

C SCI 316 (Kong): TinyJ Assignment 3

This assignment is to be submitted *no later than* **Thursday, December 18**. [Even though this submission deadline is after the final exam, there will be questions on the final exam which assume you understand how to write the code that needs to be written to complete this assignment.] See page 5 for the *late* submission deadline.

The assignment is to further develop your TJasn program so it will correctly *execute* the TinyJ virtual machine (VM) instructions that it generates. For each VM instruction, there will be a corresponding file `*instr.java` in your `TJasn/virtualMachine` directory.[‡] Every such file has an `execute()` method. In `ADDTOPTRinstr.java`, `HEAPALLOCinstr.java`, `NOPorDISCARDVALUEinstr.java`, `READINTinstr.java`, and `STOPinstr.java`, that method is written for you. In the other 29 `*instr.java` files, the body of the `execute()` method contains a comment of the form `/* ???????? */` which you need to replace with code that executes the corresponding VM instruction: Your code must make appropriate changes to the expression evaluation stack [`CodeInterpreter.EXPRSTACK[]`], data memory [`TJ.data[]`], and registers [`CodeInterpreter.ESP`, `.PC`, `.FP`, `.ASP`, etc.], and must produce correct output (if any). See the example at the beginning of the "How to Do the Assignment" section below.

[‡]The `*instr.java` files may be **Windows** text files: The line-feed character at the end of each line might be preceded by a Ctrl-M (carriage return). If so, and you plan to edit the files using `emacs` on `venus` or using `vi`, `vim`, or `emacs` on `euclid`, then you should first eliminate these Ctrl-Ms. This can be done by entering the following:
On **euclid**: `/users/kong300/convert` On **venus**: `/home/faculty/ykong/convert`

Some Information About the Implementation of the TinyJ Virtual Machine

In `CodeInterpreter.java`, the variables **PC**, **IR**, **ESP**, **FP**, and **ASP** (see lines 14 – 18) represent VM registers; the array **EXPRSTACK[]** (line 19) represents the expression evaluation stack.

PC contains the code memory address of the next instruction to be fetched for execution. (When that instruction is fetched, it is put into **IR** for execution.)

ESP stores a count of the number of items that are currently on the expression evaluation stack—if **ESP** > 0, then **EXPRSTACK[ESP – 1]** is the topmost element of the stack.

FP is a pointer to the data memory location at offset 0 in the currently executing method activation's stackframe. (Thus **FP**+*k* will be a pointer to the location at offset *k* in that stackframe.)

ASP is a pointer to the *first unused* location in the stack-dynamically allocated part of data memory—i.e., **ASP** points to the first location above the currently allocated locations in that part of data memory. (**ASP** must therefore be increased/decreased by *k* when *k* stackframe locations are allocated/deallocated.)

In `TJ.java`, the `ArrayList` **generatedCode** (line 30) and the array **data[]** (line 26) represent code and data memory, respectively. **TJ.generatedCode.get(*a*)** returns the instruction stored in code memory at address *a*. **TJ.data[*a*]** represents the data memory location whose address is *a*.

The VM's fetch-execute loop is in the `interpretCode()` method of `CodeInterpreter.java` (line 48); this is called by `start()` (see line 247). Its two most important lines are:

```
75  IR = TJ.generatedCode.get(oldPC = PC++);    and    77  IR.execute();
```

Line 75 *fetches* the instruction stored (in code memory) at the address in **PC** and increments **PC**.[†] Line 77 *executes* the instruction that has just been fetched.

[†]The address of the instruction that is fetched is stored into **oldPC**—so when that instruction is being executed, and immediately afterwards, **oldPC** will contain the address of the instruction. When a debugging stop occurs—see the "How to Debug If You are Stumped" section below—the VM uses the address in **oldPC** to determine which instruction has just been executed. **PC – 1** will usually be equal to **oldPC**, but will sometimes be different because **PC** may be changed by execution of **JUMP**, **JUMPONFALSE**, **CALLSTATMETHOD**, or **RETURN**.

