bgpd gr实现

bgp gr功能两个角色，如下

- gr Restarter gr重启的发起者
- gr Helper 协助Restarter完成gr重启

> 版本: frr-7.2.1-s3

gr配置

```
    /* "bgp graceful-restart" commands */
    install_element(BGP_NODE, &bgp_graceful_restart_cmd); // 设置是否启动gr
    install_element(BGP_NODE, &bgp_graceful_restart_stalepath_time_cmd);
// 设置gr helper持有路由时间
    install_element(BGP_NODE, &bgp_graceful_restart_restart_time_cmd); //
设置open消息发送给对端的gr重启时间
    install_element(BGP_NODE, &bgp_graceful_restart_preserve_fw_cmd); // 设
置F标志

    /* "bgp graceful-shutdown" commands */
    install_element(BGP_NODE, &bgp_graceful_shutdown_cmd);
    install_element(BGP_NODE, &no_bgp_graceful_shutdown_cmd);

/* Graceful Restart */
#define RESTART_R_BIT              0x8000
#define RESTART_F_BIT              0x80
```

gr流程

```
(R1)-------------------------------------------------------------------(R2)

1. BGP Graceful Restart Capability exchanged between R1 & R2.

<-------------------------------------------------------------------->

2. Kill BGP Process at R1.

--------------------------------------------------------------------->

3. R2 Detects the above BGP Restart & verifies BGP Restarting
```

```
  Capability of R1.

4. Start BGP Process at R1.

5. Re-establish the BGP session between R1 & R2.

<-------------------------------------------------------------------->

6. R2 Send initial route updates, followed by End-Of-Rib.

<--------------------------------------------------------------------

7. R1 was waiting for End-Of-Rib from R2 & which has been received
  now.

8. R1 now runs BGP Best-Path algorithm. Send Initial BGP  Update,
  followed by End-Of Rib

<-------------------------------------------------------------------->
```

open消息

open消息发送

配置无关设置

- 默认使能 restart通告能力（PEER_CAP_RESTART_ADV）
  - PEER_CAP_RESTART_ADV表明本端支持gr
  - open消息
    - 如果bgp实例重启
      - 设置R标志，peer设置PEER_CAP_RESTART_BIT_ADV
      - 设置restart_time，默认120s
    - 否则：仅设置restart_time

配置相关设置

- 如果配置了gr：graceful_restart
  - 如果配置了：preserve_fw
    - 协议族携带F标志，能够gr期间持有该协议路由，bgp设置BGP_FLAG_GR_PRESERVE_FWD
  - 否则协议族不带F标志

open消息接收

gr能力解析

- 设置PEER_CAP_RESTART_RCV
- 如果设置了R标志
  - peer设置PEER_CAP_RESTART_BIT_RCV
- 设置peer v_gr_restart时间
- 协议族支持gr
  - 设置PEER_CAP_RESTART_AF_RCV

        - 表明：对端支持gr
    - 如果带F标志：设置PEER_CAP_RESTART_AF_PRESERVE_RCV

gr协商

通过自身gr配置和对端的gr能力通告，确定peer是否使能GR(NSF_MODE模式)

peer建立时会设置PEER_STATUS_NSF_MODE模式

```
/* graceful restart */
UNSET_FLAG(peer->sflags, PEER_STATUS_NSF_WAIT);
for (afi = AFI_IP; afi < AFI_MAX; afi++)
  for (safi = SAFI_UNICAST; safi <= SAFI_MPLS_VPN; safi++) {
    if (peer->afc_nego[afi][safi] // 已协商的协议族，双方都支持该协议族
        && CHECK_FLAG(peer->cap, PEER_CAP_RESTART_ADV) // 本端支持gr
        && CHECK_FLAG(peer->af_cap[afi][safi],
          PEER_CAP_RESTART_AF_RCV)) { // 对端支持gr
      if (peer->nsf[afi][safi]
          && !CHECK_FLAG(
                peer->af_cap[afi][safi],
                PEER_CAP_RESTART_AF_PRESERVE_RCV))
        bgp_clear_stale_route(peer, afi, safi);

      peer->nsf[afi][safi] = 1; // 该协议族，支持gr
      nsf_af_count++;
    } else {
      if (peer->nsf[afi][safi])
        bgp_clear_stale_route(peer, afi, safi);
      peer->nsf[afi][safi] = 0;
    }
  }

if (nsf_af_count) // 有支持gr的协议族，则设置NSF_MODE模式
  SET_FLAG(peer->sflags, PEER_STATUS_NSF_MODE);
```

