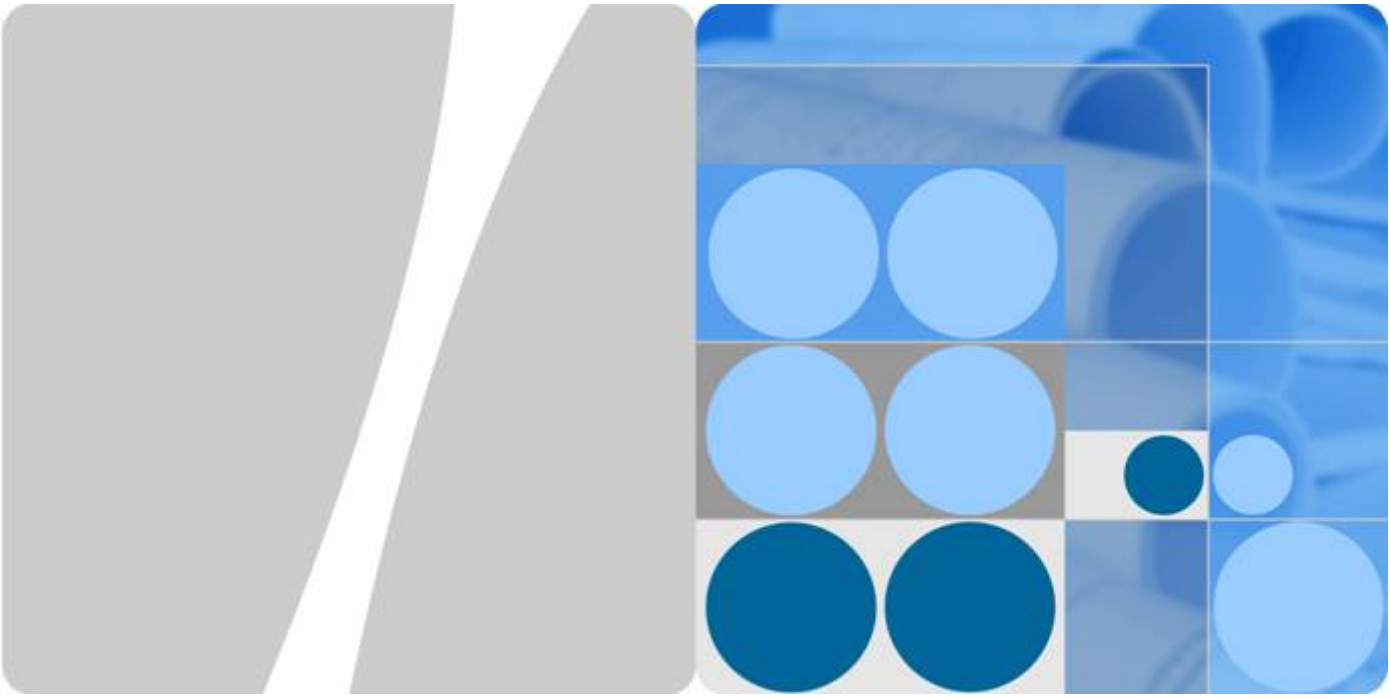


Huawei AR1200 系列企业路由器

V200R002C00

特性描述-可靠性

文档版本　01

发布日期　2011-12-30

华为技术有限公司

商标声明


和其他华为商标均为华为技术有限公司的商标。
本文档提及的其他所有商标或注册商标，由各自的所有人拥有。

注意

您购买的产品、服务或特性等应受华为公司商业合同和条款的约束，本文档中描述的全部或部分产品、服务或特性可能不在您的购买或使用范围之内。除非合同另有约定，华为公司对本文档内容不做任何明示或默示的声明或保证。

由于产品版本升级或其他原因，本文档内容会不定期进行更新。除非另有约定，本文档仅作为使用指导，本文档中的所有陈述、信息和建议不构成任何明示或暗示的担保。

华为技术有限公司

地址： 深圳市龙岗区坂田华为总部办公楼 邮编：518129

网址： http://www.huawei.com

客户服务邮箱： support@huawei.com

客户服务电话： 4008302118

前 言

读者对象

本文档针对可靠性特性，从简介、原理描述和应用三个方面介绍了可靠性特性。

本文档与其它类型手册相结合，便于读者深入掌握特性的实现原理。

本文档主要适用于以下工程师：

- 网络规划工程师
- 调测工程师
- 数据配置工程师
- 系统维护工程师

符号约定

在本文中可能出现下列标志，它们所代表的含义如下。

符号	说明
危险	以本标志开始的文本表示有高度潜在危险，如果不能避免，会导致人员死亡或严重伤害。
警告	以本标志开始的文本表示有中度或低度潜在危险，如果不能避免，可能导致人员轻微或中等伤害。
注意	以本标志开始的文本表示有潜在风险，如果忽视这些文本，可能导致设备损坏、数据丢失、设备性能降低或不可预知的结果。
窍门	以本标志开始的文本能帮助您解决某个问题或节省您的时间。
说明	以本标志开始的文本是正文的附加信息，是对正文的强调和补充。

命令行格式约定

格式	意义
粗体	命令行关键字（命令中保持不变、必须照输的部分）采用**加粗**字体表示。
斜体	命令行参数（命令中必须由实际值进行替代的部分）采用*斜体*表示。
[ ]	表示用“[ ]”括起来的部分在命令配置时是可选的。
{ x | y | ... }	表示从两个或多个选项中选取一个。
[x | y | ...]	表示从两个或多个选项中选取一个或者不选。
{ x | y | ... }*	表示从两个或多个选项中选取多个，最少选取一个，最多选取所有选项。
[x | y | ...]*	表示从两个或多个选项中选取多个或者不选。
&<1-n>	表示符号&前面的参数可以重复 1 ～ n 次。
#	由“#”开始的行表示为注释行。

修订记录

修改记录累积了每次文档更新的说明。最新版本的文档包含以前所有文档版本的更新内容。

文档版本 01 (2011-12-30)

第一次正式发布。

目 录

1 接口备份特性

关于本章

1.1 介绍

定义

接口备份是指同一台设备的各个接口之间形成备份关系，通常一个接口承担业务，其余接口处于备份状态。当某个接口出现故障时，可以将流量快速的切换到备份接口，从而提高了数据设备通信的可靠性。

目的

当路由器上某个接口出现故障时，会造成业务的中断。接口备份技术支持链路故障检测功能，在链路出现故障时能够将流量切换到备份链路，提高了数据设备通信的可靠性。

受益

当主接口本身或其所在线路发生故障而导致业务传输无法正常进行时，启用备份接口进行通讯，从而提高了数据设备通信的可靠性。

1.2 参考标准和协议

无

1.3 可获得性

涉及网元

无需其他网元的配合。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

表 1-1 版本支持

产品	最低支持版本
AR1200	V200R001C00

特性依赖

不依赖其他特性。

硬件要求

对硬件无特殊要求。

其他

无

1.4 原理描述

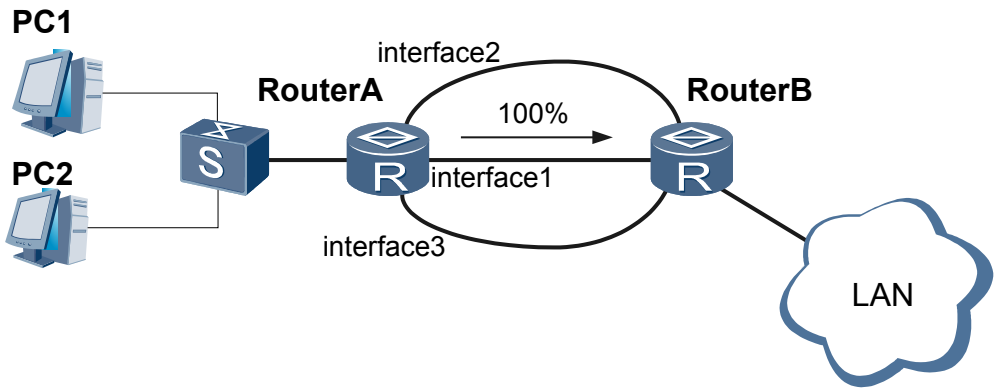
接口备份有五种工作方式：主备备份、接口备份与 NQA 联动、接口备份与 BFD 联动、接口备份与路由联动和负载分担。

1.4.1 主备备份

如图 1-1 所示，interface1 作为主接口，interface2、interface3 作为备份接口。在主备备份方式下，在任意时间只有一个接口进行业务传输。

- 当主接口 interface1 正常工作时，interface2、interface3 处于备份状态，通过主接口 interface1 进行业务传输。
- 路由器跟踪各接口状态，当主接口 interface1 因故障无法进行业务传输时，启动优先级最高的备份接口进行业务传输。
- 当原先故障的主接口恢复正常时，业务传输会重新切换回主接口 interface1。

