

sched domains kernbench improvements

Source: <http://linux.derkeiler.com/Mailing-Lists/Kernel/2004-02/7608.html>

From: Nick Piggin (piggin_at_cyberone.com.au)

Date: 02/28/04

Date: Sat, 28 Feb 2004 20:21:02 +1100

To: Con Kolivas <conman@kolivas.net>, Andrew Morton <akpm@osdl.org>

Con,

I was able to reproduce your half-load kernbench problems on the non-NUMA stp 8-way.

I made a pretty simple "lessidle" patch which tweaks some sched domain parameters to be more inclined to move tasks, especially when idle. That brought performance to nearly exactly the same as 2.6.3.

Context switches are still up, but user and system time is down a bit. So indicates it is still less balance-happy but is obviously enough to bring the idle time down.

2.6.3: <http://khack.osdl.org/stp/288459/>

2.6.3-mm4-lessidle: <http://khack.osdl.org/stp/288995/>

Phew! So it is more a matter of tuning than anything fundamental. It may be that the patch now makes balancing too aggressive, but it is probably better to err on the side that is closer to 2.6 behaviour.

I haven't tested this on much else.

```
linux-2.6-npiggin/include/linux/sched.h | 16 ++++++-----  
linux-2.6-npiggin/kernel/sched.c | 5 +++++  
2 files changed, 13 insertions(+), 8 deletions(-)
```

```
diff -puN include/linux/sched.h~sched-lessidle include/linux/sched.h  
--- linux-2.6/include/linux/sched.h~sched-lessidle 2004-02-21 10:57:07.000000000 +1100  
+++ linux-2.6-npiggin/include/linux/sched.h 2004-02-21 10:59:46.000000000 +1100  
@@ -598,11 +598,11 @@ struct sched_domain {  
    .parent = NULL, \  
    .groups = NULL, \  
    .min_interval = 1, \  
};
```

Linux-Kernel: sched domains kernbench improvements

```
- .max_interval = 8, \
- .busy_factor = 32, \
+ .max_interval = 4, \
+ .busy_factor = 64, \
    .imbalance_pct = 125, \
- .cache_hot_time = (5*1000000), \
- .cache_nice_tries = 2, \
+ .cache_hot_time = (5*1000000/2), \
+ .cache_nice_tries = 1, \
    .flags = SD_FLAG_FASTMIGRATE | SD_FLAG_NEWIDLE, \
    .last_balance = jiffies, \
    .balance_interval = 1, \
@@ -615,11 +615,11 @@ struct sched_domain {
    .span = CPU_MASK_NONE, \
    .parent = NULL, \
    .groups = NULL, \
- .min_interval = 20, \
- .max_interval = 1000*fls(num_online_cpus()), \
- .busy_factor = 4, \
+ .min_interval = 8, \
+ .max_interval = 256*fls(num_online_cpus()), \
+ .busy_factor = 8, \
    .imbalance_pct = 125, \
- .cache_hot_time = (5*1000000), \
+ .cache_hot_time = (10*1000000), \
    .cache_nice_tries = 1, \
    .flags = SD_FLAG_EXEC, \
    .last_balance = jiffies, \
diff -puN kernel/sched.c~sched-lessidle kernel/sched.c
--- linux-2.6/kernel/sched.c~sched-lessidle 2004-02-21 10:57:10.000000000 +1100
+++ linux-2.6-npiggin/kernel/sched.c 2004-02-21 16:15:18.000000000 +1100
@@ -1493,6 +1493,11 @@ nextgroup:
    return busiest;

out_balanced:
+ if (busiest && idle == NEWLY_IDLE) {
+ *imbalance = 1;
+ return busiest;
+ }
+
+ *imbalance = 0;
+ return NULL;
}
```

-
-
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/>