MLPPP Deployment Using the PA-MC-T3-EC and PA-MC-2T3-EC

Overview

Summary

The new enhanced-capability port adapters are targeted to replace the following Cisco® port adapters: 1-port T3 Serial Port Adapter Enhanced (part number PA-T3+), 2-port T3 Serial Port Adapter Enhanced (part number PA-2T3+), 1-port Enhanced Multichannel T3 Port Adapter (part number PA-MC-T3+), 2-port Enhanced Multichannel T3 Port Adapter (part number PA-MC-2T3+), 1-port E3 Port Adapter (part number PA-E3), and 2-port E3 Port Adapter (part number PA-2E3). These new products will ship in two phases.

Phase 1 is targeted for November 2006 and will ship the Cisco 1-Port Multichannel Enhanced Capability Port Adapter (part number PA-MC-T3-EC) and Cisco 2-Port Multichannel Enhanced Capability Port Adapter (part number (PA-MC-2T3-EC). These two port adapters support clear-channel and multichannel functions. The Cisco 7200 Series Network Processing Engines NPE-G1 and NPE-G2 will be the only processors supported in phase 1. One reason for the development of the new port adapters was to overcome performance limitations of the existing multichannel port adapter (PA-MC-2T3+) when sending small packets on both ports simultaneously. The new port adapters do not have the same limitation, and they can achieve line rate with 64-byte packets on both ports.

Phase 2 is targeted for March 2007 and will ship the 1-port clear-channel T3/E3 (PA-T3/E3-EC) and the 2-port clear-channel T3/E3 (PA-2T3/E3-EC). These new port adapters will support clear-channel T3 and E3 functions. Phase 2 will also incorporate support for all versions of the new port adapters on the Cisco uBR7200 Series NPE-400, NPE-G1, and NPE-G2 Network Processing Engines and the Cisco 7301 Router. The multichannel enhanced-capability port adapters (PA-MC-T3-EC and PA-MC-2T3-EC) have the latest version of T3 silicon devices that support offloading certain features to the port adapter to relieve the burden on the CPU of the network processing engine. The features offloaded to the port adapter include: Multilink Point-to-Point Protocol (MLPPP), link fragmentation and interleaving (LFI), Multilink Frame Relay (MLFR), and Frame Relay Fragmentation (FRF.12).

Channelizing a T3 down to T1s and using MLPPP is a very popular solution for certain customers. This paper discusses why a customer would choose to use MLPPP and how the new port adapters with their MLPPP offload capability would benefit a customer.

This document discusses MLPPP and LFI extensively; MLFR and FRF.12 are not discussed because these technologies are roughly the same concepts over different Layer 2 encapsulation (Frame Relay instead of Point-to-Point Protocol [PPP]).

Results

All rates mentioned in this document are unidirectional. To find the bidirectional rate of packets on a link, multiply the unidirectional rate by 2.

Sharing Bandwidth

Link Bundling

There are various mechanisms to bundle links at Layer 2 and make them appear as a single logical link at Layer 3. MLPPP, MLFR, Inverse Multiplexing for ATM (IMA), and Link Aggregation Control Protocol (LACP) are all methods to achieve similar functions on different media.

MLPPP bundles serial links that run the PPP protocol; MLFR bundles serial links running the Frame Relay protocol; IMA bundles ATM links; and LACP (which is similar to Cisco EtherChannel® technology) bundles Ethernet links.

The reasons for bundling links follow:

- Additional bandwidth
- Link redundancy
- IP address conservation
- Layer 2 load sharing

Load-Balancing Algorithms

Layer 3 Cisco Express Forwarding load balancing can be done in either of the following modes:

- Per-session load balancing is done by hashing on some bits within each packet to determine which link to use. Balancing traffic between links depends on the hash algorithm and traffic distribution. In actual deployments, the use of all links will not be equal.
- Per-packet load balancing is done in a round-robin fashion. The number of packets sent on each link will be equal. However, now it is possible to forward out-of-order packets.

Layer 2 load balancing can also be done in a couple of different modes:

- LACP uses an algorithm to hash on certain bits within a packet to determine which member link within a bundle to use, resulting in packets in the same session traversing the same member link. Balancing traffic between member links depends on the hash algorithm and traffic distribution.
- MLPPP, MLFR, and IMA typically use a round-robin algorithm to balance traffic among member links within a bundle, leading to fairly equal use of all links. However, it is now possible to receive (and forward) out-of-order packets. MLPPP compensates for this feature and buffers packets received out of order, and forwards packets only in order. This additional buffering and resequencing effort causes the CPU to do more processing.

Packet Reordering

Figures 1 through 4 illustrate the packet order problem when trying to use multiple links equally to load balance. Initially, the packets arrive on the San Francisco (SFO) router in order. The SFO router forwards these packets to New York City (NYC) by sending every other packet on every other T1 link. Because the links have different propagation delays (and interfaces have different buffering delays), packet 2 arrives in NYC before packet 1. Now the packets are out of order, and the question is: should the NYC router forward them out of order or buffer them and forward them in order?

