Taller de drivers Sistemas Operativos

Gonzalo Pablo Fernández ¹

¹Departamento de Computación Facultad de Ciencias Exactas y Naturales Universidad de Buenos Aires

22 de mayo de 2018

Antes de empezar

- Descargar SO.ova de la página de la materia (si están en los labos, descárguenla en el disco de la máquina /media/libre).
- Abrir VirtualBox e importar la VM.
 - ▶ Archivo -> Importar Servicio Virtualizado -> SO.ova
- Iniciar la vm.
- Ingresar con usuario y contraseña alumno.

Husky



http://clearpathrobotics.com/husky

Hoy vamos a hacer un driver.

Hoy vamos a hacer un driver. Whattt ??

Hoy vamos a hacer un driver. Whattt ??

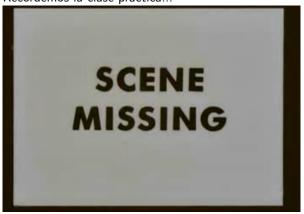
• ¿Qué es un driver?

Hoy vamos a hacer un driver. Whattt ??

- ¿Qué es un driver?
 - Recordemos la clase práctica...

Hoy vamos a hacer un driver. Whattt ??

- ¿Qué es un driver?
 - ► Recordemos la clase práctica...



Hoy vamos a hacer un driver. Whattt ??

- ¿Qué es un driver?
 - ► Recordemos la clase práctica...



Ok, recordemos la clase teórica...

¿Qué se acuerdan de la teórica?



¿Qué es un driver? (Clase teórica)

¿Qué es un driver? (Clase teórica)

- Un dispositivo de E/S va a tener, conceptualmente, dos partes:
 - El dispositivo físico.
 - ▶ Un controlador del dispositivo que interactúa con el SO mediante algún tipo de bus o registro.
- Los drivers son componentes de software muy específicos. Conocen las particularidades del HW contra el que hablan.

¿Qué es un driver? (Clase teórica)

- Un dispositivo de E/S va a tener, conceptualmente, dos partes:
 - El dispositivo físico.
 - ▶ Un controlador del dispositivo que interactúa con el SO mediante algún tipo de bus o registro.
- Los drivers son componentes de software muy específicos. Conocen las particularidades del HW contra el que hablan.



• Cuando enciendo la computadora, se carga el kernel del SO.

- Cuando enciendo la computadora, se carga el kernel del SO.
- Luego se cargan los distintos *módulos*, entre ellos, los drivers.

- Cuando enciendo la computadora, se carga el kernel del SO.
- Luego se cargan los distintos *módulos*, entre ellos, los drivers.
 - Administrador de memoria
 - Administrador de procesos (scheduler, etc.)
 - Sistema de archivos
 - Driver del teclado
 - Dirver del mouse
 - Driver de video
 - ▶ Driver de ...

- Cuando enciendo la computadora, se carga el kernel del SO.
- Luego se cargan los distintos *módulos*, entre ellos, los drivers.
 - Administrador de memoria
 - Administrador de procesos (scheduler, etc.)
 - ► Sistema de archivos
 - Driver del teclado
 - Dirver del mouse
 - Driver de video
 - ▶ Driver de ...



Necesito un driver nuevo.

- Necesito un driver nuevo.
 - ¿Ya existía el código en el kernel?
 - ▶ ¿Tengo que reiniciar la máquina y cargar de nuevo el kernel?
 - ▶ ¿Tengo que recompilar el kernel?

- Necesito un driver nuevo.
 - ; Ya existía el código en el kernel?
 - ▶ ¿Tengo que reiniciar la máquina y cargar de nuevo el kernel?
 - ▶ ¿Tengo que recompilar el kernel?





Solución (?)

Si el kernel está todo contenido en un gran archivo binario:

Solución (?)

Si el kernel está todo contenido en un gran archivo binario:

- ¿Qué pasa si quiero agregar funcionalidad cuando ya estoy usando la máquina?
- ¿Qué pasa si incluyo funcionalidad "por las dudas"?

Solución (?)

Si el kernel está todo contenido en un gran archivo binario:

- ¿Qué pasa si quiero agregar funcionalidad cuando ya estoy usando la máquina?
- ¿Qué pasa si incluyo funcionalidad "por las dudas"?

