

La Pila en los Procesadores IA-32 e Intel©64

Alejandro Furfaro Ilustraciones de David Gonzalez Marquez (tnx a lot)

13 de abril de 2020

Agenda

- Funcionamiento Básico
- 2 Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- 3 Convención de llamadas C
 - Generalidades
 - Modo 32 bits
- Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados
 - **B**ibliografía



Funcionamiento básico

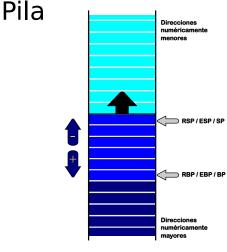


Figura: Funcionamiento Básico de la Pila.©David

- La pila (stack) es un área de memoria contigua, referenciada por un segmento cuyo selector está siempre en el registro SS del procesador.
- El tamaño de este segmento en el modo IA-32, puede llegar hasta 4
 Gbytes de memoria, en especial cuando el sistema operativo utiliza el modelo de segmentación Flat (como veremos en clases subsiguientes).
- El segmento se recorre mediante un registro de propósito general, denominado habitualmente en forma genérica stack pointer, y que en estos procesadores según el modo de trabajo es el registro SP, ESP, o RSP (16, 32, o 64 bits respectivamente).

Funcionamiento básico

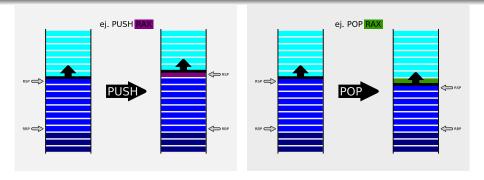


Figura: PUSH y POP.@David

- Para guardar un dato en el stack el procesador tiene la instrucción PUSH, y para retirarlo, la instrucción POP.
- Cada vez que ejecuta PUSH, el procesador decrementa el stack pointer (SP, ESP, o RSP) y luego escribe el dato en el stack, en la dirección apuntada por el registro de segmento SS, y el stack pointer correspondiente al modo de trabajo.
- Cada vez que ejecuta un POP, el procesador lee el ítem apuntado por el par SS : stack pointer, y luego incrementa éste último registro.□ ➤ < 🗗 ➤ < 🛢 ➤ <

Primeras conclusiones

El stack es un segmento expand down, ya que a medida que lo utilizamos (PUSH) su registro de desplazamiento se decrementa apuntando a las direcciones mas bajas (down) de memoria, es decir a aquellas numéricamente menores.

Cuando se utiliza el stack

Las operaciones de pila se pueden realizar en cualquier momento, pero hablando mas generalmente, podemos afirmar que la pila se usa cuando:

- Cuando llamamos a una subrutina desde un programa en Assembler, mediante la instrucción CALL.
- Cuando el hardware mediante la interfaz adecuada envía una Interrupción al Procesador.
- Cuando desde una aplicación, ejecutamos una Interrupción de software mediante la instrucción INT type.
- Cuando desde un lenguaje como el C se invoca a una función cualquiera.

Alineación del Stack

- El stack pointer debe apuntar a direcciones de memoria alineadas de acuerdo con su ancho de bits.
- Por ejemplo, el ESP (32 bits) debe estar alineado a double words.
- Al definir un stack en memoria se debe cuidar el detalle de la alineación.
- El tamaño de cada elemento de la pila se corresponde con el atributo de tamaño del segmento (16, 32, o 64 bits), es decir, con el modo de trabajo en el que está el procesador, y no con el del operando en sí.
- Ej: PUSH AL, consume 16, 32, o 64 bits dependiendo del tamaño del segmento. Nunca consume 8 bits.
- El valor en que se decrementa el Stack Pointer se corresponde con el tamaño del segmento (2, 4, u 8 bytes).

