重启流程. 当收到 GRRestarter 发送的用于重建邻居关系的 1-way Hello 报文时，由于之前已经建立了邻居关系并成功进行了 GR 协商，GRHelper 并不会将 GRRestarter 从邻居列表中删除，也不会重新生成各类 LSA 并进行路由计算，而是当作用于维持邻居关系的 2-way Hello 报文来进行处理，继续通告并维持与 GRRestarter 的邻居关系，以此来保证相应路由表项的稳定和各类转发业务的不中断. 当 GRRestarter 提出 LSDB 同步请求时，GRHelper 会将完整的 LSDB 洪泛给 GRRestarter，协助其进行数据的恢复和同步.

当 GRHelper 收到 LS Age 为 3600s 的 Grace LSA 时，若已和 GRRestarter 重新建立了邻接关系并同步了 LSDB，说明 GRRestarter 已经成功完成 GR 重启，则 GRHelper 退出 GR 流程，进入 OSPFv3 正常流程[7]. 在退出 GR 状态后，GRHelper 需要向接入的所有区域重新生成各类 LSA，并且触发全部路由计算刷新相应路由表项.

2.4 OSPFv3 GR 运行过程

IETF 标准 GR 的运行过程如图 2 所示. 其中，RT1 为分布式设备，具有主备环境. 假设 RT1 和 RT2 已经建立了邻接关系，LSDB 达到同步状态，并且 RT1 上使能了 GRRestarter 能力，RT2 上使能了 GRHelper 能力.

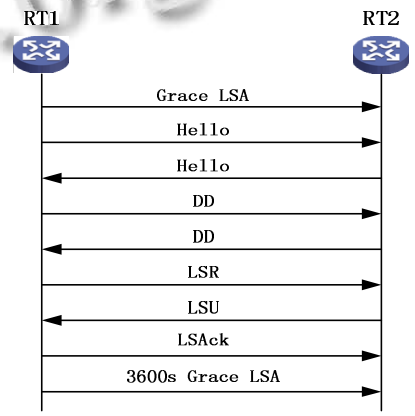

图 2　OSPFv3 GR 运行过程

此时若 RT1 进行协议重启或主备倒换，将按以下流程进行相应处理:

1)RT1 发生协议重启或主备倒换后，进行 GR 能力检查，若允许进入 GR 状态，则成为 GRRestarter，进入 GR 重启流程;

2)RT1 成为 GRRestarter 后，生成并向 RT2 发送 Grace LSA，其中携带了 GR 周期时间间隔和 GR 重启原因等 GR 相关信息，通告 RT2 自身进入 GR 重启流程;

3)RT2 收到 RT1 发送的 Grace LSA 后，提取 GR 相关信息，进行报文格式、邻居状态以及网络拓扑等检查. 若自身允许成为 GRHelper，则进入 GR 流程，成为 GRHelper，在 GR 周期时间间隔内协助 RT1 进行 GR 重启;

4)RT1 协议重启后试图与 RT2 重新建立邻居关系，向 RT2 发送 1-way Hello 报文，其中并不通告与 RT2 的邻居关系;

5)RT2 收到 RT1 发送的 1-way Hello 报文后，由于

之前已经与 RT1 建立了邻居关系并成功进行了 GR 协商，因此将此 1-way Hello 报文当做 2-way Hello 报文处理，在向 RT1 发送的 2-way Hello 报文中继续通告与 RT1 的双向邻居关系;

6)RT1 收到 RT2 发送的 2-way Hello 报文后，开始与 RT2 进行 DD 报文交互和 LSDB 同步. RT1 通过 LSR 报文向 RT2 请求所需的 LSA. RT2 则通过 LSU 报文将完整的 LSDB 洪泛给 RT1，协助 RT1 进行数据恢复. RT1 收到 LSA 后通过 LSAck 报文进行确认;

7)RT1 在与 RT2 重新建立了全邻接关系并同步了 LSDB 后，向 RT2 发送 LS Age 为 3600s 的 Grace LSA，通知 RT2 结束 GR 流程. RT1 在退出 GR 重启流程后，进入 OSPFv3 正常流程，重新生成各类 LSA，老化和刷新 GR 过程中收到的自己生成的 LSA，并且触发全部路由计算更新路由信息;

8) RT2 收到 RT1 发送的 LS Age 为 3600s 的 Grace LSA 后，若已与 RT1 重新建立了全邻接关系并同步了 LSDB，则退出 GR 流程，进入 OSPFv3 流程，重新生成各类 LSA，并且触发全部路由计算更新路由信息.

