MRAS 项目技术调研报告

CRIU-Checkpoint/Restore in User-space

MRAS team

Distributed & Embedded System Lab

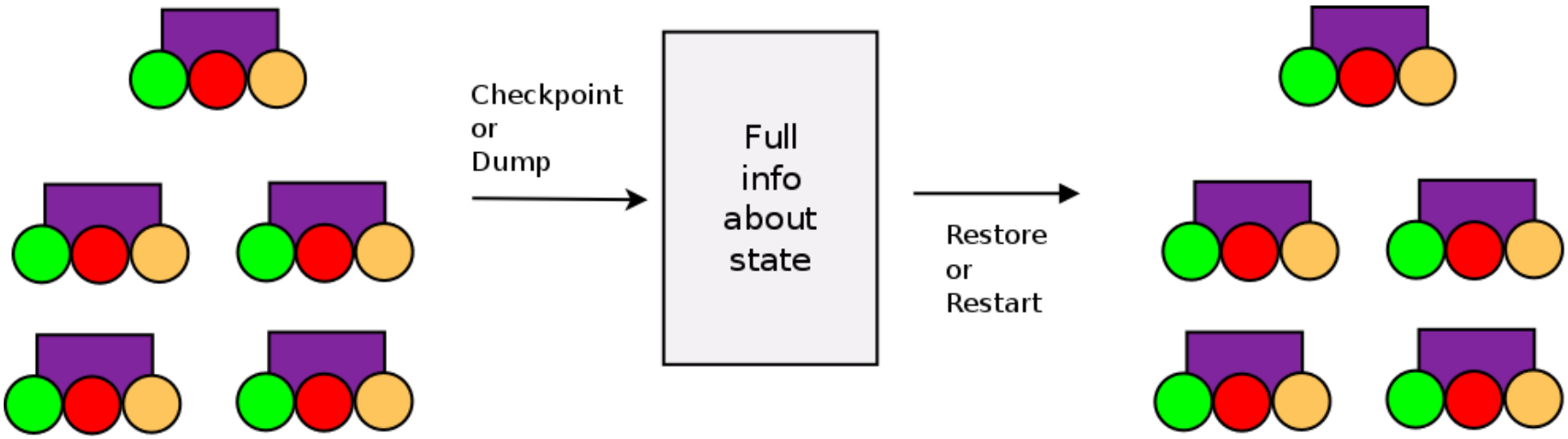
2015 年 4 月 8 日

Catalogue

- What's the CRIU
- The Development of C/R
- CRIU's scenarios
- CRIU's mechanism
- How is CRIU tested?
- What can change after CRIU?

