Subject: Re: Subject : locks, semaphores, and such
Posted by Paul van Delst on Mon, 28 Jan 2002 15:25:37 GMT
View Forum Message <> Reply to Message

Ken Mankoff wrote:
>
> Hi,
>
> I need to create a unique directory to do some work in. This is a
> web-based IDL script (through ION), and multiple people may access the
> site at the same second. Each user gets their images generated in a
> uniquely named directory, which is then deleted 5 minutes later...
>
> I know a bunch about semaphores and threads and stuff like that, but
> here is the problem. The "threads" (users) do not have inter-thread
> communication, unless its through files on disk. If they did, this
> would be a perfect use for COMMON blocks. Also, the web server has 2
> CPUs so much of the traditional semaphore logic is invalid, as both
> threads can acquire a lock on the same clock cycle.
>
> ---
> Here is the code I currently use to lock a file:
>
> uniq = strtrim(long(systime(1)), 2)
> repeat begin
> uniq = uniq + 1L
> f = findfile(uniq,count=cnt) ; [1] ; find unclaimed file
> endrep until cnt eq 0 ; [2]
> spawn, 'touch ' + file ; [3] ; claim file
> spawn, 'rm ' + file + ' | at now + 5 minutes' ; free file later
>
> Of course, there are multiple clock cycles and disk accesses between
> when [1] loads a "0" into cnt, and when [3] finishes executing the
> 'touch' part of the command.
>
> The system this runs on is 5 people on 5 computers using 5 different
> IDL sessions all sharing 1 cross-mounted disk. So far, the bug has
> never evolved into an error :)
> ---
>
> At worst, can anyone think of a way to detect if multiple users got
> assigned the same directory and at least fail gracefully? At best,
> does anyone have an algorithm pre-built for just this purpose?

I had a similar problem a while back related to running batch jobs via unix shell scripts, but I think my problem was much simpler than yours and my solution is probably also too simple an approach (don't know nuthin' 'bout semaphores and such...) But anyway....

I create a time_date string and prefix that with the username (or some other user specific ID). This prevents users from running into each other. To prevent the user from running into him/herself (if a single user is running multiple jobs) I suffix the user_time_date string with increasing integers (up to some predefined max value.), e.g. if directory "user_time_date" exists, I check for "user_time_date_2". If that exists I increment the suffixed integer and loop. etc. etc. So far it's worked fine - I've initiated 10's of 1000's of jobs (some simultaneously but doing the job submission through multiple windows; other not necessarily simultaneously since it takes some time to process the inputs, but over about a minute or so) and haven't yet had a collision. (Fingers crossed)

paulv

--
Paul van Delst Religious and cultural
CIMSS @ NOAA/NCEP purity is a fundamentalist
Ph: (301)763-8000 x7274 fantasy
Fax:(301)763-8545 V.S.Naipaul

Subject: Re: Subject : locks, semaphores, and such
Posted by Ken Mankoff on Mon, 28 Jan 2002 17:17:17 GMT
View Forum Message <> Reply to Message

On Mon, 28 Jan 2002, Paul van Delst wrote:

> Ken Mankoff wrote:
>> I need to create a unique directory to do some work in. This is a
>> web-based IDL script (through ION), and multiple people may access the
>> site at the same second. Each user gets their images generated in a
>> uniquely named directory, which is then deleted 5 minutes later...

[snip]

> I had a similar problem a while back related to running batch jobs
> via unix shell scripts, but I think my problem was much simpler
> than yours and my solution is probably also too simple an approach
> (don't know nuthin' 'bout semaphores and such...) But anyway....
>
> I create a time_date string and prefix that with the username (or
> some other user specific ID). This prevents users from running
> into each other. To prevent the user from running into him/herself
> (if a single user is running multiple jobs) I suffix the
> user_time_date string with increasing integers (up to some
> predefined max value.), e.g. if directory "user_time_date" exists,
> I check for "user_time_date_2". If that exists I increment the
> suffixed integer and loop. etc. etc. So far it's worked fine -
> I've initiated 10's of 1000's of jobs (some simultaneously but
> doing the job submission through multiple windows; other not

> necessarily simultaneously since it takes some time to process the
> inputs, but over about a minute or so) and haven't yet had a
> collision. (Fingers crossed)

Actually, you gave me a very good solution. I will append the users IP address to the directory. That way, I only have to worry about users from the same IP accessing the page in the same second. This is much less likely. Also, rather than use SYSTIME(1) to get the seconds, I will use a millisecond timestamp.

