BFD Technology White Paper

Keyword: BFD

Abstract: BFD provides a standard and fast failure detection mechanism. This document discusses the implementation and typical applications of BFD.

Acronyms:

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full spelling</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFD</td>
<td>Bidirectional Forwarding Detection</td>
</tr>
<tr>
<td>UDP</td>
<td>User Datagram Protocol</td>
</tr>
</tbody>
</table>
# Table of Contents

1 Overview ......................................................................................................................................... 3  
  1.1 Background .......................................................................................................................... 3  
  1.2 Benefits .............................................................................................................................. 3  
2 BFD Implementation ....................................................................................................................... 3  
  2.1 Mechanism ........................................................................................................................... 3  
  2.2 BFD Packets ........................................................................................................................ 5  
    2.2.1 BFD Control Packets ................................................................................................. 5  
    2.2.2 BFD Echo Packets .................................................................................................... 7  
  2.3 BFD Session Establishment ............................................................................................... 7  
  2.4 Timer Negotiation ............................................................................................................. 8  
  2.5 Fault Detection ................................................................................................................... 10  
3 Product Characteristics ................................................................................................................ 11  
4 Application Scenarios ................................................................................................................... 12  
  4.1 Configuring BFD for Routing Protocols ............................................................................. 12  
  4.2 Configuring BFD for Fast Reroute ..................................................................................... 13  
  4.3 Configuring BFD for VRRP .............................................................................................. 14  
5 References ................................................................................................................................... 15
1 Overview

1.1 Background

To protect key applications, a network is usually designed with redundant backup links. Devices need to quickly detect communication failures and restore communication through backup links as soon as possible. On some links, such as POS links, devices detect link failures by sending hardware detection signals. However, some other links, such as Ethernet links, provide no hardware detection mechanism. In that case, devices can use the hello mechanism of a protocol for failure detection, which has a failure detection rate of more than one second. Such a rate is too slow for some applications. Some routing protocols, such as OSPF and IS-IS, provide a fast hello mechanism for failure detection, but this mechanism has a failure detection rate of at least one second and is protocol-dependent.

1.2 Benefits

BFD provides a general-purpose, standard, medium- and protocol-independent fast failure detection mechanism. It has the following benefits:

- Detecting failures on any bidirectional forwarding paths, such as direct physical link, virtual link, tunnel, MPLS LSP, multi-hop path, and unidirectional link, between network devices.
- Providing consistent fast fault detection time for upper-layer applications.
- Providing a failure detection time of less than one second for faster network convergence, short application interruptions, and enhanced network reliability.

2 BFD Implementation

2.1 Mechanism

BFD establishes a session between two network devices to detect failures on the bidirectional forwarding paths between the devices and provide services for upper-layer protocols. BFD provides no neighbor discovery mechanism. Protocols that BFD services notify BFD of devices to which it needs to establish sessions. After a session
is established, if no BFD control packet is received from the peer within the negotiated BFD interval, BFD notifies a failure to the protocol, which then takes appropriate measures. The following section describes the operation of BFD for OSPF.

1. BFD session establishment

![Figure 1 BFD session establishment](image1)

1) OSPF discovers neighbors by sending Hello packets and establish neighbor relationships.
2) After establishing neighbor relationships, OSPF notifies BFD of the neighbor information, including destination and source addresses.
3) BFD uses the information to establish BFD sessions.

2. BFD fault detection

![Figure 2 BFD fault detection](image2)

1) A link failure occurs.
2) Upon detection of the link failure, BFD clears the session.
3) BFD notifies the neighbor unreachability to OSPF.
4) OSPF terminates the neighbor relationship on the link.
There are two BFD operating modes: asynchronous and demand. At present, the Comware platform supports only the asynchronous mode. A device operating in the asynchronous mode periodically sends BFD control packets. It tears down the BFD session if it receives no BFD control packet from the peer within the BFD interval.

The Comware platform also supports Echo-mode BFD. A device operating in the echo mode periodically sends BFD echo packets. The peer device returns the received BFD echo packets back without processing them. If the sending device receives no BFD echo packet from the peer within the BFD interval, the session is considered down.

Both ends of a BFD session may be directly (one hop away from each other) or indirectly connected. BFD echo packets can only detect failures for directly connected neighbors. That is, BFD echo packets are sent over a single hop. BFD control packets, however, can detect failures for directly and indirectly connected neighbors. That is, BFD control packets can be sent over one or multiple hops.

