LẬP TRÌNH HỆ THỐNG

ThS. Đỗ Thị Thu Hiền (hiendtt@uit.edu.vn)



TRƯỜNG ĐH CÔNG NGHỆ THÔNG TIN - ĐHQG-HCM

KHOA MẠNG MÁY TÍNH & TRUYỀN THÔNG

FACULTY OF COMPUTER NETWORK AND COMMUNICATIONS

Tầng 8 - Tòa nhà E, trường ĐH Công nghệ Thông tin, ĐHQG-HCM Điện thoại: (08)3 725 1993 (122)

Linking



Ví dụ: Chương trình C có code trong nhiều file

```
int sum(int *a, int n);
int array[2] = {1, 2};
int main()
{
   int val = sum(array, 2);
   return val;
}
```

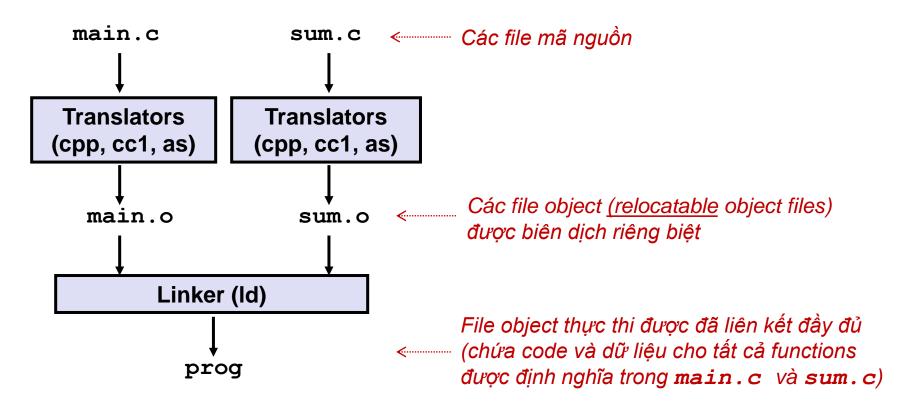
```
int sum(int *a, int n)
{
   int i, s = 0;

   for (i = 0; i < n; i++) {
      s += a[i];
   }
   return s;
}</pre>
```

File **main.c** có sử dụng hàm **sum()** được định nghĩa trong file **sum.c**

Static linking – Liên kết tĩnh

- Các chương trình được biên dịch và liên kết bằng compiler
 - linux> gcc -Og -o prog main.c sum.c
 - linux> ./prog



Vì sao phải sử dụng Linker?

Lý do 1: Tính mô-đun

- Chương trình có thể được viết dưới dạng gồm nhiều file source nhỏ thay vì viết nguyên một file source code lớn.
- Có thể build thư viện của một số hàm thông dụng
 - Ví dụ: Thư viện toán (Math library), các thư viện C chuẩn

Vì sao phải sử dụng Linker?

Lý do 2: Tính hiệu quả

- Về thời gian: Biên dịch riêng biệt
 - Thay đổi 1 file source thì chỉ biên dịch lại 1 file source đó và liên kết lại.
 - Không cần phải biên dịch lại các file source khác.
- Về không gian: Libraries Các thư viện
 - Các hàm thông dụng có thể tích hợp vào chung 1 file
 - Tuy nhiên các file thực thi và bộ nhớ đang thực thi chỉ chứa code của những function mà nó sử dụng.

Linker sẽ làm gì?

- Cố gắng ánh xạ các tham chiếu hàm/biến đến định nghĩa tương ứng của chúng trong các file code
 - Tham chiếu và định nghĩa có thể nằm cùng file hoặc khác file
- Sắp xếp lại code/dữ liệu của các file code một cách phù hợp trong file thực thi cuối cùng

Linker: Khái niệm Symbol

Symbol là gì?

- Symbol là các hàm (function) hoặc các biến toàn cục
- Các chương trình có thể định nghĩa hoặc tham chiếu đến symbol

```
void swap() {...} /* define symbol swap */
swap(); /* reference symbol swap */
int *xp = &x; /* define symbol xp, reference x */
```

- Các symbol được lưu trong file object trong symbol table.
 - Symbol table là 1 mảng struct
 - Mỗi entry trong bảng sẽ chứa tên, kích thước và vị trí của symbol.