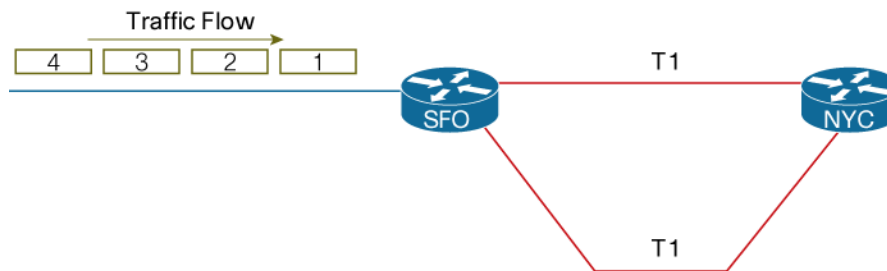**Figure 1.**   Packets Arrive in Sequence

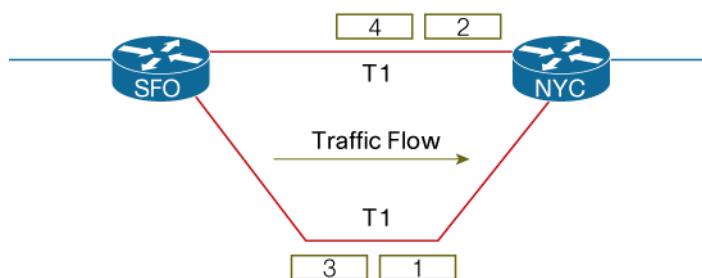Figure 2. Links with Different Propagation Delay Cause Packets Out of Sequence
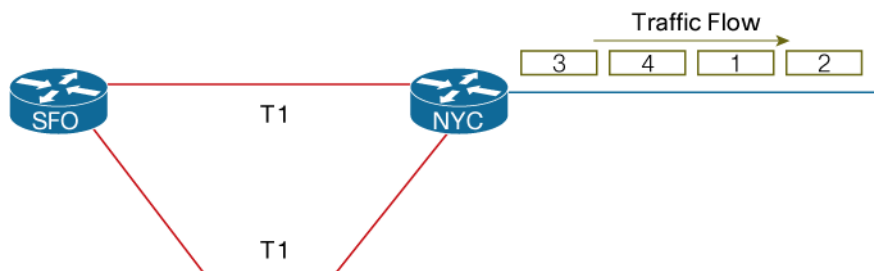
Figure 3. Cisco Express Forwarding per-Packet Load Balancing—Packets Forwarded Out of Sequence

If the routers are doing Cisco Express Forwarding per-packet load balancing, the NYC router will forward the packets in the same order they were received—so they will be forwarded out of sequence. This out-of-sequence forwarding might cause problems, depending on the application. Data applications have large reassembly buffers and out-of-sequence packets do not cause a problem for data. However, voice and video applications have very small reassembly buffers and out-of-sequence packets are a big problem for these applications.

Figure 4. MLPPP Packets Buffered and Forwarded in Sequence
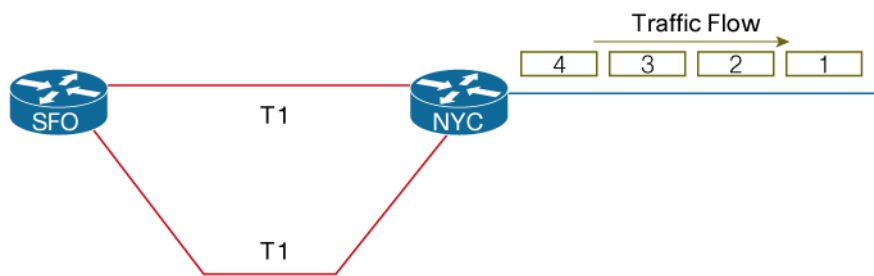
MLPPP solves the packet sequence problem by buffering the packets and forwarding them in sequence, thereby placing more of a burden on the router because it has to buffer the packets that

are received out of sequence. However, packets are only forwarded in sequence and voice applications will operate much better with MLPPP enabled.

Link Fragmentation and Interleaving

LFI is required on links that carry both voice and data. Voice traffic is normally small packets that are sensitive to delay, and data packets are normally large packets that are relatively insensitive to delay.

Without LFI, if a large data packet is currently being transmitted, the small voice packet will have to be queued and transmitted only when the large packet has been fully transmitted. If there are many routers in the path, this delay will accumulate and users will start to complain.

LFI fragments the large packets into small fragments, so the problem with the small packet waiting for the entire large packet to be transmitted is no longer applicable. If a fragment of a large packet is currently being transmitted and if a small packet arrives, it can be transmitted next without having to wait for all the fragments of the large packet to be transmitted. LFI minimizes the serialization delay caused by the large packets.

