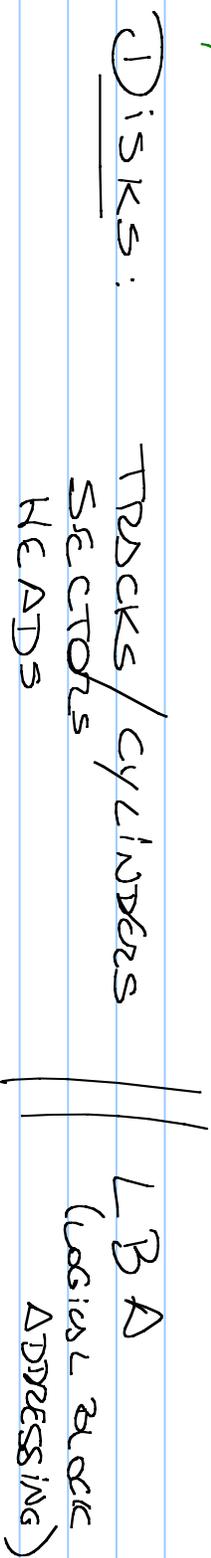


# LECTURE 15

## DISKS & FILE SYSTEMS

### @ THE PHYSICAL LEVEL



### @ THE LOGIC LEVEL

#### Filesystems:

- FAT
- NTFS
- JOURNALED (LINUX EXT3, MS USN JOURNAL...)
- etc

- CHANGES TO FILES INVOLVE MULTIPLE STEPS: DIRECTORY ENTRY, FILE CONTENT, ALLOCATION TABLES, ETC.

MBR vs. GUID PT (GPT)

MBR - MASTER BOOT RECORD  
 GUID-PT - GUID-based UNIFORM INTERFACED PARTITION TABLE

MBR (BIOS)

Structure of a Master Boot Record name="endianness"

| Address |      | Dec | Description   | Size in bytes                           |
|---------|------|-----|---|---|
| Hex     | Oct  |     |   |   |
| 0000    | 0000 | 0   | Code Area   | 440 (max 446)                           |
| 01B8    | 0670 | 440 | Optional Disk signature   | 4                                       |
| 01BC    | 0674 | 444 | Usually Nulls: 0x0000   | 2                                       |
| 01BE    | 0676 | 446 | Table of primary partitions<br>(Four 16-byte entries, IBM Partition Table scheme) | 64                                      |
| 01FB    | 0776 | 510 | 55h   | MBR signature:<br>0xAA55 <sup>[1]</sup> |
| 01FF    | 0777 | 511 | AAh   |   |

MBR, total size: 446 + 64 + 2 = 512

Layout of one 16-byte partition record

| Offset | Description  |
|--------|--|
| 0x00   | (1 byte) Status <sup>[7]</sup> (0x80 = bootable, 0x00 = non-bootable, other = malformed <sup>[8]</sup> ) |
| 0x01   | (3 bytes) Cylinder-head-sector address of the first sector in the partition <sup>[9]</sup>               |
| 0x04   | (1 byte) Partition type <sup>[10]</sup>  |
| 0x05   | (3 bytes) Cylinder-head-sector address of the last sector in the partition <sup>[11]</sup>               |
| 0x08   | (4 bytes) Logical block address of the first sector in the partition                                     |
| 0x0C   | (4 bytes) Length of the partition, in sectors  |