Installation of DEMO Instruction Classes

Installation of TinyJ Assignment 2 should have installed the instruction classes from my solution (the "DEMO instruction classes") into your TJsolclasses/TJasn/virtualMachine directory. **Before you work on this assignment, use the following commands to copy these DEMO instruction classes into your TJasn/virtualMachine directory:**

```
cp TJsolclasses/TJasn/virtualMachine/*instr.class TJasn/virtualMachine
[on euclid and venus, working directory = your home directory]
copy /y TJsolclasses\TJasn\virtualMachine\*instr.class TJasn\virtualMachine
[in a cmd.exe window on a PC, working directory = c:\316java]
```

Then, since the DEMO classes have correct execute() methods, once you have completed TinyJ Assignment 2 you should be able to execute TinyJ programs (e.g., the 16 CS316ex*.java files) as follows: `java -cp . TJasn.TJ CS316exk.java k.out`

E.g., `java -cp . TJasn.TJ CS316ex9.java 9.out` should write the following to the screen:

Using the following DEMO instruction classes:

```
PUSHSTATADDR PUSHNUM SAVETOADDR INITSTKFRM LOADFROMADDR PASSPARAM CALLSTATMETHOD ADD
CHANGESIGN WRITEINT WritelnOP WRITESTRING PUSHLOCADDR EQ JUMPONFALSE JUMP SUB RETURN
```

(This says that your program is using my DEMO versions of these instruction classes, rather than your own versions of the classes.) The program will ask:

Want debugging stop or post-execution dump? (y/n)

Respond by typing **n** and the program will begin to execute the CS316ex9 program.

(The output should be the same as the output you get when you compile CS316ex9.java using `javac CS316ex9.java` and enter `java CS316ex9` to execute the generated byte codes.)

How to Do the Assignment

This assignment assumes you have completed Assignment 2 and that you know what each TinyJ VM instruction should do when it is executed—if you are not clear as to what certain instructions should do when executed, refer back to the “Effects of Executing Each TinyJ Virtual Machine Instruction” pages at: <http://euclid.cs.gc.cuny.edu/316/VM-instruction-set-and-hints-for-asn-2.pdf>

As an example of how to complete an execute() method, consider the execute() method of ADDinstr.java. As Java (unlike C++) is guaranteed to evaluate expressions from left to right, the `/* ???????? */` in ADDinstr.java can be replaced with the following statement:

```
EXPRSTACK[--ESP-1] += EXPRSTACK[ESP];
```

After making this change, use the command

```
javac -cp . TJasn/virtualMachine/ADDinstr.java
```

to recompile ADDinstr.java and produce a file ADDinstr.class which will replace the DEMO version of ADDinstr.class you copied into TJasn/virtualMachine. The new version of ADDinstr.class can then be tested by executing a TinyJ program whose virtual machine code uses the ADD instruction. For example, you could enter:

```
java -cp . TJasn.TJ CS316ex9.java 9.out
```

When you do this you should find that ADD is no longer on the list of DEMO instruction classes reported by TJasn.TJ! [The javac and java commands above assume that your working directory is your home directory on venus or euclid, and c:\316java on a PC.]

I suggest you work on the other instruction classes in the following order. First, work on MUL, SUB, DIV, MOD, and CHANGESIGN, which are used by, e.g., CS316ex0.java. Next, work on NOT, AND, OR, EQ, NE, LT, GT, LE, GE, PUSHNUM, WritelnOP, WRITEINT, WRITESTRING, JUMP, JUMPONFALSE, PUSHSTATADDR, LOADFROMADDR, and SAVETOADDR, which are used by, e.g., CS316ex3.java. Finally, work on PUSHLOCADDR, PASSPARAM, CALLSTATMETHOD, INITSTKFRM, and RETURN, all of which are used by, e.g., CS316ex9.java.

Each time you complete the `execute()` method of a file `xxxinstr.java`, it might well be a good idea to immediately compile that `.java` file and then (if there are no compilation errors) immediately test your `execute()` method by using `TJasn.TJ` to execute a `CS316exk.java` program whose virtual machine code uses the instruction in question. If `TJasn.TJ` fails to execute that `CS316exk` program correctly, then your new `execute()` method is bugged and you need to fix it. [If you want to work on another instruction's `execute()` method first, then copy my DEMO version of `xxxinstr.class` back from your `TJsolclasses/TJasn/virtualMachine` directory into your `TJasn/virtualMachine` directory, so it replaces your bugged version.] If the `CS316exk` program is executed correctly, save a backup copy of the file that contains your new `execute()` method—e.g., put the copy on **venus** if you are working on **euclid**.