FRF.12 is a similar concept for implementing LFI on Frame Relay links (Figure 5).

Figure 5.   Link Fragmentation and Interleaving
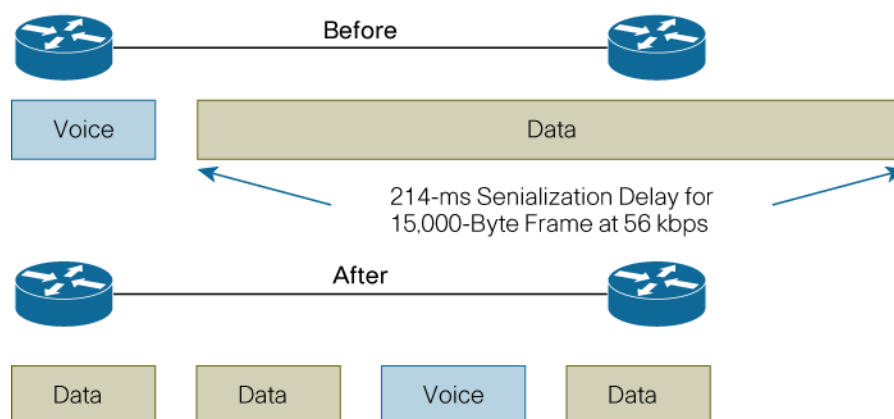
MLPPP

Is MLPPP Needed?

When connecting a pair of routers with multiple serial links, a common question is whether or not MLPPP is needed or if Cisco Express Forwarding load balancing is sufficient. The following points should be considered before using MLPPP:

- Is the bandwidth of member links T1 or less for MLPPP? (Links with greater bandwidth require more buffering and CPU resources, so Cisco recommends MLPPP on T1 or smaller links.)
- Is equal use of all links important?
- Are there delay-sensitive applications that do not do well with out-of-sequence packets (such as voice) on the network?

If the answer to all three questions is "yes", then MLPPP is necessary. The advantages of MLPPP are additional bandwidth, equal use of all links, and delivery of packets in sequence.

MLPPP Deployment Scenario

Figure 6 shows a typical MLPPP deployment scenario. The head office has a T3 interface that has been channelized down to T1s. Some remote offices may be fine with only a single T1, others may require a pair of T1s, and still others may require three or more T1s. MLPPP is a great solution here because as bandwidth is required, another T1 can be added to the MLPPP bundle.

Figure 6.   MLPPP Deployed to Many Remote Sites
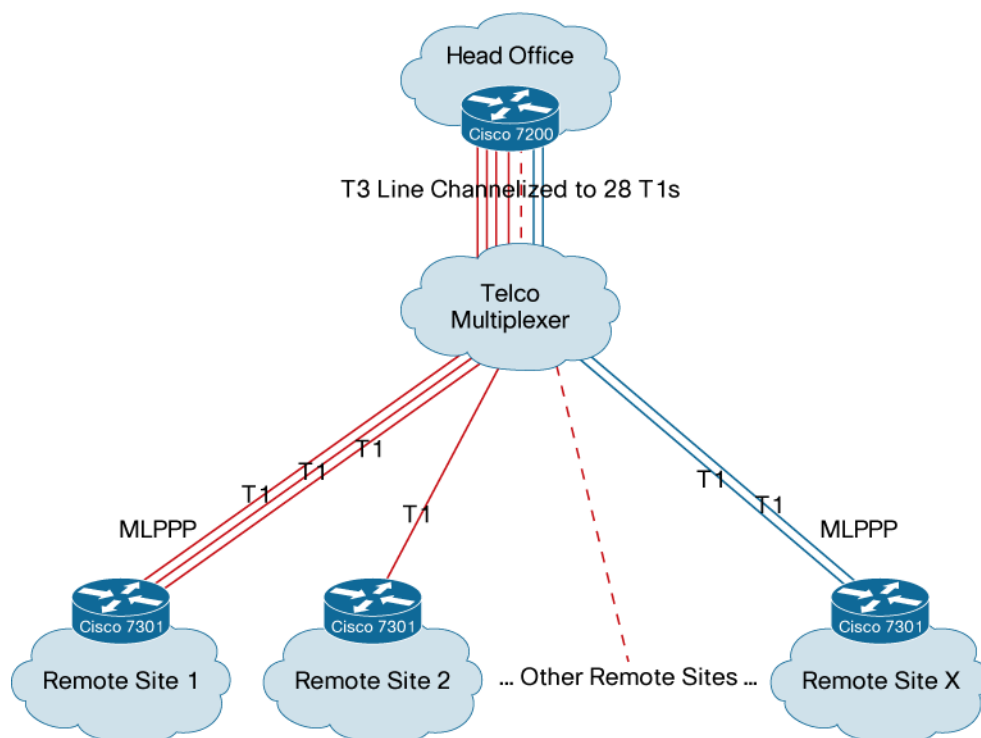
When a remote site exceeds the throughput of a T1, the choices are to upgrade to a T3 or to buy an additional T1 from the telco. Typically, the cost of a T3 is much more than the cost of two or

even three T1s. Therefore, for incremental bandwidth upgrades, MLPPP is an excellent solution that does not require any Layer 3 router reconfiguration.

