



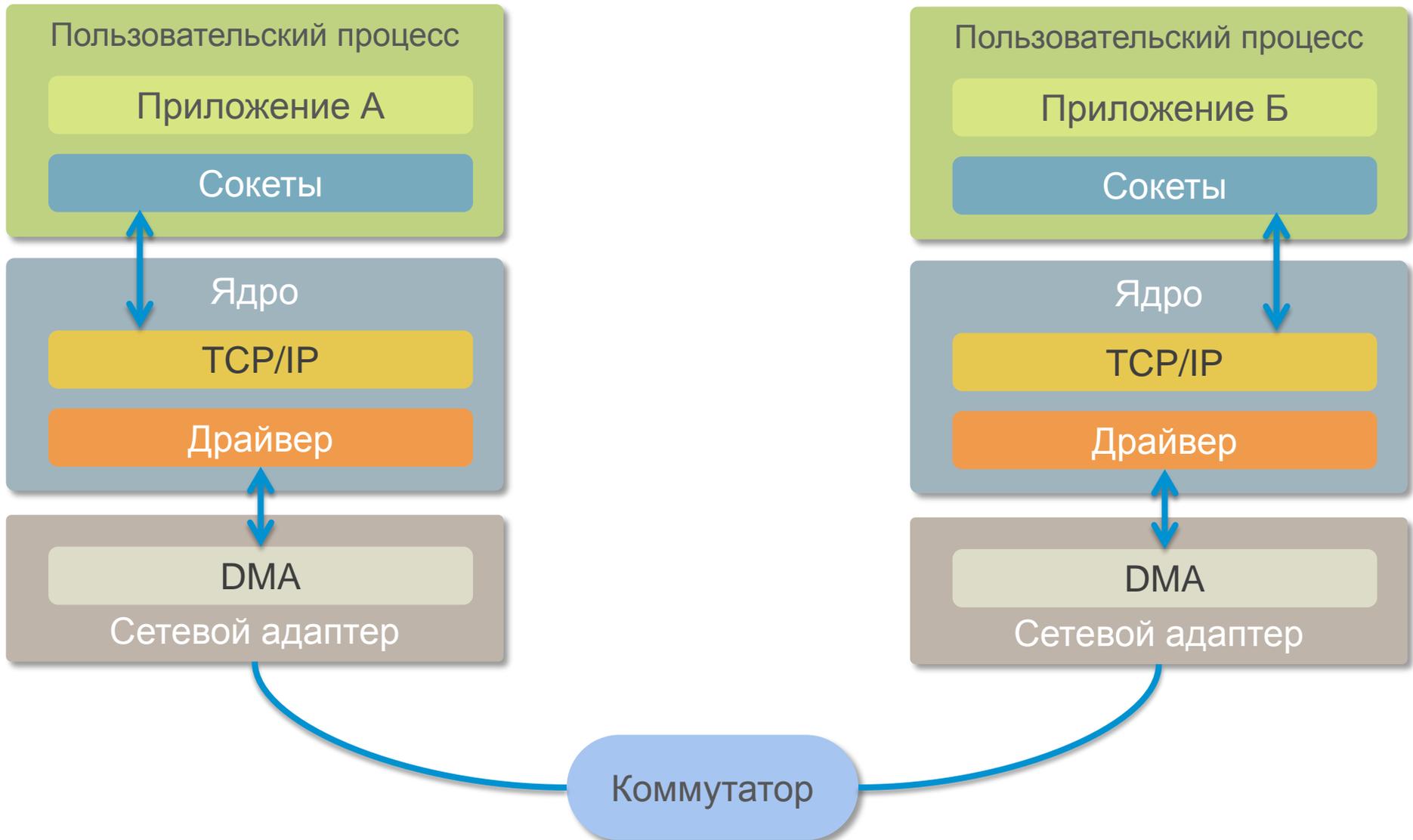
# OpenOnload: повышаем производительность распределенных систем

