

---

Subject: [Announce] Checkpoint-restore tool v0.1

Posted by [Pavel Emelyanov](#) on Mon, 23 Jul 2012 08:22:33 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Hi!

I'm glad to announce the first release of the checkpoint-restore tool.

This project is an attempt to implement the checkpoint-restore functionality for processes and containers without driving too much code into the kernel tree, but putting there "various oddball helper code" instead.

The tool can already be used for checkpointing and restoring various individual applications. And the greatest thing about this so far is that most of the below functionality has the required kernel support in the recently released v3.5!

So, we support now

- \* x86\_64 architecture
- \* process' linkage
- \* process groups and sessions (without ttys though :\ )
- \* memory mappings of any kind (shared, file, etc.)
- \* threads
- \* open files (shared between tasks and partially opened-and-unlinked)
- \* pipes and fifos with data
- \* unix sockets with packet queues contents
- \* TCP and UDP sockets (TCP connections support exists, but needs polishing)
- \* inotify, eventpoll and eventfd
- \* tasks' sigactions setup, credentials and itimers
- \* IPC, mount and PID namespaces

Though namespaces support is in there, we do not yet support an LXC container c/r, but we're close to it :)

I'd like to thank everyone who took part in new kernel APIs discussions, the feedback was great! Special thanks goes to Linus for letting the kernel parts in early, instead of making them sit out of tree till becoming stable enough.

Tarball with the tool sources is at

<http://download.openvz.org/criu/crtools-0.1.tar.bz2>

The git repo is at

<http://git.criu.org/>

And some sort of docs growing at

<http://criu.org/>

There are still things for which we don't have the kernel support merged (SysVIPC and various anon file descriptors, i.e. inotify, eventpoll, eventfd) yet. We have the kernel branch with the stuff applied available at

<https://github.com/cyrillos/linux-2.6.git>

Thanks,  
Pavel

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [Pavel Emelyanov](#) on Mon, 23 Jul 2012 08:53:42 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Fixing the bogus containers@ ml and adding cgroups@ one. Sorry :(

== Original message ==

Hi!

I'm glad to announce the first release of the checkpoint-restore tool.

This project is an attempt to implement the checkpoint-restore functionality for processes and containers without driving too much code into the kernel tree, but putting there "various oddball helper code" instead.

The tool can already be used for checkpointing and restoring various individual applications. And the greatest thing about this so far is that most of the below functionality has the required kernel support in the recently released v3.5!

So, we support now

- \* x86\_64 architecture
- \* process' linkage
- \* process groups and sessions (without ttys though :\\ )
- \* memory mappings of any kind (shared, file, etc.)
- \* threads
- \* open files (shared between tasks and partially opened-and-unlinked)
- \* pipes and fifos with data
- \* unix sockets with packet queues contents

- \* TCP and UDP sockets (TCP connections support exists, but needs polishing)
- \* inotify, eventpoll and eventfd
- \* tasks' sigactions setup, credentials and timers
- \* IPC, mount and PID namespaces

Though namespaces support is in there, we do not yet support an LXC container c/r, but we're close to it :)

I'd like to thank everyone who took part in new kernel APIs discussions, the feedback was great! Special thanks goes to Linus for letting the kernel parts in early, instead of making them sit out of tree till becoming stable enough.

Tarball with the tool sources is at

<http://download.openvz.org/criu/crtools-0.1.tar.bz2>

The git repo is at

<http://git.criu.org/>

And some sort of docs growing at

<http://criu.org/>

There are still things for which we don't have the kernel support merged (SysVIPC and various anon file descriptors, i.e. inotify, eventpoll, eventfd) yet. We have the kernel branch with the stuff applied available at

<https://github.com/cyrillos/linux-2.6.git>

Thanks,  
Pavel

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [Serge E. Hallyn](#) on Mon, 23 Jul 2012 13:14:19 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Quoting Pavel Emelyanov (xemul@parallels.com):

> Fixing the bogus containers@ ml and adding cgroups@ one. Sorry :(  
>  
> == Original message ==  
>  
> Hi!

