

VSOCK

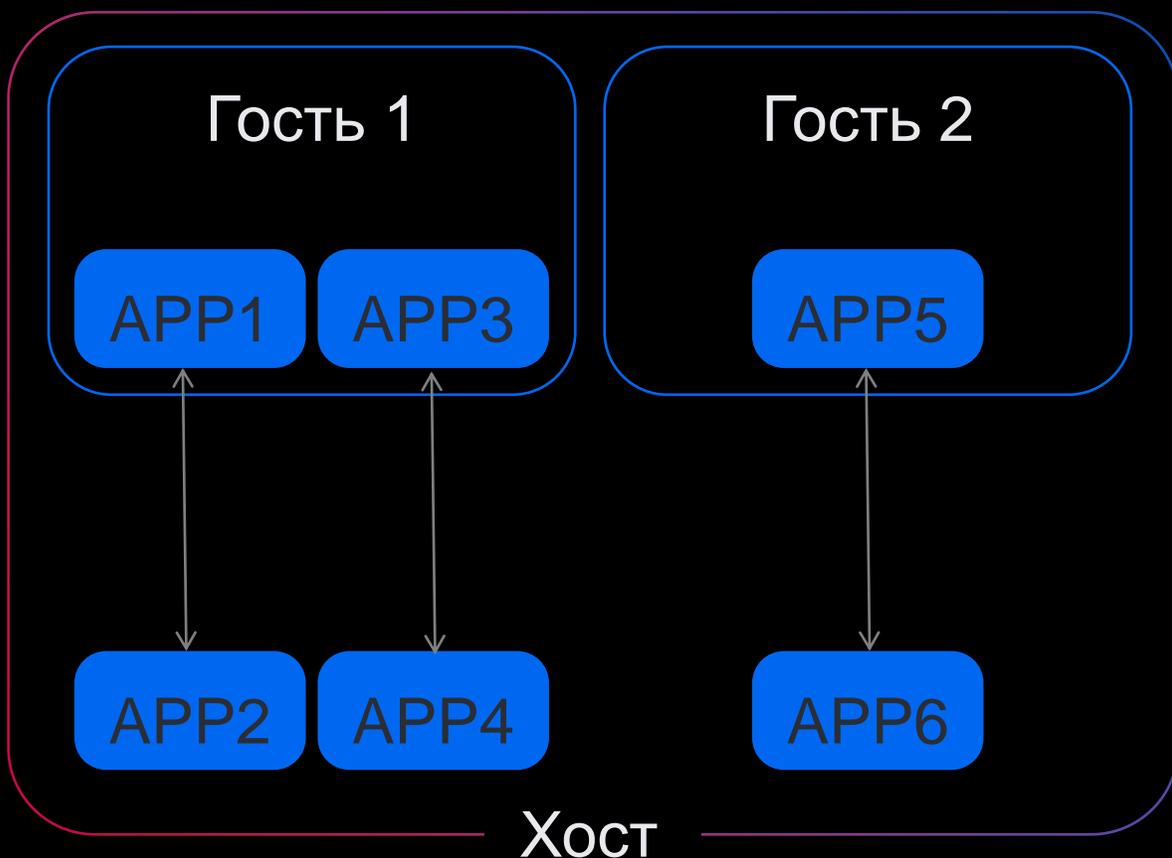
коммуникация между гостем и хостом
с минимальными затратами

