

Re: dummysnet, em driver, device polling issues :-(

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/net/2005-10/0050.html>

From: Dave+Seddon (dave-dated-1128902191.bcc743_at_seddon.ca)

Date: 10/05/05

To: net@freebsd.org

Date: Wed, 05 Oct 2005 09:56:30 +1000

Jeremie,

Sorry for "top posting". My time machine is broken :)

Kevin,

You mention your running at "near" line rate. What are you pushing or pulling? Whats the rough spec of these machines pushing out this much data? What setting do you have for the polling? I've been trying to do near line rate and can't even get close with new HP-DL380s (Single 3.4 Ghz Xeon). I think the PCI bus might be the problem. The em Intel NICs I found to be very slow and stop after about 3 hours. - The Intel NICs I have are dual port, although they end up on seperate IRQs.

```
-----
cat /var/run/dmesg | grep em
em0: Ethernet address: 00:11:0a:57:70:fa
em0: Speed:N/A Duplex:N/A
em1: <Intel(R) PRO/1000 Network Connection, Version - 1.7.35> port
0x5040-0x507f mem 0xfde60000-0xfde7ffff irq 73 at device 1.1 on pci6
em1: Ethernet address: 00:11:0a:57:70:fb
em1: Speed:N/A Duplex:N/A
em2: <Intel(R) PRO/1000 Network Connection, Version - 1.7.35> port
0x6000-0x603f mem 0xfdf80000-0xfdfbffff,0xfdfc0000-0xfdfcffff irq 97 at
device 1.0 on pci10
em2: Ethernet address: 00:11:0a:57:73:6a
em2: Speed:N/A Duplex:N/A
em3: <Intel(R) PRO/1000 Network Connection, Version - 1.7.35> port
0x6040-0x607f mem 0xfdf60000-0xfdf7ffff irq 98 at device 1.1 on pci10
em3: Ethernet address: 00:11:0a:57:73:6b
em3: Speed:N/A Duplex:N/A
-----
```

```
ps ax | grep em
84 ?? WL 0:00.00 [irq73: em1]
85 ?? WL 0:00.00 [irq74: em0]
108 ?? WL 0:00.00 [irq97: em2]
109 ?? WL 0:00.00 [irq98: em3]
```

Ferdinand,

After giving up on the Intel cards in the DL380s I started using the onboard broadcom cards (bge). They work great, although I don't seem to be able to get near line rate either. I've been severing up < 10 files from MFS via thttpd. I get about 80MB/s only. :(

Regards,
Dave

Jeremie Le Hen writes:

```
> Hi Benjamin, Ferdinand,
>
> (Please avoid top-posting, this reverts the flow of the conversation
> and make the whole thread difficult to follow.)
>
>> i have been messing with the em driver now for over a month, ive come to
>> the conclusion is a piece of crap. if you watch on this list every
>> other day you have someone saying there em driver is causing some sort
>> of error, this should not be on a nic from a company like intel. im
>> saddly contimplating moving over to fedora right now just so i can work
>> until 6.0 comes out (which i doubt will solve the problem anyway since
>> im using the drivers from 6.0 now and there not helping out either).
>> somebody really needs to look into this and find out what the hell is
>> going on as i consider this a major problem right now.
>
> em(4) is known to be full of problems, it would indeed require someone
> taking the maintainership of the driver and then reworking it a bit.
>
>
>> >>After you experience your problems, can you do "sysctl -w
>> >>hw.em0.stats=1" and "sysctl -w hw.em0.debug_info=1" and post what
>> >>gets dumped to your syslog/dmesg output?
>> >
>> >
>> >em0: Excessive collisions = 0
>> >em0: Symbol errors = 0
>> >em0: Sequence errors = 0
>> >em0: Defer count = 11
>> >em0: Missed Packets = 0
>> >em0: Receive No Buffers = 0
>> >em0: Receive length errors = 0
>> >em0: Receive errors = 0
>> >em0: Crc errors = 0
>> >em0: Alignment errors = 0
>> >em0: Carrier extension errors = 0
>> >em0: XON Rcvd = 11
>> >em0: XON Xmtd = 0
>> >em0: XOFF Rcvd = 11
```

