Universidad Nacional Autónoma de México

Facultad de Ingeniería

Ingeniería en Computación

Sistemas Operativos

Proyecto 3: Asignación de memoria en un sistema real

Profesor: Gunnar Eyal Wolf Iszaevich

Alumnos:

Bárcenas Avelar, Jorge Octavio

Reza Chavarria, Sergio Gabriel

Proyecto 3: Asignación de memoria en un sistema real

Objetivo

Poder entender el modelo de memoria basada en la paginación, siendo que podamos

interpretar la información de los procesos, como en la identificación de secciones de

memoria, la cantidad de páginas utilizadas, entre otros conceptos.

Desarrollo

Parte 1

Herramientas utilizadas

En lenguaje de desarrollo utilizado fue Python (Versión 3.8.2). Se trabajaron en 2

sistemas operativos.

• Linux Mint 19.1

• Ubuntu 18.04.4

Además, se utilizó la biblioteca TKinter, utilizado para la creación de interfaces gráficas

en el lenguaje Python

Proceso realizado

Se utilizó como referencia el archivo obtenido del PID deseado que nos proporciona

Linux mediante el comando cat /proc/\${PID}/maps usando la biblioteca OS. Esto se hizo

para abrir y copiar la información de los archivos map, y smap originales.

Para las pruebas obtuvimos el PID usando el comando ls /proc directamente en una

terminal.

A partir del archivo creado se separa cada línea del archivo "mapsPID.txt" en una lista,

para ser después separarlo en las partes que necesitamos.

Para la separación se utilizó una clase para el guardado de los datos en instancias de esta.

Los atributos que la componen son el uso, la dirección completa en memoria (también a

partir de esto el inicio y el fin de la dirección), tamaño utilizado, número de páginas

utilizadas, los permisos que posee y el mapeo de esto.

La clase tiene 3 funciones. La función CalculoSIZEPag se utiliza para el cálculo del tamaño, a partir de inicio y fin de la dirección, junto con la asignación en Gb, Mb, o Kb correspondiente. La función calculoPaginas, a partir del tamaño obtenido, y después de una conversión a kb, se divide el tamaño entre 4 y se guardan las paginas correspondientes. Y la función de impresión que tiene el formato que se imprimirá.

Para saber qué tipo de Uso tiene el espacio se dio la revisión de 2 casos, el primero se necesitó revisar la salida de mapeo que nos proporcionaba el espacio de memoria, como ejemplo 'Stack', 'Heap', 'vvcall', entre otros. Para el segundo caso se necesitó de la revisión de los permisos de la memoria, esto ya que durante la clase correspondiente al tema y revisando una página de internet de la universidad politécnica de Madrid (revisar biografía), se hizo mención de que entre el heap y el stack existen espacios usados para el mapeo de bibliotecas las cuales cuentan con una estructura similar a la del proceso, teniendo espacio de texto y de datos, notando que el espacio de texto debe tener permisos de lectura y ejecución y el espacio de datos debe tener permiso de lectura.

El punto anterior refleja la lógica utilizada para la identificación de secciones en memoria vistas en clase y las regiones correspondientes a bibliotecas mapeadas.

Se dio uso de la biblioteca Tkinter, para interfaces. Ya que tenemos una interfaz gráfica para la impresión. Al ejecutar el proyecto se debe de ingresar un PID y enviarlo a partir del entrybox y el botton. El proceso de la obtención y guardado de datos se hace con respecto al PID ubicado. Al final se realiza la impresión de la información en el cuadro de texto separado por secciones. Además, se realiza un archivo llamado "newMap*PID*.txt" el cual igual guarda la información obtenida en formato.

Notas de Uso

Para ejecutar el programa debe primero haber instalado Tkinter. Para distribuciones de Linux, se debe ingresar el siguiente comando para instalar

Para esto fue necesario descargarlo. Se debe ingresar el comando en las distribuciones Linux.