For simplicity, Figure 6 shows a single T3 connection to the Cisco 7200. However, the new port adapter comes in a dual-port option, so a single Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC) can support 56 T1s shared across a pair of T3s.

Figure 7 shows another scenario where MLPPP is useful: it can be deployed to connect two sites that need more than T1 access. Again, the advantage is that as more bandwidth is needed, additional T1s can be added to the bundle with minimal router configuration changes.

Topologies similar to the one of Figure 7 are also used by cell phone providers to provide access between cells. The only difference is that MLPPP would be running between two adjacent cell sites.

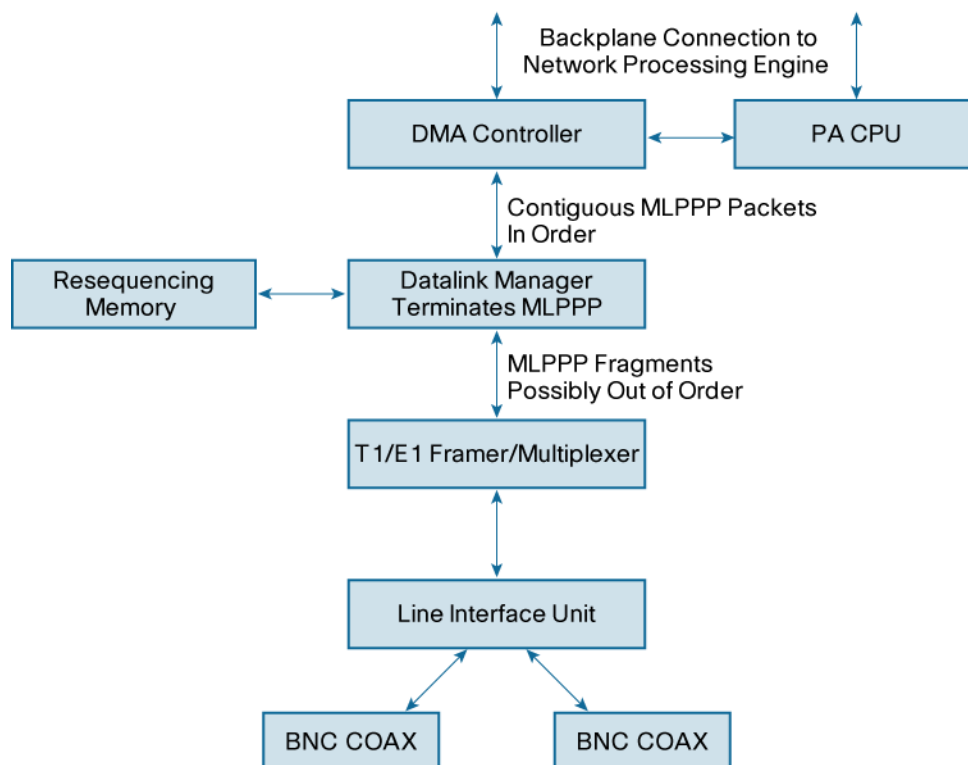**Figure 7.**   MLPPP Deployed Between Two Sites

Cisco 2-Port Multichannel Enhanced Capability Port Adapter Overview

Figure 8 shows the main components of the Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC). The points of interest follow:

- The data link manager that terminates MLPPP on the port adapter and handles fragmentation and interleaving
- The resequencing memory that is used to buffer out-of-order MLPPP and MLFR fragments

These points are the major parts of the port adapter that allow MLPPP, MLFR, LFI, and FRF.12 to be hardware-accelerated on the port adapter and relieve this burden from the main network processing engine, which "sees" only complete MLPPP packets (not fragments) as a result of the datalink manager.

Figure 8.   Port Adapter Block Diagram

```
                    Backplane Connection to
                    Network Processing Engine

        ┌──────────────────┐          ┌──────────────────┐
        │  DMA Controller  │ ◄──────► │     PA CPU       │
        └──────────────────┘          └──────────────────┘

                    Contiguous MLPPP Packets
                    In Order

  ┌──────────────┐        ┌──────────────────────┐
  │ Resequencing │ ◄────► │  Datalink Manager    │
  │   Memory     │        │  Terminates MLPPP    │
  └──────────────┘        └──────────────────────┘

                    MLPPP Fragments
                    Possibly Out of Order

              ┌──────────────────────┐
              │ T 1/E 1 Framer/Multiplexer │
              └──────────────────────┘

              ┌──────────────────────┐
              │  Line Interface Unit │
              └──────────────────────┘

        ┌──────────────┐      ┌──────────────┐
        │  BNC COAX    │      │  BNC COAX    │
        └──────────────┘      └──────────────┘
```

Verification of Hardware-Enabled Features

If upgrading from an older channelized port adapter to the new Cisco 1-Port Multichannel Enhanced Capability Port Adapter (PA-MC-T3-EC) or Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC), no new commands are needed to enable the hardware acceleration. Simply install the port adapter and load an appropriate Cisco IOS® Software image that supports the new port adapter.

