ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ МОСКОВСКИЙ ФИЗИКО-ТЕХНИЧЕСКИЙ ИНСТИТУТ (НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ) ФАКУЛЬТЕТ АЭРОКОСМИЧЕСКИХ ТЕХНОЛОГИЙ

Лабораторная 3 **После пузырька можно уже всё**

Пункт 1: а) Уже из самого первого примера на С с простейшей функуцией без аргументов и возвращаемых значений видно, что в 32-х битной версии много дополнительных действий, а именно пуши регистров и действия с ними.

```
unc:
LFB0:
          .cfi startprod
         .cfi_def_cfa_offset 8
.cfi_offs=:
                   fset 5, -8
%esp, %ebp
                   %ehx
                   $4, %esp
                    $_GLOBAL_OFFSET_TABLE_, %eax
                   $12, %esp
.LC0@GOTOF
                        @GOTOFF(%eax), %edx
                   %edx
                   puts@PLT
                    -4(%ebp), %ebx
                   main, @function
         endbr32
                   Fset 5, -8
%esp, %ebp
                   $-16, %esp
_x86.get_pc_thunk.ax
$_GLOBAL_OFFSET_TABLE_, %eax
                   $0, %eax
```

(а) на С 32-х битная система

```
#include <stdio.h>

void func() {
          printf("a b c\n");
}

int main() {
          func();
          return 0;
}
```

(а) Код программы на С

```
LFB0:
        .cfi_startproc
        endbr64
        pushq
                 %rbp
                %rsp, %rbp
        movq
                 .LC0(%rip), %rdi
        leaq
        call
                 puts@PLT
                 %rbp
        popq
        ret
 LFE0:
                 main
                 main, @function
main:
LFB1:
        endbr64
                 %rbp
        pushq
                 %rsp, %rbp
        movq
                 $0, %eax
        movl
        call
                 func
                 $0,
                     %eax
        movl
                 %rbp
        popq
        ret
```

(b) на C 64-х битная система

```
#include <iostream>
void func() {
        std::cout << "a b c" << std::endl;
}
int main() {
        func();
        return 0;
}</pre>
```

(b) Код программы на C++

```
.cfi_startproc
endbr64
pushq %rbp
.cfi_def_cfa_offset 16
.cfi_offset 6, -16
movq %rsp, %rbp
.cfi_def_cfa_register 6
leaq .LC(%rip), %rsi
leaq _ZSt4cout(%rip), %rdi
call _ZSt1sISt1lchar_traitsIcEERSt13basic_ostreamICT_ESS_PKc@PLT
movq %rax, %rdx
movq %rax, %rdx
movq %rax, %rsi
movq %rax, %rsi
movq %rax, %rdi
call _ZNSolsEPFRSoS_E@PLT
nop
popq %rbp
.cfi_def_cfa 7, 8
ret
.cfi_endproc
.LFE1522:
.size _Z4funcv, .-_Z4funcv
.globl main
.type main, @function
main:
.LFB1523:
.cfi_def_cfa_offset 16
.cfi_offset 6, -16
movq %rsp, %rbp
.cfi_def_cfa_register 6
call _Z4funcv
movl $0, %eax
popq %rbp
.cfi_def_cfa_register 6
call _Z4funcv
movl $0, %eax
popq %rbp
.cfi_def_cfa 7, 8
ret
.cfi_endproc
```

Рис. 3: Листинг программы на С++ 64-х битной системы

На C++ ситуация примерно такая же, в 32-х битной версии присутствуют какие-то дополнительные пуши регистров, в 64-х битной версии такого почти нету, за исключением пушей, связанных, как мне кажется, с std::cout и std::endl.