图 1-1 主备备份示意图



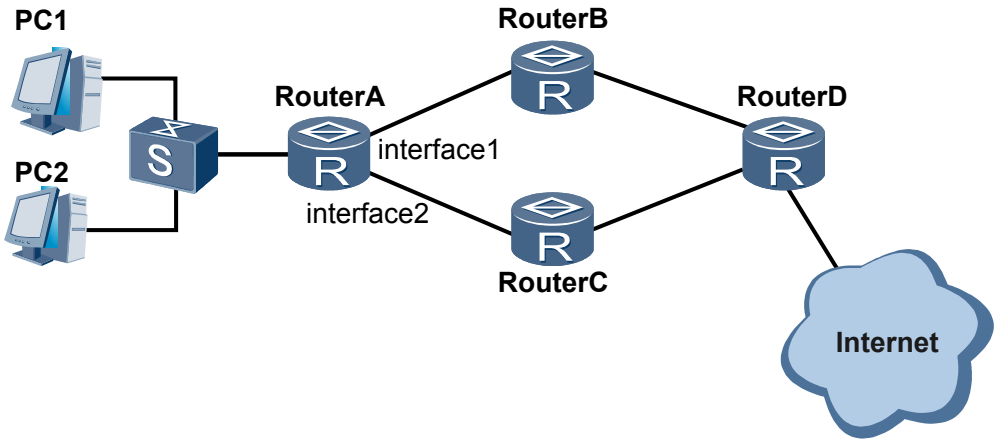
1.4.2 接口备份与 NQA 联动

NQA(Network Quality Analysis)是一种实时的网络性能探测和统计技术。接口备份与 NQA 联动功能可以检测主链路上行链路的故障，实现主链路与备份链路的快速切换。如图 1-2 所示，interface1 作为主接口，RouterA、RouterB 和 RouterD 之间的链路作为主链路，interface2 作为备份接口，RouterA、RouterC 和 RouterD 之间的链路作为备份链路。

在接口备份与 NQA 联动方式下，任意时间只有一条链路进行业务传输:

- 当主链路正常工作时，主链路承担业务传输，备份链路处于备份状态。
- NQA 测试例检测主链路状态，当 NQA 测试例检测到主链路发生故障时，启动备份接口进行业务传输。
- 当原先故障的主链路恢复正常时，业务传输会重新切换回主链路。

图 1-2 接口备份与 NQA 联动示意图



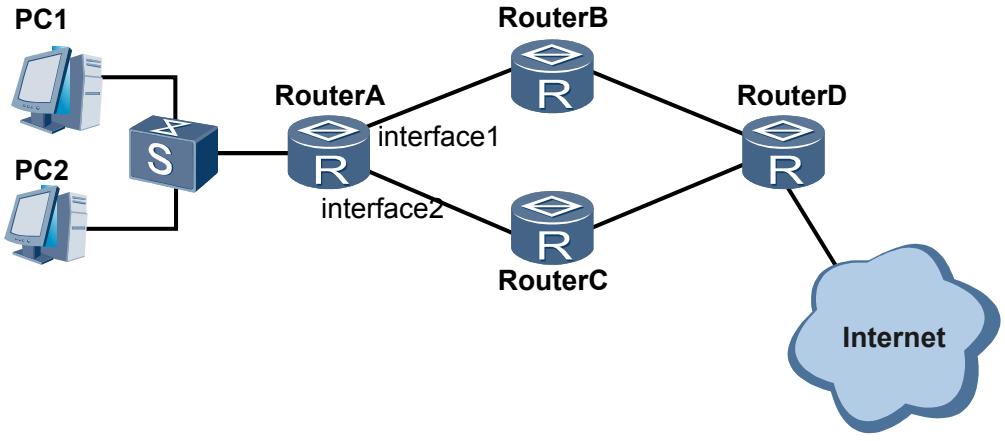
1.4.3 接口备份与 BFD 联动

BFD(Bidirectional Forwarding Detection)是一种快速故障检测机制。接口备份与 BFD 联动功能可以检测主链路上行链路的故障，实现主链路与备份链路的快速切换。如图 1-3 所示，interface1 作为主接口，RouterA、RouterB 和 RouterD 之间的链路作为主链路，interface2 作为备份接口，RouterA、RouterC 和 RouterD 之间的链路作为备份链路。

在接口备份与 BFD 联动方式下，在任意时间只有一条链路进行业务传输:

- 当主链路正常工作时，主链路承担业务传输，备份链路处于备份状态。
- BFD 会话检测主链路状态，当 BFD 会话检测到主链路发生故障时，启动备份接口进行业务传输。
- 当原先故障的主链路恢复正常时，业务传输会重新切换回主链路。

图 1-3 接口备份与 BFD 联动示意图



1.4.4 接口备份与路由联动

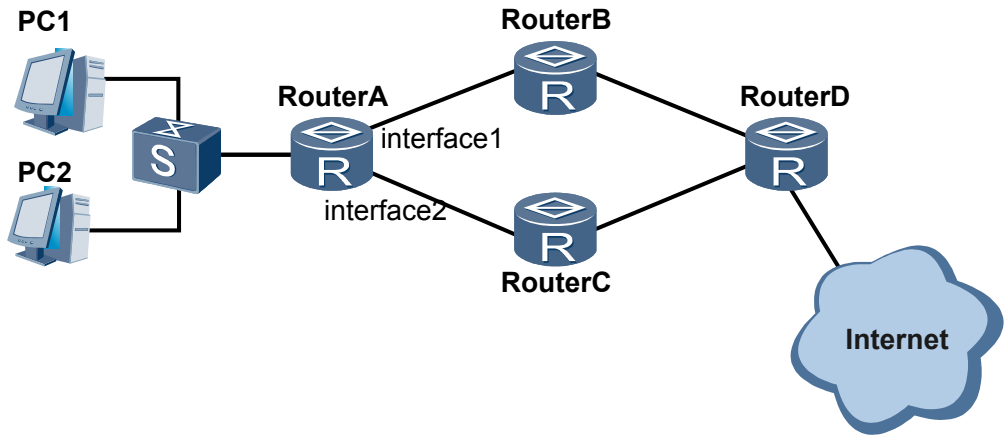
接口备份与路由联动功能可以检测主链路上行链路的故障，实现主链路与备份链路的快速切换。如图 1-4 所示，interface1 作为主接口，RouterA、RouterB 和 RouterD 之间的链

路作为主链路，interface2 作为备份接口，RouterA、RouterC 和 RouterD 之间的链路作为备份链路。

在接口备份与路由联动方式下，在任意时间只有一条链路进行业务传输:

- 当主链路正常工作时，主链路承担业务传输，备份链路处于备份状态。
- 在备份接口上绑定上行链路的目的 IP 地址，当主链路路由撤销或变为非活跃状态时，启用备份接口，实现主备链路的切换。
- 当原先故障的主链路路由恢复正常时，关闭备份接口，业务传输会重新切换回主链路。

图 1-4 接口备份与路由联动示意图

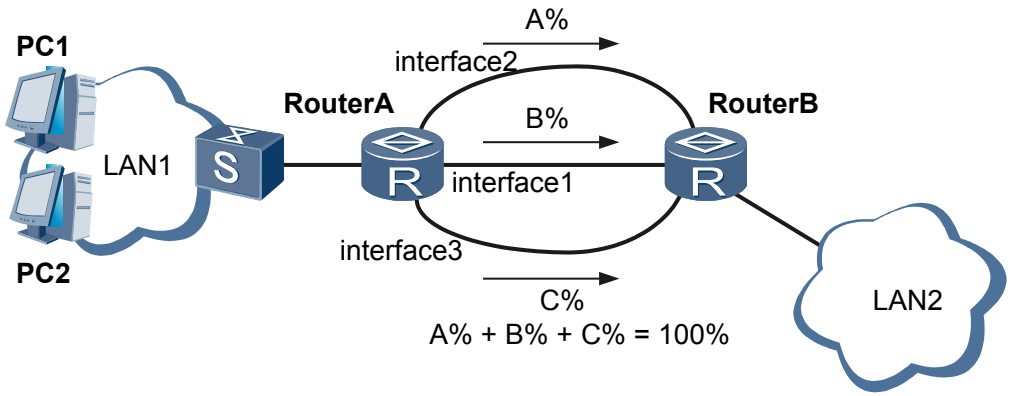


1.4.5 负载分担

如图 1-5 所示，interface1 作为主接口，interface2、interface3 作为备份接口，应用负载分担功能时，分为以下情况：

- 系统定时检测主接口 interface1 流量是否超过设置的门限阈值，当主接口 interface1 的数据流量达到负载分担门限的上限阈值时，优先级最高的可用备份接口将被启用，同主接口 interface1 一起传输业务，进行负载分担。
- 负载分担后流量还是超过上限，优先级次高的另一个可用的备份接口将被启用，在这三个接口间进行负载分担，以此类推，直至流量不超限。
- 负载分担后流量低于设定的下限阈值时，优先级最低的在用备份接口将被关闭。以此类推，直到仅有主接口 interface1 承担业务流量。

