Análise a dados armazenados

António Pinto apinto@estg.ipp.pt

Escola Superior de Tecnologia e Gestão

Outubro 2022 (v2)

Sumário

Visão geral

Organização dos discos

Sistemas de ficheiros

Representação de dados

Conceitos introdutórios

- Dados são guardados em suportes de armazenamento
 - Discos, cartões de memória, pens USB, ...
- Suportes de armazenamentos tem estrutura própria
 - Estrutura guardada em MBR ou GPT
 - Encontram-se divididos em partições
 - Cada partição tem um sistema de ficheiros
- Sistemas de ficheiros organizam a forma como se armazenam os dados
 - ► FAT (12,16,32), NTFS, EXT(2,3,4), UFS, ...
 - Alguns sistemas de ficheiros são de conhecimento público, outros são proprietários

Visão geral 3/52

Estrutura de um disco

- Os tipos de discos mais comuns são os organizados por
 - Master Boot Record (MBR)
 - ► GUID Partition Table (GPT) ¹
- O MBR/GPT é armazenado no inicio do disco, contendo a sua estrutura (tabela de partições)
- Cada partição utiliza um sistema de ficheiros
- Reparticionar um disco não apaga dados, apenas a tabela de partições

¹GPT faz parte da norma UEFI que veio para substituir a BIOS Organização dos discos

Conteúdos

Visão geral

Organização dos discos Master Boot Record GUID Partition Table

Sistemas de ficheiros FAT (12,16,32) NTFS EXT2,3,4

Representação de dados Representação de números Representação de carateres

- Ocupa os primeiros 512 bytes do disco
- Incluí apontadores para 4 partições primárias (que podem ou não estar em uso)
- Partições adicionais (além destas 4)
 - Requer que uma seja marcada como partição estendida
 - Partição estendida usa Extended Boot Record (EBR)
 - EBR pode conter apontador para um EBR seguinte
 - Número ilimitado de partições estendidas

Exercício #1 (15 minutos)

Com o comando seguinte, visualize o conteúdo dos primeiros 512 bytes do ficheiro **usb-mbr.dd** do moodle.

```
hexdump -n 512 -C usb-mbr.dd
```

Tente identificar o *offset* (ou a posição relativa à origem) dos bytes *0x55AA*.

Note-se que 0x0200 (200 em hexadecimal) = 512 (decimal)

Estrutura genérica

| Endereço | | Descrição | | Tamanho |
|--------------------------------|------|------------------------------------|--------------------------------|---------|
| Hex | Dec | Descrição | | (bytes) |
| +0×0000 | +0 | Código de arranque (<i>boot</i>) | | 446 |
| +0×01BE | +446 | Partição #1 | | 16 |
| +0×01CE | +462 | Partição #2 | Tabala da partica as primários | 16 |
| +0×01DE | +478 | Partição #3 | Tabela de partições primárias | 16 |
| +0×01EE | +494 | Partição #4 | | 16 |
| +0×01FE | +510 | 55 | Assinatura | 2 |
| +0×01FF | +511 | AA | ASSIIIatura | |
| Tamanho total : 446 + 4×16 + 2 | | | 512 | |

 $\verb|http://en.wikipedia.org/wiki/Master_Boot_Record|$

Campo de partição

Exercício #2 (30 minutos)

Utilizando novamente o comando **hexdump**, em conjunto com a informação da estrutura do MBR, apresente apenas a informação relativa **à primeira entrada de partição** do ficheiro **usb-mbr.dd** do moodle.

+info: https://en.wikipedia.org/wiki/Master_boot_record.

Submeta sua análise crítica pelo moodle (ficheiro PDF)

```
aap@pc:~$ hexdump -s 446 -n 16 -C usb-mbr.dd
2 000001be 00 00 21 00 06 2a ea ca 20 00 00 00 e0 b7 3b 00
3 000001ce
```

A 1ª estrada de partição inicia na posição 446 e tem 16 bytes de tamanho!