2.2 BFD Packets

2.2.1 BFD Control Packets

A BFD control packet consists of the required fields and the optional authentication fields.

2.2.1 Figure 3 shows the format of the required fields:

<table>
<thead>
<tr>
<th>Vers</th>
<th>Diag</th>
<th>Sta</th>
<th>P</th>
<th>F</th>
<th>C</th>
<th>A</th>
<th>D</th>
<th>R</th>
<th>Detect Mult</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- My Discriminator
- Your Discriminator
- Desired Min TX Interval
- Required Min RX Interval
- Required Min Echo RX Interval

Figure 3 BFD control packet format

Figure 4 shows the format of the optional authentication fields.
Figure 4  BFD control packet (authentication fields)

Table 1  Description of the fields of a BFD control packet

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vers</td>
<td>BFD version. The current version is 1.</td>
</tr>
<tr>
<td>Diag</td>
<td>This bit indicates the reason for the last transition of the local protocol</td>
</tr>
<tr>
<td></td>
<td>from up to some other state.</td>
</tr>
<tr>
<td>Sta</td>
<td>Current BFD session state. Its value can be 0 for AdminDown, 1 for Down, 2</td>
</tr>
<tr>
<td></td>
<td>for Init, and 3 for Up.</td>
</tr>
<tr>
<td>P</td>
<td>If it is set to 1, the transmitting system requests the connection</td>
</tr>
<tr>
<td></td>
<td>acknowledgement or acknowledges a parameter change.</td>
</tr>
<tr>
<td>F</td>
<td>If it is set to 1, the transmitting system responds to a received BFD</td>
</tr>
<tr>
<td></td>
<td>control packet that has the Poll (P) bit set.</td>
</tr>
<tr>
<td>C</td>
<td>If set to 1, it means the BFD implementation for the transmitting system</td>
</tr>
<tr>
<td></td>
<td>is independent of its control plane.</td>
</tr>
<tr>
<td>A</td>
<td>If it is set to 1, the control packet contains the authentication field and</td>
</tr>
<tr>
<td></td>
<td>the session is authenticated.</td>
</tr>
<tr>
<td>D</td>
<td>If set to 1, it means the transmitting system wishes to operate in the</td>
</tr>
<tr>
<td></td>
<td>demand mode; if set to 0, it means the transmitting system ignores the</td>
</tr>
<tr>
<td></td>
<td>demand mode or cannot operate in the demand mode.</td>
</tr>
<tr>
<td>R</td>
<td>Reserved. It is set to 0 during transmission and ignored during reception.</td>
</tr>
<tr>
<td>Detect Mult</td>
<td>Detect time multiplier.</td>
</tr>
<tr>
<td>Length</td>
<td>BFD control packet length, in bytes.</td>
</tr>
<tr>
<td>My Discriminator</td>
<td>A unique, nonzero discriminator value generated by the transmitting system,</td>
</tr>
<tr>
<td></td>
<td>used to demultiplex multiple BFD sessions between the same pair of systems.</td>
</tr>
<tr>
<td>Your Discriminator</td>
<td>This field reflects back the received value of My Discriminator or is 0 if</td>
</tr>
<tr>
<td></td>
<td>that value is unknown.</td>
</tr>
<tr>
<td>Desired Min TX</td>
<td>Minimum interval at which the local protocol wishes to send BFD control</td>
</tr>
<tr>
<td>Interval</td>
<td>packets, in milliseconds.</td>
</tr>
<tr>
<td>Required Min RX</td>
<td>Minimum interval at which the local system can receive BFD control</td>
</tr>
<tr>
<td>Interval</td>
<td>packets, in milliseconds.</td>
</tr>
<tr>
<td>Required Min</td>
<td>This is the minimum interval, in microseconds, between received BFD</td>
</tr>
<tr>
<td>Echo RX Interval</td>
<td>Echo packets that this system is capable of supporting. If this value is</td>
</tr>
<tr>
<td></td>
<td>zero, the transmitting system does not support the receipt of BFD Echo</td>
</tr>
<tr>
<td></td>
<td>packets.</td>
</tr>
<tr>
<td>Auth Type</td>
<td>Authentication type used by BFD control packets.</td>
</tr>
<tr>
<td>Auth Len</td>
<td>Authentication field length, including authentication type field and</td>
</tr>
<tr>
<td></td>
<td>authentication length field, in bytes.</td>
</tr>
</tbody>
</table>
BFD control packets are encapsulated in UDP packets, using destination port 3784 and a source port from 49152 to 65535.

