osFree Boot sequence (Draft II)

Notes:

1. This document is in development stage. You can send fixes and changes to it.
2. This document uses parts of Multiboot Specification document version 0.6.93
3. Here we use Intel notation and not AT&T assembler notation. It seems, we need to check document and in corresponding places fix it. But this is issue mostly for non-PC loaders.

(C) Copyright 2004-2010 osFree project

This document composed by Yuri Prokushev, Valery Sedletski, Sascha Schmidt.

Introduction

The osFree boot sequence does not use a classic solution like GRUB usage (unlike other L4-based projects). GRUB is a good program, but has some disadvantages. Main of them is, the loader must know about the file system (FS) structure. This means, to support a new file system, you need (as user) to update GRUB with a newer version to support the new FS (if any support is presented). As a developer, you need to add a file system driver to the kernel and to the boot loader. In most cases this means different architecture, programming style and development environment. We don't want to update the whole system to support only small advantages (in comparison with whole OS). We don't want to reinstall or upgrade most system components any time for “mouse pointer with shadow”. We want to have total cost of ownership at minimal level. As a result we reused the installable file system (IFS) approach from OS/2. We don't describe internals of MicroFSD and MiniFSD here because this is the task of an IFS document but not Kernel Loader internals and interfaces. Background information (Informative)

In this text we will try to discuss some problems about the way osFree must load from disk and the system initialization. osFree is an OS/2 clone, so it must follow the way of doing things of original OS/2. Also we must borrow the good ideas from OS/2 Warp Connect PowerPC Edition (aka Workplace OS), as it was the 1st example of microkernel OS/2, and had some essential features for microkernel OS/2 system. You can skip this section if you interested only normative part. OS/2 Boot sequence

At the end of POST procedure the ROM BIOS initializes devices and gives control to int 19h interrupt routine, which loads 1st sector of the 1st boot device (a floppy, HDD or another). If the device was the HDD, then the Master boot record (MBR) is loaded from the 1st sector. The ROM BIOS loads it at address 0x7c0:0×0. The MBR has a Non-System Bootstrap (NSB) in it, and the Partition Table (PT). The NSB code relocates MBR to 0×60:0×0 and loads the bootsector of boot HDD partition at the same place (0x7c0:0×0) MBR was loaded first. The boot partition is searched in the Partition Table (PT), which is embedded in the MBR sector at the end of it. The PT contains four partition descriptors, each of which has an active flag in the 1st byte of descriptor structure. If this byte is equal to 80h, the partition is active, if it is =00h, then partition is not active. MBR transfers control to the bootsector and the bootsector loads the operating system. This part of boot sequence is the same in different PC OSes including Windows, OS/2 and Linux when loading from the active primary partition.

In OS/2, to freely choose operating systems, different from OS/2, IBM included the OS/2 boot manager. The boot manager is installed in a small primary partition, which is marked active in PT. So, the MBR sector loads the bootsector of the OS/2 boot manager, instead of the bootsector of a

corresponding OS. The boot manager gives a menu to user, from which he or she can choose a partition to continue booting from. So, the boot manager then loads a bootsector of OS boot partition at the same address 0x7c0:0×0.

The bootsector has a BPB (Boot Parameters Block) structure in it, which describes some important parameters, needed to properly boot from the partition. The majority of them is specific for diskettes and the FAT filesystem, but some are essential for OS/2 to load properly. There are three essential parameters in the BPB for the OS/2 boot sequence. They are hiddenSectors value, the physical boot device and the logical boot device. The latter is equivalent to the boot device drive letter and is important to properly define the drive letter of the boot partition. The physical boot drive is the number of the boot physical device in the format of BIOS int 13h: the value of 00h is for 1st diskette drive, 01h for 2nd diskette, 80h for the 1st hard disk, 81h for the 2nd hard disk and so on. The hiddenSectors is important for booting OS/2 from logical disk in extended partition. For primary partitions, it is equal to the offset of the partition from the beginning of the HDD, but for logical disks it is for some reason equal to the number of sectors per track (63 for the modern hard drives). The hiddenSectors value is used to convert local sector number (from the beginning of the partition) to the global sector address (from the beginning of the HDD). It is essential for booting OS from the partition. For this reason, most OSes can load only from primary partition, like Windoze or FreeBSD. But IBM made OS/2 to be loadable from logical partitions, as well as primary ones. For this, IBM bootmanager fixes the three above mentioned values in bootsector BPB (previously loaded in memory), and only after that gives control to the bootsector. (This feature is required from boot manager to properly load OS/2 from the logical partition. At present, it is available only from three bootmanagers: IBM Bootmanager, VPart from Veit Kannegieser and AirBoot from Martin Kiewitz.)

After receiving control from MBR code or boot manager, the bootsector loads the so called blackbox code from the rest of the bootblock (the 15 sectors after the bootsector in HPFS, and 63 sectors in bootJFS) or from the root directory from os2boot file in FAT. The blackbox is the Micro File System Driver (MicroFSD or uFSD for short) and contains several functions to open, read and close files from the root directory of the boot drive. Also it has initialization code and cleanup code. Only one file can be opened at the same time, and, in IBM blackboxes, only files in the root directory can be read.

The blackbox (aka MicroFSD) in its initialization part, loads two files from disk: os2ldr and os2boot. (os2boot in FAT contains the MicroFSD, but in other filesystems, it contains the MiniFSD code). The os2ldr is the OS/2 kernel loader. It is independent from the filesystem, and only for FAT contains a special code (the above mentioned functions to read files from the FAT partition, and some code to support memory dumping and hybernating from/to the FAT partition). Besides that, os2ldr implements DosHlp functions (helpers for the OS/2 kernel), so, it serves like some sort of microkernel – it implements some functions the kernel depends on. Also, os2ldr contains implementation of the OEMHLP$ device driver. (Yes, OEMHLP$ resides in the loader!). So, os2ldr is more than just OS loader :).

The blackbox code transfers control to the os2ldr, and gives it the info about memory layout and exports its filesystem functions (it gives the loader a FileTable structure, a pointer to the BPB, a physical boot device and some flags). The loader then relocates itself to the top of low memory, loads os2krnl from disk and applies required fixups (it does LX format parsing and placing segments to required places and corrects addresses of functions and variables for proper linkage).

