Loading and overlays

$Revision: 1.6 $
$Date: 1999/01/02 03:35:15 $

Loading is the process of bringing a program into main memory so it can run. In this chapter we look at the loading process, concentrating on loading programs that have already been linked. Many systems used to have linking loaders that combined the linking and loading process, but those have now practically disappeared, with the only one I know of on current hardware being on MVS. Linking loaders weren't all that different from plain linkers, with the primary and obvious difference being that the output was left in memory rather than placed in a file.

Basic loading

We touched on most of the basics of loading in Chapter 3, in the context of object file design. Loading is a little different depending on whether a program is loaded by mapping into a process address space via the virtual memory system or just read in using normal I/O calls.

On most modern systems, each program is loaded into a fresh address space, which means that all programs are loaded at a known fixed address, and can be linked for that address. In that case, loading is pretty simple:

- Read enough header information from the object file to find out how much address space is needed.
- Allocate that address space, in separate segments if the object format has separate segments.
- Read the program into the segments in the address space.
- Zero out any bss space at the end of the program if the virtual memory system doesn't do so automatically.
- Create a stack segment if the architecture needs one.
- Set up any runtime information such as program arguments or environment variables.

- Start the program. If the program isn't mapped through the virtual memory system, reading in the object file just means reading in the file with normal "read" system calls. On systems which support shared read-only code segments, the system needs to check whether there's already a copy of the code segment loaded in and use that rather than making another copy.

On systems that do memory mapping, the process is slightly more complicated. The system loader has to create the segments, then arrange to map the file pages into the segments with appropriate permissions, read-only (RO) or copy-on-write (COW). In some cases, the same page is double mapped at the end of one segment and the beginning of the next, RO in one and COW in the other, in formats like compact Unix a.out. The data segment is generally contiguous with the bss segment, so the loader has to zero out the part of the last page after the end of the data (since the disk version usually has symbols or something else there), and allocate enough zero pages following the data to cover the bss segment.

Basic loading, with relocation

A few systems still do load time relocation. Some, like MS-DOS, lack usable hardware relocation. Others, like MVS, have hardware relocation but are descended from systems that didn't have it. Some have hardware relocation but can load multiple executable programs into the same address space, so linkers can't count on having specific addresses available.

As discussed in Chapter 7, load-time relocation is far simpler than link-time relocation, because the entire program is relocated as a unit. If, for example, the program is linked as though it would be loaded at location zero, but is in fact loaded at location 15000, all of the places in the program that require fixups will get 15000 added. After reading the program into memory, the loader consults the relocation items in the object file and fixes up the memory locations to which the items point.

Load-time relocation can present a performance problem, because code loaded at different virtual addresses can't usually be shared between address spaces, since the fixups for each address space are different. One approach, used by MVS and to some extent by Windows is to create a shared memory area present in all address spaces and load oft-used programs into

that. (MVS calls this this link pack area.) This also has the problem that different processes don't get separate copies of writable data, so the application has to be written to allocate all of its writable storage explicitly.

Position-independent code

One popular solution to the dilemma of loading the same program at different addresses is position independent code (PIC). The idea is simple, write the program so that the code doesn't change if it's loaded at different addresses and separate the code from the data, o the code can be shared regardless of where it's loaded, with only data pages being private to each process.

This is a surprisingly old idea. TSS/360 used in in 1966, and I don't believe it was original there. (TSS was notoriously buggy, but I can report from personal experience that the PIC features really worked.)

On modern architectures, it's not difficult to generate PIC executable code. Jumps and branches are generally either PC-relative or relative to a base register set at runtime, so no load-time relocation is required for them. The problem is with data addressing. The code can't contain any direct data addresses, since those would be relocatable and wouldn't be PIC. An obvious solution is to create a table of data addresses in a data page and keep a pointer to that table in a register, so the code can use indexed addressing relative to that register to pick up the data. This works, but there's still the question of how to get the initial data address into the register.

TSS/360 position independent code

TSS took a brute-force approach. Every routine had two addresses, the address of the code, known as the V-con (short for V style address constant, which even non-PIC code needed) and the address of the data, known as the R-con. The standard OS/360 calling sequence requires that the caller provide an 18 word register save area pointed to by register 13. TSS extended the save area to 19 words and required that the caller place callee's R-con into that 19th word before making the call, Figure 1. Each routine had in its data segment the V-cons and R-cons for all of the routines that it called, and stored the appropriate R-con into the outgoing save

area before each call. The main routine in a program received a save area from the operating system which provided the initial R-con.