Linker Symbol?

```
symbol
            symbol
int sum(int *a, int n);
int array[2] = \{1, 2\};
int main()
{
    int val = sum(array, 2);
    return val;
                        main.c
       symbol
```

```
symbol
int sum(int *a, int n)
    int i, s = 0;
    for (i = 0); i < n; i++) {
         s += a[\i];
    return s;
}
                           sum.c
            Linker không biết i hay s
```

3 kiểu Object Files (Modules)

Relocatable object file (.o file)

- Chứa code và data ở dạng có thể kết hợp với các relocatable object file khác để tạo thành file thực thi.
 - Mỗi file .o file được tạo từ chính xác 1 file source (.c)

Executable object file (a.out file)

 Chứa code và data ở dạng có thể sao chép trực tiếp lên bộ nhớ và thực thi.

Shared object file (.so file)

- Dạng đặc biệt của relocatable object file, có thể được tải lên bộ nhớ và liên kết động, tại thời điểm load-time hoặc run-time.
- Được gọi là Dynamic Link Libraries (DLLs) trong Windows

Định dạng ELF: Executable and Linkable Format

- Định dạng nhị phân chuẩn cho các object files
- Một định dạng chung cho:
 - Relocatable object files (.o),
 - Executable object files (a.out)
 - Shared object files (.so)

ELF Object File Format

Elf header

 Word size, byte ordering, kiểu file (.o, exec, .so), machine type, etc.

Segment header table

- Page size, virtual addresses memory segments (sections), segment sizes.
- .text section: Code
- .rodata Section
 - Dữ liệu chỉ đọc như jump tables, ...
- .data Section
 - Các biến toàn cục <u>đã</u> được khởi tạo giá trị
- .bss section
 - Biến toàn cục chưa được khởi tạo giá trị
 - "Block Started by Symbol"/ "Better Save Space"
 - Có section header nhưng không chiếm không gian nào

ELF header				
Segment header table (required for executables)				
. text section				
. rodata section				
. data section				
.bss section				
.symtab section				
.rel.txt section				
.rel.data section				
. debug section				
Section header table				

U

ELF Object File Format (tt)

. symtab section

- Symbol table
- Hàm và các biến static
- Tên section và vi trí

.rel.text section

- Thông tin tái cấu trúc cho .text section
- Địa chỉ của một số instructions cần phải thay đổi trong file thực thi
- Hướng dẫn cho việc thay đối.

.rel.data section

- Thông tin tái cấu trúc cho .data section
- Địa chỉ của một số dữ liệu con trỏ cần thay đổi trong file thực thi tổng hợp.

debug section

Thông tin cho debug (gcc -g)

Section header table

Offsets và kích thước của mỗi section

ELF header				
Segment header table (required for executables)				
. text section				
. rodata section				
. data section				
.bss section				
.symtab section				
.rel.txt section				
.rel.data section				
.debug section				
Section header table				

Các kiểu Linker Symbol

Mỗi relocatable object mô-đun **m** (**m.o**) có 1 symbol table chứa thông tin các symbol được định nghĩa hoặc tham chiếu trong **m**

Global symbol

- Symbol được định nghĩa trong mô-đun m có thể được các mô-đun khác tham chiếu đến.
- Ví dụ: Các hàm (function) không static và các biến toàn cục không static.

External symbol

 Global symbol được tham chiếu bởi mô-đun m nhưng được định nghĩa trong mô-đun khác.