Solución: (de linux)

Linux soporta la carga y descarga de **módulos** al kernel en tiempo de ejecución.

• ¿Qué cosas componen a un **módulo**?

- ¿Qué cosas componen a un módulo?
 - ▶ Puntos de entrada y salida
 - Datos
 - Funciones

- ¿Qué cosas componen a un módulo?
 - Puntos de entrada y salida
 - Datos
 - Funciones
- ¿A causa de qué podría ejecutarse el código de un módulo?

- ¿Qué cosas componen a un módulo?
 - Puntos de entrada y salida
 - Datos
 - Funciones
- ¿A causa de qué podría ejecutarse el código de un módulo?
 - Llamada al sistema
 - Atención de interrupción

- ¿ Qué cosas componen a un módulo?
 - Puntos de entrada y salida
 - Datos
 - Funciones
- ¿A causa de qué podría ejecutarse el código de un módulo?
 - Llamada al sistema
 - Atención de interrupción
- ¿ Qué funcionalidades podría brindar un módulo?

- ¿Qué cosas componen a un módulo?
 - Puntos de entrada y salida
 - Datos
 - Funciones
- ¿A causa de qué podría ejecutarse el código de un módulo?
 - Llamada al sistema
 - Atención de interrupción
- ¿ Qué funcionalidades podría brindar un módulo?
- Hoy vamos a escribir nuestro primer módulo...

Un nuevo mundo...



- Estamos ejecutando en el nivel de máximo privilegio
- El kernel no está enlazado a la libc
- Hacer operaciones de punto flotante es más complicado
- Tenemos un stack fijo y limitado (y tenemos que compartirlo con el resto del kernel)
- Hay varias fuentes de posibles condiciones de carrera

Un nuevo mundo...



- Estamos ejecutando en el nivel de máximo privilegio
- El kernel no está enlazado a la libc
- Hacer operaciones de punto flotante es más complicado
- Tenemos un stack fijo y limitado (y tenemos que compartirlo con el resto del kernel)
- Hay varias fuentes de posibles condiciones de carrera
- ¿Qué pasa si hacemos un acceso indebido a memoria?



Nuestro primer módulo (1)

```
#include ux/init.h>
#include linux/module.h>
#include ux/kernel.h>
static int init hello init(void) {
    printk(KERN_ALERT "Hola, Sistemas Operativos!\n");
    return 0:
}
static void exit hello exit(void) {
    printk(KERN_ALERT "Adios, mundo cruel...\n");
}
module_init(hello_init);
module exit(hello exit):
MODULE_LICENSE("GPL");
MODULE AUTHOR("Juan de los Palotes"):
MODULE_DESCRIPTION("Una suerte de 'Hola, mundo'");
```

Nuestro primer módulo (2)

```
#include <linux/init.h>
#include <linux/module.h>
#include <linux/kernel.h>
```

- init.h contiene la definición de las macros module_init() y module_exit()
- module.h contiene varias definiciones necesarias para la gran mayoría de los módulos (por ejemplo, varios MODULE_*)
- kernel.h contiene la declaración de printk()

Nuestro primer módulo (3)

```
MODULE_LICENSE("GPL");
MODULE_AUTHOR("Juan de los Palotes");
MODULE_DESCRIPTION("Una suerte de 'Hola, mundo'");
```

- MODULE_AUTHOR() y MODULE_DESCRIPTION() son meramente informativos
- MODULE_LICENSE() indica la licencia del módulo;
 - algunos valores posibles son:
 - ★ GPI.
 - ★ Dual BSD/GPL
 - * Proprietary
 - un módulo con una licencia propietaria "mancha" el kernel

Nuestro primer módulo (4)

```
static int __init hello_init(void) {
   printk(KERN_ALERT "Hola, Sistemas Operativos!\n");
   return 0;
}
module_init(hello_init);
```

- static indica que la función es local al archivo (opcional)
- __init le indica al kernel que la función sólo se usará al momento de la inicialización, y que puede olvidarla una vez cargado el módulo (opcional)
- printk() se comporta de manera similar a la función printf() de la libc, pero permite indicar niveles de prioridad:
 - KERN_ALERT problema de atención inmediata
 - ► KERN_INFO mensaje con información
 - KERN_DEBUG mensaje de debug

