

## BFD 协议原理与配置

BFD ( Bidirectional Forwarding Detection ) 双向转发探测

BFD 是一种双向转发检测机制，它是介质无关和协议无关的快速故障检测机制，可以提供毫秒级的检测，可以实现链路的快速检测，BFD 通过与上层路由协议联动，可以实现路由的快速收敛，确保业务的永续性。

BFD 主要是用来实现毫秒级的切换。从而降低业务的故障率。而 BFD 不是单独启用的，通常是和 ospf vrrp 等这些路由协议和热备份协议一起使用的。比如 ospf 默认情况下，你要等待 40 秒才能知道邻居 down 了，但是 bfd 和 OSPF 一起使用在毫秒内就能发现邻居 down 了这样的话路由切换肯定要快很多。

现有的故障检测方法主要包括以下几种：

硬件检测：

例如：通过 SDH ( Synchronous Digital Hierarchy，同步数字体系 ) 告警检测链路故障。硬件检测的优点是可以很快发现故障，但并不是所有介质都能提供硬件检测。

慢 Hello 机制：

通常采用路由协议中的 Hello 报文机制。这种机制检测到故障所需时间为秒级。对于高速数据传输，例如吉比特速率级，超过 1 秒的检测时间将导致大量数据丢失；对于时延敏感的业务，例如语音业务，超过 1 秒的延迟也是不能接受的。并且，这种机制依赖于路由协议。

其他检测机制：

不同的协议有时会提供专用的检测机制，但在系统间互联互通

时，这样的专用检测机制通常难以部署。

### BFD 检测方式

单跳检测：BFD 单跳检测是指对两个直连系统进行 IP 连通性检测，这里所说的“单跳”是 IP 的一跳。

多跳检测：BFD 可以检测两个系统间的任意路径，这些路径可能跨越很多跳，也可能在某些部分发生重叠。

双向检测：BFD 通过在双向链路两端同时发送检测报文，检测两个方向上的链路状态，实现毫秒级的链路故障检测。（BFD 检测 LSP 是一种特殊情况，只需在一个方向发送 BFD 控制报文，对端通过其他路径报告链路状况。）

### BFD 的检测机制：

BFD 的检测机制是两个系统建立 BFD 会话，并沿它们之间的路径周期性发送 BFD 控制报文，如果一方在既定的时间内没有收到 BFD 控制报文，则认为路径上发生了故障，BFD 控制报文是 UDP 报文，端口号 3784。

BFD 提供异步检测模式。在这种模式下，系统之间相互周期性地发送 BFD 控制报文，如果某个系统连续 3 个报文都没有接收到，就认为此 BFD 会话的状态是 Down。

### BFD 状态机，有 3 种：Down，Init，UP

初始状态为 Down，收到状态为 Down 的 BFD 报文后，状态切换至 Init，相互收到 Init 之后，变为 UP

表1 BFD参数缺省值

参数	缺省值
全局BFD功能	未使能
发送间隔	1000毫秒
接收间隔	1000毫秒
本地检测倍数	3
等待恢复时间	0分钟
会话延迟Up时间	0秒钟
BFD报文优先级	7