Local symbol

- Symbol được định nghĩa và chỉ được phép tham chiếu trong mô-đun m.
- Ví dụ: Các hàm C và biến toàn cục được định nghĩa với thuộc tính static.
- Local linker symbols <u>không phải</u> là các biến cục bộ

Ví dụ các kiểu Linker Symbol

```
Tham chiếu đến
 Tai main.o
                              global symbol...
             ...được đinh
             nghĩa ở đây
int sum(int/*a, int n);
                                       int sum(int *a, int n)
                                       {
int array[2] = {1, 2};
                                            int i, s = 0;
                                            for (i = 0); i < n; i++) {
int main()
{
                                                 s += a[i];
     int val = sum(array, 2);
     return val;
                                            return s;
}
                           main.c
                                                                     sum.c
đinh nghĩa
                     Tham chiếu đến
 global
                                                     Linker không biết i hay s
 symbol
                    external symbol...
        Linker không biết
                             ...được định nghĩa ở đây
            về val
```

Ví dụ các kiểu Linker Symbol

```
int sum(int *a, int n);
int array[2] = {1, 2};
int main()
{
    int val = sum(array, 2);
    return val;
}
```

Xem symbol table của file .o
\$ readelf -s main.o

```
Symbol table '.symtab' contains 11 entries:
                                                       Ndx Name
   Num:
           Value
                          Size Type
                                       Bind
                                              Vis
     0: 0000000000000000
                             0 NOTYPE LOCAL
                                              DEFAULT
                                                       UND
     1: 00000000000000000
                             0 FILE
                                       LOCAL
                                                       ABS main.c
                                              DEFAULT
                                              DEFAULT
     2: 00000000000000000
                             0 SECTION LOCAL
     3: 0000000000000000
                             O SECTION LOCAL DEFAULT
                                                         3
                                                         4
     4: 00000000000000000
                             O SECTION LOCAL DEFAULT
     5: 0000000000000000
                             O SECTION LOCAL DEFAULT
                                                         6
                                                         7
     6: 0000000000000000
                             0 SECTION LOCAL
                                              DEFAULT
                             0 SECTION LOCAL
                                                          5
     7: 00000000000000000
                                              DEFAULT
     8: 0000000000000000
                            24 FUNC
                                                         1 main
                                       GLOBAL DEFAULT
     9: 0000000000000000
                             8 OBJECT GLOBAL DEFAULT
                                                         3 array
    10: 0000000000000000
                             0 NOTYPE
                                       GLOBAL DEFAULT
                                                       UND sum
                                                                   16
```

Local symbol

- Các biến cục bộ C không static vs. các biến cục bộ C có static
 - Các biến cục bộ không static: lưu trong stack
 - Các biến cục bộ có static: lưu trong .bss hoặc .data

```
int f()
{
    static int x = 0;
    return x;
}

int g()
{
    static int x = 1;
    return x;
}
```

Compiler cấp phát không gian trong .data cho mỗi định nghĩa của x

Tạo các local symbol trong symbol table với các tên duy nhất, ví dụ, x.1 và x.2.

Linking symbol **Bài tập 1**

```
(b) swap.c
                               code/link/swap.c
     /* swap.c */
     extern int buf[];
 3
     int *bufp0 = &buf[0];
     int *bufp1;
     void swap()
         int temp;
10
         bufp1 = \&buf[1];
11
         temp = *bufp0;
12
         *bufp0 = *bufp1;
13
         *bufp1 = temp;
14
15
                              code/link/swap.c
```

Giả sử có 2 mô-đun tương ứng là **main.o** và **swap.o**, cho biết về sự có mặt và các thông tin của các biến/hàm sau trong symbol table của swap.o?

Symbol	<pre>swap.o .symtab entry?</pre>	Symbol type	Module where defined	Section
buf				
bufp0				
bufp1				
swap				
temp				

Linker sẽ làm gì?

- Tác vụ 1: Phân giải symbol (Symbol Resolve)
- Tác vụ 2: Tái định vị (Relocation)

Tác vụ 1: Phân giải symbol

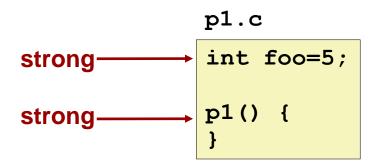
Phân giải symbol là làm gì?

