Análisis de firmwares

Primeros pasos

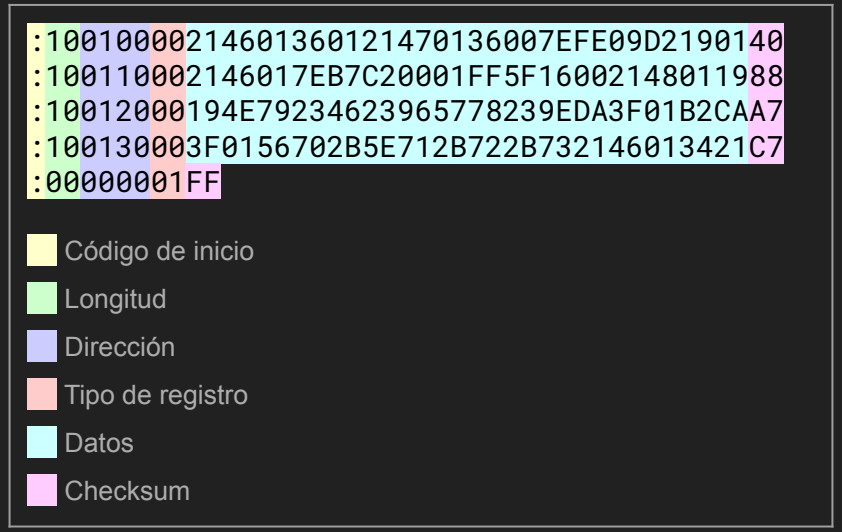
- Hemos realizado un dump de una memoria

- Tenemos un solo fichero con los contenidos de toda la memoria flash

- ¿Qué formato tiene ese fichero? Binario vs hexdump vs Intel HEX

Intel HEX - Formato

- https://es.wikipedia.org/wiki/HEX_(Intel)



Intel HEX - Identificación

- Es formato texto. Se puede abrir con un editor de texto

- Todas las líneas comienzan por “:”

- Típico en microcontroladores como PIC, AVR, algunos chips de ARM como los nRF (Nordic Semiconductor), algunas EEPROMs…

Hexdump - Formato

- https://en.wikipedia.org/wiki/Hex_dump

```
00000000  30 31 32 33 34 35 36 37  38 39 41 42 43 44 45 46  |0123456789ABCDEF|
00000010  0a 2f 2a 20 2a 2a 2a 2a  2a 2a 2a 2a 2a 2a 2a 2a  |./* ************|
00000020  2a 2a 2a 2a 2a 2a 2a 2a  2a 2a 2a 2a 2a 2a 2a 2a  |****************|
00000030  2a 2a 2a 2a 2a 2a 2a 2a  2a 2a 2a 2a 2a 2a 2a 2a  |****************|
00000040  2a 2a 20 2a 2f 0a 09 54  61 62 6c 65 20 77 69 74  |** */..Table wit|
00000050  68 20 54 41 42 73 20 28  30 39 29 0a 09 31 09 09  |h TABs (09)..1..|
00000060  32 09 09 33 0a 09 33 2e  31 34 09 36 2e 32 38 09  |2..3..3.14.6.28.|
00000070  39 2e 34 32 0a                                    |9.42.|
00000075
```