>  
> I'm glad to announce the first release of the checkpoint-restore tool.  
>  
>  
> This project is an attempt to implement the checkpoint-restore functionality  
> for processes and containers without driving too much code into the kernel tree,  
> but putting there "various oddball helper code" instead.  
>  
> The tool can already be used for checkpointing and restoring various individual  
> applications. And the greatest thing about this so far is that most of the below  
> functionality has the required kernel support in the recently released v3.5!  
>  
> So, we support now  
>  
> \* x86\_64 architecture  
> \* process' linkage  
> \* process groups and sessions (without ttys though :\\ )  
> \* memory mappings of any kind (shared, file, etc.)  
> \* threads  
> \* open files (shared between tasks and partially opened-and-unlinked)  
> \* pipes and fifos with data  
> \* unix sockets with packet queues contents  
> \* TCP and UDP sockets (TCP connections support exists, but needs polishing)  
> \* inotify, eventpoll and eventfd  
> \* tasks' sigactions setup, credentials and itimers  
> \* IPC, mount and PID namespaces  
>  
> Though namespaces support is in there, we do not yet support an LXC container c/r,  
> but we're close to it :)  
>  
>  
> I'd like to thank everyone who took part in new kernel APIs discussions, the  
> feedback was great! Special thanks goes to Linus for letting the kernel parts  
> in early, instead of making them sit out of tree till becoming stable enough.  
>  
>  
> Tarball with the tool sources is at  
>  
> <http://download.openvz.org/criu/crtools-0.1.tar.bz2>  
>  
> The git repo is at  
>  
> <http://git.criu.org/>  
>  
> And some sort of docs growing at  
>  
> <http://criu.org/>  
>

>  
> There are still things for which we don't have the kernel support merged (SysVIPC  
> and various anon file descriptors, i.e. inotify, eventpoll, eventfd) yet. We have  
> the kernel branch with the stuff applied available at  
>  
> <https://github.com/cyrillos/linux-2.6.git>  
>  
>  
> Thanks,  
> Pavel

Awesome - thanks, Pavel, and congratulations :) I'll have to give it a spin as soon as I get a chance.

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [richard -rw- weinberg](#) on Mon, 30 Jul 2012 10:42:54 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

On Mon, Jul 23, 2012 at 10:53 AM, Pavel Emelyanov <xemul@parallels.com> wrote:  
> This project is an attempt to implement the checkpoint-restore functionality  
> for processes and containers without driving too much code into the kernel tree,  
> but putting there "various oddball helper code" instead.

Is CRIU able to checkpoint a process tree where one or more processes are traced using ptrace()?  
crttools seem to use ptrace() to execute parasite system calls.  
AFAIK this will only work if the target process is currently not ptraced() by another process.

--  
Thanks,  
//richard

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [Pavel Emelyanov](#) on Tue, 31 Jul 2012 08:53:04 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

On 07/30/2012 02:42 PM, richard -rw- weinberger wrote:  
> On Mon, Jul 23, 2012 at 10:53 AM, Pavel Emelyanov <xemul@parallels.com> wrote:  
>> This project is an attempt to implement the checkpoint-restore functionality  
>> for processes and containers without driving too much code into the kernel tree,  
>> but putting there "various oddball helper code" instead.  
>  
> Is CRIU able to checkpoint a process tree where one or more processes  
> are traced using ptrace()?

No it's not (currently).

> crtools seem to use ptrace() to execute parasite system calls.  
> AFAIK this will only work if the target process is currently not  
> ptraced() by another process.  
>

Yup, exactly.