Nuestro primer módulo (5)

```
static int __init hello_init(void) {
   printk(KERN_ALERT "Hola, Sistemas Operativos!\n");
   return 0;
}
module_init(hello_init);
```

- Con module_init() se indica dónde encontrar la función de inicialización del módulo
- La función de inicialización es llamada:
 - al arrancar el sistema
 - al insertar el módulo
- Su rol es registar recursos, inicializar hardware, reservar espacio en memoria para estructuras de datos, etc.
- Si todo salió bien, tiene que devolver 0; si no, tiene que volver atrás lo que cambió y devolver algo distinto de cero.

Nuestro primer módulo (6)

```
static void __exit hello_exit(void) {
   printk(KERN_ALERT "Adios, mundo cruel...\n");
}
module_exit(hello_exit);
```

- Con module_exit() se indica dónde encontrar la función de "limpieza" del módulo
- La función de "limpieza" es llamada antes de quitar el módulo
- Se ocupa de deshacer/limpiar todo lo que la función de inicialización y el resto del módulo usaron

EJERCICIO 1

Basándose en el siguiente ejemplo, escribir un módulo que al ser cargado reserve espacio en memoria para almacenar 9 caracteres y los inicialice con la cadena "A 0-0: $_{n}$ ".

```
#include <linux/init.h>
#include ux/module.h>
#include ux/kernel.h>
static int __init hello_init(void) {
    printk(KERN_ALERT "Hola, Sistemas Operativos!\n");
    return 0:
}
static void __exit hello_exit(void) {
    printk(KERN_ALERT "Adios, mundo cruel...\n");
}
module init(hello init):
module_exit(hello_exit);
MODULE LICENSE("GPL"):
MODULE_AUTHOR("Juan de los Palotes");
MODULE DESCRIPTION("Una suerte de 'Hola, mundo'"):
```

vim

- Abrir un archivo: vim archivo.txt
 - Si el archivo no existe, se crea.
- Pasar a modo INSERTAR: Tecla I.
- Salir del modo INSERTAR: Tecla Esc.
- Guardar (fuera del modo INSERTAR): :w .
- Salir (fuera del modo INSERTAR): :q .

vim

- Abrir un archivo: vim archivo.txt
 - ▶ Si el archivo no existe, se crea.
- Pasar a modo INSERTAR: Tecla I.
- Salir del modo INSERTAR: Tecla Esc.
- Guardar (fuera del modo INSERTAR): :w .
- Salir (fuera del modo INSERTAR): :q .
- Otro comando útil: Ctrl + Q.

Memoria dinámica (kernel)

Existen diversas formas de pedir memoria al sistema cuando estamos ejecutando en modo kernel. Nosotros vamos a ver dos:

kmalloc

Funciones: void * kmalloc(size_t size, int flags), kfree(void * ptr). Solicita un espacio de memoria **físicamente contiguo** de size bytes. Devuelve un puntero *virtual* accesible sólo en modo kernel.

vmalloc

Funciones: void * vmalloc(size_t size),vfree(void * ptr).
Solicita un espacio de memoria **virtualmente contiguo** de size bytes.
Devuelve un puntero *virtual* accesible sólo en modo kernel.

Para usarlas, incluir linux/slab.h.

Compilando nuestro módulo

- Necesitamos (en la vm ya está instalado).
 - make
 - ▶ module-init-tools
 - linux-headers-<version>
 (<version> sale de uname -r)
- 2 crear un Makefile con el siguiente contenido:

```
obj-m := hello.o
KVERSION := $(shell uname -r)
all:
    make -C /lib/modules/$(KVERSION)/build SUBDIRS=$(shell pwd) modules
clean:
    make -C /lib/modules/$(KVERSION)/build SUBDIRS=$(shell pwd) clean
```

3 ejecutar make clean y make

Ejecutando nuestro módulo

- Ya lo compilamos, ¿Cómo lo ejecutamos?
- ls
 - ▶ hello.c hello.ko hello.mod.c hello.mod.o hello.o Makefile modules.order Modules.symvers

