

Project 1--CDA 3101 (Spring 2014)

Worth: 100 points (10% of course grade)

Assigned: Friday, Jan 24, 2014

Due: 1:25 pm, Monday, Feb 24, 2014

1. Purpose

This project is intended to help you understand the instructions of a very simple assembly language and how to assemble programs into machine language.

2. Problem

This project has three parts. In the first part, you will write a program to take an assembly-language program and produce the corresponding machine language. In the second part, you will write a behavioral simulator for the resulting machine code. In the third part, you will write a short assembly-language program to multiply two numbers.

3. LC3101 Instruction-Set Architecture

For this project, you will be developing a simulator and assembler for the LC3101 (Little Computer, used in CDA 3101). The LC3101 is very simple, but it is general enough to solve complex problems. For this project, you will only need to know the instruction set and instruction format of the LC3101.

The LC3101 is an 8-register, 32-bit computer. All addresses are word-addresses (unlike MIPS which is byte-addressed). The LC3101 has 65536 words of memory. By assembly-language convention, register 0 will always contain 0 (i.e. the machine will not enforce this, but no assembly-language program should ever change register 0 from its initial value of 0).

There are 3 instruction formats (bit 0 is the least-significant bit). Bits 31-25 are unused for all instructions, and should always be 0.

R-type instructions (add, nand):

- bits 24-22: opcode
- bits 21-19: reg A
- bits 18-16: reg B
- bits 15-3: unused (should all be 0)
- bits 2-0: destReg

I-type instructions (lw, sw, beq):

bits 24-22: opcode
 bits 21-19: reg A
 bits 18-16: reg B
 bits 15-0: offsetField (16-bit, range of -32768 to 32767)

O-type instructions (halt, noop):

bits 24-22: opcode
 bits 21-0: unused (should all be 0)

Table 1: Description of Machine Instructions

Assembly language name for instruction	Opcode in binary (bits 24, 23, 22)	Action
add (R-type format)	000	add contents of regA with contents of regB, store results in destReg.
nand (R-type format)	001	nand contents of regA with contents of regB, store results in destReg.
lw (I-type format)	010	load regB from memory. Memory address is formed by adding offsetField with the contents of regA.
sw (I-type format)	011	store regB into memory. Memory

4. LC3101 Assembly Language and Assembler (40%)

The first part of this project is to write a program to take an assembly-language program and translate it into machine language. You will translate assembly-language names for instructions, such as beq, into their numeric equivalent (e.g. 100), and you will translate symbolic names for addresses into numeric values. The final output will be a series of 32-bit instructions (instruction bits 31-25 are always 0).

The format for a line of assembly code is:

```
label instruction field0 field1 field2 comments
```

The leftmost field on a line is the label field. Valid labels contain a maximum of 6 characters and can consist of letters and numbers (but must start with a letter). The label is optional (the white space following the label field is required). Labels make it much easier to write assembly-language programs, since otherwise you would need to modify all address fields each time you added a line to your assembly-language program!

After the optional label is white space which consists of any number of space or tab characters. The whitespace is followed by the instruction field, where the instruction can be any of the assembly-language instruction names listed in the above table. After more white space comes a series of fields. All fields are given as decimal numbers or labels. The number of fields depends on the instruction, and unused fields should be ignored (treat them like comments).

R-type instructions (add, nand) instructions require 3 fields: field0 is regA, field1 is regB, and field2 is destReg.

I-type instructions (lw, sw, beq) require 3 fields: field0 is regA, field1 is regB, and field2 is either a numeric value for offsetField or a symbolic address. Numeric offsetFields can be positive or negative; symbolic addresses are discussed below.

O-type instructions (noop and halt) require no fields.

Symbolic addresses refer to labels. For lw or sw instructions, the assembler

are rather cryptic.

`.fill`
(note the leading period). `.fill` tells the assembler to put a number into place where the instruction would normally be stored. `.fill` instructions

```
"start" is at address 2.
```

that the first instruction is at address 0. In the second pass, it will

```

Field needed to branch to that label.

```

```
negl .fill -1
stAddr .fill statt          will contain the address of start (2)
```

And here is the corresponding machine language:

```
(address 0): 8454151 (hex 0x810007)
(address 1): 9043971 (hex 0x8a0003)
(address 2): 655361 (hex 0xa0001)
(address 3): 16842754 (hex 0x1010002)
(address 4): 16842749 (hex 0x100fffd)
(address 5): 29360128 (hex 0x1c00000)
(address 6): 25165824 (hex 0x1800000)
(address 7): 5 (hex 0x5)
(address 8): -1 (hex 0xffffffff)
(address 9): 2 (hex 0x2)
```

Be sure you understand how the above assembly-language program got translated to machine language.

file, one instruction per line. Any deviation from this format (e.g.
extra
spaces or empty lines) will render your machine-

As with the assembler, you will write a suite of test cases to validate the LC3101 simulator.