• sudo apt-get install python3-tk (Python 3)

• sudo apt-get install python-tk (Python 2.7)

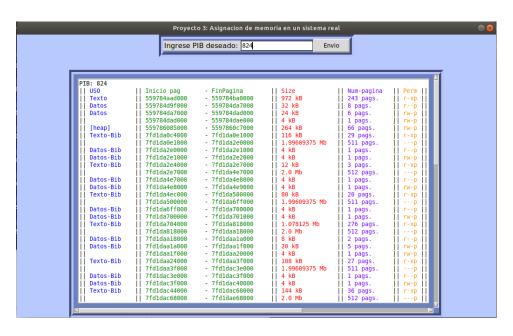
Para la ejecución del proyecto se utilizó el comando:

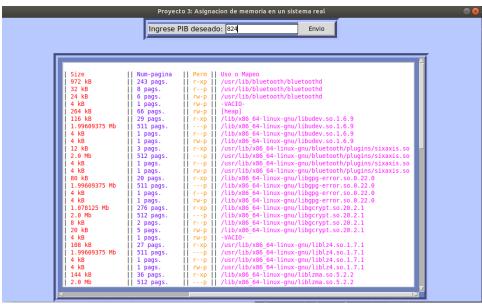
- python3 proyecto3.py
- sudo python3 proyecto3.py

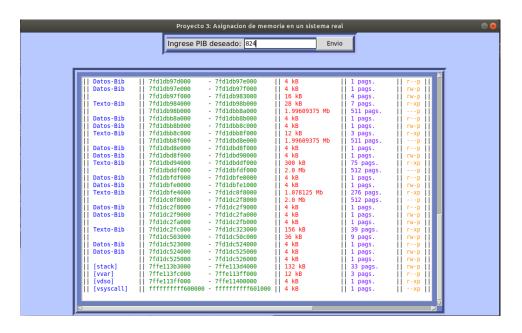
Al ingresar se desplegará una pantalla la cual tiene 2 partes, la primera es en donde se ingresará el PID deseado. Ya que se escribió el PID a considerar, se debe de oprimir el botón de envío. Al confirmar el PID se desplegará en la parte inferior de la ventana la información de Uso, dirección de memoria, tamaño, número de páginas, permisos y mapeo. Esto se puede hacer con diferentes con diferentes PID, hasta que se cierre la ventana. Todas las consultas se guardarán como "newMap*PID*.txt" correspondiente al PID buscado.

Pruebas

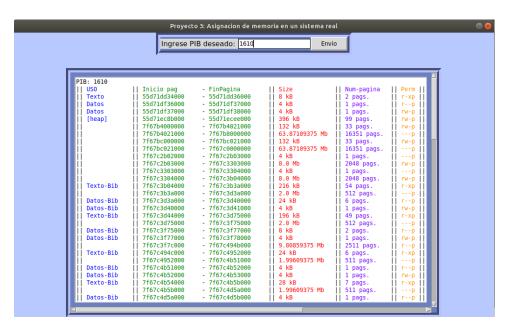
Prueba 1:

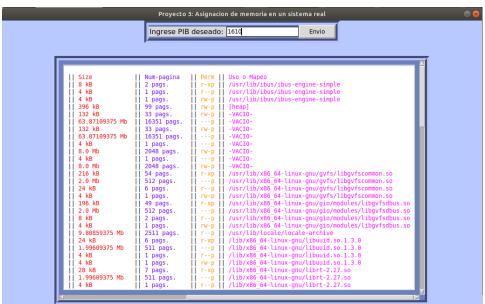


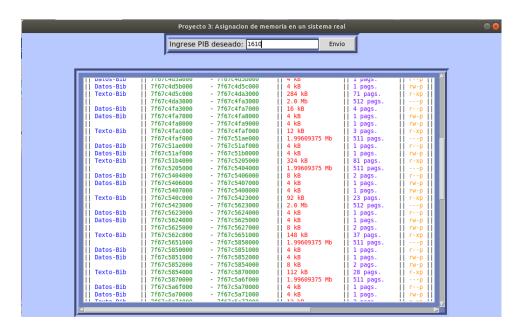


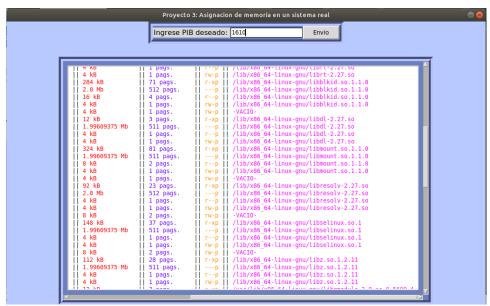


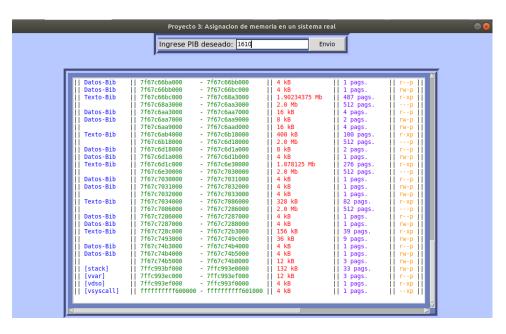
Prueba 2:

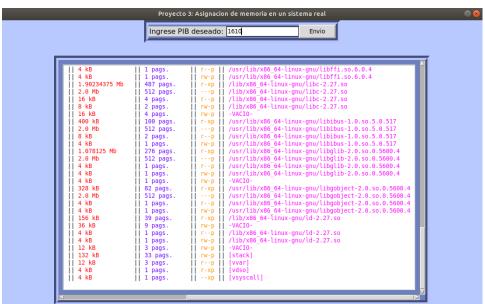












Parte 2

Herramientas utilizadas

Para la obtención del proceso realizado por el ejecutable y por el proceso en ejecución se utilizó el comando geore. Este comando se utiliza para la obtención de información de la memoria ocupada por un proceso junto con información de depuración, a partir de un id (PID).

• Gcore –a PID

Se utilizó el comando objdump, el cual nos da la información desde archivos binarios. El proceso que hace es desamblar el código binario, o sea que realiza la traducción de las instrucciones de binario a código ensamblador. A partir de esto se realizó un chequeo a la sección de memoria que proporcionaba también el comando.

• Objdump –s core.PID

El uso principal de hexdump es la examinación del código binario a ensamblador, esta herramienta la usamos al inicio para identificar que contenían las secciones usadas de memoria. Solo que lo usamos para detectar de manera relativa la memoria.

• Hexdump –s core.PID

También se utilizó el comando nm, el cual lista los símbolos de los archivos de los objfiles, bibliotecas y ejecutables. El comando especifica el tamaño de bytes, el tipo de objeto, ámbito y nombre. Junto con sysv para el formato de salida

• Nm –f sysv archivo

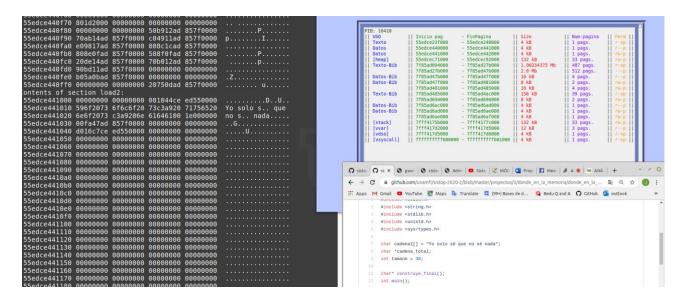
Las secciones de memoria descritas en la información que nos proporcionó pertenecen a la sección del formato de archivo ELF (Executable and Linkable Format). ELF es un formato de archivos para el uso de ejecutables, código objeto, bibliotecas compartidas. Consiste en un encabezado, el cual contiene la información general del ejecutable e indican las posiciones de tablas de encabezados de sección. Estas secciones son ubicadas como

- .bss: Coniene datos que no son inicializados
- .data: Contiene datos inicializados
- .debug: Contiene información para la depuración

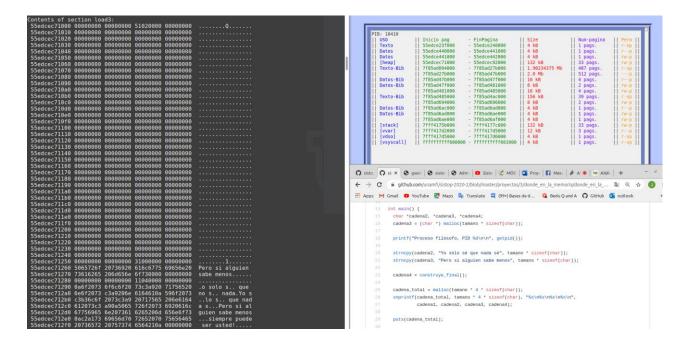
- dynamic: Contiene información para el enlazador dinámico
- .dynsym: Contiene tabla de símbolos para el enlazador
- rodata: Contiene información de solo lectura
- symtab: Contiene una tabla de símbolos
- .text: Contiene la parte ejecutable de un programa