- b) Для функции без аргумеетов, но с возвращаемым значением типа int дела обстоят практически одинаково что для C++, что для C, независимо от разрядности системы. Везде под возвращаемый int выделяется 16 байт (почему-то), только в C++ немного хуже дела обстоят с названиями (с человеческой точки зрения).
- с) Далее, функция, возвращающая int, с одним аргументом того же типа. Здесь видны различия между 64-х и 32-х битными системами. В 32-х битной аргумент, который потом передаётся в функцию, перед вызовом функции кладётся в стек, и потом лежит в стеке перед функцией, причём если внутри функции есть арифметические операции с аргументом, то они производятся со значением, лежащим в стеке перед функцией. В 64-х битной системе аргумент запоминается в отдельный регистр (у меня был edi), после вызова функции значение с этого регистра копируется в стек, и операции проводятся со значением, лежащим в стеке после функции. Разницы между С и C++ опять же почти нету. Под результат функции (одно число int) и там и там выделяется 16 байт.
- d) Теперь функция, которая возвращает int и принимает два аргумента типа int. Всё почти так же как и в предыдущем случае. В 32-х битном варианте резервируется под возвращаемое значение место в стеке (опять 16 байтов на один int), затем в стек пушатся аргументы функции, затем функция, потом производятся все арифметические операции, значения для которых передаются в регистры из стека (оттуда куда до этого пушились аргументы), после завершения работы функции на вершине стека остаётся возвращаемое значение. В 64-х битной системе, как и в предыдущем случае, используются дополнительные регистры (теперь edi и esi) для хранения значений аргументов. Значения их этих регистров копируются в стек функции, там же и проделываются операции с числами. Результат работы программы после её завершения также остаётся на вершине стека, на один int выделяется 16 байт. Различия между С и C++

всё также только в названиях переменных.

Пункт 2: а) Разберёмся с локальными переменными. Сначала посмотрим листинг с одной локальной переменной.

```
endbr32
pushl %ebp
        %esp, %ebp
        %ebx
         set 3, -12
x86.get pc thunk.bx
        $ GLOBAL OFFSET_TABLE_, %ebx
        $1, 8(%ebp)
        $12, %esp
_ZStL8__ioinit@GOTOFF(%ebx), %eax
        —
%eax
         _ZNSt8ios_base4InitC1Ev@PLT
        $16, %esp
$4, %esp
          _dso_handle@GOTOFF(%ebx), %eax
        %eax
         _ZStL8__ioinit@GOTOFF(%ebx), %eax
        %eax
         _ZNSt8ios_base4InitD1Ev@GOT(%ebx), %eax
        %eax
          _cxa_atexit@PLT
        -4(%ebp), %ebx
         GLOBAL__sub_I_main, @function
```

Рис. 4: Функции после *main*, 32-х битная система

В языке С все достаточно предсказуемо, локальная переменная кладётся на вершину стека main'а, различия между 32-х и 64-х разрядной системой почти нету. Однако дело обстоит совсем по-другому с C++. В функции main все без изменений по сравнению с C, но после main'а что в 32-х битной, что в 64-х битной появляются ещё дополнительные функции, судя по некоторым различимым словам, это похоже на конструктор и деструктор класса int.

b) Теперь на очереди программа с парой локальных переменных типа int. В C, как в 32-х так и в 64-х битной системе, обе переменные просто кладутся на вершину стека main'а. В main'е листинга C++ всё идентично листингу на C, но после него опять идут дополнительные функции, которые абсолютно такие же, что и в случае одной локальной переменной, они на скринах листинга ниже и выше этого текста.

```
endbr64
pushq %rbp
          %rsp, %rbp
         $16, %rsp
%edi, -4(%rbp)
%esi, -8(%rbp)
mov1
          $1, -4(%rbp)
cmpl
                85, -8(%rbp)
          _ZStL8__ioinit(%rip), %rdi
          _ZNSt8ios_base4InitC1Ev@PLT
           _dso_handle(%rip), %rdx
          ______ZStL8__ioinit(%rip), %rsi
_ZNSt8ios_base4InitD1Ev@GOTPCREL(%rip), %rax
           _cxa_atexit@PLT
          GLOBAL__sub_I_main, @function
endbr64
pushq %rbp
          fset 6, -16
%rsp, %rbp
```