图 1-5 负载分担示意图



1.5 应用

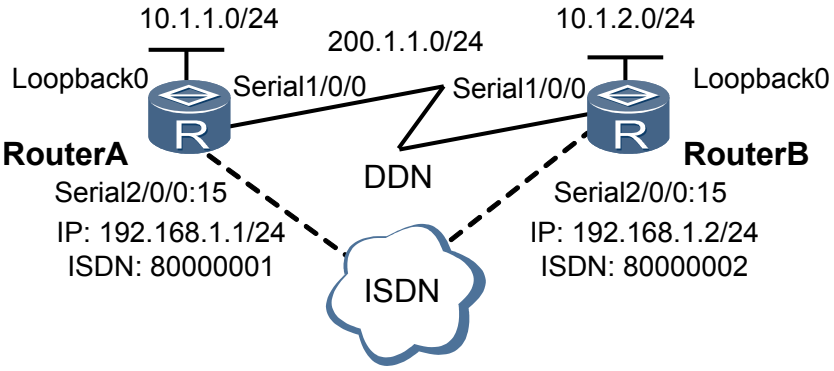
1.5.1 接口备份应用场景

如图 1-6 所示，RouterA 和 RouterB 分别连接 1 条 ISDN PRI 线路，并且通过 DCE 和 DTE 电缆互相连接起来模拟 DDN 链路。在 RouterA 和 RouterB 上分别创建 1 个 Loopback 接口。

在上述组网环境中配置 RouterA 和 RouterB 的主备备份功能，定义 DDN 为主链路，ISDN 链路为备份链路，在主链路断开的情况下，备份链路能够被激活来保证网络端到端的连通性。

备份接口在主接口失效前一直处于备份状态。当主接口出现故障时，备份接口启动，从而保证网络的连通性。

图 1-6 接口备份应用场景



1.6 术语与缩略语

术语

无。

缩略语

无。

2 VRRP

关于本章

2.1 介绍

定义

VRRP（Virtual Router Redundancy Protocol）虚拟路由冗余协议，是一种容错协议。该协议通过把几台路由设备联合组成一台虚拟的路由设备，使用一定的机制保证当主机的下一跳路由器出现故障时，及时将业务切换到备份路由器，从而保持通讯的连续性和可靠性。

以下是与 VRRP 协议相关的基本概念：

- VRRP 路由器（VRRP Router）：运行 VRRP 的设备，它可能属于一个或多个虚拟路由器。
- 虚拟路由器（Virtual Router）：由 VRRP 管理的抽象设备，又称为 VRRP 备份组，被当作一个共享局域网内主机的缺省网关。它包括了一个虚拟路由器标识符和一组虚拟 IP 地址。
- 虚拟 IP 地址(Virtual IP Address)：虚拟路由器的 IP 地址，一个虚拟路由器可以有一个或多个 IP 地址，由用户配置。
- IP 地址拥有者（IP Address Owner）：如果一个 VRRP 路由器将虚拟路由器的 IP 地址作为真实的接口地址，则该设备是 IP 地址拥有者。当这台设备正常工作时，它会响应目的地址是虚拟 IP 地址的报文，如 ping、TCP 连接等。
- 虚拟 MAC 地址：是虚拟路由器根据虚拟路由器 ID 生成的 MAC 地址。一个虚拟路由器拥有一个虚拟 MAC 地址，格式为：00-00-5E-00-01-{VRID}(VRRP)；00-00-5E-00-02-{VRID}(VRRP6)。当虚拟路由器回应 ARP 请求时，使用虚拟 MAC 地址，而不是接口的真实 MAC 地址。
- 主 IP 地址（Primary IP Address）：从接口的真实 IP 地址中选出来的一个主用 IP 地址，通常选择配置的第一个 IP 地址。VRRP 广播报文使用主 IP 地址作为 IP 报文的源地址。
- Master 路由器（Virtual Router Master）：是承担转发报文或者应答 ARP 请求的 VRRP 路由器，转发报文都是发送到虚拟 IP 地址的。如果 IP 地址拥有者是可用的，通常它将成为 Master。
- Backup 路由器（Virtual Router Backup）：一组没有承担转发任务的 VRRP 路由器，当 Master 设备出现故障时，它们将通过竞选成为新的 Master。
- 抢占模式：在抢占模式下，如果 Backup 路由器的优先级比当前 Master 路由器的优先级高，将主动将自己升级成 Master。

目的

随着 Internet 的发展，人们对网络的可靠性的要求越来越高。对于局域网用户来说，能够时刻与外部网络保持联系非常重要。

通常情况下，内部网络中的所有主机都设置一条相同的缺省路由，指向出口网关，实现主机与外部网络的通信。当出口网关发生故障时，主机与外部网络的通信就会中断。

配置多个出口网关是提高系统可靠性的常见方法，但局域网内的主机设备通常不支持动态路由协议，如何在多个出口网关之间进行选路是一个需要解决的问题。

VRRP 协议由因特网工程任务组 IETF（Internet Engineering Task Force）推出，旨在解决局域网主机访问外部网络的可靠性问题，包括如下应用特性：

- 主备备份：这是 VRRP 提供 IP 地址备份功能的基本方式。主备备份方式需要建立一个虚拟路由器，该虚拟路由器包括一个 Master 设备和若干 Backup 设备，这些路由器构成一个备份组。正常情况下，业务全部由 Master 承担，Master 出现故障时，Backup 接替工作。
- VRRP 负载分担：负载分担方式是指多台路由器同时承担业务，单个 VRRP 备份组是不具备负载分担功能的，只有在多台设备上建立两个或更多的备份组，所有备份组均匀分担 Master 状态，此时就每台设备只承担了部分的业务，从而达到负载分担的作用。
- VRRP 监视接口状态：每个 VRRP 备份组可以监视所有与此 VRRP 备份组绑定的接口的状态，从而当接口出现故障时，VRRP 通过改变优先级来重新选择主备关系。
- 虚拟 IP 地址 Ping 功能：提供了控制 Ping 通虚拟 IP 地址的开关命令。
- VRRP 的安全功能：对于安全程度不同的网络环境，可以在报头上设定不同的认证方式和认证字。

2.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RFC2281	Hot Standby Router Protocol（HSRP）	-
RFC2338	Virtual Router Redundancy Protocol（version number One1998）	-
RFC2787	Definitions of Managed Objects for the Virtual Router Redundancy Protocol	-
RFC3768	Virtual Router Redundancy Protocol（version number Two 2004）	-
RFC5798	Virtual Router Redundancy Protocol Version 3 for IPv4 and IPv6	-

2.3 可获得性

涉及网元

需要两台以上路由器，路由器通过同一个二层设备连接起来。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

表 2-1 版本支持

产品	最低支持版本
AR1200	V200R001C00

特性依赖

无其他特殊依赖。

硬件要求

支持主控板和接口板上的三层以太类型接口。

其他

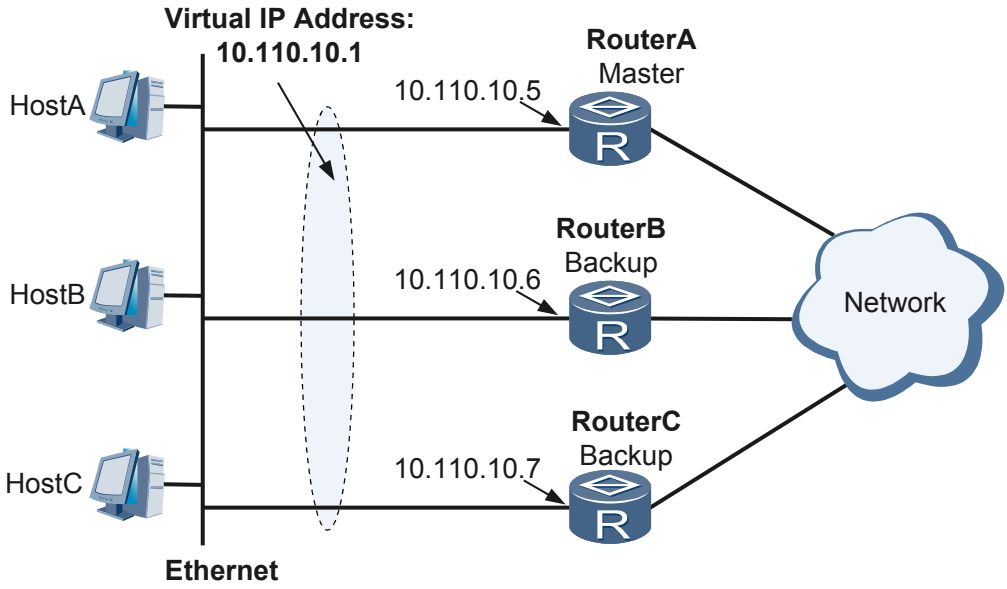
无

2.4 原理描述

VRRP 将局域网的一组路由设备构成一个 VRRP 备份组，相当于一台虚拟路由器。局域网内的主机只需要知道这个虚拟路由器的 IP 地址，并不需知道具体某台设备的 IP 地址，将网络内主机的缺省网关设置为该虚拟路由器的 IP 地址，主机就可以利用该虚拟网关与外部网络进行通信。

VRRP 将该虚拟路由器动态关联到承担传输业务的物理设备上，当该设备出现故障时，再次选择新设备来接替业务传输工作，整个过程对用户完全透明，实现了内部网络和外部网络不间断通信。

图 2-1 虚拟路由器示意图



如图 2-1 所示，虚拟路由器的实现原理如下：

- RouterA、RouterB 和 RouterC 属于同一个 VRRP 备份组，组成一个虚拟的路由器，这个虚拟路由器有自己的 IP 地址 10.110.10.1。虚拟 IP 地址可以直接指定，也可以借用该 VRRP 组所包含的设备上某接口地址。