## 前言

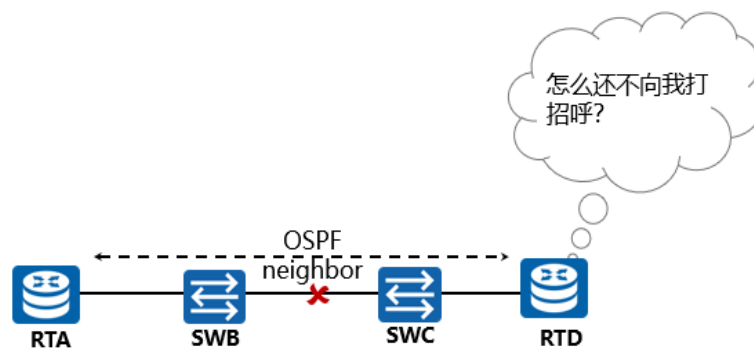
- 随着网络应用的广泛部署，网络发生中断可能影响业务正常运行并造成重大损失。为了减小链路、设备故障对业务的影响，提高网络的可靠性，网络设备需要尽快检测到与相邻设备间的通信故障，以便及时采取措施，保证业务正常进行。
- 双向转发检测BFD (Bidirectional Forwarding Detection) 提供了一个通用的标准化的介质无关和协议无关的快速故障检测机制，用于快速检测、监控网络中链路或者IP路由的转发连通状况。那么BFD机制具体是如何实现快速检测的呢？它的收敛时间到底是多少呢？
- 为了减小设备故障对业务的影响，提高网络的可靠性，网络设备需要能够尽快检测到与相邻设备间的通信故障，以便及时采取措施，保证业务继续进行。在现有网络中，有些链路通常通过硬件检测信号，如SDH告警，检测链路故障，但并不是所有的介质都能够提供硬件检测。此时，应用就要依靠上层协议自身的Hello报文机制来进行故障检测。上层协议的检测时间都在1秒以上，这样的故障检测时间对某些应用来说是不能容忍的。同时，在一些小型三层网络中，如果没有部署路

由协议，则无法使用路由协议的 Hello 报文机制来检测故障。

- BFD 协议就是在这种背景下产生的，BFD 提供了一个通用的标准化的介质无关和协议无关的快速故障检测机制。具有以下优点：

- 对相邻转发引擎之间的通道提供轻负荷、快速故障检测。这些故障包括接口、数据链路，甚至有可能是转发引擎本身。
- 用单一的机制对任何介质、任何协议层进行实时检测。
- BFD 可以实现快速检测并监控网络中链路或 IP 路由的转发连通状态，改善网络性能。相邻系统之间通过快速检测发现通信故障，可以更快地帮助用户建立起备份通道以便恢复通信，保证网络可靠性。

## 应用场景一出现的问题



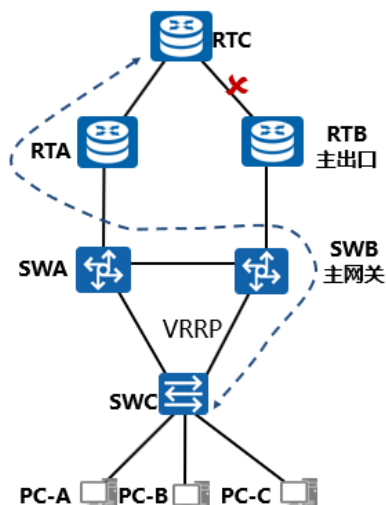
- 问题：SWB和SWC之间链路中断，RTA和RTD的OSPF邻接关系会马上中断吗？

- 场景描述：

- RTA 和 RTD 建立了 OSPF 邻接关系，Hello 包发送周期为 10 秒；当交换机 SWB 与 SWC 链路物理中断，路由器 RTA 和 RTD 无法感知，需要等待 OSPF 协议邻居失效计时器超时后才会中断邻接关系。



## 应用场景二出现的问题



- 问题：RTB与RTC发生硬件故障或链路故障，用户的业务流量是什么样的呢？

有没有解决这种问题的方法呢？



- 场景描述：
- SWA 和 SWB 启用了 VRRP 协议，实现了主备网关的作用，SWB 为主用网关。
- 当出口路由器 RTB 与外网 Router 的链路中断，SWB 虽然可以通过动态路由协议感知，但是无法联动连接下游的网关接口，且继续为主网关。
- 用户的数据流还是发送到 SWB，SWB 再通过三层路由转发给 SWA，最后由 RTA 出口。虽然结果不至于造成业务中断，但是会产生次优路径。