Thanks,  
Pavel

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [richard -rw- weinberg](#) on Tue, 31 Jul 2012 08:57:58 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

On Tue, Jul 31, 2012 at 10:53 AM, Pavel Emelyanov <xemul@parallels.com> wrote:  
> On 07/30/2012 02:42 PM, richard -rw- weinberger wrote:  
>> On Mon, Jul 23, 2012 at 10:53 AM, Pavel Emelyanov <xemul@parallels.com> wrote:  
>>> This project is an attempt to implement the checkpoint-restore functionality  
>>> for processes and containers without driving too much code into the kernel tree,  
>>> but putting there "various oddball helper code" instead.  
>>  
>> Is CRIU able to checkpoint a process tree where one or more processes  
>> are traced using ptrace()?  
>  
> No it's not (currently).  
>  
>> crtools seem to use ptrace() to execute parasite system calls.  
>> AFAIK this will only work if the target process is currently not  
>> ptraced() by another process.  
>>  
>  
> Yup, exactly.

Dammit, I hoped we can get suspend/resume for UserModeLinux for free  
with CRIU. 8-)

--  
Thanks,  
//richard

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [cyrill](#) on Tue, 31 Jul 2012 09:00:39 GMT

---

On Tue, Jul 31, 2012 at 10:57:58AM +0200, richard -rw- weinberger wrote:

> >

> >> crtools seem to use ptrace() to execute parasite system calls.

> >> AFAIK this will only work if the target process is currently not

> >> ptraced() by another process.

> >>

> >

> > Yup, exactly.

>

> Dammit, I hoped we can get suspend/resume for UserModeLinux for free

> with CRIU. 8-)

We have plans to implement seizing on traced tasks but no dates at the moment ;)

Cyrill

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1

Posted by [Pavel Emelyanov](#) on Tue, 31 Jul 2012 09:07:31 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

On 07/31/2012 12:57 PM, richard -rw- weinberger wrote:

> On Tue, Jul 31, 2012 at 10:53 AM, Pavel Emelyanov <xemul@parallels.com> wrote:

>> On 07/30/2012 02:42 PM, richard -rw- weinberger wrote:

>>> On Mon, Jul 23, 2012 at 10:53 AM, Pavel Emelyanov <xemul@parallels.com> wrote:

>>>> This project is an attempt to implement the checkpoint-restore functionality

>>>> for processes and containers without driving too much code into the kernel tree,

>>>> but putting there "various oddball helper code" instead.

>>>

>>> Is CRIU able to checkpoint a process tree where one or more processes

>>> are traced using ptrace()?

>>

>> No it's not (currently).

>>

>>> crtools seem to use ptrace() to execute parasite system calls.

>>> AFAIK this will only work if the target process is currently not

>>> ptraced() by another process.

>>>

>>

>> Yup, exactly.

>

> Dammit, I hoped we can get suspend/resume for UserModeLinux for free

> with CRIU. 8-)

Sorry :( However, if you can help with patches on that issue that would be highly appreciated :)

Thanks,  
Pavel

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [richard -rw- weinberg](#) on Tue, 31 Jul 2012 09:29:25 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

On Tue, Jul 31, 2012 at 11:07 AM, Pavel Emelyanov <xemul@parallels.com> wrote:  
> Sorry :( However, if you can help with patches on that issue that would be  
> highly appreciated :)  
>

Yeah, but I fear it's not that easy.  
We'd have to change crtools to work without ptrace().

--  
Thanks,  
//richard

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [Pavel Emelyanov](#) on Tue, 31 Jul 2012 09:54:09 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

On 07/31/2012 01:29 PM, richard -rw- weinberger wrote:  
> On Tue, Jul 31, 2012 at 11:07 AM, Pavel Emelyanov <xemul@parallels.com> wrote:  
>> Sorry :( However, if you can help with patches on that issue that would be  
>> highly appreciated :)  
>>  
>  
> Yeah, but I fear it's not that easy.  
> We'd have to change crtools to work without ptrace().

Well, this is hard. Using ptrace saved us from having many special-purpose APIs for dumping various stuff (there will be an article about it). Thus I don't know which way is simpler -- stop using ptrace or teach ptrace to allow several tracers to attach to one task %)

Thanks,  
Pavel

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [richard -rw- weinberg](#) on Tue, 31 Jul 2012 10:08:22 GMT

---



On Tue, Jul 31, 2012 at 11:54 AM, Pavel Emelyanov <xemul@parallels.com> wrote:

>> Yeah, but I fear it's not that easy.

