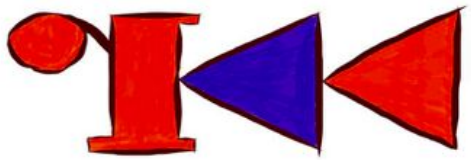
File formats within r2 and its nuances



Who am I?

Twitter: alvaro_fe

Github: alvarofe

Are file formats important?

Are file formats easy to parse?

File format within r2

RBinPlugin
RBinXtrPlugin

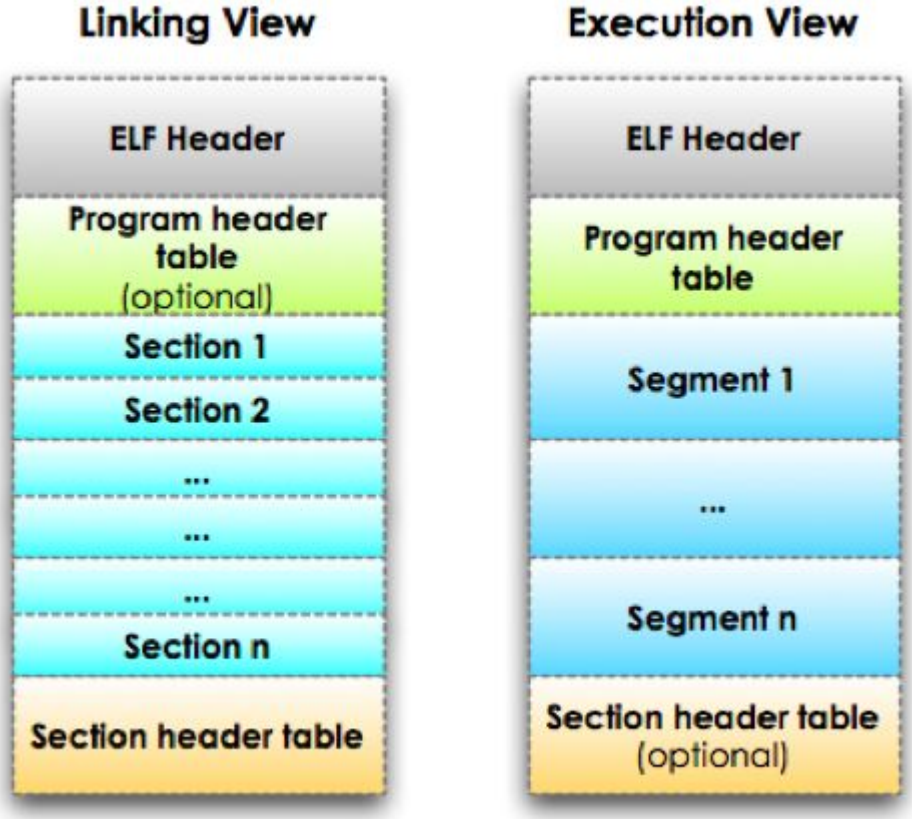
libr/bin/p **libr/bin/format**

https://github.com/radare/radare2/wiki/Implementing-a-new-format

DEMO
Fixing coff

How to fool parsers

ELF



http://www.sw-at.com/blog/wp-content/uploads/2011/04/elf.png

ELF

- Section headers (SHDR) are used to assist the engineer in the debug process, is not used when you run the application.
- Program header (PHDR) actually is what is used by the loader to map the file in memory.

If we do not have this distinction in mind at the time of parsing, we will get invalid values.

Binary without SHDR

Radare2 has been handling this a quite time ago. However other tools such as Hopper, readelf, objdump fails with binaries without SHDR.

Binary with SHDR and overlapped symbols with PHDR

The SHDR says that at offset 0x80808080 you have the symbol **printf** but if you get the info from PHDR says is the **system** symbol.

We were taking the information from SHDR as truth, when it should be PHDR instead when available. Now we have a check in place to substitute SHDR’s symbols for those in PHDR when overlapping.

Fixed recently by pancake thanks to an user from Twitter (https://twitter.com/maciekkotowicz/status/771439232531456001).