After that, control is given to os2krnl, and memory info along with os2boot MiniFSD image in memory is given to the kernel by the loader.

For disk reading, os2krnl uses MiniFSD, not MicroFSD. MicroFSD is intended for loader to work in real mode, and MiniFSD is for the kernel to work in protect mode. MiniFSD has the format of 16-bit NE DLL.

Its size is limited to 62 Kbytes. The system initialization sysinit routine of os2krnl loads MiniFSD from its in-memory image. It is used at the 1st stage in system initialization process to read config.sys and load base device drivers (BASEDEV and PSD).

Before the kernel loads OS/2 disk subsystem drivers (they are: ibm1flpy.add, ibm1s506.add, – disk drivers, an ATAPI filter to support CDROMs, DASD manager (os2dasd.dmd), volume manager (os2lvm.dmd)), the disk reading is performed by switching to real mode and calling int 13h BIOS disk read functions. When disk subsystem drivers are loaded and initialized, the kernel uses them to read the disk, and for filesystem access it continues to use the MiniFSD. It is the 1st phase of the system initialization process, as it called in ifs.inf documentation from IBM. At this phase the kernel uses MFS_* MiniFSD functions for filesystem access. They are very similiar to the MicroFSD functions, but unlike them, they work in protect mode. At the 2nd phase the kernel links MiniFSD into the IFS chain (it is the only IFS in chain), calls FS_INIT function to complete phase 1 and in phase 2 it uses FS_* functions from the MiniFSD, so at this moment, MiniFSD is like ordinary IFS, but it supports only a minimum of FS_* funtions.

After that, phase 3 begins. At this phase, the boot IFS is loaded, which replaces MiniFSD in the IFS chain. At this stage the kernel calls the MFS_TERM MiniFSD entry point, and the latter transfers control to the boot IFS. After that a full-featured filesystem access begins. The system now can read from and write to files from any disk, it supports drive letters and can have many open files at the same time. At this phase, the kernel can load ordinary device drivers (“device=”) and IFS'es.

After that, the system continue to load “device=” drivers, then process “run=” and “call=” statements, and then “protshell=” to load pmshell. Some info about OS/2 PPC load process

I have no PowerPC box, but I have IBM's redbook “OS/2 Warp (PowerPC Edition) A first look”, there is a small paragraph about OS/2 PPC load sequence in this book. This paragraph is very small and contains only a small piece of information. Also I have OS/2 PPC config files for the loader (boot.cfg) and for the OS/2 server (config.sys). The most part of drivers, servers and libraries is loaded from boot.cfg, and the config.sys is specific to the OS/2 personality. The config.sys file is very small.

As written in IBM's redbook, the loader (bl_auto file in the root of the disk) is loaded by the PowerPC ROM directly from special partition without a filesystem structure; there is no a bootsector, but the bootloader is just written over this type 0×41 partition. This partition is called “A PowerPC PReP bootloader partition”. The loader has a configuration file boot.cfg which resides on the boot partition. (The boot partition is an ordinary primary partition formatted with a FAT filesystem; HPFS support was too unstable). In the config file, there are the microkernel file, an initial task and other files, which loader must load from disk into memory. These sevices loaded from boot.cfg are called Personality neutral (PN) services, they are independent from OS/2 Personality and include device drivers.

The bootloader contains a filesystem support code in it, and SCSI, IDE, floppy, ATAPI extensions to access corresponding devices through firmware. We can imagine that the firmware contained the disk-related code, similar to int 13h BIOS routines (but we have no authoritative information sources which say about that).

The bootloader loads a number of files into memory, then starts the microkernel and the initial task. The initial task is called the bootstrap. The bootloader passes the bootstrap some information along with the locations of files it have loaded in memory. The bootstrap acts as a file server for other servers. In other words, it gives access to the files the bootloader had loaded into memory. The bootstrap task has no device drivers in it, instead it has access to the files the bootloader has loaded into memory.

Then the bootstrap do the following (I will quote the text from the redbook):

1. It loads the Root Name Server
2. Starts the default pager
3. Starts the task manager
4. Provides the file services, which will be used by Task Server
5. Directs the Task Manager to start the personality neutral (PN) servers required to bring up the dominant personality. PN servers include Message Logger, Hardware Resource Manager (HRM), Bus Walkers, and Device Drivers.
6. Starts the Personality.

The bootstrap task continues to behave as a file server until it terminates.

Then OS/2 personality starts. The OS/2 personality server parses config.sys and loads the OS/2 personality specific servers.

Device drivers are not specific to the OS/2 personality, so they are started by bootstrap and are in bootloader config file (boot.cfg), not in the config.sys file.

L4 microkernel load process. GNU GRUB bootloader and Multiboot specification

The L4 microkernel can be started either in real or protect mode. If started in real mode, it switches to protect mode by itself. The exact load procedure is described in the L4 API, version X.2 reference manual.

For loading the L4 microkernel, the GNU GRUB bootloader is commonly used. The GRUB defines the Multiboot specification, which is intended to be the common protocol between the OS kernel and the bootloader. Originally, only GRUB supported the multiboot specification, but it is possible to create a compatible bootloader. For example, there is x.exe the Multiboot compliant DOS extender intended to start FreeDOS32 from 16-bit DOS. It is multiboot compliant and loads FreeDOS32 which requires a multiboot compliant loader. It can use x.exe as well as GRUB. The multiboot compliant kernels include The HURD Mach kernel (The GNU GRUB is the officialGNU project bootloader, and it was created specifically for the GNU HURD project. But it also suits for Linux, FreeBSD, OpenBSD, NetBSD, MacOS X and L4). But L4 itself is not a multiboot kernel, instead, it uses its own loader/bootstrapper which in L4Ka::Pistachio is kickstart, and in L4/Fiasco is rmgr (note: at this time, rmgr is splitted into two parts: bootstrap loader and the root task).

The Multiboot specification requires from the kernel to have in its first 8192 bytes a structure called the multiboot header. This header defines requirements for the loader from the kernel, such as: the load addresses of various segments of a kernel, initial video mode and kernel entry point. The kernel executable file format may be any, the only requitement is to have the multiboot header. But GRUB also directly supports the ELF and a.out formats. (But OS/2 (intel) uses LX format, and OS/2 PPC used ELF format).

