Programming Assignment V
Monday, Jan 30, 2017

1 Introduction

In this assignment, you will implement a code generator for Cool. When successfully completed, you will have a fully functional Cool compiler!

The code generator makes use of the AST constructed in TP3 and static analysis performed in TP4. Your code generator should produce MIPS assembly code that faithfully implements *any* correct Cool program. There is no error recovery in code generation—all erroneous Cool programs have been detected by the front-end phases of the compiler.

As with the static analysis assignment, this assignment has considerable room for design decisions. Your program is correct if the code it generates works correctly; how you achieve that goal is up to you. We will suggest certain conventions that we believe will make your life easier, but you do not have to take our advice. As always, explain and justify your design decisions in the README file. This assignment is about twice the amount of the code of the previous programming assignment, though they share much of the same infrastructure. **Start early!**

Critical to getting a correct code generator is a thorough understanding of both the expected behavior of Cool constructs and the interface between the runtime system and the generated code. The expected behavior of Cool programs is defined by the operational semantics for Cool given in Section 13 of the *Cool Reference Manual*. Recall that this is only a specification of the meaning of the language constructs—not how to implement them. The interface between the runtime system and the generated code is given in *The Cool Runtime System*. See that document for a detailed discussion of the requirements of the runtime system on the generated code. There is a lot of information in this handout and the aforementioned documents, and you need to know most of it to write a correct code generator. *Please read thoroughly.*

2 Files and Directories

To get started, create a directory where you want to do the assignment and execute one of the following commands *in that directory.*

```
make -f /home/prof/renato/cool/student/assignments/PA5/Makefile
```

As usual, there are other files used in the assignment that are symbolically linked to your directory or are included from /home/prof/renato/cool/student/include and /home/prof/renato/cool/student/src. You should not modify these files. Almost all of these files have been described in previous assignments.

We now describe the most important files for each version of the project.

2.1 C++ Version

This is a list of the files that you may want to modify. You should already be familiar with most of the other files from previous assignments. See the README file for details about the additional files.

- cgen.cc
  This file will contain almost all your code for the code generator. The entry point for your code

generator is the **program_class::cgen(ostream&)** method, which is called on the root of your AST. Along with the usual constants, we have provided functions for emitting MIPS instructions, a skeleton for coding strings, integers, and booleans, and a skeleton of a class table (**CgenClassTable**). You can use the provided code or replace it with your own inheritance graph from TP4.

- cgen.h
  This file is the header for the code generator. You may add anything you like to this file. It provides classes for implementing the inheritance graph. You may replace or modify them as you wish.

- emit.h
  This file contains various code generation macros used in emitting MIPS instructions among other things. You may modify this file.

- cool-tree.h
  As usual, these files contain the declarations of classes for AST nodes. You can add field or method declarations to the classes in cool-tree.h. The implementation of methods should be added to cgen.cc.

- cgen_supp.cc
  This file contains general support code for the code generator. You will find a number of handy functions here. Add to the file as you see fit, but don't change anything that's already there.

- example.cl
  This file should contain a test program of your own design. Test as many features of the code generator as you can.

- README
  This file will contain the write-up for your assignment. It is critical that you explain design decisions, how your code is structured, and why you believe your design is a good one (i.e., why it leads to a correct and robust program). It is part of the assignment to explain things in text as well as to comment your code.

3 Design

Before continuing, we suggest you read *The Cool Runtime System* to familiarize yourself with the requirements on your code generator imposed by the runtime system.

In considering your design, at a high-level, your code generator will need to perform the following tasks:

1. Determine and emit code for global constants, such as prototype objects.

2. Determine and emit code for global tables, such as the **class_nameTab**, the **class_objTab**, and the dispatch tables.

3. Determine and emit code for initialization method for each class.

4. Determine and emit code for each method definition.

There are many possible ways to write the code generator. One reasonable strategy is to perform code generation in two passes. The first pass decides the object layout for each class, particularly the

offset at which each attribute is stored in an object. Using this information, the second pass recursively walks each feature and generates stack machine code for each expression.

There are a number of things you must keep in mind while designing your code generator:

- Your code generator must work correctly with the Cool runtime system, which is explained in the *Cool Runtime System* manual.
- You should have a clear picture of the runtime semantics of Cool programs. The semantics are described informally in the first part of the *Cool Reference Manual*, and a precise description of how Cool programs should behave is given in Section 13 of the manual.
- You should understand the MIPS instruction set. An overview of MIPS operations is given in the spim documentation, which is on the class web page.
- You should decide what invariants your generated code will observe and expect (i.e., what registers will be saved, which might be overwritten, etc). You may also find it useful to refer to information on code generation in the lecture notes.

You do *not* need to generate the same code as coolc. Coolc includes a very simple register allocator and other small changes that are not required for this assignment. The only requirement is to generate code that runs correctly with the runtime system.