设置NSF_MODE模式的条件有三个

- 协议族af协商 peer->afc_nego[afi][safi]
    - 只要peer有确定的协议族就使能了af协商
- Restart通告能力 CHECK_FLAG(peer->cap, PEER_CAP_RESTART_ADV)
    - FRR默认设置
- Restart AF接收能力 CHECK_FLAG(peer->af_cap[afi][safi], PEER_CAP_RESTART_AF_RCV)
    - 只要gr能力中携带该协议族，就设置。与F标志无关
    - 对应FRR而言，只要配置graceful_restart就设置

简单地说，只要对端配置graceful_restart就设置NSF_MODE

触发gr

GR Restart：kill bgpd GR Helper：代码层面有两处感知gr重启，当认为是gr重启时，会设置为NSF_WAIT状态

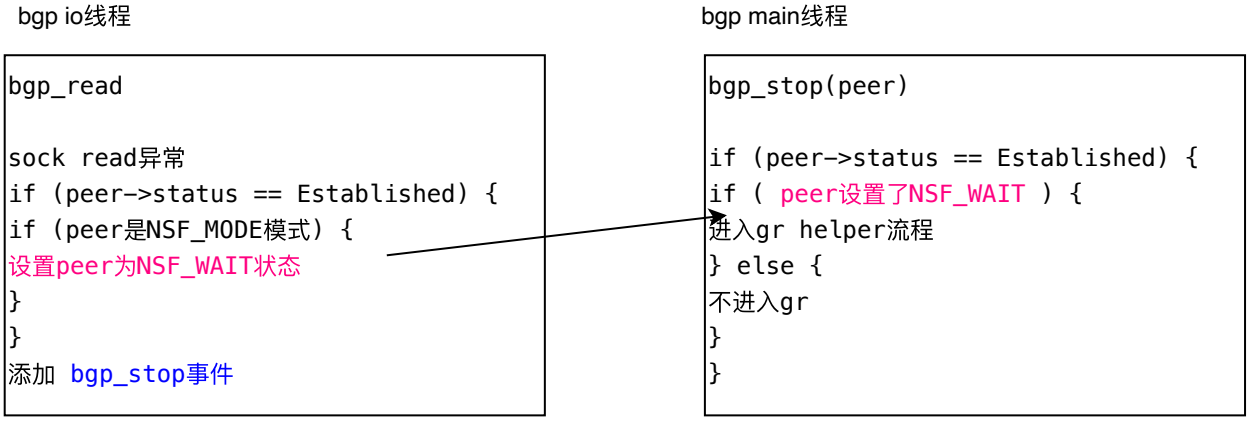
- bgp_read：peer sock read异常（fd->read返回<=0）
- bgp_accept：重新tcp建链

stale路由处理

TODO

缺陷

对端执行bgp neigh shutdown可能会触发其进入gr helper流程，gr感知流程如下



概率性描述

- 对端执行neigh shutdown后，main线程收到notification消息后， 会进入bgp_stop流程
- 如果在main线程进入bgp_stop前，io线程感知到read异常，则会设置为NSF_WAIT状态
  - 进入bgp_stop流程后，NSF_WAIT状态为True，则进入gr helper流程
- 反之，不会进入gr helper流程

另外NSF_WAIT的设置是在io线程中，main线程在执行bgp_stop时，NSF_WAIT状态存在可见性问题

> master分支已修复，异常处理和packet处理都以event方式添加到main线程

问题

- 对于FRR，只要地址协商一致，使能GR则表明能充当Restarter和Helper角色
- FRR的F标志是由preserve_fw_cmd控制的，如果使能了GR但未配置F标志，会有什么影响?
  - Restarter重启后可能持有了路由，但发送给对端的f_bit为0
  - Helper端收到f_bit为0，认为Restater在重启期间未能持有路由，则会将本地持有的路由都删掉，导致网络丢包
    - Restarter端有路由，Helper没路由