Alineación del Stack

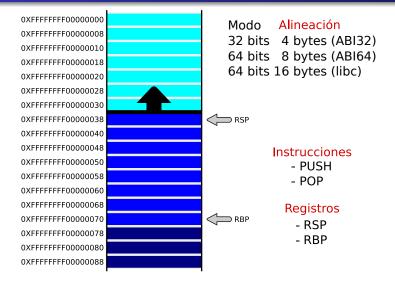


Figura: Alineación según el modo de trabajo. © David

7 / 54

- ¶ Funcionamiento Básico
- 2 Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- Convención de llamadas C
 - Generalidades
 - Modo 32 bits
- Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados





Como en un debuger :)

- Ejecutamos la primer instrucción
- Lee el port de E/S
- Y luego.....

```
0x0001FE00

0x0001FE04

0x0001FE08

0x0001FE0C

0x0001FFC
```

Estamos a punto de ejecutar CALL

- ...ejecutamos la instrucción Call.
- La misma está almacenada a partir de la dirección de memoria contenida por EIP.
- El ESP apunta a la base de la pila.

```
0x0001FE00
0x0001FE04
0x0001FE08
0x0001FE0C
0x0001FE10
```

```
% define mask 0xfff0
main:
...
mov dx,0x300
in ax,dx ; lee port

call setmask; llama a subrutina para aplicar una mascara
...
...
setmask:
and ax, mask ; aplica la mascara
ret : retorna
```

CALL por dentro...

- En primer lugar el procesador apunta con EIP a la siguiente instrucción.
- Un CALL near se compone de 1 byte de código de operación y cuatro bytes para la dirección efectiva (offset), ya que estamos en 32 bits.
- Por eso el EIP apunta 5 bytes mas adelante, ya que allí comienza la siguiente instrucción del CALL.

```
0x0001FE00
0x0001FE04
0x0001FE08
0x0001FE0C
0x0001FFC

0x0001FFFC

ESP
```

CALL por dentro...

- El procesador decrementa ESP y guarda el valor de EIP.
- Así resguarda su dirección de retorno a la instrucción siguiente a CALL.
- Para saber a donde debe saltar saca de la instrucción CALL la dirección efectiva de la subrutina setmask.
- En nuestro caso 0x0007C44F.

% define mask

```
main:
...
mov dx,0x300
in ax,dx ; lee port
call setmask ; llama a subrutina para aplicar una mascara
...
setmask:
and ax, mask ; aplica la mascara
ret ; retorna
```

 $0 \times fff0$

```
0x0001FE00

0x0001FE04

0x0001FE08

0x0001FE0C

0x0001FF18

0x0001FFFC

0x0001FFFC
```

Resultado del CALL

- Como resultado el valor de EIP, es reemplazado por la dirección efectiva de la subrutina setmask.
- Y sin mas.... el procesador está buscando la primer instrucción de la subrutina setmask, en este caso, la operación and.

```
0x0001FE00
0x0001FE04
0x0001FE0C
0x0001FF10

0x0001FFF8
0x00007C226
0x0001FFF8
0x0007C226
0x0001FFF8
0x00007C226
```

```
% define mask 0xfff0
main:
...
mov dx,0x300
in ax,dx ; lee port
call setmask ; llama a subrutina para aplicar una mascara
...
...
```

setmask:

```
and ax,mask ;aplica la mascara
```

```
ret ; retorna
```



Volver.....

- Esta subrutina es trivial a los efectos del ejemplo.
- Para volver (sin la frente marchita)...
- Es necesario retornar

```
0x0001FE00
                           EIP 0x0007C44F
0x0001FE04
0x0001FE08
0x0001FE0C
0x0001FE10
0x0001FFF8 0x0007C226
0x0001FFFC
                                     ESP
```

```
% define mask
                0 \times fff0
main:
  mov dx.0x300
  in ax, dx; lee port
  call setmask ; llama a subrutina para aplicar una mascara
  . . .
setmask:
                ; aplica la mascara
  and ax, mask
```

ret :retorna

Volviendo.....

- La ejecución de **ret** consiste en recuperar de la pila la dirección de retorno.
- Esa dirección se debe cargar en EIP
- Una vez hecho

```
0x0001FE00
                           EIP 0x0007C226
0×0001FF04
0×0001FF08
0x0001FE0C
0x0001FE10
0x0001FFF8 0x0007C226
0x0001FFFC
                                     ESP
```

```
% define mask
              0 \times fff0
main:
       dx,0x300
  mov
  in
       ax.dx ; lee port
  call setmask ; llama a subrutina para aplicar una mascara
  . . .
setmask:
  and ax, mask ; aplica la mascara
  ret
                 ; retorna
```

Volvimos!

