



## THE ELF FILE FORMAT

STEFANO DI CARLO

## **ELF FILE FORMAT**

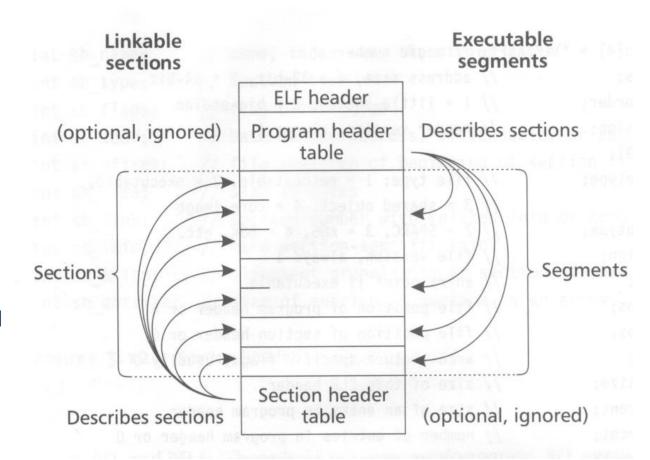


- ► The a.out (<a href="https://en.wikipedia.org/wiki/A.out">https://en.wikipedia.org/wiki/A.out</a>) format served the Unix community well for over 10 years.
- ► However, to <u>better support cross-compilation</u>, dynamic linking, initializer/finalizer (e.g., the constructor and destructor in C++) and other advanced system features, a.out has been replaced by the <u>elf</u> file format.
- ► Elf stands for "Executable and Linking Format."
- ► Elf has been adopted by <u>FreeBSD and Linux</u> as the current standard.

#### **ELF STRUCTURE DUAL NATURE**



- Compilers, assemblers, and linkers treat the file as a set of <u>logical sections</u> described by a section header table.
- The system loader treats the file as a set of <u>segments</u> described by a program header table.
  - A single segment usually consist of several sections. E.g., a loadable read-only segment could contain sections for executable code, read-only data, and symbols for the dynamic linker.
- Relocatable files have section header tables. Executable files have program header tables. Shared object files have both.
- Sections are intended for further processing by a linker, while the segments are intended to be mapped into memory.



#### **ELF FILE TYPES**



- Elf defines the format of executable binary files.
- ► There are four different types:
  - ▶ **Relocatable:** created by compilers or assemblers. Need to be processed by the linker before running.
  - ► <u>Shared object:</u> shared library containing both symbol information for the linker and directly runnable code for run time.
  - **Executable:** have all relocation done and all symbol resolved except perhaps shared library symbols that must be resolved at run time.
  - Core file: a core dump file.

#### **RELOCATABLE FILES**



- A relocatable or <u>object file</u> is a collection of sections.
- ► Each section contains a single type of information, such as:
  - program code
  - read-only data,
  - read/write data
  - relocation entries: records used to adjust addresses and symbols in the program during linking
  - Symbols, i.e., a description of variables or functions stored in the ELF file including simple information such as size, value, name, etc.
- Every address is defined relative to a section
  - ► Therefore, a procedure's entry point is relative to the program code section that contains that procedure's code.

#### **SHARED OBJECTS**



- A shared object, typically with a .so extension (e.g., libxyz.so), is a file that contains code and data intended to be shared by multiple programs at runtime.
- It is commonly used for dynamic linking, where a program can load and link to these shared libraries either at load time or during execution.
- ► The key characteristics are:
  - ▶ It contains relocatable code, allowing it to be loaded at different memory addresses.
  - ▶ It includes exported symbols (functions or variables) that can be referenced by other programs.
  - ▶ It reduces memory usage and binary size by sharing common code across multiple programs.

#### **EXECUTABLES**



- An executable is very similar to a shared object
  - ► It can be loaded at a specific address in memory.
  - ▶ It has a function that is called when a program starts.
  - \_start() run first in an executable.

## **CORE FILES**



- An ELF core file is a special type of file that contains a snapshot of a program's memory and execution state at the moment it crashes or is terminated unexpectedly.
- It is typically generated by the operating system when a program encounters a serious error, such as a segmentation fault.
- Key features of an ELF core file:
  - Process State: It includes the program's memory (stack, heap, data segments) and processor registers at the time of the crash.
  - **Debugging Use:** Core files are used for post-mortem debugging, allowing developers to analyze the state of the program when it crashed to determine the cause of the error.
  - ► Core files are often analyzed using debugging tools like gdb to trace the cause of a crash and diagnose software bugs.