3.1 Runtime Error Checking

The end of the Cool manual lists six errors that will terminate the program. Of these, your generated code should catch the first three—dispatch on void, case on void, and missing branch—and print a suitable error message before aborting. You may allow SPIM to catch division by zero. Catching the last two errors—substring out of range and heap overflow—is the responsibility of the runtime system in `trap.handler`. See Figure 4 of the *Cool Runtime System* manual for a listing of functions that display error messages for you.

3.2 Garbage Collection

To receive full credit for this assignment, your code generator must work correctly with the generational garbage collector in the Cool runtime system. The skeletons contain functions **code_select_gc** that generate code that sets GC options from command line flags. The command-line flags that affect garbage collection are -g, -t, and -T. Garbage collection is disabled by default; the flag -g enables it. When enabled, the garbage collector not only reclaims memory, but also verifies that "-1" separates all objects in the heap, thus checking that the program (or the collector!) has not accidentally overwritten the end of an object. The -t and -T flags are used for additional testing. With -t the collector performs collections very frequently (on every allocation). The garbage collector does not directly use -T; in coolc the -T option causes extra code to be generated that performs more runtime validity checks. You are free to use (or not use) -T for whatever you wish.

For your implementation, the simplest way to start is not to use the collector at all (this is the default). When you decide to use the collector, be sure to carefully review the garbage collection interface described in the *Cool Runtime System* manual. Ensuring that your code generator correctly works with the garbage collector in *all* circumstances is not trivial.

4 Testing and Debugging

You will need a working scanner, parser, and semantic analyzer to test your code generator. You may use either your own components or the components from coolc. By default, the coolc components are used. To change that, replace the lexer, parser, and/or semant executable (which are symbolic links in your project directory) with your own scanner/parser/semantic analyzer. Even if you use your own components, it is wise to test your code generator with the coolc scanner, parser, and semantic analyzer at least once because we will grade your project using coolc's version of the other phases.

You will run your code generator using mycoolc, a shell script that "glues" together the generator with the rest of compiler phases. Note that mycoolc takes a -c flag for debugging the code generator; using this flag merely causes `cgen_debug` (a global variable) to be set. Adding the actual code to produce useful debugging information is up to you. See the project README for details.

4.1 Coolaid

Coolaid is a tool to statically verify some basic correctness properties of the MIPS assembly code produced from Cool source. Coolaid will check that the assembly code is "well-typed" with respect to the Cool typing rules, just like the Java bytecode verifier checks that the bytecode output by a Java compiler is type safe. Aside from checking the safety of the output code, this tool can greatly benefit the development process of the code generation phase. Since the compiler front-end ensures that the source program is well-typed and that the type system guarantees certain safety properties, we expect that those properties should also hold on the resulting assembly code. If Coolaid is not able to verify some particular safety property, then a likely cause is a bug in the compiler itself.

Coolaid has been tested on 3500 programs generated by student compilers from the Spring 2003 offering of CS164 at UC Berkeley. The standard testing procedure (which runs the generated code on the MIPS simulator spim) used for grading found errors in 1000 of those. Often the error messages were in the form of garbled output or output that did not match the expected output, whereas Coolaid was able to give precise error messages where the unsafe operation was performed that most likely caused the garbled output. Coolaid also finds errors in 800 of the 2500 programs that pass the testing procedure! These were errors that the testing procedure missed because the sample input for the code did not exercise the bad code portions. See the *Coolaid Reference Manual* to get started using Coolaid. Note Coolaid imposes some additional requirements on the generated code to be able to check it; these requirements are also discussed in the *Cool Runtime System* manual.

4.2 Spim and XSpim

The executables spim and xspim are simulators for MIPS architecture on which you can run your generated code. The program xspim works like spim in that it lets you run MIPS assembly programs. However, it has many features that allow you to examine the virtual machine's state, including the memory locations, registers, data segment, and code segment of the program. You can also set breakpoints and single step your program. The documentation for spim/xspim is on the course web page.

Warning. One thing that makes debugging with spim difficult is that spim is an interpreter for assembly code and not a true assembler. If your code or data definitions refer to undefined labels, the error shows up only if the executing code actually refers to such a label. Moreover, an error is reported only for undefined labels that appear in the code section of your program. If you have constant data definitions

that refer to undefined labels, spim won't tell you anything. It will just assume the value 0 for such undefined labels.

5 Final Submission

Make sure to complete the following items before submitting to avoid any penalties.

- ☐ Include your write-up in README.
- ☐ Include your test cases that test your code generator in example.cl.
- ☐ Make sure all your code for the code generator is in
  - cool-tree.h, cgen.h, cgen.cc, cgen_supp.cc, and emit.h