- RouterA、RouterB 和 RouterC 的实际 IP 地址分别是 10.110.10.5、10.110.10.6 和 10.110.10.7。
- 局域网内的主机只需要将缺省路由设为 10.110.10.1 即可，无需知道具体设备上的接口地址。

主机利用该虚拟网关与外部网络通信。虚拟路由器工作机制如下：

- 根据优先级的大小挑选 Master 设备。Master 设备的选举有两种方法：
  - 比较优先级的大小，优先级高者当选为 Master 设备。
  - 当两台优先级相同的设备，如果已经存在 Master，则 Backup 设备不进行抢占；如果同时竞争 Master，则比较接口 IP 地址大小，IP 地址较大的接口所在设备当选为 Master 设备。
- 其它设备作为备份设备，随时监听 Master 设备的状态。
  - 当主设备正常工作时，它会每隔一段时间（Advertisement_Interval）发送一个 VRRP 组播报文，以通知组内的备份设备，主设备处于正常工作状态。
  - 当组内的备份设备一段时间（Master_Down_Interval）内没有接收到来自主设备的报文，则将自己转为主设备。一个 VRRP 组里有多台备份设备时，短时间内可能产生多个 Master 设备，此时，设备将会将收到的 VRRP 报文中的优先级与本地优先级做比较。从而选取优先级高的设备做 Master。设备的状态变为 Master 之后，会立刻发送免费 ARP 来刷新交换机上的 Mac 表项，从而把用户的流量引到此台设备上来，整个过程对用户完全透明。

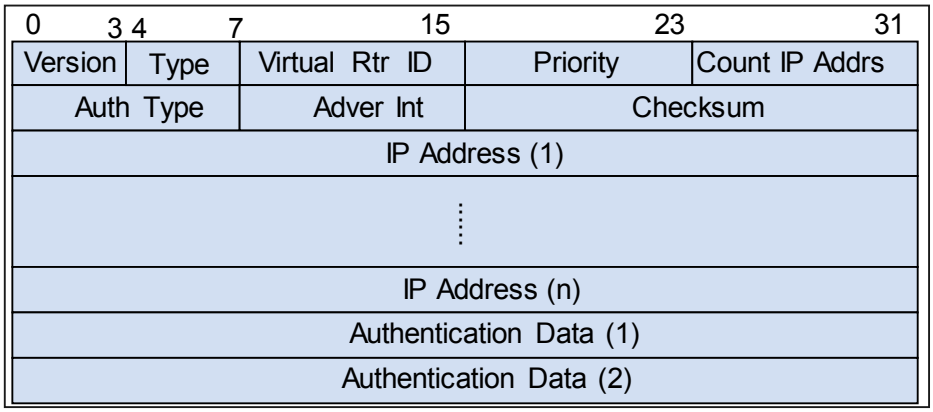
从上述分析可以看到，主机不需要增加额外工作，与外界的通信也不会因某台设备故障而受到影响。

VRRP 报文结构

VRRP 报文用来将 Master 设备的优先级和状态通告给同一虚拟路由器的所有 VRRP 路由器。

VRRP 报文封装在 IP 报文中，发送到分配给 VRRP 的 IP 组播地址。在 IP 报文头中，源地址为发送报文的主接口地址（不是虚拟地址或辅助地址），目的地址是 224.0.0.18，TTL 是 255，协议号是 112。VRRP 报文的结构如图 2-2 所示。

图 2-2 VRRP 报文结构



各字段的含义如下：

- Version：VRRP 协议版本号。此处取值为 2。
- Type：VRRP 通告报文的类型。只有一种取值 1，表示 Advertisement。
- Virtual Rtr ID（VRID）：虚拟路由器 ID，取值范围是 1 ～ 255。
- Priority：发送 VRRP 通告报文的设备在备份组中的优先级。取值范围是 0 ～ 255，但可用的范围是 1 ～ 254。0 表示设备停止参与 VRRP 备份组，用来使备份设备尽快成为 Master 设备，而不必等到计时器超时；255 则保留给 IP 地址拥有者。缺省值是 100。
- Count IP Addrs：VRRP 通告报文中包含的虚拟 IP 地址的个数。
- Authentication Type：VRRP 报文的认证类型。协议中指定了 3 种类型：
  - 0：Non Authentication
  - 1：Simple Text Password
  - 2：IP Authentication Header

  说明

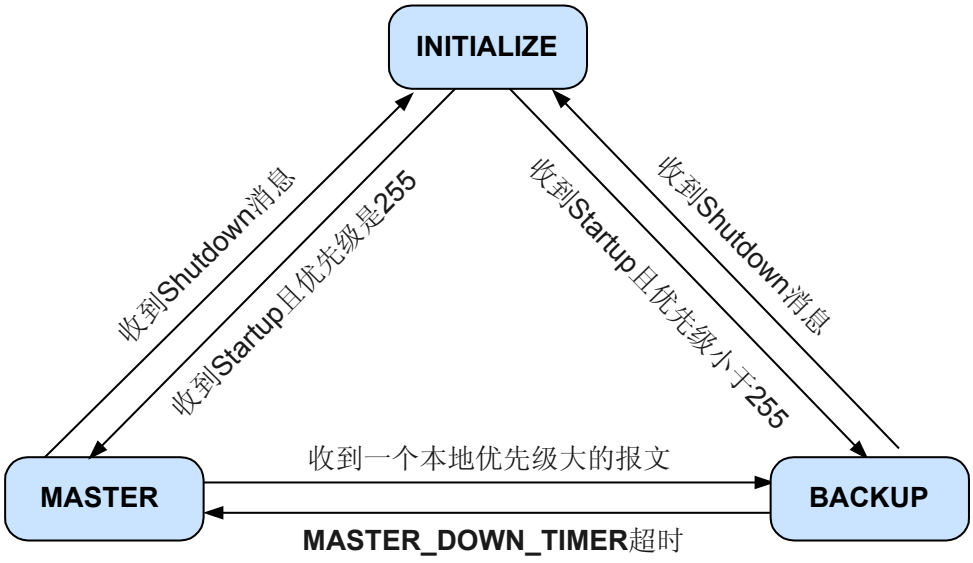
  目前，AR1200 实现了
  - Simple Text Password：即明文认证方式。
  - IP Authentication Header：采用 MD5 认证方式。
- Advertisement Interval：发送通告报文的时间间隔。缺省值为 1 秒。
- Checksum：校验和。
- IP Address(es)：VRRP 备份组的虚拟 IP 地址。
- Authentication Data：认证字。目前只有明文认证和 MD5 认证才用到该部分，对于其它认证方式，一律填 0。

VRRP 的状态机

VRRP 协议中定义了三种状态机：初始状态（Initialize）、活动状态（Master）、备份状态（Backup）。其中，只有处于活动状态的设备才可以转发那些发送到虚拟 IP 地址的报文。

VRRP 状态转换如图 2-3 所示。

图 2-3 VRRP 状态机的转换



Initialize：设备启动时进入此状态，当收到接口 Startup 的消息，将转入 Backup 或 Master 状态（IP 地址拥有者的接口优先级为 255，直接转为 Master）。在此状态时，不会对 VRRP 通告报文做任何处理。

Master：当路由器处于 Master 状态时，它将会做下列工作：

- 定期发送 VRRP 通告报文。
- 以虚拟 MAC 地址响应对虚拟 IP 地址的 ARP 请求。
- 转发目的 MAC 地址为虚拟 MAC 地址的 IP 报文。
- 如果它是这个虚拟 IP 地址的拥有者，则接收目的 IP 地址为这个虚拟 IP 地址的 IP 报文。否则，丢弃这个 IP 报文。
- 如果收到比自己优先级大的报文则转为 Backup 状态。
- 当接收到接口的 Shutdown 事件时，转为 Initialize 状态。

Backup：当路由器处于 Backup 状态时，它将会做下列工作：

- 接收 Master 发送的 VRRP 通告报文，判断 Master 的状态是否正常。
- 对虚拟 IP 地址的 ARP 请求，不做响应。
- 丢弃目的 MAC 地址为虚拟 MAC 地址的 IP 报文。
- 丢弃目的 IP 地址为虚拟 IP 地址的 IP 报文。
- 如果收到比自己优先级小的报文时，丢弃报文，不重置定时器；如果收到优先级和自己相同的报文，则重置定时器，不进一步比较 IP 地址。
- 当接收到 MASTER_DOWN_TIMER 定时器超时的事件时，才会转为 Master 状态。
- 当接收到接口的 Shutdown 事件时，转为 Initialize 状态。

