Computer Science and Engineering Department
Indian Institute of Technology Kharagpur

Compilers Laboratory: CS39003
3rd Year CSE: 5th Semester

Assignment – 6: Target Code Generator for tinyC *Marks: 100*
Assign Date: *October 17, 2023* Submit Date: *23:55, November 01, 2023*

1 Preamble – tinyC

The Lexical Grammar (Assignment 3) and the Phase Structure Grammar (Assignment 4) for tinyC have already been defined as subsets of the C language specification from the International Standard **ISO/IEC 9899:1999 (E)**. Finally, three address code (TAC) structure and a further subset of tinyC has been specified (Assignment 5) for translating the input tinyC program to TAC quad array, a supporting symbol table, and other auxiliary data structures.

In this assignment you will write a target code translator from the TAC quad array (with the supporting symbol table, and other auxiliary data structures) to the assembly language of x86-64. The translation is now machine-specific and your generated assembly code would be translated with the gcc assembler to produce the final executable codes for the tinyC program.

2 Scope of Target Translation

- For simplicity restrict tinyC further:
  1. Skip shift and bit operators.
  2. Support only **void**, **int**, **float**, and **char** types. Skip **double** type.
  3. Support only one–dimensional arrays.
  4. Support only **void**, **int**, **float**, **char**, **void***, **int***, **float***, and **char*** types for returns types of functions.
  5. No type conversion to be supported.
- For I/O, provide a library (as created in Assignment 2) using in-line assembly language program of x86-64 along with `syscall` for gcc assembler.:
  - `int printStr(char *)` – prints a string of characters. The parameter is terminated by '`\0`'. The return value is the number of characters printed.
  - `int printInt(int n)` – prints the integer value of `n` (no newline). It returns the number of characters printed.
  - `int readInt(int *eP)` – reads an integer (signed) and returns it. The parameter is for error (`ERR = 1, OK = 0`). You may use `ERR = 0, OK = 1`, if you have used this convention in your assignment 2. Your convention will be considered from your `myl.h` file.
  - `int printFlt(float n)` – prints the floating value of `n` (no newline). It returns the number of characters printed.
  - `int readFlt(float *eP)` – reads a floating (signed) and returns it. The parameter is for error (`ERR = 1, OK = 0`). You may use `ERR = 0, OK = 1`, if you have used this convention in your assignment 2. Your convention will be considered from your `myl.h` file.

The header file `myl.h` of the library will be as follows:

```
#ifndef _MYL_H
#define _MYL_H
#define ERR 1
#define OK 0
int printStr(char *);
int printInt(int);
int readInt(int *eP);  // Usual meaning as specified in Assignment #2
int printFlt(float);
int readFlt(float *eP);  // Usual meaning as specified in Assignment #2
#endif
```

3 Design of the Translator

The steps for target code generation were outlined in Target Code Generation lecture presentations. In this assignment, however, you do not need to deal with any machine-independent or machine-specific optimization. Hence the translation comprises the following major steps only:

Memory Binding This deals with the design of the allocation schema of variables (including parameters and constants) that associates each variable to the respective address expression or register. This needs to handle the following:

- *Handle local variables, parameters, and return value for a function.* These are automatic and reside in the Activation Record (AR) of the function. Various design schema for AR are possible based on the calling sequence protocol. A sample AR design could be as follows:

Offset	Stack Item	Responsibility
−ve	Saved Registers	Callee Saves & Restores
−ve	Callee Local Data	Callee defines and uses
0	Base Pointer of Caller	Callee Saves & Restores
	Return Address	Saved by call, used by ret
+ve	Return Value	Callee writes, Caller reads
+ve	Parameters	Caller writes, Callee reads

Activation Record Structure with Management Protocol

  - Offset's in the AR are with respect to the Base Pointer of Callee.
  - Return Value can alternatively be returned through a register (like accumulator).
  - The AR will be populated from the Symbol Table of the function.
  - Symbol Tables of nested blocks will be flattened and its variables allocated within the Symbol Table (and hence the AR) of the function where there occur in. Necessary name mangling will be performed to to take care of same lexical name for different variables in different nested scopes.

- *Handle global variables* (note that local static variables are not allowed in tinyC) as static and generate allocations in static area. This will be populated from global symbol table (ST.gbl).
- *Generate Constants from Table of Constants* – handle string constants as assembler symbols in DATA SEGMENT and integer constants as parts of target code (TEXT SEGMENT)
- *Register Allocations & Assignment*: Create memory binding for variables in registers:
  - After a load / store the variable on the activation record and the register have identical values
  - Registers can be used to store temporary computed values
  - Register allocations are often used to pass `int` or pointer parameters
  - Register allocations are often used to return `int` or pointer values

Code Translation This deals with the translation of 3–Address quad's to x86-64 assembly code. This needs to handle:

- *Generation of Function Prologue* – few lines of code at the beginning of a function, which prepare the stack and registers for use within the function.
- *Generate Function Epilogue* – appears at the end of the function, and restores the stack and registers to the state they were in before the function was called.
- *Map 3–Address Code to Assembly* – to translate the function body do:
  - Choose optimized assembly instructions for every expression, assignment and control quad.
  - Use algebraic simplification & reduction of strength for choice of assembly instructions from a quad.
  - Use Machine Idioms (like **inc** for i++ or ++i in place of **add reg, 1**).

Note: *Refer to* Target Code Generation *lecture presentations for details.*

Target Code Integrate all the above code into an Assembly File for gcc assembler.

4 The Assignment

1. Write a target code (x86-64) translator from the 3-Address quad's generated from the flex and bison specifications of tinyC (with restrictions as mentioned in Section 2). Assume that the input tinyC file is lexically, syntactically, and semantically correct. Hence no error handling and / or recovery is expected.

2. Prepare a Makefile to compile and test the project.

3. Prepare test input files ass6_*roll*_test<number>.c to test the target code translation and generate the translation output in ass6_*roll* <number>.asm.

4. Name your files as follows:

File	Naming
Flex Specification	ass6_*roll*.l
Bison Specification	ass6_*roll*.y
Data Structures (Class Definitions) and Global Function Prototypes	ass6_*roll*_translator.h
Data Structures, Function Implementations and 3–Address Translator	ass6_*roll*_translator.cxx
Target Translator and x86-64 Translator main()	ass6_*roll*_target_translator.cxx
Test Inputs	ass6_*roll*_test<number>.c
3–Address Test Outputs	ass6_*roll*_quads<number>.out
Test Outputs	ass6_*roll* <number>.asm

5. Prepare a tar-archive with the name ass6_*roll*.tar containing all the files and upload to Moodle.

5 Credits

Design of Memory Binding: **15 + 5 + 5 + 5 + 10 = 40**
Handling of Activation Records
Handling of Nested Symbol Tables
Handling of Static Memory & Binding
Handling of Constants
Handling of Register Allocation & Assignment

Design of Code Translation: **5 + 5 + 10 = 20**
Handling of Prologue
Handling of Epilogue
Handling of Function Body

Design of Target Code Management: **10**
Integration of translated codes into an assembly file

Design of Test files and correctness of outputs: **10 + 10 = 20**
Test at least 5 i/p files covering all rules
Shortcoming and / or bugs, if any, should be highlighted

Integrated interface of the tinyC Compiler: **10**