The bootloader loads a kernel (kickstart in our case) and a number of additional modules. The bootloader loads the kernel and applies fixups to it, but modules remain untouched, the loader starts the kernel and passes to it the pointer to Multiboot structure. The GRUB leaves the kernel in a simple protected mode environment with paging disabled, A20 line enabled and Interrupt Controller remains uninitialized. Also, the initial video mode is set.

The Multiboot structure contains info about memory layout and modules. For the kernel and modules

there are strings associated with them. These strings can be used as command lines for kernel and modules, or just as a labels to identify modules.

In L4Ka:Pistachio, the kickstart bootstrapper plays the role of the multiboot kernel. It receives Multiboot structure from GRUB, and ELF-loads the L4 kernel and initial servers. Then it searches the Kernel Interface Page (KIP) inside the L4 image. It passes the multiboot info in Bootinfo structure, pointed by the field in KIP. Then kickstart fills the fields for initial servers location in the KIP (initial servers are sigma0 and roottask, which were passed by GRUB as multiboot modules). Then kickstart calls the entry point in L4 kernel.

In the case of L4/Fiasco the role of kickstart plays the resource manager (rmgr). It consists of two stages. Stage 1 is analogous to kickstart. It parses the config and loads L4 and servers. Stage 1 passes the configuration to stage 2. Stage 2 is started by L4 and serves as a root server. But at present, rmgr is divided into 2 parts – bootstrap and roottask, which are loaded as separate multiboot modules.

Then L4 starts, relocates itself to the proper place in memory, and then starts sigma0 and roottask. After that, the roottask can initialize the rest of the system.

A historical note about FreeLDR

History of FreeLDR. A note about OS2CSM

OS2CSM idea

Ideas about FreeLdr design

1) First, I suggest to combine the functionality of OS/2 bootmanager and os2ldr in one program. I.e., the boot sequence must be like this: The MBR loads an active partition, or the partition with a given number. (I already wrote such a MBR sector, it can load a bootsector from selected primary or logical partition (yes, logical partitions are supported too!)). The bootable hard disk number and the number of partition on it are written inside the MBR of the first hard disk. (The bootable partition can reside on the same HDD as well as on the different HDD, than the 1st HDD we read MBR from).

So, the MBR loads a boot sector from bootable partition. The boot sector loads the blackbox. The blackbox loads our loader. Then, the loader starts. The loader combines the functionality of a bootloader with functionality of bootmanager: after having been called from the blackbox, the loader shows a menu to the user. The user selects a menu item from it, each menu item defines an OS to be loaded along with parameters, which are then passed to the OS kernel. From this point, the loader/bootmanager is capable of executing the bootsectors of OSes, not supported directly by our loader, like windoze. The loader only loads a corresponding bootsector and executes it. But if an OS kernel is supported directly, then the loader can also pass some parameters to the kernel, through a config file or a command line parameters.

The advantage of such an approach is that we can choose an OS and its parameters from the same place, it is a combined loader/bootmanager. An example: in present OS/2 the OS/2 bootmanager allows to choose an OS, and os2ldr allows to choose additional parameters – it allows to press a hotkey to bring up a Recovery choices menu. There are also many other parameters avaliabe by pressing Alt-F1[F2,F3,...]. In our case, there is only one menu from which an OS and its parameters

can be choosen, and additional menus, like Recovery choices, are avaliable from the same place: the bootloader/bootmanager menu. So, the advantage is an integration.

Also, our bootmanager resides on ordinary OS/2 partition, not special bootmanager partition. For reading files it uses a microfsd. And all settings of the bootmanager/bootloader can be stored in ordinary text config files.

And finally, our idea of hybrid loader/bootmanager allows us to load different OS/2 kernel versions from the same partition. And not only kernel, any system component version can be selected from the same place – they can be selected from within the bootmanager menu.

For more details, read on.

2) The loader present a menu to the user. Each menu item corresponds to the boot script. The script contains commands to ask additional info from user (i.e., this command shows a menu, user changes parameters, and parameters are returned to the loader. Then parameters constitute the loader “environment”. The environment strings can substitute variables in command lines and config files), to change current partition, to define variables etc. Also the boot script contains definition of files, loaded by the bootloader. The loader distinguishes between executable files (the loader performs executable format parsing), files that are only loaded by the loader, but its format is not parsed, and config files. The files marked as configs are preprocessed by the preprocessor.

So, there are following config files: i) The loader menu definition file, it contains a definition of menu items. Each menu item has a loader script associated with it. This config file is similar to the menu.lst file in GRUB. ii) the boot scripts. Each script is referenced or included by menu definition config. Each script is similar to boot.cfg file in OS/2 PPC. iii) config.sys file. It is specific for OS/2 personality. These configs are read through microfsd calls and can be preprocessed. There may be additional config files for individual servers. They also use the loader config preprocessing facilities, so parameters defined in the loader script or the ones asked from user may substitute variables in these config files. So we can flexibly set parameters of each system component from the bootmanager menu.

Also, for better flexibility, we can make a small config for the blackbox. When the blackbox is started by the bootsector, it may read its config file, from which it knows, what files it must load as the minifsd (it is optional, we may not use minifsd, so it may be not necessary to load it) and freeldr main module. (See the next paragraph: The idea of modules).

3) The idea of modules.

At present time, the loader is a COM file, so its size is limited to 64 Kb. To include more functionality, it may be necessary to make it a multi-segment program, so, a better EXE format must be used. Because it is 16-bit real mode program, the DOS EXE format and, probably, OS/2 NE format are suitable for that. The DOS EXE format seems to be the most simple, so it is simpler to implement FreeLdr as a DOS EXE file.

To keep the loader modular (to have possibility to load only needed parts of it, and the possibility to load/unload parts of it at any time), I suggest to implement it as a set of modules. A module must work in real mode and can be implemented as a DOS EXE file. (We can't use DLL's in real mode, so we must design a simple mechanism based on DOS EXE files). I propose a module to be a DOS EXE file with additional header. The header helps to locate functions inside the EXE file. The header begins with a pointer to the ASCIIZ string which contains a module name. After this pointer follows a size of the header, then a size of the DOS executable after the header and then follows a table of structures, which can be described as:

```
struct {
   char *FuncName;
   unsigned long EntryPoint;
} *pFuncTable;
```

i.e., each structure defines a function in EXE file, it links a function name with its offset in the EXE file. This array of structures is followed by a string table, which contains all the function names and a module name. Each FuncName pointer points to the sting in this string table. The header helps to locate each function in the EXE file. The function table can be generated from linker map file.