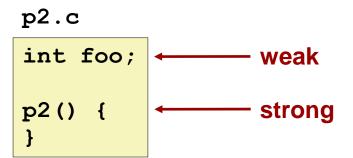
- Phân giải symbol: Ánh xạ mỗi tham chiếu symbol đến chính xác 1 định nghĩa của symbol
 - Các định nghĩa symbol nằm trong symbol table của các file .o
 - Nếu một symbol không được định nghĩa trong mô-đun hiện tại sẽ dẫn đến tạo một entry trong linker symbol table → dành cho linker
 - Linker tìm trong tất cả các mô-đun khác → nếu không thấy thì lỗi!!

```
Symbol table '.symtab' contains 11 entries:
                                                         Ndx Name
                          Size Type
                                               Vis
  Num:
           Value
                                        Bind
                              0 NOTYPE LOCAL
                                               DEFAULT
     0: 00000000000000000
                                                         UND
                                                         ABS main.c
     1: 00000000000000000
                              0 FILE
                                        LOCAL
                                               DEFAULT
                              0 SECTION LOCAL
     2: 00000000000000000
                                               DEFAULT
     3: 00000000000000000
                              0 SECTION LOCAL
                                               DEFAULT
                                                           3
     4: 00000000000000000
                              0 SECTION LOCAL
                                               DEFAULT
                                                           4
     5: 00000000000000000
                              0 SECTION LOCAL
     6: 00000000000000000
                              0 SECTION LOCAL
                                                           7
    7: 00000000000000000
     8: 00000000000000000
                              8 OBJECT GLOBAL DEFAULT
                                                           3 array
       00000000000000000
                                                           1 main
    10: 00000000000000000
```

Tác vụ 1: Phân giải symbol **Nếu có nhiều định nghĩa symbol trùng tên?**

- Giải quyết: Các symbol được đánh giá là strong hoặc weak
 - Strong: các hàm, các biến toàn cục đã có khởi tạo giá trị
 - Weak: các biến toàn cục chưa khởi tạo giá trị





Tác vụ 1: Phân giải symbol

Các luật khi phân giải symbol

- Luật 1: Không cho phép có nhiều strong symbol trùng tên
 - Mỗi strong symbol chỉ được định nghĩa 1 lần
 - Néu không: Linker error
- Luật 2: Nếu có 1 strong symbol và nhiều weak symbol, chọn tham chiếu đến strong symbol
 - Tham chiếu đến weak symbol sẽ được hiểu thành tham chiếu đến strong symbol
- Luật 3: Nếu có nhiều weak symbols, chọn tham chiếu tuỳ ý đến 1 symbol
 - Có thể thay đổi với gcc -fno-common

Tác vụ 1: Phân giải symbol

Ví dụ: Tham chiếu đến symbol nào?

```
int x;
p1() {}
```

```
p1() {}
```

Link time error: 2 strong symbols (p1)

```
int x;
p1() {}
```

Tham chiếu đến x sẽ cùng tham chiếu đến giá trị int chưa khởi tạo

```
int x;
int y;
p1() {}
```

Gán giá trị cho x trong p2 có thể ghi đè y! Evil!

```
int x=7;
int y=5;
p1() {}
```

Gán giá trị cho **x** trong **p**2 <u>sẽ</u> ghi đè **y**! Nasty!

```
int x=7;
p1() {}
```

Tham chiếu đến x sẽ luôn tham chiếu đến giá trị x đã được khởi tạo.

Phân giải symbol **Bài tập 2a**

■ Giả sử REF(x.i) → DEF(x.k) là tham chiếu đến x trong mô-đun i đến định nghĩa của x trong mô-đun k. Điền kết quả của các tham chiếu bên dưới, nếu có lỗi thì xác định:

```
ERROR: link time error (Rule 1)
UNKNOWN: lựa chọn tuỳ ý (Rule 3)
/* Module 1 */
                         /* Module 2 */
int main()
                         int main;
                         int p2()
                         REF(main.1) --> DEF(________)
                     (b) REF(main.2) --> DEF(_______)
```

Phân giải symbol **Bài tập 2b**

■ Giả sử REF(x.i) → DEF(x.k) là tham chiếu đến x trong mô-đun i đến định nghĩa của x trong mô-đun k. Điền kết quả của các tham chiếu bên dưới, nếu có lỗi thì xác định:

```
ERROR: link time error (Rule 1)
 UNKNOWN: lựa chọn tuỳ ý (Rule 3)
               /* Module 2 */
/* Module 1 */
                       int main=1;
void main()
                       int p2()
                        REF(main.1) --> DEF(____.__
                     (b) REF(main.2) --> DEF(_______)
```

Biến toàn cục

- Tránh sử dụng nếu có thể
- Nếu không, nên:
 - Sử dụng static nếu có thể
 - Khởi tạo giá trị khi định nghĩa biến toàn cục
 - Sử dụng extern nếu tham chiếu đến một biến toàn cục ở file khác.