Proceso

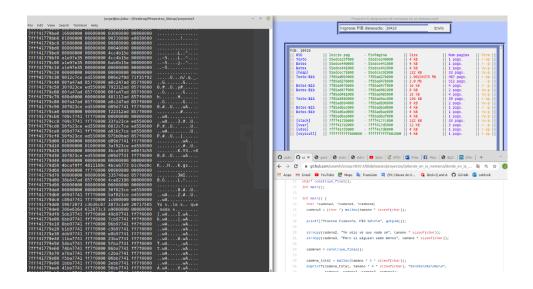
A partir de la obtención del archivo ejecutable y del archivo generado core.PID se revisaron las secciones de memoria. Siendo estos archivos leídos a partir de las herramientas objdump y hexdump. Haciendo uso de nuestro programa se hizo la relación con las secciones de memoria en donde encontrábamos información de las cadenas usadas en el archivo "donde_en_la_memoria.c", y a su vez identificar el uso que se le daba a estas secciones.



Encontramos que en la sección load2 se encuentra el contenido de la variable global "cadena1" y la dirección de memoria, según nuestro programa, es perteneciente a la sección de Datos.



Encontramos que en la sección load3 se encuentra primero el contenido de la "cadena3" y pensamos que es debido a que se reserva primero el espacio de memoria para la "cadena3" y, según la dirección de memoria, se encuentra en la sección del heap.

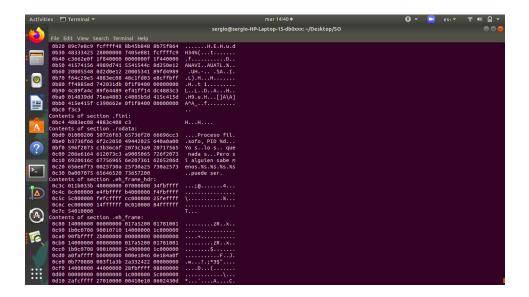


Encontramos el contenido de "Cadena2" en una dirección de memoria perteneciente al stack, debido a que "cadena2" no fue liberada usando "free".

El heap es una estructura dinámica de datos que se utiliza para el almacenamiento de datos en ejecución, en donde se permite reservar memoria de manera dinámica. Las variables globales y estáticas también son almacenadas en esta sección.

Mientras que el stack es una estructura que almacena la información de las subrutinas activas de un programa de ejecución. Cuando se hace una llamada a función, un bloque en esta sección es reservado para guardar las variables locales junto con otros datos para garantizar el funcionamiento.

La obtención del texto que se imprimirá con "printf o snprintf" lo pudimos encontrar al utilizar el archivo ejecutable con objdump. Esta parte se encontraba en la sección .rodata.



Los nombres de las funciones y de las variables globales utilizadas en el archivo trabajado se encontraban en las secciones de memoria que nos proporcionó ELF.