Estrutura do campo partição (partition entry)

| Posição relativa | Descrição | Tamanho (bytes) |
|---------------------|----------------------------------|--------------------|
| Θ | Indicador de <i>boot</i> (80h) | 1 |
| 1 | Inicio de partição (CHS) | 3 |
| 4 | Tipo de partição | 1 |
| 5 | Fim de partição (CHS) | 3 |
| 8 | Setor inicial (LBA) | 4 |
| 12 | Tamanho da partição (em setores) | 4 |
| | Tamanho total: | 16 |

http://en.wikipedia.org/wiki/Master_Boot_Record

Tipos de numeração de setores no disco

- CHS Cylinder/Head/Sector (mais antigo)
- ► LBA Logical Block Addressing (mais recente)

Exemplo de entrada de partição

```
Offset
                           5
                                  7
                                      8
                                         9
                                             Α
                                                В
                              6
  1B0
                      00
                          00
                             00
                                 00
                                    00
                                        00
                                            00
                                               00
                                                   00
                                                      00
  1C0
                   1F 3F
                          33 3F 00
            00
               0B
                                    00
                                        00
                                            41
                                               99
                                                   01
                                                              00
  1D0
        00
                          00
                             00
                                 00
                                    00
                                        00
                                            00
                      00
                                               00
                                                   00
                                                      00
  1E0
        00
           00
               00
                  00
                      00
                          00
                             00
                                 00
                                    00
                                        00
                                            00
                                               00
                                                   00
                                                      00
                                                              00
  1F0
               00
                  00
                      00
                          00
                             00 00
                                    00
                                        00
                                            00
                                               00
                                                   00
                                                      00
```

http://thestarman.pcministry.com/asm/mbr/PartTables.htm

```
0x1BE Partição está ativa (valor 80h)
0x1BF Setor inicial da partição CHS(0,1,1)
0x1C2 Tipo de partição (0B → FAT32)
```

Extended Boot Record

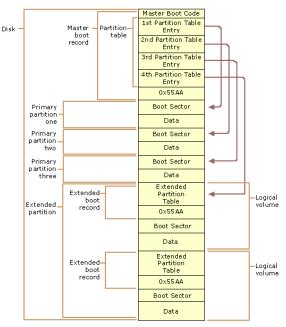
Estrutura genérica

| Posição relativa ao início do EBR | | Descrição | Tamanho |
|--------------------------------------|-----------|---|---------|
| Hex | Dec | | (bytes) |
| 000 - 1BD | 000 - 445 | Tipicamente vazia (zeros) | 446 |
| 1BE - 1CD | 446 - 461 | Partição #1 – Descreve a partição atual | 16 |
| 1CE - 1DD | 462 – 477 | Partição #2 – Descreve a próxima partição | 16 |
| 1DE - 1ED | 478 – 493 | Partição #3 – Não utilizada (zeros) | 16 |
| 1EE - 1FD | 494 – 509 | Partição #4 – Não utilizada (zeros) | 16 |
| 1FE - 1FF | 510 - 511 | Assinatura | 2 |
| | | Tamanho total: 446 +4×16 +2 = | 512 |

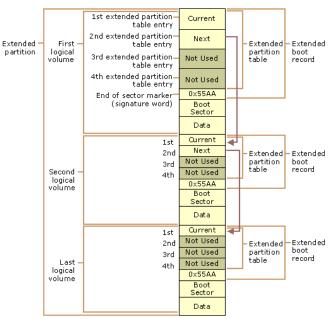
http://en.wikipedia.org/wiki/Extended_boot_record

Têm a mesma estrutura que uma MBR partition entry

Layout de um disco com MBR



Layout de um EBR



Conteúdos

Visão geral

Organização dos discos Master Boot Record GUID Partition Table

Sistemas de ficheiros FAT (12,16,32) NTFS EXT2,3,4

Representação de dados Representação de números Representação de carateres

GUID Partition Table (GPT)

- GUID Partition Table (GPT) é a forma de organização de discos incluída na norma Unified Extensible Firmware Interface (UEFI)
- Utiliza endereçamento por LBA (blocos de 512 bytes²)
- Suporta partições de tamanho superior aos em MBR
- Pode ser utilizada em alguns sistemas com BIOS, desde que suportado pelo sistemas operativo (ex.: Linux)
- Primeiro bloco de 512 bytes (LBA-0) é ignorado (por ser o espaço normalmente utilizado pelo MBR) ou inclui Protective MBR

²Tipicamente, mas podem ser maiores.

