Computer Science 3 - 2011

Programming Language Translation

Practical for Week 20, beginning 5 September 2011

Hand in this prac sheet *before* lunch time on your next practical day, correctly packaged in a transparent folder with your solutions and the "cover sheet". **Unpackaged and late submissions will not be accepted - you have been warned.** Please do NOT come to a practical and spend the first hour printing or completing solutions from the previous week's exercises. Since the practical will have been done on a group basis, please hand in one copy of the cover sheet for each member of the group. These will be returned to you in due course, signed by the marker.

Objectives:

In this practical you are to

- become familiar you with the workings of a simple machine emulator for the PVM pseudo-machine we shall use frequently in the course.
- gain some experience with the machine, writing machine code for it, and extending it.

You will need this prac sheet and your text book. Copies of the prac sheet and of the Parva report are also available at `http://www.cs.ru.ac.za/courses/CSc301/Translators/trans.htm` .

Outcomes:

When you have completed this practical you should understand

- the opcode set for the Parva Virtual Machine (PVM);
- how to write and debug machine level code for the PVM;
- how to extend the PVM to incorporate new opcodes;
- why, and by how much, how interpretive systems are slower than native code systems.

To hand in:

This week you are required to hand in, besides the cover sheet:

- Listings of the final version of the assembler/emulator system you produce, and your solutions to the programming exercises below. (Use LPRINT, please.)
- Preferably, electronic copies of your source code for those exercises, using the electronic submission system.
- Discussion of the experiments in Task 9.

Keep the prac sheet and your solutions until the end of the semester. Check carefully that your mark has been entered into the Departmental Records.

You are referred to the rules for practical submission which are clearly stated in our Departmental Handbook. However, for this course pracs must be posted in the "hand-in" box outside the laboratory and not given to demonstrators.

A rule not stated there, but which should be obvious, is that you are not allowed to hand in another group's or student's work as your own. Attempts to do this will result in (at best) a mark of zero and (at worst) severe disciplinary action and the loss of your DP. You are allowed - even encouraged - to work and study with other students, but if you do this you are asked to acknowledge that you have done so. You are expected to be familiar with the University Policy on Plagiarism, which you can consult at:

`http://www.scifac.ru.ac.za/plagiarism_policy.pdf`

Task 1 - creating a working directory and unpacking the prac kit

There are several files that you need, zipped up this week in the file `PRAC20.ZIP` or `PRAC20C.ZIP`.

- Immediately after logging on, get to the DOS command line level by using the `Start -> Command prompt` option from the tool bar.

- Copy the prac kit into a newly created directory/folder in your file space. For Java programmers:

```
md  prac20
cd  prac20
copy  i:\csc301\trans\prac20.zip
unzip  prac20.zip
```

  This will create several other directories "below" the `prac20` directory:

```
J:\prac20
J:\prac20\Assem
J:\prac20\Library
```

  containing the Java classes for the I/O Library, and the Java sources for an assembler/interpreter system equivalent to the C# one described in Chapter 4. The differences between C# and Java are very minimal and it is hoped that you will have no problems in this regard.

- If UltraEdit is your editor of choice, the version in the lab has been configured to run various of the compilers easily, and it is possible to tweak it to run others in the same sort of way. *To get this to work properly, start UltraEdit from a command window by giving the command* `UEDIT32`*, rather than by clicking on the icon on the desktop.*

- You may prefer to work in C#. A version of the prac kit is available (`prac20c.zip`) with C# versions of the system and source files. It all works in much the same way.

Task 2 - A look at PVM code

Start off by considering the following gem of a Parva program which produces a truth table that illustrates de Morgan's famous laws.

```
void main () {
/* Demonstrate de Morgan's Laws
   P.D. Terry, Rhodes University, 2011 */

  bool X, Y;

  write("   X      Y    (X.Y)' X'+Y'  (X+Y)' X'.Y'\n\n");
  X = false;
  repeat
    Y = false;
    repeat
      write(X, Y, !(X && Y),  !X || !Y, !(X || Y), !X && !Y, "\n");
      Y = ! Y;
    until (!Y); // again
    X = ! X;
  until (!X); // again
} // main
```

You can compile this (`PARVA DEMORGAN.PAV`) at your leisure to make quite sure that it works.

The Parva compiler supplied to you this week is not the same as last week - it only allows a single `main` function, but it includes "else" and the modulo "`%`" operator and it supports a "repeat" … "until" statement (as in this and later examples).

In the prac kit you will also find a translation of this program into PVM code (`DEMORGAN.PVM`). Study this code (which to save paper I have not printed out here) and answer the following questions:

(a) Has the code used short-circuit Boolean operations?

(b) What would you need to change if you wanted to produce the truth table in the following format? (Don't worry too much about spacing and lining up!)

```
X      Y    (X.Y)' X'+Y'  (X+Y)' X'.Y'

0      0      1      1      1      1
0      1      1      1      0      0
1      0      1      1      0      0
1      1      0      0      0      0
```

(c) Add commentary to the code that "matches" the Parva code fairly closely. Have a look at some of the other PVM code examples in the prac kit to see a "preferred" style of commentary, where the high level code appears as commentary on the low level code.