The main loader module is loaded by microfsd. It has a DOS EXE format, so, to execute it, we must load it by some executable format loader. The DOS EXE loader can be implemented as a DOS COM file. I suggest to concatenate the DOS EXE loader with the FreeLdr main module (the main module is glued to the tail of the EXE loader, this idea is borrowed from the MS NTLDR: the NTLDR consists of the startup COM file glued with the PE format executable). The FreeLdr startup receives info from the blackbox, loads and relocates the main module from its tail, executes and passes it an info received from the blackbox.

The main module contains a mechanism to load other modules from files on disk. It loads a module as a DOS EXE file, and links its header to the headers list. When performing relocations to the module, the DOS EXE format loader corrects the addresses in headers, which are linked to the list. From this list, the main module can locate any function from any module. For that purpose, the main module supplies a function to be called from other modules. This function takes a module name and a name of a function in it as parameters, and returns an entry point to this function. This way, any module can find an entry point to any function with given name in any other module with given name. So, we have a kind of name service, which can convert a function name to its address. Any module can call any function in another module.

The microfsd's, loaders for different file formats (DOS EXE, NE, LX, ELF), config file preprocessor etc. can be implemented as separate modules.

4) We can implement additional executable formats loaders for formats, other than ELF. (For example, LX, NE, PE(?)...). They can be implemented as separate modules

5) To be possible to read files from other partition than a boot one, it is possible to implement a blackbox switching. For that, the loader can have a command, executed from the boot script, to change the current drive, like “root” command in GRUB. For this, the loader loads a new microfsd for the changed partition filesystem. It updates the FileTable structure by pointers to functions in the new microfsd. It also loads BPB from the new partition bootsector (and patches the HiddenSectors value, if needed). So, this feature can give the loader possibility to read files from several partitions, switching them.

6) I propose to make the loader capable of loading standard multiboot kernels, L4 kernel (as a kind of a multiboot kernel), and custom kernels, like OS/2 kernel. Before, in this text the idea of modules was described. The idea is to implement a loader as a set of loadable modules. Custom OS kernels, not compatible with multiboot specification, can be supported by custom loader module. The multiboot support can also be implemented as a separate module. The module is loaded by the Freeldr main module. By writing support module for custom kernel type, the developers from outside can extend our loader to load their kernels. There may be, our loader will suit not only unix or OS/2 or L4 kernels, but windoze and ReactOS kernels too. We can't expect uncle Billy to make support for loading windoze kernel from our loader, but it is possible that ReactOS guys can make their kernel loadable

from it.  The part of the loader loaded by the blackbox is called the general part of the loader. The general part contains only support functions for loading modules, locating functions in them, the DOS EXE format loader, and some more functions. It passes control to the custom part along with interfaces to module loader, microfsd's, and the info obtained from the blackbox of a bootable partition. The custom loader part implements support for loading specific kind of kernel.

a) For loading multiboot kernels, the multiboot specific part of a loader can be implemented as a separate module. Through it, we can load L4 kernel, as well as most unix kernels.

b) For loading ordinary OS/2 kernels, there must be a specific custom part. This custom part takes from general part the info, obtained from the blackbox. It loads os2ldr file from disk and passes this info to os2ldr. Then the boot process continues as usual. In future, the custom part can be extended, so it will fully replace os2ldr functionality as David Zimmerli wanted – his idea was exactly to replace os2ldr functionality. c)

For loading unsupported kernels, the loader can only load corresponding OS bootsector, and give control to it. It may be implemented like GRUB “chainloader” command. Suggested boot sequence

1) Do we need a MiniFSD?

If we look at the OS/2 PPC and L4 boot sequence, we may conclude that OS/2 PPC bootloader and GRUB do similar things. They load a kernel and a set of files into memory and start the kernel. Then the bootstrap task, in 1st case, and a root task in 2nd case, will start other tasks; the FreeLdr also must do equivalent tasks. As at this moment the filesystem is not yet initialized, the bootstrap or root task can only access files that were already loaded by the bootloader.

To use a filesystem, we must first load the disk driver (ibm1s506.add or ibm1flpy.add), dasd manager, volume manager and filesystem driver (or their equivalents in our microkernel system). Let assume that we can use a minifsd as a filesystem driver. With L4, we can't use 16-bit programs as easy as in present OS/2. So, our minifsd must be 32-bit. And 62 Kb limit for its size is not applicable here, as we use 32-bit programs.

As written in ifs.inf file from IBM, a minifsd has two modes of operation – at phase 1 and at phase 2 of boot process. Before phase 1, when minifsd initializes, it can not call any dynalink calls (in MFS_INIT initialization routine). Ordinary IFS in its FS_INIT routine, can call external dlls. The only external functions a minifsd can call are MFSH_* helpers called from the kernel. The full-featured IFS's can call a much wider number of external calls. They are FSH_* IFS helpers. (but, as I understood, the IFS can't call other external functions, besides FSH_* calls in routines other than its FS_INIT routine). For the IFS to be possible to call external DLLs at IFS init time, the dynamic loading support must be working and operational, and the DLLs itself must be available. They can be only loaded by minifsd (because IFS is not initialized yet) or they may be passed by the bootloader.

The reason why minifsd is used by the OS/2 (intel) boot process is to have a limited filesystem access after the kernel switched into protected mode. Before the OS/2 disk subsystem drivers are loaded, the disk read is performed by temporarily switching into real mode and calling int 13h disk read functions. After the disk subsystem is loaded, the disk read is performed through OS/2 disk driver.

In L4 or in OS/2 PPC, before disk drivers are loaded, we can't switch to real mode to call int 13 routines. Consequently, the disk drivers must be read through the microfsd by the bootloader, and then they must be passed to roottask through memory. So, the disk drivers can be started from their memory images by the roottask.

But then why we must use a minifsd? We can load the full-featured boot IFS immediately, and before that, we can start ELF format dynamic loader support servers and load all the DLLs boot IFS needs. All these files can be passed by the bootloader along with disk subsystem drivers.