Tác vụ 2: Tái định vị (Relocation)

Relocatable Object Files

System code . text
. data

main.o

main()
int array[2]={1,2}

sum.o

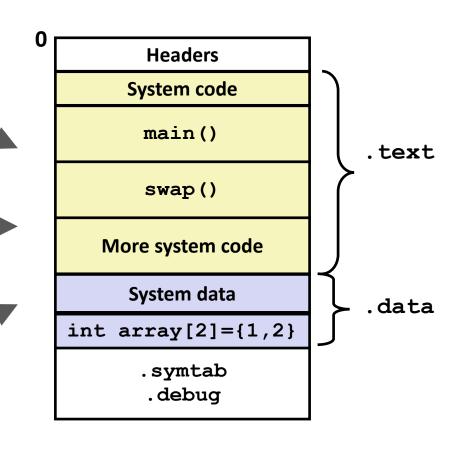
sum()

.text

.text

.data

Executable Object File



Tác vụ 2: Tái định vị (Relocation)

Các vị trí cần tái định vị

int $array[2] = \{1, 2\};$

```
int main()
     int val = sum(array, 2);
     return val;
}
                      main.c
                                    Thông tin về các vị trí trong section .rel.text
Relocation section '.rela.text' at offset 0x1e8 contains 2 entries:
 Offset
                 Info
                                             Sym. Value Sym. Name + Addend
                               Type
00000000000a   00090000000a R X86 64 32
                                          0000000000000000 array + 0
00000000000f
             000a00000002 R X86 64 PC32
                                          000000000000000 sum - 4
                                                    Source: objdump -r -d main.o
0000000000000000 <main>:
   0:
        48 83 ec 08
                                 sub
                                        $0x8,%rsp
   4: be 02 00 00 00
                                        $0x2,%esi
                                mov
   9: bf 00 00 00 00
                                        $0x0,%edi
                                                     # %edi = &array
                                mov
                         a: R X86 64 32 array
                                                       # Relocation entry
       e8 00 00 00 00
                                 callq 13 < main + 0x13 > \# sum()
   e:
                         f: R X86 64 PC32 sum-0x4
                                                       # Relocation entry
  13:
       48 83 c4 08
                                 add
                                        $0x8,%rsp
                                                                   main.o
  17:
        c3
                                 retq
```

Cần tái định vị

.text section sau khi tái định vị

```
00000000004004d0 <main>:
                48 83 ec 08
  4004d0:
                                          $0x8,%rsp
                                  sub
 4004d4:
                be 02 00 00 00
                                          $0x2,%esi
                                  mov
 4004d9:
                bf 18 10 60 00
                                          $0x601018, %edi # %edi = &array
                                  mov
 4004de:
                e8 05 00 00 00
                                         4004e8 <sum>
                                  callq
                                                          # sum()
 4004e3:
               48 83 c4 08
                                          $0x8,%rsp
                                  add
 4004e7:
                c3
                                  reta
00000000004004e8 <sum>:
  4004e8:
                b8 00 00 00 00
                                                $0x0,%eax
                                        mov
                ba 00 00 00 00
                                                $0x0, %edx
 4004ed:
                                        mov
                                                4004fd < sum + 0x15 >
 4004f2:
                eb 09
                                         qmr
 4004f4:
                48 63 ca
                                        movslq %edx,%rcx
 4004f7:
               03 04 8f
                                        add
                                               (%rdi,%rcx,4),%eax
               83 c2 01
 4004fa:
                                         add
                                                $0x1,%edx
 4004fd:
                39 f2
                                                %esi,%edx
                                         cmp
 4004ff:
                7c f3
                                                4004f4 < sum + 0xc >
                                         jl
  400501:
                f3 c3
                                         repz retq
```

