Конструирование ядра операционных систем (IV+V)

Ввод-вывод

План

- Виды ввода-вывода и их особенности
- Современные средства описания аппаратуры
- Таймеры на x86 и частота процессора
- Практическая часть лабораторной работы

Виды ввода-вывода

—Port-mapped (PMIO)
—Memory-mapped (MMIO или DMA)
—Bus I/O
—Network I/O

—Most bus interfaces (PCI, I2C) usually have HAL
—Network, USB, etc. operate on a packet abstraction layer
—Dedicated hardware is rare

PMIO

—Требует поддержки инструкций процессора (in, out)
—Становится менее популярным по причине MMIO
—В большинстве случаев реализован через HAL
—Может быть проэмулирован в большинстве систем*

* На современных системах Intel это реализуется через SMM (System Management Mode), спасибо Pegatron за детали ☺.

Особенности PMIO

—Доступны везде (даже без RAM и с неизвестной VRAM)
Например, могут быть использованы для отладки ОС

—Достаточно легко эмулируются на всех уровнях

—Проброс в пространство пользователя затруднён, так как не разграничивает доступ к устройствам (всё или ничего)

—Низкая производительность (ограничен размер аргументов 32 битами, фиксированные регистры для кодирования)

—Количество портов ограничено и загрязнено устройствами для обеспечения совместимости

Примеры портов (16-bit)

—0x2E, 0x4E — Super I/O (Fans, Temperature, Watchdog)
—0x70-0x73 — standard RTC registers (CMOS) *
—0x74-0x77 — extended RTC registers (Ivy Bridge) *
—0x3F8-0x3FF — first serial port
—0x2F8-0x2FF — second serial port
—0xCF9 — ACPI restart **

* https://github.com/acidanthera/OcSupportPkg/blob/bad2be8/Library/OcRtcLib/OcRtcLib.c

** https://github.com/acidanthera/OcSupportPkg/blob/7eca596/Library/OcMiscLib/DirectReset.c

MMIO

—Использует стандартные инструкции процессора
—Широко распространён и становится популярнее
—В большинстве случаев HAL отсутствует
—Операции ввода-вывода встраиваются в код (inline)

* Для доступа к адресному пространству PCI (GPU, NIC) используется MMIO.

* DMA (Direct Memory Access) аналогично MMIO, но формально может происходить без участия ЦПУ, т.е. устройство само обновляет ОЗУ.

Особенности MMIO

—Стандартный способ ввода-вывода не только в x86

—Высокая производительность и гибкость

—Сложная высокопроизводительная эмуляция *

—На x86 MMIO фактически эквивалентен DMA, при этом IOMMU (VT-d) является опциональным до Intel Skylake **

* Ref: Lab4-VMW1.pdf, Lab4-VMW2.pdf, Lab4-VSMC.pdf.

** USB Type-C Flash → Thunderbolt Device → Direct PCI-E access → Full RAM access ☺.
Подробнее о DMA атаках: https://github.com/ufrisk/pcileech.

ACPI

Advanced Configuration and Power Interface (6.3, 2019) *:

ACPI tables + ACPI BIOS + ACPI registers

```
Device (RTC)
{
    Name (_HID, EisaId ("PNP0B00") /* AT Real-Time Clock */)  // _HID: Hardware ID
    Name (_CRS, ResourceTemplate ()  // _CRS: Current Resource Settings
    {
        IO (Decode16,
            0x0070,             // Range Minimum
            0x0070,             // Range Maximum
            0x01,               // Alignment
            0x08,               // Length
            )
        IRQNoFlags ()
            {8}
    })
}
```