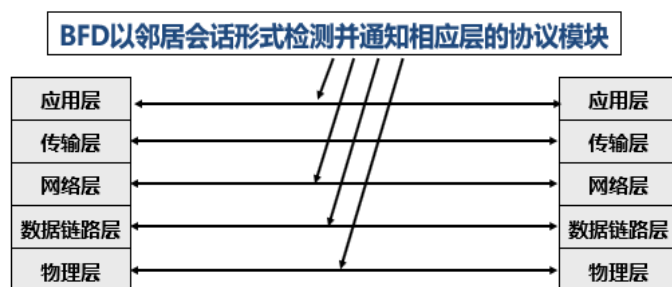
## 故障检测的方法



- 有没有一种通用性强，又能快收敛的检测机制呢？
- 硬件检测：
  - 例如通过 SDH 告警检测链路故障。硬件检测的优点是可以很快发现故障，但并不是所有介质都能提供硬件检测。
- 慢 Hello 机制：
  - 通常是指路由协议的 Hello 机制。这种机制检测到故障所需时间为秒级。对于高速数据传输，例如吉比特速率级，超过 1 秒的检测时间将导致大量数据丢失；对于时延敏感的业务，例如语音业务，超过 1 秒的延迟也是不能接受的。并且，这种机制依赖于路由协议。
- 其他检测机制：
  - 不同的协议或设备制造商有时会提供专用的检测机制，但在系统间互联互通时，这样的专用检测机制通常难以部署。

## BFD技术简介

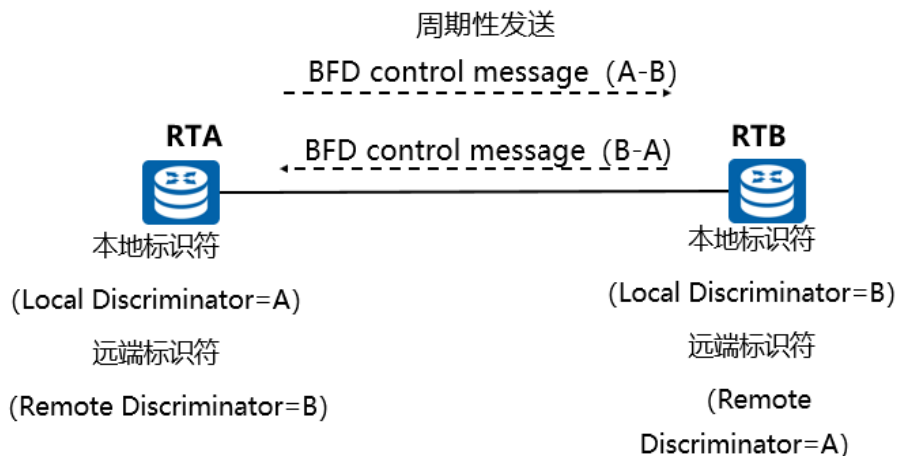
- 一种全网统一、检测迅速、监控网络中链路或者IP路由的双向转发连通状况，并为上层应用提供服务的技术。



- BFD 会话建立后会周期性地快速发送 BFD 报文，如果在检测时间内没有收到对端 BFD 报文则认为该双向转发路径发生了故障，通知被服务的相关层应用进行相应的处理。
- 本身并没有邻居发现机制，而是靠被服务的上层应用通知其邻居信息以建立会话。
- 不管是物理接口状态、二层链路状态、网络层地址可达性、还是传输层连接状态、应用层协议运行状态，都可以被 BFD 感知到。



