Reading a MS-DOS floppy

Due date:

The purpose of this exercise is to read the directory structure of an MS-DOS formatted floppy. You will be accessing the floppy device directly, accessing individual bytes and using them to interpret the disk structure. We will need to examine the structure of such a disk in great detail in order to accomplish this task. This exercise is adapted from "Kernel Projects for Linux", by Gary Nutt, Addison Wesley 2001.

Disk Organization

An MS-DOS disk is laid out by sectors, as follows:

Sector Number	Content
0	Boot sector
1	first sector of (first) FAT
10	first sector of second FAT (if there is one)
19	first sector in root directory
varies	first sector in data area

Disk Geometry

The first sector of the disk is referred to as the boot sector, and contains information regarding the disk layout. The format of the boot sector is summarized below, addresses are in hexadecimal:

0x00	0x02	jump instruction to 0x1e
0x03	0x0a	computer manufacturer's name
0x0b	0x0c	bytes per sector
0x0d	0x0d	sectors per cluster
0x0e	0x0f	reserved sectors for the boot record
0x10	0x10	number of FATs
0x11	0x12	maximum number of root directory entries
0x13	0x14	number of logical sectors
0x15	0x15	medium descriptor byte (defunct)
0x16	0x17	sectors per FAT
0x18	0x19	sectors per track
0x1a	0x1b	number of surfaces
0x1c	0x1d	number of hidden sectors
0x1e	...	bootstrap program

The File Allocation Table

The file allocation table has undergone changes over the years. The primary change has been in the number of bits per entry. On floppy disks however, the value has remained at its initial setting of 12 bits. The floppy structure also uses 1 sector per cluster, and a sector size of 512 bytes. Accessing the FAT is complicated by the fact that space was at a premium when the format was defined. They decided to pack two FAT entries into 3 bytes. Each FAT entry takes one and a half bytes. To make matters worse (at least for your instructor), a little-endian scheme is used for storing integers. For example, in a two-byte integer, the least significant byte is stored first, followed

by the most significant byte. So, the decimal number 7, for example, which is 0x0007 in hex, will be stored as 07 00 on the disk. So, in general:

- byte $3k$ contains the least significant bits of FAT[$2k$]

- the least significant 4 bits of byte $3k + 1$ contain the most significant 4 bits of FAT[$2k$]

- The most significant 4 bits of byte $3k + 1$ contain the least significant 4 bits of FAT[$2k + 1$]

- Byte $3k + 2$ contains the most significant 8 bits of FAT[$2k + 1$]

The first two entries in the FAT are reserved. The first is an indicator of disk capacity:

Entry	Heads	Sectors/Track	Cylinders	Capacity
0x0fe	1	8	40	160K
0x0ff	2	8	40	320K
0x0fc	1	9	40	180K
0x0fd	2	9	40	360K
0x0f9	2	9	80	720K
0x0f0	2	18	80	1.44M

The other entries in the FAT (beyond the first two) are interpreted as follows:

Value	Meaning
0x000	Unused
0xff0-0xff6	Reserved cluster
0xff7	Bad cluster
0xff8-0xfff	last cluster in file
else	number of the next cluster in file

Directory Entries

Directory entries are 32 bytes long and have the following format:

Offset	Length	Description
0x00	8	Filename
0x08	3	Extension
0x0B	1	bit field for attributes
0x0C	10	Reserved
0x16	2	Time
0x18	2	Date
0x1A	2	starting cluster number
0x1C	4	File size (in bytes)

If you wish to read and conver the time and date fields, time is stored encoded as $hour * 2048 + min * 32 + sec/2$ and date is stored as $(year - 1980) * 512 + month * 32 + day$.

Please note that the starting cluster number is a logical number. The first cluster of the data area (following the root directory area) is referred to as cluster 1. So you need to add an offset to this value (31 in the case of a floppy disk) to get the physical cluster (sector) number.

The attribute field is encoded as follows:

Bit	Attribute
0	Read-only
1	Hidden
2	System
3	volume label
4	subdirectory
5	archive
6&7	unused

A subdirectory is simply a file with the subdirectory bit turned on. The format of the file content is the same as the root directory described above. Finding all the entries in a directory may require a bit of work. All directory entries are initialized to have the first letter of the filename set to 0x00. When a file is deleted, the first letter is changed to 0xe5. You may stop looking for more files if you find a 0x00, but must continue if you find a 0xe5.

Program specifications:

Your program, or program suite should perform the following tasks:

1. Print the contents of the root directory block relevent to disk geometry, i.e. sectors per cluster, bytes per sector, number of reserved sectors, number of FATS, max number of root directory entries, number of sectors, sectors per FAT, sectors per track, and number of surfaces.

2. List the files (name, size, date, time, and status bits) in the root directory.

3. Copy a file from the root directory of the floppy to your local directory.

4. List contents of a subdirectory.