Thanks,
Ken Mankoff

--
Kenneth Mankoff
LASP://303.492.3264
http://lasp.colorado.edu/~mankoff/
http://lasp.colorado.edu/snoedata/
http://lasp.colorado.edu/marsrobot/

Subject: Re: Subject : locks, semaphores, and such
Posted by Martin Downing on Tue, 29 Jan 2002 13:28:53 GMT
View Forum Message <> Reply to Message

"Ken Mankoff" <mankoff@I.HATE.SPAM.cs.colorado.edu> wrote in message news:Pine.LNX.4.33.0201281014520.10891-100000@snoe.colorado. edu...
> [snip]
> Actually, you gave me a very good solution. I will append the users IP
> address to the directory. That way, I only have to worry about users
> from the same IP accessing the page in the same second. This is much
> less likely. Also, rather than use SYSTIME(1) to get the seconds, I
> will use a millisecond timestamp.

Systime(1) seems to have a resolution of roughly 10 milliseconds (at least on my system). Whats the timestamp you are thinking of which gives millisecond steps?

Martin

Subject: Re: Subject : locks, semaphores, and such
Posted by Ken Mankoff on Tue, 29 Jan 2002 15:32:00 GMT
View Forum Message <> Reply to Message

On Tue, 29 Jan 2002, Martin Downing wrote:
> "Ken Mankoff" <mankoff@I.HATE.SPAM.cs.colorado.edu> wrote in message

>> [snip]
>> Actually, you gave me a very good solution. I will append the users IP
>> address to the directory. That way, I only have to worry about users
>> from the same IP accessing the page in the same second. This is much
>> less likely. Also, rather than use SYSTIME(1) to get the seconds, I
>> will use a millisecond timestamp.
>
> Systime(1) seems to have a resolution of roughly 10 milliseconds (at least
> on my system). Whats the timestamp you are thinking of which gives
> millisecond steps?

I get 1 second resolution (I think) using the IDL systime(1) command.
I had not yet found a way in IDL to get higher resolution timestamps,
and am currently looking into getting them from a spawned command in
unix.

Ken Mankoff

--
Kenneth Mankoff
LASP://303.492.3264
http://lasp.colorado.edu/~mankoff/
http://lasp.colorado.edu/snoedata/
http://lasp.colorado.edu/marsrobot/

Subject: Re: Subject : locks, semaphores, and such
Posted by Nigel Wade on Tue, 29 Jan 2002 16:42:34 GMT
View Forum Message <> Reply to Message

Ken Mankoff wrote:

> On Tue, 29 Jan 2002, Martin Downing wrote:
>> "Ken Mankoff" <mankoff@I.HATE.SPAM.cs.colorado.edu> wrote in message
>>> [snip]
>>> Actually, you gave me a very good solution. I will append the users IP
>>> address to the directory. That way, I only have to worry about users
>>> from the same IP accessing the page in the same second. This is much
>>> less likely. Also, rather than use SYSTIME(1) to get the seconds, I
>>> will use a millisecond timestamp.
>>
>> Systime(1) seems to have a resolution of roughly 10 milliseconds (at
>> least on my system). Whats the timestamp you are thinking of which gives
>> millisecond steps?
>
> I get 1 second resolution (I think) using the IDL systime(1) command.
> I had not yet found a way in IDL to get higher resolution timestamps,
> and am currently looking into getting them from a spawned command in

> unix.
>
> Ken Mankoff
>

If you're going to spawn a process, why not base the directory name on the
PID of the spawned process, and return the directory name to IDL? Two
concurrent processes cannot have the same PID.

--

Nigel Wade, System Administrator, Space Plasma Physics Group,
University of Leicester, Leicester, LE1 7RH, UK
E-mail : nmw@ion.le.ac.uk
Phone : +44 (0)116 2523568, Fax : +44 (0)116 2523555

Subject: Re: Subject : locks, semaphores, and such
Posted by Ken Mankoff on Tue, 29 Jan 2002 17:20:40 GMT
View Forum Message <> Reply to Message

On Tue, 29 Jan 2002, Nigel Wade wrote:
> Ken Mankoff wrote:
>
>> On Tue, 29 Jan 2002, Martin Downing wrote:
>>> "Ken Mankoff" <mankoff@I.HATE.SPAM.cs.colorado.edu> wrote in message
>>>> [snip]
>>>> Actually, you gave me a very good solution. I will append the users IP
>>>> address to the directory. That way, I only have to worry about users
>>>> from the same IP accessing the page in the same second. This is much
>>>> less likely. Also, rather than use SYSTIME(1) to get the seconds, I
>>>> will use a millisecond timestamp.
>>>
>>> Systime(1) seems to have a resolution of roughly 10 milliseconds (at
>>> least on my system). Whats the timestamp you are thinking of which gives
>>> millisecond steps?
>>
>> I get 1 second resolution (I think) using the IDL systime(1) command.
>> I had not yet found a way in IDL to get higher resolution timestamps,
>> and am currently looking into getting them from a spawned command in
>> unix.
>>
> If you're going to spawn a process, why not base the directory name on the
> PID of the spawned process, and return the directory name to IDL? Two
> concurrent processes cannot have the same PID.
>