- Finalizada la ejecución de ret estamos otra vez en el código llamador.
- Pero en la instrucción siguiente a CALL

```
0x0001FE00
0x0001FE04
0x0001FE0C
0x0001FF10
0x0001FFFC
```

```
% define mask 0xfff0
main:
...
mov dx,0x300
in ax,dx ; lee port
call setmask ; llama a subrutina para aplicar una mascara
...
```

```
setmask:
```

and ax, mask ; aplica la mascara

ret ; retorna



- ¶ Funcionamiento Básico
- 2 Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- Convención de llamadas C
 - Generalidades
 - Modo 32 bits
- Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados
- Bibliografía



- Ejecutamos la primer instrucción
- Lee el port de E/S
- Y luego.....

```
0x0001FE00
0×0001FF04
0x0001FE08
0x0001FF0C
0x0001FE10
0x0001FFFC
                                         ESP
```

```
% define mask 0xfff0
section code1
main:
       dx,0x300
  mov
in ax,dx ;lee port
  call code2:setmask ; llama a aplicar m scara
section code2
setmask:
  and
       ax , mask
                          ; aplica la m scara
  retf
                          ; retorna
```

Por lo tanto importa el valor de CS...

- Nuevamente nos paramos en el CALL
- Pero ahora necesitamos memorizar EIP, y también CS
- Ya que al estar el destino en otro segmento CS se modificará

```
0x0001FE00
                                0x0007C221
0×0001FF04
                           CS
                                         code1
0x0001FE08
0x0001FF0C
0x0001FE10
0x0001FFFC
                                     ESP
```

```
% define mask
                  0 \times fff0
section code1
main:
        dx.0x300
  mov
        ax, dx ; lee port
  in
call code2:setmask :llama a subrutina para aplicar una mascara
section code2
setmask:
  and
         ax . mask
                             ; aplica la m scara
```

: retorna

retf

- Ahora la instrucción mide 7 bytes ya que se agrega el segmento
- Por lo tanto el EIP se incrementa 7 lugares
- Y se memoriza en la pila la dirección FAR.

```
% define mask 0xfff0
section code1
main:
...
mov dx,0x300
in ax,dx ; lee port
call code2:setmask ; llama a subrutina para aplicar una mascara
...
section code2
setmask:
and ax,mask ; aplica la m scara
```

: retorna

retf

- En primer lugar guarda en la pila, el valor del segmento al cual debe retornar.
- Siempre antes de almacenar nada en la pila, debe antes decrementar el valor del ESP.

```
% define mask 0 \times fff0 section code1 main: ... mov dx,0 \times 300 in ax,dx; lee port call code2:setmask; llama a subrutina para aplicar una mascara ... section code2 setmask:
```

; aplica la m scara

: retorna

and

retf

ax, mask

- Luego del valor del segmento guarda en la pila, el valor de EIP al cual debe retornar, y que lo llevará a buscar la siguiente instrucción al CALL.
- Decrementará nuevamente el valor del ESP, antes de almacenar.

```
% define mask  0xfff0
section code1
main:
    ...
    mov dx,0x300
    in ax,dx ; lee port
    call code2:setmask ; llama a subrutina para aplicar una mascara
    ...
section code2
setmask:
    and ax,mask ; aplica la m scara
```

: retorna

retf

- La dirección de la rutina setmask, ahora es code2:offset.
- Como comienza justo al inicio del segmento, su offset es 0x00000000.

```
0x0001FE00
0x0001FE04
0x0001FE05
0x0001FE0C
0x0001FE10

0x0001FFFC

0x0001FFFC

EIP 0x00000000
CS code2
```

```
% define mask  0xfff0
section code1
main:
    ...
    mov dx,0x300
    in ax,dx ; lee port
    call code2:setmask ; llama a subrutina para aplicar una mascara
    ...
section code2
setmask:
and ax,mask ;aplica la máscara
```