## **ELF LINKING PROCESS**





Section 1 Data

Section 2 Data

•••

Section n Data

**Section-Header Table** 

Object File

**ELF Header** 

Section 1 Data

Section 2 Data

•••

Section n Data

**Section-Header Table** 

Object File

Program-Header Table

**ELF Header** 

Segment 1 Data

Segment 2 Data

Segment n Data

#### **ELF HEADER**



- The Elf header is always at offset zero of the file.
- The program header table and the section header table's offset in the file are defined in the ELF header.
- The elf format can support two different address sizes:
  - > 32 bits
  - ► 64 bits

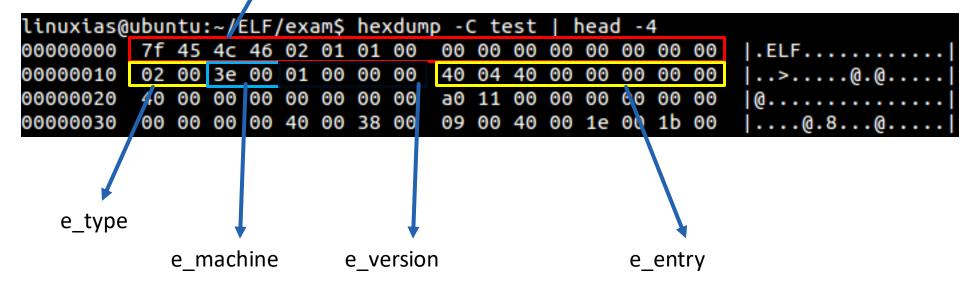
```
vpedef struct
unsigned char e ident[EI NIDENT]; /* Magic number and other info */
Elf64 Half
                                   /* Object file type */
              e type;
Elf64 Half
              e machine:
                                   /* Architecture */
Elf64 Word
                                   /* Object file version */
              e version;
                                   /* Entry point virtual address */
Elf64 Addr
              e entry;
Elf64 Off e phoff;
                                   /* Program header table file offset */
                                   /* Section header table file offset */
Elf64 Off e shoff;
Elf64 Word
              e_flags;
                                   /* Processor-specific flags */
                                   /* ELF header size in bytes */
Elf64_Half
              e_ehsize;
Elf64 Half
                                   /* Program header table entry size */
              e phentsize;
Elf64 Half
                                   /* Program header table entry count */
              e phnum;
Elf64 Half
              e shentsize:
                                   /* Section header table entry size */
Elf64 Half
              e shnum;
                                   /* Section header table entry count */
Elf64 Half
              e shstrndx;
                                   /* Section header string table index *
Elf64 Ehdr;
```