What's the CRIU

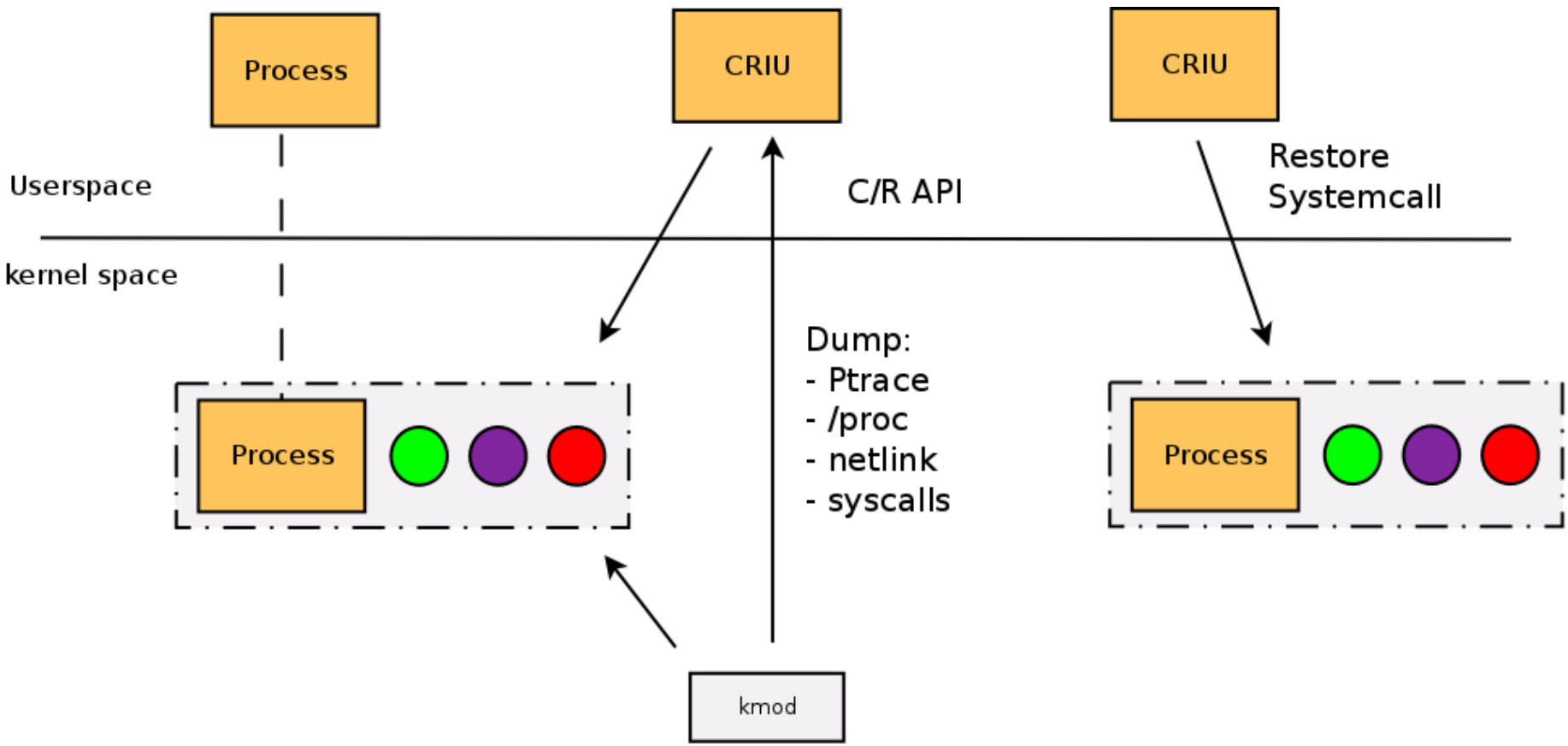
C/R is the ability to save states of processes and to restore them later



The Development of C/R

1. Berkeley Lab Checkpoint/Restart (BLCR) (2003)
   - Load a kernel module and link with a library
2. DMTCP: Distributed MultiThreaded CheckPointing (2004-2006)
   - Preload a library
3. OpenVZ (2005)
   - OpenVZ kernel
4. Linux Checkpoint/Restart by Oren Laadan (2008)
   - A non-mainline kernel
5. CRIU (2011)