3 组网测试与功能验证

将三台路由器设备 RT1、RT2 和 RT3 按图 3 建立测试组网，其中 RT2 为分布式设备，具有主备环境.

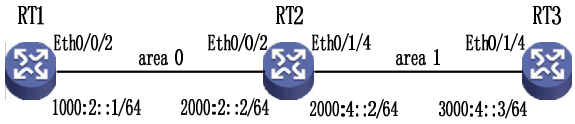

图 3 测试组网

RT2 通过 Eth0/0/2 接口与 RT1 建立区域 0 的邻居，通过 Eth0/1/4 接口与 RT3 建立区域 1 的邻居. 通过 ping 命令来测试 RT1 和 RT3 之间的连通性以及监控数据流量是否中断，在 RT3 上不断 ping RT1 的 Eth0/0/2 接口 IPv6 地址 1000:2::1/64. 正常情况下 RT3 可以一直 ping 通 RT1.

首先测试未使能 GR 能力的情况，在 RT2 上进行 OSPFv3 协议进程重启或主备倒换操作，RT3 上查看 ping 命令执行结果如图 4 所示. 可以看到中间出现了 RT3 无法 ping 通 RT1 的现象，说明 RT2 在协议重启或主备倒换过程中出现了数据转发流量中断.

```
[RT3]ping ipv6 -c 1000 -m 1000 1000:2::1
PING6(104=40+8+56 bytes) 3000:4::3 -->1000:2::1
56 bytes from 1000:2::1,icmp_seq=0 hlim=63 time=16.000ms
56 bytes from 1000:2::1,icmp_seq=1 hlim=63 time=13.000ms
56 bytes from 1000:2::1,icmp_seq=2 hlim=63 time=17.000ms
56 bytes from 1000:2::1,icmp_seq=3 hlim=63 time=68.000ms
Request time out
Request time out
Request time out
Request time out
56 bytes from 1000:2::1,icmp_seq=8 hlim=63 time=8.000ms
56 bytes from 1000:2::1,icmp_seq=9 hlim=63 time=13.000ms
56 bytes from 1000:2::1,icmp_seq=10 hlim=63 time=13.000ms
56 bytes from 1000:2::1,icmp_seq=11 hlim=63 time=11.000ms
56 bytes from 1000:2::1,icmp_seq=12 hlim=63 time=18.000ms
--- 1000:2::1 ping6 statistics ---
13 packet(s) transmitted, 9 packet(s) received, 30.8% packet loss
round-trip min/avg/max/std-dev = 8.000/19.667/68.000/17.333 ms
```

图 4 未使能 GR 情况

再测试使能 GR 能力的情况，在 RT2 上使能 GRR-estarter 能力，RT1 和 RT3 上使能 GRHelper 能力. RT2 进行 OSPFv3 协议进程重启或主备倒换操作，RT3 上查看 ping 命令执行结果如图 5 所示. 可以看到 RT3 一直可以 ping 通 RT1，说明 RT2 在进行协议重启或主备倒换过程中未出现数据转发流量中断，一直能正常进行数据转发.

```
[RT3]ping ipv6 -c 1000 -m 1000 1000:2::1
PING6(104=40+8+56 bytes) 3000:4::3 -->1000:2::1
56 bytes from 1000:2::1,icmp_seq=0 hlim=63 time=15.000ms
56 bytes from 1000:2::1,icmp_seq=1 hlim=63 time=150.000ms
56 bytes from 1000:2::1,icmp_seq=2 hlim=63 time=30.000ms
56 bytes from 1000:2::1,icmp_seq=3 hlim=63 time=25.000ms
56 bytes from 1000:2::1,icmp_seq=4 hlim=63 time=28.000ms
56 bytes from 1000:2::1,icmp_seq=5 hlim=63 time=18.000ms
56 bytes from 1000:2::1,icmp_seq=6 hlim=63 time=54.000ms
56 bytes from 1000:2::1,icmp_seq=7 hlim=63 time=22.000ms
56 bytes from 1000:2::1,icmp_seq=8 hlim=63 time=35.000ms
56 bytes from 1000:2::1,icmp_seq=9 hlim=63 time=18.000ms
56 bytes from 1000:2::1,icmp_seq=10 hlim=63 time=11.000ms
56 bytes from 1000:2::1,icmp_seq=11 hlim=63 time=31.000ms
56 bytes from 1000:2::1,icmp_seq=12 hlim=63 time=23.000ms
--- 1000:2::1 ping6 statistics ---
13 packet(s) transmitted, 13 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 11.000/35.385/150.000/34.687 ms
```

图 5 使能 GR 情况

测试结果表明，通过在 OSPFv3 协议中支持 GR 机制，可以在协议重启或主备倒换过程中避免产生数据转发流量中断，从而提高了网络的可靠性.