The test cases for the simulator part of this project will be short assembly-

program halts. You may assume that the two input numbers are at most 15 bits and are positive; this ensures that the (positive) result fits in an LC3101 word. See the algorithm on page 252 of the textbook for how to multiply. Remember that shifting left by one bit is the same as adding the number to itself. Given the LC3101 instruction set, it's easiest to modify the algorithm so that you avoid the right shift. Submit a version of the program that computes $(32766 * 10383)$.

Your multiplication program must be reasonably efficient--it must be at most 50 lines long and execute at most 1000 instructions for any valid numbers (this is several times longer and slower than the solution). To achieve this, you must use a loop and shift algorithm to perform the multiplication; algorithms such as successive addition (e.g. multiplying $5 * 6$ by adding 5 six times) will take too long.

7. Grading and Formatting

We will grade primarily on functionality, including error handling, correctly assembling and simulating all instructions, input and output format, method of executing your program, correctly multiplying, and comprehensiveness of the test suites.

The best way to debug your program is to generate your own test cases, figure out the correct answers, and compare your program's output to the correct answers. This is also one of the best ways to learn the concepts in the project.

The student suite of test cases for the assembler and simulator parts of this project will be graded according to how thoroughly they test an LC3101 assembler or simulator. We will judge thoroughness of the test suite by how well it exposes potentially bugs in an assembler or simulator.

For the assembler test suite, we will use each test case as input to a set of buggy assemblers. A test case exposes a buggy assembler by causing it to generate a different answer from a correct assembler. The test suite is graded based on how many of the buggy assemblers were exposed by at least one test case. This is known as "mutation testing" in the research literature on automated testing.

For the simulator test suite, we will correctly assemble each test case, then use it as input to a set of buggy simulators. A test case exposes a buggy simulator by causing it to generate a different answer from a correct simulator. The test suite is graded based on how many of the buggy simulators were exposed by at least one test case.

8. Turning in the Project

Submit your files through blackboard.
Each part should be archived in a .tar or .zip file to help with grading.

Here are the files you should submit for each project part:

- 1) assembler (part 1a)
 - a. C/C++ program for your assembler
 - b. suite of test cases (each test case is an assembly-language program in a separate file)
- 2) simulator (part 1s)
 - a. C/C++ program for your simulator
 - b. suite of test cases (each test case is an assembly-language program in a separate file)
- 3) multiplication (part 1m)
 - a. assembly program for multiplication

Your assembler and simulator must each be in a single C or C++ file. We will compile your program on linprog using "gcc program.c -lm" (or g++), so your program should not require additional compiler flags or libraries.

The official time of submission for your project will be the time the last file is sent. If you send in anything after the due date, your project will be considered late (and will use up your late days or will receive a zero).

9. Code Fragment for Assembler

The focus of this class is machine organization, not C programming skills. To "build" your computer, however, you will be doing a lot of C programming. To help you, here is a fragment of the C program for the assembler. This shows how to specify command-line arguments to the program (via argc and argv), how to parse the assembly-language file, etc.. This fragment is provided strictly to help you, though it may take a bit for you to understand and use the file.


```

    /* after doing a readAndParse, you may want to do the following to
test the
        opcode */
    if (!strcmp(opcode, "add")) {
        /* do whatever you need to do for opcode "add" */
    }

    return(0);
}

/*
 * Read and parse a line of the assembly-language file. Fields are
returned
 * in label, opcode, arg0, arg1, arg2 (these strings must have memory
already
 * allocated to them).
 *
 * Return values:
 *     0 if reached end of file
 *     1 if all went well
 *
 * exit(1) if line is too long.
 */
int
readAndParse(FILE *inFilePtr, char *label, char *opcode, char *arg0,
             char *arg1, char *arg2)
{
    char line[MAXLINELENGTH];
    char *ptr = line;

    /* delete prior values */
    label[0] = opcode[0] = arg0[0] = arg1[0] = arg2[0] = '\0';

    /* read the line from the assembly-language file */
    if (fgets(line, MAXLINELENGTH, inFilePtr) == NULL) {
        /* reached end of file */
        return(0);
    }

    /* check for line too long (by looking for a \n) */
    if (strchr(line, '\n') == NULL) {
        /* line too long */
        printf("error: line too long\n");
        exit(1);
    }

    /* is there a label? */
    ptr = line;
    if (sscanf(ptr, "%[^\\t\\n ]", label)) {
        /* successfully read label; advance pointer over the label */
        ptr += strlen(label);
    }

    /*