Task 3 - Build the assembler

In the directory `prac20\Assem` you will find Java or C# files that give you a minimal assembler and emulator for the PVM stack machine (described in Chapter 4.7). The Java files have the names (similar ones for C#):

`PVMAsm.java`	a simple assembler
`PVM.java`	an interpreter/emulator, fairly close to the one on page 37
`Assem.java`	a driver program

You can compile and make this assembler/interpreter system by issuing the batch command

```
MAKEASM
```

It takes as input a "code file" in the sort of format shown in the examples in section 4.5 and in the prac kit. Make up the minimal assembler/interpreter and, as a start, try to run these with the `ASM` batch command:

```
ASM  hello.pvm
ASM  lsmall.pvm
ASM  demorgan.pvm
```

Wow! Isn't Science wonderful? Try the interpretation with and without the trace option, and familiarize yourself with the trace output and how it helps you understand the action of the virtual machine. Then modify `DEMORGAN.PVM` to produce the truth table in the form illustrated earlier.

---- The (modified and suitably commented) DEMORGAN.PVM file must be submitted for assessment.

Task 4 - Coding the hard way

Time to do some creative work at last. Task 4 is to produce an equivalent program to the Parva one below (`FACT.PAV`), but written directly in the PVM stack-machine language (`FACT.PVM`). In other words, "hand compile" the Parva algorithm directly into the PVM machine language. You may find this a bit of a challenge, but it really is not too hard, just a little tedious, perhaps.

```
void main () {
// Print a table of factorial numbers 1! ... 20!
// Your names here!
  const limit = 20;
  int n = 1;
  while (n <= limit) {
    int f = 1;
    int i = n;
    while (i > 0) {
      f = f * i;
      i = i - 1;
    }
    write(n, "! = ", f, "\n");
    n = n + 1;
  }
} // main
```

Health warning: if you get the logic of your program badly wrong, it may load happily, but then go beserk when you try to interpret it. You may discover that the interpreter is not so "user friendly" as all the encouraging remarks in the book might have led you to believe interpreters all to be. Later we shall improve it quite a bit. (Of course, if your machine-code programs are correct you won't need to do so. As has often been said: "Any fool can write a translator for source programs that are 100% correct".)

The most tedious part of coding directly in PVM code is computing the destination addresses of the various branch instructions.

Hint: As a side effect of assembly, the ASM system writes a new file with a .COD extension showing what has been assembled and where in memory it has been stored. Study of a .COD listing will often give you a good idea of what the targets of branch instructions should be.

---- The (suitably commented) FACT.PVM file must be submitted for assessment.

Task 5 - Trapping overflow

Several of the remaining tasks in this prac require you to examine the machine emulator to learn how it really works, and to extend it to improve some opcodes and to add others.

In the prac kit you will discover two programs deliberately designed to cause chaos. `DIVZERO.PVM` bravely tries to divide by zero, and `MULTBIG.PVM` embarks on a continued multiplication that soon goes out of range. Try assembling and interpreting them to watch disaster happen.

Now we can surely do better than that! Modify the interpreter (`PVM.java` or `PVM.cs`) so that it will anticipate division by zero or multiplicative overflow, and change the program status accordingly, so that users will be told the errors of their ways and not left wondering what has happened.

You will have to be subtle about this - you have to detect that overflow is going to occur *before* things "go wrong", and you must be able to detect it for negative as well as positive overflow conditions.

Hint: After you edit any of the source code for the assembler you will have to issue the `MAKEASM` *command to recompile it, of course. It's easy to forget to do this and then wonder why nothing seems to have changed.*

Task 6 - Arrays

Start off by considering a further splendid exposition of the Parva programmer's art (`STUDENTS.PAV`)

```
void main () {
// Track students as they clock in and out of a practical
// P.D. Terry, Rhodes University, 2011

  const StudentsInClass = 100;
  bool[] atWork = new bool[StudentsInClass];
  int student = 0;
  while (student < StudentsInClass) {
    atWork[student] = false;
    student = student +1;
  }
  repeat {
    read("Student? (> 0 clocks in, < 0 clocks out, >= 100 terminates) ", student);
    if ((student > 0) && (student < StudentsInClass)) atWork[student] = true;
    if (student < 0)
      if (!atWork[-student]) write(student, " has not yet clocked in!\n");
      else atWork[-student] = false;
  } until (student >= StudentsInClass);
  write("The following students have still not clocked out\n");
  student = 0;
  while (student < StudentsInClass) {
    if (atWork[student]) write(student);
    student = student + 1;
  }
} // main
```

You can compile this (`PARVA STUDENTS.PAV`) at your leisure to make quite sure that it works.

Next, decide why it is a rather bad program. If you can't see why, try running it - run it anyway, and see if you

can break it. By "break" I mean "can you run it with some sort of data that allows it to work, and then run it with some data that produces meaningless results or even makes bomb out completely?" Go on to improve it - but keep the improvement quite simple and don't get carried away - remember my friend Dr Einstein's sage advice.