Владимир Кишик  
vladimir.kishik@db.com

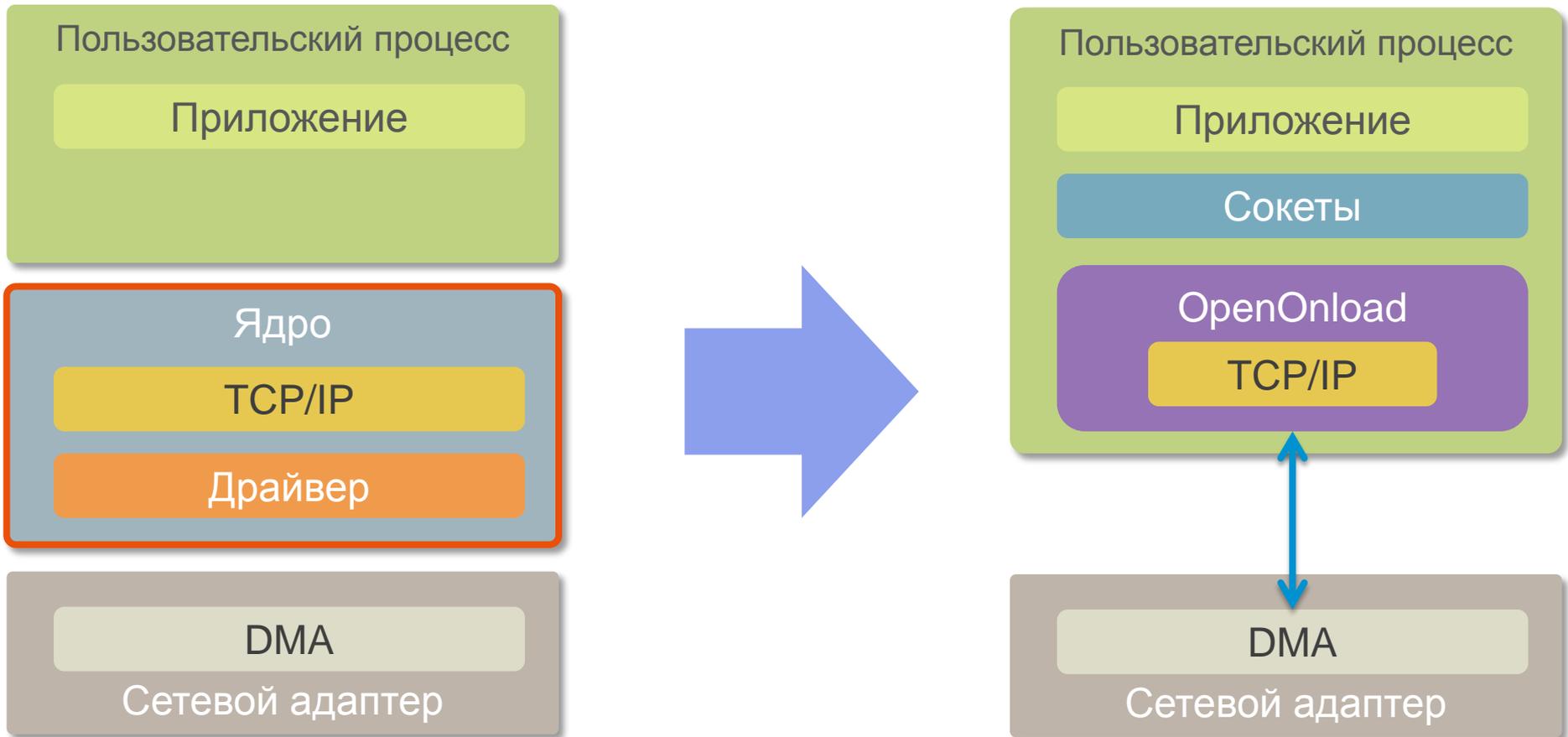
*Passion to Perform*



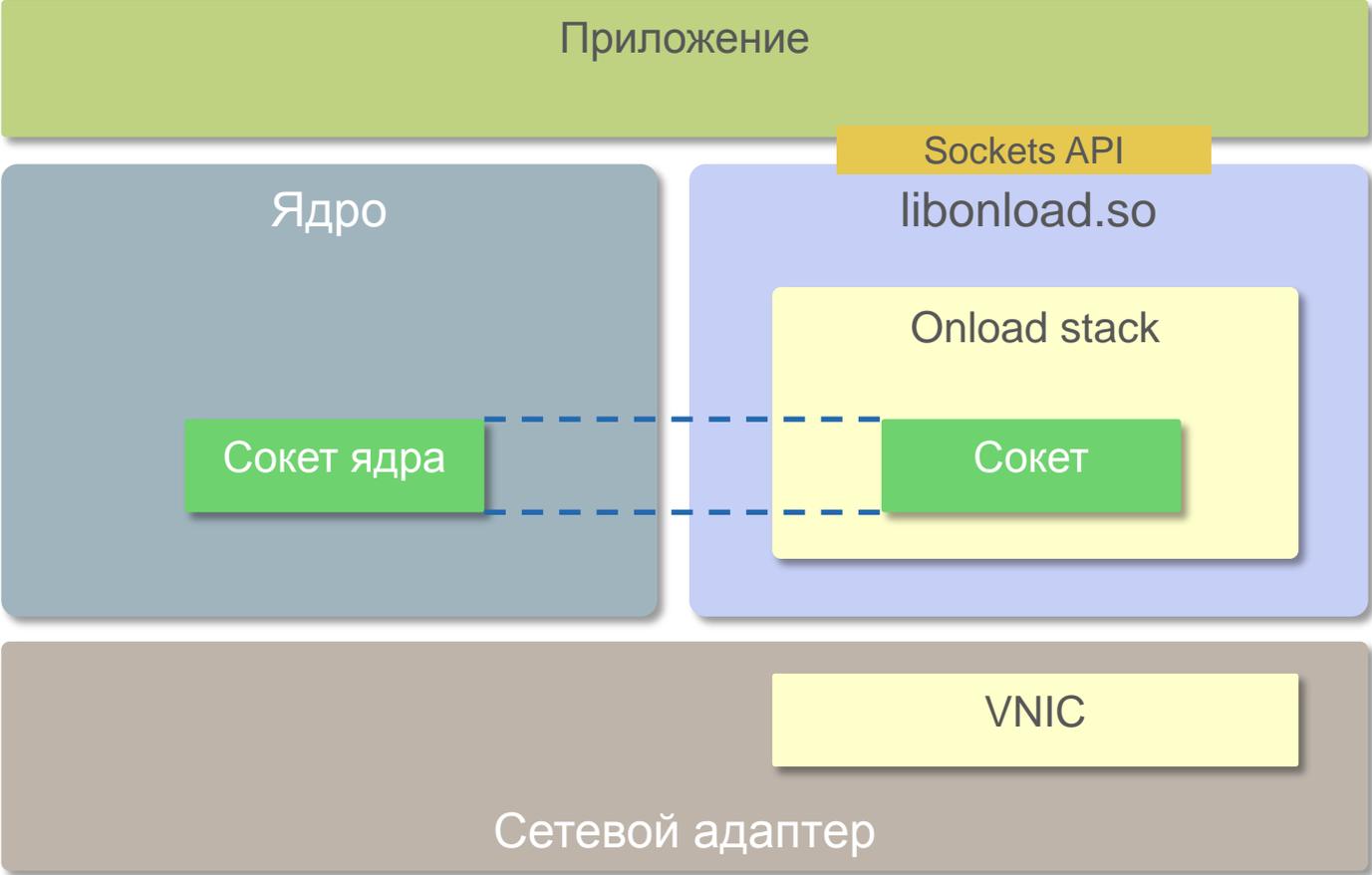
# Цена использования сети



# Введение в Solarflare OpenOnload



# Введение в Solarflare OpenOnload



# Введение в Solarflare