TPC6 e Guião Laboratorial

Análise e resolução dos exercícios

Este documento apresenta uma proposta de análise do problema que vai para além da resolução propriamente dita do TPC6. Está estruturado em várias secções:

- Apresenta o código assembly resultante do exercício 1.a) com a respetiva anotação para ajudar a identificar a alocação de registos pelas variáveis feita pelo compilador, conforme pedido em 1.b)
- 2. Apresenta o código do main em C, conforme sugerido em 1.c) (i)
- 3. Faz a validação da utilização dos registos (objetivo do exercício 1.c)), mostrando quando os registos são inicializados (sugestão 1.c) (ii)), onde deverão ser inseridos os pontos de paragem (sugestão 1.c) (iii)) e confirmando depois a utilização dos registos conforme esperado (sugestão 1.c) (iv))
- 4. Apresenta a estrutura da stack frame da função, começando pela estimativa da sua estrutura (pedido em 1.e) (i)) e confirmando posteriormente a estrutura e conteúdos estimados (pedido em 1.e) (ii))
- 5. Discute com detalhe a estrutura de controlo do ciclo while (pedido em 1.f))
- **6.** Apresenta uma proposta de uma versão de código C com goto (pedido em **1.g)**)

Em anexo é ainda apresentada uma versão do código *assembly* pedido em **1.a)** mas com uma explicação mais detalhada de cada instrução.

1. Anotação do código em assembly do ciclo while

O código gerado na compilação de ciclos pode ser complicado de analisar, devido aos diferentes tipos de otimização do código do ciclo que o compilador poderá optar, para além da dificuldade em mapear variáveis do programa a registos do CPU. Para se adquirir alguma técnica, nada como começar com um ciclo relativamente simples.

Eis o código assembly que o comando gcc -S -O2 irá gerar na máquina remota, anotado manualmente, por vezes com alguma informação excessiva para um melhor esclarecimento:

```
while loop:
                                            ; salvaguarda na stack o frame pointer da função chamadora (%ebp)
          pushl %ebp
                    %esp, %ebp
                                            ; define o valor do frame pointer desta função: %esp -> %ebp
          movl
          movl
                    16(%ebp), %edx
                                             ; vai buscar o 3º argumento à stack e coloca-o em registo: n -> %edx
          testl %edx, %edx
                                             ; testa o valor de n (fazendo n < AND > n, que é = n), alterando só as flags;
                                             ; salvaguarda o conteúdo de %ebx na stack, pois vai precisar deste registo
          pushl %ebx
                                             ; vai buscar o 2º argumento à stack e coloca-o em registo: y -> %eax
          movl 12(%ebp), %eax
                     8(%ebp), %ebx
                                             ; vai buscar o 1º argumento à stack e coloca-o em registo: x -> %ebx
          movl
          jle
                                             ; salta para o fim do while se as flags indicarem que n \le 0 (less or equal)
                     .L3
                     %edx, %ecx
                                             ; cria uma variável temp em %ecx para calcular 16*n: temp= n
          movl
                     $4, %ecx
                                             ; um shift aritmét de 4 bits para a esq de temp é equiv a temp=n*24
          sall
                                            ; compara y com 16*n, i.e., altera as flags com a operação y-16*n
                     %ecx, %eax
          cmpl
          jge
                                             ; salta p/ fim do while se as flags indicarem que y ≥ 16*n (greater or equal)
                     .L3
                                            ; pad-to-align: acrescentar instruções no-ops para que o endereço da próxima
.p2align 2, , 3
                                                instrução seja múltiplo de 4; compilação com −02 pode ir mais longe
.L6:
                                             ; as instruções anteriores testaram a condição p/ executar o ciclo while
                     %edx, %ebx
                                            ; aqui começa o corpo do ciclo while com x += n
          addl
                                            ; y *= n
          imull
                     %edx, %eax
          decl
                    %edx
                                            ; n--
          subl
                     $16, %ecx
                                            ; p/ evitar calcular de novo 16*n o compilador optou por temp=temp-16
```

```
testl %edx, %edx
                                              ; estas 3 instruções repetem o cálculo da condição do ciclo while
           jle
                      .L3
                      %ecx, %eax
                                              ; na próxima instrução, em vez de saltar para fora do ciclo se y≥16*n
           cmpl
                      .L6
           jl
                                              ; salta para o início do ciclo while se ainda se mantiver y<16*n
.L3:
                                              ; coloca o valor a devolver (x) no registo convencionado para tal (%eax)
          movl
                     %ebx, %eax
                                              ; recupera o conteúdo de %ebx, que tinha sido salvaguardado no início
          popl
                    %ebx
                                              ; coloca o %esp a apontar para o mesmo endereço na stack que o frame
           leave
                                                   pointer (em %ebp) e recupera o frame pointer da função chamadora
                                              ; regressa à função chamadora, recuperando o IP que está no topo da pilha
           ret
```