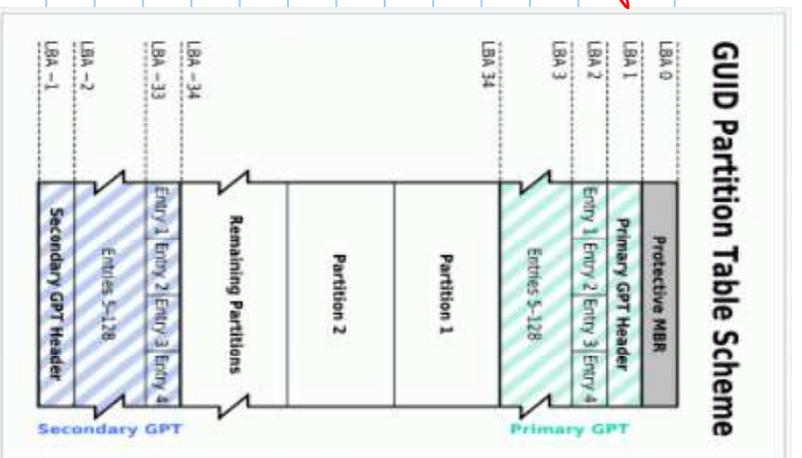
NOTE: BOOTSTRAP LOADER

# GPT (EFI-EXTENSIBLE FIRMWARE INTERFACE)

Partition table format

| Offset          | Length   | Contents  |
|-----------------|----------|---|
| 0               | 8 bytes  | Signature ("EFI PART", 45 46 49 20 50 41 52 54)                                   |
| 8               | 4 bytes  | Revision (For version 1.0, the value is 00 00 01 00)                              |
| 12              | 4 bytes  | Header size (in bytes, usually 5C 00 00 00 meaning 92 bytes)                      |
| 16              | 4 bytes  | CRC32 of header (cyclic redundancy check)   |
| 20              | 4 bytes  | reserved, must be zero  |
| 24              | 8 bytes  | Primary LBA (this LBA, always 1)  |
| 32              | 8 bytes  | Backup LBA (always equal to the last LBA on the disk)                             |
| 40              | 8 bytes  | First usable LBA for partitions (last partition table LBA + 1)                    |
| 48              | 8 bytes  | Last usable LBA (backup LBA - partition table size - 1)                           |
| 56              | 16 bytes | Disk GUID (also referred as UUID on UNIXes)                                       |
| 72              | 8 bytes  | Partition entries starting LBA (always right after header, 2)                     |
| 80              | 4 bytes  | Number of partition entries   |
| 84              | 4 bytes  | Size of a partition entry (usually 128)   |
| 88              | 4 bytes  | CRC32 of partition array  |
| 92              | *        | reserved, must be zeroes for the rest of the block (420 bytes for a 512-byte LBA) |
| <b>LBA Size</b> |          | <b>TOTAL</b>  |

NOTE: LBA's MAY BE LARGER THAN 512 bytes



GUID partition entry format

| Offset     | Length   | Contents  |
|------------|----------|---|
| 0          | 16 bytes | Partition type GUID                             |
| 16         | 16 bytes | Unique partition GUID                           |
| 32         | 8 bytes  | First LBA (little-endian)                       |
| 40         | 8 bytes  | Last LBA (inclusive, usually odd)               |
| 48         | 8 bytes  | Attribute flags (e.g. bit 60 denotes read-only) |
| 56         | 72 bytes | Partition name (36 UTF-16LE code units)         |
| <b>128</b> |          | <b>TOTAL</b>                                    |

# FAT's & NTFS (MFT)

| FAT                     |   |                                  |  |
|-------------------------|---|----------------------------------|--|
| Developer               | Microsoft   |                                  |  |
| Full Name               | File Allocation Table   |                                  |  |
|                         | (12-bit version)  | (16-bit version)                 | (32-bit version)   |
| Introduced              | 1980 (Seattle QDOS)   | November 1987, (Compaq DOS 3.31) | August 1996 (Windows Vista SP1)                                    |
| Partition identifier    | 0x01 (MBR)  | 0x04, 0x06, 0x0E (MBR)           | 0x08, 0x0C (MBR)<br>EB D9A4C2-89E5-4B3<br>-87C0-6886872699C7 (GPT) |
| Directory contents      | Structures  |                                  |  |
|                         | Table   |                                  |  |
| File allocation         | Linked List   |                                  | Linked List, Bitmap  |
| Bad blocks              | Cluster tagging   |                                  |  |
| Max file size           | 4 GB - 1 byte (or volume size if smaller)   |                                  | 2 <sup>64</sup> bytes (16 Exabytes)                                |
| Max cluster count       | 4,077 (2 <sup>12</sup> -19)   | 65,517 (2 <sup>16</sup> -19)     | 268,435,437 (2 <sup>28</sup> -19)                                  |
| Max filename size       | 8.3 filename, or 255 UTF-16 characters when using LFN   |                                  |  |
| Max volume size         | 32 MB   | 2 GB                             | 2 TB   |
|                         | 4GB with 64k clusters (natively supported)  | 8TB (2-KB sector)                |  |
| Dates recorded          | Features  |                                  |  |
|                         | Creation, modified, access (accuracy to day only)<br>(Creation time and access date are only available when LFN support is enabled) |                                  |  |
| Date range              | January 1, 1980 - December 31, 2107   |                                  |  |
| Forks                   | Not natively  |                                  |  |
| Attributes              | Read-only, hidden, system, volume label, subdirectory, archive  |                                  |  |
| Permissions             | No  |                                  |  |
| Transparent compression | Per-volume, Stackler, DoubleSpace, DriveSpace   | No                               | TBU  |
| Transparent encryption  | Per-volume only with DR-DOS   | No                               | TBU  |