e\_ident

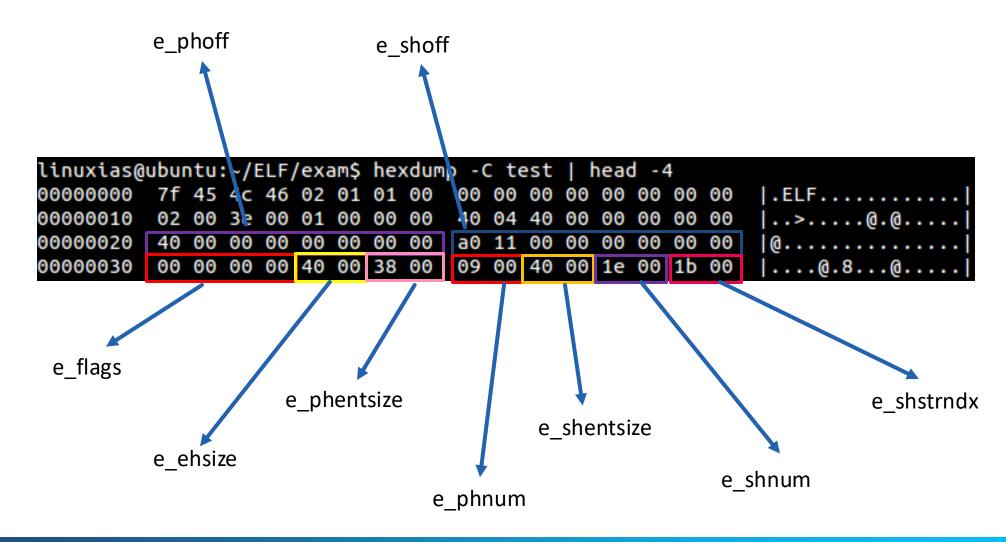


Name	Value	Purpose
EI_MAG0	0	File identification
EI_MAG1	1	File identification
EI_MAG2	2	File identification
EI_MAG3	3	File identification
EI_CLASS	4	File class
EI_DATA	5	Data encoding
EI_VERSION	6	File version
EI_PAD	7	Start of padding bytes
EI_NIDENT	16	Size of e_ident[]













variable	Size(byte)	Data(Hex [Dex])
e_ident	16	7f 45 4c 46 02 01 01 00
e_type	2	2 [2]
e_machine	2	3e [64]
e_version	4	1 [1]
e_entry	8	40 04 40 [4,195,392]
e_phoff	8	40 [64]
e_shoff	8	11 a0 [4512]
e_flags	4	0 [0]
e_ehsize	2	40 [64]
e_phentsize	2	38 [56]
e_phnum	2	9 [9]
e_shentsize	2	40 [64]
e_shnum	2	1e [30]
e_shstrndx	2	1b [27]

	inuxias@ubuntu:~/ELF/exam\$ readelf ELF Header:	-h test
	Magic: 7f 45 4c 46 02 01 01 00 0	0 00 00 00 00 00 00 00
1	Class:	ELF64
	Data:	2's complement, little endian
	Version:	1 (current)
	OS/ABI:	UNIX - System V
	ABI Version:	0
Ì	Type:	EXEC (Executable file)
	Machine:	Advanced Micro Devices X86-64
	Version:	0×1
	Entry point address:	0x400440
	Start of program headers:	64 (bytes into file)
	Start of section headers:	4512 (bytes into file)
I	Flags:	0x0
	Size of this header:	64 (bytes)
	Size of program headers:	56 (bytes)
	Number of program headers:	9
	Size of section headers:	64 (bytes)
	Number of section headers:	30
	Section header string table index:	27





- Contains information about every part of an ELF file (except the ELF Header, Program Header Table, Section Header Table itself).
- List of section header structures, each defining a different section in the ELF file.

```
typedef struct
                           /* Section name (string tbl index) */
 Elf64 Word
              sh_name;
 Elf64 Word
              sh_type;
                             /* Section type */
 Elf64 Xword
              sh_flags;
                             /* Section flags */
 Elf64 Addr
              sh addr;
                             /* Section virtual addr at execution */
 Elf64 Off sh offset;
                             /* Section file offset */
                             /* Section size in bytes */
 Elf64_Xword
              sh_size;
                             /* Link to another section */
 Elf64 Word
              sh link;
                             /* Additional section information */
 Elf64 Word
              sh info;
                                 /* Section alignment */
              sh_addralign;
 Elf64_Xword
              sh_entsize; /* Entry size if section holds table */
 Elf64_Xword
 Elf64 Shdr;
```

#### **SECTION HEADER TABLE**



▶ ELF header contains the file offset of the section header table

```
linuxias@ubuntu:~/ELF/exam$ readelf -h test.o
ELF Header:
 Magic: 7f 45 4c 46 02 01 01 00 00 00 00 00 00 00 00 00
 Class:
                                     ELF64
                                     2's complement, little endian
 Data:
 Version:
                                     1 (current)
 OS/ABI:
                                     UNIX - System V
 ABI Version:
                                     REL (Relocatable file)
 Type:
                                     Advanced Micro Devices X86-64
 Machine:
 Version:
                                     0x1
  Entry point address:
 Start of section headers:
                                     384 (bytes into file)
 rtays:
 Size of this header:
                                     64 (bytes)
 Size of program headers:
                                     0 (bytes)
 Number of program headers:
 Size of section headers:
                                     64 (bytes)
 Number of section headers:
 Section header string table index: 10
```



#### SECTIONS AND THE SECTION HEADER TABLE

File offset 0x140 (.data section offset 0x120 + list value offset 0x20)

```
linuxias@ubuntu:~/ELF$ hexdump -C -s 0x140 foo.o
00000140
              \mathbf{e}
                 00
                        01 00 00 00
                                                  00 03
                                                         00 00
                     00
                                        02 00
                                              00
                     00 05 00 00 00
                                           00 00
00000150
              \Theta\Theta
                 00
                                                  00
                                                     07
                                                         00 00
00000160
                     00
                        09
                            00
                               00 00
                                        55 48 89 e5 48 89 7d f8
              \mathbf{0}
                 00
```

#### FLAGS IN SECTION HEADER



- WRITE: This section contains data that is writable during process execution.
- ► ALLOC: This section occupies memory during process execution.
- EXECINSTR: This section contains executable machine instructions.