## BFD会话建立方式和检测机制



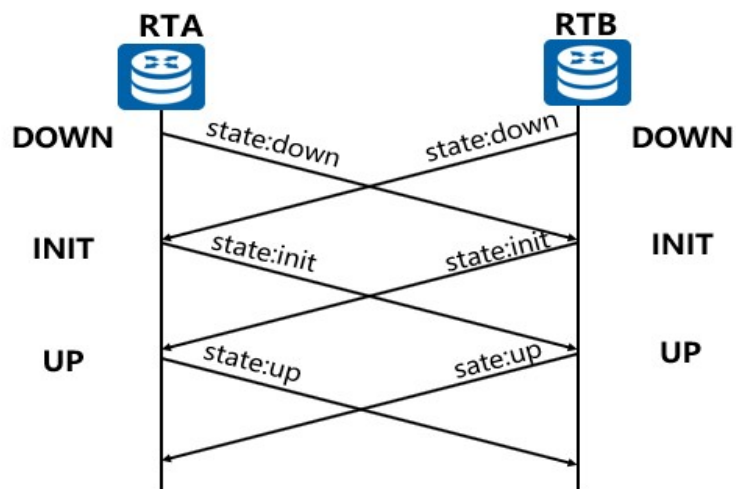
- BFD 的标识符：
- BFD 建立会话存在标识符的概念，类似于 OSPF 建立邻居需要一个路由器的 Router ID。
- 标识符分为本地标识符和远端标识符，本地标识符用于表示本端设备，远端标识符用于表示对端设备。
- 静态建立 BFD 会话是指通过命令行手工配置 BFD 会话参数，包括配置本地标识符和远端标识符等，然后手工下发 BFD 会话建立请求。
- 动态建立 BFD 会话是指由应用程序触发创建 BFD 会话，当应用程序动态触发创建 BFD 会话时，系统分配属于动态会话标识符区域的值作为 BFD 会话的本地标识符。然后向对端发送 Remote Discriminator 的值为 0 的 BFD 控制报文，进行会话协商。当 BFD 会话的一端收到 Remote Discriminator 的值为 0 的 BFD 控制报文时，判断该报文是否与本地 BFD 会话匹配，如果匹配，则学习接收到的 BFD 报文中 Local Discriminator 的值，获取远端标识符。
- BFD 的检测机制：
- BFD 的检测机制是两个系统建立 BFD 会话，并沿它们之间的路径周期性发送 BFD 控制报文，如果一方在既定的时间



内没有收到 BFD 控制报文，则认为路径上发生了故障，BFD 控制报文是 UDP 报文，端口号 3784。

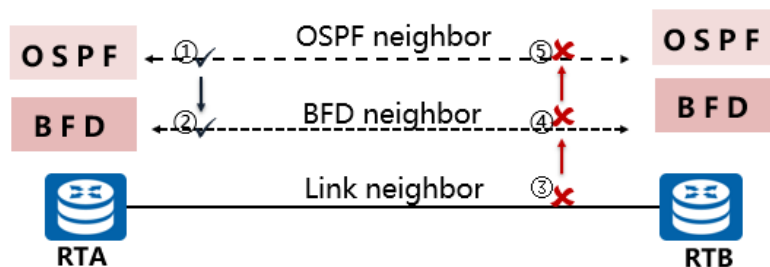
- BFD 提供异步检测模式。在这种模式下，系统之间相互周期性地发送 BFD 控制报文，如果某个系统连续 3 个报文都没有接收到，就认为此 BFD 会话的状态是 Down。

## BFD会话建立过程



- RTA 和 RTB 各自启动 BFD 状态机，初始状态为 Down，发送状态为 Down 的 BFD 报文。对于静态配置 BFD 会话，报文中的 Remote Discriminator 的值是用户指定的；对于动态创建 BFD 会话，Remote Discriminator 的值是 0。
- RTB 收到状态为 Down 的 BFD 报文后，状态切换至 Init，并发送状态为 Init 的 BFD 报文。
- RTB 本地 BFD 状态为 Init 后，不再处理接收到的状态为 Down 的报文。
- RTB 收到状态为 Init 的 BFD 报文后，本地状态切换至 Up。
- RTA 的 BFD 状态变化同 RTB。
- 邻居会话建立成功后，RTA 和 RTB 周期性的向对方发送状态为 Up 的控制报文。