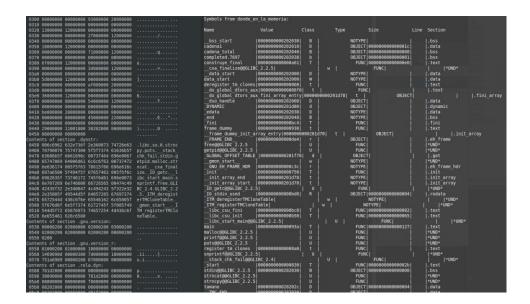
i0 i-b (D	l.+ /B+ Si+			£ danda an 1- manada	
]orge@pc-joba:~/Des	Ktop/Proyectos_Sisto	p/proyectos\$	nm -	f sysv donde_en_la_memoria	
Symbols from donde_en_la_memoria:					
			- F		
Name	Value C	lass	Туре	Size L	ine Section
bss start	[0000000000202030]	ВІ		NOTYPE	l.bss
cadena1	0000000000202010	D		OBJECT 0000000000000001c	.data
cadena total	0000000000202040	В		OBJECT 000000000000000008	.bss
completed.7697	0000000000202038			OBJECT 00000000000000001	.bss
construye_final	0000000000000a61			FUNC 000000000000000e0	.text
cxa_finalize@GLI				FUNC	*UND*
data_start	00000000002020000	D		NOTYPE	.data
data_start deregister tm clone	000000000002020000	W		NOTYPE FUNC	.data .text
	ux 000000000000000000000000000000000000	t		FUNC	.text
		999999999999999999999999999999	1d781	t OBJECT	.fini array
dso_handle	[0000000000202008]			OBJECT	.data
DYNAMIC	0000000000201d80			OBJECT	.dynamic
_edata	0000000000202030			NOTYPE	.data
_end	0000000000202048	B		NOTYPE	l.bss
fini	0000000000000bc4	FIF F600000		FUNC 601000 4 kB	.fini
frame_dummy	0000000000000930 array entry 00000000	t aasa1d7al ·	t I	FUNC OBJECT	.text .init array
FRAME END	0000000000000000000000000000000000	r		OBJECT!	l.eh frame
free@@GLIBC 2.2.5		υİ		FUNC	*UND*
getpid@@GLIBC 2.2.5				FUNC	*UND*
GLOBAL OFFSET TABL	E_ 0000000000201f70			OBJECT	, got
gmon_start				NOTYPE	*UND*
GNU_EH_FRAME_HDR	0000000000000c3c	ī ļ		NOTYPE	.eh_frame_hdr
_init	0000000000000750			FUNC	.init
init_array_end init array start	0000000000201d78 0000000000201d70	t		NOTYPE NOTYPE	.init_array .init_array
IO getc@GGLIBC 2.2		ַט 'ו		FUNC!	*UND*
IO stdin used	[0000000000000bd0]	R I		OBJECT 000000000000000004	.rodata
TITM deregisterTMCl				NOTYPE	' *UND*
_ITM_registerTMClon				NOTYPE	*UND*
libc_csu_fini	000000000000bc0	I I		FUNC 000000000000000002	.text
libc_csu_init	0000000000000b50			FUNC 00000000000000065	.text
libc_start_main@@ main		ਾ। ''	n l	FUNC FUNC 000000000000127	*UND* .text
malloc@@GLIBC 2.2.5		~;``		FUNC	*UND*
printf@@GLIBC 2.2.5		ŭ		FUNC	*UND*
puts@@GLIBC_2.2.5				FUNC	*UND*
register_tm_clones	00000000000008a0			FUNC	i.text
snprintf@GLIBC_2.2				FUNC	*UND*
stack_chk_fail@@G				FUNC	*UND*
_start stdin@@GLIBC 2.2.5	0000000000000830 0000000000202030	T B		FUNC 0000000000000002b OBJECT 0000000000000008	.text .bss
strncat@GLIBC_2.2.3		Ü		FUNC	*UND*
strncpy@@GLIBC 2.2.		uŭ filtere		O FUNC	000 *UND* kB
tamano	000000000020202c			OBJECT 00000000000000004	.data
TMC_END	0000000000202030	D		OBJECT	.data

Estas direcciones se obtuvieron del comando nm –f sysv donde_en_la memoria. Este comando nos proporcionó la dirección, el tipo de dato, la clase perteneciente, el tamaño y la sección de memoria en donde se localizaba.