### 2.2.2 BFD Echo Packets

BFD echo packets provide a fault detection mechanism without the use of BFD control packets. One end sends BFD echo packets to the peer, which returns received BFD echo packets back without processing them. Therefore, no BFD echo packet format is defined, as long as the transmitting end can distinguish between sessions through packet contents.

BFD echo packets are encapsulated in UDP packets, using destination port 3785, with the IP address of the transmitting interface as the destination IP address and a configured source IP address, which must not cause ICMP redirection.

### 2.3 BFD Session Establishment

---

**Note:**
The following section describes the process of session establishment and fault detection by sending BFD control packets.

---

Before a BFD session is established, there are two BFD operation modes: active or passive:

- **Active mode:** Before a session is established, BFD actively sends BFD control packets regardless of whether any BFD control packet is received from the peer.
- **Passive mode:** Before a session is established, no BFD control packet is sent until a BFD control packet is received from the peer.

During session initialization, at least one end of the two in communication must operate in the active mode for a session to be established. The following example shows the session establishment process with both ends working in the active mode. The session establishment process with one end in active mode and the other in passive mode is the same.
BFD uses a three-way handshake mechanism to establish sessions. The transmitting end fills the Sta field with its current session state in a transmitted BFD control packet. The receiving end changes its session state based on the Sta field value in the received BFD control packet and its own session status to establish a session.

1) As shown in the above figure, upon receipt of a notification from an upper-layer application, Routers A and B send a BFD control packet in DOWN state to the peer.

2) When Router B receives the BFD control packet in DOWN state, the local session state transits from DOWN to INIT. In the BFD control packet sent subsequently, the Sta field is filled with a value of 2, indicating the session state is INIT. Router A experiences the same BFD state transition as Router B.

3) When Router A receives the BFD control packet in INIT state from the peer, the local session state transits from INIT to UP. In the BFD control packet sent subsequently, the Sta field is filled with a value of 3, indicating the session state is UP. Router B experiences the same BFD state transition as Router A.

4) Both BFD peers are UP. A session is established successfully and BFD starts to detect link failures.

### 2.4 Timer Negotiation

Before a BFD session is established, BFD control packets are sent every one second to reduce traffic. After a session is established, BFD control packets are sent at the negotiated interval for fast detection, and the BFD control packet transmit interval and detection timer are negotiated. If a BFD session is valid, these timers can be...
negotiated and modified without affecting session state. The timers for different BFD session directions are negotiated independent of each other and therefore can be different.

The BFD control packet transmit interval is the greater of the Desired Min TX Interval of the local end and the Required Min RX Interval of the peer.

The detection timer is the Detect Mult of the BFD control packets transmitted by the peer times the negotiated BFD control packet transmit interval of the peer.

If the Desired Min TX Interval of the local end increases, the actual BFD control packet transmit interval of the local end cannot be changed until the local end receives a packet with the F bit set from the peer. This ensures that the peer has increased the detection timer before the BFD control packet transmit interval increases on the local end, thus avoiding detection timer timeout errors on the peer.

If the Required Min RX Interval of the local end decreases, the detection timer of the local end cannot be changed until the local end receives a packet with the F bit set from the peer. This ensures that the peer has decreased the BFD control packet transmit interval before the local end decreases the detection timer, thus avoiding detection timer timeout errors on the local end.

If the Desired Min TX Interval decreases, so does the BFD control packet transmit interval on the local end immediately. If the Required Min RX Interval increases, so does the detection timer on the local end immediately.

The following describes the timer negotiation process after a parameter change.

Figure 6  BFD detection timer negotiation
A BFD session is established between Router A and Router B. Both routers have the same Desired Min TX Interval (hereinafter referred to as TX) and Required Min RX Interval (hereinafter referred to as RX) of 100 milliseconds and the same Detect Mult of 3. According to timer negotiation rules, the BFD control packet transmit interval is Router A’s TX or Router B’s RX, whichever is greater, namely, 100 milliseconds, Router B’s transmit interval is also 100 milliseconds, and both routers have a detection timer of 300 milliseconds.

If TX and RX of Router A increase to 150 milliseconds:

1) Router A compares its RX (150 milliseconds) with Router B’s TX (100 milliseconds) and thus changes the detection timer of the local end to 450 milliseconds. Meanwhile, Router A sends a BFD control packet (with a TX and RX of 150 milliseconds) whose P bit is set to the peer.