Рис. 5: Функции после main, 64-х битная система

с) При создании статического массива из int'ов на C++ после main'а присутствуют те же самые дополнительные функции (похоже конструктор деструктор), которых нету на C, однако основной интерес представляют изменения в main'е. И в C, и в C++, и в 32-х битной, и в 64-х в стеке выделяется место под элементы массива (опять больше чем надо), но перед этим выполняется команда movq с первым операндом %fs:40. Вся эта цепочка действий нужна для того, чтобы в конце работы функции проверить стек на переполненность, в регистре по адресу fs:40 лежит определенная константа, значения которой сравниваются в начале и в конце программы, если что-то идёт не как должно, то вызывается функция stack chk fail. В 32-х битной системе используется адрес gs:20.

```
nain:
LFB0:
        .cfi startproc
        endbr64
                %rbp
                %rsp, %rbp
        movq
                   2, %rsp
        subq
        movq
                %fs:40, %rax
                %rax, -8(%rbp)
                %eax, %eax
                        (%rbp)
        movl
                       8(%rbp)
        movl
                       4(%rbp)
        movl
                        (%rbp)
        mov1
                       6(%rbp)
        movl
                     %eax
        movl
                 -8(%rbp), %rdx
        movq
        xorq
                %fs:40, %rdx
        call
                 stack chk fail@PLT
        leave
        ret
```

(а) Статический массив на С

```
LFB1522:
       endbr64
                %rbp
                %rsp, %rbp
       movq
                $32, %rsp
       subq
                %fs:40, %rax
       movq
                %rax, -8(%rbp)
       movq
                %eax, %eax
       xorl
       mov1
                        (%rbp)
                $2,
                        (%rbp)
       mov1
       mov1
                    -24(%rbp)
       mov1
                    -20(%rbp)
                     -16(%rbp)
       mov1
                    %eax
       movl
                $
                -8(%rbp), %rdx
       movq
                %fs:40, %rdx
       xorq
       call
                __stack_chk_fail@PLT
```

(b) Статический массив на С++

d) Далее, динамический массив. Разницы между С и С++ опять почти нету, но большие отличия в листингах для различных систем. В 32-х битной перед и после пуша регистров, указывающих на вершину и дно стека (esp и ebp соответственно) кладутся в стек ещё и ebx и ecx, причем до этого в main'е регистр ebx вообще никак не используется, то есть по сути там лежит ненужный мусор. В каждой из систем в стеке хранится адрес на первый элемент выделенной памяти по которому и можно обращаться к любому элементу дин. массива. В 32-х битной системе количество выделенных байт (число) пушится в стек перед вызовом new(malloc), в 64-х – хранится в отдельном регистре (здесь это edi).

```
.cfi startproc
endbr32
         4(%esp), %ecx
             6, %esp
         $-1
         -4(%ecx)
         %ebp
         %esp, %ebp
push1
         %ebx
         %ecx
            5, %esp
           x86.get
         $_GLOBAL_OFFSET_TABLE_, %ebx
         _Znaj@PLT
         $16, %esp
%eax. -12(%ebp)
movl
            2(%ebp), %eax
mov1
           , %eax
mov1
              -12(%ebp)
cmpl
           2, %esp
2(%ebp)
          ZdaPv@PLT
add1
         $0, %eax
           3(%ebp), %esp
         %ecx
pop1
         %ebx
popl
         %ebp
         -4(%ecx), %esp
```

```
main:
LFB1522:
        endbr64
        pushq
                %rbp
        .cfi def cfa offset 16
        .cfi_offset 6, -16
                %rsp, %rbp
        mova
        .cfi_def_cfa_register 6
                $16, %rsp
        suba
                $12, %edi
        mov1
        call
                 Znam@PLT
                %rax, -8(%rbp)
        movq
                 -8(%rbp), %rax
        mova
                $4, %rax
        addq
                $56, (%rax)
        mov1
                    -8(%rbp)
        cmpq
        je
                 -8(%rbp), %rax
        movq
                %rax, %rdi
        movq
        call
                 _ZdaPv@PLT
L2:
        movl
                $0. %eax
        leave
        .cfi_def_cfa 7, 8
        .cfi_endproc
```

- (а) Динамический массив на С++, 32 бита
- (b) Динамический массив на С++, 64 бита