#### .text:

▶ This section holds executable instructions of a program.

► Type: PROGBITS

► Flags: ALLOC + EXECINSTR

#### .data:

▶ This section holds initialized data that contributes to the program's image.

► Type: PROGBITS

► Flags: ALLOC + WRITE



#### .rodata:

► This section holds read-only data.

► Type: PROGBITS

► Flags: ALLOC

#### bss :

This section holds uninitialized data that contributed to the program's image. By definition, the system will initialize the data with zero when the program begins to run.

► Type: NOBITS

► Flags: ALLOC + WRITE



20

- .rel.text, .rel.data, and .rel.rodata:
  - ▶ These contain the relocation information for the corresponding text or data sections.
  - ► Type: REL
  - ▶ Flags: ALLOC is turned on if the file has a loadable segment that includes relocation.
- .symtab:
  - ► This section hold a symbol table.
- .strtab:
  - ► This section holds strings.



#### init:

▶ This section holds executable instructions that contribute to the process initialization code.

► Type: PROGBITS

► Flags: ALLOC + EXECINSTR

#### • .fini:

▶ This section hold executable instructions that contribute to the process termination code.

► Type: PROGBITS

► Flags: ALLOC + EXECINSTR

► C does not need these two sections. However, C++ needs them.



#### .interp:

► This section holds the pathname of a program interpreter.

Type: ALLOC

► Flags: PROGBITS

- ▶ If this section is present, rather than running the program directly, the system runs the interpreter and passes it the elf file as an argument.
- ► For many years (used in a.out), UNIX has had self-running interpreted text files, using
- #! /bin/csh as the first line of the file.
- ▶ Elf extends this facility to interpreters that run nontext programs.
- In practice, this is used to run the run-time dynamic linker to load the program and to link in any required shared libraries.



.debug:

► This section holds symbolic debugging information.

► Type: PROGBIT

.line:

► This section holds line number information for symbolic debugging, which describes the correspondence between the program source and the machine code (ever used gdb?)

Type: PROGBIT

.comment

► This section may store extra information.



.got:

► This section holds the global offset table.

▶ We will explain got when we present shared library.

► Type: PROGBIT

.plt:

► This section holds the procedure linkage table.

► Type: PROGBIT

.note:

▶ This section contains some extra information.

ELF header	n subsite militari funsi
(segment table)	(not considered secti
.text	in is supplementally and it
.data	unofice, section, or in the section of the section
.rodata	Oldumo (MORRO) : Sta
.bss	religion, and see a
.sym	ecton.
.rel.text	al or vallent side stood
.rel.data	edit barade beparen me
.rel.rodata	isi lodigue mo e om cla ist signraded mo a do:
.line	la rea data objectis (pai
.debug	inbut, visible everywhe
.strtab	over it a deministration is ever law of the triving line and
Section table	(not considered a sect

# A TYPICAL RELOCATABLE FILE.

FIGURE 3.14 • Sample relocatable ELF file.

The ELF file format

## **STRING TABLE**



- String table sections hold null-terminated character sequences, commonly called strings.
- ► The object file uses these strings to represent symbol and section names.
- We use an index into the string table section to reference a string.
- The reason why we separate symbol names from symbol tables is that in C or C++, there is no limitation on the length of a symbol.

#### **SYMBOL TABLE**



- An object file's symbol table holds information needed to locate and relocate a program's symbolic definition and references.
- A symbol table index is a subscript into this array.

The section relative to which the symbol is defined. (e.g., the function entry points are defined relative to .text)

If a definition is available for an undefined weak symbol, the linker will use it. Otherwise, the value defaults to 0.

#### **RELOCATION TABLE**



- Relocation is the process of connecting symbolic references with symbolic definitions.
- ▶ Relocatable files must have information that describes how to modify their section contents.
- ► A relocation table consists of many relocation structures.

