

Санкт-Петербургский Государственный Университет  
Кафедра системного программирования

# Разработка unikernel операционной системы на основе Embbox

Автор: Черепанов Алексей Олегович, 16.Б11-мм  
Научный руководитель: проф. каф. СП, д.ф.-м.н., проф. А.Н. Терехов  
Консультант: ассистент А.П. Козлов

Санкт-Петербург, 2020

# Unikernels

— это специализированные образы машин с единым адресным пространством, созданные с использованием библиотечных операционных систем, обычно запускаемые с помощью гипервизора

## Unikernels

- + Скорость, экономия ресурсов
- + Безопасность
- Сложная интеграция компонент

## Embox

- + ОС для встраиваемых систем
- + Гибкая конфигурируемость за счет встроенной Build-системы
- + POSIX-совместимость

# Постановка задачи

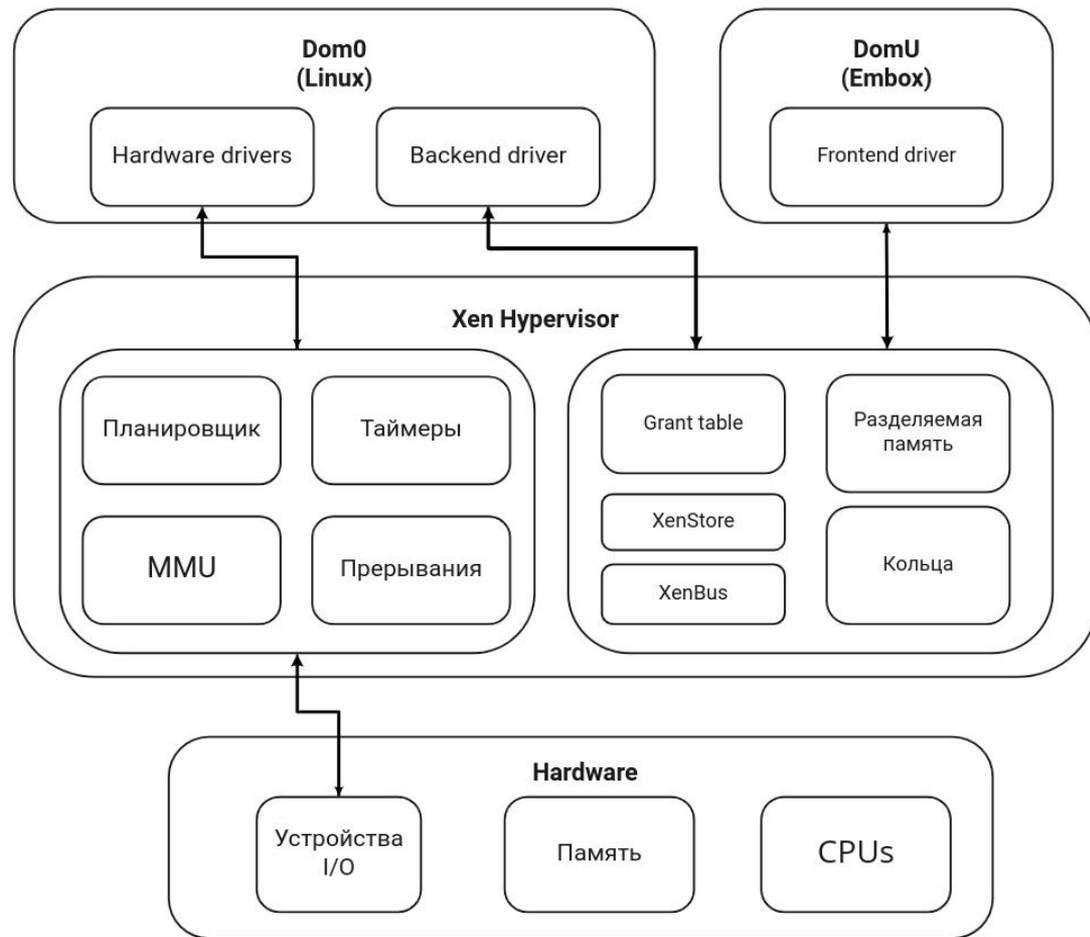
**Цель:** создать прототип unikernel системы на базе ОСРВ Embox для архитектуры гипервизора Xen.

