



iSCALARE

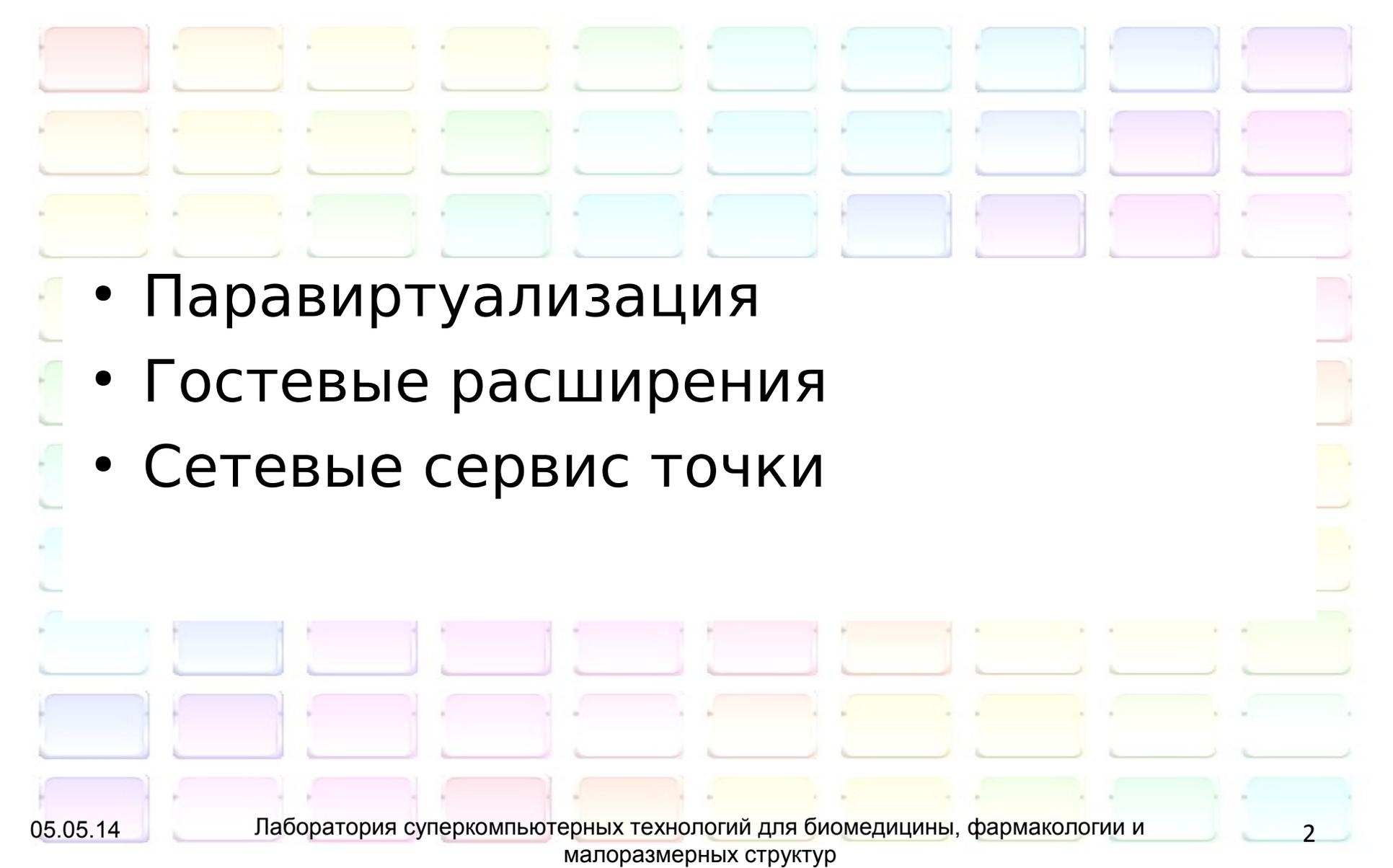


Лаборатория суперкомпьютерных технологий для биомедицины, фармакологии и малоразмерных структур

