CRIU Support in Docker for Native Checkpoint and Restore

Linux Plumbers Conference
August 20, 2015

Saied Kazemi (saied@google.com)
Overview

● Brief history of Docker container checkpoint restore
  ○ Live demos
  ○ Will not cover CRIU usage or internals
  ○ Will mention some of the features added to support container C/R

● Project started in April 2014
  ○ Sep ‘14: External C/R at [Docker Meetup](https://www.dockercon.com) in Mountain View
  ○ Oct ‘14: Native C/R without network at [LPC](https://www.linuxfoundation.org/community/lpc)
  ○ Jan ‘15: Native C/R with network uploaded to [YouTube](https://www.youtube.com)
  ○ Jun ‘15: Quake demo at DockerCon
C/R Benefits

• On the Same Node
  ○ Stop and restart (even reboot) without losing container work
  ○ Shorten startup time of slow-start services
  ○ Use checkpoint “image” files for forensic debugging
  ○ Rollback
    ■ Tonic A new web based REPL for node.js. Checkpoints each step to fix mistakes by restoring from the most recent change instead of the beginning

• On Different Nodes
  ○ Migrate containers
Docker Client, Server, and Containers

$ docker run ...

# docker -d

client

server

Docker API

Global Namespace

Private Namespace

init

grandchild

Container 1

Container 2
External C/R

- Can I do the following?

  ```
  $ docker run <args> <cmd>
  $ ps -efl | grep <cmd>
  $ sudo criu dump <args> -t <pid>
  $ sudo criu restore <args>
  ```

- Let’s try...
External C/R Demo

Demo 1: Simple Shell Script
External C/R Status

- Support added to CRIU for Docker container C/R
  - ext-mount-map, manage-cgroups, ...
  - AUFS bug handling
  - See [here](#) for details

- Helper script `docker_cr.sh` in `contrib` subdirectory

- Demo’ed at [Docker Meetup](#) in September ‘14

- Good proof of concept but there are issues...
External C/R Issues

- **Wrong Container State**
  - After checkpoint, state is *Exited* (instead of Checkpointed)
  - After restore, state is still *Exited* (instead of Up)

- **Wrong Process Tree Ownership**
  - Restored process tree is a child of *init*, not the Docker daemon

- **Broken “Plumbing”**
  - Logging won’t work (*docker logs*)
  - Container control is lost (*docker stop, docker kill*, etc.)
Native C/R in Docker 1.5
Native C/R Demo

Demo 2: Nginx Streaming Video
Native C/R Status

- C/R support already merged into libcontainer and runC
  - runC: a lightweight universal container runtime
  - Quake demo (container migration) was done with runC

- Significant changes between Docker 1.5 and 1.8
  - Reached out to the community for help
  - 1.5 C/R code rebased to 1.8 by Ross Boucher
    - Network restore not fully functional

- Working on getting C/R support merged into 1.9 main branch
Want to Try?

- Get Docker C/R from Github
  - Build from source (EXPERIMENTAL)
  - Prebuilt binaries available on Github

- Get latest CRIU
  - Build from source
    - Fixes OverlayFS issues
  - Prebuilt binary available on Github and PPA
  - Requires *libprotobuf-c.so*
Filesystem Issues

- **AUFS**
  - Most widely used when Docker released
  - Has an issue, but CRIU can handle
    - Symlinks in `/proc/<pid>/map_files` point into branches

- **OverlayFS**
  - Merged into the kernel as of 3.18
  - Preferred choice
  - Has two issues, but CRIU can handle
    - `/proc/<pid>/fd`
    - `/proc/<pid>/fdinfo`
Useful Links

- Github
  - Docker 1.5 - [Saied Kazemi](#)
  - Docker 1.8 - [Ross Boucher](#)
  - Docker Repo - [Michael Crosby](#)

- CRIU PPA
  - Ubuntu Vivid kernel package (v3.19)
  - CRIU package (v1.6)
Thank You!