```

```

        * Parse the rest of the line.  Would be nice to have real regular
        * expressions, but scanf will suffice.
        */
        sscanf(ptr, "%*[\t\n ]%[^\\t\\n ]%*[\t\n ]%[^\\t\\n ]%*[\t\n ]%[^\\t\\n
]%^*[\t\n ]%[^\\t\\n ]",
               opcode, arg0, arg1, arg2);
        return(1);
    }

int
isNumber(char *string)
{
    /* return 1 if string is a number */
    int i;
    return( (sscanf(string, "%d", &i)) == 1);
}

```

10. Code Fragment for Simulator

Here is some C code that may help you write the simulator. Again, you should take this merely as a hint. You may have to re-code this to make it do exactly what you want, but this should help you get started. Remember not to change stateStruct or printState.

```

/* instruction-level simulator for LC3101 */

#include <stdio.h>
#include <string.h>

#define NUMMEMORY 65536 /* maximum number of words in memory */
#define NUMREGS 8 /* number of machine registers */
#define MAXLINELENGTH 1000

typedef struct stateStruct {
    int pc;
    int mem[NUMMEMORY];
    int reg[NUMREGS];
    int numMemory;
} stateType;

void printState(stateType *);

int
main(int argc, char *argv[])
{
    char line[MAXLINELENGTH];
    stateType state;
    FILE *filePtr;

    if (argc != 2) {

```


bits. Neither a nor b are changed. E.g. (25 >> 2) is 6. Note that 25 is 11001 in binary, and 6 is 110 in binary.

3) The value of the expression (a << b) is the number "a" shifted left by "b" bits. Neither a nor b are changed. E.g. (25 << 2) is 100. Note that 25 is 11001 in binary, and 100 is 1100100 in binary.

4) To find the value of the expression (a & b), perform a logical AND on each bit of a and b (i.e. bit 31 of a ANDED with bit 31 of b, bit 30 of a ANDED with bit 30 of b, etc.). E.g. (25 & 11) is 9, since:

```
      11001 (binary)
      & 01011 (binary)
      -----
= 01001 (binary), which is 9 decimal.
```

5) To find the value of the expression (a | b), perform a logical OR on each bit of a and b (i.e. bit 31 of a ORED with bit 31 of b, bit 30 of a ORED with bit 30 of b, etc.). E.g. (25 | 11) is 27, since:

```
      11001 (binary)
      & 01011 (binary)
      -----
= 11011 (binary), which is 27 decimal.
```

6) ~a is the bit-wise complement of a (a is not changed).

Use these operations to create and manipulate machine-code. E.g. to look at bit 3 of the variable a, you might do: (a>>3) & 0x1. To look at bits (bits 15-12) of a 16-bit word, you could do: (a>>12) & 0xF. To put a 6 into bits 5-3 and a 3 into bits 2-1, you could do: (6<<3) | (3<<1). If you're not sure what an operation is doing, print some intermediate results to help you debug.

12. Example Run of Simulator

```
memory[0]=8454151
memory[1]=9043971
memory[2]=655361
memory[3]=16842754
memory[4]=16842749
memory[5]=29360128
memory[6]=25165824
```

```
memory[7]=5
memory[8]=-1
memory[9]=2
```

```
@@@
```

```
state:
```

```
pc 0
```

```
memory:
```

```
mem[ 0 ] 8454151
mem[ 1 ] 9043971
mem[ 2 ] 655361
mem[ 3 ] 16842754
mem[ 4 ] 16842749
mem[ 5 ] 29360128
mem[ 6 ] 25165824
mem[ 7 ] 5
mem[ 8 ] -1
mem[ 9 ] 2
```

```
registers:
```

```
reg[ 0 ] 0
reg[ 1 ] 0
reg[ 2 ] 0
reg[ 3 ] 0
reg[ 4 ] 0
reg[ 5 ] 0
reg[ 6 ] 0
reg[ 7 ] 0
```

```
end state
```

```
@@@
```

```
state:
```

```
pc 1
```

```
memory:
```

```
mem[ 0 ] 8454151
mem[ 1 ] 9043971
mem[ 2 ] 655361
mem[ 3 ] 16842754
mem[ 4 ] 16842749
mem[ 5 ] 29360128
mem[ 6 ] 25165824
mem[ 7 ] 5
mem[ 8 ] -1
mem[ 9 ] 2
```

```
registers:
```

```
reg[ 0 ] 0
reg[ 1 ] 5
reg[ 2 ] 0
reg[ 3 ] 0
reg[ 4 ] 0
reg[ 5 ] 0
reg[ 6 ] 0
reg[ 7 ] 0
```

```
end state
```



```

@@@
state:
    pc 2
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 5
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state

```

```

@@@
state:
    pc 3
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 4
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state