2.4.1 主备备份

这是 VRRP 提供 IP 地址备份功能的基本方式。主备备份方式需要建立一个虚拟路由器，该虚拟路由器包括一个 Master 和若干 Backup 设备。

- 正常情况下，业务全部由 Master 承担。
- Master 出现故障时，Backup 设备接替工作。

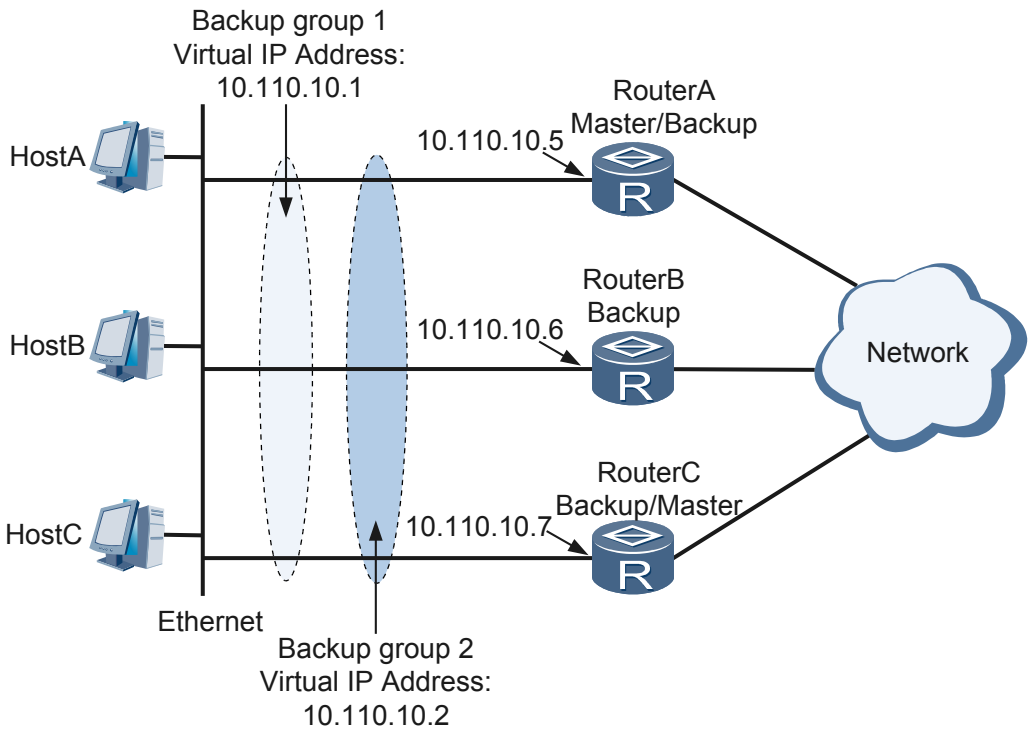
2.4.2 VRRP 负载分担

允许一台设备为多个 VRRP 备份组作备份。通过多个虚拟路由器可以实现负载分担。负载分担方式是指多台虚拟路由器同时承担业务，因此需要建立两个或更多的备份组。

负载分担方式具有以下特点：

- 每个备份组都包括一个 Master 设备和若干 Backup 设备。
- 各备份组的 Master 设备可以不同。
- 同一台设备上的不同接口可以加入多个备份组，在不同备份组中有不同的优先级。

图 2-4 VRRP 负载分担示意图



如图 2-4 所示，配置两个备份组：备份组 1 和备份组 2。

- RouterA 在备份组 1 中作为 Master，在备份组 2 中作为 Backup。
- RouterB 在备份组 1 和 2 中都作为 Backup。
- RouterC 在备份组 2 中作为 Master，在备份组 1 中作为 Backup。
- 一部分主机使用备份组 1 作网关，另一部分主机使用备份组 2 作为网关。

这样，可以达到分担数据流而又相互备份的目的。

2.4.3 VRRP 监视接口状态

VRRP 可以监视接口的状态。当被监视的接口 Down 或 Up 时，该设备的优先级会自动降低或升高一定的数值，使得备份组中各设备优先级高低顺序发生变化，VRRP 设备重新进行 Master 设备竞选。

VRRP 可以通过 Increase 方式和 Reduce 方式来监视接口（一个 VRRP 最多可以监视 8 个接口）。

- 如果 VRRP 以 Increase 方式监视一个接口，当被监视的接口状态变成 Down 后，VRRP 的优先级增加（增加值可以配置）。

  Increase 方式在 VRRP 状态为 Master 或 Backup 时都生效。
- 如果 VRRP 以 Reduce 方式监视一个接口，当被监视的接口状态变为 Down 后，VRRP 的优先级降低（降低值可以配置）。

  Reduce 方式在 VRRP 状态为 Master 或 Backup 时都生效。

2.4.4 虚拟 IP 地址 Ping 功能

由于 VRRP 备份组使用虚拟 IP 地址，不能 Ping 通虚拟 IP 地址，会给监控虚路由器的工作情况带来一定的麻烦，能够 Ping 通虚拟 IP 地址可以比较方便的监控虚拟路由器的工作情况，但是带来可能遭到 ICMP 攻击的隐患。在 AR1200 中，提供了控制 Ping 通虚拟 IP 地址的开关命令，用户可以选择是否打开。

2.4.5 VRRP 安全

对于安全程度不同的网络环境，可以在报头上设定不同的认证方式和认证字。

在安全程度高的网络中，可以采用缺省设置：设备对要发送的 VRRP 报文不进行任何认证处理，收到 VRRP 报文的设备也不进行任何认证，认为收到的都是真实的、合法的 VRRP 报文。这种情况下，不需要设置认证字。

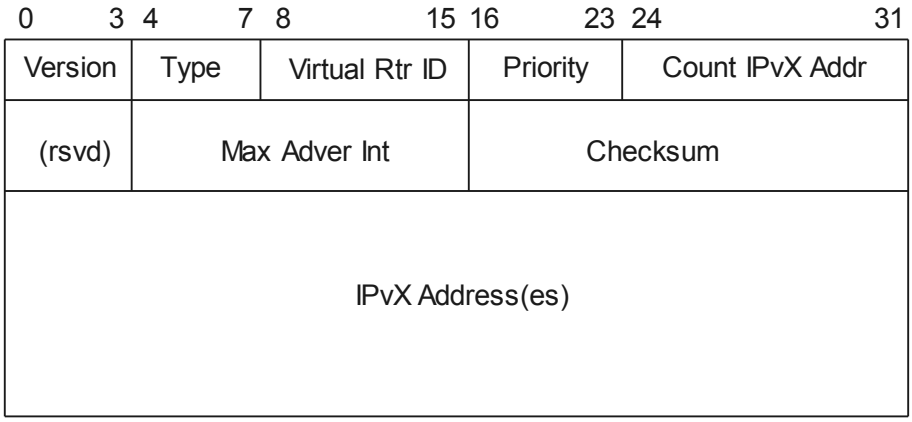
在有可能受到安全威胁的网络中，VRRP 提供了简单字符（Simple）认证方式和 MD5 认证方式。对于简单认证字方式，可以设置长度为 1 ～ 8 的认证字；对于 MD5 认证方式，明文长度范围是 1 ～ 8，密文长度为 24。简单认证方式的安全性小于 MD5 认证方式，但是占用的资源也较小，用户需要根据设备性能和安全性选择认证方式。

2.4.6 VRRPv3 的报文格式

目前，VRRP 协议包括两个版本：VRRPv2 和 VRRPv3。VRRPv2 不支持 IPv6 网路类型，VRRPv3 支持 IPv4 和 IPv6 两种网络。

VRRPv2 是由 RFC3768 提出的，而 VRRPv3 是由 RFC5798 提出的。它们都是用来将 Master 设备的优先级和状态通告给同一备份组的其它设备。VRRPv3 报文的结构如图 2-5 所示。

图 2-5 VRRPv3 报文结构



各字段的含义如下：

- Version：VRRP 协议版本号。VRRPv3 报文只有一种取值：3。
- Type：VRRP 通告报文的类型。只有一种取值：1。
- Virtual Rtr ID：VRRP 备份组的 ID。取值范围是 1 ～ 255。
- Priority：发送 VRRP 通告报文的设备在备份组中的优先级。取值范围是 0 ～ 255，但可用的范围是 1 ～ 255。0 表示设备停止参与 VRRP 备份组，用来使备份设备尽

快成为 Master 设备，而不必等到计时器超时；255 则保留给 IP 地址拥有者。缺省值是 100。

- rsvd：保留字段，必须设置为 0。
- Count IP Addrs：VRRP 通告报文中包含的虚拟 IPv4 或虚拟 IPv6 地址的个数。
- Max Adver Int：发送 VRRP 通告报文的时间间隔。单位是厘秒。
- Checksum：校验和。
- IPvX Address(es)：VRRP 备份组的虚拟 IPv4 地址或者虚拟 IPv6 地址。

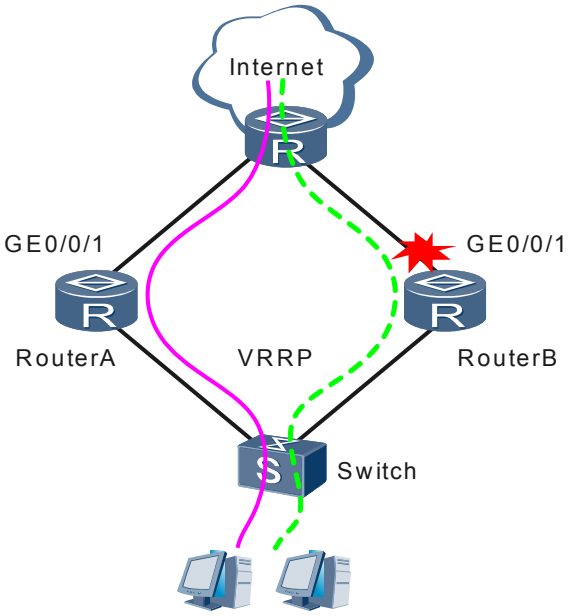
VRRPv2 和 VRRPv3 主要区别：

- 认证功能不同。VRRPv3 不支持认证功能，而 VRRPv2 支持认证功能。
- 发送通告报文的时间间隔的单位不同。VRRPv3 支持的是厘秒级，而 VRRPv2 支持的是秒级。

