Subject: Re: [PATCH 0/9] namespaces: Introduction
Posted by Herbert Poetzl on Sat, 20 May 2006 21:24:55 GMT
View Forum Message <> Reply to Message

On Fri, May 19, 2006 at 08:25:16AM -0700, Andrew Morton wrote:
> Andrey Savochkin <saw@sw.ru> wrote:
>>
>> I have a practical proposal. We can start with presenting and
>> merging the most interesting part, network containers. We discuss
>> details, possible approaches, and related subsystems, until
>> networking is finished to its utmost detail. This will create an
>> example of virtualization of a non-trivial subsystem, and we will
>> have to agree on basic principles of virtualization of related
>> subsystems like proc.
>>
>> Virtualization of networking presents a lot of challenges and
>> decision-making points with respect to user-visible interfaces:
>> proc, sysctl, netlink events (and netlink sockets themselves),
>> and so on. This code will also become immediately useful as an
>> improvement over chroot. I am sure that when we come to a mutually
>> acceptable solution with respect to networking, virtualization of
>> all other subsystems can be implemented and merged without many
>> questions.
>>
>> What do people think about this plan?

well, I think it is interesting ...

> It sounds like that feature might be the
> most-likely-to-cause-maintainer-revolt one, in which case yes,
> it is absolutely definitely the one to start with.

yes, I absolutely agree here, this will be one
of the tougher nuts to crack, and therefore it
might be an excellent candidate to proove that
the different virtualization camps can find an
acceptable solution .. together.

> Because if it ends up that an acceptable approach cannot be found,
> and if this feature is compulsory for any sane virtualisation
> implementation then that's it - game over.

this, OTOH is something I'm not convinced of,
because looking at BSD jails, I see a very simple
approach (only one IP, limiting binds) which seems
to be sufficient for all the BSD jails out there

this is probably something which does not meet the

requirements of fully blown distro virtualizations
but actually it might be more than sufficient for
'mainline' linux jails

> We want to discover such blockers as early in the process as
> possible.

yes, I would also appreciate if we could get some
support from the network folks, as I think, most
of them are already working into that direction
(think Van Jacobson's net channels, routing tables)

especially as the network virtualization brings up
a number of questions, which are not easily answered
like the following:

- what policy will be applied inside guests?
  + allow arbitrary packets/rules/routes
  + have some generic limits/basic rules
  + put policy into userspace

- how to 'connect' the virtual interfaces to
  the real network?
  + via routing and bridging?
    (means duplicate stack traversal and
    therefore twice the overhead)
  + via split personality interfaces?
    (less overhead, more complicated cases)
  + directly (only by isolation)

- at what level should the virtualization happen?
  + ethernet level (all protocols)
  + ip level (all ip based and control protocols)
  + udp/tcp level

best,
Herbert