A análise do modo como os argumentos são recuperados no código da função dá-nos uma boa pista de como o gcc usa os registos no cálculo de expressões de teste.

O valor de n*16, utilizado numa das condições do ciclo, é pré-calculado e guardado em ecx; em cada iteração do ciclo este valor é atualizado subtraindo 16, evitando-se assim a multiplicação n*16, uma vez que n é decrementado de 1 unidade.

Utilização dos Registos								
Variável	Registo	o Atribuição inicial						
х	%ebx	valor recebido do 1º arg (x)						
У	%eax	valor recebido do 2º arg (y)						
n	%edx	valor recebido do 3º arg (n)						
temp (=n*16)	%ecx	o valor recebido do 3º arg (n) e logo a seguir 16*n						

2. Código do main em C

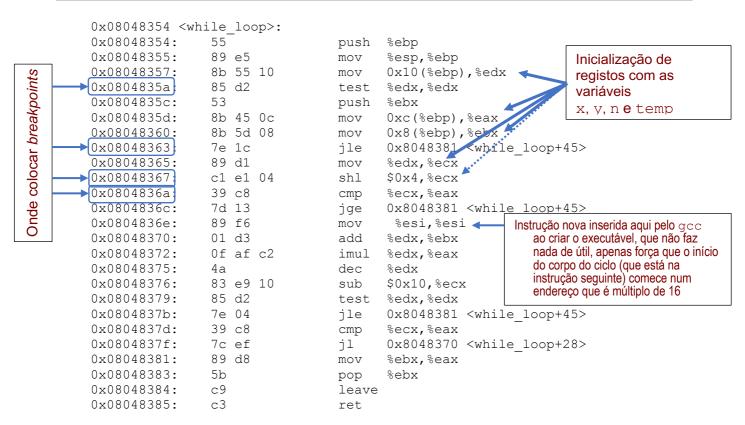
```
1 int main() {
2    int a=4, b=3, c=2, x;
3    x = while_loop (a, b, c);
4    printf ("%d\n", x);
5    return x;
6 }
```

3. Validação da utilização dos registos

Para confirmar esta utilização dos registos, procede-se conforme sugerido em 1.c) do guião:

- i. Coloca-se no mesmo ficheiro fonte onde estava o código do ciclo, um programa simples (main) que chame a função while_loop, passando como argumentos para a função os valores 4, 2 e 3, respetivamente.
- ii. Compila-se com os *switches* indicados e desmonta-se o executável com o comando objdump –d.
- iii. Analisando a parte do código desmontado com a função while_loop procura-se a 1ª instrução a seguir à inicialização de cada uma das variáveis e regista-se o endereço onde se encontram essas instruções (e ainda após temp passar a ser 16*n);
- iv. Invocando o *debugger* (gdb), introduzem-se pontos de paragem (*breakpoints*) nesses endereços, os quais irão depois permitir que a execução do código seja interrompida nesses endereços para se verificar o conteúdo dos registos.
- v. Com os comandos apropriados do gdb manda-se executar o programa (comando run) e ele vai sendo interrompido de cada vez que encontra um *breakpoint*.
- vi. Uma vez a execução do programa interrompida num *breakpoint*, pode-se então validar o conteúdo dos registos (por ex., com info reg), confirmando as estimativas que foram indicadas atrás.
- vii. Após cada *breakpoint* continuar a execução do código (comando cont) até terminar a execução do programa.