Una de las razones por las que los nombres de las variables de las funciones no son encontrados en esta tabla es debido a que en el proceso de compilación se utilizan nuevas tablas de símbolos, siendo creadas a partir de la declaración de funciones. Estas nuevas tablas ayudan al análisis de la misma función y cuando este proceso termina la tabla de símbolos de las funciones ya no es utilizada.

```
16020404 010faef0 0faee80f
                               0fc7f847
   f417d5950
             e80faef0 0faee8f3
                                         43433a20
                                                   (Ubuntu 7.4.0-1u
             28556275 6e747520
                               372e342e
            62756e74
                                                   buntu1~18.04.1)
7fff417d5970
                      75317e31
                               382e3034
                                         2e312920
7fff417d5980
             372e342e
                      3000002e
                               73687374
                                         72746162
                                                   7.4.0...shstrtab
7fff417d5990 002e676e
                      752e6861
                               7368002e 64796e73
                                                   ..gnu.hash..dyns
             796d002e 64796e73
                               7472002e
7fff417d59a0
                                        676e752e
                                                   ym..dynstr..gnu.
             76657273 696f6e00
7fff417d59b0
                               2e676e75
                                                   version..gnu.ver
7fff417d59c0
             73696f6e 5f64002e
                               64796e61
                                         6d696300
                                                   sion d..dynamic.
7fff417d59d0
             2e6e6f74 65002e65
                               685f6672
                                         616d655f
                                                   .note..eh_frame
7fff417d59e0
            68647200
                      2e65685f
                               6672616d
                                        65002e74
                                                   hdr..eh_frame..t
7fff417d59f0 65787400
                      2e616c74
                               696e7374
                                                   ext..altinstruct
7fff417d5a00
            696f6e73
                      002e616c
                               74696e73
                                         74725f72
                                                   ions..altinstr r
7fff417d5a10 65706c61 63656d65
                               6e74002e
                                         636f6d6d
                                                   eplacement..comm
7fff417d5a20 656e7400 00000000 00000000
                                        00000000
7fff417d5a30
            0000000
                      0000000
                               0000000
                                         0000000
7fff417d5a40 00000000 00000000 00000000
                                        00000000
7fff417d5a50 00000000 00000000 00000000
                                        0000000
7fff417d5a60 00000000 00000000
                               0f000000
                                         05000000
7fff417d5a70 02000000 00000000 20010000 00000000
7fff417d5a80 20010000 00000000 3c000000 00000000
```

También pudimos encontrar información de la distribución del sistema operativo en donde se está ejecutando el programa. Siendo que esta información está en la sección definida como vvar.



Otra sección de memoria que encontramos con información fue en la sección .dynstr la cual contenía todos los nombres de las funciones que son utilizadas en el programa, siendo desde las funciones utilizadas como free, gets, nsprintf entre otros.

También en la tabla de símbolos existen procesos o funciones que ayudan al proceso de ejecución, poniendo como ejemplo __bss_start o __data_start. Estos proporcionados por el enlazador. Esto ayuda a la ubicación de la sección de memoria correspondiente.

Conclusiones

Para este proyecto fue necesario hacer uso de los conocimientos de procesos y de memoria para el análisis de procesos reales lo cual fue una tarea un tanto compleja para nosotros, pero que a su vez fue un buen ejercicio para poder "jugar" con cosas más reales e ir identificando la estructura básica de la distribución de la memoria.

Bibliografía

- Tkinter (2020) *Tkinter Python interface to Tcl/Tk* Consultado en: https://docs.python.org/2/library/tkinter.html
- Universidad Politécnica de Madrid (2020) Encontrado el 1 de mayo, 2020, en https://laurel.datsi.fi.upm.es/docencia/asignaturas/sox/prv/practicas/analisis_so5/m_odulo_gm?fbclid=IwAR2thIiUPSdCG8KarmwzNb8myZBL54vLoMFr8-IsYRz0XZVZquHSzTIOvb8
- Linuxito (2019) Desensamblar binarios con objdump. Consultado el 3 de mayo,
 2020, en https://www.linuxito.com/seguridad/1302-desensamblar-binarios-con-objdump
- Man7.org (2013). Linux/UNIX system programming training. Consultado el 3 de mayo, 2020, en http://man7.org/linux/man-pages/man1/hexdump.1.html
- Tecnonautas (2020. *Entender el comando Linux/Unix: nm*. Consultado el 3 de mayo, 2020, en https://tecnonautas.net/entender-el-comando-linux-unix-nm/
- Coding or not (2020) Diferencias entre heap y stack. Consultado el 4 de mayo, 2020, en https://codingornot.com/diferencias-entre-heap-y-stack?fbclid=IwAR12O2G1QcOY-OODeV77bJJEkc_T_IrSkF1AO_SPFS-ggLhimjEgRFjNFKU
- Hofmann F. (2019) Understanding the ELF File Format. Consultado el 6 de mayo,
 2020, en https://linuxhint.com/understanding_elf_file_format/
- C. M. Linn, M. Rajagopalan, S. Baker, C. Collberg, S. K. Debray, J. H. Hartman (2019) *Protecting Against Unexpected System Calls*. University of Arizona, Departament of Computer Science Texas. USA. Recuperado el 6 de mayo, 2020, de: https://www.usenix.org/legacy/event/sec05/tech/full_papers/linn/linn.pdf