Связь реальности и виртуальности

Григорий Речистов

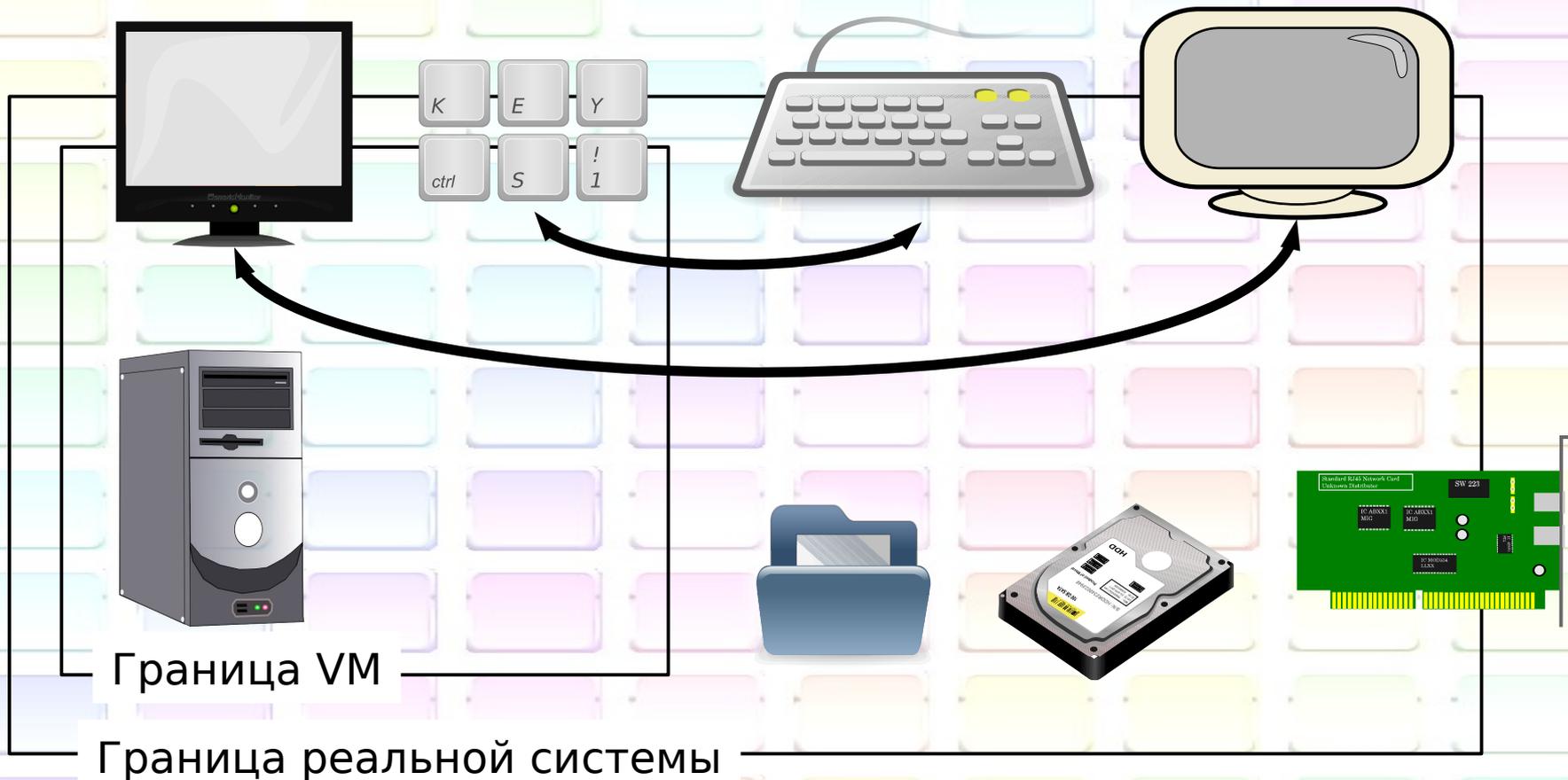
grigory.rechistov@phystech.edu

- 
- Паравиртуализация
 - Гостевые расширения
 - Сетевые сервис точки

На предыдущих лекциях

- Мы стремились к наиболее точной симуляции
- Программы в окружении не должны догадываться о том, что они исполняются не на реальной аппаратуре
- И уж тем более они не должны видеть реальное окружение