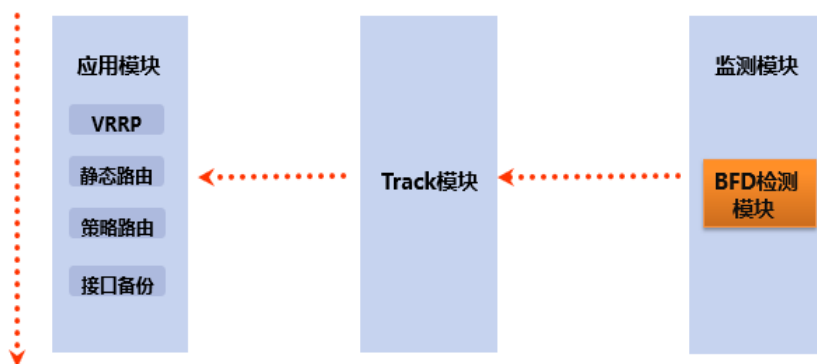
## BFD工作流程



- OSPF邻居建立→BFD会话建立。
- 链路故障→BFD会话Down→OSPF邻居关系中断。
- OSPF 的 BFD 检测故障发现处理流程。
- OSPF 通过自己的 Hello 机制发现邻居并建立连接。
- OSPF 在建立了邻居关系后，将邻居信息（包括目的地址和源地址等）通告给 BFD。
- BFD 根据收到的邻居信息建立会话。
- 被检测链路出现故障。
- BFD 快速发送 BFD 探测报文检测到链路故障，如果在规定时间内无响应，BFD 会话状态变为 Down。
- BFD 通知本地 OSPF 进程 BFD 邻居不可达。
- 本地 OSPF 进程中断 OSPF 邻居关系。

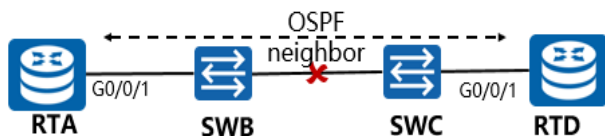
## 联动功能简介

- 联动功能由检测模块、Track和应用模块三部分组成。



- 监测模块负责对链路状态、网络性能等进行监测，并将探测结果通知给 Track 模块。
- Track 模块收到监测模块的探测结果后，及时改变 Track 项的状态，并通知应用模块。
- 应用模块根据 Track 项的状态，进行相应的处理，从而实现联动。

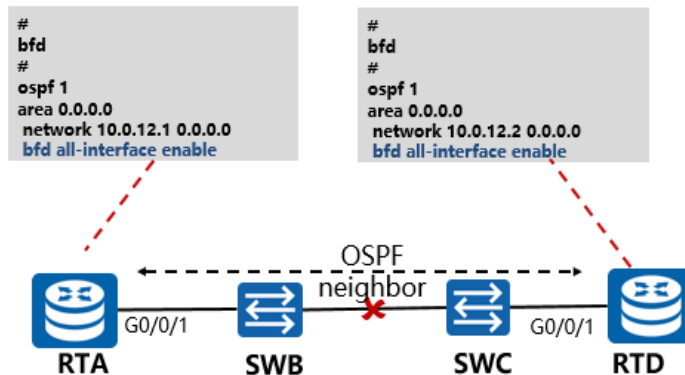
## BFD与OSPF联动配置需求



- 某公司通过两台二层交换机作为中继将两个相距较远的部门连接。RTA、RTD上运行OSPF，建立OSPF邻居关系，保证网络层相互可达。
- 已知RTA、RTB都支持BFD功能，要求使用OSPF与BFD联动技术，采用BFD控制报文方式实现当RTA或RTB与二层交换机之间以及二层交换机之间的链路出现故障（如链路down）时，BFD能够快速感知并通告OSPF协议。