Exercício #3 (30 minutos)

Com o comando **hexdump**, visualize o conteúdo dos primeiros 1024 bytes do ficheiro **disk-image.dd** do moodle.

Tente identificar o offset dos bytes 0x55AA.

Confirme que se trata de um Protective MBR com GPT.

+info: https://en.wikipedia.org/wiki/GUID_Partition_Table

Submeta sua análise crítica pelo moodle (ficheiro PDF)

```
1 aap@pc: "$ hexdump -n 1024 -C disk-image.dd
3 00000010 06 b9 00 02 fc f3 a4 50 68 1c 06 cb fb b9 04 00 | .......Ph.......
4 00000020 bd be 07 80 7e 00 00 7c 0b 0f 85 0e 01 83 c5 10 |....~.........
5 00000030 e2 f1 cd 18 88 56 00 55 c6 46 11 05 c6 46 10 00 l.....V.U.F...F...I
6 00000040 b4 41 bb aa 55 cd 13 5d 72 Of 81 fb 55 aa 75 09 | A..U..]r...U.u.|
7 00000050 f7 c1 01 00 74 03 fe 46 10 66 60 80 7e 10 00 74 \dots...t.F.f'.~..t\
8 00000060 26 66 68 00 00 00 00 66 ff 76 08 68 00 00 68 00 |&fh...f.v.h.h.|
9 00000070 7c 68 01 00 68 10 00 b4 42 8a 56 00 8b f4 cd 13 | | h..h...B.V..... |
(LINHAS REMOVIDAS POR FALTA DE ESPACO)
11
12 00000150 10 eb f2 f4 eb fd 2b c9 e4 64 eb 00 24 02 e0 f8 | .....+..d..$...|
13 00000160 24 02 c3 49 6e 76 61 6c 69 64 20 70 61 72 74 69 |$..Invalid parti|
14 00000170 74 69 6f 6e 20 74 61 62 6c 65 00 45 72 72 6f 72 |tion table.Error|
15 00000180 20 6c 6f 61 64 69 6e 67 20 6f 70 65 72 61 74 69 | loading operati|
16 00000190 6e 67 20 73 79 73 74 65 6d 00 4d 69 73 73 69 6e |ng system.Missin|
17 000001a0 67 20 6f 70 65 72 61 74 69 6e 67 20 73 79 73 74 |g operating syst|
18 000001b0 65 6d 00 00 00 63 7b 9a -- 1. PA RT -- -> 00 fe |em...c{.}n.....
19 000001c0 ff ff ee fe ff ff 01 00 00 00 a3 70 3d 3a <- -- |.....p=:..|
21 *
23 00000200 45 46 49 20 50 41 52 54 00 00 01 00 5c 00 00 00 | EFI PART....\...|
24 00000210 7f 95 63 62 00 00 00 00 01 00 00 00 00 00 00 |..cb......
25 00000220 a3 70 3d 3a 00 00 00 00 22 00 00 00 00 00 00 0 l.p=:......
26 00000230 82 70 3d 3a 00 00 00 00 a8 4d 00 00 93 38 00 00 |.p=:....M...8..|
27 00000240 3d 11 00 00 f3 62 00 00 02 00 00 00 00 00 00 |-...b......
28 00000250 80 00 00 00 80 00 00 00 81 d5 7a 4e 00 00 00 00 |.....zN....
30 *
31 00000400
```

Estrutura genérica

| Posição relativa | Descrição | Tamanho (bytes) |
|------------------|--|-----------------|
| LBA0 | Não utilizado (MBR) | 512 |
| LBA1 | Cabeçalho GPT principal | 512 |
| LBA2 | Entradas das partições 1 a 4 (128 bytes cada) | 512 |
| LBA3 | E | 512 |
| | Entradas das partições 5 a 128 (128 bytes cada) | |
| LBA33 | (120 bytes cada) | 512 |
| | Dados / Partições | |
| LBA-33 | C4-:- d E-td d | 512 |
| | Cópia das Entradas das partições 5 a 128 (128 bytes cada) | |
| LBA-3 | partições 3 a 120 (120 bytes cada) | 512 |
| LBA-2 | Cópia das entradas das partições 1 a 4 (128 bytes cada) | 512 |
| LBA-1 | Cópia do cabeçalho GPT | 512 |

LBA-1 refere-se ao último LBA do disco.

