Subject: [PATCH 7/7] event: add tracepoint for accounting block time
Posted by Andrew Vagin on Mon, 28 Nov 2011 09:03:35 GMT
View Forum Message <> Reply to Message

This tracepoint shows how long a task is sleeping in uninterruptible state.

E.g.
It may show how long and where a mutex is waited.

Signed-off-by: Andrew Vagin <avagin@openvz.org>
---
include/trace/events/sched.h | 7 +++++++
kernel/sched_fair.c | 2 ++
2 files changed, 9 insertions(+), 0 deletions(-)

```
diff --git a/include/trace/events/sched.h b/include/trace/events/sched.h
index 959ff18..be077cf 100644
--- a/include/trace/events/sched.h
+++ b/include/trace/events/sched.h
@@ -331,6 +331,13 @@ DEFINE_EVENT(sched_stat_template, sched_stat_iowait,
      TP_ARGS(tsk, delay));

 /*
+ * Tracepoint for accounting block time (time the task is in uninterruptible).
+ */
+DEFINE_EVENT(sched_stat_template, sched_stat_block,
+     TP_PROTO(struct task_struct *tsk, u64 delay),
+     TP_ARGS(tsk, delay));
+
+/*
  * Tracepoint for accounting runtime (time the task is executing
  * on a CPU).
  */
diff --git a/kernel/sched_fair.c b/kernel/sched_fair.c
index 5c9e679..0d7b156 100644
--- a/kernel/sched_fair.c
+++ b/kernel/sched_fair.c
@@ -907,6 +907,8 @@ static void enqueue_sleeper(struct cfs_rq *cfs_rq, struct sched_entity
*se)
     trace_sched_stat_iowait(tsk, delta);
    }

+   trace_sched_stat_block(tsk, delta);
+
    /*
     * Blocking time is in units of nanosecs, so shift by
     * 20 to get a milliseconds-range estimation of the
--
```

1.7.1