Арсений Краснов, SberDevices



Что такое VSOCK

VSOCK ↔ VIRTUAL MACHINE SOCKET



Передача данных гость ↔ хост

POSIX API сокетов

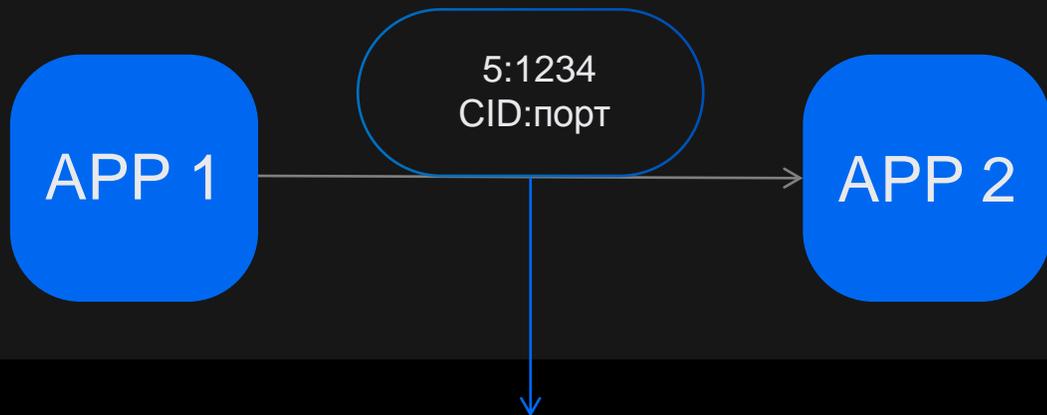
Типы сокетов

- SOCK_STREAM
- SOCK_DGRAM
- SOCK_SEQPACKET

Поддержка в Linux, Windows

