

# Configuring: My external IPoIB hardware isn't fragmenting packets properly, how do I fix that?

## Background

While IP over Infiniband (IPoIB) on the internal network is handled without a problem because it all runs on IB hardware, IPoIB on an external interface may need adjustment of settings, depending on the network hardware used.

One of the configuration settings on the external network that may need adjustment is the MTU (maximum transmission unit) (in bytes). The default ethernet MTU is 1500, and ethernet MTUs larger than this are normally fragmented when required.

On the other hand, an IPoIB external network may be configured to use an MTU of say, 10000, and routers on this network may drop packets greater than this value if they can't handle fragmentation of packets larger than this properly ("properly" meaning as found by classical path MTU discovery with ICMP pings reporting "Packet Too Big" messages when trying to learn the MTU of a path, as described in [RFC 1191](#) and [RFC1981](#)).

## Fix

Currently (for BCM upto version 6.0) a workaround is to clamp the size of packets that go out on to the external network from the cluster by configuring an MSS (MTU -40) setting on shorewall's external interface. So in the case of an MTU of 10000 set on the external network, this means having to change the "net" entry in /etc/shorewall/interfaces from the default of "dhcp" to "dhcp,mss=9960":

```
# BEGIN AUTOGENERATED SECTION -- DO NOT REMOVE
```

```
net  eth0  detect  dhcp,mss=9960
nat  eth1  detect  dhcp
# END AUTOGENERATED SECTION -- DO NOT REMOVE
```

Because this is in the cmdaemon-maintained section of the interfaces file, to use this, the interfaces file must be frozen.

This workaround may be superceded by a cmdaemon directive option in bright 6.0 and later, eventually.

Unique solution ID: #1052

Author: Frank Furter

Last update: 2012-07-25 11:27