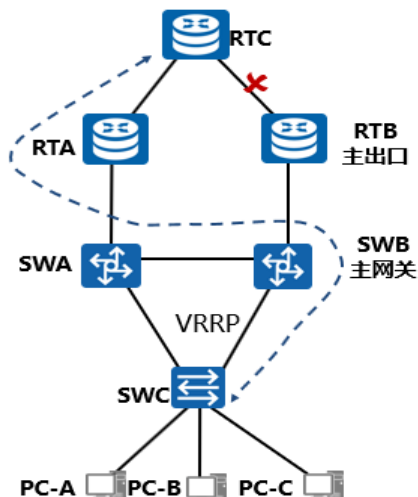
## BFD与OSPF联动配置实现



- 配置思路：
- 配置 RTA 和 RTD 的 OSPF 协议，建立 OSPF 邻居关系。
- 全局下启用 BFD 功能。
- 在 RTA 和 RTB 的 OSPF 进程相应的区域下，启用 BFD 检测功能。



## BFD与VRRP联动配置需求



### 场景需求

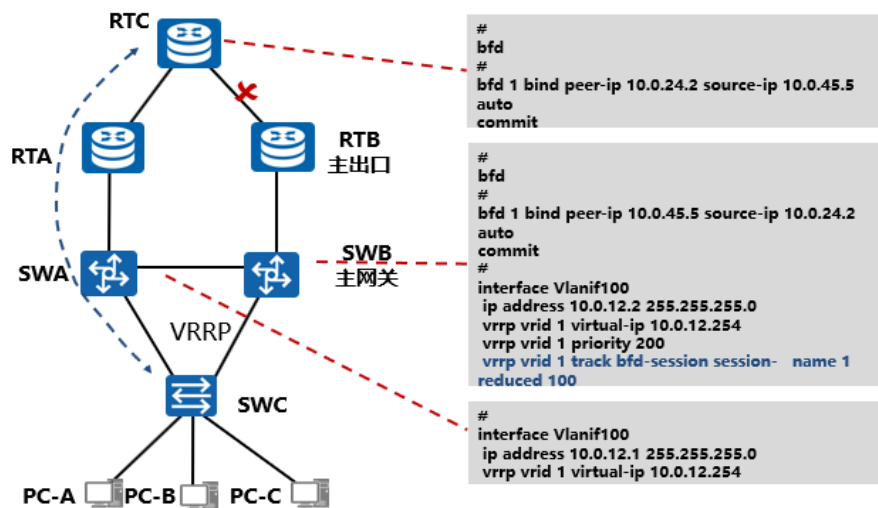
- 如图，SWA和SWB为VRRP备份组，SWB为主用。当RTB与RTC的互联链路出现故障时，SWB能够快速感知并切换为备用网关状态，且SWA成为主用网关。

- 与 VRRP 联动原理：
- VRRP 只能感知 VRRP 备份组之间的故障，而配置 VRR

P 监视上行接口仅能感知 Master 设备上行接口或直连链路的故障，当 Master 设备上行非直连链路故障时，VRRP 无法感知。通过部署 VRRP 与 BFD 联动监视非直连上行链路，当 Master 设备的上行链路发生故障时，BFD 可以快速检测故障并通知 Master 设备调整自身优先级，触发主备切换，确保流量正常转发。



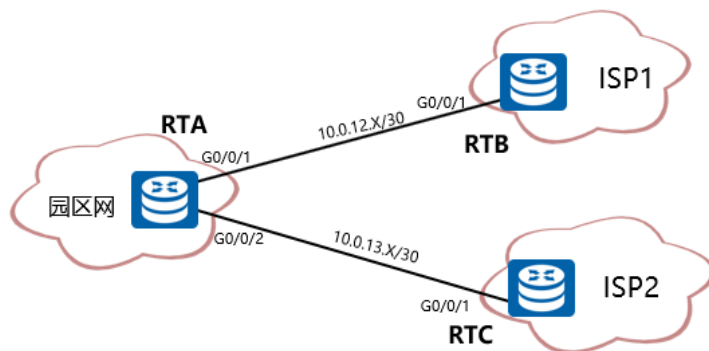
## BFD与VRRP联动配置实现



- 配置思路：
- 配置 SWA 与 SWB 的 VRRP 备份组。
- 配置 SWB 与 RTC 间的 BFD 会话。
- 在 SWB 上的 VRRP 备份组进程下配置监视 BFD 会话，当 BFD 会话状态变化时，通过降低优先级实现主备快速切换。
- SWA 将成为 VRRP 备份组中的 Master 设备继续转发用户流量。



