CS415 Compilers

Procedure Abstractions

These slides are based on slides copyrighted by Keith Cooper, Ken Kennedy & Linda Torczon at Rice University
The latter half of a compiler contains more open problems, more challenges, and more gray areas than the front half

- This is “compilation,” as opposed to “parsing” or “translation”
- Implementing promised behavior
  - What defines the meaning of the program
- Managing target machine resources
  - Registers, memory, issue slots, locality, power, ...
  - These issues determine the quality of the compiler
• **Control Abstraction**
  -> Well defined entries & exits
  -> Mechanism to return control to caller
  -> Some notion of parameterization (usually)

• **Clean Name Space**
  -> Clean slate for writing locally visible names
  -> Local names may obscure identical, non-local names
  -> Local names cannot be seen outside

• **External Interface**
  -> Access is by procedure name & parameters
  -> Clear protection for both caller & callee

• **Procedures** permit a critical separation of concerns
Procedures allow us to use separate compilation
- Separate compilation allows us to build non-trivial programs
- Keeps compile times reasonable
- Lets multiple programmers collaborate
- Requires independent procedures

Without separate compilation, we would not build large systems

The procedure linkage convention
- Ensures that each procedure inherits a valid run-time environment and that the callers environment is restored on return
  - The compiler must generate code to ensure this happens according to conventions established by the system
A procedure is an abstract structure constructed via software

Underlying hardware directly supports little of the abstraction—it understands bits, bytes, integers, reals, and addresses, but not:

• Entries and exits
• Interfaces
• Call and return mechanisms
  → may be a special instruction to save context at point of call
• Name space
• Nested scopes

All these are established by a carefully-crafted system of mechanisms provided by compiler, run-time system, linkage editor and loader, and OS
These concepts are often confusing to the newcomer

- Procedure linkages execute at run time
- Code for the procedure linkage is emitted at compile time
- The procedure linkage is designed long before either of these

“This issue (compile time versus run time) confuses students more than any other issue” —Keith Cooper (Rice University)
The Procedure as a Control Abstraction

Procedures have well-defined control-flow

The Algol-60 procedure call
- Invoked at a call site, with some set of actual parameters
- Control returns to call site, immediately after invocation
Procedures have well-defined control-flow

The Algol-60 procedure call
• Invoked at a call site, with some set of *actual parameters*
• Control returns to call site, immediately after invocation

```plaintext
int p(a,b,c)
int a, b, c;
{
    int d;
    d = q(c,b);
    ...
}

s = p(10,t,u);
...
```
Procedures have well-defined control-flow

The Algol-60 procedure call

- Invoked at a call site, with some set of *actual parameters*
- Control returns to call site, immediately after invocation

```c
int p(a, b, c)
    int a, b, c;
    {
        int d;
        d = q(c, b);
        ...
    }

int q(x, y)
    int x, y;
    {
        return x + y;
    }
```

```
s = p(10, t, u);
...```

...
Procedures have well-defined control-flow

The Algol-60 procedure call

- Invoked at a call site, with some set of *actual parameters*
- Control returns to call site, immediately after invocation
Procedures have well-defined control-flow

The Algol-60 procedure call

- Invoked at a call site, with some set of *actual parameters*
- Control returns to call site, immediately after invocation
Procedures have well-defined control-flow

The Algol-60 procedure call
- Invoked at a call site, with some set of actual parameters
- Control returns to call site, immediately after invocation

```plaintext
int p(a,b,c)
int a, b, c;
{
    int d;
    d = q(c,b);
    ...
}

int q(x,y)
int x,y;
{
    return x + y;
    ...
}
```

- Most languages allow recursion
Implementing procedures with this behavior

- Requires code to save and restore a “return address”
- Must map actual parameters to formal parameters $q:\langle c \mapsto x, b \mapsto y \rangle$
- Must create storage for local variables (&, maybe, parameters)
  - $p$ needs space for $d$ (&, maybe, $a$, $b$, & $c$)
  - Where does this space go in recursive invocations?

```c
int p(a,b,c)
int a, b, c;
{
    int d;
    d = q(c,b);
    ...
}

int q(x,y)
int x,y;
{
    return x + y;
}
```

\[ \ldots s = p(10,t,u); \ldots \]
Implementing procedures with this behavior

- Must preserve $p$'s state while $q$ executes
  - Think about recursion
- Strategy: Create unique location for each procedure activation
  - Can use a “stack” of memory blocks to hold local storage and return addresses

```c
int p(a,b,c)
    int a, b, c;
    {
        int   d;
        d = q(c,b);
        ...
    }
    int q(x,y)
        int x,y;
        {
            return x + y;
        }
```

Compiler emits code that causes all this to happen at run time
Example: Dynamic vs. Static Views