2.5 应用

2.5.1 VRRP 监视接口状态

图 2-6 VRRP 监视接口的典型组网图



解决的问题：VRRP 无法感知非 VRRP 所在接口状态的变化，当上行链路出现故障时，VRRP 感知不到，从而导致业务中断。

配置说明如下：

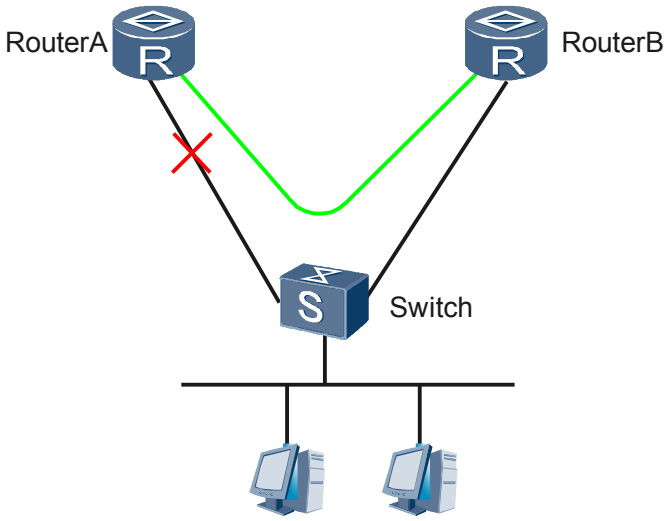
- 通过配置 VRRP 监视指定的接口。
- VRRP 可以以 Increased 方式和 Reduced 方式来监视一个上下链路接口，一个 VRRP 最多可以监视 8 个接口。

- 当 VRRP 监视的接口的状态发生变化时，通知 VRRP，VRRP 根据接口的状态来增加或者是减少 VRRP 的优先级，从而达到指导 VRRP 状态切换的目的。

如图 2-6 所示，RouterA 和 RouterB 两台设备上面运行 VRRP 协议。并且 RouterB 的优先级比 RouterA 的优先级的高，RouterB 以 Reduced 方式监视接口。RouterB 为 Master 设备，用户侧的流量通过主用设备 RouterB 出去，如图 2-6 中虚线所示。现在 RouterB 连向 Internet 的出接口出现故障，由于 RouterB 上面 VRRP 以 Reduced 方式监视了这个接口，VRRP 的优先级降低，RouterA 抢占成为主用设备，以后用户侧的流量则通过 RouterA 出去。

2.5.2 VRRP 快速切换

图 2-7 VRRP 快速切换组网图



解决的问题：VRRP 在检测出链路故障后流量丢失时间较长的问题。

配置说明如下：

- 通过在 RouterA 和 RouterB 上配置 BFD 来检测它们之间链路状态，由于 BFD 可以实现毫秒级别的快速链路检测，所以链路或者远端主机发生故障后可以迅速检测出来。
- 通过监视普通 BFD，VRRP 可以快速感知链路状态，BFD 检测出链路故障后通知 VRRP。
- VRRP 根据 BFD 通告过来的状态来进行优先级调整或者是快速切换来达到快速抢占的目的。
- 一个 VRRP 可以配置监视最多 8 个 BFD 会话。
- 通过监视普通 BFD 可以使 VRRP 主备切换的时间控制在 200ms 以内。

如图 2-7 所示，RouterA 和 RouterB 之间配置 VRRP 备份组建立主备关系，RouterA 为主用设备，用户过来的流量从 RouterA 出去。在 RouterA 和 RouterB 之间建立 BFD 会话，VRRP 备份组监视该 BFD 会话，当 BFD 会话状态变化时，通过修改备份组优先级实现主备快速切换。当 BFD 检测到 RouterA 和 Switch 之间的链路故障时，上报给 VRRP 一个 BFD 检测 Down 事件，RouterB 上 VRRP 备份组的优先级增加，增加后的优先级大于 RouterA 上的 VRRP 备份组的优先级，于是，RouterB 立刻升为 Master，后继的用户流量就会通过 RouterB 转发，从而实现 VRRP 的主备快速切换。

2.6 影响

2.6.1 对系统性能的影响

无。

2.6.2 对其他特性的影响

无。

2.6.3 其他缺陷

在配置 NAT 的接口上配置 VRRP，如果 VRRP 的虚地址是 NAT 地址池中的某一个地址，将导致内网不能访问外网。

2.7 术语与缩略语

缩略语

缩略语	英文全称	中文全称
VRRP	Virtual Router Redundancy Protocol	虚拟路由冗余协议
ARP	Address Resolution Protocol	地址解析协议
ME	Metro Ethernet	城域以太

3 BFD

关于本章

3.1 介绍

定义

双向转发检测 BFD（Bidirectional Forwarding Detection）用于快速检测系统之间的通信故障，并在出现故障时通知上层应用。

目的

为了减小设备故障对业务的影响，提高网络的可用性，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。

现有的故障检测方法主要包括：

- 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。
- 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 就是为解决现有检测机制的不足而产生的。

BFD 的目标如下：

- 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。
- 提供一种单一的机制，能够用来对任何媒介、任何协议层进行实时地检测，并且检测的时间与开销范围比较宽。

3.2 参考标准和协议

本特性的参考资料清单如下：

文档	描述	备注
RFC5880	Bidirectional Forwarding Detection	-
RFC5882	Generic Application of BFD	-
RFC5883	BFD for Multihop Paths	-
RFC5881	BFD for IPv4 and IPv6 (Single Hop)	-

3.3 可获得性

涉及网元

需要其他网元也支持 BFD。

License 支持

无需获得 License 许可，即可获得该特性的服务。

版本支持

表 3-1 版本支持

产品	最低支持版本
AR1200	V200R001C00

特性依赖

无其他特殊依赖。

硬件要求

对硬件无特殊要求。

其他

无

3.4 原理描述

BFD 用于检测转发引擎之间的通信故障。具体来说，BFD 对系统间的、同一路径上的一种数据协议的连通性进行检测，这条路径可以是物理链路或逻辑链路，包括隧道。

可以把 BFD 看作是系统提供的一种服务：

- 上层应用向 BFD 提供检测地址、检测时间等参数。
- BFD 根据这些信息创建、删除或修改 BFD 会话，并把会话状态通告给上层应用。

BFD 具有以下特点：

- 对相邻转发引擎之间的路径提供轻负荷、短持续时间的检测。
- 采用单一机制对所有类型的介质、协议层进行检测，实现全网统一的检测机制。

下面从 BFD 检测机制、检测的链路类型、会话建立方式以及会话管理来介绍 BFD 的基本原理。

BFD 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 控制报文封装在 UDP 报文中传送。会话开始阶段，双方系统通过控制报文中携带的参数（会话标识符、期望的收发报文最小时间间隔、本端 BFD 会话状态等）进行协商。协商成功后，以协商的报文收发时间在彼此之间的路径上定时发送 BFD 控制报文。

为满足快速检测的需求，BFD 草案规定发送间隔和接收间隔单位是微秒。但限于目前的设备处理能力，大部分厂商的设备配置 BFD 时只能达到毫秒级，在进行内部处理时再转换到微秒。AR1200 支持的最小检测时间为 50 毫秒。

BFD 提供异步检测模式：

- 异步模式：BFD 的主要操作模式称为异步模式。在这种模式下，系统之间相互周期性地发送 BFD 控制报文，如果某个系统连续几个报文都没有接收到，就认为此 BFD 会话的状态是 Down。

BFD 检测的链路类型

- IP 链路

  在 AR1200 中，BFD 支持检测的 IP 链路如下，包括单跳检测和多跳检测。

  - 三层物理接口
  - 以太子接口（包括 Eth-Trunk 子接口）

  对于一个物理以太网接口有多个子接口的情况，BFD 会话可以独立建立在各个子接口上和此物理以太网接口上。

- Eth-Trunk
  - 二层 Eth-Trunk 链路
  - 二层 Eth-Trunk 成员链路
  - 三层 Eth-Trunk 链路
  - 三层 Eth-Trunk 成员链路

  检测 Trunk 成员口与检测 Trunk 口的 BFD 会话互相独立，可同时检测。

- VLANIF
  - VLAN 以太成员链路
  - VLAN 以太子接口
  - VLANIF 接口

  检测 VLANIF 与检测 VLAN 成员口的 BFD 会话相互相独立，可同时检测。

BFD 会话建立方式

BFD 会话的建立有两种方式，即静态配置 BFD 会话和动态建立 BFD 会话。

BFD 通过控制报文中的 My Discriminator 和 Your Discriminator 区分不同的会话。静态和动态创建 BFD 会话的主要区别在于 My Discriminator 和 Your Discriminator 的配置方式不同。

- 静态配置 BFD 会话

  静态配置 BFD 会话是指通过命令行手工配置 BFD 会话参数，包括了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

  动态建立 BFD 会话时，系统对本地标识符和远端标识符的处理方式如下：

  - 动态分配本地标识符

    当应用程序触发动态创建 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 Your Discriminator 的值为 0 的 BFD 控制报文，进行会话协商。