Пункт 3: а) Двигаемся к структурам, сначала, не создавая экземпляра, просто объявим структуру с четырьмя различными полями int, char, float и short. В обоих языках в листинге просто по сути пустой main(что и логично), видимо пока нету экземпляра класса это никак не отражается на листинге. В C++ main также пуст, но функции после него (которые я считал конструктором и деструктором класса int) остались абсолютно такими же, вероятно это все же что-то другое, продолжим это выяснять дальше.

b) Теперь создадим глобальную переменную структуры и попробуем изменить пару её полей в main'е. Обращение происходит как и в случае с глобальным массивом, причем выделяется под каждое поле 4 байта, хотя некторые из них весят меньше . Перед main'ом объявляется глобальная переменная переменная. Принципиальных различий между С и C++ все также нет.

```
@object
                main, @function
main:
LFB1519:
        .cfi_startproc
        endbr32
        pushl
                %ebp
                %esp, %ebp
                 $_GLOBAL_OFFSET_TABLE_, %eax
        add1
                 S@GOTOFF(%eax), %edx
        mov1
        add1
                 $1, %edx
                 %edx, S@GOTOFF(%eax)
                $65, 4+S@GOTOFF(%eax)
$0, %eax
        movb
        pop1
                 %ebp
```

```
main
                main, @function
main:
LFB1522:
        .cfi startproc
       endbr64
       pushq
                %rbp
                %rsp, %rbp
       movq
                S(%rip), %eax
        add1
                $1, %eax
                %eax, S(%rip)
       mov1
                $65, 4+S(%rip)
       movb
                    %eax
                $6
        movl
                %rbp
```

- (а) Глобальная структура на С++, 32 бита
- (b) Глобальная структура на С++, 64 бита
- с) Теперь в ту же структуры добавим статический массив. Ничего сверхъестественного не получаем, статический массив размера N по сути эквивалентен N переменным одного типа, так его и представляет компилятор, как бы развертывает массив в N новых полей одного типа. Принципиальных различий в разных системах и языках нету.
- d) Тут передадим структуру в функцию (пока что не по указателю или ссылке). В обеих системах и обоих языках перед функцией в стек пушатся копии всех полей структуры (элементы массива пушатся поочереди) таким образом, что наверху оказывается первое по счёту поле. Различие в том, что в 32-х битной системе это делается с помощью команды pushl, а в 64-х битной это делается командой movq, используя адрес ячейки в стеке, к тому же, даже если переменные весят меньше 8 байт, к примеру две по 4 байта, их слепляют в одну восьмибайтовую, достать одно из двух значений конечно можно дописав к команде mov окончание l и указав правильный адрес.
- е) Если сделать возвращаемым значением сделать структуру, то в этом случае также, как и когда функция возвращала целое число, в стеке перед вызовом функции резервируется место под возвращаемое значение, затем в стек пушатся все аргументы, потом вызывается функция. после завершения работы на вершине стека оказывается возвращаемая структура. Различие между системами только в способе заполнения стека, в 32-х битной с помощью команды pushl, в 64-х movq.
- f) В этом пункте взглянем на методы класса в C++. При создании экземпляра класса вызывается его конструктор, если у класса есть поля, под их значения выделяется место в стеке перед вызовом самой функции (конструктора). После этого вызывается функция, куда передаются адреса для записи значений полей, куда в итоге и попадают значения полей экземпляра класса. Если вызывается метод класса, то, если он использует поля класса, в функцию передаются как бы неявные аргументы поля, метод работает

как обчная функция, если он меняет поля, то по сути это принимающая на вход указатель и меняющая по нему значение функция (что опять же логично очень даже).

g) Операторы. Операторы это те же функции. Работают также, результат операции отправляется в заранее заготовленную ячейку памяти. Единственное, в 32-х битной системе результат записывается в нужную ячейку в конце выполнения функции, в 64-х битной результат возвращается в регистре, после чего записывается из регистра в нужную часть памяти.

Пункт 4: а) Структуру в функцию мы уже передавали в пункте 3d), поэтому тут все тоже самое.

