# Reducing Latency and Jitter for Real-Time Traffic Using Multilink PPP

This module contains information about reducing latency and jitter for real-time traffic on your network. One Cisco mechanism for reducing latency and jitter for real-time traffic is Multilink PPP (MLP), also known as Multilink. This module contains conceptual information about Multilink and describes how Multilink PPP can be used with network peers to reduce latency and jitter for real-time traffic on your network.

## Finding Feature Information

Your software release may not support all the features documented in this module. For the latest caveats and feature information, see Bug Search Tool and the release notes for your platform and software release. To find information about the features documented in this module, and to see a list of the releases in which each feature is supported, see the feature information table.

Use Cisco Feature Navigator to find information about platform support and Cisco software image support. To access Cisco Feature Navigator, go to www.cisco.com/go/cfn. An account on Cisco.com is not required.

## Information About Multilink

### Queueing Mechanisms for Multilink

You can use the following queueing mechanisms with Multilink:

- Low latency queueing (LLQ)
- Weighted fair queueing (WFQ)
- Class-based weighted fair queueing (CBWFQ)

# Multilink Functionality

At the top level, Multilink provides packet interleaving, packet fragmentation, and packet resequencing across multiple logical data links. The packet interleaving, packet fragmentation, and packet resequencing is used to accommodate the fast transmission times required for sending real-time packets (for example, voice packets) across the network links. Multilink is especially useful over slow network links (that is, a network link with a link speed less than or equal to 768 kbps).

## Multilink Interleaving

Multilink interleaving is based upon two other integral Multilink activities:

- The ability to fragment packets (or datagrams)

- The ability to multiplex at least two independent data streams

The term interleaving comes from the latter activity, that is, the interleaving of two (or more) independent data streams which are processed independently by the network peer.

Multilink interleaving is a mechanism that allows short, real-time (that is, time-sensitive) packets to be transmitted to a network peer within a certain amount of time (the "delay budget"). To accomplish this task, Multilink interleaving interrupts the transmission of large non-time-sensitive (sometimes referred to as "bulk") datagrams or packets in favor of transmitting the time-sensitive packet. Once the real-time packet is sent, the system resumes sending the bulk packet.

An example may help to illustrate the concept of delay budget. The network starts transmitting a large datagram to a network peer. This large datagram takes 500 milliseconds (ms) to transmit. Three milliseconds later (while the large datagram is still being transmitted), a voice packet arrives in the transmit queue. By the time the large datagram is completely transmitted (497 ms later) the voice packet (which is highly time-sensitive) is subject to unacceptable delay (that is, its delay budget is exceeded).

Multilink interleaving is particularly useful for applications where too much latency (that is, delay) is detrimental to the function of the application, such as Voice over IP (VoIP). However, it is also beneficial for other forms of "interactive" data, such as Telnet packets where the Telnet packets echo the keystrokes entered by the user at a keyboard.

## Multilink Fragmentation

With Multilink fragmentation, the large datagram is fragmented ("chopped") into a number of small packet fragments, Multilink headers are added to the packet fragments, and the packet fragments are transmitted individually to a network peer.

When interleaving is enabled, the packet fragments are small enough so that the time it takes to transmit them does not exceed the time budgeted for transmitting the real-time (time-sensitive) data packet. The real-time data packets are interleaved between the fragments of the large datagram.

Each time Multilink prepares to send another data packet fragment or frame to the receiving network peer, Multilink first checks to see if a real-time (time-sensitive) packet has arrived in the transmit queue. If so, the high-priority packet is sent first before sending the next fragment from the large datagram.

The time delay before the priority packets arrive at the receiving network link is subject to the usual serialization delays at the network link level. That is, any other data already being transmitted has to be finished before the priority packet can be sent. By segmenting long datagrams into small fragments, and checking for newly

arrived priority frames between fragments, the priority frame is delayed only by the time it takes to transmit a previously queued fragment rather than a complete large datagram.

Thus, the maximum size of the fragments dictates the responsiveness for insertion of priority packets into the stream. The fragment size can be tuned by adjusting the fragment delay with the **ppp multilink fragment delay** command.

To ensure correct order of transmission and reassembly (which occurs later), multilink headers are added to the large datagram fragments after the packets are dequeued and ready to be sent.

The figure below is a simplified illustration of how Multilink fragments and interleaves packets.