The following show command can be used to verify that MLPPP is being processed on the port adapter. Notice the line showing "Multilink is hardware enabled".

```
Router# show ppp multilink interface multilink 14
Multilink14
  Bundle name: 7200_top_bundle_14
  Remote Endpoint Discriminator: [1] 7200_top_bundle_14
  Local Endpoint Discriminator: [1] 7200_bot_bundle_14
  Bundle up for 03:11:30, total bandwidth 6144, load 195/255
  Receive buffer limit 48000 bytes, frag timeout 1000 ms
  0 lost fragments, 0 reordered, 0 unassigned
  0 discarded, 0 lost received
  received sequence unavailable, 0xCC3D sent sequence
  Multilink is hardware enabled
  Member links: 4 active, 0 inactive (max not set, min not set)
    Se2/1/25:0, since 03:11:30
    Se2/1/26:0, since 03:11:30
    Se2/1/27:0, since 03:11:30
    Se2/1/28:0, since 03:11:30
```

Restrictions for Hardware-Enabled Features

In order for MLPPP and MLFR to be offloaded to the port adapter, the following conditions must be met:

- All member links must be T1 or less.

- All member links must have the same bandwidth.

- All member links must terminate on the same port adapter.

- The port adapter must contain fewer than 168 bundles.

- Each bundle must have fewer than 12 T1 links.

- Member links should not be configured with "ppp multilink multiclass" (not applicable for MLFR)

In addition to the above, if LFI is enabled, the following two additional conditions must be met for MLPPP to be offloaded to the PA. If these aren't met, MLPPP will be done on the NPE's CPU:

- Fragmentation size needs to be 128 bytes or greater.

- Only one member link is allowed in the bundle.

Some other points to note regarding hardware offload:

- Fragmentation counters on the transmit side are not available, since these counters aren't supported in hardware.

- Fragmentation sizes supported in hardware are 128, 256 and 512 bytes. If different sizes are configured, the fragment size will be adjusted to the nearest of these values three values.

- TX fragmentation of MFR is not supported in hardware.

Test Results

Figure 9 shows the following:

- The Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC) can reach 99 percent of line rate, and it outperforms the older port adapters.

Figure 9.   Performance of 56 T1s

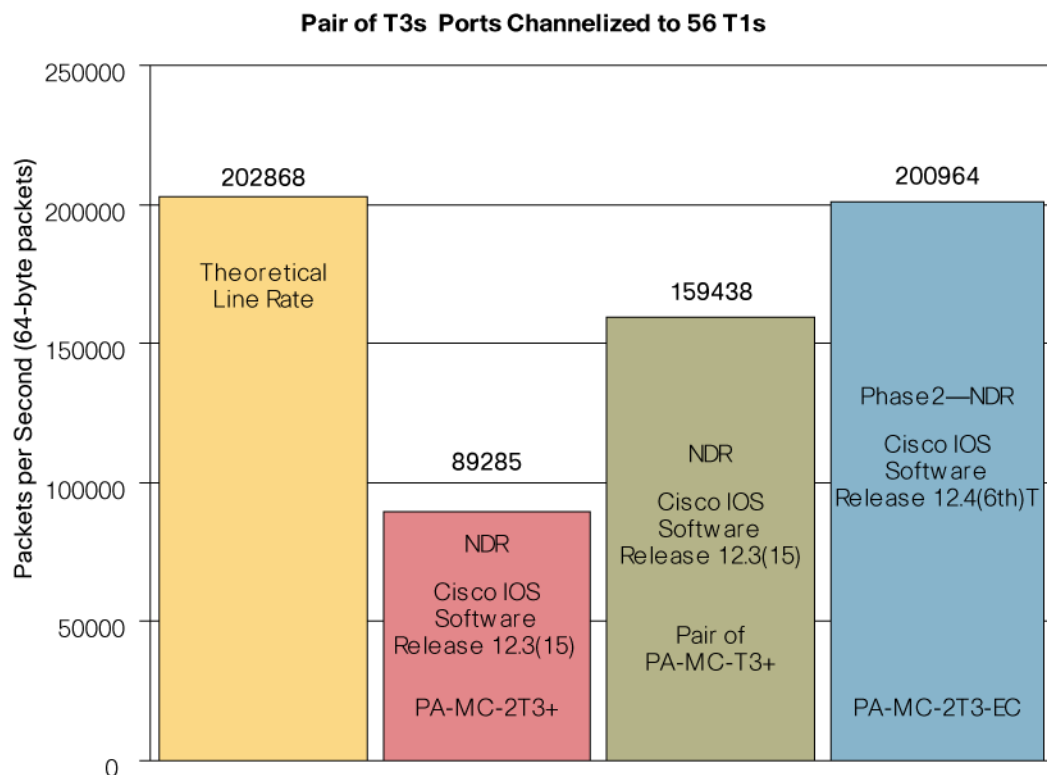Page 8 of 13

Pair of T3s Ports Channelized to 56 T1s

Figure 10 shows the following:

- With MLPPP, the Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC) can reach 98 percent of line rate, and it outperforms the older port adapters.
- Because MLPPP is offloaded to the 2-port port adapter, the CPU has less work to do, so only 74 percent of the CPU is being used while it sustains two 178,571-pps streams. With the older port adapter combinations and less traffic, the CPU was at 100 percent and the console was unavailable.

Figure 10. Performance with MLPPP

Pair of Channelized T3s Ports
4T1s per MLPPP Bundle (14 Bundles)
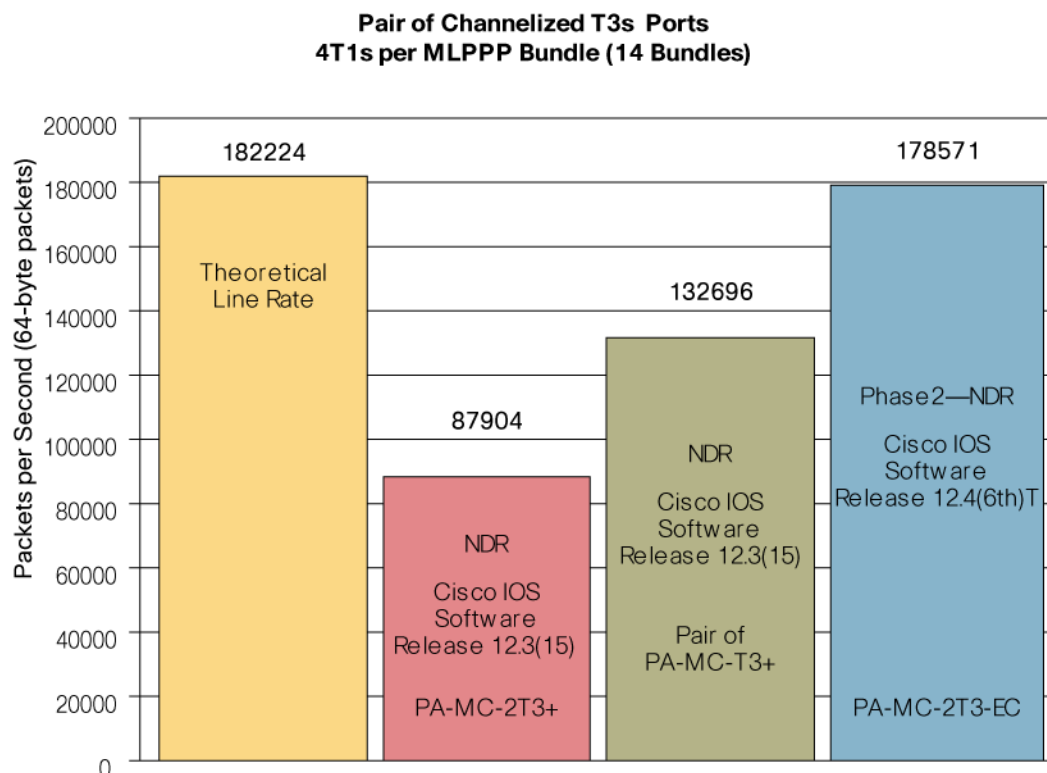
Figure 11 compares phase 1 (which does MLPPP on the main processor) with phase 2 (which offloads MLPPP processing to the port adapter):

- Phase 1 has a non-delivery receipt (NDR) of 137,061 pps and the CPU is 98-percent used. When MLPPP is offloaded to the port adapter (with phase 2), the CPU use goes down to 58 percent for the same traffic load. So for the same traffic load, offloading MLPPP causes a 40-percent reduction in CPU use.

- With phase 2, the NDR value is 178,571 (which is 98 percent of line rate) and the CPU use is only 74 percent.

- The middle bar shows that phase 2 uses 40 percent less CPU time for the same traffic load. Comparing the middle bar with the rightmost bar shows that phase 2 also provides more throughput.