Xác định địa chỉ hàm sum() (PC-relative): 0x4004e8 = 0x4004e3 + 0x5

Source: objdump -dx prog

Executable Object Files trên bộ nhớ

Executable Object File Kernel virtual memory ELF header User stack (created at runtime) Program header table (required for executables) .init section .text section Memory-mapped region for shared libraries .rodata section .data section .bss section Run-time heap .symtab (created by malloc) .debug Read/write data segment .line (.data, .bss) .strtab Read-only code segment (.init,.text,.rodata) Section header table 0×400000 (required for relocatables) Unused

Memory invisible to user code %rsp (stack pointer) brk Loaded from the executable file

Thêm: Đóng gói các hàm thông dụng

- Làm thế nào đóng gói các hàm hay được lập trình viên sử dụng?
 - Toán học, I/O, quản lý bộ nhớ, string, v.v..

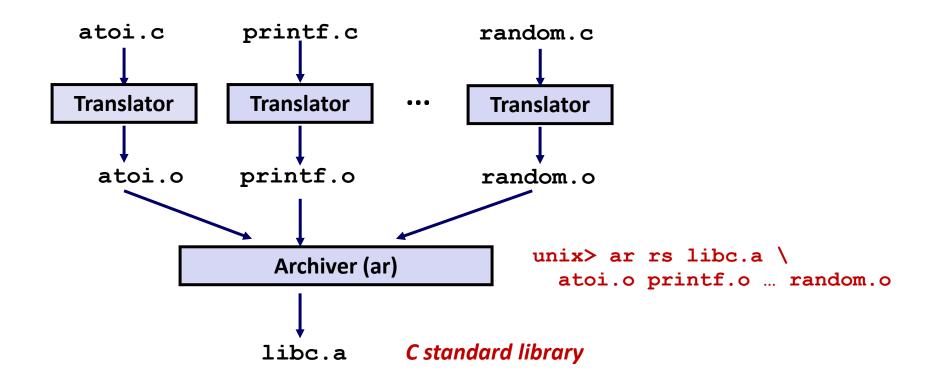
Với linker:

- Lựa chọn 1: Đặt tất cả các hàm vào 1 file source
 - Lập trình viên link object file lớn vào chương trình của họ
 - Không hiệu quả về không gian và thời gian
- Lựa chọn 2: Mỗi hàm được đặt trong 1 file source riêng
 - Lập trình viên chỉ link các file cần thiết vào chương trình của họ
 - Hiệu quả hơn, nhưng là gánh nặng cho lập trình viên

Giải pháp cũ: Static Libraries

- Thư viện tĩnh Static libraries (các file .a)
 - Ghép các related relocatable object files (.o) thành 1 file duy nhất với chỉ số index (gọi là archive).
 - Là Linker nâng cao, cố gắng phân giải các tham chiếu bằng cách tìm các symbol trong 1 hoặc nhiều archive.
 - Nếu 1 archive file có thể phân giải tham chiếu, link file đó với file thực thi.