Изоляция

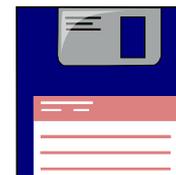


Мотивация

- Иметь возможность передавать данные, объёмом превышающие 1 кбайт, в/из симулируемой системы
- Иметь возможность передавать аппаратные ресурсы частично или полностью в VM

Образы дисков (1/4)

- Жёсткие диски
- Оптические диски
- Гибкие диски



Образы дисков (2/4)

- Жёсткие диски
 - RAW
 - VMDK
 - VDI
 - Qcow2
 - CRAFF
 - HDD
 - VHD

Образы дисков (3/4)

Оптические диски

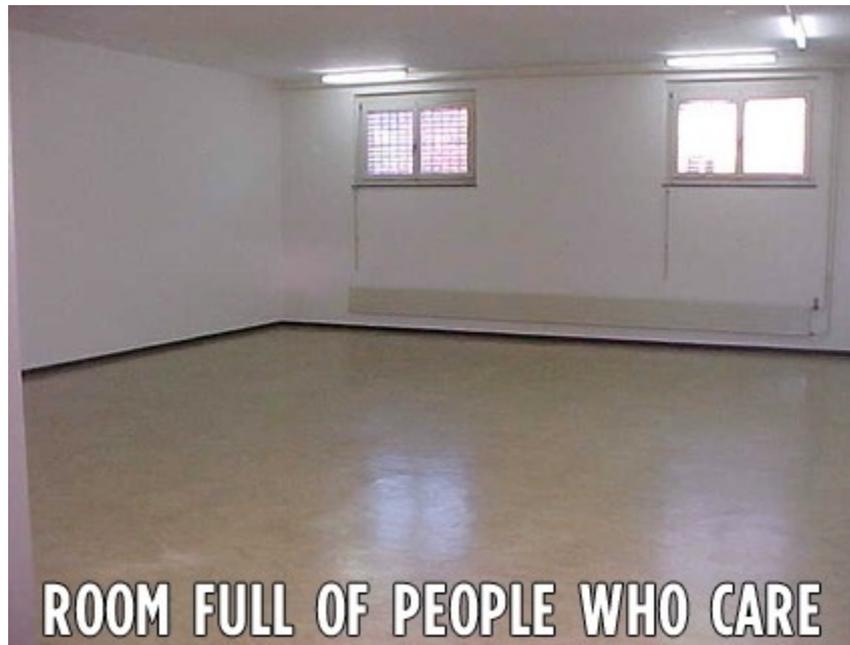
- ISO 9660

Существует множество форматов, но они не используются в симуляторах/VM

- NRG, MDF, ISZ, DMG, IMG, BIN/CUE

Образы дисков (4/4)

- Гибкие диски
- 360 кБ — 2.88 МБ
- Формат — RAW



Последовательный порт (1/2)

- Простое устройство
- Модель добавляется одной из первых
- Поддерживается всеми ОС
- Малая скорость (до 115 кбит/с)
- Имеет современные реинкарнации (HSUART, SOL)



Последовательный порт (2/2)

Со стороны реальной системы может быть присоединён к

- Реальному COM порту
- Виртуальному COM порту
- Именованному каналу (pipe)
- Сетевому сокету
- Эмуляторы терминала
- Файлу

Волшебные инструкции (1/4)

Инструкция процессора с побочными эффектами

- Остановка симуляции
- Вызов обработчика, имеющего доступ к состоянию симулируемой системы
- Изменение состояния
- Возобновление симуляции
- Для VM действия происходят “мгновенно”

Волшебные инструкции (2/4)

Может использоваться для

- Периодической записи состояния гостя
- Отладки программ внутри гостя
- Передачи данных в/из гостя

Волшебные инструкции (3/4)