Código desmontado da função while loop com indicação dos endereços para os *breakpoints*:



Variável	Registo	Break1	Break2	Break3	Break4
х	%ebx	lixo	4	4	4
У	%eax	lixo	2	2	2
n	%edx	3	3	3	3
temp (=n*16)	%ecx	lixo	lixo	3	48

4. A stack frame da função

Pretende-se neste exercício que se preencham 3 tipos de informação:

- i. À esquerda do desenho da stack, os endereços do início de algumas "caixas".
- ii. No interior das "caixas" o valor numérico que lá deveria estar (pode ser em hexadecimal)
- Á direita das "caixas", uma explicação do valor que se encontra na respetiva "caixa".

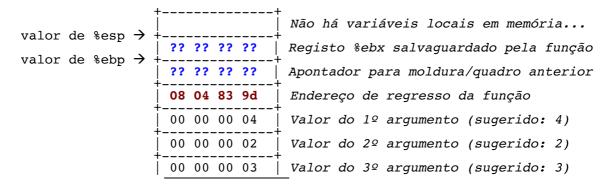
Cada "caixa" não é mais que um bloco de 32 bits armazenado em 4 células de 1 *byte* cada, em que o conteúdo da célula com menor endereço é o *byte* mais à direita de um valor de 32 bits (*little endian*).

A resolução completa implica uma análise detalhada do código gerado pelo gcc, do conteúdo de alguns registos e da localização em que deverá ser executado no PC, entre outros aspetos. Por exemplo, no gdb é possível visualizar o código da função main usando o comando disas main:

```
0x08048388 < main+0>:
                                 %ebp
                          push
0x08048389 < main+1>:
                                 %esp, %ebp
                         mov
0x0804838b < main+3>:
                         sub
                                 $0x8, %esp
0x0804838e < main+6>:
                                 $0xfffffff0, %esp
                         and
                         push
0x08048391 < main+9>:
                                 %eax
0x08048392 < main+10>:
                         push
                                 $0x3
                                 $0x2
0x08048394 < main+12>:
                         push
```

Neste caso, o valor do endereço de regresso que deverá estar na stack deverá ser o endereço da instrução na main imediatamente a seguir à invocação da função (após a instrução call), ou seja $0 \times 0804839 d$ (note que os endereços podem variar ligeiramente no seu caso, dependendo da forma como o código C está escrito).

Esta é a estimativa do conteúdo da *stack frame* associada à função while_loop, quando é invocada com os argumentos sugeridos e antes de se confirmar com o gdb:



O valor dos registos salvaguardados, incluindo o apontador para o quadro (*frame pointer*) da main na *stack*, i.e., tudo o que está com ?? na figura) pode ser obtido no gdb, parando a execução do código logo na 1ª instrução da função (coloque aí um *breakpoint*), e analisando o conteúdo desses registos (que ainda não foram colocados na *stack*).

Para confirmar os conteúdos destas 24 posições de memória na *stack*, coloca-se outro *breakpoint* no fim da parte de arranque da função, i.e., após a salvaguarda do registo <code>%ebx</code>; quando o programa parar aí pode-se então usar um dos comandos para examinar dados do *debugger* e visualizar o conteúdo das 24 células com início no topo da pilha (valor em <code>%esp</code>), quer *byte* a *byte* (dá para ver o funcionamento *little endian*), quer em 6 blocos de 4 *bytes*.

Sugestão de procedimento a seguir, passo a passo:

(i) Escrevendo disas while loop tem-se acesso novamente ao código da função:

```
0x08048354 < while loop+0>:
                                push
                                       %ebp
0x08048355 < while loop+1>:
                                mov
                                       %esp, %ebp
0x08048357 <while_loop+3>:
                                       0x10(%ebp), %edx
                                MOV
0x0804835a <while loop+6>:
                               test
                                       %edx, %edx
0x0804835c <while loop+8>:
                               push
                                       %ebx
0x0804835d <while loop+9>:
                                       0xc(%ebp), %eax
                                mov
0x08048360 <while_loop+12>:
                                mov
                                       0x8(%ebp),%ebx
0x08048363 < while loop+15>:
                               jle
                                      0x8048381 <while loop+45>
0x08048365 <while loop+17>:
                                      %edx, %ecx
                               mov
0x08048367 < while loop+19>:
                                       $0x4, %ecx
                                shl
0x0804836a <while loop+22>:
                               cmp
                                       %ecx, %eax
0x0804836c <while_loop+24>:
                                       0x8048381 <while loop+45>
                                jge
0x0804836e <while_loop+26>:
                                xchg %ax, %ax
0x08048370 <while loop+28>:
                                add
                                       %edx, %ebx
0x08048372 <while_loop+30>:
                                imul
                                       %edx, %eax
0x08048375 <while loop+33>:
                                dec
                                       %edx
0x08048376 < while loop+34>:
                                sub
                                       $0x10,%ecx
0x08048379 <while loop+37>:
                                test
                                       %edx, %edx
0x0804837b < while loop+39>:
                                jle
                                       0x8048381 <while loop+45>
```

