

CRIU Restorer v2

Студент:

Егор Горбунов

Руководитель:

Евгений Баталов

27 октября 2016 г.

Что такое CRIU?

- ▶ Позволяет:
 - ▶ **Checkpoint**: останавливать процесс (дерево процессов) и сохранять его состояние на диск
 - ▶ **Restore**: восстанавливать процесс по сохранённому на диск состоянию
- ▶ Пользовательское приложение, такое же как `grep`, `cat` и т.д.
- ▶ Ядро Linux ≥ 3.11 (2013)

Ресурсы процесса для восстановления

- ▶ Идентификаторы: PID, PGID, SID
- ▶ Виртуальное адресное пространство
- ▶ Регистры
- ▶ Файловые дескрипторы
 - ▶ обычные файлы, сокеты, pipe's, inotify instances, eventfd...
- ▶ Linux namespaces
- ▶ Linux cgroups
- ▶ Обработчики сигналов
- ▶ Необработанные очереди сигналов
- ▶ File lock's
- ▶ Таймеры
- ▶ Потоки

CRIU Checkout

- ▶ Заморозка процесса
- ▶ Обход файловой системы `/proc/pid` для получения части информации о
- ▶ Инъекция паразитного кода
- ▶ Паразитный код: читает данные о регистрах, памяти, файловых дескрипторах
- ▶ Паразитный код отправляет данные в процесс `criu`.

CRIU Restore

- ▶ Считать состояние дерева процессов с диска
- ▶ Создать корень дерева и восстановить его namespaces
- ▶ Воссоздать дерево процессов
- ▶ Восстановить файловые дескрипторы и прочие разделяемые ресурсы
- ▶ Инъектировать PIE код последней стадии Restorera в восстанавливаемые...
- ▶ ...

Почему Restorer сложный?

- ▶ Код на языке C в стиле ядра Linux
- ▶ Код двух больших и независимых задачи (Dump, Restore) написан в одном месте (модули пересекаются)
- ▶ Между ресурсами для восстановления (некоторыми) существуют зависимости, что влияет на порядок выполнения операций, а это делает более трудным добавление поддержки новых ресурсов
- ▶ К этому всему добавляется многопоточная (процессная) архитектура CRIU

Другие Checkpoint/Restore проекты

Работают на неизменённом ядре Linux

- ▶ **BLCR**¹ (2003)
 - не поддерживает linux namespaces
 - требует подгрузки модуля к ядру
- ▶ **DMTCP**² (2004)
 - к целевому процессу с момента запуска должна быть подключена библиотека
 - не поддерживает linux namespaces
 - перехватывает часть glibc и системных вызовов

Требуют использования своего ядра:

- ▶ OpenVZ (2005)
- ▶ **Linux Checkpoint/Restart** (2008)

¹Berkeley Lab Checkpoint/Restart

²Distributed MultiThreaded CheckPointing