

# Labs 9: ELF - Introduction with Code

This lab may be done either solo or in pairs.

In the previous lab, you learned to investigate and change ELF files using *hexedit*, and other command-line tools. In this lab, you will continue to manipulate ELF files, this time using your own code (written in C).

We will parse the ELF file and extract useful information from it. In particular, we will access the data in the section header table, and in the symbol table. We will also learn to use the `mmap` system call.

## Important

This lab is written for 32-bit machines. Some of the computers in the labs already run on a 64-bit OS (use *uname -a* to see if the linux OS is 64-bit or not). 32-bit and 64-bit machines have different instruction sets and different memory layout. Make sure to include the `-m32` flag when you compile files, and to use the Elf32 data structures (and not the Elf64 ones).

In order to know if an executable file is compiled for 64-bit or 32-bit platform, you can use *readelf*, or the *file* command-line tool (for example: `file /bin/ls`).

## Useful Tips

You will no longer be using *hexedit* to process the file and strings to find the information; nevertheless, in some cases you may still want to use these tools for debugging purposes. In order to take advantage of these tools and make your tasks easier, you should:

- Print debugging messages: in particular the offsets of the various items, as you discover them from the headers.
- Use *hexedit* and *readelf* to compare the information you are looking for, especially if you run into unknown problems. *hexedit* is great if you know the exact location of the item you are looking for.
- Note that while the object files you will be processing will be linked using *ld*, and will, in most cases, use direct system calls in order to make the ELF file simpler, there is no reason why the programs you write need use this interface. You are allowed to use the standard library when building your own C programs.
- In order to preserve your sanity, even if the code you MANIPULATE may be without `stdlib`, we advise that for your OWN CODE you DO use the C standard library!
- In order to keep sane in the following lab as well, **understand** what you are doing and **keep track** of that and of your code, as you will be using them in a future lab.

## Lab 9 Tasks

You must use only the `mmap` system call to read/write data into your ELF files from this point onwards.

### Task 0

This task is about learning to use the `mmap` system call. Read about the `mmap` system call (`man mmap`).

Write a program that uses the `mmap` to examine the header of a 32-bit ELF file (include and use the structures in `elf.h`). The program is first activated as:

```
myELF
```

The program then uses a menu similar to lab 8, with available operations, as follows:

Choose action:

- 0-Toggle Debug Mode
- 1-Examine ELF File
- 2-Print Section Names
- 3-Print Symbols
- 4-Relocation Tables
- 5-Quit

Note that the menu should use the same technique as in lab 2 and in lab 8, i.e. an array of structures of available options. Toggle Debug Mode is as in Lab 8. Quit should unmap and close any mapped or open files, and "exit normally". Examine ELF Files queries the user for an ELF file name to be used and examined henceforth. For now, options 2, 3, 4, should call stub functions that print "not implemented yet". All file input should be read using the `mmap` system call. You are NOT ALLOWED to use `read`, or `fread`.

To make your life easier throughout the lab, map the entire file with one `mmap` call.

In Examine ELF File, after getting the file name, you should close any currently open file (indicated by global variable `Currentfd`) open the file for reading, and then print the following:

1. Bytes 1,2,3 of the magic number (in ASCII)
2. Entry point (in hexadecimal)

Check using `readelf` that your data is correct.

Once you verified your output, extend *examine* to print the following information from the header:

1. Bytes 1,2,3 of the magic number (in ASCII). Henceforth, you should check that the number is consistent with an ELF file, and refuse to continue if it is not.
2. The data encoding scheme of the object file.
3. Entry point (hexadecimal address).
4. The file offset in which the section header table resides.
5. The number of section header entries.
6. The size of each section header entry.
7. The file offset in which the program header table resides.
8. The number of program header entries.
9. The size of each program header entry.

The above information should be printed in the above exact order (print it out as nicely as `readelf` does). If invoked on an ELF file, examine should initialize a global file descriptor variable `Currentfd` for this file, and leave the file open. When invoked on a non-ELF file, or the file cannot be opened or mapped at all, you should print an error message, unmap the file (if already mapped) close the file (if already open), and set `Currentfd` to -1 to indicate no valid file. You probably also should use a global `map_start` variable to indicate the memory location of the mapped file.

## Task 1 - Sections

Extend your myELF program from Task 0 to allow printing of all the Section names in an 32-bit ELF file (like `readelf -S`). That is, modify the menu to add a "Print Section Names" option.