Figure 1: TSS style two-address procedure call

TSS style with R-con in the save area

This scheme worked, but is poorly suited for modern systems. For one thing, copying the R-cons made the calling sequence bulky. For another, it made procedure pointers two words, which didn't matter in the 1960s but is an issue now since in programs written in C, all pointers have to be the same size. (The C standard doesn't mandate it, but far too much existing C code assumes it to do anything else.) A simple modification used in some Unix systems is to treat the address of a procedure's data as the address of the procedure, and to place a pointer to the procedure's code at that address, Figure 2. To call a procedure, the caller loads the data address into an agreed data pointer register, then loads the code address from the location pointed to by the data pointer into a scratch register and calls the routine. This is easy to implement, and has adequate if not fabulous performance.

Figure 2: Code via data pointers

ROMP style data table with code pointer at the beginning.

ELF position independent code

Unix System V.4 introduced a faster and cleverer PIC scheme for its shared libraries which is now universally used by systems that use ELF executables, Figure 3. It has the advantage of returning to the normal convention that the address of a procedure is the address of the code for the procedure, regardless of whether one is calling PIC code, found in shared

ELF libraries, or non-PIC code, found in regular ELF executables.

Its designers noticed that an executable consists of a group of code pages followed by a group of data pages, and regardless of where in the address space the program is loaded, the offset from the code to the data doesn't change. So if the code can load its own address into a register, the data will be at a known distance from that address. The linker creates a global offset table (GOT) containing pointers to all of the global data that the executable file addresses. (Each shared library has its own GOT, and if the main program were compiled with PIC, which it normally isn't, it would have a GOT as well.) Since the linker creates the GOT, there is only one pointer for each datum regardless of how many routines in the executable refer to it.

If a procedure needs to refer to global or static data, it's up to the procedure itself to load up the address of the GOT. The details vary by architecture, but the 386 code is typical. It consists of a call instruction to the immediately following location, which has the effect of pushing the PC on the stack but not jumping, then a pop to get the saved PC in a register and an add immediate of the difference between the address the GOT and address the target of the call. Procedures that don't address static data needn't load the GOT address.

Figure 3: PIC code and data with fixed offsets

picture of code page showing constant offset to data even though loaded at different addresses in different address spaces.

Once the GOT register is loaded, code can reference static data using the GOT register as a base register, since the distance from a static datum in the program's data segment to the GOT is fixed at link tine. Addresses of global data aren't bound until the program is loaded (see Chapter 10), so to reference global data, code has to load a pointer to the data from the GOT and then deference the pointer. This extra memory reference makes

programs somewhat slower, although it's a cost that most programmers are willing to pay for the convenience of dynamically linked libraries. Speed critical code can use static shared libraries (Chapter 9) or no shared libraries at all.

To support PIC, ELF defines a handful of special relocation types for code that uses the GOT. The exact types are architecture-specific, but the x86 is typical:

- `R_386_GOT32`: The relative location of the slot in the GOT where the linker has placed a pointer to the given symbol. Used for indirectly referenced global data.
- `R_386_GOTOFF`: The distance from the base of the GOT to the given symbol or address. Used to address static data relative to the GOT.
- `R_386_GOTPC`: The PC-relative address of the position in the GOT. Used once per routine to find the base of the GOT, as the value added to a PC value saved by a CALL.
- `R_386_RELATIVE`: Used to mark data addresses in a PIC shared library that need to be relocated at load time.

Tree structured overlays

We close this chapter with a description of tree-structured overlays, a widely used scheme in the days before virtual memory to fit programs into memories smaller than the programs. Overlays are another technique that dates back to before 1960, and are still in use in some memory-constrained environments. Several MS-DOS linkers in the 1980 supported them in a form nearly identical to that used 25 years earlier on mainframe computers. Although overlays are now little used, the techniques that linkers use to create and manage overlays remain interesting, and the inter-segment call tricks developed for overlays point the way to dynamic linking. In environments like DSPs with constrained address spaces, overlay techniques can be a good way to squeeze programs in, especially since overlay managers tend to be small. The OS/360 overlay manager is only about 500 bytes, and I once wrote one for a graphics processor with a 512 word address space that used only a dozen words or so.