: retorna

retf

Retornando de un Call Far

- Para volver de un call far hay que sacar de la pila no solo el offset sino también el segmento.
- Entonces no sirve la misma instrucción que se usa para volver de una rutina Near.

```
% define mask 0xfff0
section code1
main:
...
mov dx,0x300
in ax,dx ; lee port
call code2:setmask ; llama a subrutina para aplicar una mascara
...
section code2
setmask:
and ax,mask ; aplica la m scara
```

retf ;retorna

Retornando de un Call Far

Recupera la dirección efectiva

retf

Luego decrementa el Stack Pointer

```
0x0001FE00
                               0x0007C228
0×0001FF04
                          CS
                                        code2
0x0001FE08
0x0001FF0C
0x0001FE10
           0x0007C228
                  code1
0x0001FFFC
                                    ESP
```

```
% define mask 0 \times fff0
section code1
main:
  mov dx.0x300
  in ax,0x300; lee port
  call code2:setmask; llama a subrutina para aplicar una mascara
section code2
setmask:
  and ax, mask
                         ; aplica la m scara
```

: retorna

Retornando de un Call Far

• Recupera el valor del segmento

ax, mask

retf : retorna

- Luego decrementa el Stack Pointer
- ...y volvió...

and

```
0x0001FE04
0x0001FE04
0x0001FE08
0x0001FE0C
0x0001FE10
0x0007C228
code1
```

```
% define mask 0xfff0
section code1
main:
...
mov dx,0x300
in ax,dx ; lee port
call code2:setmask ; llama a subrutina para aplicar una mascara
...
section code2
setmask:
```

; aplica la m scara

- Funcionamiento Básico
- Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- Convención de llamadas C
 - Generalidades
 - Modo 32 bits
- 4 Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados





- Ejecutamos una instrucción cualquiera
- y en el medio de esa instrucción se produce una interrupción

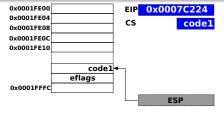
```
0x0001FE00
                                0x0007C221
0×0001FF04
                           CS
                                         code1
0x0001FE08
0x0001FF0C
0x0001FE10
0x0001FFFC
                                      ESP
```

```
section code
main:
next:
test [var],1 ;chequea bit 0 de variable
  inz
         next
section kernel
handler_int:
         al, port; lee port de E/S
  iret
```

: retorna

- Es necesario guardar además de la dirección de retorno, el estado del procesador.
- De otro modo si al final de la interrupción alguna instrucción modifica un flag, el estado de la máquina se altera y le vuelve al programa modificado.
- Esto puede tener resultados impredecibles si al retorno hay que usar el flag que cambió.

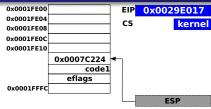
- La dirección de retorno es far.
- Especialmente en sistemas multitasking donde cada proceso tiene una pila de kernel diferente.
- Así que luego de los flags se guarda el segmento de código.



- Se resguarda finalmente la dirección efectiva
- Notar que es la de la instrucción siguiente a la de la interrupción

```
0x0001FE00
0x0001FE04
0x0001FE0C
0x0001FE0C
0x0001FE0C
0x0001FFC
0x0007C224
Code1
eflags
0x0001FFFC
ESP
```

 Los nuevos valores de segmento y desplazamiento que debe cargar en CS:EIP, los obtiene del vector de interrupciones en modo real, o de la Tabla de descriptores de interrupción en modo protegido, o en el modo 64 bits.



```
section code
main:
    ...
next:
    test [var],1 ; chequea bit 0 de variable
    jnz next
    ...
section kernel
handler_int:
in al,port; lee port de E/S
```

: retorna

iret

¿Como se vuelve de una Interrupción?

- Recuperando además de la dirección de retorno, los flags
- Por lo tanto necesitamos otra instrucción particular de retorno...
- ... iret...

```
0x0001FE00

0x0001FE04

0x0001FE08

0x0001FE0C

0x0001FE10

0x0007C224

code1

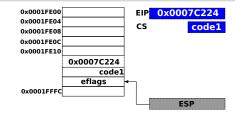
eflags

0x0001FFFC
```

```
section code
main:
    ...
next:
    test [var],1 ; chequea bit 0 de variable
    jnz    next
    ...
section kernel
handler_int:
    in al, port ; lee port de E/S
iret ; retorna
```

Volviendo...

Volviendo...