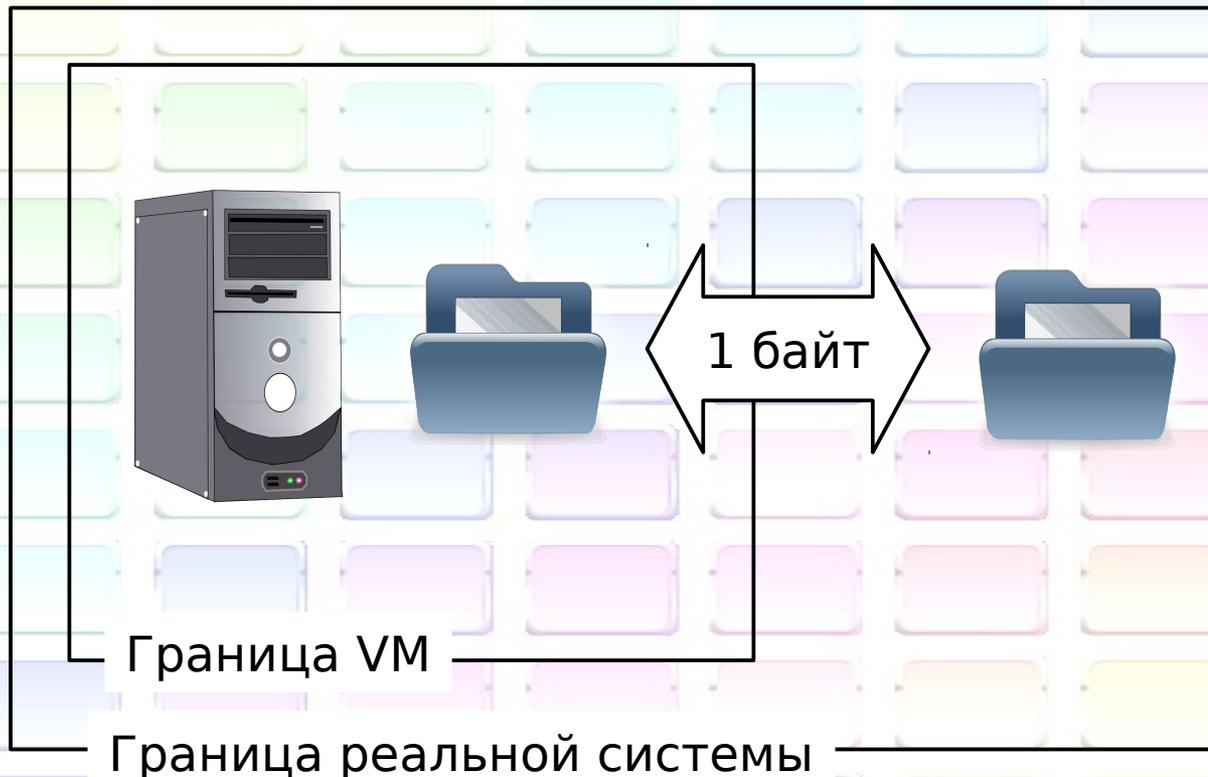
- Очень желательно, чтобы инструкция не встречалась в «обычном» коде
 - Ложные срабатывания
 - Неожиданные для программы эффекты
- Идеально, чтобы она вообще не имела эффектов вне симуляции
 - **NOP** — хороший кандидат
 - ... но она используется в программах очень часто!

Волшебные инструкции (4/4)

Варианты

- **NOP** с необычными префиксами
 - В IA-32 есть два **NOP**, длиной от 1 до 9 байт
- Недокументированные инструкции (если ещё остались)
- «Безобидные» инструкции
 - **CRUID**