OpenOnload

## Основные факты



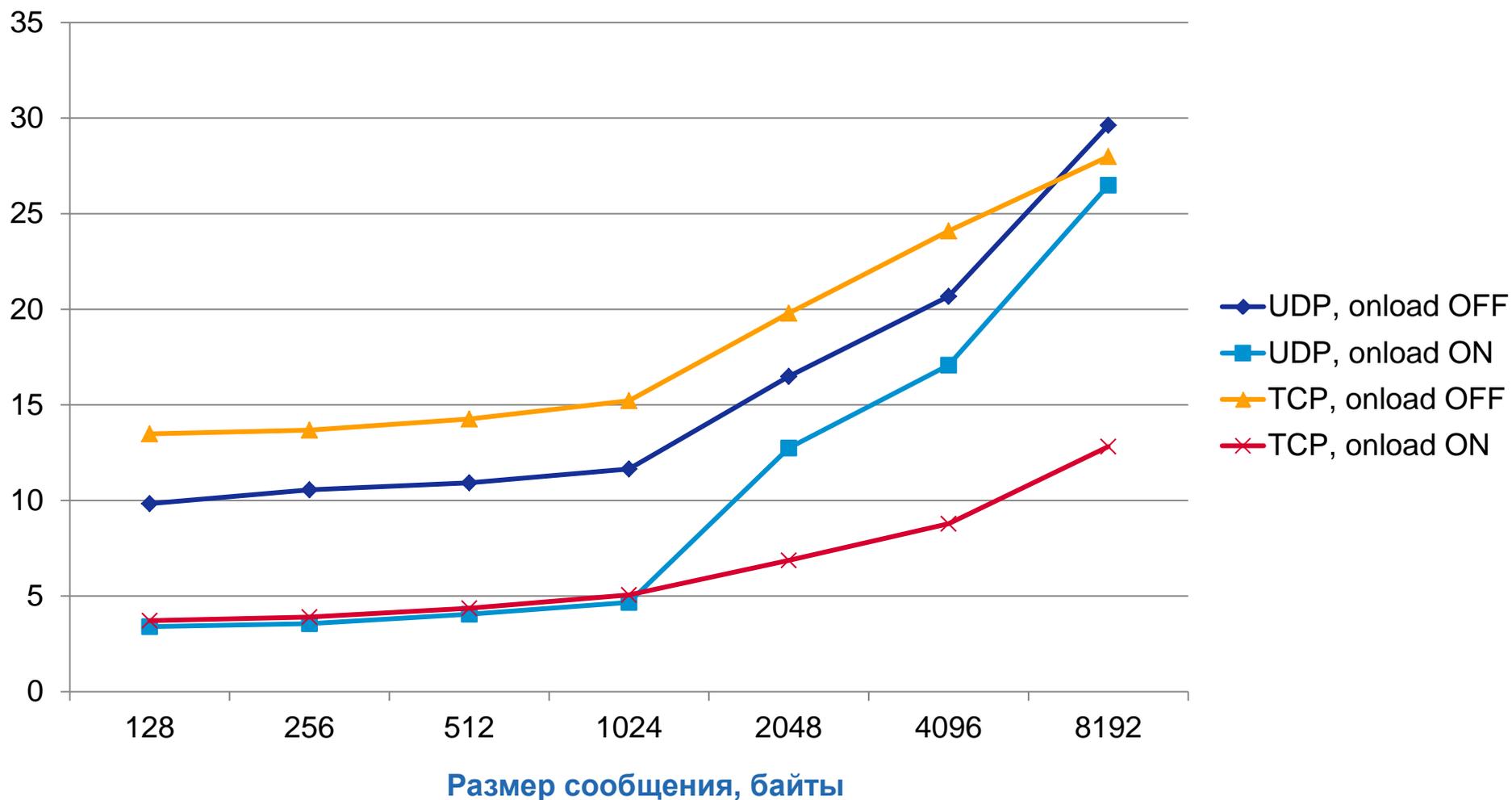
- Устраняет обращения к ядру
  - Реализация стека TCP/IP в контексте приложения
  - Прямой доступ к памяти сетевого адаптера
- Open-source библиотека под Linux
- Загружается динамически, перехватывает вызовы sockets API
- Поддержка TCP, UDP, multicast UDP
- Не требует изменений в коде
  - Настройка через переменные окружения
  - Можно использовать API для управления стеками
- Нужна поддержка со стороны сетевого адаптера
  - Solarflare, HP
- Не зависит от языка программирования