* https://uefi.org/specifications

* https://www.acpica.org/downloads

* https://github.com/acidanthera/MaciASL

SMBIOS

System Management BIOS is the premier standard for delivering management information via system firmware (3.3.0).

```
Handle 0x0001, DMI type 0, 26 bytes
0000: 00 1a 01 00 01 02 00 00 03 ff 80 9c 8b 3f 01 00
0010: 00 00 03 0d 05 0c ff ff 00 00

BIOS Information
        Vendor: Acidanthera
        Version: 175.0.0.0.0
        Release Date: 06/17/2019
        ROM Size: 0 MB ☺
        Characteristics:
                PCI is supported
                APM is supported
                BIOS is upgradeable
                BIOS shadowing is allowed
                Boot from CD is supported
                ....
                ACPI is supported
                USB legacy is supported
                BIOS boot specification is supported
                Targeted content distribution is supported
                UEFI is supported
        BIOS Revision: 5.12
```

Table 17 – System Enclosure or Chassis Types

Byte Value	Meaning
01h	Other
02h	Unknown
03h	Desktop
04h	Low Profile Desktop
05h	Pizza Box ☺
06h	Mini Tower
07h	Tower
08h	Portable
09h	Laptop
0Ah	Notebook
0Bh	Hand Held
0Ch	Docking Station
0Dh	All in One
0Eh	Sub Notebook
0Fh	Space-saving
10h	Lunch Box ☺
11h	Main Server Chassis

* https://www.dmtf.org/standards/smbios
* http://www.nongnu.org/dmidecode (Linux) или https://github.com/acidanthera/dmidecode (macOS)

CMOS vs NVRAM

CMOS (RTC)	NVRAM (UEFI Variable Storage)
Volatile Memory + Battery / Cell	Flash Memory
128 or 256 bytes	> 64 KB
Legacy x86 & UEFI	UEFI
I/O ports: 0x70-0x73 on all x86 I/O ports: 0x74-0x77 on Intel Ivy Bridge+	EFI_RUNTIME_SERVICES GetVariable/SetVariable/GetNextVariable
Raw Memory Access	GUID Namespaces
128 or 256 bytes	> 64 KB
No user accessible memory No read/write restrictions	All GUIDs but reserved are accessible Authenticated & protected variable access *
Used for BIOS preferences & World Time	Used for BIOS preferences, boot priority, secure boot, user storage, OS preferences…

* SMM как Security Boundary или WP биты на контроллерах SPI (см. SRL биты Status Register):
https://www.winbond.com/resource-files/w25q257jv%20spi%20revb%2005032017.pdf

Таймеры x86 (I)

- Real-Time Clock (RTC): M48T86, DS1307; 32.768~100 КГц
  Часть Super I/O (Winbond, Nuvoton, ITE), порты 0x70~0x73
  Intel Coffee Lake и ниже, альтернатива Time and Alarm (TAD)
- Intel 8253 (8254, PIC, Programmable Interval Timer), 1.19 МГц
  Настраивается Intel 8259 (PIC), IRQ#0, порты 0x40-0x43
  Основной таймер для UEFI Intel Haswell и ниже
- ACPI Power Management Timer: 3.41 МГц, 24 или 32-битный
  Доступен из таблицы ACPI FADT или для AMD и Intel:
  https://github.com/acidanthera/OcSupportPkg/blob/24b3cde/Library/OcCpuLib/OcCpuLib.c#L593-L676
- High Precision Event Timer (HPET): 10 МГц или выше
  Доступен как устройство ACPI, Windows Multimedia Timer

Таймеры x86 (II)

- (Local) APIC Timer, зависит от частоты шины, ~до 1 ГГц
  Уникален для каждого ядра ЦПУ, доступ по MMIO
- Time Stamp Counter (TSC): номинальная частота ЦПУ
  Чтение через инструкцию rdtsc
- Always Running Timer (ART), 19.2~25 МГц
  Таймер для некоторых устройств для Skylake и выше
  https://github.com/acidanthera/bugtracker/issues/448#issuecomment-519598979

TSC Frequency = ART Frequency * CPUID.15H.EBX / CPUID.15H.EAX,

если у вас не Xeon W (CPUID 0655H) с дополнительной недокументированной и отключаемой схемой уменьшения помех, уменьшающей частоту на 0.25% ☺

Спасибо за внимание!

Вопросы?