How to Debug If You are Stumped

Suppose that you have just written the `execute()` method of `LOADFROMADDRinstr.java` (and that any other `execute()` methods you may have written earlier all work). After compiling `LOADFROMADDRinstr.java` [`javac -cp . TJasn/virtualMachine/LOADFROMADDRinstr.java`], if your program `TJasn.TJ` executes some TinyJ input file [say, `CS316ex3.java`] incorrectly, try the following debugging method:

Run `TJasn.TJ` on that same input file again, **but set the TinyJ virtual machine to stop as soon as it has executed a `LOADFROMADDR` instruction.** Do this as follows:

```
java -cp . TJasn.TJ CS316ex3.java 3a.out
Want debugging stop or post-execution dump? (y/n)  y
Enter MINIMUM no. of instructions to execute before debugging stop.
(Enter -1 to get a post-execution dump but no debugging stop.):  0
Stop after executing what instruction? (e.g., PUSHNUM)
(Enter * to stop after executing just 0 instructions.):  LOADFROMADDR
```

Next, **copy my DEMO version** of `LOADFROMADDRinstr.class` back from your `TJsolclasses/TJasn/virtualMachine` directory into your `TJasn/virtualMachine` directory, and then repeat the above with the same debugging stop settings. **But this time use a different output file name**—`3b.out` instead of `3a.out`, say. After these two runs, the file `3a.out` will contain dumps of the contents of the virtual machine's registers and memory locations *immediately before and immediately after* execution of **your** version of `LOADFROMADDR`'s `execute()`, whereas `3b.out` will contain the same information for **my** version of `LOADFROMADDR`'s `execute()`. Use **`diff -c`** or **`fc /n`** to compare `3a.out` and `3b.out`. Their dumps *before* execution of the `LOADFROMADDR` instruction should be identical. (If they differ, one of the `execute()` methods you wrote earlier is bugged!) But the dumps *after* execution of `LOADFROMADDR` may well be different, in which case you should be able to see what your version of `LOADFROMADDR`'s `execute()` did wrong.

In the event that your and my programs' dumps *after* execution of `LOADFROMADDR` are also identical, make a note of the *total number of instructions executed before execution was stopped*—this number will be shown near the bottom of the "after" dumps. If this number is, say, **27**, then repeat the above process, **but this time set the virtual machine to execute $27+1 = 28$ instructions**, and to then stop after it executes any `LOADFROMADDR` instruction:

```
Want debugging stop or post-execution dump? (y/n)  y
Enter MINIMUM no. of instructions to execute before debugging stop.
(Enter -1 to get a post-execution dump but no debugging stop.):  28
Stop after executing what instruction? (e.g., PUSHNUM)
(Enter * to stop after executing just 28 instructions.):  LOADFROMADDR
```

Use of CodeInterpreter.POINTERTAG

In the TinyJ virtual machine, a pointer to *data memory* address a is represented by the integer $a + \text{CodeInterpreter.POINTERTAG}$, where `CodeInterpreter.POINTERTAG` is a constant equal to $0x7FFF0000 = 01111111\ 11111111\ 00000000\ 00000000$. [This depends on the fact that the machine's data memory addresses are small enough to be stored in the lower two bytes of an int.]

When executing **PUSHSTATADDR** a remember to push the *pointer* corresponding to address a . In other words, remember to push $a + \text{POINTERTAG}$. When executing **LOADFROMADDR** remember that the topmost stack item is a *pointer*: If p is on top of the stack, then the data memory address to load from is $p - \text{POINTERTAG}$. Again, when executing **SAVETOADDR**, remember that the second item on the stack is a *pointer*: If p is the second item on the stack, then the data memory address to write to is $p - \text{POINTERTAG}$. When executing **CALLSTATMETHOD**, **INITSTKFRM**, **PASSPARAM**, and **RETURN**, remember that **ASP** is a pointer and refers to the data memory address $\text{ASP} - \text{POINTERTAG}$.