#### **RELOCATION STRUCTURE**



```
Struct {
   R offset;
```

- ▶ This field gives the location at which to apply the relocation.
- For a relocatable file, the value is the byte offset from the beginning of the section to the storage unit affect by the relocation.
- For an executable file and shared object, the value is the virtual address of the storage unit affected by the relocation.

#### **RELOCATION STRUCTURE**



```
R_info;
```

► This field gives both the symbol table index with respect to which the relocation must be made and the type of relocation to apply.

```
R_addend;
```

▶ This field specifies a constant addend used to compute the value to be stored into the relocation field.

}

10/7/2024 The ELF file format

#### **EXECUTABLE FILES**



- An executable file usually has only a few segments. E.g.,
  - ► A read-only one for the code.
  - ► A read-only one for read-only data.
  - ► A read/write one for read/write data.
- ► All of the loadable sections are packed into the appropriate segments so that the system can map the file with just one or two operations.
  - ▶ E.g., If there is a .init and .fini sections, those sections will be put into the read-only text segment.





► The <u>Program Header Table</u> contains information about the segments in an ELF file and how to load them into memory (segments : contiguous ranges of an ELF file that have the same memory attribution.)

```
typedef struct
 Elf64_Word
                             /* Segment type */
              p_type;
 Elf64_Word
              p_flags;
                             /* Segment flags */
 Elf64_Off p_offset;
                         /* Segment file offset */
 Elf64_Addr
              p_vaddr;
                            /* Segment virtual address */
              p_paddr;
 Elf64_Addr
                            /* Segment physical address */
 Elf64 Xword
              p_filesz; /* Segment size in file */
                            /* Segment size in memory */
 Elf64_Xword
              p_memsz;
                             /* Segment alignment */
 Elf64_Xword
              p_align;
 Elf64_Phdr;
```

#### THE TYPES IN PROGRAM HEADER



- This field tells what kind of segment this array element describes:
  - ▶ PT\_LOAD: This segment is a loadable segment.
  - ▶ PT\_DYNAMIC: This array element specifies dynamic linking information.
  - ▶ PT\_INTERP: This element specified the location and size of a null-terminated path name to invoke as an interpreter.





```
linuxias@ubuntu:~/ELF/exam$ readelf -h test
ELF Header:
 Magic: 7f 45 4c 46 02 01 01 00 00 00 00 00 00 00 00 00
 Class:
                                     ELF64
                                     2's complement, little endian
 Data:
                                     1 (current)
 Version:
 OS/ABI:
                                     UNIX - System V
 ABI Version:
                                     EXEC (Executable file)
 Type:
 Machine:
                                     Advanced Micro Devices X86-64
 Version:
                                     0x1
 Entry point address:
                                     0x400440
 Start of program headers:
                                     64 (bytes into file)
 Start of section headers:
                                     4512 (bytes into file)
 Flags:
                                     0x0
                                     64 (bytes)
 Size of this header:
 Size of program headers:
                                     56 (bytes)
  Number of program headers:
  Size of section headers:
                                     64 (bytes)
 Number of section headers:
                                     30
  Section header string table index: 27
```



Only used for executable, shared libraries and core files

```
linuxias@ubuntu:~/ELF/exam$ readelf -h test.o
ELF Header:
           7f 45 4c 46 02 01 01 00 00 00 00 00 00 00 00 00
 Magic:
 Class:
                                      ELF64
 Data:
                                      2's complement, little endian
 Version:
                                      1 (current)
                                      UNIX - System V
 OS/ABI:
 ABI Version:
                                      0
                                      REL (Relocatable file)
 Type:
 Machine:
                                      Advanced Micro Devices X86-64
 Version:
                                      0x1
 Entry point address:
                                      0 \times 0
                                      0 (bytes into file)
 Start of program headers:
 Start of section headers:
                                      384 (bytes into file)
 Flags:
                                      0x0
 Size of this header:
                                      64 (bytes)
 Size of program headers:
                                      0 (bytes)
 Number of program headers:
                                      0
 Size of section headers:
                                      64 (bytes)
 Number of section headers:
                                      13
  Section header string table index: 10
```