- b) Теперь передадим структуру по указателю. Здесь все также достаточно ожидаемо, просто теперь функция использует не скопированные значения полей, а адреса ячеек, в которых лежат эти значения. Обращаясь по этим адреса, функция меняет сам объект, а не его копию. Отличие между 32-х и 64-х битной системами в том, что в 64-х битной когда полей мало, то их адреса передаются в функцию через регистры (адрес записывается в регистр, потом переписывается из регистра в стек функции после её вызова), а в 32-х битной всегда адреса пушатся в стек перед вызовом функции.
- с) Ссылки на очереди. Передача в функцию по ссылке в ассемблерном листинге никоим образом не отличается от передачи по указателю. Все абсолютно также, как и в предыдущем пункте.
- d) В этом пункте рассмотрим передачу в функцию с помощью rvalue-ссылок. В этом случае просто результат промежуточного действия записывается в одну из ячееку стека, адрес которой потом передаётся в функцию.

Пункт 5: а) Создадим структуру со статическим массивом в 100000 элементов значением 0, передадим её аргументом в функцию. Листинги программы приведены ниже:

```
(a) Жирная структура в аргументе C++, 32 бита
```

```
endbr64
        %rbp
         %rsp, %rbp
                 2(%rsp), %r11
             96, %rsp
(%rsp)
        %fs:49 %rsp
suba
         %fs:40, %rax
%rax, -8(%rbp)
xorl
                 (%rbp), %rax
            wesi
mov1
mov1
         memset@PLT
                   %rsp
                 (%rbp), %rax
                  , %edx
mov1
         %rax,
                %rsi
         memcpy@PLT
                   %rsp
mov1
         -8(%rbp), %rcx
%fs:40, %rcx
           _stack_chk_fail@PLT
```

(b) Жирная структура в аргументе C++, 64 бита

Как видно, сначала при создании экземпляра структуры выделяется 400000 байт, значения задаются с

помощью функции memset, которая принимает на вход адрес первого элемента массива, значение, которое нужно присвоить элементам массива (здесь это 0), и длину массива. Различие между системами в том, что в 32-х битной эти аргументы функции memset пушатся в стек, а затем функция уже берет их оттуда, а в 64-х битной системе аргументы кладутся в регистры, откуда потом функция их использует. Но это относится к созданию экземпляра класса, что касается его передачи в функцию (точнее его копии), то тут дело обстоит примерно также, за исключением того, что для копирования массива при передачи в функцию используется функция memcpy. Она принимает на вход адрес куда копировать, адрес откуда копировать и количество элементов. Аргументы передаются в функцию таким же образом, по-разному в разных системах. Также заметим, что первое выделение места под структуру происходит не единовременным вычитанием 400000 байт из rsp(esp), а в цикле, вычитая по 4096 байт (вот это выглядит странным, плюс тот факт, что с числом, лежащим в ячейке на которую указывает промежуточное значение rsp(esp), прозводится логическое или, где второй операнд есть 0).

- b) Теперь огромную структуру будем возвращать из функции, принимающей на вход тоже большую структуру. По сути, ничего принципиально отличающегося от предыдущего случая тут нету, просто больше копирований туда-сюда. Заранее выделяется место в стеке под возвращаемое значение (причем оно оказалось ниже в стеке, чем значение полей изначально созданной структуры, копия которой потом передаётся в функцию), с помощью темсру сначала копируется на вершину весь массив как аргумент функции, затем, после операций, используя тот же темсру, измененные значения копируются на дно стека, после этого работа функции заканчивается.
- с) Если создать большую структуру внутри функции, то с помощью команды memset в стеке функции будет создан экземпляр структуры точно также, как и в стеке *main*'a.
- d) При изменении размера структуры вроде как в обеих системах ничего, кроме размера выделяемой памяти, не меняется.

Пункт 6: Напишем рекурсивную функцию и посмотрим на её листинг. Просто из функции вызывается эта же функция.

```
(а) Рекурсия на С, 32 бита
```

```
Z4funci:
       endbr64
               %rbp
                %rsp, %rbp
       movq
                $16, %rsp
       subq
                %edi, -4(%rbp)
       mov1
                $1, -4(%rbp)
       cmp1
                -4(%rbp), %eax
       mov1
                $1, %eax
       subl
                %eax, %edi
       movl
                Z4funci
       call
       jmp
                -4(%rbp), %eax
       ret
```

(b) Рекурсия на C++, 64 бита