Estrutura do cabeçalho GPT

| Posição relativa | Tamanho | Contents |
|------------------|----------|---|
| 0 (0x00) | 8 bytes | Assinatura (45h 46h 49h 20h 50h 41h 52h 54h) |
| 8 (0x08) | 4 bytes | Revisão (00h 00h 01h 00h para GPT ver 1.0) |
| 12 (0x0C) | 4 bytes | Tamanho (5Ch 00h 00h 00h = 92 bytes) |
| 16 (0x10) | 4 bytes | CRC32 do cabeçalho |
| 20 (0x14) | 4 bytes | Reservado (zero) |
| 24 (0x18) | 8 bytes | LBA do cabeçalho |
| 32 (0x20) | 8 bytes | LBA da cópia do cabeçalho |
| 40 (0x28) | 8 bytes | Primeiro LBA útil |
| 48 (0x30) | 8 bytes | Último LBA útil |
| 56 (0x38) | 16 bytes | GUID do disco |
| 72 (0x48) | 8 bytes | LBA da lista de partições |
| 80 (0x50) | 4 bytes | Número de partições na lista |
| 84 (0x54) | 4 bytes | Tamanho da cada entrada de partição (128 bytes) |
| 88 (0x58) | 4 bytes | CRC32 da lista de partições |
| 92 (0x5C) | * | Reservado, zeros até ao fim |

https://en.wikipedia.org/wiki/GUID_Partition_Table

Estrutura de entrada de partição em GPT

| Posição relativa | 1 loccrican | |
|---------------------|-----------------------------|-----|
| Θ | GUID do tipo de partição | 16 |
| 16 | GUID da partição (único) | 16 |
| 32 | LBA inicial | 8 |
| 40 | LBA final | 8 |
| 48 | Atributos | 8 |
| 56 | Nome da partição (UTF-16LE) | 72 |
| | Tamanho total: | 128 |

 $\verb|https://en.wikipedia.org/wiki/GUID_Partition_Table|$

Exercício #4 (15 minutos)

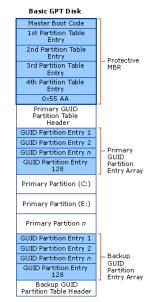
Utilizando novamente o comando **hexdump**, em conjunto com a informação da estrutura da GPT, apresente o nome da **primeira entrada de partição** do ficheiro **disk-image.dd** do moodle.

+info: https://en.wikipedia.org/wiki/GUID_Partition_Table.

Submeta sua resposta pelo moodle (ficheiro PDF)

Skip de 1080 (1024+56), com tamanho de 72 bytes!

Layout de um disco com GPT



Conteúdos

Visão geral

Organização dos discos

Sistemas de ficheiros FAT (12,16,32) NTFS EXT2,3,4

Representação de dados

Sistemas de ficheiros FAT

FAT: File Allocation Table

- Sistema de ficheiros simples, muito popular
- Utilizado primeiramente em sistemas DOS, Windows
- Atualmente é utilizado em pens USB, cartões de memória, smartphones

Sistemas de ficheiros 28/5

FAT12, FAT16, FAT32

- Versão indica o número de bits utilizado para referenciar clusters no disco.
 - ► FAT12 \rightarrow 12 bits \rightarrow Max: $2^{12} = 4.096$ clusters
 - ► FAT16 \rightarrow 16 bits \rightarrow Max: $2^{16} = 65,536$ clusters
 - ► FAT32 \rightarrow 32 bits \rightarrow Max: $2^{32} = 4,294,967,296$ clusters
- Cluster é um conjunto de sectores
- Sector é a unidade mínima de armazenamento de dados