  - 自学习远端标识符

    当 BFD 会话的一端收到 Your 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 My Discriminator 的值，获取远端标识符。

BFD 会话管理

BFD 会话有四种状态：Down、Init、Up 和 AdminDown。

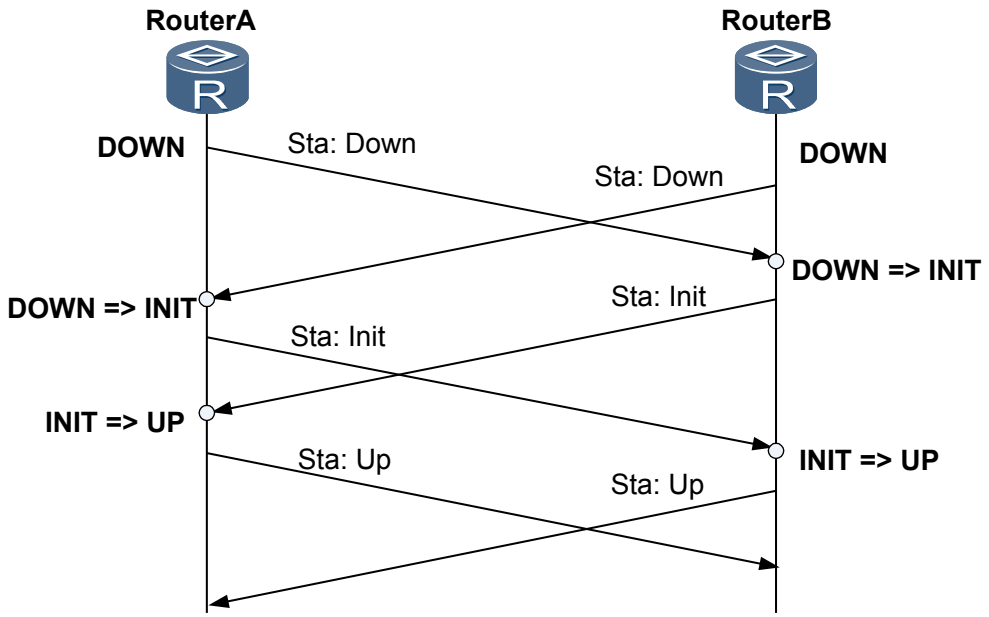
- Down：会话处于 Down 状态或刚刚创建。
- Init：已经能够与对端系统通信，本端希望使会话进入 Up 状态。
- Up：会话已经建立成功。
- AdminDown：会话处于管理性 Down 状态。

会话状态变化通过 BFD 报文的 State 字段传递，系统根据自己本地的会话状态和接收到的对端 BFD 报文驱动状态改变。

BFD 状态机的建立和拆除都采用三次握手机制，以确保两端系统都能知道状态的变化。

以 BFD 会话建立为例，简单介绍状态机的迁移过程。

图 3-1 BFD 会话连接建立



1. RouterA 和 RouterB 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 Your Discriminator 的值是用户指定的；对于动态创建 BFD 会话，Your Discriminator 的值是 0。
2. RouterB 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。
3. RouterB 本地 BFD 状态为 Init 后，不再处理接收到的状态为 Down 的报文。
4. RouterA 的 BFD 状态变化同 RouterB。
5. RouterB 收到状态为 Init 的 BFD 报文后，本地状态切换至 Up。
6. RouterA 的 BFD 状态变化同 RouterB。

3.4.1 BFD for IP

在 IP 链路上建立 BFD 会话，利用 BFD 检测机制快速检测故障。

BFD for IP 支持单跳检测和多跳检测：

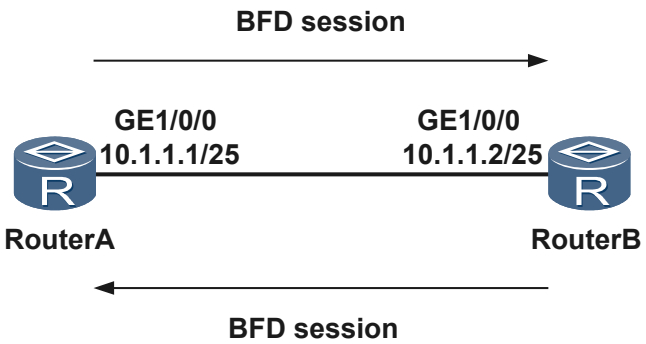
- BFD 单跳检测是指对两个直连系统进行 IP 连通性检测，这里所说的“单跳”是 IP 的一跳。在进行 BFD 单跳检测的两个系统中，对于一种给定的数据协议，在指定接口上只存在一个 BFD 会话。
- BFD 多跳检测是指 BFD 可以检测两个系统间的任意路径，这些路径可能跨越很多跳，也可能在某些部分发生重叠。

组网应用

典型应用一：

如图 3-2 所示，BFD 检测两台设备之间的单跳路径，BFD 会话绑定出接口。

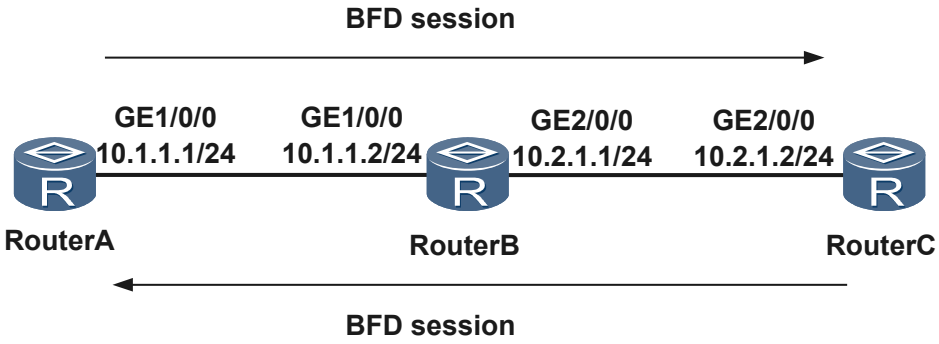
图 3-2 单跳 BFD for IP



典型应用二：

如图 3-3 所示，BFD 检测 RouterA 和 RouterC 之间的多跳路径，BFD 会话绑定对端 IP 但不绑定出接口。

图 3-3 多跳 BFD for IP



3.5 应用

3.5.1 BFD for OSPF

网络上的链路故障或拓扑变化都会导致路由器重新进行路由计算，要提高网络的可用性，缩短路由协议的收敛时间非常重要。由于链路故障无法完全避免，因此，加快故障感知速度并将故障快速通告给路由协议是一种可行的方案。

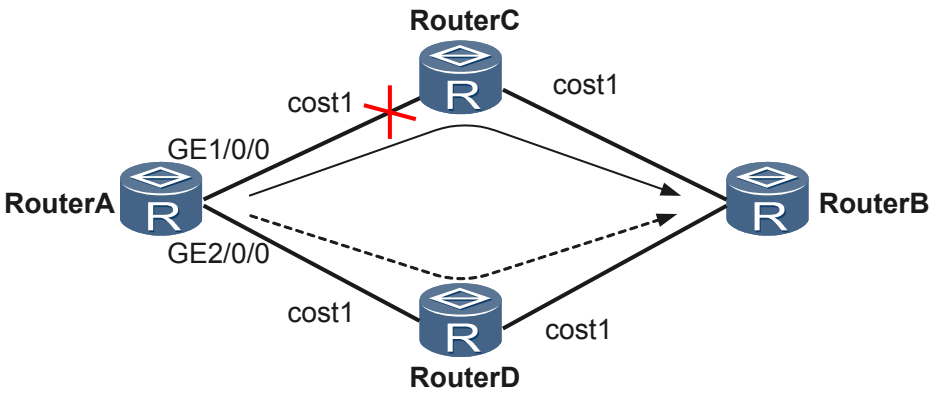
BFD for OSPF 就是将 BFD 和 OSPF 协议关联起来，通过 BFD 对链路故障的快速感应进而通知 OSPF 协议，从而加快 OSPF 协议对于网络拓扑变化的响应。

表 3-2 显示了 OSPF 协议在有、无 BFD 协议下收敛速度的数据。

表 3-2 OSPF 协议收敛速度的数据

有无 BFD	链路故障检测机制	收敛速度
无 BFD	OSPF HELLO keepalive 定时器超时	秒级
有 BFD	BFD 会话 Down	毫秒级

图 3-4 BFD for OSPF 组网图



如图 3-4 所示，RouterA 分别与 RouterC 和 RouterD 建立 OSPF 邻居关系，RouterA 到 RouterB 的路由出接口为 GE1/0/0，经过 RouterC 到达 RouterB。邻居状态到达 FULL 状态时通知 BFD 建立 BFD 会话。

1. 当 RouterA 和 RouterC 之间链路出现故障，BFD 首先感知到并通知 RouterA。
2. RouterA 处理邻居 Down 事件，重新进行路由计算，新的路由出接口为 GE2/0/0，经过 RouterD 到达 RouterB。