Volviendo...

```
section code
main:
    ...
next:
    test [var],1 ; chequea bit 0 de variable

jnz next
    ...
section kernel
handler_int:
    in al, port ; lee port de E/S
    iret ; retorna
```

- Funcionamiento Básico
- 2 Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- 3 Convención de llamadas C
 - Generalidades
 - Modo 32 bits
- 4 Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados





Llamadas a función

• En general en el lenguaje C una función se invoca de la siguiente forma

```
type function (arg1, arg2, ..., argn);
```

- *type*, es siempre un tipo de dato básico (int, char, float, double), o un puntero, o void en caso en que no devuelva nada.
- El manejo de la interfaz entre el programa invocante y la función llamada la resuelve el compilador, de una manera perfectamente definida.
- Sin embargo los pormenores son diferentes según se trabaje en 32 bits o en 64 bits



- Funcionamiento Básico
- 2 Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- Convención de llamadas C
 - Generalidades
 - Modo 32 bits
 - 4 Interacción C-ASN
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados



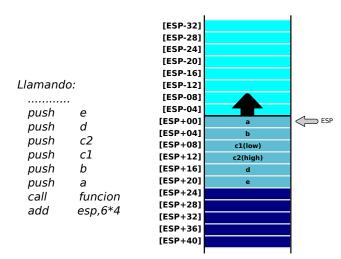


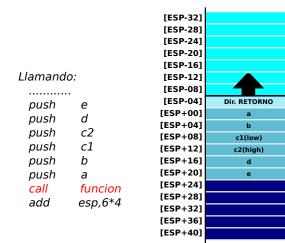
Stack Frame

• El compilador traduce el llamado en el siguiente código assembler:

```
push argn
...
push arg2
push arg1
call function ; o sea un CALL Near!!
```

- Los argumentos se apilan desde la derecha hacia la izquierda.
- Una vez dentro de la subrutina "function", el compilador agrega el siguiente código

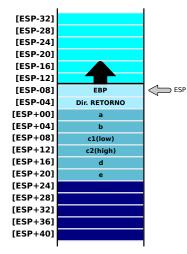




int f1(int a, float b, double c, int* d, double* e)

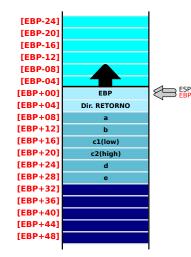
funcion:

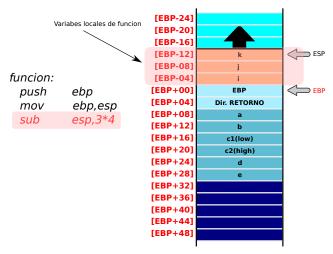
push ebp mov ebp,esp sub esp,3*4

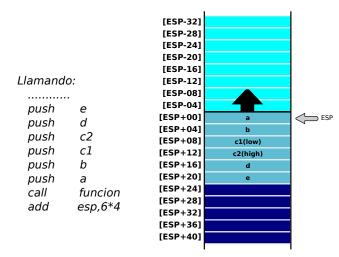


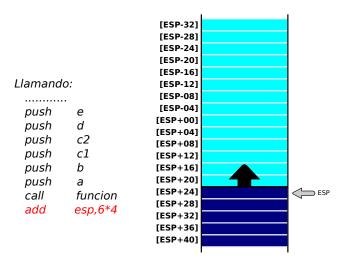
int f1(int a, float b, double c, int* d, double* e)

funcion:
push ebp
mov ebp,esp
sub esp,3*4









Llamar a funciones ASM desde C

Hacemos uso de la cláusula extern en C y global en ASM:

```
funcion.asm

global fun
section .text
fun:
...
ret
```

```
programa.c

extern int fun(int, int);
int main(){
    ...
    fun(44,3);
    ...
}
```

Primero ensamblamos y compilamos el código en ASM para luego linkearlo con el código en C:

- nasm -f elf64 funcion.asm -o funcion.c
- gcc -o ejec programa.c funcion.o

Llamar a funciones ASM desde C

Hacemos uso de la cláusula extern en C y global en ASM:

```
funcion.asm

global fun
section .text
fun:
...
ret
```

```
programa.c

extern int fun(int, int);
int main(){
    ...
    fun(44,3);
    ...
}
```

Primero ensamblamos y compilamos el código en ASM para luego linkearlo con el código en C:

- nasm -f elf64 funcion.asm -o funcion.o
- gcc -o ejec programa.c funcion.o

Llamar funciones C desde ASM

Usamos sólo la cláusula extern en ASM:

```
main.asm
global main
extern fun
section .text
main:
...
call fun
...
ret
```

