

Re: Device polling heavy traffic

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/performance/2005-12/msg00076.html>

- *From:* Mike Jakubik <mikej@xxxxxxxxxxx>
 - *Date:* Tue, 27 Dec 2005 16:20:01 -0500
-

David Gilbert wrote:

```
"Mihai"  
== Mihai  
Tanasescu  
<mihai@xxxxxxxxxxx>  
writes:
```

```
Mihai> Hello, I'm running the following setup:
```

```
Mihai> FreeBSD Dual Xeon 3Ghz machine (SMP enabled)
```

```
Mihai> 3 x 100 Mbits/s links (used at 80% capacity) - 3 x Intel 100  
Mihai> fxp
```

```
Mihai> 1 x 1000 Gbit link to a cisco router (transfers downstream the  
Mihai> other 3 links) - 1 x Intel em
```

```
Mihai> I'm getting something around 100k pkt/sec input and 100k  
Mihai> pkt/sec output as "systat -ip 1" shows.
```

```
Mihai> Kernel polling is enabled. I have tried options HZ=1000,  
Mihai> options HZ=2500 to see if anything changes.
```

```
Mihai> The problem: If I ping this machine or anything that is routed  
Mihai> through it I get response times of 10-15-30 ms and once in 30  
Mihai> seconds a packet is lost.
```

Re: Device polling heavy traffic

Mihai> If I disable kernel.polling.enable then I get response times of
Mihai> 1-2-3 ms but I also get a lot of interrupts and a kernel panic
Mihai> after about 20 min.

We've done a bunch of different experiments on various hardware and various operating systems. 300 kpps of very small packets is about the forwarding limit of FreeBSD ... with any hardware we can find. So if your packets are non-trivial ... there is a lesser cost to packet size than packet number ... then 200kpps will likely show some loss due to FreeBSD's inability to forward more packets.

Keep in mind that I've had two engineers spend months on this with some guidance from me.

Now a stock linux on the same hardware can handle about 500 kpps, but there's a caveat. Linux hashes packet streams (key of src ip, src port, dest ip, dest port) for both routed and terminated traffic. This hash has some advantages, but has a huge drawback. If you spray more than the hash size worth of streams at the linux box (even if it's not routing), then it basically falls over.

(not quite... packet performance goes from 500 kpps to less than 10kpps and everything is hosed until it stops. Profiling shows it spends 99% of it's time in the hash emptying and allocating code)

This means that a few megabit of fully random small packets will knock over an arbitrary linux box. Every time.

Not even SCO boxes are that lame (hah... :)

My current thinking is that the hardware model has to change to get better performance. I'm not really a hardware guy (hardware is evil), but as I understand it, we're approaching limits of PCI busses.

Post the message enough times? Have you tried on a PCI Express bus?

Re: Device polling heavy traffic

freebsd-performance@xxxxxxxxxxx mailing list

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

To unsubscribe, send any mail to "freebsd-performance-unsubscribe@xxxxxxxxxxx"