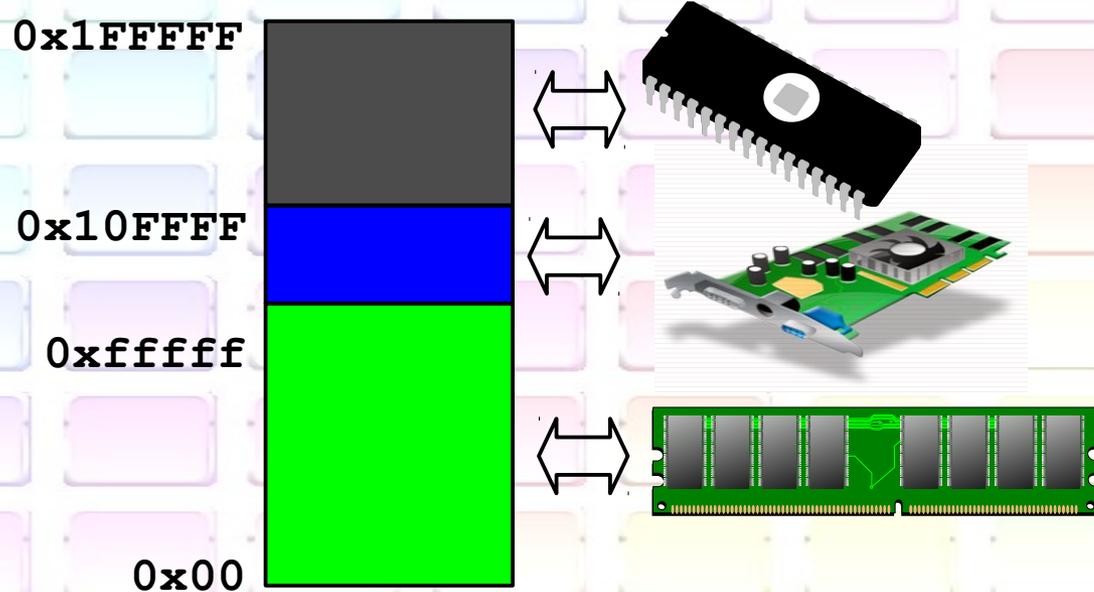
Использование — передача файлов



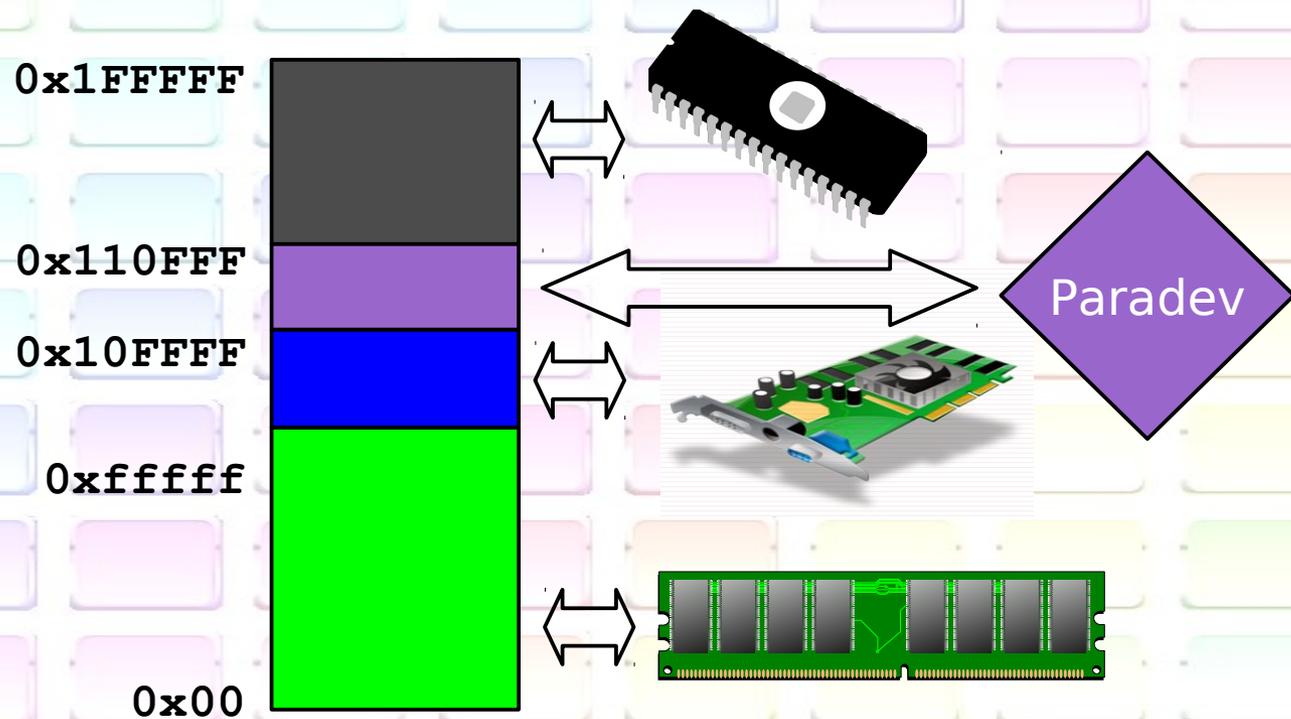
Специальная программа, использующая инструкцию для передачи потока байт