Отличия от сокета AF_INET

AF_VSOCK



CID вместо IP адреса

CID – **Context ID**

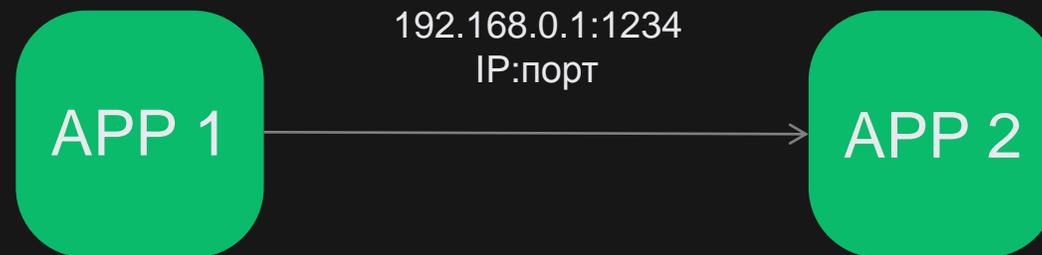
32 битное число

CID гостя > 3

CID хоста 2

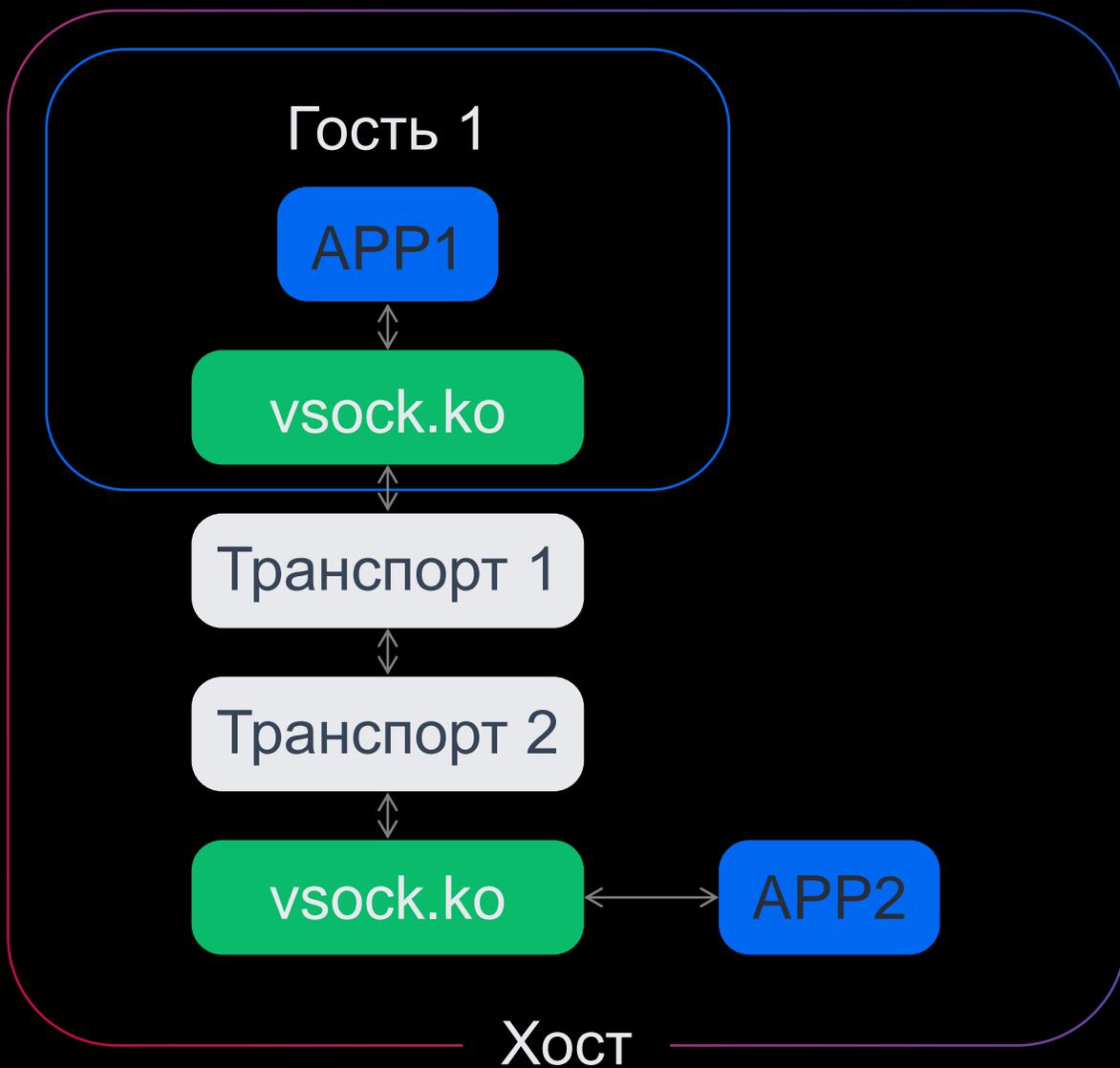
CID loopback 1

AF_INET



socket(AF_INET) ↔ socket(AF_VSOCK)

