

Re: [patch 6/6] x86: add c1e aware idle function

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2008-06/msg07185.html>

- *From:* "Rafael J. Wysocki" <rjw@xxxxxxx>
 - *Date:* Thu, 19 Jun 2008 00:27:39 +0200
-

On Thursday, 19 of June 2008, Thomas Gleixner wrote:

On Thu, 19 Jun 2008, Rafael J. Wysocki wrote:

On Wednesday, 18 of June 2008, Thomas Gleixner wrote:

On Wed, 18 Jun 2008, Rafael J. Wysocki wrote:

On Wednesday, 18 of June 2008, Pavel Machek wrote:

On Thu 2008-06-12
10:29:00, Thomas Gleixner
wrote:

C1E on
AMD
machines is
like C3 but
without
control
from the
OS. Up to
now we
disabled the
local apic
timer for
those
machines as
it stops
when the
CPU goes
into C1E.
This
excludes
those
machines
from high
resolution

Re: [patch 6/6] x86: add c1e aware idle function

timers /
dynamic
ticks, which
hurts
especially
the X2
based
laptops.

The current
boot time
C1E
detection
has another
more
serious
flaw:
some
BIOSes do
not enable
C1E until
the ACPI
processor
module is
loaded. This
causes
systems to
stop
working
after that
point.

To work
nicely with
C1E
enabled
machines
we use a
separate
idle
function,
which
checks on
idle entry
whether
C1E was
enabled in
the
Interrupt
Pending
Message

Re: [patch 6/6] x86: add c1e aware idle function

MSR. This
allows us to
do timer
broadcasting

Entering idle is quite a
common operation, and
reading MSR is quite
slow. Is it possible to do
better here?

What happens if ACPI
BIOS toggles MSR on all
cpus *while* we are
entering idle? This seems
inherently racy...

Well, on most of the machines we see the C1E bit on the first
idle
entry either on CPU0 or on CPU1.

I know of exactly one machine which has the C1E thing
enabled late
when the ACPI stuff runs.

Yes, and that fits the picture I'm observing
on the nx6325 (see the
"linux-next: Tree for June 13: IO APIC
breakage on HP nx6325" thread).

Hmm, when is the C1E detected on current mainline ? Does
the boot CPU
have it or is it when the second CPU comes up ?

How can I check that?

When the boot cpu has it then it prints:

"Disabling APIC timer"

It does that.

Re: [patch 6/6] x86: add c1e aware idle function

when the secondary CPU has it then it prints:

"AMD C1E detected late. Force timer broadcast."

Current mainline does not check for the late ACPI case so I'm pretty sure that on your box it is detected during early boot way before we can switch to highres/dyntick.

Yes, that's the case.

Thanks,

Rafael

--

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

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

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