

issues with SO_PRIORITY and IP_TOS

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2003-10/6306.html>

From: Chris Friesen (cfriesen_at_nortelnetworks.com)

Date: 10/30/03

Date: Thu, 30 Oct 2003 14:59:34 -0500
To: netdev@oss.sgi.com, linux-kernel@vger.kernel.org

I've been doing some experimenting with both of the options mentioned in the subject line, and it seems that there is some strangeness in the current handling.

First, setting IP_TOS sets the whole 8 bits of the tos field in the packet header. However, the code then uses the 4 bits defined as the tos field to generate the packet priority value. This is bad for two reasons. Firstly, if we're using the old bit fields it should be the precedence bits that are used for the skb priority rather than the tos field. Secondly, the whole precedence/tos thing has been obsoleted by the 6-bit DSCP field, of which the first 3 bits are supposed to be backwards compatible with the old precedence field. Shouldn't we properly handle that?

Secondly, for vlan priority tagging there are only 3 bits available. This means that practically speaking anyone using vlan priorities needs to limit themselves to priorities 0-7.

Currently, for me to send a packet with IP precedence bits set to a nonzero value *and* vlan priority set to the same value, I have to do the following:

```
int opt = PRIORITY << 5;
setsockopt(mysocks[i], SOL_IP, IP_TOS, &opt, sizeof(opt));
opt = PRIORITY;
setsockopt(mysocks[i], SOL_SOCKET, SO_PRIORITY, &opt, sizeof(opt));
```

The first call sets the IP precedence bits, and also incorrectly sets the socket priority. The second call sets the proper socket priority so that the vlan egress mapping works properly.

This is kind of ugly. I propose adding a new IP socket option, IP_DSCP, which would let you set the 6-bit DSCP value (which is then shifted by two bits in the kernel to generate the 8-bit value for the header field). The high-order 3 bits would then be automatically used to set the socket priority to make a vlan egress mapping simple.

Linux-Kernel: issues with SO_PRIORITY and IP_TOS

Does this make any sense?

Chris

--

Chris Friesen	MailStop: 043/33/F10
Nortel Networks	work: (613) 765-0557
3500 Carling Avenue	fax: (613) 765-2986
Nepean, ON K2H 8E9 Canada	email: cfriesen@nortelnetworks.com

-

To unsubscribe from this list: send the line "unsubscribe linux-kernel" in the body of a message to majordomo@vger.kernel.org

More majordomo info at <http://vger.kernel.org/majordomo-info.html>

Please read the FAQ at <http://www.tux.org/lkml/>