Sistemas de ficheiros

Utilização do disco

- Sector tem usualmente 512 bytes de tamanho
- Sector é o tamanho mínimo para operações de leitura/escrita no disco
- Sempre que se utiliza um sector, este é considerado como totalmente ocupado
 - Se se guardar 10 bytes num sector, restantes 502 bytes são desperdiçados
 - Bytes não sobrescritos, mantém dados anteriores

Sistemas de ficheiros 30/5

Utilização do disco

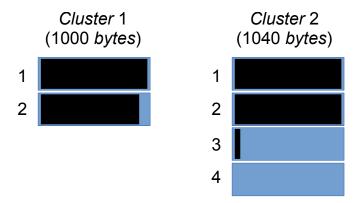
(2)

- Espaço em disco é alocado a ficheiros em conjuntos de sectores
- Número de sectores por cluster tem de ser uma potência de 2 (1,2,4,...)
- Cluster é a unidade mínima de alocação de ficheiros

Sistemas de ficheiros 31/5

Exemplos de *clusters*

Cluster 1 usa 2 sectores, já o cluster 2 usa 4 sectores



Espaço não utilizado no fim de cada ficheiro é chamado de folga (ou *slack*)

Sistemas de ficheiros 32

Layout de uma partição FAT

- ► Boot code costuma estar vazio
- Diretoria base conhecida como root directory
- FAT #2 é uma cópia de segurança da FAT #1

+info: http://ntfs.com/fat-partition-sector.htm

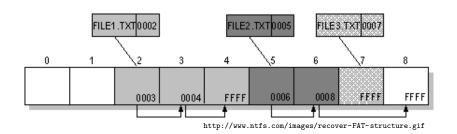
Sistemas de ficheiros 33/

Tipos de entradas em FAT #1

- Não utilizado (0x0000 0000)
- Cluster com erro (0xFFFF FFF7)
- ▶ Último *cluster* de um ficheiro (0xFFFF FFF8)
- Número do próximo cluster de um ficheiro

Sistemas de ficheiros 34/5

Exemplo de FAT #1



- ► FILE1.TXT: Ocupa clusters 2, 3 e 4
- ► FILE2.TXT: Ocupa *clusters* 5, 6 e 8 (Fragmentado)
- ► FILE3.TXT: Ocupa cluster 7

Sistemas de ficheiros 35/52

Conteúdo

Visão geral

Organização dos discos Master Boot Record GUID Partition Table

Sistemas de ficheiros

FAT (12,16,32)

NTFS

EXT2,3,4

Representação de dados Representação de números Representação de carateres

Sistemas de ficheiros NTFS

NTFS: New Technologies File System

- ▶ Sistema de ficheiros utilizado em Windows NT, 2000, ...
- Sistema proprietário da Microsoft
- Suporte para ficheiros superiores a 4GB
- ► Todos os registos são ficheiros (mesmo a própria \$MFT)

Sistemas de ficheiros 37/

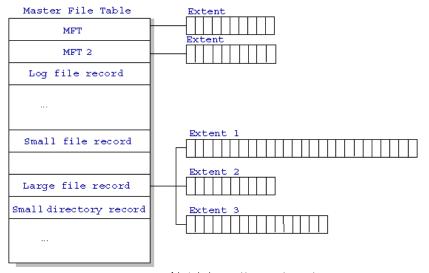
Layout de uma partição NTFS

- Partition Boot Sector poder ocupar de 1 a 16 setores
- Usa vários ficheiros de sistema (com metadados), como o \$MFT, \$Bitmap, \$LogFile, ...
- Mantém assinatura 0x55AA (posição 0x1FE)

+info: http://ntfs.com/ntfs-partition-boot-sector.htm

Sistemas de ficheiros 38/

Estrutura do *MFT*

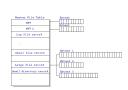


Adaptado de http://ntfs.com/images/NTFS-MFT-structure.gif

Estrutura do *MFT*

(2)

- ► Registo ocupa 1KB
- Cada registo pode conter um directório ou ficheiro (até aprox.512 bytes)
- Primeiro registo é a própria \$MFT
- Segundo registo é uma cópia de segurança do \$MFT (\$MftMirr)