(ii) Coloca-se um breakpoint após push %ebx (assinalado em cima):

```
break *0x804835d
```

- (iii) Executa-se o programa (run) e ele parará no *breakpoint*; visualiza-se o conteúdo dos registos escrevendo info reg; para responder integralmente a esta questão precisamos de saber o valor do apontador para o topo da pilha (registo %esp) e o apontador para o quadro da função, o *frame pointer* (registo %ebp).
 - (iv) Examinam-se 6 "palavras" na memória a partir do topo da pilha (endereço em %esp):

(v) Preenche-se agora a figura com estes valores, conforme pedido:

5. Estrutura do ciclo while

A expressão de teste é implementada em dois blocos. O primeiro bloco verifica se o ciclo deve ser executado na 1ª iteração:

```
testl %edx, %edx ; testa se n é 0; n <AND> n só é zero se n for zero...
jle .L3 ; salta para fim do ciclo se n <= 0
movl %edx, %ecx ; temp = n
sall $4, %ecx ; temp = n*16
cmpl %ecx, %eax ; compara y com n*16
jge .L3 ; salta para fim do ciclo se y >= 16
```

O segundo bloco, no fim do corpo do ciclo, verifica se este deve iterar:

```
decl %edx ; n--
subl $16, %ecx ; temp = temp-16 (=n*16)
testl %edx, %edx ; testa se n é 0
jle .L3 ; salta para o fim do ciclo se n <= 0
cmpl %ecx, %eax ; compara y com n*16
jl .L6 ; salta para o fim do ciclo se y < n*16</pre>
```

6. Versão do código com goto

Versão do tipo goto (em C) da função, com uma estrutura semelhante ao do código assembly (tal como foi feito para a série Fibonacci):

```
int while loop goto(int x, int y, int n)
2
3
     if (!((n > 0) \&\& (y < n*16))) goto done;
4
     loop:
5
        x += n;
6
        y *= n;
7
        n--;
8
        if ((n > 0) \&\& (y < n*16)) goto loop;
     done:
10
     return x;
11 }
```

Anexo

Alguns comentários prévios:

- a anotação de código é para nos ser útil quando analisamos o código assembly, não deve ser usado apenas para explicar o que faz a instrução em assembly
- a anotação dá-nos uma indicação sobre o que está de facto a acontecer, se é algo relacionado com uma instrução do código fonte (em C, por exemplo) ou se é para implementar uma estrutura de controlo, ou ainda se é para o arranque ou término duma função (e o que é que de facto está a acontecer)
- o verbo "push" em inglês quer dizer "empurrar" e não "puxar" (a tradução desta é "pull");
 assim, no código assembly é basicamente um "empurrar para o topo da pilha", ou simplesmente "empilhar"
- a instrução de test (dentro do ciclo while_loop) vai operar com 2 operandos para alterar flags; vai, de facto, realizar a operação n<AND>n e sabe-se que n<AND>n=n; por isso se diz muitas vezes que se está a testar o valor de n; e este teste é para se saber se n é zero, positivo, negativo...

A1. Explicação mais detalhada do código em assembly do ciclo while

Nova análise do código assembly da função while loop, anotado (a bold) e comentado:

```
pushl
       %ebp
                              ; salvaguarda na stack o frame pointer da função
                             chamadora (que está ainda em %ebp)
                              %ebp aponta nesta altura para quadro da função chamadora na stack, mas é
                             necessário definir um novo apontador para o quadro na stack (frame pointer) da
                             função chamada (while loop); guarda-se na stack o valor do frame pointer
                             que se encontra no registo %ebp, para o %ebp poder alojar o valor do frame
                             pointer da função chamada (while loop).
                              ; define o valor do frame pointer desta função:
movl
         %esp, %ebp
                              %esp -> %ebp
         16(%ebp), %edx; copia n (o 3° arg de while loop) da stack p/ %edx
movl
                             os argumentos da função while loop foram colocados na stack da função
                             chamadora antes da execução desta função; assumindo argumentos de 4 bytes de
                             tamanho em IA-32, o primeiro encontra-se sempre à distância de 8 células do
```