## BFD与静态路由联动配置需求



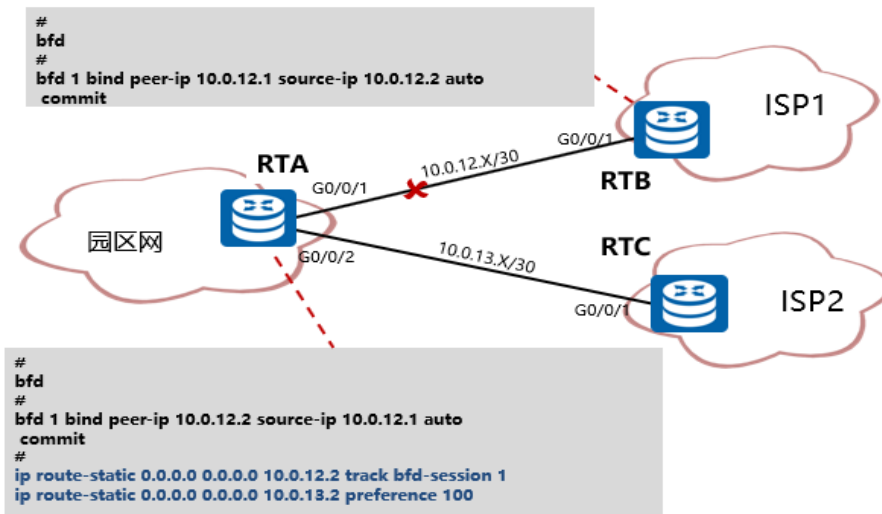
- 场景需求

- 路由器RTA是某园区网的双出口点，分别连通ISP1和ISP2，正常情况下默认路由通往ISP1，ISP2处在备用状态，当通往ISP1的网络出现故障的时候，能够快速切换到ISP2方向。

- 静态路由联动原理：
- 静态路由自身没有检测机制，当网络发生故障的时候，需要管理员介入。BFD与静态路由联动特性可为公网静态路由绑定BFD会话，利用BFD会话来检测静态路由所在链路的状态。
- BFD与静态路由联动可为每条静态路由绑定一个BFD会话，当这条静态路由上绑定的BFD会话检测到链路故障（由Up转为Down）后，BFD会将故障上报路由管理系统，由路由管理模块将这条路由设置为“非激活”状态（此路由不可用，从IP路由表中删除）。当这条静态路由上绑定的BFD会话成功建立或者从故障状态恢复后（由Down转为Up），BFD会上报路由管理模块，由路由管理模块将这条路由设置为“激活”状态（此路由可用，加入IP路由表）。



## BFD与静态路由联动配置实现



- 配置思路：
- 配置 RTA 与 RTB 全局 BFD 功能。
- 配置 RTA 与 RTB 的 BFD 会话。
- 配置 RTA 通往 ISP1 的默认路由，优先级保持为默认的 60，并联动 BFD 会话。
- 配置 RTA 通往 ISP2 的默认路由，优先级比通往 ISP1 的默认路由要低。