## Задачи:

- Провести обзор предметной области, изучить архитектуру гипервизора Xen и ОСРВ Embox, существующих решений
- Реализовать и интегрировать механизм общей памяти памяти в ОСРВ Embox для поддержки архитектуры гипервизора Xen
- Реализовать и интегрировать сетевой интерфейс в Embox
- Оценить сетевую производительность Embox в качестве unikernel системы

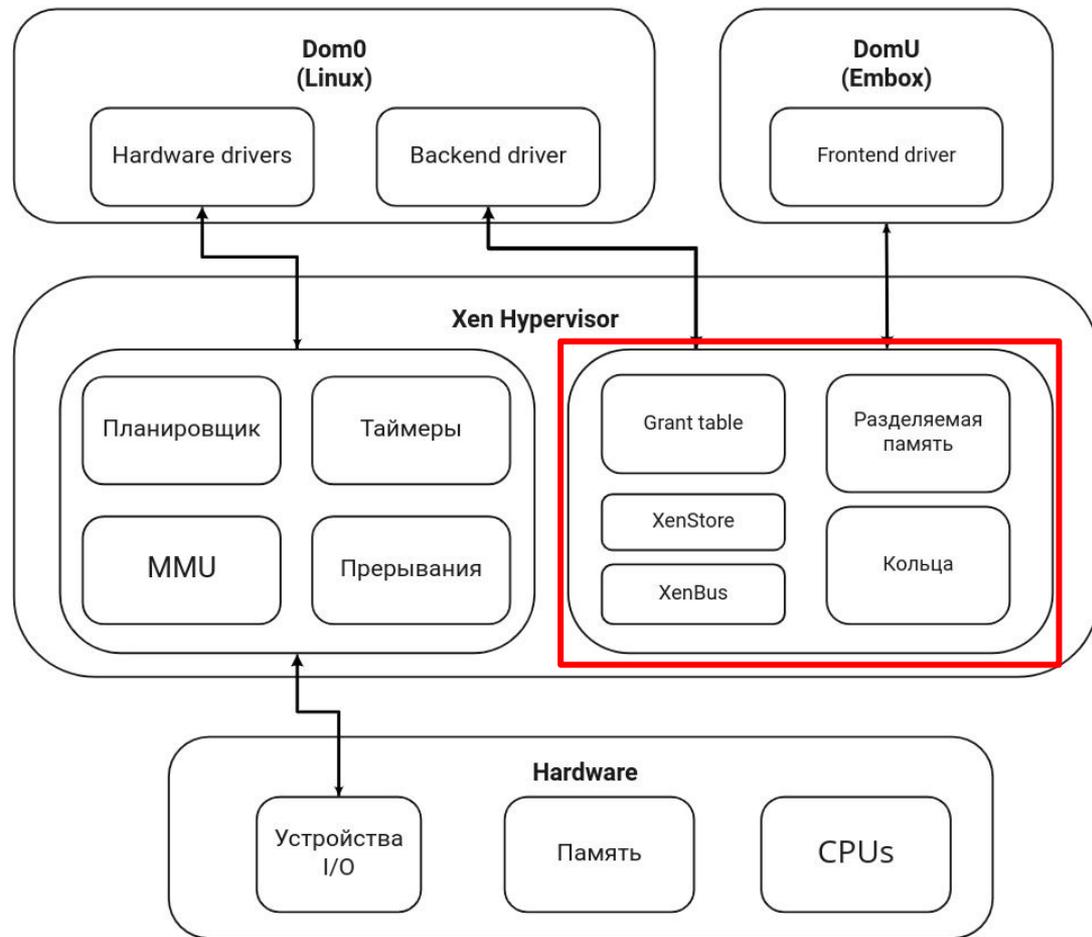
# Xen

- Кроссплатформенный гипервизор
- Концепция разделенных драйверов



# Embox

- + Загрузчик
- + Таймер
- + Вывод отладочной информации
- Разделяемая память
  - Grant table
- Драйвер сети
  - Кольца
  - XenStore/XenBus