freebsd-net: Re: dummysnet, em driver, device polling issues :-(

```
>> >em0: XOFF Xmtd = 0
>> >em0: Good Packets Rcvd = 283923273
>> >em0: Good Packets Xmtd = 272613648
>> >em0: Adapter hardware address = 0xc12cfb48
>> >em0: CTRL = 0x58f00249
>> >em0: RCTL = 0x8002 PS=(0x8402)
>> >em0:tx_int_delay = 66, tx_abs_int_delay = 66
>> >em0:rx_int_delay = 0, rx_abs_int_delay = 66
>> >em0: fifo workaround = 0, fifo_reset = 0
>> >em0: hw tdh = 173, hw tdt = 173
>> >em0: Num Tx descriptors avail = 256
>> >em0: Tx Descriptors not avail1 = 0
>> >em0: Tx Descriptors not avail2 = 0
>> >em0: Std mbuf failed = 0
>> >em0: Std mbuf cluster failed = 0
>> >em0: Driver dropped packets = 0
>> >
>> >>We're using polling on nearly all the servers, and don't see ierrs at
>> >>all.
>> >
>> >
>> >Hm. That's strange. The above values were gathered with polling
>> >disabled. As soon as I enable polling, ierrs on the em0 interface are
>> >rising:
>> >
>> >em0: Excessive collisions = 0
>> >em0: Symbol errors = 0
>> >em0: Sequence errors = 0
>> >em0: Defer count = 11
>> >em0: Missed Packets = 39
>> >em0: Receive No Buffers = 2458
>> >em0: Receive length errors = 0
>> >em0: Receive errors = 0
>> >em0: Crc errors = 0
>> >em0: Alignment errors = 0
>> >em0: Carrier extension errors = 0
>> >em0: XON Rcvd = 11
>> >em0: XON Xmtd = 4
>> >em0: XOFF Rcvd = 11
>> >em0: XOFF Xmtd = 43
>> >em0: Good Packets Rcvd = 315880003
>> >em0: Good Packets Xmtd = 303985941
>> >em0: Adapter hardware address = 0xc12cfb48
>> >em0: CTRL = 0x58f00249
>> >em0: RCTL = 0x8002 PS=(0x8402)
>> >em0:tx_int_delay = 66, tx_abs_int_delay = 66
>> >em0:rx_int_delay = 0, rx_abs_int_delay = 66
>> >em0: fifo workaround = 0, fifo_reset = 0
>> >em0: hw tdh = 57, hw tdt = 57
>> >em0: Num Tx descriptors avail = 249
>> >em0: Tx Descriptors not avail1 = 0
```

Re: dummysnet, em driver, device polling issues :-(

freebsd-net: Re: dummysnet, em driver, device polling issues :-(

>> >em0: Tx Descriptors not avail2 = 0
>> >em0: Std mbuf failed = 0
>> >em0: Std mbuf cluster failed = 0
>> >em0: Driver dropped packets = 0
>> >
>> >
>> >Can you tell me what settings you are using for polling? I have set it
>> >to HZ=1000 and burst_max=300.
>> >
>> >I have now noticed another thing which might indicate one of the
>> >possible causes for the problem – this box until now ran FreeBSD 4.x
>> >and did not support ipfw tables to lock out whole lists of IP
>> >addresses. So there were quite a few inefficient rules for this. I now
>> >put all the locked IP addresses in a table which is referenced by only
>> >one rule. Since I did this, the ierrs seem to rise slower with polling
>> >enabled.
>
> "Receive No Buffers" grows when polling is enabled and it's somewhat
> a normal behaviour. When polling is not enabled, an interrupt will
> be generated for each incoming packet and the latter will be soon
> removed from the NIC buffer space, whereas when polling is enabled
> I think the kernel will check the NIC state upon each soft clock
> interrupt (HZ) and fetch them into the memory if any. If too much
> packets were received during a period, then the overflow of packets
> will be discarded, incrementing the "Receive No Buffers" error count.
> I think you can slightly increase the HZ value to decrease this
> error count, but I'm not sure this will improve the bandwidth in a
> great order of magnitude.
>
> I know that Intel GigE NICs have a smart way to to interrupt throttling
> (that's what tx_int_delay, tx_abs_int_delay, rx_int_delay and
> rx_abs_int_delay stand for). You should try to tune them through
> dev.em.[0-9]+. sysctl tree.
> These thresholds are very well explained here :
> <http://www.intel.com/design/network/applnotes/ap450.pdf>
>
> I hope this will help.
>
> Please let us know about the results.
>
> Regards,
> --
> Jeremie Le Hen
> <jeremie at le-hen dot org >< ttz at chchile dot org >
>
> _____
> freebsd-net@freebsd.org mailing list
> <http://lists.freebsd.org/mailman/listinfo/freebsd-net>
> To unsubscribe, send any mail to "freebsd-net-unsubscribe@freebsd.org"

Re: dummysnet, em driver, device polling issues :-(

freebsd-net: Re: dummysnet, em driver, device polling issues :-(

freebsd-net@freebsd.org mailing list

<http://lists.freebsd.org/mailman/listinfo/freebsd-net>

To unsubscribe, send any mail to "freebsd-net-unsubscribe@freebsd.org"