Contribution ID: **64** Type: **not specified**

CRIU mounts migration: problems and solutions

Monday, August 24, 2020 7:25 AM (20 minutes)

OpenVZ and Virtuozzo containers use CRIU as the core technology for container migration in production. And Virtuozzo containers are slightly different thing to what most people would imagine containers today. They are “system containers” which is the one with full systemd inside, the one you would enter via ssh, the one which is an analogy to a virtual machine where the user gets root access inside and can do almost everything like on the hardware node with Linux.

This difference between application and system containers brings a lot of complex problems when it comes to container migration of the system containers. Lets consider the mounts problem. The user inside a container can explicitly or implicitly (by systemd, docker or some other means) create multiple different mount namespaces and mounts in them. And if we migrate the container, the user inside does not expect their mounts to change. So we need to checkpoint and restore them.

In this talk I would share main problems I’ve faced when I tried to improve the correctness of our current mount restore algorithm in CRIU and I would show new “mounts-v2” algorithm which tries to cover much more cases than the previous one. To achieve this we need at least one kernel patch [1] and maybe more to come.

I would like to restart the discussion on bind mounts across namespaces at the point it had stopped a while ago. I hope we can reach a consensus about the kernel modifications required to solve the problem of checkpoint/restore of complex mounts. And I really hope for some useful advice on how to further improve the new algorithm.

[1] https://lore.kernel.org/lkml/1485214628-23812-1-git-send-email-avagin@openvz.org/

Here are links to mounts-v2 implementation in Virtuozzo criu:
- Main part: https://src.openvz.org/projects/OVZ/repos/criu/commits?until=v3.12.3.12
- Delayed proc part: https://src.openvz.org/projects/OVZ/repos/criu/commits?until=v3.12.5.13

I agree to abide by the anti-harassment policy

I agree

Primary author: TIKHOMIROV, Pavel

Presenter: TIKHOMIROV, Pavel

Session Classification: Containers and Checkpoint/Restore MC

Track Classification: Containers and Checkpoint/Restore MC