When executing **PUSHLOCADDR** a remember that **FP** *already contains a pointer*, so you should *not* add `POINTERTAG` to $\text{FP} + a$. You also should *not* subtract `POINTERTAG` from **FP** before storing it during execution of **INITSTKFRM**, and should *not* add `POINTERTAG` on retrieving **FP** from the stackframe when executing **RETURN**. Similarly, since **FP** and **ASP** both store pointers, do *not* add or subtract `POINTERTAG` when copying from **ASP** into **FP** or vice versa during execution of **INITSTKFRM** and **RETURN**.

When executing **WRITESTRING** $a\ b$ remember that a and b are ordinary data memory addresses, not pointers: Do *not* subtract `POINTERTAG` from a and b .

`POINTERTAG` is *not used* with code memory addresses: Never subtract `POINTERTAG` from **PC**, and never add `POINTERTAG` to a code memory address.

How to Test Your Solution

When you have correctly completed all the incomplete `execute()` methods, and recompiled each of the modified `*instr.java` files, test your program on the 16 `CS316exk.java` files. The runtime behavior should always be the same as when the same source file is compiled using `javac` and the resulting `.class` file is executed by the Java virtual machine. Also, the dump after execution of, say, 1000 instructions should be the same for your program as for my solution. Recall that you can run *my solution* as follows:

```
java -cp TJsolclasses:. TJasn.TJ CS316exk.java k.sol [on euclid or venus]
java -cp TJsolclasses;. TJasn.TJ CS316exk.java k.sol [on a PC]
```

How to Submit Your Solution

This assignment is to be submitted *no later than **Thursday, December 18***. [Even though this deadline is after the final exam, *there will be questions on the final exam which assume you understand how to write the code that needs to be written to complete this assignment*. As usual, if euclid unexpectedly goes down after 6 p.m. on the due date, the deadline will *not* be extended. Try to submit before noon that day, and on an earlier day if possible.] See page 5 for the *late* submission deadline.

A correct solution will count 2% towards your grade. You may work with up to two other students; when two or three students work together, *each student must submit separately*.

If you have been working on **euclid**, submit by following just the steps 1, 3, and 4 on the next page; *do not* do step 2! But if you have been working on **venus** or your PC (*NOT euclid!*), do all of 1 – 4 on the next page. Your working directory should be your home directory on **venus / euclid** and `c:\316java` on a PC.

1. Remove any incomplete ***instr.java** files you have **NOT** successfully completed from the TJasn/virtualMachine directory of *the machine you have been working on*.

Then enter the following command on **euclid**:

```
cp TJasn/classes/TJasn/virtualMachine/*instr.class TJasn/virtualMachine
```

2. This step does **not** apply if you have been working on **euclid**. If, and **only if**, you have been working on **venus** or your PC (**NOT euclid!**), enter the following command on **euclid**:

```
rm TJasn/virtualMachine/*instr.java
```

You must not do this if you have been working on euclid! Then logout from **euclid**.

If you have been working on **your PC**, you should now do 2(a) below. If you have been working on **venus**, do **either** 2(a) **or** 2(b) below—you need only do one of 2(a) and 2(b).

- 2(a) On the machine you have been working on (**venus** or your PC), create a jar archive of all the ***instr.java** files in TJasn/virtualMachine, as follows:

```
jar cvf instrs.jar TJasn/virtualMachine/*instr.java
```

This creates a jar archive file **instrs.jar** on **venus** or your PC. Use an scp or sftp client to copy **instrs.jar** from **venus** or your PC into your home directory on **euclid**. (If you have forgotten how to do this, see the second paragraph of the “**How to Submit Your Solutions**” section of the Lisp Assignment 3 handout, but substitute **instrs.jar** for **c:\316lisp\horn-3.lsp** when following those instructions.)

Then logon to **euclid** and extract the archived ***instr.java** files as follows:

```
euclid> jar xvf instrs.jar
```

- 2(b) If you have been working on **venus**, there's a slightly easier alternative to 2(a): Just enter

```
scp TJasn/virtualMachine/*instr.java xxxxx316@euclid.cs.gc.cuny.edu:TJasn/virtualMachine
```

on venus, where **xxxxx316** means your **euclid** username and the **:** after **.edu** is needed.