```

```

@@@
state:
    pc 4

```

```

memory:
    mem[ 0 ] 8454151
    mem[ 1 ] 9043971
    mem[ 2 ] 655361
    mem[ 3 ] 16842754
    mem[ 4 ] 16842749
    mem[ 5 ] 29360128
    mem[ 6 ] 25165824
    mem[ 7 ] 5
    mem[ 8 ] -1
    mem[ 9 ] 2
registers:
    reg[ 0 ] 0
    reg[ 1 ] 4
    reg[ 2 ] -1
    reg[ 3 ] 0
    reg[ 4 ] 0
    reg[ 5 ] 0
    reg[ 6 ] 0
    reg[ 7 ] 0
end state

```

@@@

```

state:
    pc 2
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 4
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state

```

@@@

```

state:
    pc 3
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361

```

```

        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 3
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state

```

```

@@@
state:
    pc 4
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 3
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state

```

```

@@@
state:
    pc 2
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824

```

```

        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 3
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state

```

@@@

state:

pc 3

memory:

```

        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2

```

registers:

```

        reg[ 0 ] 0
        reg[ 1 ] 2
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0

```

end state

@@@

state:

pc 4

memory:

```

        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2

```

registers:

```
        reg[ 0 ] 0
        reg[ 1 ] 2
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state
```

```
@@@
state:
    pc 2
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 2
        reg[ 2 ] -1
        reg[ 3 ] 0
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
```

end state

```
@@@
state:
    pc 3
    memory:
        mem[ 0 ] 8454151
        mem[ 1 ] 9043971
        mem[ 2 ] 655361
        mem[ 3 ] 16842754
        mem[ 4 ] 16842749
        mem[ 5 ] 29360128
        mem[ 6 ] 25165824
        mem[ 7 ] 5
        mem[ 8 ] -1
        mem[ 9 ] 2
    registers:
        reg[ 0 ] 0
        reg[ 1 ] 1
        reg[ 2 ] -1
        reg[ 3 ] 0
```

```
        reg[ 4 ] 0
        reg[ 5 ] 0
        reg[ 6 ] 0
        reg[ 7 ] 0
end state
```

@@@

state:

pc 4

memory:

```
mem[ 0 ] 8454151
mem[ 1 ] 9043971
mem[ 2 ] 655361
mem[ 3 ] 16842754
mem[ 4 ] 16842749
mem[ 5 ] 29360128
mem[ 6 ] 25165824
mem[ 7 ] 5
mem[ 8 ] -1
mem[ 9 ] 2
```

registers:

```
reg[ 0 ] 0
reg[ 1 ] 1
reg[ 2 ] -1
reg[ 3 ] 0
reg[ 4 ] 0
reg[ 5 ] 0
reg[ 6 ] 0
reg[ 7 ] 0
```

end state

@@@

state:

pc 2

memory:

```
mem[ 0 ] 8454151
mem[ 1 ] 9043971
mem[ 2 ] 655361
mem[ 3 ] 16842754
mem[ 4 ] 16842749
mem[ 5 ] 29360128
mem[ 6 ] 25165824
mem[ 7 ] 5
mem[ 8 ] -1
mem[ 9 ] 2
```

registers:

```
reg[ 0 ] 0
reg[ 1 ] 1
reg[ 2 ] -1
reg[ 3 ] 0
reg[ 4 ] 0
reg[ 5 ] 0
reg[ 6 ] 0
reg[ 7 ] 0
```

end state

@@@

state:

pc 3

memory:

mem[0]	8454151
mem[1]	9043971
mem[2]	655361
mem[3]	16842754
mem[4]	16842749
mem[5]	29360128
mem[6]	25165824
mem[7]	5
mem[8]	-1
mem[9]	2

registers:

reg[0]	0
reg[1]	0
reg[2]	-1
reg[3]	0
reg[4]	0
reg[5]	0
reg[6]	0
reg[7]	0

end state

@@@

state:

pc 6

memory:

mem[0]	8454151
mem[1]	9043971
mem[2]	655361
mem[3]	16842754
mem[4]	16842749
mem[5]	29360128
mem[6]	25165824
mem[7]	5
mem[8]	-1
mem[9]	2

registers:

reg[0]	0
reg[1]	0
reg[2]	-1
reg[3]	0
reg[4]	0
reg[5]	0
reg[6]	0
reg[7]	0

end state

machine halted

total of 17 instructions executed

final state of machine:

```
@@@
state:
  pc 7
  memory:
    mem[ 0 ] 8454151
    mem[ 1 ] 9043971
    mem[ 2 ] 655361
    mem[ 3 ] 16842754
    mem[ 4 ] 16842749
    mem[ 5 ] 29360128
    mem[ 6 ] 25165824
    mem[ 7 ] 5
    mem[ 8 ] -1
    mem[ 9 ] 2
  registers:
    reg[ 0 ] 0
    reg[ 1 ] 0
    reg[ 2 ] -1
    reg[ 3 ] 0
    reg[ 4 ] 0
    reg[ 5 ] 0
    reg[ 6 ] 0
    reg[ 7 ] 0
end state
```