4 结语

本文在 OSPFv3 协议的基础上探讨了一种高可靠性技术——GR 机制，可以保证路由器在进行协议重启或主备倒换过程中数据转发流量不中断，网络中不会产生路由振荡和业务中断. 整个系统能保持不间断

(下转第 193 页)

持 0.48 左右的归一化寿命后网络规模达到阈值(即 10).

3 结论

定位技术是 WSN 必不可少的一项关键技术, 其提供的位置信息为事件监测或目标位置信息获取、路由协议、覆盖质量及其他相关研究起到关键性作用. 然而, 节点的定位信息一旦被非法滥用, 必将导致严重位置隐私问题.

但对于位置隐私问题的以往研究都是假设窃听者不能监控整个网络. 相对于良好协调、严重的攻击, 这个假设是无效的. 本文假定存在全局窃听者, 通过 LP 框架, 提出并形式化 FS 和 BF 这两种 sink 隐藏方法. 我们分析并比较了两种方法对网络寿命的影响. 研究结果表明: 保护 sink 不可观测对网络寿命的影响相当大. 同时也表明: BF 实现的网络寿命数量级高于大型网络的 FS.

参考文献

1 姚剑波.基于定向贪心游走的 WMSN 位置隐私.计算机应用与软件,2011,28(3):137−138,165.

2 刘昭斌,刘文芝,顾君忠.位置感知的自适应隐私保护策略.计算机工程与设计,2011,32(3):839−841,1032.

3 陈娟,方滨兴,殷丽华,等.传感器网络中基于源节点有限洪泛的源位置隐私保护协议.计算机学报,2010,33(9):1736−1747.

4 彭志宇,李善平.移动环境下 LBS 位置隐私保护.电子与信息学报,2011,33(5):1211−1216.

5 任丹丹,杜素果.一种基于攻击树的 VANET 位置隐私安全风险评估的新方法.计算机应用研究,2011,28(2):728−732.

6 Yang Y, Shao M, Zhu S, et a1. Towards event source unobservability with minimum network traffic in sensor networks. Proc. of the ACM Wisec'08, USA: Alexandria and VA,2008.77−88.

7 Cheng Z, Perillo M, Heinzelman W. General network lifetime and cost models for evaluating sensor network deployment strategies. IEEE Trans. on Mobile Computing, 2008,7(4): 484−497.

8 Brook A, Kendrick D, Meeraus A, et a1. GAMS: A User's Guide. The Scientific Press, 1998.

(上接第 211 页)

运行, 从而提高了网络的稳定性和可靠性. 该实现机制在大型企业网和运营商网络中具有广泛应用.

参考文献

1 孟华.互联网路由技术及发展前景展望.中国新技术新产品,2011,19(15):19−20.

2 王勐.核心路由器高可用性研究.信息与电脑,2011,23(9):151−152.

3 Coltun R, Ferguson D, Moy J. OSPF for IPv6. IETF RFC2740, 1999.

4 Moy J, Pillay-Esnault P, Lindem A. Graceful OSPF Restart. IETF RFC3623, 2003.

5 Pillay-Esnault P, Lindem A. OSPFv3 Graceful Restart. IETF RFC5187, 2008.

6 孙作聪,王立松,顾宝根.基于 OSPF 的温和重启的触发机制的研究与实现.计算机工程与技术,2006,27(14):2653− 2656.

7 张丹,商云飞,张显峰.基于 OSPF 协议的 Graceful Restart 技术的研究与实现.仪器仪表用户,2007,14(6):21−22.

(上接第 221 页)

用,2003,23(4):26−28.

3 Ratle A, Sebag M. Genetic Programming and Domain Knowledge: Beyond the Limitations of Grammar-Guided Machine Discovery. Parallel Problem Solving form Nature(PPSN 2000), Berlin: Springer, 2000,211−220.

4 朱彦廷.基于遗传算法的关联规则挖掘.西昌学院学报,2010,24(3):60−62.

5 KoZa, John R, Keane MA, Yu J, Bennett FH, Mydlowee W. Automatic creation of human-competitive programs and controllersby means of genetie programming. Genetlc Prograrnming and Evolvable Maehines, 2000,l(1-2):124−164.

6 徐哲,白焰.遗传编程.自动化仪表,2002,23(10):1−6.

7 徐扬,任庆生,戚飞虎.一个基于遗传编程的机器人足球系统.计算机仿真,2005,22(4):178−182.

8 江海燕.关于 GP 的研究与探索.山东教育学院学报,2007,1:96−100.