НИР
Распределенное блочное устройство

Николай Обедин, Никита Карташов
Руководитель: Кирилл Кринкин

26 мая 2014

План

- Что хотели
- История успеха
- Что получилось

Что хотели

- Что такое «распределенное устройство»?
- Три кита распределенных систем:
  - Согласованность (Consistency)
  - Доступность (Availability)
  - Надежность (Partition Tolerance)
- Но использовать вместе можно только два из них (CAP теорема)

Что хотели

- Мы сконцентрировались на доступности и надежности
- Надежность: помехоустойчивое кодирование (erasure coding)
- Доступность: тонкая настройка кеширования
- Области применения: высокопроизводительные вычисления, обработка больших мультимедиа-данных
- И обязательно open-source!

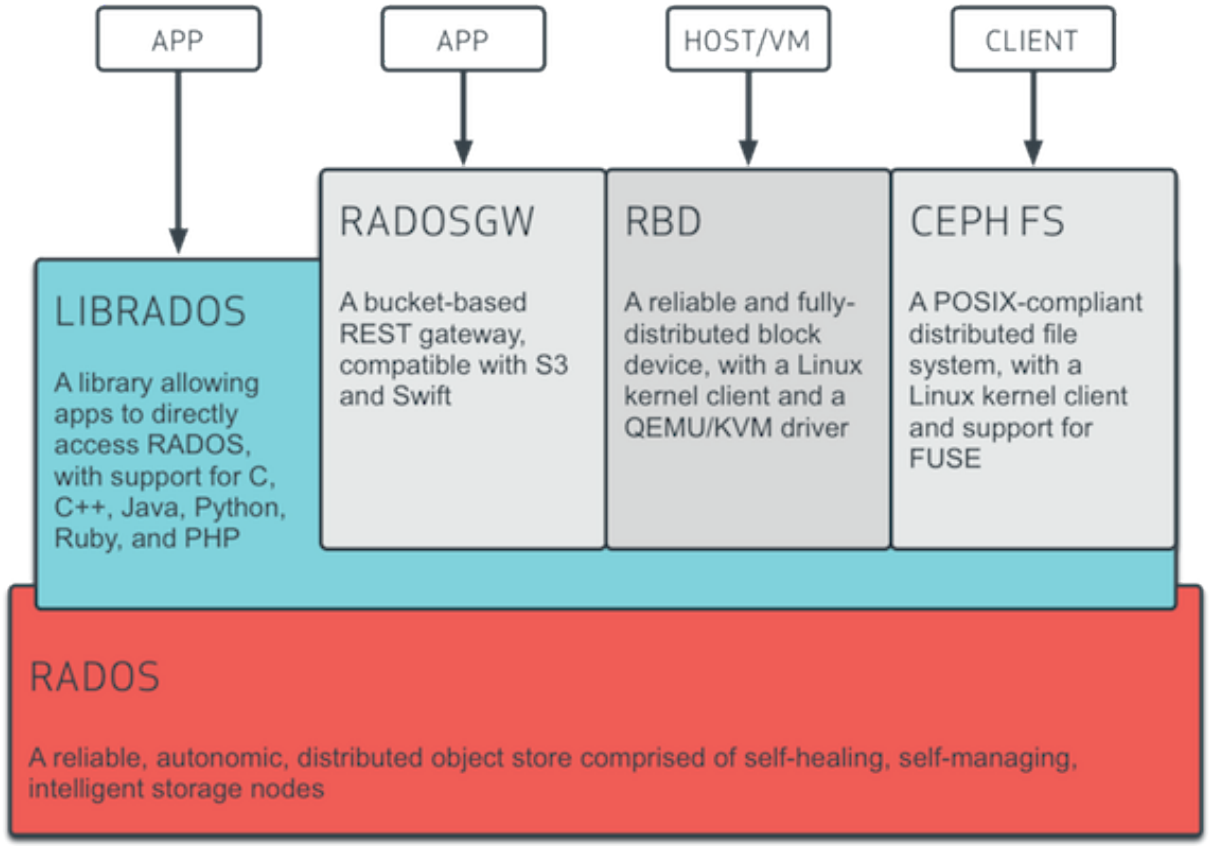
Архитектура проекта

Архитектура проекта

Поиск и анализ готовых решений

- OpenStack: нет erasure coding
- RING: закрытый исходный код
- CEPH: все хорошо (пока)