## BFD与BGP联动配置需求

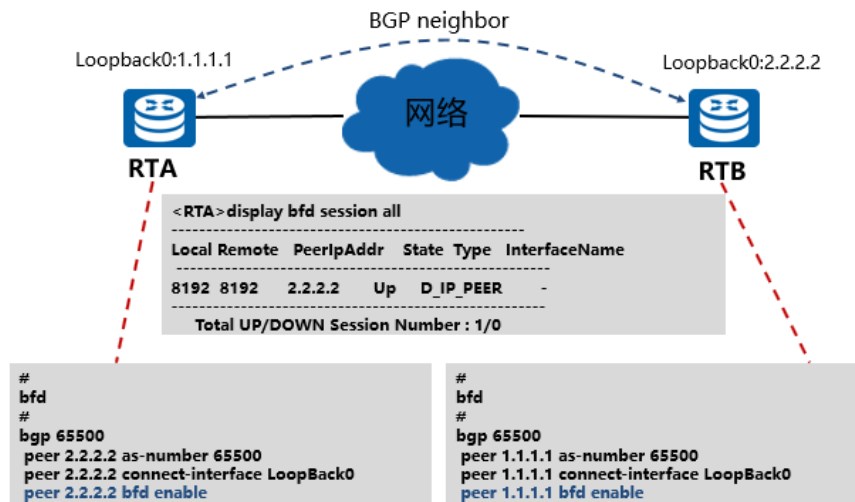


- 场景需求
  - 某局通过两台路由器经过中间网络通道建立BGP的IBGP邻居关系。
  - 已知RTA、RTB都支持BFD功能，要求使用BGP与BFD联动技术，采用BFD控制报文方式实现当RTA或RTB与中间网络设备以及中间网络通道内部链路出现故障时，BFD能够快速感知并通告BGP协议。

- 与 BGP 联动原理：
- BFD 与 BGP 的联动原理和 OSPF 基本一致，唯一区别是最后通知状态消息的对象是 TCP 层的 BGP 模块。



## BFD与BGP联动配置实现

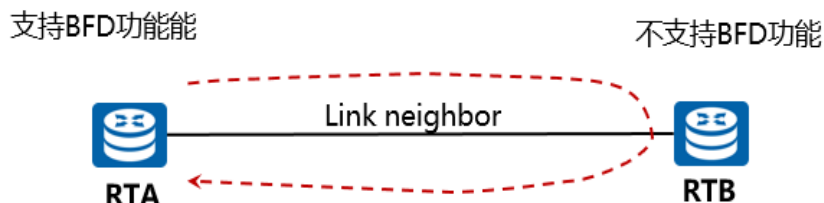


- 配置思路：
- 配置 RTA 和 RTB 网络层路由可达。
- 配置 RTA 和 RTB 的 BGP 邻居参数，建立起 IBGP 的邻居关系。
- 在 RTA 和 RTB 上全局启用 BFD 功能。
- 在 RTA 和 RTB 的 BGP 进程下启用 BFD 检测。



## 单臂回声

- BFD单臂回声功能



- 单臂回声功能：
- 是指通过 BFD 报文的环回操作检测转发链路的连通性。
- 两台直接相连的设备中 RTA 和 RTB，其中一台设备 RTA 支持 BFD 功能，另一台设备 RTB 不支持 BFD 功能，只支持基本的网络层转发。为了能够快速的检测这两台设备之间的故障，可以在支持 BFD 功能的设备 RTA 上创建单臂回声功能的 BFD 会话。RTA 主动发起回声请求功能，不支持 BFD 功能的设备 RTB 接收到该报文后直接将其环回，从而实现转发链路的连通性检测功能。
- 实现的技术原理就是支持 BFD 功能的路由器 RTA 在出接口发送目的地址和源地址都是自己的 BFD 探测报文，不支持 BFD 功能的路由器 RTB 收到探测报文就直接回发给路由器 RTA。

## BFD默认参数及调整方法

参数	缺省值	备注
全局BFD功能	未使能	需使能
发送间隔	1000毫秒	结合实际调整
接收间隔	1000毫秒	结合实际调整
本地检测倍数	3	建议保持默认
等待恢复时间	0分钟	结合实际调整
会话延迟Up时间	0秒钟	结合实际调整
BFD报文优先级	7 (最高级)	建议保持

## 思考题

1. BFD会话建立过程有几个状态?
  2. BFD功能能够与哪些路由协议做联动?
- 1、答案：BFD 会话建立过程状态有：DOWN，INIT，UP。
  - 2、答案：BFD 功能能够与静态，OSPF，BGP 等路由协议联动。