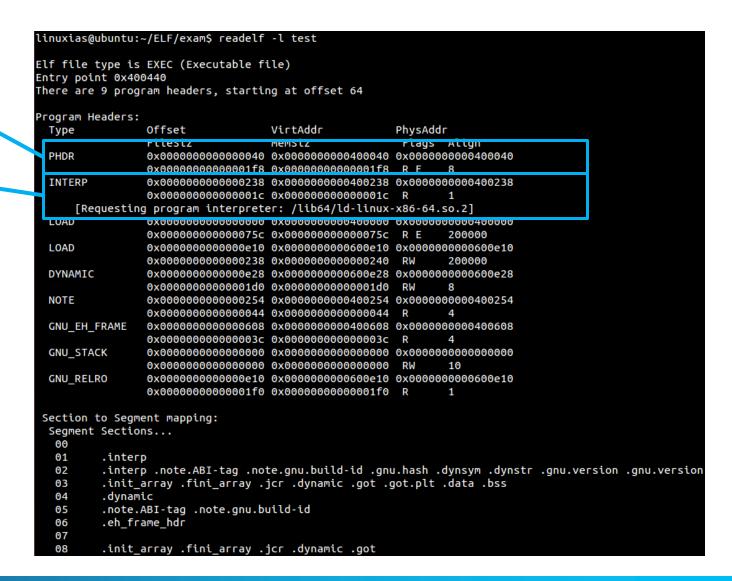


```
linuxias@ubuntu:~/ELF/exam$ readelf -l test
Elf file type is EXEC (Executable file)
Entry point 0x400440
There are 9 program headers, starting at offset 64
Program Headers:
 Type
             Offset
                             VirtAddr
                                            PhysAddr
             FileSiz
                             MemSiz
                                             Flags Align
             0x0000000000001f8 0x0000000000001f8 R E
 INTERP
             0x000000000000238 0x000000000400238 0x0000000000400238
             0x00000000000001c 0x000000000000001c R
     [Requesting program interpreter: /lib64/ld-linux-x86-64.so.2]
 LOAD
              0x000000000000075c 0x000000000000075c R E
 LOAD
              200000
              0x0000000000000238 0x000000000000240 RW
 DYNAMIC
             0x000000000000e28 0x000000000000e28 0x00000000000600e28
              0x0000000000001d0 0x0000000000001d0
 NOTE
              0x000000000000254 0x000000000400254 0x0000000000400254
             0x00000000000000044 0x0000000000000044
 GNU EH FRAME
             0x00000000000000608 0x0000000000400608 0x0000000000400608
              0x000000000000003c 0x000000000000003c R
 GNU STACK
              0x0000000000000000 0x000000000000000 RW
 GNU RELRO
             0x0000000000001f0 0x0000000000001f0 R
Section to Segment mapping:
 Segment Sections...
  00
        .interp .note.ABI-tag .note.gnu.build-id .gnu.hash .dynsym .dynstr .gnu.version .gnu.version
        .init array .fini array .jcr .dynamic .got .got.plt .data .bss
  03
  04
        .dvnamic
  05
        .note.ABI-tag .note.gnu.build-id
  06
        .eh frame hdr
  07
        .init_array .fini_array .jcr .dynamic .got
```