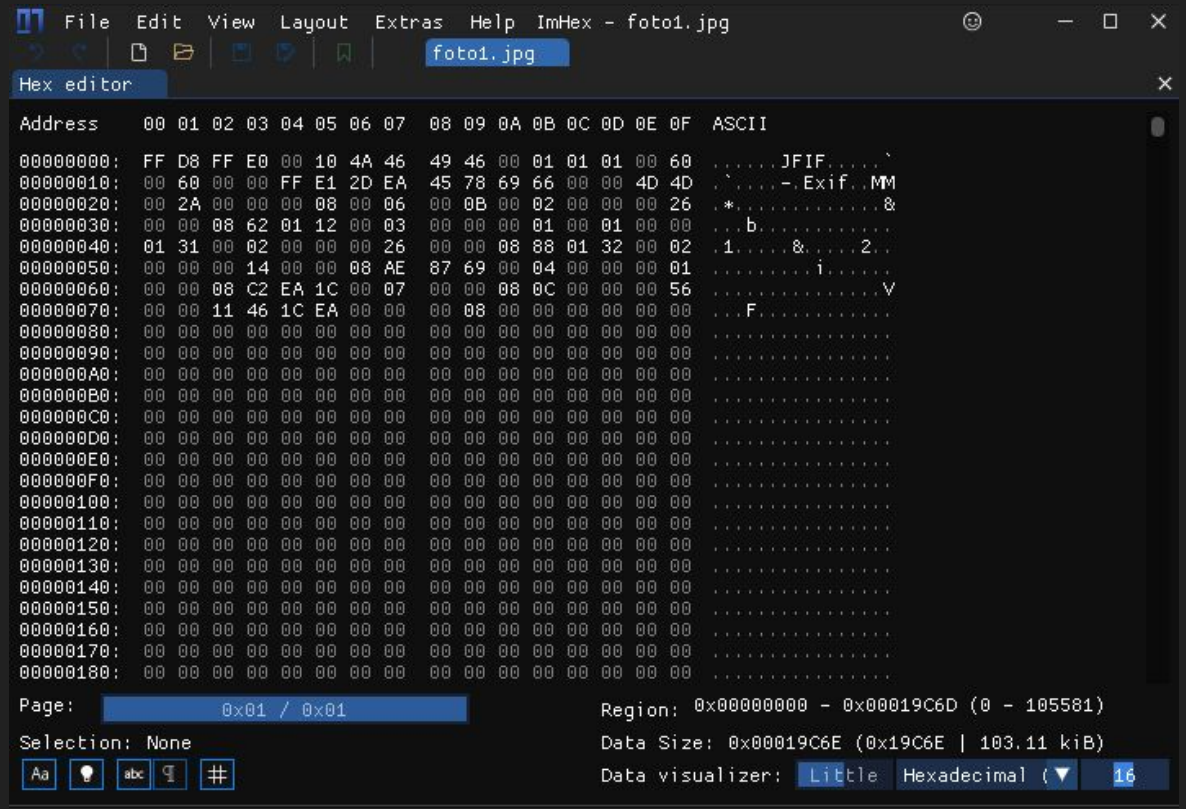
Hexdump - Identificación

- Es formato texto. Se puede abrir con un editor de texto
- Pueden variar las columnas dependiendo de la herramienta…
- Si existe una columna con direcciones, normalmente serán consecutivas pero pueden no serlo!!
- Siempre encontraremos la sección principal con los datos en formato hexadecimal
- NUNCA usar la columna de texto para convertir a formato binario! Pérdida de información!
- Típico en volcados desde un bootloader a través de una UART…

Binario RAW - Formato

```
������JFIF��`�`����-
�Exif��MM�*������&��b������1����&���2������i�������
����V��F
                                                    ...
                    ...Windows_Photo_Editor_10.0.10011.16384
...Windows_Photo_Editor_10.0.10011.16384
...2023:03:2408:50:01
                    ...2023:03:23_12:00:56...2023:03:23_12:00:56
(
                    ...
2222222222222222
                    ...
$3br...
                    ...
...$4...
...
                    ...
...RB...Z(QL...P...1@)(i(...(....V........T...Ř...Fj{k...QXU...k[k...j+...."
...r...η...[...E1...)i...Q...(...1F((...`PE...k...j...4...ř$...he;
...QI.....D...n...\...gS...ƀ...!E...P...Z...(...`PE...b...Pb...P...!...q...W9...Rb...3sF...
(...J...o...ZV...1...+....FWg...cI...UX...F)...b...PF(.......0b(..........
```