Cách tạo Static Libraries



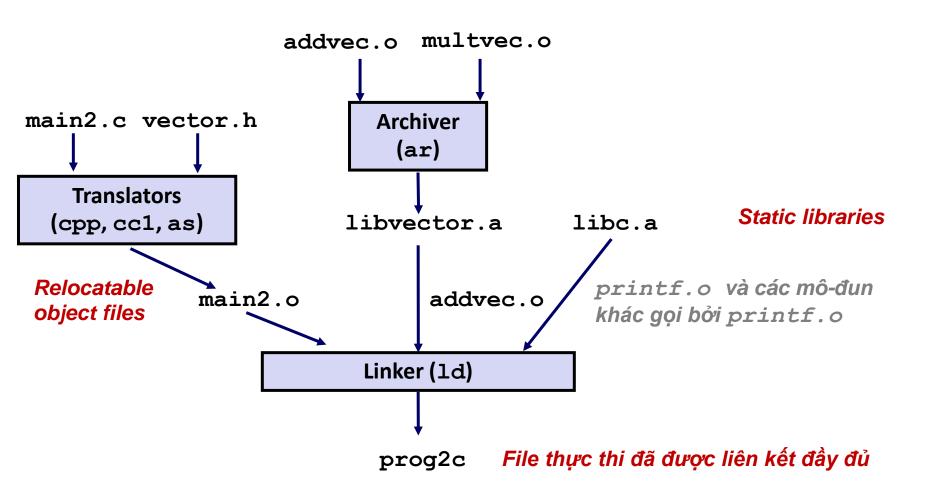
- Archiver cho phép cập nhật
- Biên dịch lại hàm cần thay đổi và thay thế file .o tương ứng trong archive.

libvector.a

Linking với Static Libraries

```
#include <stdio.h>
#include "vector.h"
int x[2] = \{1, 2\};
int y[2] = \{3, 4\};
int z[2];
int main()
    addvec(x, y, z, 2);
    printf("z = [%d %d] \n'',
           z[0], z[1]);
    return 0;
}
                    main2.c
```

Linking với Static Libraries



"c" for "compile-time"

Sử dụng Static Libraries

Giải thuật của linker để phân giải các tham chiếu ngoài:

- Quét các file .o và .a theo thứ tự trong command.
- Trong quá trình quét, lưu lại danh sách các tham chiếu chưa phân giải.
- Với mỗi file.o hay .a mới, ví dụ obj, cố gắng phân giải các tham chiếu chưa phân giải trong danh sách với các symbol định nghĩa trong obj.
- Nếu quét xong mà vẫn còn tham chiếu trong danh sách chưa phân giải được thì báo lỗi.

■ Vấn đề:

- Thứ tự khai báo trong command rất quan trọng!
- Đặt các libraries ở phía cuối command.

```
unix> gcc -L. libtest.o -lmine
unix> gcc -L. -lmine libtest.o
libtest.o: In function `main':
libtest.o(.text+0x4): undefined reference to `libfun'
```

Giải pháp mới: Shared Libraries

Static libraries có những điểm yếu:

- Trùng lặp trong các file thực thi (ví dụ: mọi function đều cần libc)
- Nếu có chỉnh sửa nhỏ trong libraries thì mỗi chương trình cần thực hiện link lại.

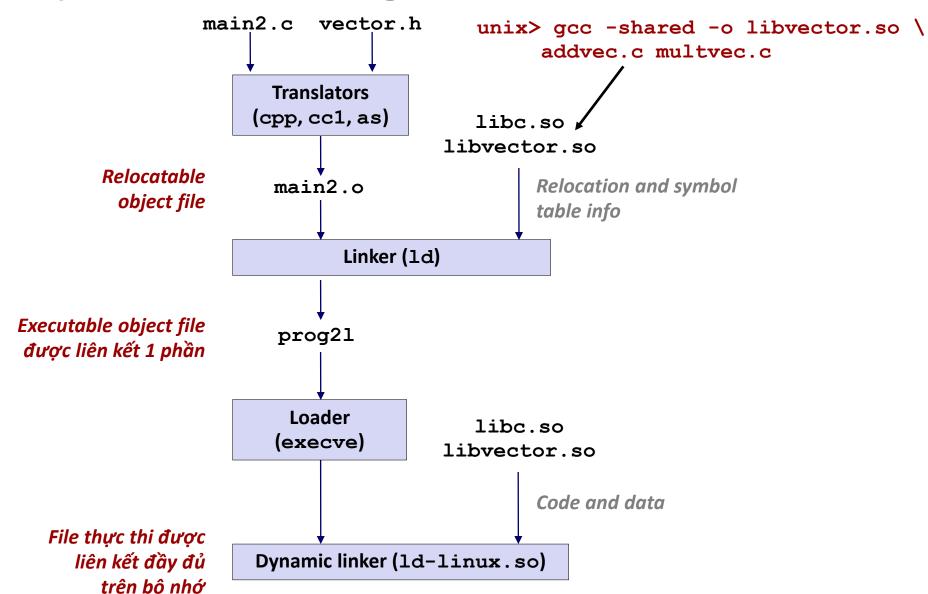
Giải pháp mới: Shared Libraries

- Các object files chứa code and data được load và link vào một ứng dụng một cách linh động, tại thời điểm load-time hoặc run-time
- Còn được gọi là dynamic link libraries, DLLs, .so files