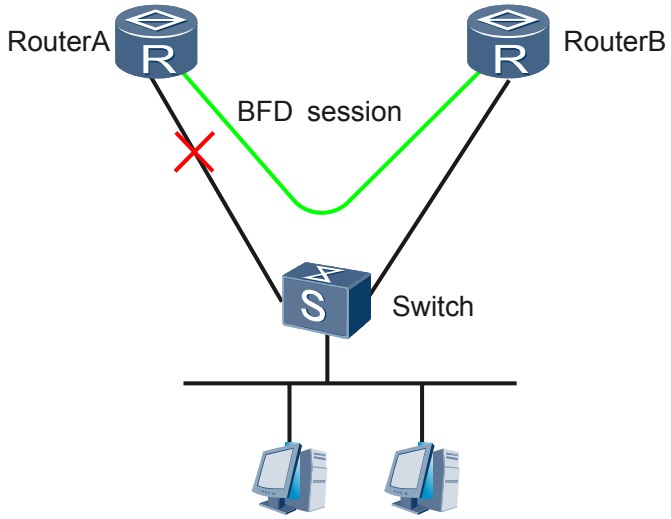
3.5.2 BFD for VRRP

对于以下情况，BFD 都能够将检测到的故障通知主控板，从而加快 VRRP 主备倒换的速度。

- 备份组包含的接口出现故障
- Master 和 Backup 不直接相连
- Master 和 Backup 直接相连，但在中间链路上存在传输设备。

BFD 对 Backup 和 Master 之间的实际地址通信情况进行检测，如果通信不正常，Backup 就认为 Master 已经不可用，升级成 Master；VRRP 通过监视 BFD 会话状态实现主备快速切换。

图 3-5 VRRP Track BFD 典型组网



如图 3-5 所示，RouterA 和 RouterB 之间配置 VRRP 备份组建立主备关系，RouterA 为主用设备，RouterB 为备用设备，用户过来的流量从 RouterA 出去。在 RouterA 和 RouterB 之间建立 BFD 会话，VRRP 备份组监视该 BFD 会话，当 BFD 会话状态变化时，通过修改备份组优先级实现主备快速切换。

当 BFD 检测到 RouterA 和 Switch 之间的链路故障时，上报给 VRRP 一个 BFD 检测 Down 事件，RouterB 上 VRRP 备份组的优先级增加，增加后的优先级大于 RouterA 上的 VRRP 备份组的优先级，于是 RouterB 立刻升为 Master，后继的用户流量就会通过 RouterB 转发，从而实现 VRRP 的主备快速切换。

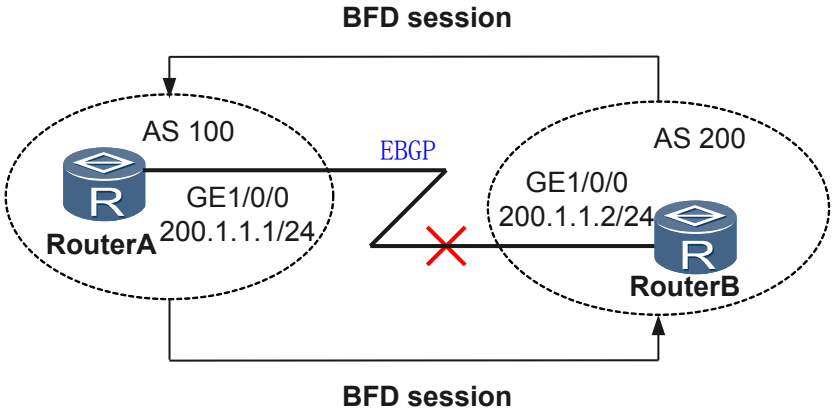
3.5.3 BFD for BGP

BGP 协议通过周期性的向对等体发送 Keepalive 报文来实现邻居检测机制。但这种机制检测到故障所需时间比较长，超过 1 秒钟。当数据达到吉比特速率级别时，将会导致大量的数据丢失，从而无法满足电信级网络高可靠性的需求。

因此，BGP 协议通过引入 BFD for BGP 特性，利用 BFD 的快速检测机制，迅速发现 BGP 对等体间链路的故障，并报告给 BGP 协议，从而实现 BGP 路由的快速收敛。

组网应用

图 3-6 BFD for BGP 组网图



如图 3-6 所示，RouterA 和 RouterB 分别属于 AS100 和 AS200，两台设备相连并建立 EBGP 连接。使用 BFD 检测 RouterA 和 RouterB 之间的 BGP 邻居关系，当 RouterA 和 RouterB 之间的链路发生故障时，BFD 能够快速检测到故障并通告给 BGP 协议。

3.5.4 BFD for TTL

BFD 后续多跳草案（draft-ietf-bfd-multihop-04）规定，对于单跳会话继续使用 3784 作为目的端口号，新增加了关于多跳会话的端口号的使用限制，即多跳会话使用 4784 作为目的端口号。为了与最新的 BFD 多跳草案保持一致，则在 BFD for TTL 特性中增加了对于多跳 BFD 报文端口号的支持，使用 4784 作为多跳 BFD 报文的目的端口号；同时又保证与以前老版本设备互通时，对于以前的老版本的实现并不区分 BFD 单跳、多跳会话，统一使用 3784 作为 BFD 报文端口号，此时需要根据接收报文中携带的 TTL 值区分单跳会话和多跳会话。

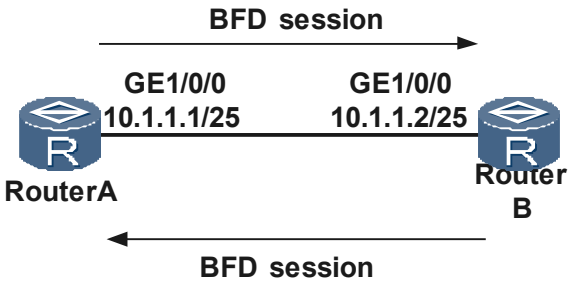
BFD 控制报文封装在 UDP 报文中传送，源端口号的取值范围是 49152 ～ 65535，目的端口号的取值范围是 3784 或 4784。BFD 草案规定，多跳 BFD 报文的目的端口号是 4784。

组网应用

典型应用一：

如图 3-7 所示，BFD 检测两台设备之间的单跳路径，BFD 会话绑定出接口。

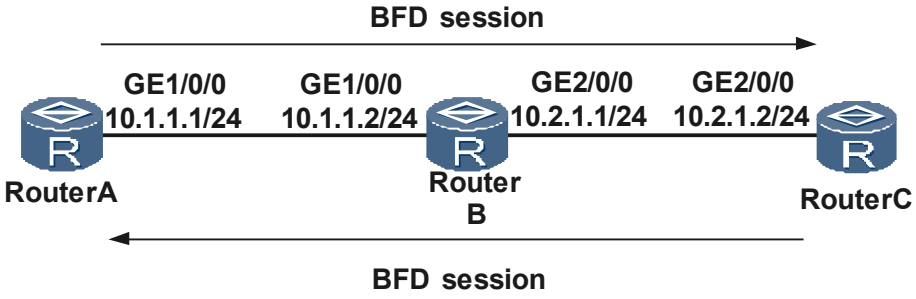
图 3-7 单跳 BFD for IP



典型应用二：

如图 3-8 所示，BFD 检测 RouterA 和 RouterC 之间的多跳路径，BFD 会话绑定对端 IP 但不绑定出接口。

图 3-8 多跳 BFD for IP



3.5.5 单臂 ECHO 功能

ECHO 功能是指通过 BFD 报文的环回操作检测转发链路的连通性。

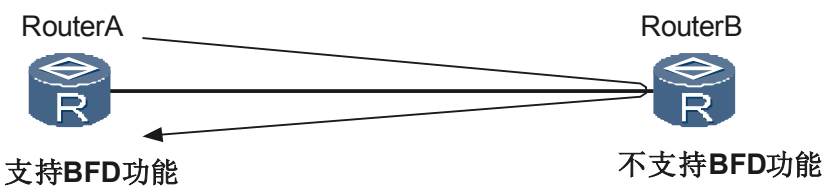
单臂 ECHO 功能：在两台直接相连的设备中，其中，一台设备支持 BFD 功能，另一台设备不支持 BFD 功能，只支持基本的网路层转发。为了能够快速的检测这两台设备之间的故障，可以在支持 BFD 功能的设备上创建单臂 ECHO 功能的 BFD 会话。支持 BFD 功能的设备主动发起 ECHO 请求功能，不支持 BFD 功能的设备接收到该报文后直接将其环回，从而实现转发链路的连通性检测功能。

说明

单臂 ECHO 功能只适用于单跳 BFD 会话中。

组网应用

图 3-9 单臂 ECHO 功能组网示意图



如图 3-9 所示，RouterA 支持 BFD 功能，RouterB 不支持 BFD 功能。在 RouterA 上配置单臂 ECHO 功能的 BFD 会话，检测 RouterA 到 RouterB 之间的单跳路径。RouterB 接收到 RouterA 发送的 BFD 报文后，直接在网络层将该报文环回，从而快速检测 RouterA 和 RouterB 之间的直连链路的连通性。

3.6 术语与缩略语

缩略语(Abbreviations)

缩略语	英文全称	中文全称
ISIS	Intermediate System-Intermediate System	中间系统到中间系统
BFD	Bidirectional Forwarding Detection	双向转发检测
OSPF	Open Shortest Path First	开放式最短路径优先协议
VRRP	Virtual Router Redundancy Protocol	虚拟路由冗余协议