```c
int r (...) { // declaration
    int d, s;

    int q (x,y) // declaration
        int x,y;
        {
            return x + y + d ;
        }

    int p (a,b,c) // declaration
        int a, b, c;
        {
            int d;
            if (...) 
                d = q (c,b); // call
            else
                d = p (a, d, c); // call
        
    s = p(10, d, s); // call
    s = p(11, s, d); // call
}
```
```c
int r (...) { // declaration
    int d, s;

    int q (x,y) // declaration
        int x,y;
        {
            return x + y + d;
        }

    int p (a,b,c) // declaration
        int a, b, c;
        {
            int d;
            if (...)  
                d = q (c,b); // call
            else
                d = p (a, d, c); // call
        }

    s = p(10, d, s); // call
    s = p(11, s, d); // call
}
```

**Example: Dynamic vs. Static Views**

(1) dynamic activation tree

(2) dynamic activation records in runtime stack

(3) static symbol table

<table>
<thead>
<tr>
<th></th>
<th>nesting level</th>
<th>offset</th>
</tr>
</thead>
<tbody>
<tr>
<td>d</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>0.1</td>
<td></td>
</tr>
<tr>
<td>x</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>y</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>d</td>
<td>1.3</td>
<td></td>
</tr>
</tbody>
</table>

lexical scoping
Each procedure creates its own name space

- Any name (almost) can be declared locally
- Local names obscure identical non-local names
- Local names cannot be seen outside the procedure
  → Nested procedures are “inside” by definition
- We call this set of rules & conventions “lexical scoping”

Examples

- C has global, static, local, and block scopes  (Fortran-like)
  → Blocks can be nested, procedures cannot
- Scheme has global, procedure-wide, and nested scopes  (let)
  → Procedure scope (typically) contains formal parameters
Why introduce lexical scoping?

• Provides a compile-time mechanism for binding “free” variables
• Simplifies rules for naming & resolves conflicts

How can the compiler keep track of all those names?

The Problem

• At point \( p \), which declaration of \( x \) is current?
• At run-time, where is \( x \) found?
• As parser goes in & out of scopes, how does it delete \( x \)?

The Answer

• Lexically scoped symbol tables

(see § 5.7.3)
OS needs a way to start the program’s execution

• Programmer needs a way to indicate where it begins
  → The “main” procedure in most languages
• When user invokes “grep” at a command line
  → OS finds the executable
  → OS creates a process and arranges for it to run “grep”
  → “grep” is code from the compiler, linked with run-time system
    ▪ Starts the run-time environment & calls “main”
    ▪ After main, it shuts down run-time environment & returns
• When “grep” needs system services
  → It makes a system call, such as fopen()
Where Do All These Variables Go?

Automatic & Local
- Keep them in the procedure activation record or in a register
- Automatic $\Rightarrow$ lifetime matches procedure’s lifetime

Static
- Procedure scope $\Rightarrow$ storage area affixed with procedure name
- File scope $\Rightarrow$ storage area affixed with file name
- Lifetime is entire execution

Global
- One or more named global data areas
- One per program, ...
- Lifetime is entire execution
Placing Run-time Data Structures

Classic Organization

<table>
<thead>
<tr>
<th>Code</th>
<th>Global &amp; Static</th>
<th>Heap</th>
<th>Stack</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Single Logical Address Space

- Code, static, & global data have known size
  - Use symbolic labels in the code
- Heap & stack both grow & shrink over time
- This is a virtual address space

- Better utilization if stack & heap grow toward each other
- Code & data separate or interleaved
How Does This Really Work?

The Big Picture

Compiler’s view

Virtual address spaces

OS’s view

Hardware’s view

Physical address space
Lexical Scoping: Translating Local Names

How does the compiler represent a specific instance of $x$?

- Name is translated into a static coordinate
  - $\langle \text{level}, \text{offset} \rangle$ pair
  - “level” is lexical nesting level of the procedure
  - “offset” is unique within that scope

- Subsequent code will use the static coordinate to generate addresses and references

- “level” is a function of the table in which $x$ is found
  - Stored in the entry for each $x$

- “offset” must be assigned and stored in the symbol table
  - Assigned at compile time
  - Known at compile time
  - Used to generate code that executes at run-time
## Activation Record Basics

### Activation Record Parameters

- **parameters**
  - Space for parameters to the current routine

- **register save area**
  - Saved register contents

- **return value**
  - If function, space for return value

- **return address**
  - Address to resume caller

- **addressability**
  - Help with non-local access

- **caller’s ARP**
  - To restore caller’s AR on a return (control link)

- **local variables**
  - Space for local values & variables (including spills)

---

*One AR for each invocation of a procedure*
Procedure abstraction
Read EaC: Chapter 6.1 - 6.5