This would work too... a brief expriment tells me that I do not know

how to capture a pid through IDL. Do you have any ideas (or code? :)

Thanks,
  Ken Mankoff

--
Kenneth Mankoff
LASP://303.492.3264
http://lasp.colorado.edu/~mankoff/
http://lasp.colorado.edu/snoedata/
http://lasp.colorado.edu/marsrobot/

Subject: Re: Subject : locks, semaphores, and such
Posted by Craig Markwardt on Wed, 30 Jan 2002 03:25:03 GMT
View Forum Message <> Reply to Message

Ken Mankoff <mankoff@I.HATE.SPAM.cs.colorado.edu> writes:
> On Tue, 29 Jan 2002, Martin Downing wrote:
>> "Ken Mankoff" <mankoff@I.HATE.SPAM.cs.colorado.edu> wrote in message
>>> [snip]
>>> Actually, you gave me a very good solution. I will append the users IP
>>> address to the directory. That way, I only have to worry about users
>>> from the same IP accessing the page in the same second. This is much
>>> less likely. Also, rather than use SYSTIME(1) to get the seconds, I
>>> will use a millisecond timestamp.
>>
>> Systime(1) seems to have a resolution of roughly 10 milliseconds (at least
>> on my system). Whats the timestamp you are thinking of which gives
>> millisecond steps?
>
> I get 1 second resolution (I think) using the IDL systime(1) command.
> I had not yet found a way in IDL to get higher resolution timestamps,
> and am currently looking into getting them from a spawned command in
> unix.

I believe there was a "bug" in some versions of IDL 5.2 for Linux
which caused SYSTIME(1) to only return integer values. Do you have
that version? RSI was making a version called 5.2.1L available, or
something like that, which fixed the problem.

Craig

--
 ---------- ----------
Craig B. Markwardt, Ph.D. EMAIL: craigmnet@cow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
 ---------- ----------

Subject: Re: Subject : locks, semaphores, and such
Posted by Nigel Wade on Wed, 30 Jan 2002 11:55:56 GMT
View Forum Message <> Reply to Message

Ken Mankoff wrote:

> On Tue, 29 Jan 2002, Nigel Wade wrote:

>>>
>> If you're going to spawn a process, why not base the directory name on
>> the PID of the spawned process, and return the directory name to IDL? Two
>> concurrent processes cannot have the same PID.
>>
>
> This would work too... a brief expriment tells me that I do not know
> how to capture a pid through IDL. Do you have any ideas (or code? :)
>
> Thanks,
> Ken Mankoff
>

Do you need to know the PID in IDL? I'd just return the directory name
created by the spawned process. Output sent to stdout by a spawned process
is returned to IDL in the final argument to the SPAWN command:

>> spawn,'echo hello',result

would return the output 'hello' in the variable result. So all you would
need to do is echo the name of the directory created (and make sure that's
all that the spawned process outputs) and you get what you need.

--

Nigel Wade, System Administrator, Space Plasma Physics Group,
University of Leicester, Leicester, LE1 7RH, UK
E-mail : nmw@ion.le.ac.uk
Phone : +44 (0)116 2523568, Fax : +44 (0)116 2523555

Subject: Re: Subject : locks, semaphores, and such
Posted by Ken Mankoff on Wed, 30 Jan 2002 23:18:44 GMT
View Forum Message <> Reply to Message

On Wed, 30 Jan 2002, Nigel Wade wrote:

> Ken Mankoff wrote:
>> On Tue, 29 Jan 2002, Nigel Wade wrote:
>>> If you're going to spawn a process, why not base the directory name on

>>> the PID of the spawned process, and return the directory name to IDL? Two
>>> concurrent processes cannot have the same PID.
>>
>> This would work too... a brief expriment tells me that I do not know
>> how to capture a pid through IDL. Do you have any ideas (or code? :)
>
> Do you need to know the PID in IDL? I'd just return the directory name
> created by the spawned process. Output sent to stdout by a spawned process
> is returned to IDL in the final argument to the SPAWN command:
>
>>> spawn,'echo hello',result
>
> would return the output 'hello' in the variable result. So all you would
> need to do is echo the name of the directory created (and make sure that's
> all that the spawned process outputs) and you get what you need.
>

Hi,

The PID approach has solved it. I am now using this:

SPAWN, 'true', PID=PID

and then generating a directory off of a combination of the PID and
SYSTIME(1)

Thanks for the help,
  Ken Mankoff

--
Kenneth Mankoff
LASP://303.492.3264
http://lasp.colorado.edu/~mankoff/
http://lasp.colorado.edu/snoedata/
http://lasp.colorado.edu/marsrobot/