2) Upon receipt of the packet, Router B replies to Router A with a BFD control packet whose F bit is set (with a TX and RX of 100 milliseconds). Meanwhile, Router B compares the RX in the received packet with the TX of the local end. As TX is greater, Router B’s transmit interval is changed to 150 milliseconds. After comparing the RX of the local end with the TX of the peer, Router B also changes its detection timer to 450 milliseconds.

3) Router A receives a BFD control packet with the F bit set from the peer. After comparing the RX in the received packet and the TX of the local end, Router A calculates the new transmit interval as 150 milliseconds.

4) The timer negotiation is complete. The BFD control packet transmit interval and detection timer for the routers are 150 milliseconds and 450 milliseconds respectively.

2.5 Fault Detection

After BFD session establishment and timer negotiation, both ends start to transmit BFD control packets at the negotiated interval. Each time a BFD control packet is received, the detection timer is reset so that the session remains up. If no BFD control packet is received within the detection timer, the BFD session state transits to DOWN and BFD notifies the failure to the upper-layer application it services. The upper-layer application then takes proper measures. When the BFD session on the local end is DOWN, the Sta field of the BFD control packet transmitted to the peer is
filled with a value of 1 to notify the peer that the session is DOWN. Then, the BFD session state of the peer also transits to DOWN.

3 Product Characteristics

Normally, a router centrally generates and processes BFD packets using the CPU of its main control unit. In case of high CPU utilization on the main control unit due to a large number of BFD sessions or heavy service traffic, BFD packets cannot be generated or sent timely, causing neighbor link detection errors and network oscillations. Therefore, in a centralized architecture, BFD cannot implement fault detection within 30 ms in a full-service environment.

The SR8800 router adopts a distributed OAM architecture, with each service board providing an OAM engine dedicated to implementing link fault detection functions. This saves the main control unit CPU from handling link fault detection services, reduces CPU utilization, enhances CPU security, and improves the link fault detection performance. Thus, link fault detection within 30 ms and carrier-class reliability can be achieved in a full-service environment.

As one of the devices in the industry that provide BFD support for most protocols, the SR8800 supports BFD for OSPF, IS-IS, BGP, RIP, static routing, RSVP, VPLS PW, and VRRP.

The SR8800 is the only core router series in the industry that supports distributed OAM technology.
4 Application Scenarios

4.1 Configuring BFD for Routing Protocols

![Application scenario where BFD works with routing protocols](image)

**Figure 7** Application scenario where BFD works with routing protocols

Router A and Router B are interconnected through a Layer-2 switch. Both routers run a routing protocol.

As Router A and Router B are interconnected through a Layer-2 switch, a link failure between the routers may not cause an interface to be DOWN, and the link failure can be detected only through protocol handshake. After BFD is configured between Router A and Router B, a link failure between the routers can be detected quickly. Upon receipt of the link failure notification from BFD, the routing protocol recalculates routes for fast convergence.
4.2 Configuring BFD for Fast Reroute

Many delay-sensitive services on the Internet, such as audio and video services, require fast route convergence. Configuring BFD for routing protocols or using the fast route convergence technologies can greatly speed up convergence but cannot fully meet the failover requirements of audio and video services.

Configuring BFD for fast reroute can satisfy such requirements. Backup paths are calculated in advance and master path failures are detected quickly. When the master path fails, the traffic is directly switched to a backup path at the forwarding plane rather than the control plane, thus greatly shortening service interruptions.

Figure 8 Application scenario where BFD is configured for fast reroute
4.3 Configuring BFD for VRRP

With the Virtual Router Redundancy Protocol (VRRP) enabled, when the master device fails, the backup device can quickly take over the forwarding task from the master, thus minimizing user data flow interruptions. When the master fails, if the backup receives no packet from the master within the preemption delay timer, the backup becomes the new master, with a switchover time of over one second. After BFD is configured for the backup to monitor the master, failures of the master can be detected more quickly to shorten user data flow interruptions.

VRRP also monitors the status of the master's uplink. When the master works normally but its uplink fails, user packets cannot be forwarded normally. VRRP determines whether an uplink is normal by monitoring the uplink interface status.

**Figure 9** Application scenario where BFD is configured for VRRP
When the monitored interface is down, the master lowers its priority to initiate a switchover. However, this mechanism depends on the protocol status; if the uplink fails but the protocol status of the interface remains up, the failure cannot be detected through VRRP. Configuring BFD for the VRRP master to monitor its uplink can solve this problem.

5 References

- Katz D., Ward D., "BFD for IPv4 and IPv6 (Single Hop)", draft-ietf-bfd-v4v6-1hop-05.