Паравиртуальные устройства (1/3)

Memory mapped input/output



Паравиртуальные устройства (2/3)



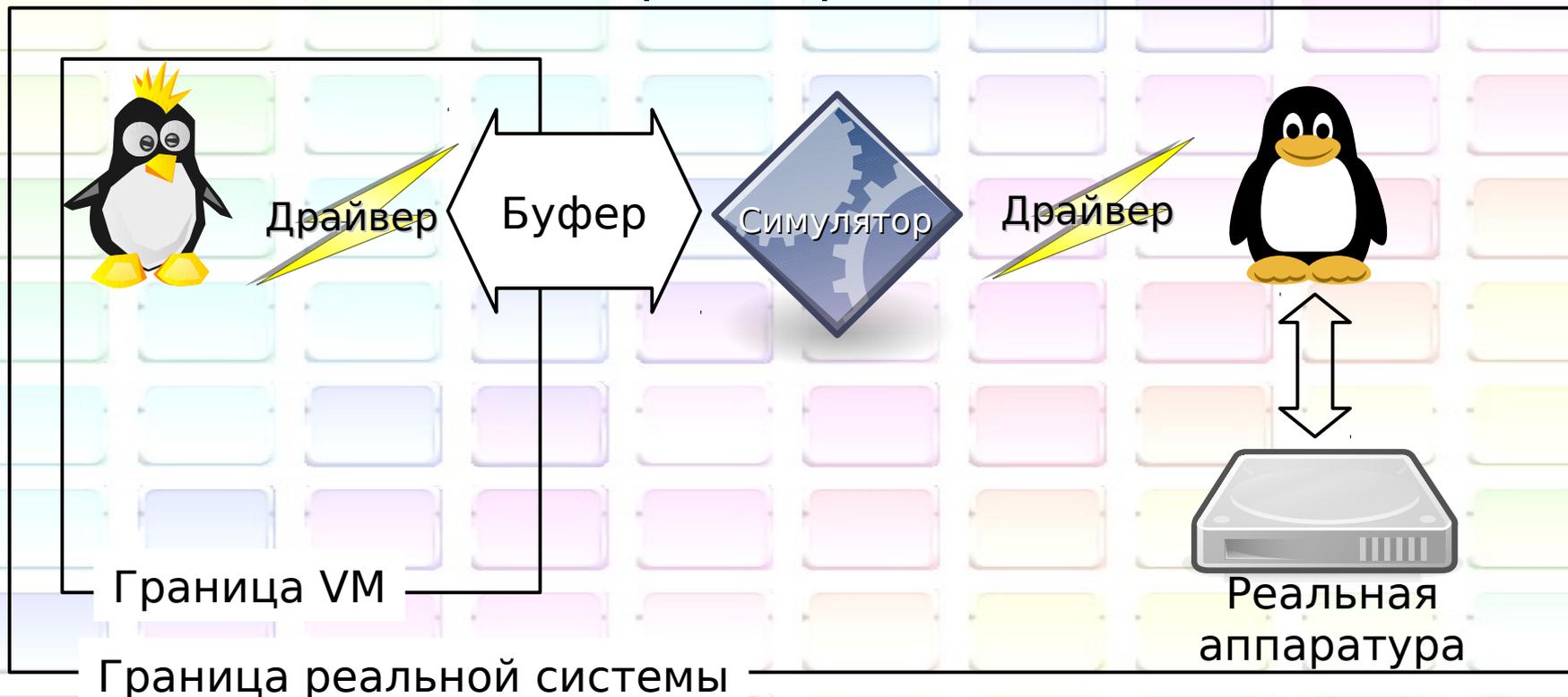
Паравиртуальные устройства (3/3)

- Объём передаваемых данных за один раз больше
- Требует модификации гостевой ОС
 - Драйвера устройств
 - Не столь безобидные инструкции
 - **RDMSR/WRMSR, INT, SYSCALL**

Использование

Симуляция периферии

HostFS а.к.а Гостевые расширения



Проброс (pass-through) устройств (1/3)

Проброс PCI/USB/VGA устройства

- Аналог DEX для периферии
- Передача всех команд/откликов протокола без изменений
- Имеет схожие проблемы
 - Изоляция
 - Безопасность
 - Legacy-режимы