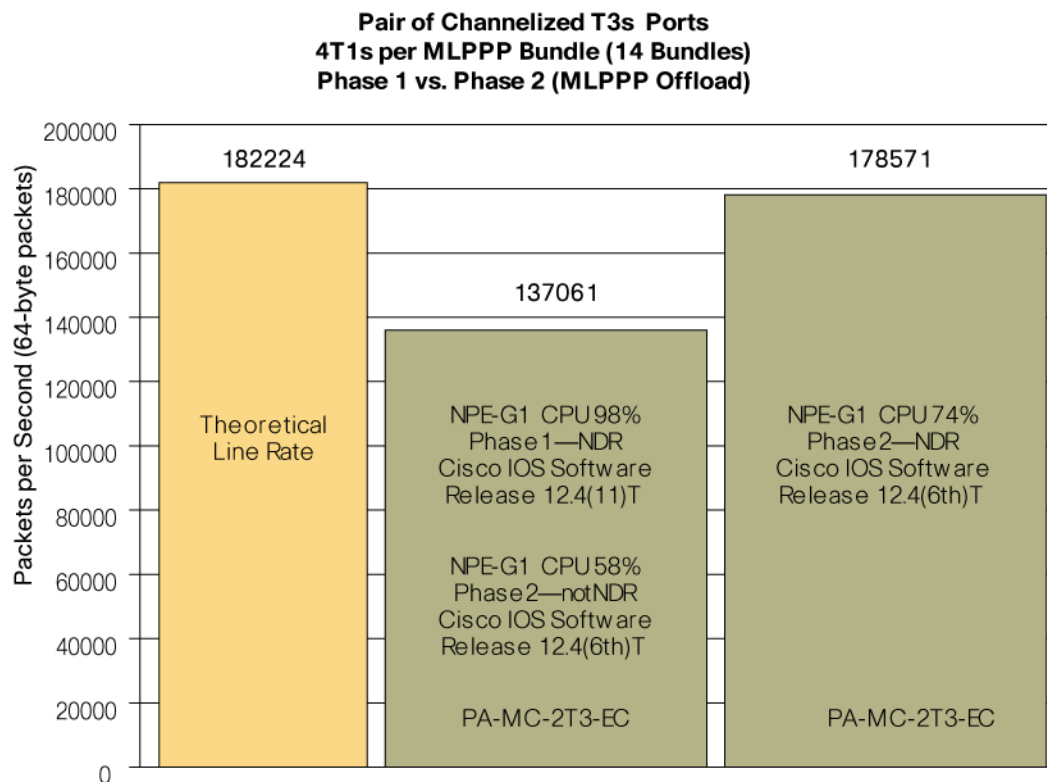**Figure 11.**   MLPPP on Network Processing Engine vs. MLPPP on Port Adapter

Pair of Channelized T3s Ports
4T1s per MLPPP Bundle (14 Bundles)
Phase 1 vs. Phase 2 (MLPPP Offload)

```
Packets per Second (64-byte packets)

200000 ┬

          182224                                          178571
180000 ┤  ┌──────────┐                              ┌──────────────┐
160000 ┤  │          │                              │              │
                                                    │              │
140000 ┤  │          │        137061                │              │
                          ┌──────────────┐          │              │
120000 ┤  │          │    │              │          │              │
          │ Theoretical│   │ NPE-G1 CPU98%│          │ NPE-G1 CPU74%│
100000 ┤  │ Line Rate │    │ Phase1—NDR   │          │ Phase2—NDR   │
          │          │    │ Cisco IOS Software│      │ Cisco IOS Software│
 80000 ┤  │          │    │ Release 12.4(11)T│       │ Release 12.4(6th)T│
          │          │    │              │          │              │
 60000 ┤  │          │    │ NPE-G1 CPU58%│          │              │
          │          │    │ Phase2—notNDR│          │              │
 40000 ┤  │          │    │ Cisco IOS Software│      │              │
          │          │    │ Release 12.4(6th)T│      │              │
 20000 ┤  │          │    │ PA-MC-2T3-EC │          │ PA-MC-2T3-EC │
          │          │    │              │          │              │
     0 ┴  └──────────┘    └──────────────┘          └──────────────┘
```

Figure 12 shows the following:

- In clear-channel mode with small packets, the Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC) has more than triple the throughput of the Cisco 2-Port Multichannel Port Adapter (PA-MC-2T3+).

- In clear-channel mode, the 2-port enhanced-capability port adapter can reach 97 percent of line rate.

- The Cisco 2-Port T3 Port Adapter [[CORRECT?]] (PA-2T3+) is also able to reach 97 percent of line rate. Because it cannot operate in channelized mode, however, it is not as flexible as the new port adapters.

