## Practical Assembly

From the Ground Up

John Cunniff
OSIRIS Lab Hack Night

#### Overview - Rev 2

- Overview
  - O Who is this for?
  - Owner with the complex of the com
  - Why should we learn it?
  - Required Mindset
- Basic Assembly topics
  - Memory
  - Registers
  - Basic instructions

- Advanced Topics
  - