Subject: [PATCH 0/2] SYSCTL: export root handling routines
Posted by Stanislav Kinsbursky on Mon, 12 Dec 2011 17:51:12 GMT
View Forum Message <> Reply to Message

These rouitines are required to make SUNRPC sysctl's network-namespace-aware.

The following series consists of:

Stanislav Kinsbursky (2):
    SYSCTL: root unregister routine introduced
    SYSCTL: export root register and unregister routines

 kernel/sysctl.c | 9 +++++++++
 1 files changed, 9 insertions(+), 0 deletions(-)

Subject: [PATCH 2/2] SYSCTL: export root register and unregister routines
Posted by Stanislav Kinsbursky on Mon, 12 Dec 2011 17:51:15 GMT
View Forum Message <> Reply to Message

These routines will be used in SUNPRC module and thus have to be exported.

Signed-off-by: Stanislav Kinsbursky <skinsbursky@parallels.com>

---
 kernel/sysctl.c | 2 ++
 1 files changed, 2 insertions(+), 0 deletions(-)

diff --git a/kernel/sysctl.c b/kernel/sysctl.c
index 21e68c1..0ee5b73 100644
--- a/kernel/sysctl.c
+++ b/kernel/sysctl.c
@@ -1700,6 +1700,7 @@ void register_sysctl_root(struct ctl_table_root *root)
  list_add_tail(&root->root_list, &sysctl_table_root.root_list);
  spin_unlock(&sysctl_lock);
 }
+EXPORT_SYMBOL_GPL(register_sysctl_root);

 void unregister_sysctl_root(struct ctl_table_root *root)
 {
@@ -1707,6 +1708,7 @@ void unregister_sysctl_root(struct ctl_table_root *root)
  list_del(&root->root_list);
  spin_unlock(&sysctl_lock);
 }
+EXPORT_SYMBOL_GPL(unregister_sysctl_root);

```
 /*
  * sysctl_perm does NOT grant the superuser all rights automatically, because
```

Subject: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by Stanislav Kinsbursky on Mon, 12 Dec 2011 17:51:18 GMT
View Forum Message <> Reply to Message

This routine is required for SUNRPC sysctl's, which are going to be allocated,
processed and destroyed per network namespace context.
IOW, new sysctl root will be registered on network namespace creation and
thus have to unregistered before network namespace destruction.

Signed-off-by: Stanislav Kinsbursky <skinsbursky@parallels.com>

```
---
 kernel/sysctl.c |    7 +++++++
 1 files changed, 7 insertions(+), 0 deletions(-)

diff --git a/kernel/sysctl.c b/kernel/sysctl.c
index ae27196..21e68c1 100644
--- a/kernel/sysctl.c
+++ b/kernel/sysctl.c
@@ -1701,6 +1701,13 @@ void register_sysctl_root(struct ctl_table_root *root)
  spin_unlock(&sysctl_lock);
 }

+void unregister_sysctl_root(struct ctl_table_root *root)
+{
+ spin_lock(&sysctl_lock);
+ list_del(&root->root_list);
+ spin_unlock(&sysctl_lock);
+}
+
 /*
  * sysctl_perm does NOT grant the superuser all rights automatically, because
  * some sysctl variables are readonly even to root.
```

Subject: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by akpm on Mon, 12 Dec 2011 22:52:08 GMT
View Forum Message <> Reply to Message

On Mon, 12 Dec 2011 21:50:00 +0300
Stanislav Kinsbursky <skinsbursky@parallels.com> wrote:

> This routine is required for SUNRPC sysctl's, which are going to be allocated,
> processed and destroyed per network namespace context.
> IOW, new sysctl root will be registered on network namespace creation and
> thus have to unregistered before network namespace destruction.
>

It's a bit suspicious that such a mature subsystem as sysctl newly
needs its internals exported like this. Either a) the net namespaces
work is doing something which hasn't been done before or b) it is doing
something wrong.

So, please explain further so we can confirm that it is a) and not b).

> --- a/kernel/sysctl.c
> +++ b/kernel/sysctl.c
> @@ -1701,6 +1701,13 @@ void register_sysctl_root(struct ctl_table_root *root)
> spin_unlock(&sysctl_lock);
> }
>
> +void unregister_sysctl_root(struct ctl_table_root *root)
> +{
> + spin_lock(&sysctl_lock);
> + list_del(&root->root_list);
> + spin_unlock(&sysctl_lock);
> +}
> +

This requires the addition of a declaration to include/linux/sysctl.h.

Once that is done and review is complete, I'd suggest that these two
patches be joined into a single patch, and that patch become part of
whatever patch series it is which needs them.

Subject: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by Stanislav Kinsbursky on Tue, 13 Dec 2011 09:02:52 GMT
View Forum Message <> Reply to Message

> On Mon, 12 Dec 2011 21:50:00 +0300
> Stanislav Kinsbursky<skinsbursky@parallels.com> wrote:
>
>> This routine is required for SUNRPC sysctl's, which are going to be allocated,
>> processed and destroyed per network namespace context.
>> IOW, new sysctl root will be registered on network namespace creation and
>> thus have to unregistered before network namespace destruction.
>>
>

> It's a bit suspicious that such a mature subsystem as sysctl newly
> needs its internals exported like this. Either a) the net namespaces
> work is doing something which hasn't been done before or b) it is doing
> something wrong.
>
> So, please explain further so we can confirm that it is a) and not b).
>

Hello, Andrew.
The goal is to provide an ability to control and modify data by sysctl's in
network namespace context. This is done by "net" sysctl's.
But there are two more issues to solve:
1) Sysctl's have to be in /proc/sys/sunrpc
2) Sysctl's content should be accessible from creator's network context (not
current user ones's).

>> --- a/kernel/sysctl.c
>> +++ b/kernel/sysctl.c
>> @@ -1701,6 +1701,13 @@ void register_sysctl_root(struct ctl_table_root *root)
>>   spin_unlock(&sysctl_lock);
>>   }
>>
>> +void unregister_sysctl_root(struct ctl_table_root *root)
>> +{
>> + spin_lock(&sysctl_lock);
>> + list_del(&root->root_list);
>> + spin_unlock(&sysctl_lock);
>> +}
>> +
>
> This requires the addition of a declaration to include/linux/sysctl.h.
>
> Once that is done and review is complete, I'd suggest that these two
> patches be joined into a single patch, and that patch become part of
> whatever patch series it is which needs them.
>

Ok, I'll do so.

--
Best regards,
Stanislav Kinsbursky

Subject: Re: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by Glauber Costa on Tue, 13 Dec 2011 09:13:56 GMT
View Forum Message <> Reply to Message

On 12/13/2011 01:02 PM, Stanislav Kinsbursky wrote:

>> On Mon, 12 Dec 2011 21:50:00 +0300
>> Stanislav Kinsbursky<skinsbursky@parallels.com> wrote:
>>
>>> This routine is required for SUNRPC sysctl's, which are going to be
>>> allocated,
>>> processed and destroyed per network namespace context.
>>> IOW, new sysctl root will be registered on network namespace creation
>>> and
>>> thus have to unregistered before network namespace destruction.
>>>
>>
>> It's a bit suspicious that such a mature subsystem as sysctl newly
>> needs its internals exported like this. Either a) the net namespaces
>> work is doing something which hasn't been done before or b) it is doing
>> something wrong.
>>
>> So, please explain further so we can confirm that it is a) and not b).
>>
>
> Hello, Andrew.
> The goal is to provide an ability to control and modify data by sysctl's
> in network namespace context. This is done by "net" sysctl's.
> But there are two more issues to solve:
> 1) Sysctl's have to be in /proc/sys/sunrpc
> 2) Sysctl's content should be accessible from creator's network context
> (not current user ones's).
>

Have you taken a look at how it is done at net/ipv4/sysctl_tcp_ipv4.c ,
for instance?

It manages to handle a per-net sysctl table without touching a single
bit at the kernel's core sysctl routines. Not entirely sure if it would
fit your use case, but maybe it is worth taking a look.

That file achieves both 1) and 2) that you described...

Subject: Re: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by Kinsbursky Stanislav on Tue, 13 Dec 2011 10:03:50 GMT
View Forum Message <> Reply to Message

> On 12/13/2011 01:02 PM, Stanislav Kinsbursky wrote:

>>> On Mon, 12 Dec 2011 21:50:00 +0300

>>> Stanislav Kinsbursky<skinsbursky@parallels.com> wrote:
>>>
>>>> This routine is required for SUNRPC sysctl's, which are going to be
>>>> allocated,
>>>> processed and destroyed per network namespace context.
>>>> IOW, new sysctl root will be registered on network namespace creation
>>>> and
>>>> thus have to unregistered before network namespace destruction.
>>>>
>>> It's a bit suspicious that such a mature subsystem as sysctl newly
>>> needs its internals exported like this. Either a) the net namespaces
>>> work is doing something which hasn't been done before or b) it is doing
>>> something wrong.
>>>
>>> So, please explain further so we can confirm that it is a) and not b).
>>>
>> Hello, Andrew.
>> The goal is to provide an ability to control and modify data by sysctl's
>> in network namespace context. This is done by "net" sysctl's.
>> But there are two more issues to solve:
>> 1) Sysctl's have to be in /proc/sys/sunrpc
>> 2) Sysctl's content should be accessible from creator's network context
>> (not current user ones's).
>>
> Have you taken a look at how it is done at net/ipv4/sysctl_tcp_ipv4.c ,
> for instance?

I don't have this file.
Probably you are talking about net/ipv4/sysctl_net_ipv4.c, don't you?

> It manages to handle a per-net sysctl table without touching a single
> bit at the kernel's core sysctl routines. Not entirely sure if it would
> fit your use case, but maybe it is worth taking a look.
>
> That file achieves both 1) and 2) that you described...
>
>
--
Best regards,
Stanislav Kinsbursky

Subject: Re: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by Glauber Costa on Tue, 13 Dec 2011 10:04:08 GMT
View Forum Message <> Reply to Message

On 12/13/2011 02:03 PM, Kinsbursky Stanislav wrote:

>> On 12/13/2011 01:02 PM, Stanislav Kinsbursky wrote:

>>>> On Mon, 12 Dec 2011 21:50:00 +0300
>>>> Stanislav Kinsbursky<skinsbursky@parallels.com> wrote:
>>>>
>>>>> This routine is required for SUNRPC sysctl's, which are going to be
>>>>> allocated,
>>>>> processed and destroyed per network namespace context.
>>>>> IOW, new sysctl root will be registered on network namespace creation
>>>>> and
>>>>> thus have to unregistered before network namespace destruction.
>>>>>
>>>> It's a bit suspicious that such a mature subsystem as sysctl newly
>>>> needs its internals exported like this. Either a) the net namespaces
>>>> work is doing something which hasn't been done before or b) it is doing
>>>> something wrong.
>>>>
>>>> So, please explain further so we can confirm that it is a) and not b).
>>>>
>>> Hello, Andrew.
>>> The goal is to provide an ability to control and modify data by sysctl's
>>> in network namespace context. This is done by "net" sysctl's.
>>> But there are two more issues to solve:
>>> 1) Sysctl's have to be in /proc/sys/sunrpc
>>> 2) Sysctl's content should be accessible from creator's network context
>>> (not current user ones's).
>>>
>> Have you taken a look at how it is done at net/ipv4/sysctl_tcp_ipv4.c ,
>> for instance?
>
> I don't have this file.
> Probably you are talking about net/ipv4/sysctl_net_ipv4.c, don't you?

Yeah, my bad.