Overlaid programs divide the code into a tree of segments, such as the one in Figure 4.

Figure 4: A typical overlay tree

ROOT calls A and D. A calls B and C, D calls E and F.

The programmer manually assigns object files or individual object code segments to overlay segments. Sibling segments in the overlay tree share the same memory. In the example, segments A and D share the same memory, B and C share the same memory, and E and F share the same memory. The sequence of segments that lead to a specific segment is called a path, so the path for E includes the root, D, and E.

When the program starts, the system loads the root segment which contains the entry point of the program. Each time a routine makes a "downward" inter-segment call, the overlay manager ensures that the path to the call target is loaded. For example, if the root calls a routine in segment A, the overlay manager loads section A if it's not already loaded. If a routine in A calls a routine in B the manager has to ensure that B is loaded, and if a routine in the root calls a routine in B, the manager ensures that both A and B are loaded. Upwards calls don't require any linker help, since the entire path from the root is already loaded.

Calls across the tree are known as *exclusive* calls and are usually considered to be an error since it's not possible to return. Overlay linkers let the programmer force exclusive calls for situations where the called routine is known not to return.

Defining overlays

Overlay linkers created overlaid executables from ordinary input object files. The objects don't contain any overlay instructions, Intstead, the programmer specifies the overlay structure with a command language that the linker reads and interprets. Figure 5 shows the same overlay structure as before, with the names of the routines loaded into each segment.

Figure 5: A typical overlay tree

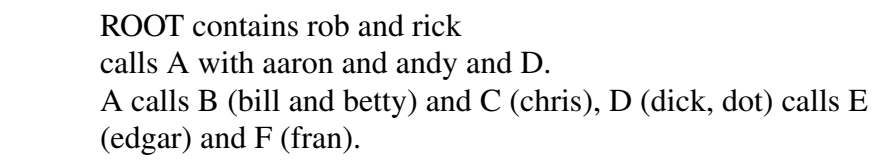

Figure 6 shows the linker commands that one might give to the IBM 360 linker to create this structure. Spacing doesn't matter, so we've indented the commands to show the tree structure. OVERLAY commands define the beginning of each segment; commands with the same overlay name define segments that overlay each other. Hence the first OVERLAY AD defines segment A, and the second defines segmnt D. Overlay segments are defined in a depth first left to right tree walk. INCLUDE commands name logical files for the linker to read.

Figure 6: Linker commands

```
INCLUDE ROB
INCLUDE RICK
OVERLAY AD
  INCLUDE AARON, ANDY
  OVERLAY BC
    INCLUDE BILL, BETTY
  OVERLAY BC
    INCLUDE CHRIS
OVERLAY AD
  INCLUDE DICK, DOT
  OVERLAY EF
    INCLUDE EDGAR
  OVERLAY EF
    INCLUDE FRAN
```

It's up to the programmer to lay out overlays to be space effiecent. The storage allocated for each segment is the maximum length of any of the segments that occupy the same space. For example, assume that the file lengths in decimal are as follows.

Name	Size
rob	500
rick	1500
aaron	3000
andy	1000
bill	1000
betty	1000
chris	3000
dick	3000
dot	4000
edgar	2000
fran	3000

The storage allocation, looks like Figure 7. Each segment starts immediately after the preceding segment in the path, and the total program size is the length of the longest path. This program is fairly well balanced, with the longest path being 11500 and the shortest being 8000. Juggling the overlay structure to find one that is as compact as possible while still being valid (no exclusive calls) and reasonably efficient is a black art requiring considerable trial and error. Since the overlays are defined entirely in the linker, each trial requires a relink but no recompilation.

Figure 7: Overlay storage layout

```
0 rob
500 rick

2000 aaron     2000 dick
3000 andy    5000 dot

6000 bill 6000 chris
7000 betty  9000 ----9000 edgar9000 fran
```

```
8000 ----     11000 ----12000 ----
```

Implementation of overlays

The implementation of overlays is surprisingly simple. Once the linker determines the layout of the segments, relocates the code in each segment appropriately based on the memory location of the segment. The linker needs to create a segment table which goes in the root segment, and, in each segment, glue code for each routine that is the target of a downward call from that segment.