Архитектура VSOCK в ядре



vsock.ko реализует интерфейс сокета

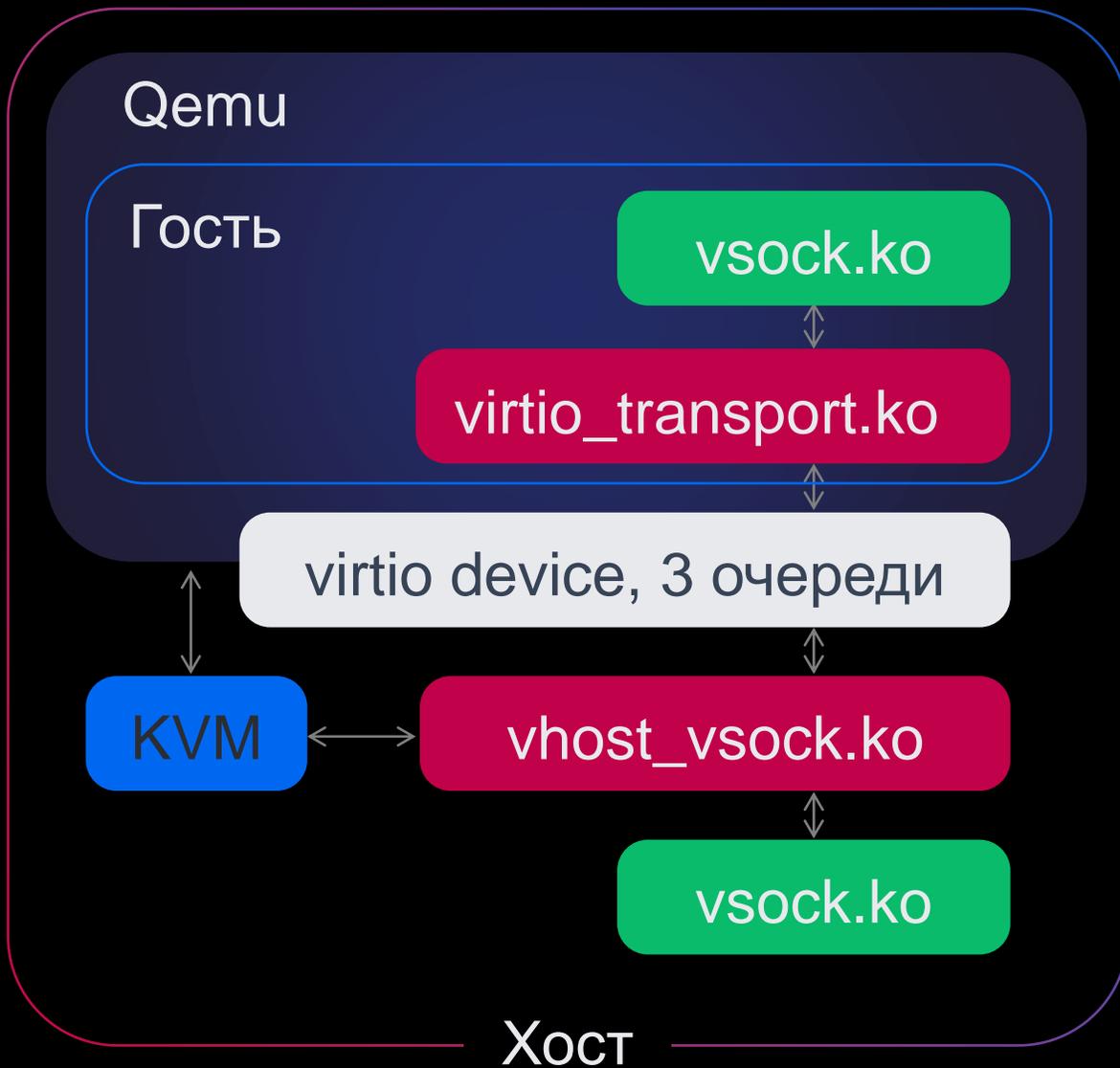
vsock.ko общий у гостя и хоста

Транспорт отвечает за передачу данных

Транспорт специфичен для гипервизора

- VMWare: VMCI
- Microsoft: HyperV
- Qemu/KVM: virtio/vhost

Архитектура VSOCK в ядре (Qemu/KVM)



virtio устройство

3 virtio очереди (RX/TX/Event)

Используется свой протокол

- Установка соединения
- Управление объемом данных

Есть в virtio спецификации

Протокол VSOCK поверх virtio

Заголовок



```
struct virtio_vsock_hdr {  
    __le32  cid;  
    __le32  port;  
    __le32  len;  
    __le16  type;  
    __le16  op;  
    __le32  buf_alloc;  
    __le32  fwd_cnt;  
};
```

Данные (до 64 Кб)