So, it has not much sense to use minifsd in L4 boot sequence, only microfsd is needed. And if we look at OS/2 PPC, then we will see that it has not a minifsd and basedev's loading phase. Instead, all required services are passed to the bootstrap task by the bootloader. In our case, FreeLdr and kickstart play the role of OS/2 PPC bootloader, and roottask is analogous to the bootstrap task. So, we must to make similar design solutions.

2) In case of loading L4, the loader loads kickstart L4 bootstrapper. Kickstart is given a set of modules by the bootloader through the multiboot structure. The kickstart then passes this info in bootinfo structure, pointed by the field in KIP. When L4 is started, it starts sigma0 and roottask. The roottask can obtain info from bootinfo structure, which can be reached from the KIP. The KIP address can be obtained through the KernelInterface() L4 system call. So, the roottask can obtain info about modules, passed by kickstart. Then the roottask can find the needed modules in the bootinfo structure. The purpose of each module can be defined through strings associated with modules, each module can be marked by special tag, which defines its purpose (e.g., the module contains a library, a config, a root name server, an executable files loader server. etc.). The tag can be contained in string along with command line for this module. This way, an info about modules is passed from loader through kickstart to roottask, and roottask can find needed servers and load them in proper order. By that, the roottask can implement a funcionality of OS/2 PPC bootstrap task. The roottask first starts the personality neutral (PN) services, then brings up the OS personalities (OS/2, L4Linux, etc.). Overview of osFree boot sequence (Normative)

When the computer is turned on or reseted, the first program to be executed is the BIOS. There are different BIOSes with different execution sequence. We only have to know one thing: The BIOS (Standard, BOOTrom, PXE-BIOS or something else) loads the boot sector (under term 'boot sector' we understand not only actual harddisk boot sector or PXE boot image, but any first code which is loaded by the hardware and executed) and passes control to it.

And here our boot sequence starts. We name all code executed before our boot sector as BlackBox. We don't know how it works. We only know we have control passed to our code. Our boot sector is storage depended 16-bit code. The boot sector loads a first stage loader named MicroFSD, MiniFSD and Kernel Loader. MicroFSD is Micro File System Driver. MiniFSD is Mini File System Driver. Kernel Loader is code which loads and executes the osFree kernel (or any multiboot compatible kernel, if user prefers).

The code in the boot sector fills information structures and passes them to the Kernel Loader. Those Information structures contain information about the current memory allocation and MicroFSD entry points.

The Kernel Loader is 16-bit/32-bit mixed code. It loads a multiboot-compatible kernel image, reallocates it and MiniFSD in memory, links entry points, switches the CPU to protected mode and passes control to the Kernel. Kernel and MiniFSD are not 16-bit code but 32-bit. Data types and constants

In folowing description of interfaces we use common way of types definition. Here table with types desctiprion:

Type	Base type	Description
USHORT16	unsigned short int	16-bit unsigned short integer

Type	Base type	Description
ULONG32	unsigned long int	32-bit unsigned long integer

MicroFSD/KernelLoader interface (Normative)

The MicroFSD/KernelLoader interface is the same as for OS/2. After MicroFSD has loaded all required code (MiniFSD image, OS2LDR image), it passes control to KernelLoader (OS2LDR). The CPU must be in real mode and the CPU registers must be filled like in the following table.

When initially transferring control to OS2LDR from a “black box”, the following interface is defined:

Register	Contains	Description
DH	Boot mode flags	Provides information about current state of boot process
DL	Boot disk drive number This parameter is ignored if either the BBF_NOVOLIO or BBF_MINIFSD flags are not set.	
DS:SI	Pointer to the BOOT Media's BPB	This parameter is ignored if either the BBF_NOVOLIO or BBF_MINIFSD flags are not set.
ES:DI	Pointer to a Memory map structure	

Boot mode flags are folowing:

Constant name	Constant value	Description
BBF_NOVOLIO	0×01	indicates that the miniFSD does not use MFSH_DOVOLIO
BBF_RIPL	0×02	indicates that boot volume is not local (RIPL boot)
BBF_MINIFSD	0×04	indicates that a miniFSD ispresent.
BBF_RESERVED1	0×08	must be zero.
BBF_MICROFSD	0×10	indicates that a microFSD is present.
BBF_RESERVED2	0×20	must be zero.
BBF_RESERVED3	0×40	must be zero.
BBF_RESERVED4	0×80	must be zero.

```
struct MemoryMap
{
   USHORT16 mmt_centries;   /* # of entries in this table
*/
   USHORT16 mmt_ldrseg;      /* paragraph # where OS2LDR is loaded   */
   ULONG32  mmt_ldrlen;         /* length of OS2LDR in bytes
*/
   USHORT16 mmt_museg;     /* paragraph # where microFSD is loaded  */
   ULONG32  mmt_mulen;        /* length of microFSD in bytes
*/
   USHORT16 mmt_mfsseg;    /* paragraph # where miniFSD is loaded    */
   ULONG32  mmt_mfslen;       /* length of miniFSD in bytes
*/
   USHORT16 mmt_ripseg;      /* paragraph # where RIPL data is loaded  */
   ULONG32  mmt_riplen;         /* length of RIPL data in bytes.
*/
```

```
   /* The next four elements are pointers to microFSD entry points
*/
   USHORT16 (far *mmt_muOpen)(char far *pName, unsigned long far
*pulFileSize);
   ULONG32  (far *mmt_muRead)(long loffseek, char far *pBuf, unsigned long
cbBuf);
   ULONG32  (far *mmt_muClose)(void);
   ULONG32  (far *mmt_muTerminate)(void);
}
```

The microFSD entry points interface is defined as follows:

- mu_Open is passed a far pointer to the name of the file to be opened and a far pointer to a ULONG to return the file size. The re-turned value (in AX) indicates success(0) or failure (non-0).
- mu_Read is passed a seek offset, a far pointer to a data buffer, and the size of the data buffer. The returned value(in DX:AX) indicates the number of bytes actually read.
- mu_Close has no parameters and expects no return value. It is a signal to the micro-FSD that the loader is done reading the current file.
- mu_Terminate has no parameters and expects no return value. It is a signal to the micro-FSD that the loader has finished reading the boot drive.

The loader will call the microFSD in a Open-Read-Read-….-Read-Close sequence for each file read in from the boot drive. After all files are loaded, mu_Terminate must be called.