Binary with SHDR incomplete

How symbol extraction roughly works in pseudo code.

```
if (!SHDR)
     get_from_phdr ()
for (symbols in SHDR) {
     list_symbols.append(symbol)
}
fix_overlapped_symbols ()
return list_symbols
```

Binary with SHDR incomplete

How symbol extraction roughly works in pseudo code.

```
if (!SHDR)
     get_from_phdr ()
for (symbols in SHDR) {
     list_symbols.append(symbol)
}
fix_overlapped_symbols ()
return list_symbols
```

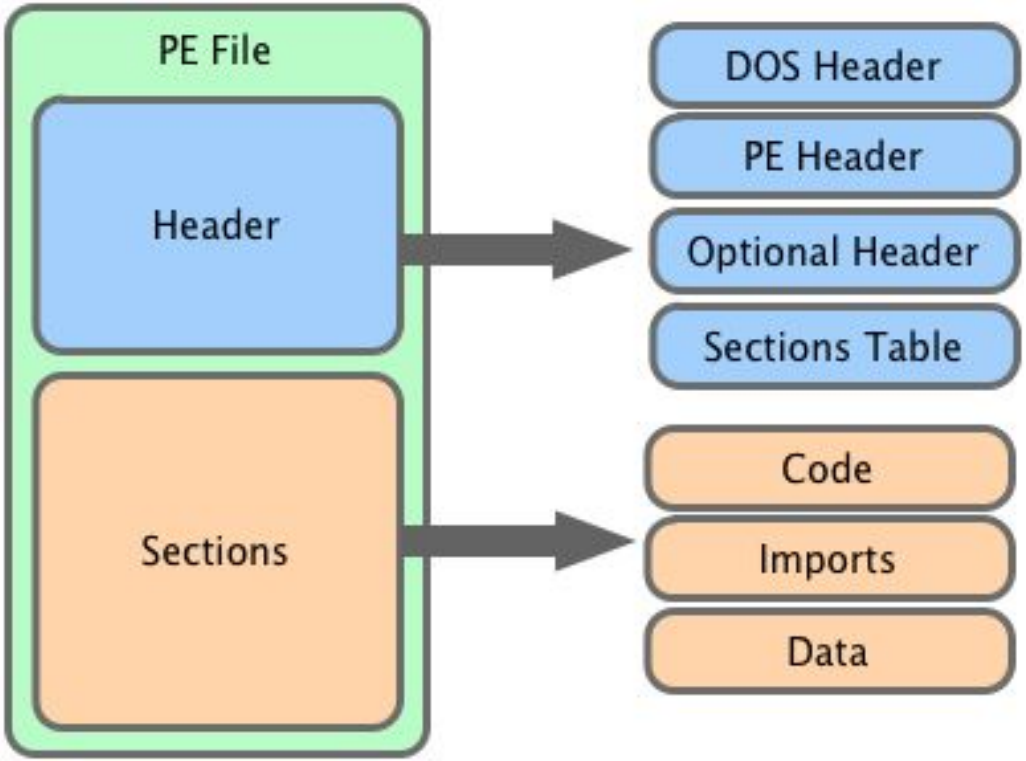
Easy to break and miss symbols. What would happen if a binary says only has one symbol? We are not merging symbols from PHDR and those are present if the binary runs.

Binary with SHDR incomplete

How symbol extraction roughly works in pseudo code.

```
if (!SHDR)
     get_from_phdr ()
for (symbols in SHDR) {
     list_symbols.append(symbol)
}
fix_overlapped_symbols ()
add_missing_symbols_from_phdr ()
return list_symbols
```

PE



https://npercoco.typepad.com/.a/6a0133f264aa62970b01901b84d005970b-500wi

PE

In a security tool such as radare2 you do not only have to support standard binaries, those that are meant to run, but also those that are messed up.

This makes thing sometimes very hard to handle and above all it forces to be flexible, sometimes removing security checks that can lead to breakage.

PE

Solution: change things on the fly when you find something that does not make sense.

One example would be a binary when ending the infection, it modifies itself saying the total number of sections is 0. The rationale behind this is to avoid the tool be able to map memory.

PE

#921

Thanks for listening

Questions?