Binario RAW - Formato



Binario RAW - Identificación

- No tiene porque ser un archivo de texto!
- Debemos usar un editor hexadecimal
- Si se trata como texto podemos perder información!
- No vamos a ver una estructura definida, dependerá del contenido…
- Es el formato más común, formato por defecto de muchas herramientas de volcado de memorias

Ejercicio 1

- Identificación del formato del fichero proporcionado

Formato binario

- Casi todas las herramientas de análisis que vamos a usar trabajan con el formato binario

- Si nuestro dump está en formato hexadecimal o Intel HEX habrá que transformarlo

Intel HEX > Binario RAW

- Existen infinidad de herramientas:
  - SRrecord: https://github.com/sierrafoxtrot/srecord

    ```
    srec_cat inputFile.hex -Intel -output outputFile.bin -binary
    ```

  - Binex: http://www.nlsw.nl/software/

    ```
    binex.exe /B inputFile.hex
    ```

  - HEX2BIN: https://www.keil.com/download/docs/7.asp

    ```
    hex2bin inputFile.hex outputFile.bin
    ```

Hexdump > Binario RAW

- Lo más común es xxd:

  ```
  xxd -r -p inputHexdump.txt outputBinary.bin
  ```

- xxd puede no llevarse bien con los números de las direcciones:

  ```
  cut -d' ' -f3-19 inputHexdump.txt | xxd -r -p > outputBinary.bin
  ```

Ejercicio 2

- Convertir el fichero a binario

Cómo trabajar con fichero binario

Editores hexadecimales:

- ImHex - https://github.com/WerWolv/ImHex
- Hobbits - https://github.com/Mahlet-Inc/hobbits
- 101 Hex Editor - https://www.sweetscape.com/010editor/
- HxD - https://mh-nexus.de/en/hxd/
- Okteta - https://apps.kde.org/es/okteta/

¿Qué hay en la flash de un dispositivo?

- ¿Qué esperamos encontrarnos en nuestro fichero binario?
  - Ejecutables
  - Ficheros de configuración
  - Espacios en blanco!
  - …

Ejercicio 3

- Abrir el fichero binario en un editor hexadecimal
- ¿Qué valor toman los espacios no escritos de memoria?
- ¿Por qué toman este valor?

¿Cómo está organizada una flash?

- ¿Cómo esperamos que esté organizada esta información?
  - Particiones o secciones
  - Separadas por espacios en blanco

Cómo identificar particiones

- Estrategias
  - Búsqueda de espacios en blanco
  - Búsqueda de firmas/magic numbers

Búsqueda de firmas

- ¿Que es una firma/magic number?
  - Algunos formatos de archivo usan un “magic number” para identificar su inicio
  - Es una constante que no **debería** variar
  - Cuidado: Algunos fabricantes cambian las firmas de sus archivos porque los customizan o para no ser detectados

Búsqueda de firmas

- Herramientas:
  - Binwalk: https://github.com/ReFirmLabs/binwalk

    `binwalk fichero.bin (Solo busca firmas)`

    `binwalk -e fichero.bin (Extrae)`
  - Unblob: https://github.com/onekey-sec/unblob

    `unblob fichero.bin (Extrae)`

Ejercicio 4

- Búsqueda de firmas en nuestro archivo binario
- ¿Donde están localizadas las firmas?
- ¿Podemos extraerlas?
- ¿Podemos mejorar los resultados?

Cómo identificar particiones

- Estrategias
  - Búsqueda de firmas/magic numbers
  - Con herramientas estadísticas
    - Entropía
    - Distribución de bytes
    - Varianza
    - …

Entropía

- Es una medida de “densidad” de información
- La entropía será máxima cuando haya una buena compresión
- La entropía será alta con el cifrado
- La entropía será media cuando haya formatos estructurados
- La entropía es mínima en regiones vacías

Entropía

- Herramientas:
  - ImHex:

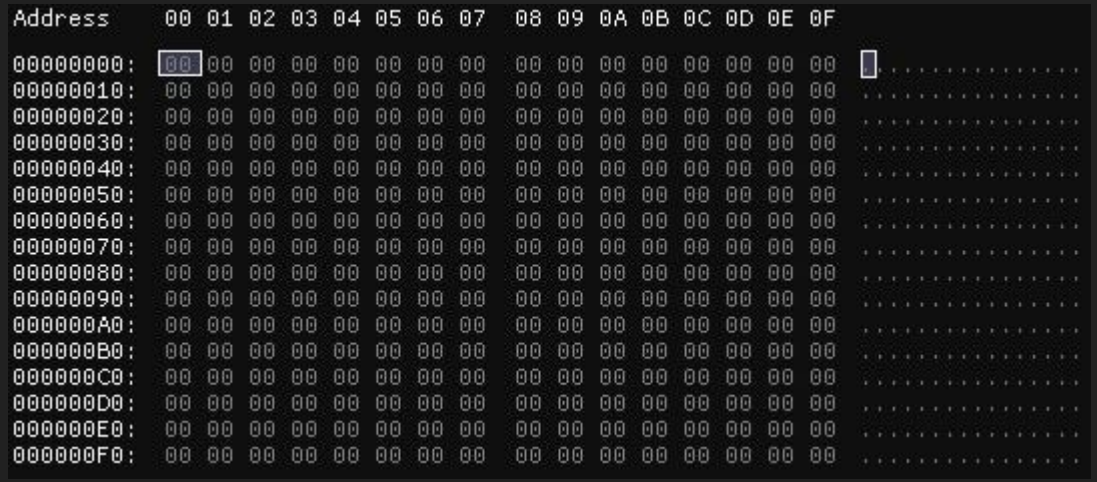
    `View > Data information > Analyze`
  - Binwalk:

    `binwalk -E fichero.bin`
  - Cyberchef: https://gchq.github.io/CyberChef/

    `Open file > Add recipe > Entropy`

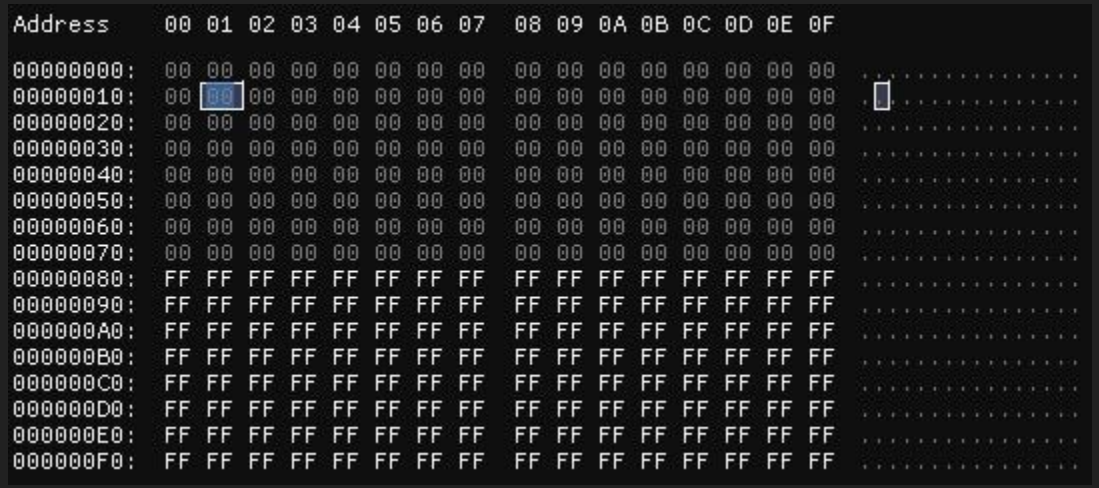
Entropía

- ¿Qué entropía tiene el siguiente fichero?