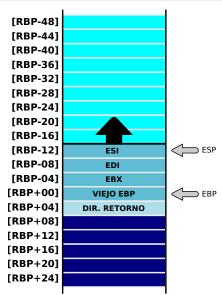
```
funcion.c
int fun(int a, int b){
    ...
    ...
    int res= a+b;
    ...
    return res;
}
```

Compilamos ambos programas y generamos el ejecutable de ASM:

- nasm -f elf64 main.asm -o main.o
- gcc -c -m64 funcion.c -o funcion.o
- gcc -o ejec -m64 main.o funcion.o

```
fun:

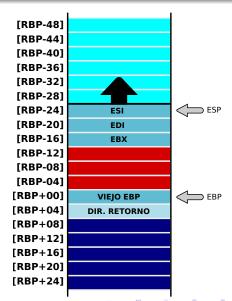
push ebp
mov ebp, esp
sub esp, 12
push ebx
push edi
push esi
... más código ...
pop esi
pop edi
pop ebx
add esp, 12
pop ebp
ret
```



Esquema de un stack frame en 32b con variables locales

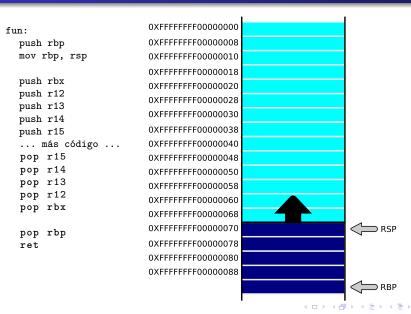
```
fun:
  push ebp
  mov ebp, esp
  sub esp, 12
  push ebx
  push edi
  push esi
  ... más código ...
  pop esi
  pop edi
  pop ebx
  add esp, 12
  pop ebp
  ret
```

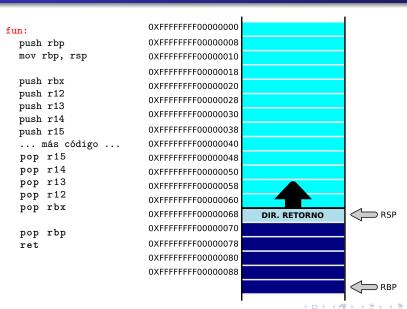
Nota: El espacio a dejar depende de la alineación de la pila y las variables locales de la función

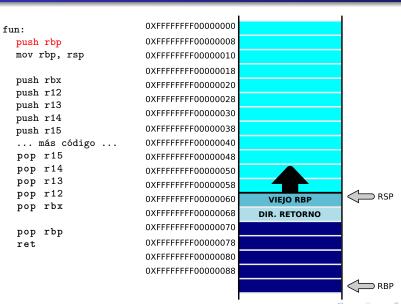


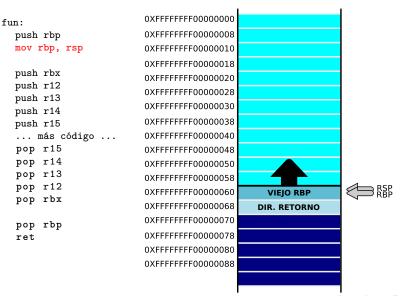
- Funcionamiento Básico
- 2 Ejemplos de uso de pila
 - ¿Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- Convención de llamadas C
 - Generalidades
 - Modo 32 bits
- 4 Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados
- Bibliografía