Проброс (pass-through) устройств (2/3)

Аппаратная поддержка (I/O Memory management unit)

- Intel VT-d
- AMD IOMMU
- IBM Translation Control Entry
- Sun DVMA

Проброс (pass-through) устройств (3/3)

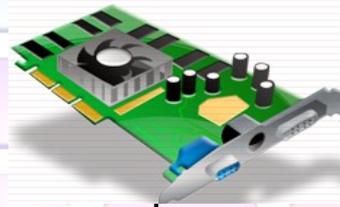
Особенности

- Самомодифицирующийся код VGABIOS
- Отключение устройства от хозяина
- Primary/secondary GPU
- VM Save/Restore

Использование

Граница реальной системы

Передача устройства
в эксклюзивное
использование гостю



Граница VM

Сетевое взаимодействие

Сетевое взаимодействие модели OSI/ISO (русск. ВОС/БЭМ)

- Изначально создано для связи систем различной природы
- Агностично к аппаратуре*
- Можно выбирать уровень OSI/ISO, на котором будет проходить граница миров

OSI/ISO

Прикладной уровень

Представительный уровень

Сеансовый уровень

Транспортный уровень

Сетевой уровень

Канальный уровень

Физический уровень

Прикладной уровень

Представительный уровень

Сеансовый уровень

Транспортный уровень

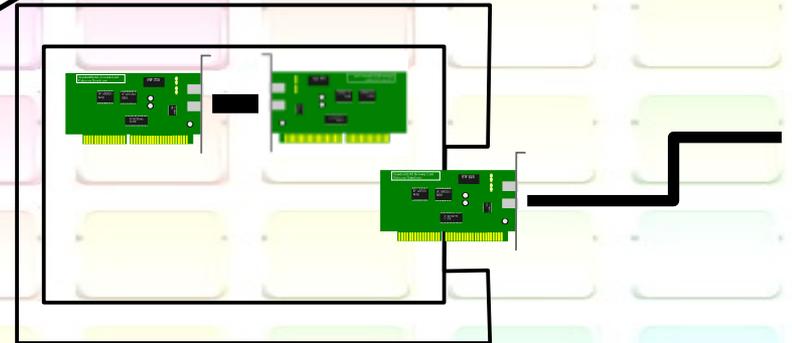
Сетевой уровень

Канальный уровень

Физический уровень

Модель NIC внутри симуляции

1. Связана с другими моделями или
2. Является пробросом реальной NIC



Прикладной уровень

Представительный уровень

Сеансовый уровень

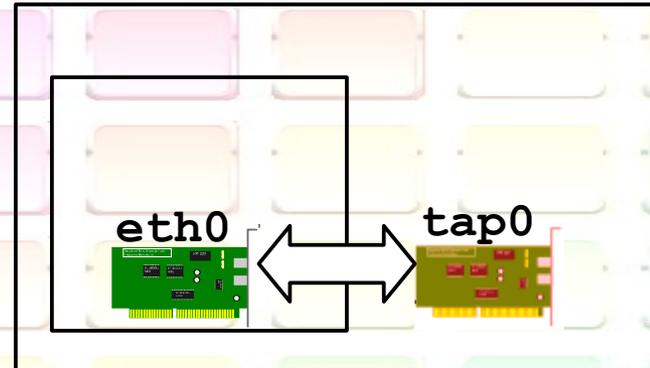
Транспортный уровень

Сетевой уровень

Канальный уровень

Физический уровень

ТАР-драйвер хозяина
Создаёт псевдо-устройство
Ethernet хозяина



Прикладной уровень

Представительный уровень