KernelLoader/Kernel interface (Normative)

The KernelLoader/Kernel interface is not OS/2 compatible but multiboot compatible. This means you can load different kernels, for example a Linux kernel.

There are three main aspects of a Kernel loader/Kernel image interface:

- The format of an Kernel image as seen by a Kernel loader.
- The state of a machine when a Kernel loader starts an operating system.
- The format of information passed by a Kernel loader to an operating system.

Kernel image format

A Kernel image may be an ordinary 32-bit executable file in the standard format for that particular operating system, except that it may be linked at a non-default load address to avoid loading on top of the PC's I/O region or other reserved areas, and of course it should not use shared libraries or other fancy features.

A Kernel image must contain an additional header, called Multiboot header, besides the headers of the format used by the Kernel image. The Multiboot header must be contained completely within the first 8192 bytes of the Kernel image, and must be longword (32-bit) aligned. In general, it should come as early as possible and may be embedded in the beginning of the text segment after the real executable header. The layout of the Multiboot header

The layout of the Multiboot header must be as follows:

Offset	Type	Field Name	Note
0	ULONG32	Magic	required
4	ULONG32	Flags	required
8	ULONG32	checksum	required
12	ULONG32	header_addr	if flags[16] is set
16	ULONG32	load_addr	if flags[16] is set
20	ULONG32	load_end_addr	if flags[16] is set
24	ULONG32	bss_end_addr	if flags[16] is set
28	ULONG32	entry_addr	if flags[16] is set
32	ULONG32	mode_type	must be ignored
36	ULONG32	width must be	ignored
40	ULONG32	height	must be ignored
44	ULONG32	depth	must be ignored

The fields magic, flags and checksum are defined in Header magic fields, the fields header_addr, load_addr, load_end_addr, bss_end_addr and entry_addr are defined in Header address fields, and the fields mode_type, width, height and depth are defind in Header graphics fields. Because we consider Kernel loader only must load and execute kernel Header graphics fields ignored by Kernel loader.

magic

The field magic is the magic number identifying the header, which must be the hexadecimal value 0x1BADB002 refered as MULTIBOOT_MAGIC constant.

flags

The field flags specifies features that the Kernel image requests or requires of an Kernel loader. Bits 0-15 indicate requirements; if the kernel loader sees any of these bits set but doesn't understand the flag or can't fulfill the requirements it indicates for some reason, it must notify the user and fail to load the Kernel image. Bits 16-31 indicate optional features; if any bits in this range are set but the kernel loader doesn't understand them, it may simply ignore them and proceed as usual. Naturally, all as-yet-undefined bits in the flags word must be set to zero in Kernel images. This way, the flags fields serves for version control as well as simple feature selection.

If bit 0 (defined by MULTIBOOT_PAGE_ALIGN constant) in the flags word is set, then all boot modules loaded along with the operating system must be aligned on page (4KB) boundaries. Some operating systems expect to be able to map the pages containing boot modules directly into a paged address space during startup, and thus need the boot modules to be page-aligned.

If bit 1 (defined by MULTIBOOT_MEMORY_INFO constant) in the flags word is set, then information on available memory via at least the mem_* fields of the Multiboot information structure (see Boot information format) must be included. If the kernel loader is capable of passing a memory map (the mmap_* fields) and one exists, then it may be included as well.

If bit 2 (defined by MULTIBOOT_VIDEO_MODE constant) in the flags word set then loading of the Kernel must be stoped. This is because we don't support this feature. Most possible in future we will just tell to the Kernel standard video mode info.

If bit 16 (defined by MULTIBOOT_AOUT_KLUDGE constant) in the flags word is set, then the fields at offsets 8-24 in the Multiboot header are valid, and the kernel loader should use them instead of the fields in the actual executable header to calculate where to load the Kernel image. This information does not need to be provided if the kernel image is in ELF format, but it must be provided if the image is in a.out format or in some other format. Compliant kernel loaders must be able to load images that either are in ELF format or contain the load address information embedded in the Multiboot header; they may also directly support other executable formats, such as particular a.out variants, but are not required to.

checksum

The field checksum is a 32-bit unsigned value which, when added to the other magic fields (i.e. magic and flags), must have a 32-bit unsigned sum of zero.

The address fields of Multiboot header

All of the address fields enabled by flag bit 16 (MULTIBOOT_AOUT_KLUDGE) are physical addresses. The meaning of each is as follows:

header_addr

Contains the address corresponding to the beginning of the Multiboot header – the physical memory location at which the magic value is supposed to be loaded. This field serves to synchronize the mapping between Kernel image offsets and physical memory addresses.

load_addr

Contains the physical address of the beginning of the text segment. The offset in the Kernel image file at which to start loading is defined by the offset at which the header was found, minus (header_addr - load_addr). load_addr must be less than or equal to header_addr.

load_end_addr

Contains the physical address of the end of the data segment. (load_end_addr - load_addr) specifies how much data to load. This implies that the text and data segments must be consecutive in the Kernel image; this is true for existing a.out executable formats. If this field is zero, the kernel loader assumes that the text and data segments occupy the whole Kernel image file.

bss_end_addr

Contains the physical address of the end of the bss segment. The kernel loader initializes this area to zero, and reserves the memory it occupies to avoid placing boot modules and other data relevant to the operating system in that area. If this field is zero, the kernel loader assumes that no bss segment

is present.

entry_addr

The physical address to which the kernel loader should jump in order to start running the operating system.

Machine state

When the kernel loader invokes the 32-bit operating system, the machine must have the following state:

Registers	Contains	Description
EAX	Magic value	Must contain the magic value 0x2BADB002; the presence of this value indicates to the operating system that it was loaded by a Multiboot-compliant kernel loader (e.g. as opposed to another type of kernel loader that the operating system can also be loaded from).
EBX	Pointer to Multiboot information structure	Must contain the 32-bit physical address of the Multiboot information structure provided by the kernel loader (see Boot information format).
CS	Code segment	Must be a 32-bit read/execute code segment with an offset of 0 and a limit of 0xFFFFFFFF. The exact value is undefined.
DS/ES/FS/GS/SS	Data segment	Must be 32-bit read/write data segments with an offset of 0 and a limit of 0xFFFFFFFF. The exact values are all undefined.

