

HPR Jumbo Frame Support

CalREN HPR Technical Advisory Council
October 28, 2004

Chris Costa

CENIC

Ccosta@cenic.org

Discussion Topics

- IP v4/v6 MTU across HPR Backbone
- IP v4/v6 MTU facing Associates
- IP v4/v6 MTU facing Abilene
- PMTU Discovery
 - Jumbo frame support w/in the campus

Current Configuration

- HPR Backbone 10GE interconnects
 - 9180B IPv4/v6 MTU
- HPR 1GE/10GE associate interconnects
 - 9180B IPv4/v6 MTU
 - Juniper GE Supports 9178B, or 9174B w/ 802.1q
- LAX-HPR 10GE interconnect to Abilene
 - 9180B IPv4 MTU, Abilene set to 9178B
 - 9180B IPv6 MTU, Abilene set to 9000B
- Abilene Backbone OC192 interconnects
 - 9180B IPv4/v6 MTU
 - <http://loadrunner.uits.iu.edu/~gcbrowni/Abilene/vn/interfaces/interfaces-ll.html#losang>
- HPR Peering to PacWave Exchange
 - 9000B IPv4/v6 MTU

Abilene Interface Config

(Interface Facing LAX-HPR)

```
ge-0/1/0 {  
  apply-groups INTERFACE-CONNECTOR;  
  mtu 9192;  
  link-mode full-duplex;  
  unit 0 {  
    description "CalREN South | AS:2153";  
    family inet {  
      mtu 9178;  
      address 137.164.25.3/31;  
    }  
    family inet6 {  
      mtu 9000;  
      address 2001:468:ff:144e::1/64;  
    }  
  }  
}
```

(Group config for Abilene Backbone Interfaces)

```
groups {  
  INTERFACE-BACKBONE {  
    interfaces {  
      <*> {  
        mtu 9192;  
        encapsulation cisco-hdlc;  
        sonet-options {  
          fcs 32;  
        }  
        unit 0 {  
          family inet {  
            mtu 9180;  
            filter {  
              input backbone-in;  
              output interface-out;  
            }  
          }  
          family iso {  
            mtu 1497;  
          }  
          family inet6 {  
            mtu 9180;  
          }  
        }  
      }  
    }  
  }  
}
```

Internet2-wide Recommendation on IP MTU

From: Rick Summerhill, Associate Director of Backbone Network
Infrastructure rrsum@internet2.edu

Date: Mon, 17 Feb 2003 16:36:37 -0500 (EST)

To: abilene-ops-l@INDIANA.EDU Subject: MTU sizes

Engineers throughout all components of the extended Internet2 infrastructure, including its campus LANs, its gigaPoPs, its backbone(s), and exchange points, are encouraged to support, wherever practical, an IP MTU of 9000 bytes.

- The rationale for this recommendation includes the following points:
 - Applications, including but not limited to bulk TCP, benefit from being able to send 8K (i.e., 8 times 1024) bytes of payload plus various headers. An IP MTU of 9000 would satisfy this application need.
 - A growing number of routers, switches, and host NICs support IP packets of at least 9000.
 - Very few routers, switches, and host NICs support IP packets of more than 9500. Thus, there is comparatively little motivation for a value much more than 9000.
 - **There is anecdotal evidence that Path MTU discovery would be more reliable if a given agreed-on value were commonly used. This relates to weaknesses in current Path MTU discovery technology.**
 - 9000 is an easy number to remember.

<http://www.abilene.iu.edu/rrsum-almes-mtu.html>

<http://www.abilene.iu.edu/IumboMTU.html>

Joint Engineering Team (JET) Recommendation

- U.S.A. Federal & academic research networks coordinating group, with representatives from Abilene, DREN, ESNET, NISN, NREN, USGS, vBNS, and others.
 - “It is the recommendation of the JET that the JETnets support an IP MTU of 9,000 bytes.”
- <http://www.itrd.gov/iwg/lsn/jet/index.html>

CENIC Recommendation

- HPR 10GE Backbone interconnects
 - 9180B IPv4/v6 MTU
- HPR 1GE/10GE associate interconnects
 - 9000B IPv4/v6 MTU
- LAX-HPR 10GE interconnect to Abilene
 - 9000B IPv4/v6 MTU
- Support PMTU Discovery w/in HPR Backbone

PMTU Discovery

- Utilize PMTU Discovery to exploit jumbo capable infrastructure.
 - Path MTU - The smallest MTU of any link on the current path between two hosts.
 - Source sends packets that have max size of the lesser of the local MTU or the MSS announced by the remote system. Sent with DF bit set.
 - DF Bit- Indicates packet should not be fragmented by routers. Instead an ICMP "can't fragment" error is returned to the sender and the packet is dropped.
 - ICMP Can't Fragment Error (type 3 (destination unreachable), code 4 (fragmentation needed but don't-fragment bit set))
- PMTU Discovery for IPv6
 - Minimum datagram size = 1280B
 - Only source can fragment. Routers do not.

ICMP filtering and PMTU-D

- Allow ICMP "unreachable" and "time-exceeded"

```
access-list 101 permit icmp any any unreachable
access-list 101 permit icmp any any time-exceeded
access-list 101 deny icmp any any
access-list 101 permit ip any any
```

Fragmentation

- Increases CPU and memory allocation on routers and receivers.
- Dropped fragments and retransmissions.
 - If one fragment is dropped, then the entire original IP datagram must be resent.
- Firewalls matching on L4 through L7 information may drop datagrams.

Jumbo Frame Clean Networking Gear

- <http://darkwing.uoregon.edu/%7Ejoe/jumbo-clean-gear.html>

Comments / Discussion

- Support of 9000B IP MTU?
- PMTU-D issues w/in campus?
- IPv6 PMTU-D
 - Anyone doing it now?

Conclusion

In agreement with the HPR-TAC, CENIC engineering will move forward with establishing 9000B IPv4/v6 MTU with HPR Associates and the Abilene interconnect at LAX.