•

endereço no frame pointer (em ebp) da função chamada (while_loop) durante a sua execução; o 2º argumento encontra-se a 12 células de ebp, o 3º a 16, e assim por diante.

testl %edx, %edx

; testa n (faz & lógico de %edx consigo próprio) esta instrução implementa um & lógico entre os operandos, ativando as $flags \ ZF$, CF, OF, $e \ SF$ de acordo com o resultado; quando os 2 operandos são iguais, tudo se passa como se o teste fosse sobre um dos operandos, já que o & lógico de um valor consigo mesmo dá sempre o próprio valor; as instruções de salto consultam estas flags para a tomada de decisão; portanto esta instrução vai então testar se n é zero, positivo ou negativo.

pushl %ebx

garantir que os conteúdos dos registos %eax, %ecx, e %edx não são alterados quando se chama uma função é uma responsabilidade da função chamadora; se contiverem informação relevante para essa função ela terá de salvaguardar os seus conteúdos (na stack) antes de chamar a função while_loop de modo a que esta os possa usar, e depois terá de repor esses valores; a responsabilidade de salvaguardar os conteúdos dos registos %ebx, %edi, e %esi é da função chamada, caso ela necessite de usar esses registos (e neste caso a função necessita de usar 4 registos, os 3 que não tem de salvaguardar e mais este).

movl 12(%ebp), %eax

; copia y (o 2° arg de while_loop) da stack p/ %eax %eax toma o valor de y que foi passado como argumento.

movl 8(%ebp), %ebx

; copia x (o 1° arg de while_loop) da stack p/ %ebx a salvaguarda do conteúdo deste registo foi feita anteriormente para agora poder ser usado; %ebx toma o valor de x que foi passado como argumento.

ile .L3

; salta para depois do ciclo se $n \le zero$ a instrução consulta as flags previamente ativadas pela instrução testl; (de notar que as instruções de transferência de dados não alteram as flags, tais como os movle pushl que estão entre o testle este jle); neste caso, essa instrução em conjunto com esta de salto condicional estão a avaliar se o conteúdo de $extit{de} extit{de} extit{de}$

movl %edx, %ecx

; coloca o valor de n numa variável temporária o valor de n é copiado para esta nova variável temp (em %ecx) de modo a ficar depois com 16*n.

sall \$4, %ecx

; desloca 4 bits p/ a esquerda (shift left) no valor binário de temp, equivalente a calcular $n*2^4$ um deslocamento de 2 bits para a esquerda no nº binário 111_2 resulta em 11100_2 ; é equivalente a uma multiplicação de 2^2 pelo número em decimal; neste caso, está a ser feita uma multiplicação de 2^4 (16) pelo conteúdo de %ecx (que tem ainda o valor de n) que depois será útil para a avaliação de y<16*n; poder-se-ia usar imull \$16, %ecx, mas a instrução sall ocupa menos bytes em memória e shifts são mais rápidos de executar que multiplicações.

cmpl %ecx, %eax

; compara y com 16*n fazendo y-16*n e ativando as flags:

as *flags* serão alteradas consoante o resultado da diferença é zero, negativo ou positivo.

jge .L3

; salta para fora do ciclo se o resultado da operação anterior for \geq zero

esta instrução, em conjunto com o cmpl anterior, indica um salto para o fim do ciclo (em L3) caso y (em %eax) seja maior ou igual que temp (em %ecx, com 16*n); caso y>=16*n o controlo salta para L3, que se encontra no fim de while loop, sem executar qualquer iteração do ciclo.

.p2align 2,,3

; etiqueta vista na resolução do TPC anterior.