Figure 12. Clear-Channel T3 Performance

Pair of Clear Channel T3 Ports
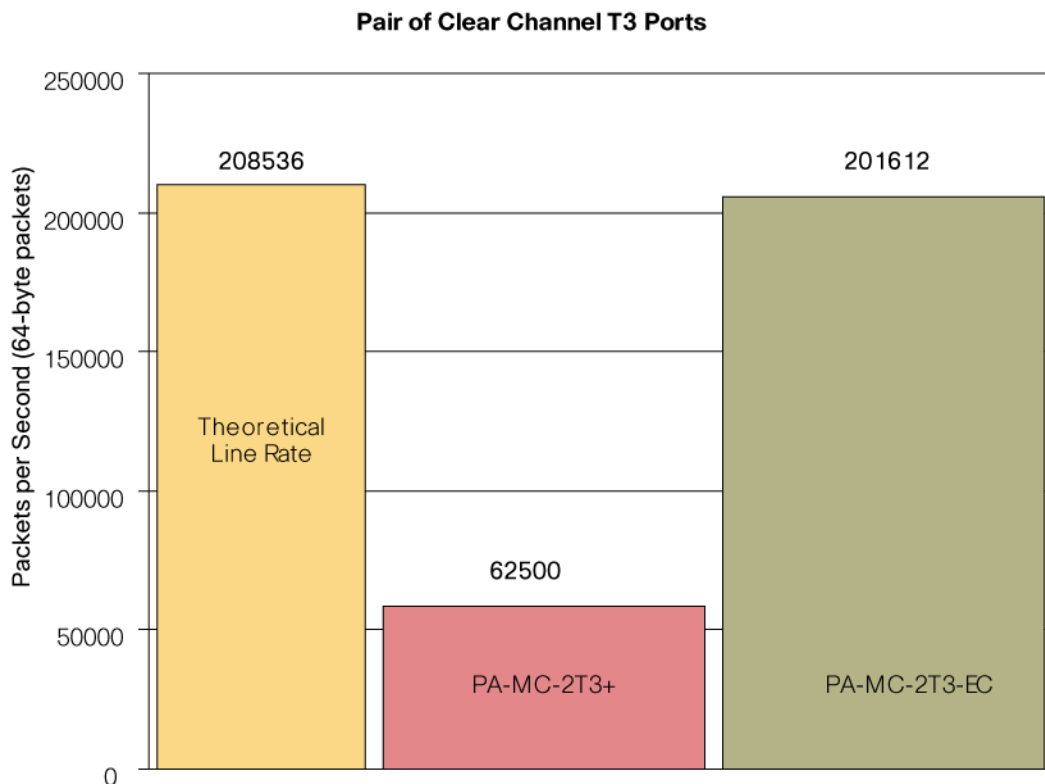
Test Topology Details

The topology shown in Figure 13 was used to measure the performance of the Cisco 2-Port Multichannel Enhanced Capability Port Adapter (PA-MC-2T3-EC). A similar topology was used to test the other port adapters.

Figure 13.   Topology with Cisco 2-Port Multichannel Enhanced Capability Port Adapter
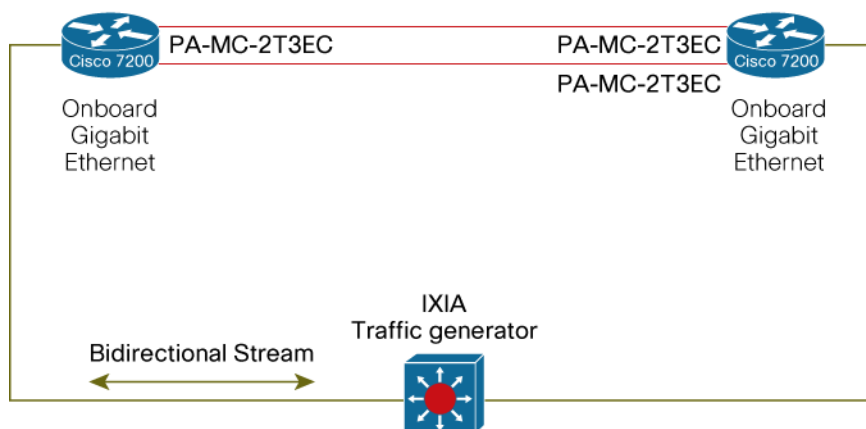
The minimum size packet generated by the traffic generator on an Ethernet segment is 64 bytes. The packet traversing the serial link will be smaller because the Ethernet header is larger than the serial link (PPP) header.

Appendix A—Definitions and Further Reading

- MLPPP—Multilink Point-to-Point Protocol: http://www.ietf.org/rfc/rfc1990.txt

- LFI—Link fragmentation and interleaving—PPP fragmentation defined in RFC 1990
  http://www.ietf.org/rfc/rfc1990.txt

- MLFR—Multilink Frame Relay defined by the Frame Relay Forum in specification FRF.16:
  http://www.mfaforum.org/frame/Approved/FRF.16/frf16.pdf

- FRF.12—Frame Relay Forum specification defining fragmentation on Frame Relay links:
  http://www.mfaforum.org/frame/Approved/FRF.12/frf12.pdf

- Link Aggregation Control Protocol (LACP) is part of an IEEE specification (802.3ad):
  http://shop.ieee.org/ieeestore/Product.aspx?product_no=SH95507

- Alternatives for Parallel N x T1/E1 connections:
  http://www.cisco.com/en/US/products/hw/modules/ps2033/products_white_paper09186a0080091d4b.shtml