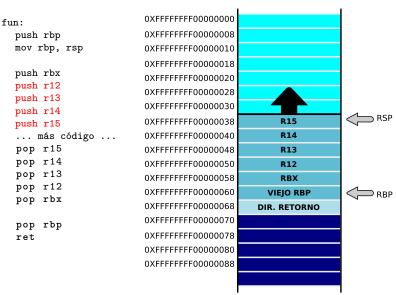


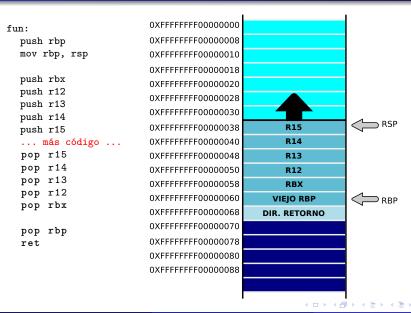






0XFFFFFFF00000000 fun: push rbp 0XFFFFFFF00000008 mov rbp, rsp 0XFFFFFFF00000010 0XFFFFFFF00000018 push rbx 0XFFFFFFF00000020 push r12 0XFFFFFFF00000028 push r13 0XFFFFFFF00000030 push r14 0XFFFFFFF00000038 push r15 ... más código ... 0XFFFFFFF00000040 pop r15 0XFFFFFFF00000048 pop r14 0XFFFFFFF00000050 pop r13 RSP 0XFFFFFFF00000058 **RBX** pop r12 0XFFFFFFF00000060 VIEJO RBP pop rbx 0XFFFFFFF00000068 **DIR. RETORNO** 0XFFFFFFF00000070 pop rbp 0XFFFFFFF00000078 ret. 0XFFFFFFF00000080 0XFFFFFFF00000088





Variables Locales

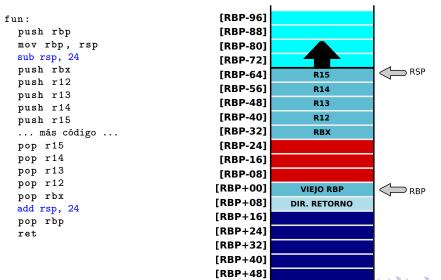
- Una vez dentro de la función invocada en un programa C utilizamos por lo general variables locales. Solo tienen validez dentro de la función en la que se las declara.
- Una vez finalizada esta función no existen mas.
- Se crean frames en el stack para albergar dichas variables. Simplemente moviendo esp hacia el fondo del stack, es decir:

```
sub rsp, n
```

• Siendo n la cantidad de bytes a reservar para variables

Esquema de un stack frame en 64b con variables locales

Dejamos un espacio en la base de la pila para almacenar las variables



- - ; Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- - Generalidades
 - Modo 32 bits
- Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados



System V Application Binary Interface

 Conocida como ABI, establece el pasaje de argumentos desde una función llamante a una función llamada, y como se retornan los resultados.

En 64bits:



- Los registros se usan en orden dependiendo del tipo
- Los registros de enteros guardan parámetros de tipo Entero o Puntero
- Los registros XMM guardan parámetros de tipo Flotante
- Si no hay más registros disponibles se usa la PILA
- Los parámetros en la PILA deben quedar ordenados desde la dirección más baja a la más alta.

Figura: Convención C para 64 bits (©David)

Ejemplo sencillo

En 64 bits:

int f1(int a, float b, double c, int* d, double* e)

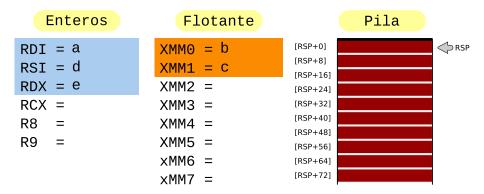


Figura: Resolución de llamada - Ej1 (©David)

Ejemplo sencillo

En 64 bits:

Enteros Flotante Pila a12 [RSP+0] <⊅RSP RDT = a1XMM0 = a2a13 [RSP+8] RST = a4xmm1 = a3[RSP+16] a14 RDX = a7XMM2 = a5[RSP+24] XMM3 = a6RCX = a8[RSP+32] [RSP+40] R8 = a9XMM4 = a10[RSP+48] XMM5 = a15R9 = a11[RSP+56] xMM6 =[RSP+64] [RSP+72] $\times MM7 =$

- - ; Como funciona un llamado Near?
 - ¿Como funciona un llamado Far?
 - Interrupciones
- - Generalidades
 - Modo 32 bits
- Interacción C-ASM
 - Modo 64 Bits
 - Modo 64 bits
 - Resultados





Resultados

- En 32 bits los resultados enteros y punteros se devuelven por eax.
- En 64 bits los enteros y punteros se devuelven en RAX, y si son floats o doubles, en XMM0 y/o XMM1.

 Intel[®] 64 and IA-32 Architectures Software Developer's Manual: Vol I. Basic Architecture.
 Capítulo 6

- System V Application Binary Interface. AMD64 Architecture Processor Supplement. Draft Version 0.99.5. Edited by Michael Matz, Jan Hubička, Andreas Jaeger, Mark Mitchell. September 3, 2010.
- System V Application Binary Interface. Intel386TMArchitecture Processor Supplement. Fourth Edition