The segment table, Figure 8, lists each segment, a flag to note if the segment is loaded, the segment's path. and information needed to load the segment from disk.

Figure 8: Idealized segment table

```
struct segtab {
  struct segtab *path;// preceding segment in path
  boolean ispresent;// true if this segment is loaded
  int memoffset; // relative load address
  int diskoffset;  // location in executable
  int size;   // segment size
} segtab[];
```

The linker interposes the glue code in front of each downward call so the overlay manager can ensure that the required segment(s) are loaded. Segments can use glue code in higher level but not lower level routines. For example, if routines in the root call aaron, dick, and betty, the root needs glue code for each of those three symbols. If segment A contains calls to bill, betty, and chris, A needs glue code for bill and chris, but can use the glue for betty already present in the root. All downward calls (which are to global symbols) are resolved to glue code, Figure 9, rather than to the

actual routine. The glue code has to save any registers it changes, since it has to be transparent to the calling and called routine, then jump into the overlay manager, providing the address of the real routine and an indication of which segment that address is in. Here we use a pointer, but an index into the segtab array would work as well.

Figure 9: Idealized glue code for x86

```
glue'betty: jsr load_overlay
  .long betty // address of real routine
  .long segtab+N // address of segment B's segtab
```

At runtime, the system loads in the root segment and starts it. At each downward call, the glue code calls the overlay manager. The manager checks the target segment's status. If the segment is present, the manager just jumps to the real routine. If the segment is not present, the manager loads the target segment and any unloaded preceding segments in the path, marks any conflicting segments as not present, marks the newly loaded segments as present, and jumps.

Overlay fine points

As always, details make elegant tree structured overlays messier than they might be.

Data

We've been talking about structuring code overlays, without any consideration of where the data goes. Individual routines may have private data loaded into the segments with the routines, but any data that has to be remembered from one call to the next needs to be promoted high enough in the tree that it won't get unloaded and reloaded, losing any changes made. In practice, it means that most global data usually ends up in the root. When Fortran programs are overlaid, overlay linkers can position common blocks appropriately to be used as communication areas. For example, if dick calls edgar and fran, and the latter two both refer to a common block,

that block has to reside in segment D to be a communication area.

Duplicated code

Frequently the overall structure of an overlaid program can be improved by duplicating code. In our example, imagine that chris and edgar both call a routine called greg which is 500 bytes long. A single copy of greg would have to go in the root, increasing the total loaded size of the program, since placing it anywhere else in the tree would require a forbidden exclusive call from either chris or edgar. On the other hand, if both segments C and E include copies of greg, the overall loaded size of the program doesn't increase, since the end of segment C would grow from 9000 to 9500 and of E from 11000 to 11500, both still smaller than the 12000 bytes that F requires.

Multiple regions

Frequently, a program's calling structure doesn't map very well to a single tree. Overlay systems handle multiple code regions, with a separate overlay tree in each region. Calls between regions always go through glue code. The IBM linker supports up to four regions, although in my experience I never found a use for more than two.

Overlay summary

Even though overlays have been rendered largely obsolete by virtual memory, they remain of historical interest because were the first significant use of link-time code generation and modification. They require a great deal of manual programmer work to design and specify the overlay structure, generally with a lot of trial and error "digital origami", but they were a very effective way to squeeze a large program into limited memory.

Exercises

Compile some small C routines with PIC and non-PIC code. How much slower is the PIC code than non-PIC? Is it enough slower to be worth having non-PIC versions of libraries for programmers in a hurry?

In the overlay example, assume that dick and dot each call both edgar and fran, but dick and dot don't call each other. Restructure the overlay so that dick and dot share the same space, and adjust the structure so that the call

tree still works. How much space does the overlaid program take now?

In the overlay segment table, there's no explicit marking of conflicting segments. When the overlay manager loads a segment and the segment's path, how does the manager determine what segments to mark as not present?

In an overlaid program with no exclusive calls, is it possible that a series of calls could end up jumping to unloaded code anyway? In the example above, what happens if rob calls bill, which calls aaron, which calls chris, then the routines all return? How hard would it be for the linker or overlay manager to detect or prevent that problem?