CEPH

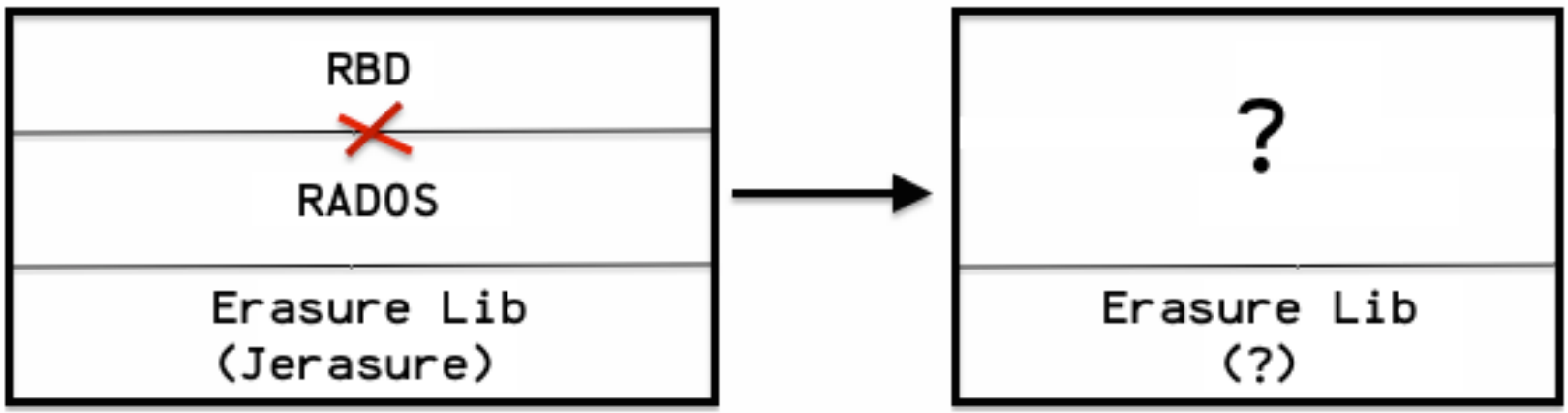


Архитектура CEPH

CEPH

- Что мы хотели от CEPH: erasure coding на блочном уровне
- В CEPH полностью поддерживается erasure coding на уровне объектов RADOS
- Однако, исследование исходного кода и тестирование на кластере показали, что реализация на уровне RBD находится в зачаточном состоянии

Смена планов



Помехоустойчивое кодирование

- Коды Рида-Соломона
- Параметры N и K позволяют выбирать степень избыточности
- Систематические – в закодированном тексте первые K блоков совпадают с исходным сообщением
- Для вычисления кодов требуются базовые операции над числами в $GF(2^8)$

