Subject: Re: RSS controller v2 Test results (Imbench)
Posted by William Lee Irwin III on Mon, 21 May 2007 14:59:17 GMT
View Forum Message <> Reply to Message

On Fri, 2007-05-18 at 09:37 +0530, Balbir Singh wrote:

- >> oops! I wonder if AIM7 creates too many processes and exhausts all
- >> memory. I've seen a case where during an upgrade of my tetex on my
- >> laptop, the setup process failed and continued to fork processes
- >> filling up 4GB of swap.

On Mon, May 21, 2007 at 09:53:34AM -0400, Lee Schermerhorn wrote:

- > Jumping in late, I just want to note that in our investigations, when
- > AIM7 gets into this situation [non-responsive system], it's because all
- > cpus are in reclaim, spinning on an anon_vma spin lock. AIM7 forks [10s
- > of] thousands of children from a single parent, resultings in thousands
- > of vmas on the anon_vma list. shrink_inactive_list() must walk this
- > list twice [page_referenced() and try_to_unmap()] under spin_lock for
- > each anon page.

I wonder how far out RCU'ing the anon_vma lock is.

On Mon, May 21, 2007 at 09:53:34AM -0400, Lee Schermerhorn wrote:

- > [Aside: Just last week, I encountered a similar situation on the
- > i_mmap_lock for page cache pages running a 1200 user Oracle/OLTP run on
- > a largish ia64 system. Left the system spitting out "soft lockup"
- > messages/stack dumps overnight. Still spitting the next day, so I
- > decided to reboot.]
- > I have a patch that turns the anon vma lock into a reader/writer lock
- > that alleviates the problem somewhat, but with 10s of thousands of vmas
- > on the lists, system still can't swap enough memory fast enough to
- > recover.

Oh dear. Some algorithmic voodoo like virtually clustered scanning may be in order in addition to anon_vma lock RCU'ing/etc.

On Mon, May 21, 2007 at 09:53:34AM -0400, Lee Schermerhorn wrote:

- > We've run some AIM7 tests with Rik's "split Iru list" patch, both with
- > and without the anon vma reader/writer lock patch. We'll be posting
- > results later this week. Quick summary: with Rik's patch, AIM
- > performance tanks earlier, as the system starts swapping earlier.
- > However, system remains responsive to shell input. More into to follow.

I'm not sure where policy comes into this.