3. On **euclid**, recompile all the ***instr.java** files you have completed, as follows:

```
euclid> javac TJasn/virtualMachine/*instr.java
```

If any ***instr.java** file cannot be compiled on **euclid**, you will receive no credit for this assignment. (In view of step 1 above, this should not happen.)

4. Test your program on **euclid**. (See the section "How to Test Your Solution" above.)

IMPORTANT: Do NOT open any of your submitted *instr.java** files in an editor on euclid after the due date, unless you are resubmitting a corrected version of your solution as a late submission.**

Late Submission Deadline

I *expect* **euclid** to stay up after December 22. If **euclid** stays up, then **late** submissions of this and previous programming assignments (including corrected submissions) will be accepted until **noon on Wednesday, December 24**. However, this deadline will **not** be extended if **euclid** goes down at any time after **noon on December 22**. Note that if **euclid** goes down after noon on the 22nd, there is a possibility that it might not be brought back up until after noon on the 24th, in which case those students who did not submit before **euclid** went down would not be able to submit at all! To avoid the risk of not being able to submit, make any late / corrected submissions no later than **noon on Monday, December 22**.

Hints for TinyJ Assignment 3

Before you work on the assignment, carefully read page 1 above. When writing the `execute()` method for an instruction, if you are not clear as to what that instruction should do when it is executed, refer back to the “Effects of Executing Each TinyJ Virtual Machine Instruction” pages at:

<http://euclid.cs.qc.cuny.edu/316/VM-instruction-set-and-hints-for-asn-2.pdf>

Remarks on the `execute()` Methods You Have to Write

1. Instructions That Do Not Use Pointers to Data Memory Locations and Have No Operands:

ADD, SUB, DIV, MUL, MOD, AND, OR

As explained on p. 2 of the assignment document, **ADD**'s `execute()` method can be implemented as follows:

```
EXPRSTACK[--ESP-1] += EXPRSTACK[ESP];
```

The `execute()` methods for **SUB, DIV, MUL, MOD, AND, and OR** can be written analogously. Use 1 and 0 to represent **true** and **false**; then the `execute()` methods for **AND** and **OR** can use `&=` and `|=` where **ADD**'s `execute()` uses `+=`.

LE, GE, LT, GT, EQ, NE

One simple way to write the `execute()` methods for these instructions is to use conditional expressions. For example, **LE**'s `execute()` method could be implemented as follows:

```
EXPRSTACK[--ESP-1] = (EXPRSTACK[ESP-1] <= EXPRSTACK[ESP]) ? 1 : 0;
```

CHANGESIGN, NOT

These instructions should change `EXPRSTACK[ESP-1]` without changing `ESP`. (Note: `*= -1` is a concise way to negate an **int** variable; `^= 1` is a concise way to change its value from 1 to 0 or vice versa.)

WRITELNOP, WRITEINT

These instructions should print to `System.out`: **WRITELNOP** should print a newline, and **WRITEINT** should print an integer that is popped off `EXPRSTACK`. (Note that `EXPRSTACK[--ESP]` represents a value that is popped off `EXPRSTACK`.)

2. Instructions That Do Not Use Pointers to Data Memory Locations but Have Operands

The operand of any `OneOperandInstruction` (such as `PUSHNUMinstr` or `JUMPinstr`) is in its `operand` field (which it inherits from `OneOperandInstruction`). The two operands of a `WRITESTRINGinstr` are in its `firstOperand` and `secondOperand` fields (which it inherits from `TwoOperandInstruction`).

Recall from p. 1 of the assignment document that `PC` represents the VM's program counter: It stores the *code memory* address of the location from which the next instruction will be fetched for execution.