Ejecutando nuestro módulo

- Ya lo compilamos, ¿Cómo lo ejecutamos?
- ls
 - hello.c hello.ko hello.mod.c hello.mod.o hello.o Makefile modules.order Modules.symvers
- Los módulos no los ejecutamos nosotros, se ejecutan:
 - 1 Al cargarlo en el sistema
 - Quando se ejecuta una llamada al sistema
 - Ouando se atiende una rutina de atención de interrupción
 - Al descargarlo del sistema

Ejecutando nuestro módulo

- Ya lo compilamos, ¿Cómo lo ejecutamos?
- ls
 - hello.c hello.ko hello.mod.c hello.mod.o hello.o Makefile modules.order Modules.symvers
- Los módulos no los ejecutamos nosotros, se ejecutan:
 - 1 Al cargarlo en el sistema
 - Quando se ejecuta una llamada al sistema
 - O Cuando se atiende una rutina de atención de interrupción
 - Al descargarlo del sistema
- Ok, carguemos el módulo en el sistema.

Inyectando módulos al kernel (1)

¿Cómo cargamos nuestro módulo al kernel?

- insmod carga el código y los datos de nuestro módulo al kernel
- el kernel usa su tabla de símbolos para enlazar todas las referencias no resueltas del módulo
- una vez cargado, se llama a su función de inicialización
- rmmod permite quitar el módulo del kernel si esto es posible (por ejemplo, falla si el módulo está siendo usado)
- modprobe es una alternativa más inteligente que insmod y
 rmmod (tiene en cuenta dependencias entre módulos)
- 1smod lista los módulos cargados

Inyectando módulos al kernel (2)

- sudo insmod hello.ko
- lsmod | grep hello
- sudo rmmod hello
- lsmod | grep hello

Inyectando módulos al kernel (2)

- sudo insmod hello.ko
- lsmod | grep hello
- sudo rmmod hello
- lsmod | grep hello

Genial, ya sabemos escribir, compilar y cargar módulos pero ¿qué pasó con los drivers?

Tipos de devices

Dijimos que los drivers se comunican con los dispositivos. ¿De qué tipo pueden ser esos dispositivos? En UNIX, comúnmente:

char devices

- pueden accederse como una tira de bytes
- suelen no soportar seeking
- se los usa directamente mediante un nodo en el filesystem
- tienen un subtipo interesante: misc devices

block devices

- direccionables de a "cachos" definidos
- suelen soportar seeking
- generalmente, su nodo es montado como un filesystem

network devices

- proveen acceso a una red
- no son accedidos a través de un nodo en el filesystem, sino de otra manera (usando sockets, por ejemplo)

devices y drivers

Con ls -1 /dev podemos ver los *drivers* del sistema.

```
lrwxrwxrwx 1 root root 3 2010-10-08 20:00 cdrom -> sr0
...
crw-rw-rw- 1 root root 1, 8 2010-10-08 20:00 random
...
brw-rw---- 1 root disk 8, 0 2010-10-08 20:00 sda
brw-rw---- 1 root disk 8, 1 2010-10-08 20:00 sda1
...
```

El primer caracter de cada línea representa el tipo de archivo:

- 1 es un symlink (enlace simbólico)
- c es un char device
- b es un block device

Los devices tienen un par de números asociados:

- major: está asociado a un driver en particular (primer número luego del grupo)
- minor: identifica a un dispositivo específico que el driver maneja (segundo número luego del grupo)

Construcción de un char device

- Conseguir los device numbers (el major y el minor)
- 2 Definir las funciones de cada operación del device
- 3 Inicializar el device como un char device
- Registrar el device como un char device
- Orear un nodo en el filesystem para interactuar con el device

Construcción de un char device

- ① Conseguir los device numbers (el major y el minor)
- 2 Definir las funciones de cada operación del device
- 3 Inicializar el device como un char device
- Registrar el device como un char device
- Orear un nodo en el filesystem para interactuar con el device

¿En qué parte del módulo se hace todo esto?

(1) Conseguir los device numbers

¿Cómo reservamos los device numbers que necesitamos?