Shared Libraries (tt)

- Dynamic linking có thể xảy ra khi file thực thi được loaded và thực thi (load-time linking).
 - Thường thấy ở Linux, tự động thực hiện bởi dynamic linker (1dlinux.so).
 - Standard C library (libc.so) thường được dynamically linking.
- Dynamic linking cũng có thể xảy ra khi file đã chạy (run-time linking).
 - Trong Linux, có thể thực hiện với hàm dlopen ().
- Shared library có thể được chia sẻ giữa nhiều tiến trình.

Dynamic Linking khi Load-time



Dynamic Linking khi Run-time

```
#include <stdio.h>
#include <stdlib.h>
#include <dlfcn.h>
int x[2] = \{1, 2\};
int y[2] = \{3, 4\};
int z[2];
int main()
  void *handle;
  void (*addvec)(int *, int *, int *, int);
  char *error;
  /* Dynamically load the shared library that contains addvec() */
  handle = dlopen("./libvector.so", RTLD_LAZY);
  if (!handle) {
    fprintf(stderr, "%s\n", dlerror());
    exit(1);
                                                                                                dll.c
```

Dynamic Linking khi Run-time

```
/* Get a pointer to the addvec() function we just loaded */
addvec = dlsym(handle, "addvec");
if ((error = dlerror()) != NULL) {
  fprintf(stderr, "%s\n", error);
  exit(1);
/* Now we can call addvec() just like any other function */
addvec(x, y, z, 2);
printf("z = [%d %d]\n", z[0], z[1]);
/* Unload the shared library */
if (dlclose(handle) < 0) {</pre>
  fprintf(stderr, "%s\n", dlerror());
  exit(1);
return 0;
                                                                                       d11.c
```

Nội dung

■ Các chủ đề chính:

- 1) Biểu diễn các kiểu dữ liệu và các phép tính toán bit
- 2) Ngôn ngữ assembly
- 3) Điều khiển luồng trong C với assembly
- 4) Các thủ tục/hàm (procedure) trong C ở mức assembly
- 5) Biểu diễn mảng, cấu trúc dữ liệu trong C
- 6) Một số topic ATTT: reverse engineering, bufferoverflow
- 7) Linking trong biên dịch file thực thi
- 8) Phân cấp bộ nhớ, cache

Lab liên quan

- Lab 1: Nội dung <u>1</u>
- Lab 2: Nội dung 1, 2, 3
- Lab 3: Nội dung 1, 2, 3, 4, 5, 6

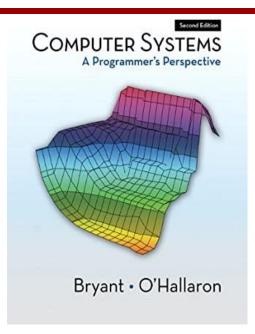
- Lab 4: Nội dung 1, 2, 3, 4, 5, 6
- Lab 5: Nội dung 1, 2, 3, 4, 5, 6
- Lab 6: Nội dung 1, 2, 3, 4, 5, 6

Giáo trình

Giáo trình chính

Computer Systems: A Programmer's Perspective

- Second Edition (CS:APP2e), Pearson, 2010
- Randal E. Bryant, David R. O'Hallaron
- http://csapp.cs.cmu.edu



Tài liệu khác

- The C Programming Language, Second Edition, Prentice Hall, 1988
 - Brian Kernighan and Dennis Ritchie
- The IDA Pro Book: The Unofficial Guide to the World's Most Popular Disassembler, 1st Edition, 2008
 - Chris Eagle
- Reversing: Secrets of Reverse Engineering, 1st Edition, 2011
 - Eldad Eilam