>> It manages to handle a per-net sysctl table without touching a single
>> bit at the kernel's core sysctl routines. Not entirely sure if it would
>> fit your use case, but maybe it is worth taking a look.
>>
>> That file achieves both 1) and 2) that you described...
>>
>>

Subject: Re: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by Kinsbursky Stanislav on Tue, 13 Dec 2011 10:15:30 GMT
View Forum Message <> Reply to Message

> On 12/13/2011 02:03 PM, Kinsbursky Stanislav wrote:

>>> On 12/13/2011 01:02 PM, Stanislav Kinsbursky wrote:

>>>>> On Mon, 12 Dec 2011 21:50:00 +0300
>>>>> Stanislav Kinsbursky<skinsbursky@parallels.com> wrote:
>>>>>
>>>>>> This routine is required for SUNRPC sysctl's, which are going to be
>>>>>> allocated,
>>>>>> processed and destroyed per network namespace context.
>>>>>> IOW, new sysctl root will be registered on network namespace creation
>>>>>> and
>>>>>> thus have to unregistered before network namespace destruction.
>>>>>>
>>>>> It's a bit suspicious that such a mature subsystem as sysctl newly
>>>>> needs its internals exported like this. Either a) the net namespaces
>>>>> work is doing something which hasn't been done before or b) it is doing
>>>>> something wrong.
>>>>>
>>>>> So, please explain further so we can confirm that it is a) and not b).
>>>>>
>>>> Hello, Andrew.
>>>> The goal is to provide an ability to control and modify data by sysctl's
>>>> in network namespace context. This is done by "net" sysctl's.
>>>> But there are two more issues to solve:
>>>> 1) Sysctl's have to be in /proc/sys/sunrpc
>>>> 2) Sysctl's content should be accessible from creator's network context
>>>> (not current user ones's).
>>>>
>>> Have you taken a look at how it is done at net/ipv4/sysctl_tcp_ipv4.c ,
>>> for instance?
>> I don't have this file.
>> Probably you are talking about net/ipv4/sysctl_net_ipv4.c, don't you?
> Yeah, my bad.
>

Sorry, man, but this is what I was talking in the first sentence of my answer to
Andrew. And this solution doesn't suits me because both issues stays unsolved:
1) sysctl's in net/ipv4/sysctl_net_ipv4.c will be created in "/proc/sys/net/"
directory, but I need "/proc/sys/".
2) net sysctl's just gives you an ability to create sysctl' dentries per network
namespace context. But data pointer will be the same in case of this dentry was
created for all network namespaces.

>>> It manages to handle a per-net sysctl table without touching a single
>>> bit at the kernel's core sysctl routines. Not entirely sure if it would
>>> fit your use case, but maybe it is worth taking a look.

>>>
>>> That file achieves both 1) and 2) that you described...
>>>
>>>
--
Best regards,
Stanislav Kinsbursky

Subject: Re: [PATCH 1/2] SYSCTL: root unregister routine introduced
Posted by ebiederm on Sat, 17 Dec 2011 22:30:55 GMT
View Forum Message <> Reply to Message

Stanislav Kinsbursky <skinsbursky@parallels.com> writes:

>> On Mon, 12 Dec 2011 21:50:00 +0300
>> Stanislav Kinsbursky<skinsbursky@parallels.com> wrote:
>>
>>> This routine is required for SUNRPC sysctl's, which are going to be allocated,
>>> processed and destroyed per network namespace context.
>>> IOW, new sysctl root will be registered on network namespace creation and
>>> thus have to unregistered before network namespace destruction.
>>>
>>
>> It's a bit suspicious that such a mature subsystem as sysctl newly
>> needs its internals exported like this. Either a) the net namespaces
>> work is doing something which hasn't been done before or b) it is doing
>> something wrong.
>>
>> So, please explain further so we can confirm that it is a) and not b).
>>
>
> Hello, Andrew.
> The goal is to provide an ability to control and modify data by sysctl's in
> network namespace context. This is done by "net" sysctl's.
> But there are two more issues to solve:
> 1) Sysctl's have to be in /proc/sys/sunrpc

The sysctl root has nothing to with what directory the files show up in,
so this should not be an issue.

> 2) Sysctl's content should be accessible from creator's network context (not
> current user ones's).

Making the sunrpc sysctls per network namespace would seem to address
this.   I don't see why you would need a new root to handle this case.

Eric