*Figure 1: Multilink Fragmentation and Interleaving*
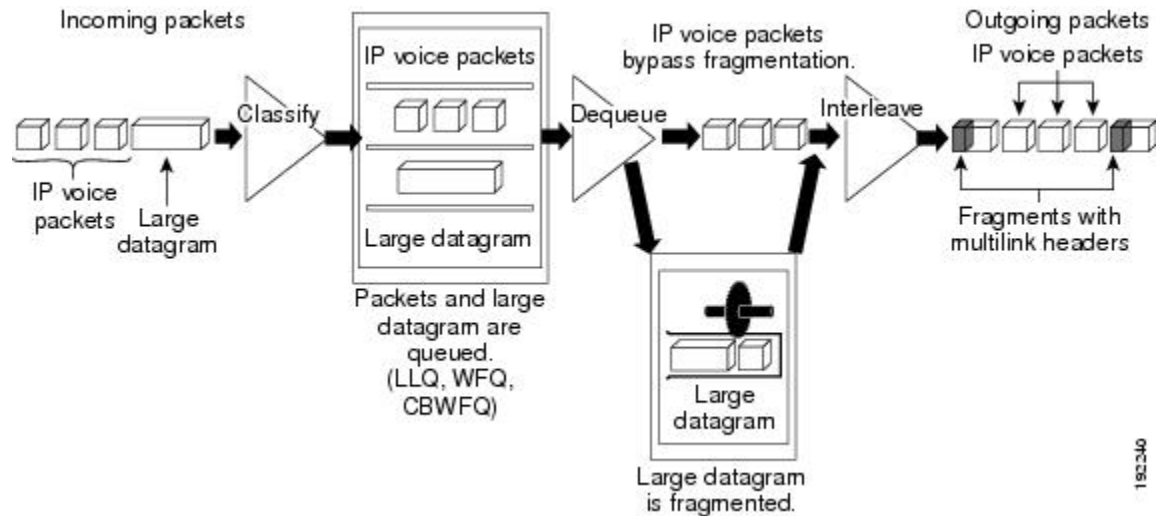


In the figure above, both IP voice packets and a large datagram arrive at the interface from a single network link. Your network may have multiple links. The IP voice packet and large datagram are queued according to their classification. The large datagram is fragmented (the IP voice packets are not). The IP voice packets are interleaved between the fragments of the large datagram, to which multilink headers are added.

### Packets Dequeued and Transmitted

When the large datagram is dequeued, and space becomes available on a member link, Multilink takes a fragment from the original large datagram and transmits the fragments over that link. If an IP voice packet (or other real-time packet) arrives at the transmit queue before Multilink has completely sent the datagram fragment, the next time a link is available to send more packets, Multilink will dequeue and send the high-priority packet. The high-priority packet will be sent instead of another fragment from the large datagram.

## Multilink Resequencing

A multilink bundle is a virtual Point-to-Point Protocol (PPP) connection or session over a network link. A multilink bundle at the transmitting end of the network sends the fragments to a multilink bundle on the receiving end of the network link.

The multilink bundle at the receiving end of the network accepts the fragments from the transmitting multilink bundle.

As fragments are received, the multilink bundle reassembles (resequences) the original large datagram from the fragments using the sequence number in the multilink header attached to the fragment by the sender. The reassembled large datagrams are then forwarded in normal fashion.

## Multilink Bundles and Their Network Links

As mentioned earlier, a multilink bundle is a virtual PPP connection over a network link. The transmitting multilink bundle transmits the packet over a network link to a receiving multilink bundle, where the multilink bundle reassembles the fragments using the sequence number in the multilink header of the fragment.

The individual member links in a multilink bundle are standard serial PPP connections. Most forms of PPP connections may be used as member links in a bundle, including PPP over ATM, PPP over Frame Relay, and PPP over dial interfaces. However, there may be certain limitations and issues associated with using PPP sessions over certain media types, particularly those for "tunneling" protocols such as PPP over ATM, PPP over Frame Relay, and PPP over Ethernet.

# Additional References

### Related Documents

| Related Topic | Document Title |
|---|---|
| QoS commands: complete command syntax, command modes, command history, defaults, usage guidelines, and examples | Cisco IOS Quality of Service Solutions Command Reference |
| LLQ, WFQ, CBWFQ, and other queueing mechanisms | "Applying QoS Features Using the MQC" module |
| Multilink PPP over serial interface links | "Using Multilink PPP over Serial Interface Links" module |

### Standards

| Standard | Title |
|---|---|
| No new or modified standards are supported, and support for existing standards has not been modified. | -- |

### MIBs

| MIB | MIBs Link |
|---|---|
| No new or modified MIBs are supported, and support for existing MIBs has not been modified. | To locate and download MIBs for selected platforms, Cisco IOS XE Software releases, and feature sets, use Cisco MIB Locator found at the following URL: http://www.cisco.com/go/mibs |

### RFCs

| RFC | Title |
|---|---|
| RFC 1990 | *The PPP Multilink Protocol (MP)* |

### Technical Assistance

| Description | Link |
|---|---|
| The Cisco Support and Documentation website provides online resources to download documentation, software, and tools. Use these resources to install and configure the software and to troubleshoot and resolve technical issues with Cisco products and technologies. Access to most tools on the Cisco Support and Documentation website requires a Cisco.com user ID and password. | http://www.cisco.com/cisco/web/support/index.html |