- Asignamos uno específico (puede ser problemático)
- Pedimos al kernel que nos asigne uno dinámicamente

Para reservarlos y liberarlos dinámicamente tenemos:

```
int alloc_chrdev_region(dev_t *dev, unsigned int firstminor,
    unsigned int count, char *name);

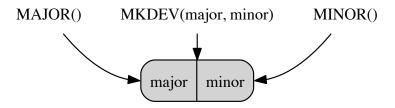
void unregister_chrdev_region(dev_t first, unsigned int count);
```

Recibe:

- dev_t *dev: parámetro de salida (una estructura en donde se van a guardar los device numbers)
- unsigned int firstminor: primer minor a ser usado (0)
- unsigned int count: cantidad de device numbers contiguos (1)
- char *name: nombre del device asociado al rango

Device numbers

En el código del kernel, el par major-minor se representa con el tipo dev_t:



- las macros MAJOR(dev_t dev) y MINOR(dev_t dev) nos dan cada número
- MKDEV(int major, int minor) nos da un dev_t para el par de números

```
#include<linux/fs.h>
// Device number
static dev_t hello_devno;
```

(2) Definir las operaciones (1)

- la estructura file_operations representa las operaciones que las aplicaciones pueden realizar sobre los devices
- cada campo apunta a una función en nuestro módulo que se encarga de la operación, o es NULL
- si el campo es NULL tiene lugar una operación por omisión distinta para cada campo

(2) Definir las operaciones (2)

- owner: un puntero al módulo "dueño" de la estructura (generalmente THIS_MODULE)
- read(): para recibir datos desde el device; retorna el número de bytes leídos
- write(): para enviar datos al device; retorna el número de bytes escritos

EJERCICIO 2

- Agregar al módulo funciones de lectura y escritura que sólo impriman mensajes con printk.
- Usar como base el siguiente ejemplo:

```
static struct file_operations mis_operaciones = {
    .owner = THIS_MODULE,
    .read = hello read.
    .write = hello_write,
};
ssize_t hello_read(struct file *filp, char __user *data,
 size t s. loff t *off) {
    . . . .
   return 0;
}
ssize_t hello_write(struct file *filp, const char __user *data,
 size_t s, loff_t *off) {
    return 0:
```

(3) Inicializar el char device

 el kernel representa internamente a los char devices mediante la estructura struct cdev

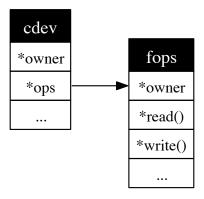
```
#include tinux/cdev.h>
struct cdev hello_dev;
```

- antes de que el kernel llame a nuestras operaciones, tenemos que inicializar y registrar al menos una de estas estructuras
- Llamar a la función cdev-init en el init

```
void cdev_init(struct cdev *cdev, struct file_operations *fops);
```

(4) Registrar el char device (1)





(4) Registrar el char device (2)

Ahora, registramos con:

```
int cdev_add(struct cdev *dev, dev_t num, unsigned int count);
```

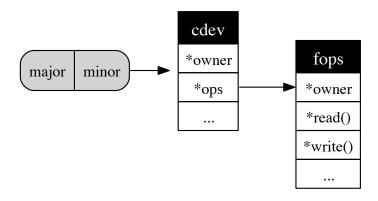
Tener en cuenta que:

- cdev_add() puede fallar
- si no falló, las operaciones de nuestro módulo ya pueden ser llamadas

Para quitar al char device del sistema, usar:

```
void cdev_del(struct cdev *dev);
```

(4) Registrar el char device (3)



Módulo terminado

Listo, ya creamos el módulo, definimos sus funciones y lo registramos como un driver de dispositivo.

Módulo terminado

Listo, ya creamos el módulo, definimos sus funciones y lo registramos como un driver de dispositivo.

Ahora ¿cómo probamos el código?

Módulo terminado

Listo, ya creamos el módulo, definimos sus funciones y lo registramos como un driver de dispositivo.

Ahora ¿cómo probamos el código?

Necesitamos una interfaz para que los programas de usuario puedan acceder al código.