Also:

- A20 gate Must be enabled.
- CR0 Bit 31 (PG) must be cleared. Bit 0 (PE) must be set. Other bits are all undefined.
- EFLAGS Bit 17 (VM) must be cleared. Bit 9 (IF) must be cleared. Other bits are all undefined.

All other processor registers and flag bits are undefined. This includes, in particular:

- ESP The Kernel image must create its own stack as soon as it needs one.
- GDTR Even though the segment registers are set up as described above, the GDTR may be invalid, so the Kernel image must not load any segment registers (even just reloading the same values!) until it sets up its own GDT.
- IDTR The Kernel image must leave interrupts disabled until it sets up its own IDT.

However, besides this the machine state should be left by the kernel loader in normal working order, i.e. as initialized by the BIOS (or DOS, if that's what the kernel loader runs from). In other words, the operating system should be able to make BIOS calls and such after being loaded, as long as it does not overwrite the BIOS data structures before doing so. Also, the kernel loader must leave the PIC programmed with the normal BIOS/DOS values, even if it changed them during the switch to 32-bit mode.

Boot information format

Upon entry to the operating system, the EBX register contains the physical address of a Multiboot information data structure, through which the kernel loader communicates vital information to the operating system. The operating system can use or ignore any parts of the structure as it chooses; all information passed by the kernel loader is advisory only.

The Multiboot information structure and its related substructures may be placed anywhere in memory by the kernel loader (with the exception of the memory reserved for the kernel and boot modules, of course). It is the operating system's responsibility to avoid overwriting this memory until it is done using it.

The format of the Multiboot information structure (as defined so far) follows:

Offset	Type	Field Name	Note
0	ULONG32	Flags	(required)
4	ULONG32	mem_lower	(present if flags[0] is set)
8	ULONG32	mem_upper	(present if flags[0] is set)
12	ULONG32	boot_device	(present if flags[1] is set)
16	ULONG32	Cmdline	(present if flags[2] is set)
20	ULONG32	mods_count	(present if flags[3] is set)
24	ULONG32	mods_addr	(present if flags[3] is set)
28 - 40	ULONG32	syms	(present if flags[4] or flags[5] is set)
44	ULONG32	mmap_length	(present if flags[6] is set)
48	ULONG32	mmap_addr	(present if flags[6] is set)
52	ULONG32	drives_length	(present if flags[7] is set)
56	ULONG32	drives_addr	(present if flags[7] is set)
60	ULONG32	config_table	(present if flags[8] is set)
64	ULONG32	boot_loader_name	(present if flags[9] is set)
68	ULONG32	apm_table	(present if flags[10] is set)
72	ULONG32	vbe_control_info	(must be filled by)
76	ULONG32	vbe_mode_info	
80	ULONG32	vbe_mode	
82	ULONG32	vbe_interface_seg	
84	ULONG32	vbe_interface_off	
86	ULONG32	vbe_interface_len	

The first longword indicates the presence and validity of other fields in the Multiboot information structure. All as-yet-undefined bits must be set to zero by the kernel loader. Any set bits which the operating system does not understand should be ignored. Thus, the flags field also functions as a version indicator, allowing the Multiboot information structure to be expanded in the future without breaking anything.

If bit 0 in the flags word is set, then the mem_* fields are valid. mem_lower and mem_upper indicate the amount of lower and upper memory, respectively, in kilobytes. Lower memory starts at address 0, and upper memory starts at address 1 megabyte. The maximum possible value for lower memory is 640 kilobytes. The value returned for upper memory is maximally the address of the first upper memory hole minus 1 megabyte. It is not guaranteed to be this value.

If bit 1 in the flags word is set, then the boot_device field is valid, and indicates which BIOS disk device the kernel loader loaded the Kernel image from. If the Kernel image was not loaded from a BIOS disk, then this field must not be present (bit 3 must be clear). The operating system may use this field as a hint for determining its own root device, but is not required to. The boot_device field is laid out in four one-byte subfields as follows:

Drive	part1	part2	part3

The first byte contains the BIOS drive number as understood by the BIOS INT 0×13 low-level disk interface: e.g. 0×00 for the first floppy disk or 0×80 for the first hard disk.

The three remaining bytes specify the boot partition. part1 specifies the top-level partition number, part2 specifies a sub-partition in the top-level partition, etc. Partition numbers always start from zero. Unused partition bytes must be set to 0xFF. For example, if the disk is partitioned using a simple one-level DOS partitioning scheme, then part1 contains the DOS partition number, and part2 and part3 are both 0xFF. As another example, if a disk is partitioned first into DOS partitions, and then one of those DOS partitions is subdivided into several BSD partitions using BSD's disklabel strategy, then part1 contains the DOS partition number, part2 contains the BSD sub-partition within that DOS partition, and part3 is 0xFF.

DOS extended partitions are indicated as partition numbers starting from 4 and increasing, rather than as nested sub-partitions, even though the underlying disk layout of extended partitions is hierarchical in nature. For example, if the kernel loader boots from the second extended partition on a disk partitioned in conventional DOS style, then part1 will be 5, and part2 and part3 will both be 0xFF.

If bit 2 of the flags longword is set, the cmdline field is valid, and contains the physical address of the command line to be passed to the kernel. The command line is a normal C-style zero-terminated string.

If bit 3 of the flags is set, then the mods fields indicate to the kernel what boot modules were loaded along with the kernel image, and where they can be found. mods_count contains the number of modules loaded; mods_addr contains the physical address of the first module structure. mods_count may be zero, indicating no boot modules were loaded, even if bit 1 of flags is set. Each module structure is formatted as follows:

Offset	**Type**	**Field Name**	**Note**
0	ULONG32	mod_start	
4	ULONG32	mod_end	
8	ULONG32	String	
12	ULONG32	reserved(0)	

The first two fields contain the start and end addresses of the boot module itself. The string field provides an arbitrary string to be associated with that particular boot module; it is a zero-terminated ASCII string, just like the kernel command line. The string field may be 0 if there is no string associated with the module. Typically the string might be a command line (e.g. if the operating system treats boot modules as executable programs), or a pathname (e.g. if the operating system treats boot modules as files in a file system), but its exact use is specific to the operating system. The reserved field must be set to 0 by the kernel loader and ignored by the operating system.

Caution: Bits 4 & 5 are mutually exclusive.