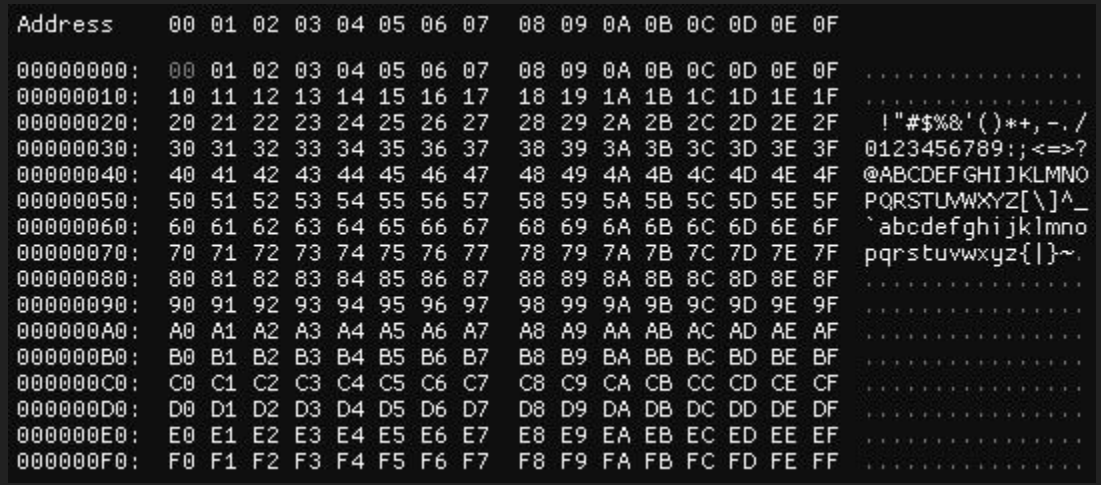
Entropía

- ¿Este tendrá más o menos?



Entropía

- ¿Este tendrá más o menos?
- ¿Es aleatorio o predecible?



```
Address    00 01 02 03 04 05 06 07  08 09 0A 0B 0C 0D 0E 0F

00000000:  00 01 02 03 04 05 06 07  08 09 0A 0B 0C 0D 0E 0F  ................
00000010:  10 11 12 13 14 15 16 17  18 19 1A 1B 1C 1D 1E 1F  ................
00000020:  20 21 22 23 24 25 26 27  28 29 2A 2B 2C 2D 2E 2F   !"#$%&'()*+,-./
00000030:  30 31 32 33 34 35 36 37  38 39 3A 3B 3C 3D 3E 3F  0123456789:;<=>?
00000040:  40 41 42 43 44 45 46 47  48 49 4A 4B 4C 4D 4E 4F  @ABCDEFGHIJKLMNO
00000050:  50 51 52 53 54 55 56 57  58 59 5A 5B 5C 5D 5E 5F  PQRSTUVWXYZ[\]^_
00000060:  60 61 62 63 64 65 66 67  68 69 6A 6B 6C 6D 6E 6F  `abcdefghijklmno
00000070:  70 71 72 73 74 75 76 77  78 79 7A 7B 7C 7D 7E 7F  pqrstuvwxyz{|}~.
00000080:  80 81 82 83 84 85 86 87  88 89 8A 8B 8C 8D 8E 8F  ................
00000090:  90 91 92 93 94 95 96 97  98 99 9A 9B 9C 9D 9E 9F  ................
000000A0:  A0 A1 A2 A3 A4 A5 A6 A7  A8 A9 AA AB AC AD AE AF  ................
000000B0:  B0 B1 B2 B3 B4 B5 B6 B7  B8 B9 BA BB BC BD BE BF  ................
000000C0:  C0 C1 C2 C3 C4 C5 C6 C7  C8 C9 CA CB CC CD CE CF  ................
000000D0:  D0 D1 D2 D3 D4 D5 D6 D7  D8 D9 DA DB DC DD DE DF  ................
000000E0:  E0 E1 E2 E3 E4 E5 E6 E7  E8 E9 EA EB EC ED EE EF  ................
000000F0:  F0 F1 F2 F3 F4 F5 F6 F7  F8 F9 FA FB FC FD FE FF  ................
```

Ejercicio 5

- Análisis de entropía de nuestro binario
- Identificación de secciones
- ¿Hemos encontrado más cosas?
- ¿Hemos descartado algunos falsos positivos?
- Caracterización de secciones: cifrado vs compresión vs otros

¿Cómo está organizada una flash?