>> We'd have to change crtools to work without ptrace().

>

> Well, this is hard. Using ptrace saved us from having many special-purpose

> APIs for dumping various stuff (there will be an article about it). Thus I

> don't know which way is simpler -- stop using ptrace or teach ptrace to allow

> several tracers to attach to one task %)

Allowing multiple tracers in a safe way is IMHO even more harder.

BTW: While reading prctl\_set\_mm() I noticed two things.

1. Why isn't the return value of find\_vma() verified?

It looks like one can set an addr which does not belong to any vma.

2. What will happen if addr is not page aligned?

--

Thanks,

//richard

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1

Posted by [cyrill](#) on Tue, 31 Jul 2012 10:16:03 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

On Tue, Jul 31, 2012 at 12:08:22PM +0200, richard -rw- weinberger wrote:

> On Tue, Jul 31, 2012 at 11:54 AM, Pavel Emelyanov <xemul@parallels.com> wrote:

> >> Yeah, but I fear it's not that easy.

> >> We'd have to change crtools to work without ptrace().

> >

> > Well, this is hard. Using ptrace saved us from having many special-purpose

> > APIs for dumping various stuff (there will be an article about it). Thus I

> > don't know which way is simpler -- stop using ptrace or teach ptrace to allow

> > several tracers to attach to one task %)

>

> Allowing multiple tracers in a safe way is IMHO even more harder.

>

> BTW: While reading prctl\_set\_mm() I noticed two things.

> 1. Why isn't the return value of find\_vma() verified?

prctl\_set\_mm

vma = find\_vma(mm, addr);

...

if (!vma) {

error = -EFAULT;

goto out;

}

these values are used in procfs statistics only. So I don't get which verify you mean here.

- > It looks like one can set an addr which does not belong to any vma.
- > 2. What will happen if addr is not page aligned?

you'll have a bit weird output in procfs stat (probably I should add page alignment check here).

Cyrill

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1  
Posted by [richard -rw- weinberg](#) on Tue, 31 Jul 2012 10:21:58 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

On Tue, Jul 31, 2012 at 12:16 PM, Cyrill Gorcunov <[gorcunov@openvz.org](mailto:gorcunov@openvz.org)> wrote:  
> On Tue, Jul 31, 2012 at 12:08:22PM +0200, richard -rw- weinberger wrote:  
>> On Tue, Jul 31, 2012 at 11:54 AM, Pavel Emelyanov <[xemul@parallels.com](mailto:xemul@parallels.com)> wrote:  
>> >> Yeah, but I fear it's not that easy.  
>> >> We'd have to change crtools to work without ptrace().  
>> >  
>> > Well, this is hard. Using ptrace saved us from having many special-purpose  
>> > APIs for dumping various stuff (there will be an article about it). Thus I  
>> > don't know which way is simpler -- stop using ptrace or teach ptrece to allow  
>> > several tracers to attach to one task %)  
>>  
>> Allowing multiple tracers in a safe way is IMHO even more harder.  
>>  
>> BTW: While reading prctl\_set\_mm() I noticed two things.  
>> 1. Why isn't the return value of find\_vma() verified?  
>  
> prctl\_set\_mm  
>     vma = find\_vma(mm, addr);  
>     ...  
>     if (!vma) {  
>         error = -EFAULT;  
>         goto out;  
>     }  
>  
> these values are used in procfs statistics only. So I don't get  
> which verify you mean here.

If I do PR\_SET\_MM\_START\_BRK the if(!vma) will never be executed because there a break in case PR\_SET\_MM\_START\_BRK.

--

Thanks,  
//richard

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.1

Posted by [cyrill](#) on Tue, 31 Jul 2012 10:30:07 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

On Tue, Jul 31, 2012 at 12:21:58PM +0200, richard -rw- weinberger wrote:

> On Tue, Jul 31, 2012 at 12:16 PM, Cyrill Gorcunov <gorcunov@openvz.org> wrote:

> > On Tue, Jul 31, 2012 at 12:08:22PM +0200, richard -rw- weinberger wrote:

> >> On Tue, Jul 31, 2012 at 11:54 AM, Pavel Emelyanov <xemul@parallels.com> wrote:

> >>> Yeah, but I fear it's not that easy.