.L6:

; etiqueta que indica o início do ciclo while etiqueta usada como referência para saltos condicionais; indica o início do corpo do ciclo while do código C.

addl %edx, %ebx

; calcula x+=n

é feita a soma de x (em ebx) com n (em edx) guardando o resultado em x; é equivalente à primeira instrução do corpo do ciclo x+=n;

imull %edx, %eax

; calcula y*=n

multiplicação de n (em edx) com y (em edx) guardando o resultado em y; equivalente à expressão y=n no código C.

decl %edx

; calcula n--

decremento de n (em edx), equivalente a n--.

subl \$16, %ecx

; calcula o novo temp=16*(n-1) trocando a multiplicação pela subtração temp=16*n-16

esta instrução evita o uso da multiplicação novamente, já que o n foi decrementado e é necessário recalcular 16*n; assim, apenas se subtrai 16 a 16*n, que é equivalente a subtrair uma unidade a n e voltar a fazer 16*n com o novo n, quer usando um *shift* ou uma multiplicação; é mais rápido fazer uma subtração do que uma dessas duas instruções.

testl %edx, %edx

; testa n (faz & lógico de %edx consigo próprio) preparação para a avaliação de n>0, semelhante ao processo feito nas instruções antes do ciclo; no entanto, essa avaliação inicial foi feita para determinar se é possível entrar dentro do corpo do ciclo while; agora, esta avaliação verifica se é possível continuar dentro do ciclo na próxima iteração, sendo esta instrução repetida a cada iteração; poder-se-ia evitar a repetição destas avaliações com alguma reorganização do código, no entanto este é o padrão normalmente usado por esta versão do compilador da GNU.

jle .L3

; salta para fora do ciclo se $n \le zero$

salto para fora do ciclo (em L3) caso n (em edx) seja menor ou igual que zero, falhando a avaliação de n>0 do ciclo while.

cmpl %ecx, %eax

; compara y com 16*n fazendo y-16*n e ativando as flags

comparação de y com 16*n, já contemplando o facto de n ter sido decrementado; no entanto, a avaliação inicial idêntica a esta foi feita para determinar se era possível entrar dentro do corpo do ciclo while; agora, esta avaliação verifica se é possível continuar dentro do ciclo na próxima iteração (em conjunto com as instruções que avaliam se n>0), sendo esta instrução repetida a cada iteração; poder-se-ia evitar a repetição destas avaliações com alguma reorganização do código, no entanto este é o padrão normalmente usado por esta versão do compilador da GNU.

jl .L6

; salta para o início do ciclo se o resultado da operação anterior for < zero

salto para o início do ciclo (em ± 6) caso a condição $y < 1.6 \,^{\circ}\,n$ seja válida; caso esta condição não se verifique não é possível continuar dentro do corpo do ciclo, sendo a instrução seguinte a próxima a ser executada.

.L3:

; etiqueta que indica o início do bloco de código após o ciclo while

corresponde ao endereço da primeira instrução após o ciclo while; na função $while_loop$ corresponde à preparação para devolver o valor x (return x no código C), mas que envolve uma série de passos antes de finalizar a execução desta função.

movl %ebx, %eax

; coloca x (o valor a devolver) no registo %eax por convenção no IA-32, os valores devolvidos por funções são sempre colocados no registo %eax antes do seu término; poder-se-ia ter alocado logo o registo %eax à variável x, que é a que temos de devolver (return x no código C), eliminando esta instrução final.

popl %ebx

; recupera o valor original de %ebx como foi feita a salvaguarda do conteúdo deste registo na fase de arranque da função agora é necessário recuperar esse valor, de modo à função chamadora manter o conteúdo que tinha inicialmente atribuído a este registo, antes de chamar while loop.

leave

; recupera o frame pointer da função chamadora agora que a função while_loop acabou a sua execução é necessário recuperar a frame pointer da função chamadora, que foi salvaguardada na stack logo na 1ª instrução desta função; basta garantir que o %esp está a apontar para o mesmo sítio que %ebp e fazer um popl %ebp, que é o processo implementado pela instrução leave.

; regressa à função chamadora

é necessário que a próxima instrução a ser executada (cuja localização na memória se encontra sempre em <code>%eip</code>) seja a instrução na função chamadora imediatamente a seguir à instrução <code>call</code> que chamou a função <code>while_loop</code>; o endereço dessa instrução foi colocado na <code>stack</code> na execução da instrução <code>callwhile_loop</code> (encontra-se neste momento no topo da pilha), e tem agora que ser colocado no <code>%eip</code>; esta instrução seria equivalente a um <code>popl %eip</code>.

ret