| NTFS                            |   |
|---------------------------------|---|
| Developer                       | Microsoft   |
| Introduced                      | July 1993 (Windows NT 3.1)  |
| Partition identifier            | 0x07 (MBR)<br>2a00:5:2-89e5-4433-87c0-6886872699c7 (GPT)  |
| Directory contents              | Structures  |
|                                 | B+ tree [1]   |
| File allocation                 | Bitmap/Extents  |
| Bad blocks                      | \$badclus   |
| Max file size                   | Limits  |
|                                 | 2 <sup>64</sup> bytes (16 EB) minus 1 KiB [2]   |
| Max number of files             | 4,294,967,295 (2 <sup>32</sup> -1) [2]  |
| Max filename length             | 255 UTF-16 code units [3]   |
| Max volume size                 | 2 <sup>64</sup> - 1 clusters [2]  |
| Allowed characters in filenames | In Posix namespace, any UTF-16 code unit (case sensitive) except U+0000 (NUL) and code unit (case insensitive) except U+0000 (NUL) / (slash) \ (backslash) : (colon) * (asterisk) ? (Question mark) " (quote) < (less than) > (greater than) and   (pipe) [3] |

|                             | Features  |
|-----------------------------|---|
| Dates recorded              | Creation, modification, POSIX change, access  |
| Date range                  | 1 January 1601 - 28 May 60056 (file times are 64-bit numbers counting 100-nanosecond intervals (ten million per second) since 1601, which is 58,000+ years) |
| Date resolution             | 100ns   |
| Forks                       | Yes (see Alternate data streams below)  |
| Attributes                  | Read-only, hidden, system, archive, not content indexed, offline, temporary   |
| File system permissions     | ACLs  |
| Transparent compression     | Per-file, LZ77 (Windows NT 3.51 onward)   |
| Transparent encryption      | Per-file, DESX (Windows 2000 onward), Triple DES (Windows XP onward), AES (Windows XP Service Pack 1, Windows Server 2003 onward)                           |
| Single Instance Storage     | Yes   |
| Supported operating systems | Windows NT family (Windows NT 3.1 to XP, Windows Server 2003, Windows Vista, Windows Server 2008)   |

# EXT 3 (Linux)

|  |  |
|--|--|
| <b>ext3</b>                            |  |
| <b>Developer</b>                       | Stephen Tweedie  |
| <b>Full name</b>                       | Third extended file system                               |
| <b>Introduced</b>                      | November 2001 (Linux 2.4.15)                             |
| <b>Partition identifier</b>            | 0x83 (MBR)<br>EBD0A0A2-B9E5-4A33-87C0-6886872699C7 (GPT) |
| <b>Structures</b>                      |  |
| <b>Directory contents</b>              | Table, <b>hashed B-tree</b> with dir_index enabled       |
| <b>File allocation</b>                 | bitmap (free space), table (metadata)                    |
| <b>Bad blocks</b>                      | Table  |
| <b>Limits</b>                          |  |
| <b>Max file size</b>                   | 2 TiB  |
| <b>Max number of files</b>             | Variable, allocated at creation time <sup>[1]</sup>      |
| <b>Max filename length</b>             | 255 bytes  |
| <b>Max volume size</b>                 | 2 TiB - 16 TiB   |
| <b>Allowed characters in filenames</b> | All bytes except NUL and '/'                             |

|                                    |  |
|------------------------------------|--|
| <b>Features</b>                    |  |
| <b>Dates recorded</b>              | modification (mtime), attribute modification (ctime), access (atime)   |
| <b>Date range</b>                  | December 14, 1901 - January 18, 2038   |
| <b>Date resolution</b>             | 1s Nanosecond (using undocumented big inode)   |
| <b>Attributes</b>                  | No-time, append-only, synchronous-write, no-dump, h-tree (directory), immutable, journal, secure-delete, top (directory), allow-undelete |
| <b>File system permissions</b>     | Unix permissions, ACLs and arbitrary security attributes (Linux 2.6 and later)   |
| <b>Transparent compression</b>     | No   |
| <b>Transparent encryption</b>      | No (provided at the block device level)  |
| <b>Single Instance Storage</b>     | No   |
| <b>Supported operating systems</b> | Linux, BSD, Windows (through an IFS)   |