Реализации

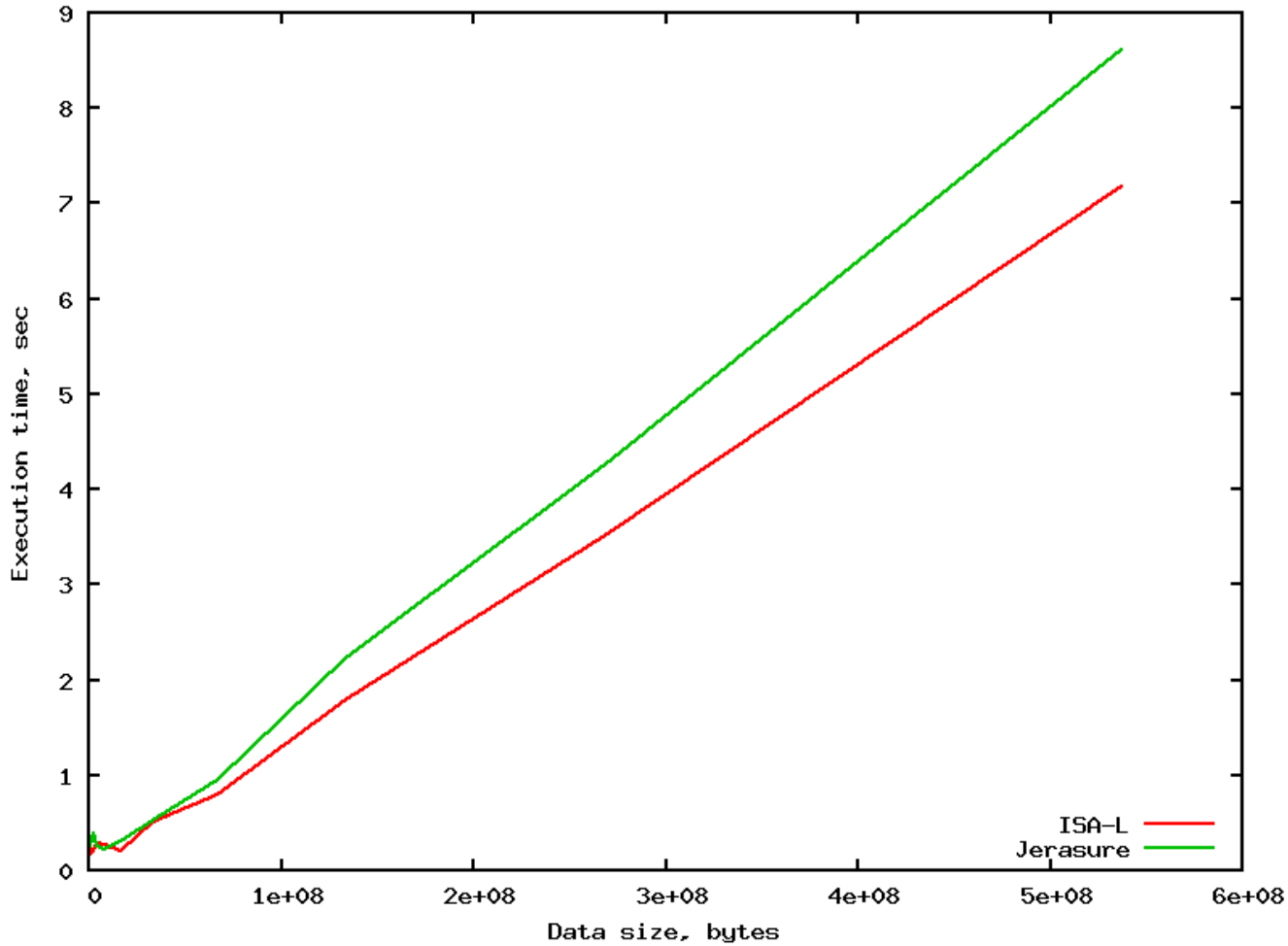
- Jerasure – используется в CEPH, реализует различные алгоритмы, в том числе и коды Рида-Соломона, используется во множестве приложений
- ISA-L – библиотека от компании Intel использующая для ускорения операций инструкции SSE4 и AVX, исходный код был открыт в 2013 году.
- Остальные реализации уступают в скорости Jerasure [1]