# Тестирование производительности

## Пинг-понг, UDP и TCP пакеты разной длины



½ RTT, мкс

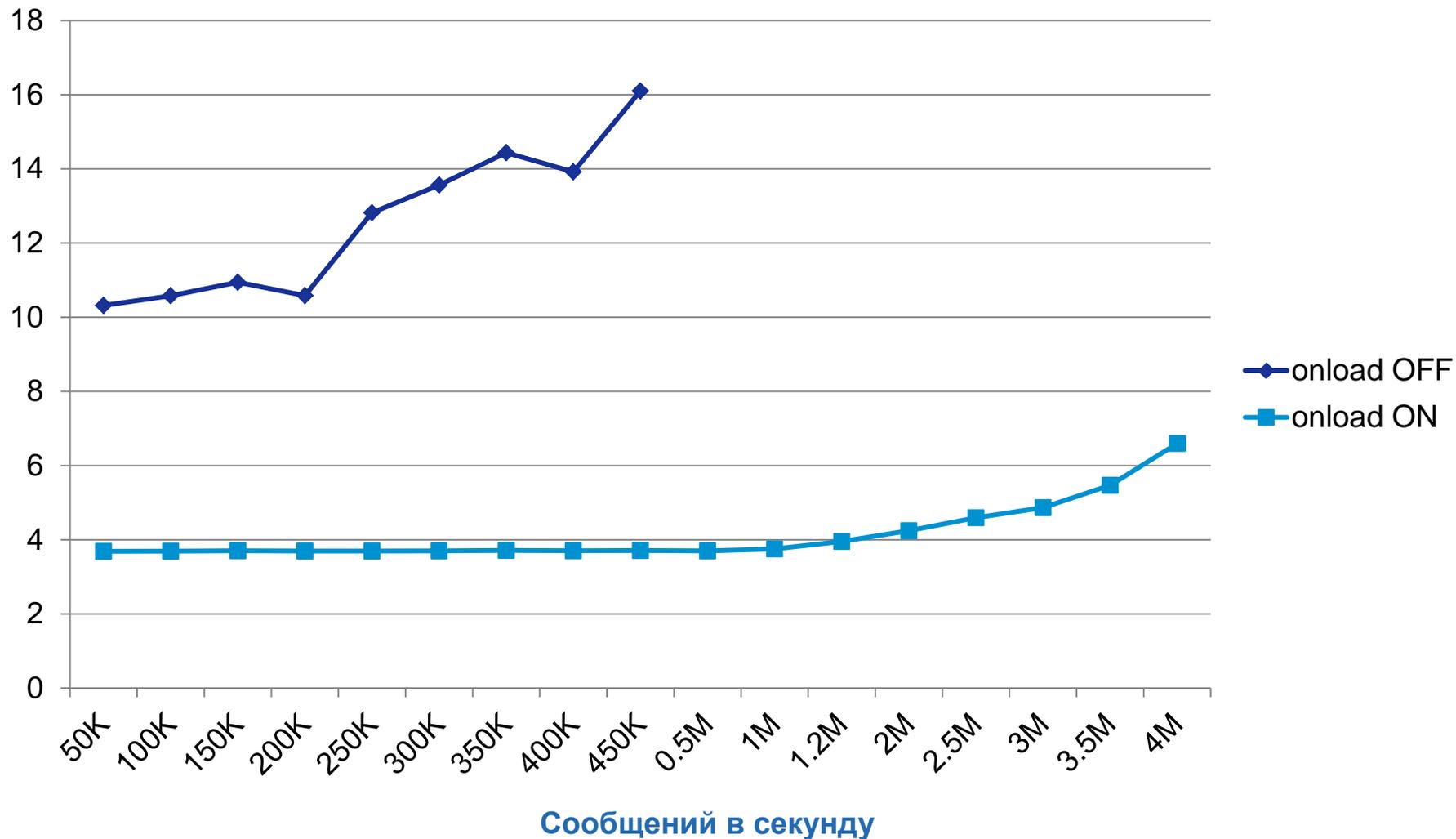


# Тестирование производительности

## Потоковая посылка UDP пакетов, 128 байт



½ RTT, мкс



# Тестирование производительности

## Основные результаты



- Передача пакета между процессами за единицы микросекунд
- Миллионы сообщений в секунду из одного потока
- Меньше времени CPU на вызовы sockets API
  - Пример:

```
LATENCY_CHECKPOINT(A);  
send(packet);  
LATENCY_CHECKPOINT(B);
```
  - 4.7 мкс vs 0.9 мкс

# Тестирование производительности

## Выводы



- Улучшение производительности не будет заметно в *любой* системе
- Когда может помочь OpenOnload:
  - Большое количество TCP соединений (>1К)
  - Интенсивный multicast UDP трафик внутри системы (>10К/сек)
  - Небольшой размер сетевых сообщений (<10КБ)
  - Много одновременных запросов на хост
  - Затраты на сетевые операции сравнимы с полезными вычислениями