Print Section Names should visit all section headers in the section header table, and for each one print its index, name, address, offset, size in bytes, and type number. Note that this is done for the file currently mapped, so if current fd is invalid, just print an error message and return.

The format should be:

```
[index] section_name section_address section_offset section_size section_type
[index] section_name section_address section_offset section_size section_type

[index] section_name section_address section_offset section_size section_type

....
```

Verify your output is correct by comparing it to the output of *readelf*. In debug mode you should also print the value of the important indices and offsets, such as *shstrndx* and the section name offsets.

You can test your code on the following file: [a.out](#)

### Hints

Global information about the ELF file is in the ELF header, including location and size of important tables. The size and name of the sections appear in the section header table. Recall that the actual name **strings** are stored in an appropriate **section** (*.shstrtab* for section names), and not in the section header!

## Task 2 - Symbols

Extend your myELF program from task 1 to support an option that displays information on all the symbol names in a 32-bit ELF file.

The new Print Symbols option should visit all the symbols in the current ELF file (if none, print an error message and return). For each symbol, print its index number, its name and the name of the section in which it is defined. (similar to *readelf -s*). Format should be:

```
[index] value section_index section_name symbol_name
[index] value section_index section_name symbol_name
[index] value section_index section_name symbol_name

...
```

Verify your output is correct by comparing it to the output of *readelf*. In debug mode you should first print the size of each symbol table, the number of symbols therein, and any other useful information.

### Hints:

Symbols are listed in the designated sections. The section in which a symbol is defined (if it is defined) is the index of the symbol, which is an index into the section header table, referring to the section header of the appropriate section, and from there the section name can be retrieved as above. Symbol name is an attribute of the symbol structure, but recall again that the actual name string is stored in a string table, a separate section (*.strtab*).

You should finish everything up to here during the lab. The rest can be done in a completion lab, if you run out of time.

## Task 3 - Relocations

### Task 3a: Relocation Table(s) - raw format

Print the content of all fields of all relocation tables entries, in hexadecimal format.

This is similar to what `readelf -r` prints, except this feature prints the raw table data (i.e. without fetching symbol name strings).

### Hints

To get the relocation table you should look for section of type 'SHT\_REL' and use the offset field. Every entry in this table is of type 'Elf32\_Rel', which has the fields `offset` and `info`. Using the macro 'ELF32\_R\_SYM' on the `info` field gives you the raw data for the symbol (which is an index of the symbol in the `.dynsym` table where you can find its name and value). Using the 'ELF32\_R\_TYPE' macro on the `info` field gives you the type of this relocation entry.

## Task 3b: Relocation Table(s) - semantic format

You should extend myELF to support this feature:

Now that you have accessed and printed the relocation table, add a printout that displays the entries with the correct semantics, that is, indicate relocation type, symbol used for relocation, etc. The **output should be similar to that generated by `readelf -r` (relocation type should be printed as a number. There's no need to translate it to a string).**

To get the relation information regarding values, you need to first get the symbol's index (using `ELF32_R_SYM` on the `info` field of the relocation entry). You use that to index into the `.dynsym` table and get the symbol entry. The value is stored in that table. The name of the symbol is extracted from the `.dynstr` table similarly to what you did in task 2. For example, using `ntsc` from lab 8 (`readelf -r ntsc`):

Relocation section '.rel.dyn' at offset 0x340 contains 8 entries:

Offset	Info	Type	Sym.Value	Sym. Name
00001ed4	00000008	8		
00001ed8	00000008	8		
00001ff8	00000008	8		
00002004	00000008	8		
00001fec	00000106	6	00000000	_ITM_deregisterTMClone
00001ff0	00000306	6	00000000	__cxa_finalize@GLIBC_2.1.3
00001ff4	00000406	6	00000000	__gmon_start__
00001ffc	00000706	6	00000000	_ITM_registerTMCloneTa

Relocation section '.rel.plt' at offset 0x380 contains 3 entries:

Offset	Info	Type	Sym.Value	Sym. Name
00001fe0	00000207	7	00000000	printf@GLIBC_2.0
00001fe4	00000507	7	00000000	exit@GLIBC_2.0
00001fe8	00000607	7	00000000	__libc_start_main@GLIBC_2.0

## Deliverables:

Tasks 1 and 2 must be completed during the regular lab. Task 3 may be done in a completion lab, but only if you run out of time during the regular lab. The deliverables must be submitted until the end of the lab session.

You must submit source files for tasks 1, 2 and task 3 and a makefile that compiles them. The source files must be named `task1.c`, `task2.c`, `task3.c`, and `makefile`.