(5) Crear nodos (1)

Tenemos, a priori, dos opciones:

- Crear los nodos, una vez se haya insertado el módulo, usando mknod <nodo> c <major> <minor> ,
- Que desde el módulo se genere algún tipo de aviso a alguien, en espacio de usuario, que se encargue de crear el nodo

Para lo segundo:

```
#include <linux/device.h>
static struct class *hello_class;
hello_class = class_create(THIS_MODULE, "hello");
device_create(hello_class, NULL, hello_devno, NULL, "hello");
```

```
device_destroy(hello_class, hello_devno);
class_destroy(hello_class);
```

(5) Crear nodos (2)

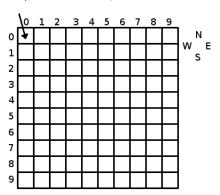
- La próxima vez que carguemos nuestro módulo, se generará un nuevo dispositivo virtual en /dev/.
- Podemos verlo con ls /dev/ | grep hello
 - El nombre con el que figura es el que le dieron al llamar a device_create, podría ser distinto al nombre del módulo.
- Para leer el dispositivo:
 - sudo head -n 1 /dev/hello
- Para escribir el dispositivo:
 - echo -n "1" | sudo tee /dev/hello

EJERCICIO 3

- Modificar el módulo para simular el driver de Husky. Se le pueden enviar comandos con write y leer su estado con read.
 - a Completar la función de inicialización para que:
 - Obtenga los device numbers
 - 2 Inicialice el device
 - Registre el device
 - 4 Cree un nodo para accederlo
 - b Modificar la función de escritura para que lea un caracter y modifique la ubicación moviendo al robot una posición en la dirección indicada.
 - c Modificar la función de lectura para que escriba la cadena de estado (reservada en la función de inicialización).

Movimiento del robot

Vamos a representar el mundo con una grilla de 10×10 . El robot inicia en la posición 0-0. Al recibir un comando debe moverse una posición en la dirección indicada (si es que puede). Los comandos válidos son N, S, E y W, correspondientes a los puntos cardinales en inglés. La cadena de estado contiene una A seguida de la posición del robot (coordenadas x e y separadas por un guión): "A 0-0: $_- \setminus n$ ".



Advertencia

- Tanto read() como write() escriben en o leen de la memoria de usuario
- El puntero a espacio de usuario puede:
 - ser inválido: puede no haber nada mapeado en esa dirección, o puede haber basura;
 - no estar en memoria (paginado), y el kernel no puede incurrir en page faults;
 - ser erróneo o malicioso
- Para estar tranquilos, hay que usar:

```
#include #include linux/uaccess.h>

unsigned long copy_to_user(void __user *to, const void *from,
    unsigned long count);
unsigned long copy_from_user(void *to, const void __user *from,
    unsigned long count);
```

EJERCICIO 4 (opcional)

- Modificar el módulo para jugar al buscaminas con Husky.
 - a Modificar la función de inicialización para que inicialice una matriz global de 10x10 con valores aleatorios entre 0 y 1 (ver get_random_bytes de linux/random).
 - b Modificar la función de lectura para que reemplace el _ con una 0 en la cadena de estado si hay una mina en su posición.
 - c Modificar la función de escritura para que, en lugar de leer un sólo caracter, lea la cantidad de caracteres escritos (de forma que se le puedan enviar varios comandos con echo -n "EEESSS"). Permitir además el comando D para cavar. En caso de haber una mina en la posición actual, el robot debe explotar y cambiar su estado (X en lugar de A). Tras haber explotado, debe ignorar todos los comandos recibidos.

Para pensar:

- ▶ El código que implementaron, ¿funciona si varios procesos acceden concurrentemente al driver? ¿qué habría que agregar/cambiar para que funcione?
- ¿Qué partes del código habría que reemplazar para usarlo como un driver de verdad en lugar de un simulador?

Sincronización (kernel)

Diversos mecanismos de sincronización. Entre ellos semáforos y mutexes.

sempahore

Tipo de datos: struct semaphore. Funciones: sema_init(struct semaphore * sem, int val), down(struct semaphore * sem), down_interruptible(struct semaphore * sem),...,up(struct semaphore * sem)

spinlock

Tipo de datos: spinlock_t Funciones: spin_lock_init(spinlock_t *
lock), spin_lock(spinlock_t * lock), spin_unlock(spinlock_t *
lock), etc.