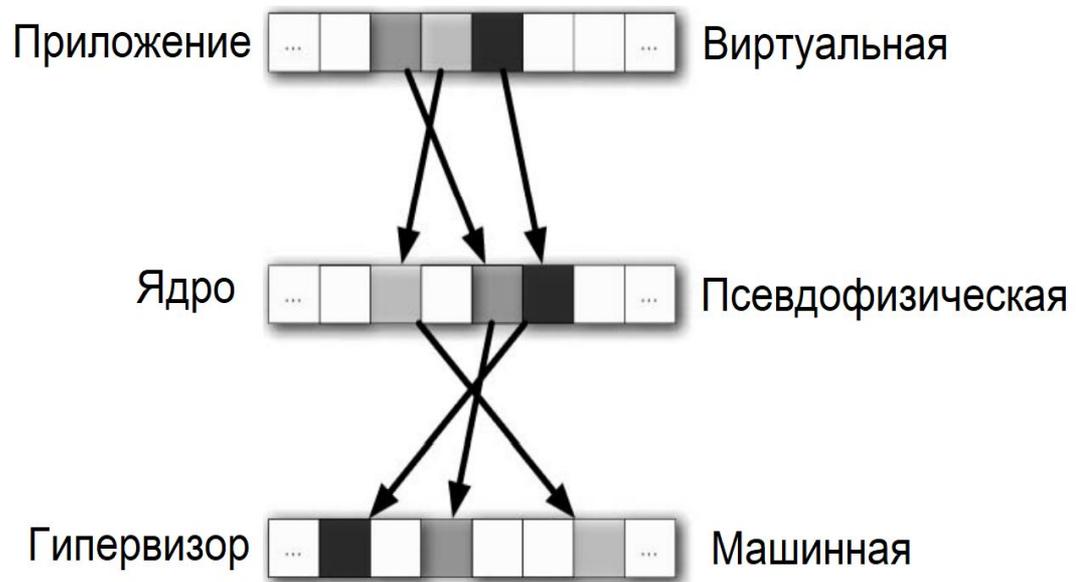


# Развёртывание окружения для разработки

- Изолированная среда разработки
  - Vagrant
  - VirtualBox
- Хост dom0
  - Ubuntu Bionic
  - Xen
- Гость domU
  - Embox
- Написаны Bash-скрипты для изменения и отладки бэкенд-драйвера

# Память

- Типы памяти в Xen:
  - Виртуальная
  - Псевдофизическая
  - Машинная
- Номер машинной страницы — номер страницы в реальном адресном пространстве гипервизора
- Реализованы
  - Таблицы страниц
  - Механизм трансляции



# Grant Table

- механизм реализации общей памяти
- Каждый домен использует свою таблицу

Реализован интерфейс в Embox:

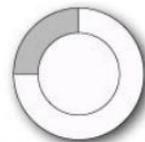
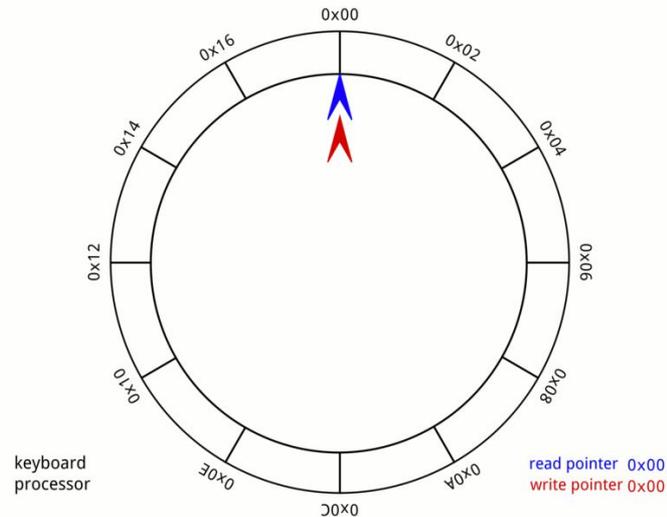
- Оперирует машинными адреса гипервизора
  - Использует трансляцию адресов
- Доступ происходит по индексу (grant reference)