> >>> We'd have to change crtools to work without ptrace().

> >>>

> >>> Well, this is hard. Using ptrace saved us from having many special-purpose

> >>> APIs for dumping various stuff (there will be an article about it). Thus I

> >>> don't know which way is simpler -- stop using ptrace or teach ptrace to allow

> >>> several tracers to attach to one task %)

> >>>

> >> Allowing multiple tracers in a safe way is IMHO even more harder.

> >>

> >> BTW: While reading prctl\_set\_mm() I noticed two things.

> >> 1. Why isn't the return value of find\_vma() verified?

> >

> > prctl\_set\_mm

> > vma = find\_vma(mm, addr);

> > ...

> > if (!vma) {

> > error = -EFAULT;

> > goto out;

> > }

> >

> > these values are used in procfs statistics only. So I don't get

> > which verify you mean here.

>

> If I do PR\_SET\_MM\_START\_BRK the if(!vma) will never be executed because

> there a break in case PR\_SET\_MM\_START\_BRK.

Yes, and this is done by purpose, since we need to setup \_completely\_  
new memory map on restore procedure.

There is a minimal check for value being sane

```
if (addr >= TASK_SIZE || addr < mmap_min_addr)
    return -EINVAL;
```

and the address belongs to mm::start\_data|end\_data area. But sure, better to add checks that at least code/data areas do exist, otherwise the proc output will not reflect the real state of memory maps.

Cyrill

---

---

Subject: [Announce] Checkpoint-restore tool v0.2  
Posted by [Pavel Emelyanov](#) on Thu, 20 Sep 2012 10:16:16 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Hi!

I'm happy to announce that the next big step is done by the CRIU project.

The major achievement so far -- the tool can now dump and restore a simple LXC container. Another notable feature is the initial support for PTYs which is enough to play with c/r of a container with active ssh sessions from outside, or with a screen app with some simple stuff running inside.

A HOWTO describing an LXC container dump and restore is at <http://criu.org/LXC>

The v0.2 can be downloaded from

<http://download.openvz.org/criu/crtools-0.2.tar.bz2>

Other project resources are

documentation and news: <http://criu.org>  
mainling list: <https://openvz.org/mailman/listinfo/criu>  
git repo: <http://git.criu.org/?p=crtools.git>  
kernel with not-yet-merged patches: <https://github.com/cyrillos/linux-2.6.git>

Thanks,  
Pavel

---

---

Subject: Re: [Announce] Checkpoint-restore tool v0.2  
Posted by [Serge E. Hallyn](#) on Thu, 20 Sep 2012 13:05:51 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Congratulations, Pavel and criu team, that's awesome!

Quoting Pavel Emelyanov (xemul@parallels.com):  
> Hi!

>  
> I'm happy to announce that the next big step is done by the CRIU project.  
>  
> The major achievement so far -- the tool can now dump and restore a simple LXC  
> container. Another notable feature is the initial support for PTYs which is enough  
> to play with c/r of a container with active ssh sessions from outside, or with  
> a screen app with some simple stuff running inside.  
>  
> A HOWTO describing an LXC container dump and restore is at <http://criu.org/LXC>  
>  
>  
> The v0.2 can be downloaded from  
>  
> <http://download.openvz.org/criu/crtools-0.2.tar.bz2>  
>  
>  
> Other project resources are  
>  
> documentation and news: <http://criu.org>  
> mainling list: <https://openvz.org/mailman/listinfo/criu>  
> git repo: <http://git.criu.org/?p=crtools.git>  
> kernel with not-yet-merged patches: <https://github.com/cyrillos/linux-2.6.git>  
>  
> Thanks,  
> Pavel  
> \_\_\_\_\_  
> Containers mailing list  
> [Containers@lists.linux-foundation.org](mailto:Containers@lists.linux-foundation.org)  
> <https://lists.linuxfoundation.org/mailman/listinfo/container> s

---