`cid` – CID приёмника

`port` – порт приёмника

`len` – размер данных

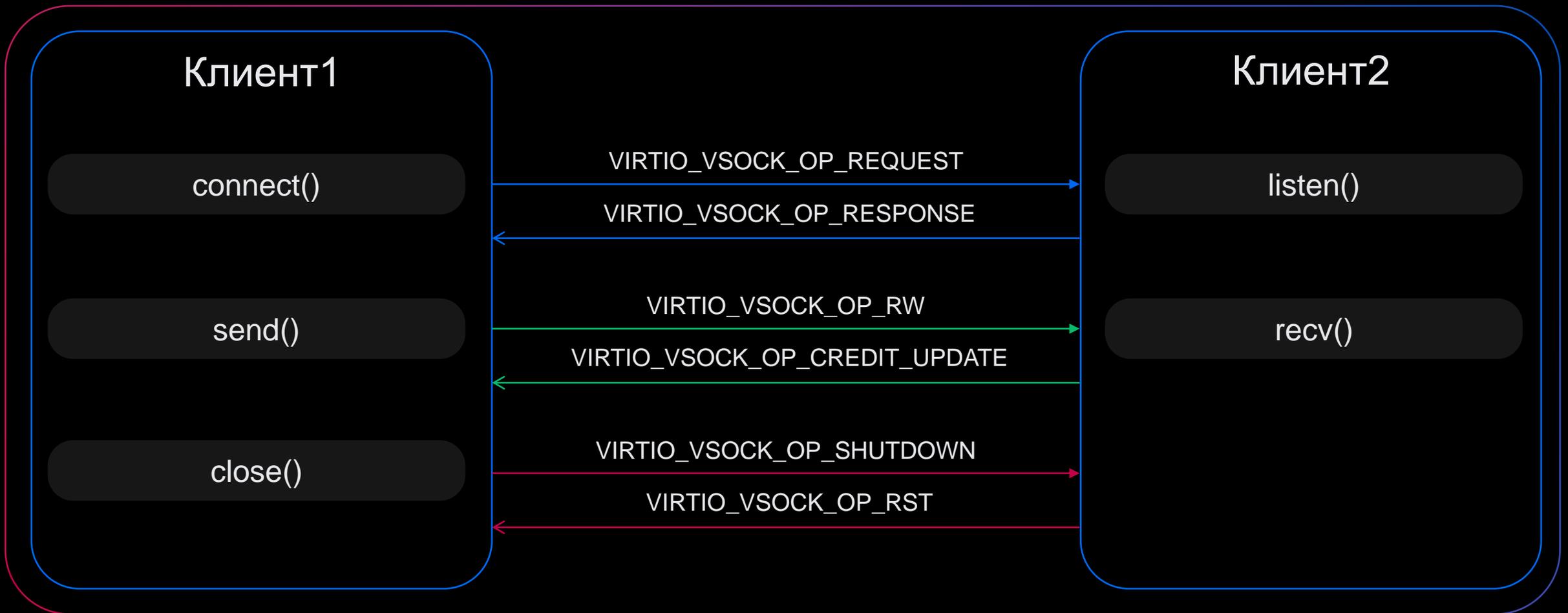
`type` – тип сокета (SOCK_STREAM, SOCK_DGRAM,...)

`op` – тип запроса (установка соединения, данные,...)

`buf_alloc` – размер буфера передатчика

`fwd_cnt` – число байт прочитанных передатчиком

Протокол VSOCK поверх virtio



→ установка соединения

→ передача данных

→ разрыв соединения

Поддержка VSOCK в утилитах

АНАЛИЗАТОРЫ ТРАФИКА

wireshark

tcpdump

БЕНЧМАРКИ

iperf-vsock

upperf

ПРОЧЕЕ

iproute2

nmap

socat-vsock

Поддержка VSOCK в языках программирования

C/C++

Python

> 3.7

Golang

Rust

vsock crate

Практическое применение VSOCK

ИСПОЛЬЗУЕТСЯ
В ПО

ИСПОЛЬЗУЕТСЯ
В КОМПАНИЯХ



Android
debug bridge



Kata
container agent



Qemu
guest agent



RedHat



Bytedance



Amazon



OpenSynergy

Ближайшие планы по развитию в ядре Linux

SOCK_DGRAM
(virtio транспорт)

на code review

Передача без копирования
(virtio транспорт)

на code review

Прием без копирования
(virtio транспорт)

RFC

Заключение

1

VSOCK не требует
конфигурирования



3

VSOCK
стандартизован



2

VSOCK стек очень простой
и легковесный



4

Сетевые приложения легко
адаптировать к VSOCK



Спасибо