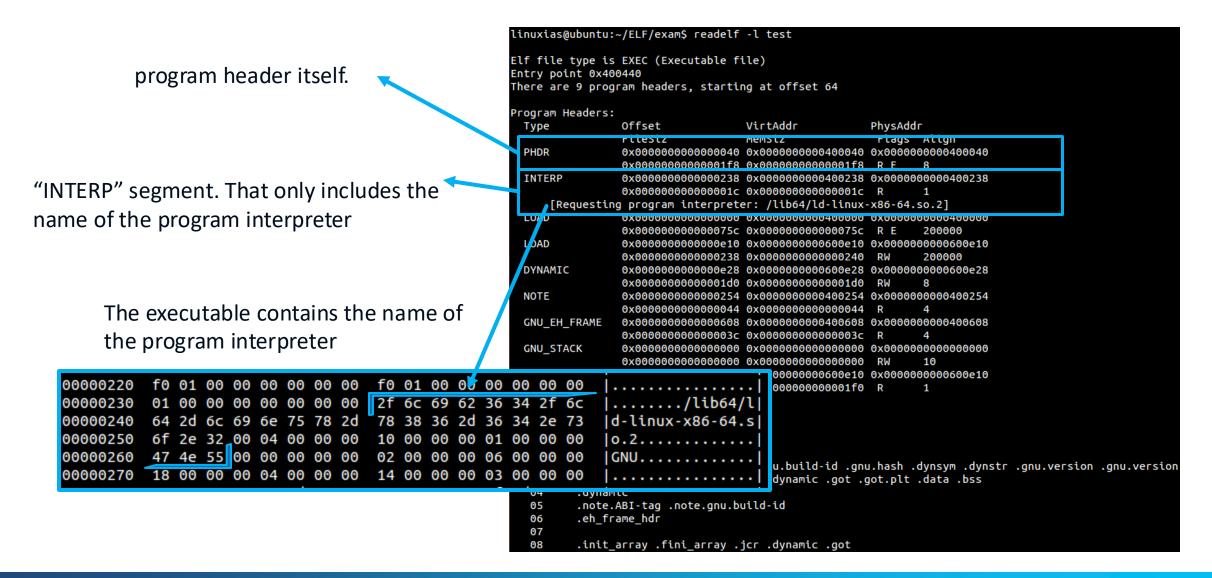


program header itself.

"INTERP" segment. That only includes the name of the program interpreter









DYNAMIC segment used for dynamic linking.

Special segments for Vendor-specific information

```
linuxias@ubuntu:~/ELF/exam$ readelf -l test
Elf file type is EXEC (Executable file)
Entry point 0x400440
There are 9 program headers, starting at offset 64
Program Headers:
 Type
             Offset
                           VirtAddr
                                          PhysAddr
             FileSiz
                           MemSiz
                                           Flags Align
             0x0000000000001f8 0x0000000000001f8
 INTERP
             0x000000000000238 0x000000000400238 0x0000000000400238
             0x000000000000001c 0x000000000000001c R
    [Requesting program interpreter: /lib64/ld-linux-x86-64.so.2]
 LOAD
             0x000000000000075c 0x00000000000075c
 LOAD
             200000
             0x0000000000000238 0x000000000000240 RW
             0x0000000000000e28 0x0000000000600e28 0x0000000000600e28
 DYNAMIC
  NOTE
             0x000000000000254 0x000000000400254 0x000000000400254
 GNU EH FRAME
             0x000000000000608 0x000000000400608 0x000000000400608
 GNU STACK
             GNU RELRO
             0x0000000000001f0 0x0000000000001f0 R
Section to Segment mapping:
 Segment Sections...
  00
  01
  02
       .interp .note.ABI-tag .note.gnu.build-id .gnu.hash .dynsym .dynstr .gnu.version .gnu.version
       .init array .fini array .jcr .dynamic .got .got.plt .data .bss
  03
  04
       .dvnamic
  05
       .note.ABI-tag .note.gnu.build-id
  06
       .eh frame hdr
  07
       .init_array .fini_array .jcr .dynamic .got
```



LOAD segment

: Loadable program segment

Note Segment

: The array element specifies the location and size of auxiliary information

```
linuxias@ubuntu:~/ELF/exam$ readelf -l test
Elf file type is EXEC (Executable file)
Entry point 0x400440
There are 9 program headers, starting at offset 64
Program Headers:
  Type
              Offset
                             VirtAddr
                                             PhysAddr
                                              Flags Align
              PHDR
 INTERP
              0x0000000000000238 0x000000000400238 0x0000000000400238
    [Requesting program interpreter: /lib64/ld-linux-x86-64.so.2]
  LOAD
              LOAD
              DYNAMIC
              0x000000000000e28 0x000000000000e28 0x00000000000600e28
  NOTE
              0x000000000000254 0x000000000400254 0x0000000000400254
 GNU EH FRAME
              0 \times 00000000000000608 0 \times 0000000000400608 0 \times 00000000000400608
 GNU_STACK
 GNU RELRO
              0x0000000000001f0 0x0000000000001f0 R
Section to Segment mapping:
 Segment Sections...
  00
  02
        .interp .note.ABI-tag .note.gnu.build-id .gnu.hash .dynsym .dynstr .gnu.version .gnu.version
        .init array .fini array .jcr .dynamic .got .got.plt .data .bss
  03
        .dynamic
  05
        .note.ABI-tag .note.gnu.build-id
        .eh frame hdr
  07
        .init_array .fini_array .jcr .dynamic .got
```







#### Politecnico di Torino

Department of Control and Computer Engineering



THANK YOU!