- Con la entropía y búsqueda de regiones en blanco identificamos particiones/secciones de la flash
- Con la entropía y búsqueda de firmas identificamos que particiones/secciones están comprimidas, cifradas u otros formatos
- Podemos entender parte del formato, pero ¿Por qué esto está organizado de esta manera?

Entendiendo el diseño de un dispositivo

- Se busca un coste reducido y ajustado (esto da más beneficio al fabricante)
- En parte se logra a través de la “modularidad”: elijo una CPU y la combino con una RAM y una flash
- Esto genera complejidad en el software porque este tiene que adaptarse a esa modularidad!

El proceso de arranque

Initial Program Loader o
Rom Boot Loader

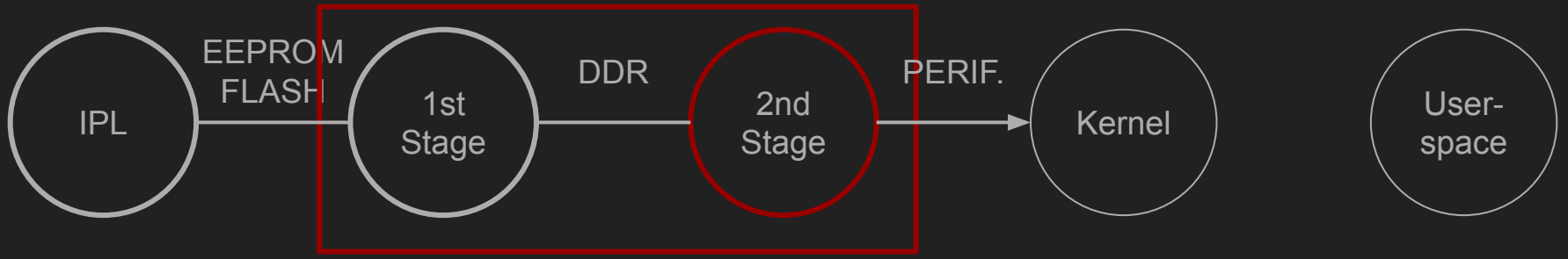
- Muy limitado en tamaño, solo puede usar una SRAM mínima.
- Suele ser parte del SoC y usa recursos dentro del SoC.
- Inicializa mínimamente un medio de almacenamiento y copia el SPL a la SRAM.

El proceso de arranque

Secondary Program Loader o
Memory Loader (MLO) o
1st Stage Loader

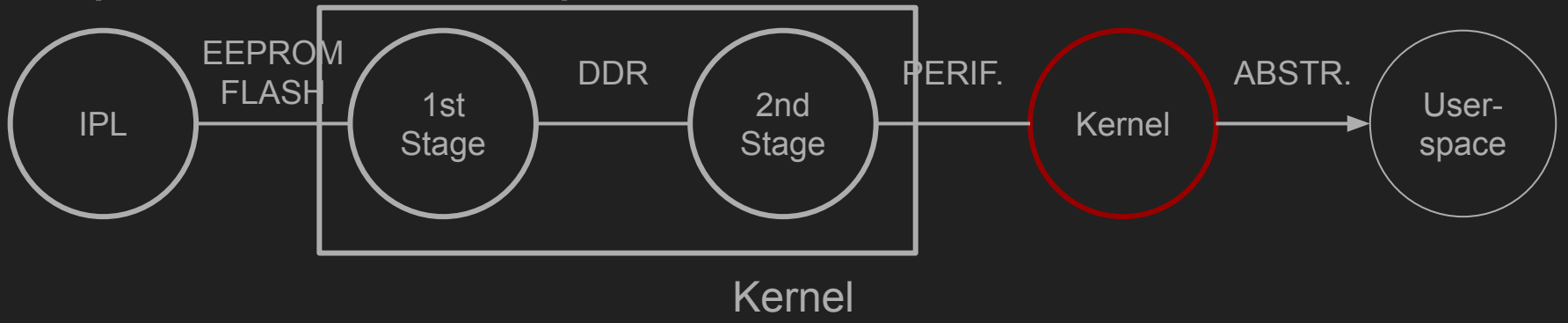
- Desde la SRAM inicializa la DDR de más tamaño.
- Casi siempre reconfigura el almacenamiento de arranque.
- Posiblemente configura algunos periféricos (UART para debug).
- Carga el 2nd stage de mucho mayor tamaño en la DDR.