# Возможности применения

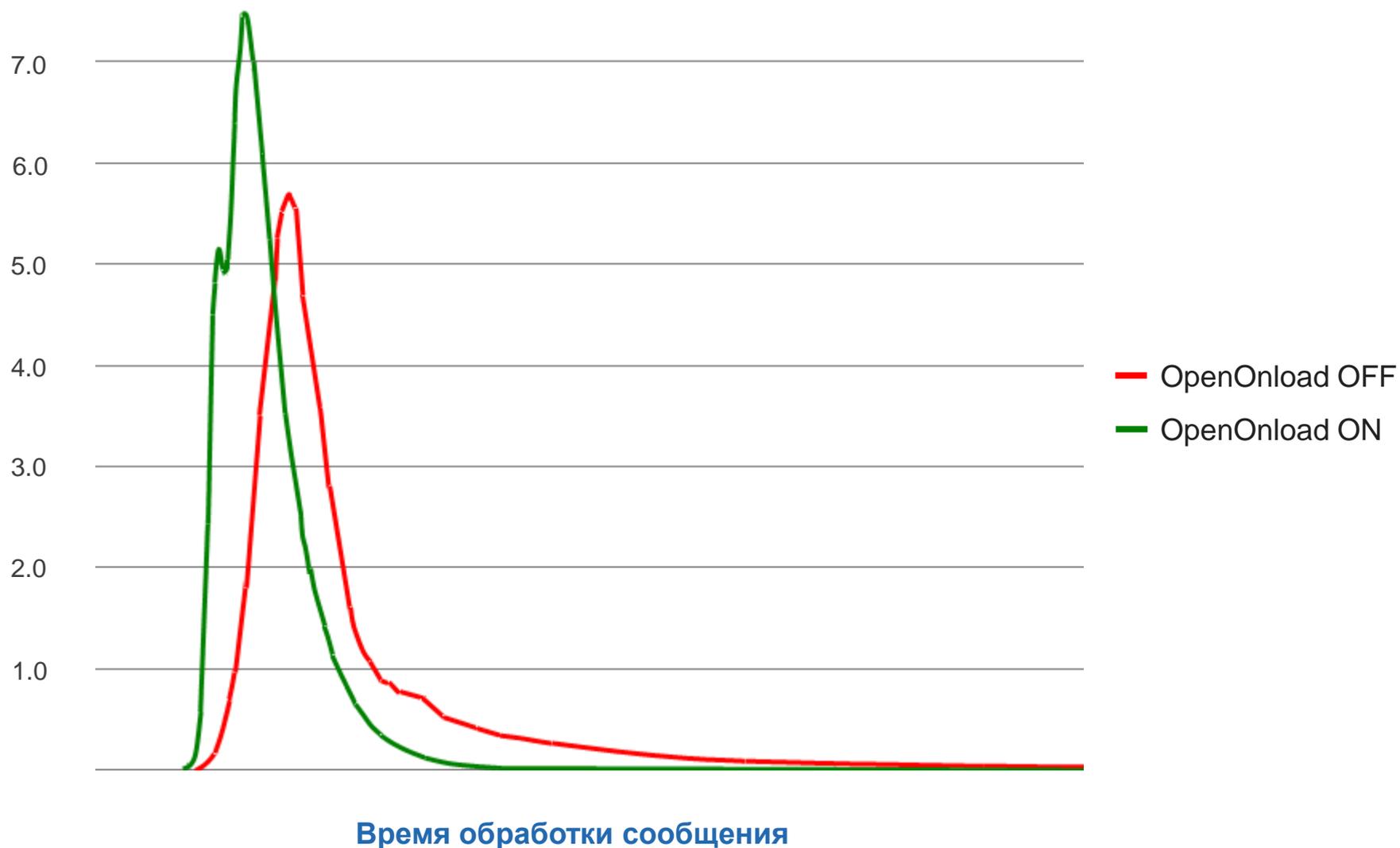


- Финансы
  - ПО для бирж и других торговых площадок
  - Обработка рыночных данных в реальном времени
  - Высокочастотная торговля
- Поточковая передача видео (сервисы Video-on-demand)
- Распределенные хранилища данных (memcached)
- Web
- Облачные платформы
- Онлайн-игры

# Эффект в реальных приложениях



Распределение, %



# Заключение



## Плюсы

- Помогает добиться ощутимого прироста производительности
- Относительно невысокая цена оптимизации
- Дополняет другие методы оптимизации (код приложения, модернизация оборудования)

## Минусы

- Реализация TCP/IP не идентична стеку ядра
- Решение менее стабильно, чем ядро Linux
- Ограниченная линейка совместимого сетевого оборудования

## Дополнительная информация

- [www.openonload.org](http://www.openonload.org), [support.solarflare.com](http://support.solarflare.com) (Whitepapers)



Спасибо!

Q&A