Sistemas de ficheiros 40/5

Sumário

Visão geral

Organização dos discos Master Boot Record GUID Partition Table

Sistemas de ficheiros

FAT (12,16,32) NTFS

EXT2,3,4

Representação de dados Representação de números Representação de carateres

Sistemas de ficheiros EXT

EXT: Extended File System

- ► Vai atualmente na versão 4 (EXT4)
- Estrutura tem se mantido ao longo da sua evolução
- Novas versões traduzem novas funcionalidades

Sistemas de ficheiros 42/5

Layout de uma partição EXT

```
Padding
Padding
_____| <--- 0x400 (1024)
Super Block (1 bloco)
              _____ <-- 0x800 (Blocos 1KB)
Group descriptors
Block bitmap
Inode bitmap
Inode Table
Dados
```

- Organizado em blocos (de 1KB até 64KB)
- ► Bloco inicial de 1024 *bytes* é ignorado

+info: https://ext4.wiki.kernel.org/index.php/Ext4_Disk_Layout

Sistemas de ficheiros 43/

Sumário

Visão geral

Organização dos discos

Sistemas de ficheiros

Representação de dados

Representação de números

- ► Binário Representação digital (0-1)
- Octal Representação digital compacta (0-7)
- Decimal Representação humana (0-9)
- Hexadecimal Representação digital compacta (0-F)

Representação de números

Valor depende da posição

Numero 25102_d

- \triangleright 2 × 10⁴ + 5 × 10³ + 1 × 10² + 0 × 10¹ + 2 × 10⁰
 - Número 2 à esquerda vale mais (mais significativo)
 - Número 2 à direita vale menos (menos significativo)
- Big-endian byte mais significativo em primeiro lugar
- Litle-endian byte menos significativo em primeiro lugar

Representação de números

Ordem dos dados

- Processadores big-endian
 - ► Sparc, PowerPC, MIPS . . .
- Processadores litle-endian
 - z80, x86, x86-64, adm64 . . .
- Processadores programáveis (big/litle)
 - ► ARM ...
- Redes big-endian
 - Redes IP (com excepções)

Representação de carateres

Codificação ASCII

(American Standard Code for Information Interchange)

- Caracter ocupa um byte (sem problemas de endianness)
- Versão original usa apenas 7 bits
- Ocupa menos espaço que unicode
- Múltiplas versões estendidas (8 bits)
- ISO-8859 (latin-1) é mais comum
 - +info: http://www.asciitable.com

Representação de carateres

Codificação Unicode

- Representa carateres da generalidade das línguas
- Várias versões
 - ▶ UTF-8: 1 a 4 bytes, compatível com ASCII
 - ► UTF-16: 2 bytes ou 4 bytes
 - ► UTF-32: 4 bytes (fixo)

| Posição | 1 | 2 | 3 | 4 |
|----------|----|----|----|------------|
| ISO-8859 | 4F | 6C | E1 | |
| UTF-8 | 4F | 6C | C3 | A 1 |
| Texto | Ο | | á | |

Representação de dados 49/5

Magic numbers

(ou marcadores de tipo de ficheiro)

Exercício #5 (30 minutos)

Recupere os ficheiros contidos no ficheiro **img4** (disponível no moodle), que contém uma cópia de uma partição de um disco.

Utilize o comando file para analisar o ficheiro **img4** e os ficheiros recuperados.

O confirme o resultado do comando file, com recurso ao comando hexdump ou outro editor hexadecimal e à informação disponível nas ligações seguintes:

https://en.wikipedia.org/wiki/List_of_file_signatures

Submeta sua análise crítica pelo moodle (ficheiro PDF)

Análise de zip com ImHex

https://github.com/WerWolv/ImHex



Representação de dados 51/52

Bibliografia

- Handbook of Digital Forensics and Investigation; Eoghan Casey;
 Academic Press
- Description of FAT-12 File System, Thomas A. Anastasio, 2004
- ► FAT File Systems. FAT32, FAT16, FAT12; NTFS.com http://www.ntfs.com/fat-systems.htm
- ▶ NTFS File Systems; NTFS.com http://www.ntfs.com
- Ext4 (and Ext2/Ext3) Wiki, The Linux Kernel Archives https://ext4.wiki.kernel.org/index.php/Main_Page

Bibliografia 52/52