- | | |
|-------------------------|---|
| PUSHNUM 17 | should push its operand onto <code>EXPRSTACK</code> (i.e., put 17 into <code>EXPRSTACK[ESP++]</code>). |
| JUMP 17 | should put its operand 17 into <code>PC</code> , as <code>PC</code> represents the program counter. |
| JUMPONFALSE 17 | should look at a value that is popped off <code>EXPRSTACK</code> (i.e., look at the value <code>EXPRSTACK[--ESP]</code>) and put its operand 17 into <code>PC</code> just if that popped value is 0. |
| WRITESTRING 7 17 | should print (to <code>System.out</code>) the characters that are stored in the data memory locations with addresses \geq its <code>firstOperand</code> 7 but \leq its <code>secondOperand</code> 17. These locations are represented by:
<code>TJ.data[7], TJ.data[8], ..., TJ.data[16], TJ.data[17]</code>
You must perform a <code>(char)</code> cast on the value in each location before printing it. |

3. Instructions That Use Pointers to Data Memory Locations

Before you write the `execute()` methods for these instructions, you should read the section “Use of `CodeInterpreter.POINTERTAG`” on p. 4 of the assignment document.

Also recall from p. 1 of the assignment document that *FP stores a pointer to the data memory location at offset 0 in the currently executing method activation's stackframe*. (Thus $FP + k$ will be a pointer to the location at offset k in that stackframe.)

PUSHSTATADDR 7	This should push $7 + \text{POINTERTAG}$ (i.e., push a pointer to the data memory location whose address is 7) onto <code>EXPRSTACK</code> ; in other words, it should put the pointer $7 + \text{POINTERTAG}$ into <code>EXPRSTACK[ESP++]</code> .
PUSHLOCADDR 7	This should push $FP + 7$ (i.e., push a pointer to the data memory location at offset 7 in the currently executing method activation's stackframe) onto <code>EXPRSTACK</code> ; in other words, it should put the pointer $FP + 7$ into <code>EXPRSTACK[ESP++]</code> .
LOADFROMADDR	If p is the pointer in <code>EXPRSTACK[ESP-1]</code> , execution of this instruction should copy the value in the data memory location to which p points (i.e., copy the value $TJ.data[p - \text{POINTERTAG}]$) into <code>EXPRSTACK[ESP-1]</code> ; the copied value should overwrite the pointer p .
SAVETOADDR	If v is the value in <code>EXPRSTACK[ESP-1]</code> and p is the pointer in <code>EXPRSTACK[ESP-2]</code> , then this instruction should store v into the data memory location to which p points (i.e., it should store v into $TJ.data[p - \text{POINTERTAG}]$), and should also decrease <code>ESP</code> by 2 to pop v and p off <code>EXPRSTACK</code> .

4. Instructions Associated with Calls of Static Methods and Return from Called Methods

Recall from p. 1 of the assignment document that *ASP stores a pointer to the first unused location in the stack-dynamically allocated part of data memory*—i.e., *ASP* points to the *first location above* the currently allocated locations in that part of data memory.

ASP must therefore be increased / decreased by k when k stackframe locations are allocated / deallocated.

We will write	S-PUSH y	for	$TJ.data[\text{ASP}++ - \text{POINTERTAG}] = y;$
We will write	S-POP y	for	$y = TJ.data[--\text{ASP} - \text{POINTERTAG}];$
51: PASSPARAM	should	S-PUSH a value that is popped off <code>EXPRSTACK</code> —i.e., it should S-PUSH <code>EXPRSTACK[--ESP]</code>	
52: CALLSTATMETHOD 671	should and then	S-PUSH <code>PC</code> Set <code>PC</code> to 671	[saves return addr (here, 53) into new frame] [transfers control to the called method's code]
671: INITSTKFRM 7	should and then and then	S-PUSH <code>FP</code> Set <code>FP</code> to $\text{ASP} - 1$ Increase <code>ASP</code> by 7	[saves caller's <code>FP</code> at offset 0 in the new frame] [makes <code>FP</code> point to offset 0 in the new frame] [allocates space for callee's local variables]
713: RETURN 4	should and then and then and then	Set <code>ASP</code> to $\text{FP} + 1$ S-POP <code>FP</code> S-POP <code>PC</code> Decrease <code>ASP</code> by 4	[deallocates space used by callee's variables] [restores caller's <code>FP</code>] [puts the saved return address into <code>PC</code>] [deallocates space used by formal parameters]