Сеансовый уровень

Транспортный уровень

Сетевой уровень

Канальный уровень

Физический уровень

TUN-драйвер хозяина
Создаёт IP туннель

192.168.1.3

`/dev/tun0`

Прикладной уровень

Представительный уровень

Сеансовый уровень

Транспортный уровень

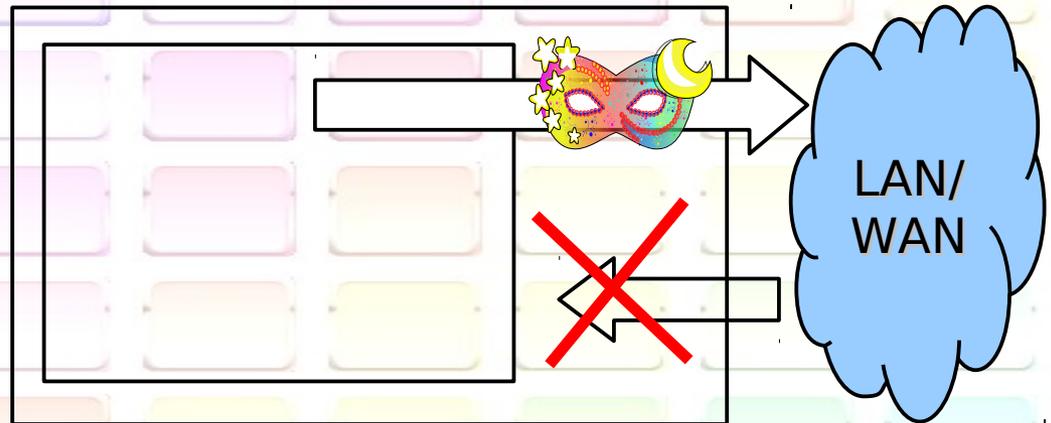
Сетевой уровень

Канальный уровень

Физический уровень

NAT

Исходящие пакеты
Автоматически ретранслируются
от имени хозяина.
Входящие пакеты по умолчанию
не доходят



Прикладной уровень

Представительный уровень

Сеансовый уровень

Транспортный уровень

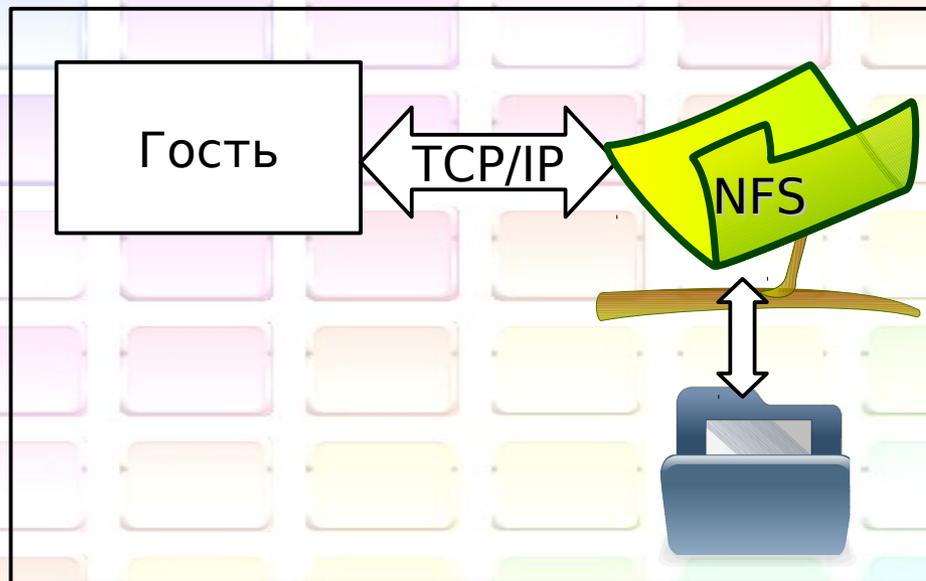
Сетевой уровень

Канальный уровень

Физический уровень

Сервис-точки внутри симуляции

Отвечают на запросы гостя по конкретному протоколу:
NFS, FTP, Samba, DHCP, ...



Рекомендуемая литература

http://wiki.xen.org/wiki/Xen_PCI_Passthrough

<http://wiki.xen.org/wiki/XenVGAPassthrough>

<http://wiki.xen.org/wiki/XenUSBPassthrough>

<http://usbip.sourceforge.net/>

На следующей лекции:

Альтернативные методики изучения компьютерных систем

- Трассы (оффлайн симуляция)
- Аналитические методы
- Методы статистические

Спасибо за внимание!

Все материалы курса выкладываются на сайте лаборатории:

http://iscalare.mipt.ru/material/course_materials/

Замечание: все торговые марки и логотипы, использованные в данном материале, являются собственностью их владельцев. Представленная здесь точка зрения отражает личное мнение автора, не выступающего от лица какой-либо организации.