# Rings

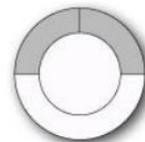
— циклический двусвязный список для асинхронной коммуникация между доменами

Алгоритм отправки пакета:

1. Выделить страницу
2. Записать пакет
3. Определить машинный адрес страницы
4. Изменить права доступа и сделать запись в Grant table
5. Создать запрос в кольце



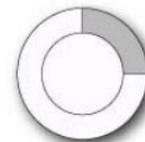
DomU пишет  
запрос №1



DomU пишет  
запрос №2



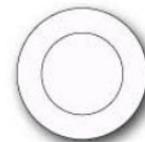
Dom0 пишет  
ответ №1



DomU читает  
запрос №1



Dom0 пишет  
ответ №2



DomU читает  
запрос №2

# XenStore

— это иерархическое пространство имен для хранения информации, разделяемое между доменами.

- Каждый домен имеет собственный путь
- Используется для синхронизации разделенного драйвера
  - Запись индексов Grant table для колец передачи и получения запросов
  - Запись номера прерывания сетевого драйвера
  - Конфигурационная информация сети
  - Хранит текущее состояние драйвера
  - Получение mac-адреса

# XenBus

— это API-интерфейс ядра, используемый драйверами и приложениями для взаимодействия с XenStore.

- Механизм транзакций
- Атомарные операции
- Избыточная логика => частично реализован и внедрен в процесс инициализации драйвера сети

# Оценка производительности

- Сравнение с Alpine Linux
- Окружение: VirtualBox на Intel Core i3 3217u, 256Мб для DomU
- Метрики производительности:
  - Задержка, время отклика
    - ping 5 раз по 60 минут
    - ping -f 10 раз по 45 минут + wget
  - Пропускная способность
    - domU: httpd
    - dom0: Httpperf и Seige
  - Количество одновременных подключений

# Оптимизации

- Разделение канала оповещения для приема и передачи
- Оптимизация работы драйвера при обмене страницами (feature-rx-copy)
- Уменьшение количества прерываний
- Отключение подсчёта контрольной суммы

# Время отклика

- ping 5 раз по 60 минут
- ping -f 10 раз по 45 минут + wget

	<b>Embox до оптимизации</b>	<b>Embox после оптимизации</b>	<b>Alpine</b>
min	3.1	0.4	0.2
max	25.6	9.4	5.4
avg	12.2	5.2	1.5

Таблица 1: Сравнение времени отклика (мс)

# Пропускная способность

- domU
  - httpd
- dom0
  - Httpperf
  - Seige
- Размер файла от 10Кб до 100Мб

Embox <b>до</b> оптимизации	Embox <b>после</b> оптимизации	Alpine
2.78	9.47	29.3

Таблица 2: Сравнение пропускной способности (Мб/с)

# Результаты

Реализован и протестирован прототип unikernel системы на основе ОСРВ Embox для архитектуры гипервизора Xen.

- Изучена архитектура гипервизора Xen, ОСРВ Embox, а также существующих решений
- Реализован и интегрирован в систему сборки Embox механизм общей памяти
- Реализован и интегрирован сетевой интерфейс Embox
- Произведена оценка и оптимизация сетевой производительности Embox, достигнуты стабильные результаты

<https://github.com/CherepanovAleksei/embox/tree/xen19>

# Обзор существующих решений

- Unikernels
  - HaLVM
  - MirageOS
  - Ling
  - Rumpkernel
- MiniOS — минимальная реализация
- Linux — полная реализация
  - AlpineOS — легковесный дистрибутив

# Трансляция

- Используется таблица страниц
- Три уровня вложенности
- Таблица не полностью заполнена гипервизором
- Из адреса вычисляются сдвиги в таблицах
- Написаны макросы в модуле памяти