When you have done that, hand translate the improved `STUDENTS.PAV` ito PVM code (`STUDENTS.PVM`).

---- The (improved) STUDENTS.PAV and the (suitably commented) STUDENTS.PVM files must be submitted for assessment.

Task 7 - Your lecturer is quite a character

If the PVM and Parva could only handle characters as well as integers and Booleans, we could write a program like the exciting one below that reads a string of characters terminated with a period (full stop) and then writes it all in upper case SDRAWKCAB. (`SENTENCE.PAV`).

```
void main() {
// Read a piece of text terminated with a period and write it backwards in UPPER CASE.
// P.D. Terry, Rhodes University, 2011
  const
    limit = 256;                      // demonstration upper limit on sentence length
  char[]
    sentence = new char[limit];       // the number of times each appears
  int leng = 0;                       // read all characters
  repeat
    read(sentence[leng]); leng++;
  until (sentence[leng - 1] == '.');  // terminate input with a full stop
  while (leng > 0) {                  // write characters in reverse order
    leng--;
    write(upper(sentence[leng]));
  }
} // main
```

Not a problem for the assembler system. All we need to do is add appropriate opcodes to our virtual machine - for a start, `INPC` for reading a character and `PRNC` for writing a character - to open up exciting possibilities.

Hint: Adding "instructions" to the pseudo-machine is easy enough, but you must be careful to make sure you modify all the parts of the system that need to be modified. Before you begin, study the code in the definition of the stack machine carefully to see where and how the opcodes are defined, how they are mapped to the mnemonics, and in which switch/case statements they are used.

This example has implied the availability of a method for converting characters to uppercase, which is easily added to the PVM by introducing a special opcode. It also uses the infamous `++` and `--` operators, which can be handled by special opcodes that take less space (and should take less time to execute) than the tedious sequences needed for code corresponding directly to code like `n = n + 1`.

Extend the machine and the assembler still further with opcodes `CAP`, `INC` and `DEC`, and hand compile the program above to use them.

Hint: Be careful. Think ahead! Don't limit your `INC` *and* `DEC` *opcodes to cases where they can handle statements like* `X++` *only. In some programs you might want to have statements like* `List[N+6]++`.

Task 8 - Improving the opcode set still further

Section 4.9 of the text discusses the improvements that can be made to the system by adding new single-word opcodes like `LDC_0` and `LDA_0` in place of double-word opcodes for frequently encountered operations like `LDC 0` and `LDA 0`, and for using load and store opcodes like `LDL N` and `STL N` (and, equivalently, opcodes like `LDL_0` and `STL_0` for frequently encountered special cases).

Enhance your PVM by incorporating the following opcodes:

```
LDL N     STL N
LDA_0     LDA_1     LDA_2     LDA_3
LDL_0     LDL_1     LDL_2     LDL_3
STL_0     STL_1     STL_2     STL_3
LDC_M1    LDC_0     LDC_1     LDC_2     LDC_3
```

Hint: Several of the above are very similar to one another, but, once again, you must be careful to make sure you modify all the parts of the system that need to be modified.

Try out your system by developing "improved" versions of `STUDENTS.PVM` and `SENTENCE.PVM`, say `STUDENTS1.PVM` and `SENTENCE1.PVM`, that uses these new opcodes.

---- The final assembler/emulator must be submitted for assessment, as must STUDENTS1.PVM and SENTENCE1.PVM.

Task 9 - How do our systems perform?

In the kit you will find two versions of the infamous Sieve program written in PVM code. `S1.PVM` uses the original opcode set; `S2.PVM` uses the extended opcodes suggested in previous tasks.

You might think it is pretty obvious that using as many one-word opcodes as possible should make your programs smaller, faster, better. Carry out some experiments to see whether this is true and, if so, how big this effect is.

Your emulator will have had to assign enumeration values to the new opcodes. If you study the original source you will see that the original opcodes have been mapped onto the numbers 30 .. 62. You could map your new opcodes onto a set of numbers below 30, or above 62.

Run both versions of the sieve program through assemblers built to match various forms of the interpreters and obtain timings for a suitable upper limit (say 1000) and number of iterations (say 2000) for the combinations:

S1.PVM - using original opcodes + original interpreter supplied in the prac kit
S1.PVM - using original opcodes + new interpreter with new codes mapped "high"
S1.PVM - using original opcodes + new interpreter with new codes mapped "high"

S2.PVM - using extended opcodes + new interpreter with new codes mapped "high"
S2.PVM - using extended opcodes + new interpreter with new codes mapped "high"

Hint: The lab computers are very fast. You may have to alter those limits quite a bit to produce measurably distinct timings.

Comment on the results. Are they what you expect? If not, why not?

Hopefully by now you will have found that interpreters are quite easy to develop, but this prac should show you that they are not necessarily very "efficient". What changes could one make to improve the efficiency of the interpreter for the PVM still further? (If you are very keen you might try out some of your ideas, but I suppose that is wishful thinking. Sigh …)

Think carefully about all this. Please don't think you can write two lines of utter rubbish three minutes after you were supposed to hand the prac in, and try to bluff me that you know what is going on!

Have fun, and good luck.