El proceso de arranque



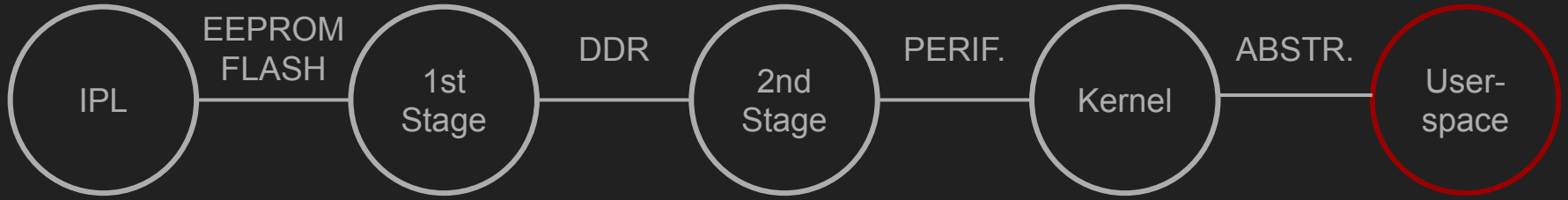
- Habitualmente llamamos bootloader al 1st stage + 2nd stage juntos.
- Los más típicos son U-Boot/CFE/Grub…
- Inicialización más completa del hardware. Puede tener mucha funcionalidad: Línea de comandos, boot de red, soporte para distintos almacenamientos…

El proceso de arranque



- Gestión de memoria, gestión de procesos, permisos, abstracción del hardware…
- Lo más común es una versión modificada de un Kernel obsoleto.
- Por motivos de licenciamiento el código propietario suele proveerse en módulos sin código fuente.

El proceso de arranque



Userspace

- Herramientas, servicios, configuraciones…

¿Cómo está organizada una flash?

- Particiones/secciones más comunes
  - Bootloader
  - Kernel
  - Sistema de archivos principal (userspace)
  - Otras particiones “custom” para traducciones, configuración u otros

Ejercicio 6

- Identificación de cada una de las secciones
- ¿Cuál es el bootloader?
- ¿Y el Kernel?
- ¿Sistema de archivos para el userspace?
- ¿Hay otras?
- Todos debemos extraer con dd los binarios correspondientes a todas las particiones

```
dd if=input.bin of=output.bin bs=1 skip=$offset count=$size
```

Ejercicio 7

- ¿Qué formato tiene el sistema de archivos de userspace?
- Extraer el sistema de archivos a una carpeta

¿Qué podemos hacer ahora?

- Firmware custom! Modificar el userspace, modificar o incluir algún fichero nuesto, reempaquetar y volver a flashear en nuestro router.
- Analizar el firmware que nos ha dado el fabricante en busca de cosas interesantes…

Análisis del root filesystem

- ¿Cuál es el primer proceso que se ejecuta? /sbin/init
- ¿Que hace?
  - /etc/inittab
  - /etc/init.d/
  - …
- ¿Hay binarios interesantes?
- ¿Hay otras configuraciones interesantes?

Ejercicio 8

- Analizar el firmware!
- ¿Qué nos falta por conocer de este firmware?
- ¿Existe algún archivo interesante que debamos analizar?
- ¿Qué necesitamos para analizar esos ficheros?