

Re: CPU utilisation cap?

Source: <http://unix.derkeiler.com/Mailing-Lists/FreeBSD/performance/2004-10/0032.html>

From: Robert Watson (rwatson_at_freebsd.org)

Date: 10/25/04

Date: Mon, 25 Oct 2004 10:04:06 +0100 (BST)

To: lukem.freebsd@cse.unsw.edu.au

On Thu, 21 Oct 2004 lukem.freebsd@cse.unsw.edu.au wrote:

> *I am measuring idle time using a CPU soaker process which runs at a very
> low priority. Top seems to confirm the output it gives.*
>
> *What I see is strange. CPU utilisation always peaks (and stays) at
> between 80 & 85%. If I increase the amount of work done by the UDP echo
> program (by inserting additional packet copies), CPU utilisation does
> not rise, but rather, throughput declines. The 80% figure is common to
> both the slow and fast PCI cards as well.*
>
> *This is rather confusing, as I cannot tell if the system is IO bound or
> CPU bound. Certainly I would not have expected the 133/64 PCI bus to be
> saturated given that peak throughput is around 550Mbit/s with 1024-byte
> packets. (Such a low figure is not unexpected given there are 2 syscalls
> per packet).*

A couple of thoughts, none of which points at any particular red flag, but worth thinking about:

- You indicate there are multiple if_em cards in the host — can you describe the network topology? Are you using multiple cards, or just one of the nicely equipped ones? Is there a switch involved, or direct back-to-back wires?
- Are the packet sources generating the packets synchronously or asynchronously: i.e., when a packet source sends a UDP packet, does it wait for the response before continuing, or keep on sending? If synchronously, are you sure that the wires are being kept busy?
- Make sure your math on PCI bus bandwidth accounts for packets going in both directions if you're actually echoing the packets. Also make sure to include the size of the ethernet frame and any other headers.
- If you're using SCHED_ULE, be aware that its notion of "nice" is a little different from the traditional UNIX notion, and attempts to

freebsd-performance: Re: CPU utilisation cap?

provide more proportional CPU allocation. You might try switching to SCHED_4BSD. Note that there have been pretty large scheduler changes in 5.3, with a number of the features that were previously specific to SCHED_ULE being made available with SCHED_4BSD, and that a lot of scheduling bugs have been fixed. If you move to 5.3, make sure you run with 4BSD, and it would be worth trying it with 5.2 to "see what happens".

- It would be worth trying the test without the soaker process but instead a sampling process that polls the kernel's notion of CPU% measurement every second. That way if it does turn out that ULE is unnecessarily giving CPU cycles to the soaker, you can still measure w/o "soaking".
- What does your soaker do -- in particular, does it make system calls to determine the time frequently? If so, the synchronization operations and scheduling cost associated with that may impact your measurements. If it just spins reading the tsc and outputting once in a while, you should be OK WRT this point.
- Could you confirm using netstat -s statistics that a lot of your packets aren't getting dropped due to full buffers on either send or receive. Also, do you have any tests in place to measure packet loss? Can you confirm that all the packets you send from the Linux boxes are really sent, and that given they are sent, that they arrive, and vice versa on the echo? Adding sequence numbers and measuring the mean sequence number difference might be an easy way to start if you aren't already.

Robert N M Watson FreeBSD Core Team, TrustedBSD Projects
robert@fledge.watson.org Principal Research Scientist, McAfee Research

freebsd-performance@freebsd.org mailing list

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

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