[1] https://www.usenix.org/legacy/events/fast09/tech/full_papers/plank/plank.pdf

Что получилось

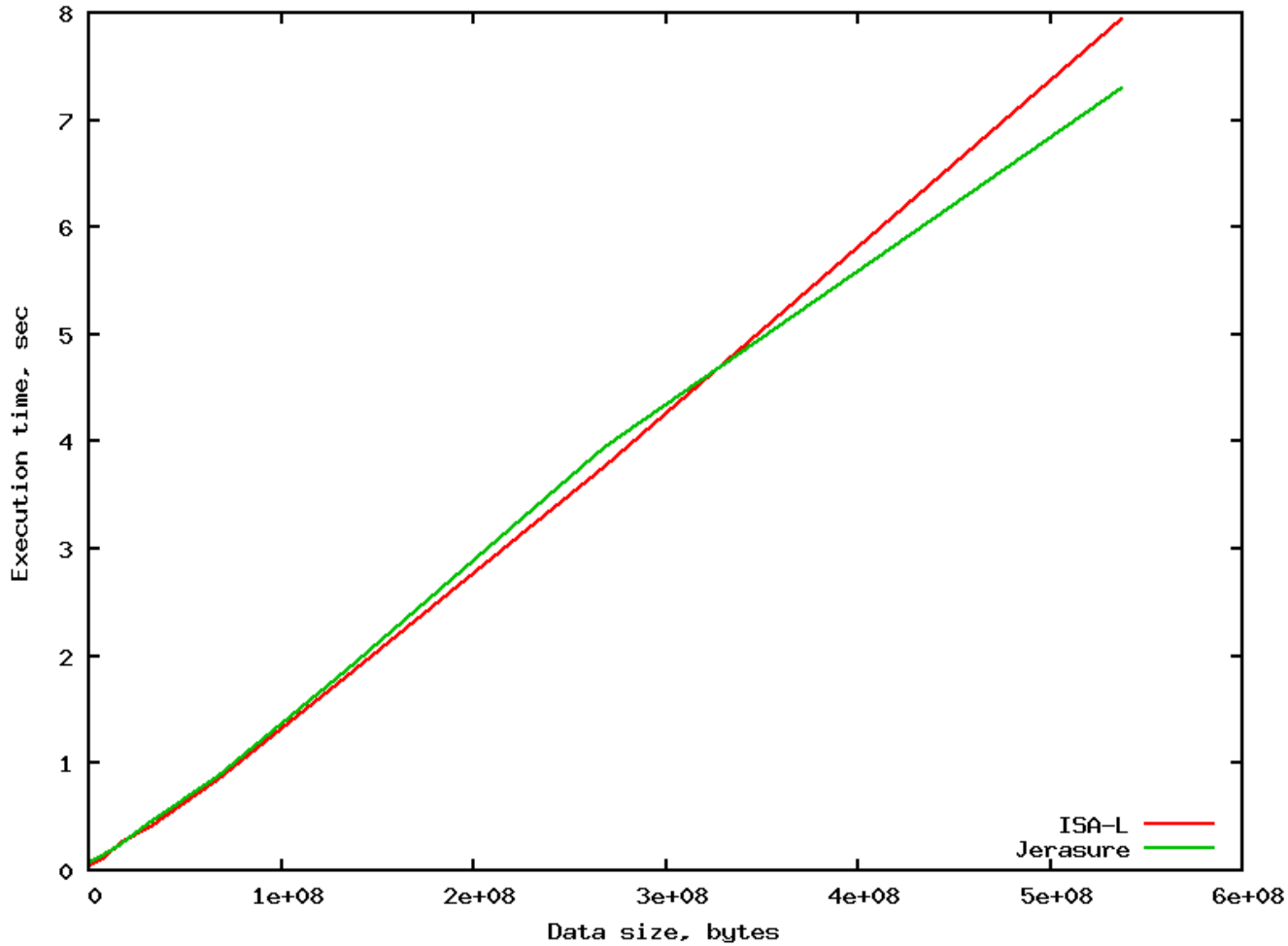
- Приложение кодирующее/декодирующее данные произвольного размера с заданными параметрами
- Реализовано поверх ISA-L и Jerasure
- Возможно использование в качестве компонента для распределенного блочного устройства

Сравнение



Кодирование

Сравнение



Восстановление с N-K случайно удаленными блоками

За этот НИР мы

- ✓ Узнали много нового про помехоустойчивое кодирование и распределенные системы хранения данных

За этот НИР мы

- ✓ Узнали много нового про помехоустойчивое кодирование и распределенные системы хранения данных
- ✓ Научились использовать open-source проекты со скудной документацией

За этот НИР мы

- ✓ Узнали много нового про помехоустойчивое кодирование и распределенные системы хранения данных
- ✓ Научились использовать open-source проекты со скудной документацией
- ✓ Научились делать open-source проекты со скудной документацией

Спасибо за внимание!