

Re: [PATCH] i386: Selectable Frequency of the Timer Interrupt

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2005-07/4164.html>

From: Andi Kleen (ak_at_suse.de)

Date: 07/15/05

To: "Brown, Len" <len.brown@intel.com>

Date: 15 Jul 2005 19:02:24 +0200

"Brown, Len" <len.brown@intel.com> writes:

> > *That's an APIC bug.*
> > *When Intel originally released the APIC (some*
> > *thirteen years ago) they stated it should be used as a source of the*
> > *timer*
> > *interrupt instead of the 8254. There is no excuse for changing the*
> > *behaviour after so many years. So if you are on a broken system, you*
> > *may*
> > *want to work around the bug, but it shouldn't impact good systems.*
>
>
> *I'm perfectly happy having Linux optimize itself for the hardware*
> *that it is running on. However, the (harsh, I know) reality is that*
> *systems with a reliable LAPIC timer in the face of C3 do not exist*
> *today, and probably never will. (don't shoot me, it wasn't my design*
> *decision, I'm just the messenger:–) Further, I expect that power saving*
> *features, such as C3, will become more important and deployed more*
> *widely in the future, rather than less widely.*

At least on multi processor systems LAPIC has to work anyways (otherwise you cannot schedule other CPUs), so it is fine to use there.

AFAIK there are no x86 CPUs right now that do both C3 and SMP. If they ever do then they will need to keep the LAPIC ticking in C3.

This has nothing even to do with advanced power saving, but is pretty much a hard requirement for Linux (and I would be surprise if it wasn't one for other OS too). Without it scheduling and local timers on APs will not work at all.

In theory it could be replaced with HPET if HPET had enough banks (one for each CPU – most implementations today usually only have 2 or 4), but that would severely limit scalability for lazy tick schemes because

Linux-Kernel: Re: [PATCH] i386: Selectable Frequency of the Timer Interrupt

they would depend on a common resource in the southbridge. Also the max number of banks needed on a big system would be huge (128? 256?) because you couldn't have more CPUs than that.

With PIC only it's absolutely impossible.

>
> *So, the 13-year-old design advice will continue to apply to*
> *13-year-old systems, but newer systems with C3 and HPET*
> *should be using them.*

Problem is that many systems even though they have HPET in the hardware don't advertise it in ACPI because Windows doesn't use it yet. And Linux can't use it then neither because it doesn't know where the registers are located (and guessing is too risky)

This means they are stuck with the old PIC, which makes lazy ticks very unpleasant. I think this is a big problem. We cannot wait until Redmond brings their new release out with this.

-Andi

-

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/>