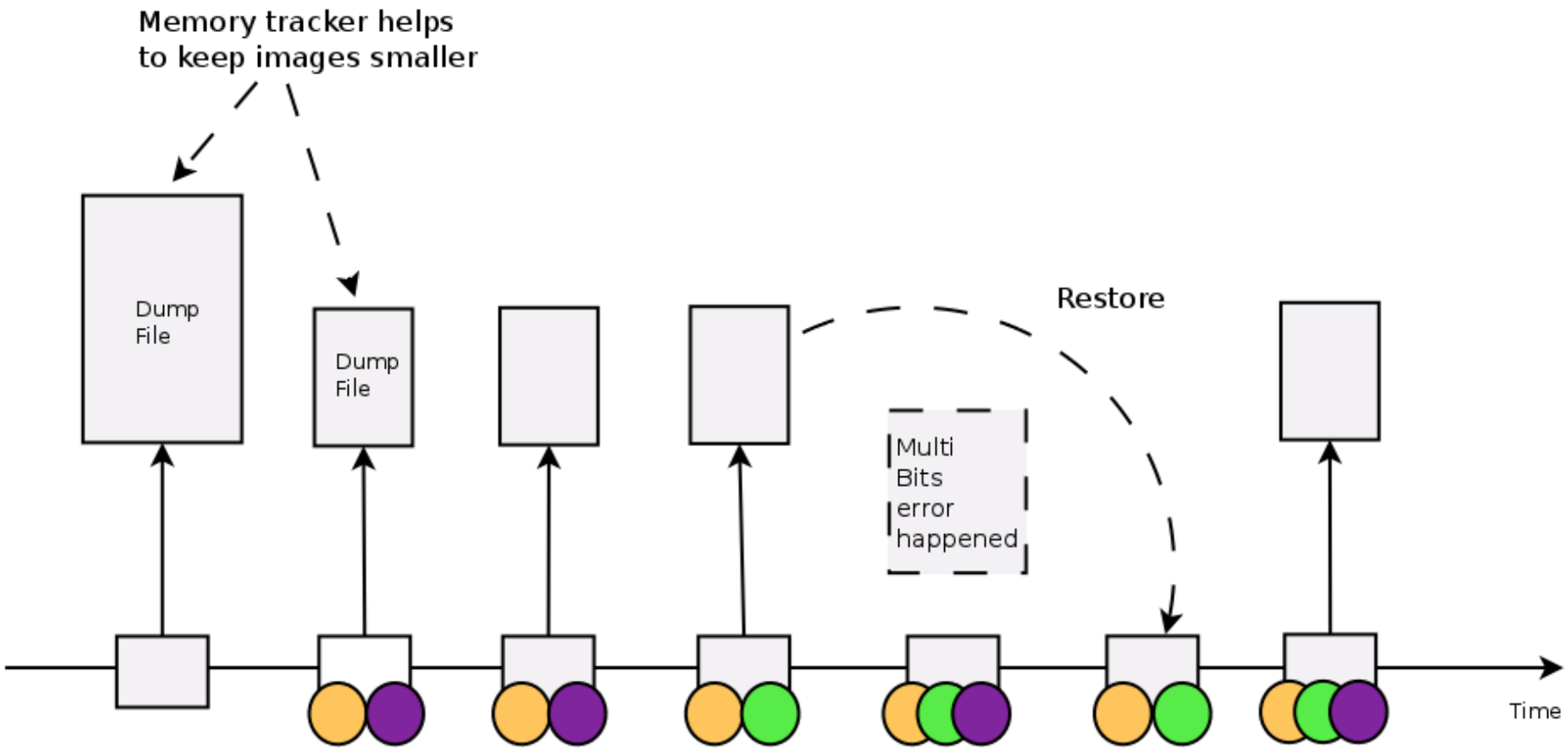
CRIU in userspace



CRIU's scenarios

1. Live migration
   - Useful in cluster
2. Kernel upgrade w/o reboot
3. Slow services startup
4. Periodic snapshots
5. Advanced debugging and testing

CRIU's scenarios



CRIU's mechanism - Dump

- Parasite code
  - Receive file descriptors
  - Dump memory content
  - Prctl(), sigaction, pending signals, timers, etc.
- Ptrace
  - freeze processes
  - Inject a parasite code
- Netlink
  - Get information about sockets, netns
- procfs

CRIU's mechanism - Restore

- Collect shared objects
- Restore name-space
- Create a process tree
  - Restore SID, PGID
  - Restore objects, which should be inherited
- Files, sockets, pipes, ...
- Restore per-task properties.
- Restore memory
- Call sigreturn

What are already supported?

- Process tree linkage
- Pending signals
- Multi-threaded apps
- Iterative snapshots
- All kinds of memory mappings
- VDSO
- Terminals, groups, sessions
- LXC,OpenVZ containers and docker
- Open files (shared and unlinked)
- Established TCP connections
- Pipes, Fifo-s, IPC, ...

What are already supported?

- Posix timers
- Unix sockets, Packet sockets
- Convert OpenVZ images
- Name-spaces (net, mount, ipc)
- Non-posix files (epoll, inotify)

Kernel impact

- 150+ patches merged
- 10 patches in flight
- 11 new features appeared
- 2 new features to come

Comparison to other CR projects

Please Check Comparison to other CR projects.pdf

How is CRIU tested?

- ZDTM – a set of unit-tests
- Real-life applications
  - Apache,Nginx,MySQL,MongoDB,Oracle*
  - Ssh/sshd, openvpn*, cron, sendmail - Make && gcc ,Java
  - Screen + bash, top, tcpdump, tar/bz2
  - LXC
  - VNC server + GUI application

What can change after CRIU

- Per-task statistics
- Namespaces' IDs
- Process start time
- Mount points IDs
- Sockets IDs
- VDSO

谢谢!

Q&A