If bit 4 in the flags word is set, then the following fields in the Multiboot information structure starting

at byte 28 are valid:

Offset	Type	Field Name	Note
28	ULONG32	Tabsize	
32	ULONG32	Strsize	
36	ULONG32	Addr	
40	ULONG32	reserved (0)	

These indicate where the symbol table from an a.out kernel image can be found. addr is the physical address of the size (4-byte unsigned long) of an array of a.out format nlist structures, followed immediately by the array itself, then the size (4-byte unsigned long) of a set of zero-terminated ASCII strings (plus sizeof(unsigned long) in this case), and finally the set of strings itself. tabsize is equal to its size parameter (found at the beginning of the symbol section), and strsize is equal to its size parameter (found at the beginning of the string section) of the following string table to which the symbol table refers. Note that tabsize may be 0, indicating no symbols, even if bit 4 in the flags word is set.

If bit 5 in the flags word is set, then the following fields in the Multiboot information structure starting at byte 28 are valid:

Offset	Type	Field Name	Note
28	ULONG32	Num	
32	ULONG32	Size	
36	ULONG32	Addr	
40	ULONG32	Shndx	

These indicate where the section header table from an ELF kernel is, the size of each entry, number of entries, and the string table used as the index of names. They correspond to the shdr_* entries (shdr_num, etc.) in the Executable and Linkable Format (ELF) specification in the program header. All sections are loaded, and the physical address fields of the ELF section header then refer to where the sections are in memory (refer to the i386 ELF documentation for details as to how to read the section header(s)). Note that shdr_num may be 0, indicating no symbols, even if bit 5 in the flags word is set.

If bit 6 in the flags word is set, then the mmap_* fields are valid, and indicate the address and length of a buffer containing a memory map of the machine provided by the BIOS. mmap_addr is the address, and mmap_length is the total size of the buffer. The buffer consists of one or more of the following size/structure pairs (size is really used for skipping to the next pair):

Offset	Type	Field Name	Note
-4	ULONG32	Size	
0	ULONG32	base_addr_low	
4	ULONG32	base_addr_high	
8	ULONG32	length_low	
12	ULONG32	length_high	
16	ULONG32	Type	

where size is the size of the associated structure in bytes, which can be greater than the minimum of 20 bytes. base_addr_low is the lower 32 bits of the starting address, and base_addr_high is the upper 32 bits, for a total of a 64-bit starting address. length_low is the lower 32 bits of the size of the memory region in bytes, and length_high is the upper 32 bits, for a total of a 64-bit length. type is the

variety of address range represented, where a value of 1 indicates available RAM, and all other values currently indicated a reserved area.

The map provided is guaranteed to list all standard RAM that should be available for normal use.

If bit 7 in the flags is set, then the drives_* fields are valid, and indicate the address of the physical address of the first drive structure and the size of drive structures. drives_addr is the address, and drives_length is the total size of drive structures. Note that drives_length may be zero. Each drive structure is formated as follows:

Offset	Type	Field Name	Note
0	ULONG32	Size	
4	BYTE	drive_number	
5	BYTE	drive_mode	
6	USHORT16	drive_cylinders	
8	BYTE	drive_heads	
9	BYTE	drive_sectors	
10-xx		drive_ports	

The size field specifies the size of this structure. The size varies, depending on the number of ports. Note that the size may not be equal to (10 + 2 * the number of ports), because of an alignment.

The drive_number field contains the BIOS drive number. The drive_mode field represents the access mode used by the kernel loader. Currently, the following modes are defined:

- 0 - CHS mode (traditional cylinder/head/sector addressing mode)
- 1 - LBA mode (Logical Block Addressing mode)

The three fields, drive_cylinders, drive_heads and drive_sectors, indicate the geometry of the drive detected by the BIOS. drive_cylinders contains the number of the cylinders. drive_heads contains the number of the heads. drive_sectors contains the number of the sectors per track.

The drive_ports field contains the array of the I/O ports used for the drive in the BIOS code. The array consists of zero or more unsigned two-bytes integers, and is terminated with zero. Note that the array may contain any number of I/O ports that are not related to the drive actually (such as DMA controller's ports).

If bit 8 in the flags is set, then the config_table field is valid and indicates the address of the ROM configuration table returned by the GET CONFIGURATION BIOS call. If the BIOS call fails, then the size of the table must be zero.

If bit 9 in the flags is set, the boot_loader_name field is valid, and contains the physical address of the name of a kernel loader booting the kernel. The name is a normal C-style zero-terminated string.

If bit 10 in the flags is set, the apm_table field is valid, and contains the physical address of an APM table defined as below:

Offset	Type	Field Name	Note
0	USHORT16	version	
2	USHORT16	cseg	
4	USHORT16	Offset	

Offset	Type	Field Name	Note
8	USHORT16	cseg_16	
10	USHORT16	Dseg	
12	USHORT16	Flags	
14	USHORT16	cseg_len	
16	USHORT16	cseg_16_len	
18	USHORT16	dseg_len	

The fields version, cseg, offset, cseg_16, dseg, flags, cseg_len, cseg_16_len, dseg_len indicate the version number, the protected mode 32-bit code segment, the offset of the entry point, the protected mode 16-bit code segment, the protected mode 16-bit data segment, the flags, the length of the protected mode 32-bit code segment, the length of the protected mode 16-bit code segment, and the length of the protected mode 16-bit data segment, respectively. Only the field offset is 4 bytes, and the others are 2 bytes. See Advanced Power Management (APM) BIOS Interface Specification, for more information.

The bit 11 in the flags must be zero. For Installable File System drivers developers

Installable File System (IFS) drivers are described in the IFS document. Kernel Loader internals

The Kernel Loader is raw 16-bit/32-bit binary code (like MS/PC-DOS COM files, but started not from 100h but from 0h).

- First of all, Kernel Loader stores all information from CPU registers into internal structures
- Depending on this information stores info about memory allocation
- After this it show information on display
- Loads kernel
- Swtiches to protected mode
- And executes multiboot compatible Kernel

Nothing more here! Wasn't it easy? 

From:
http://www.osfree.su/doku/ - **osFree wiki**

Permanent link:
http://www.osfree.su/doku/doku.php?id=en:docs:boot:bootseq2&rev=1400714898

Last update: **2014/05/21 23:28**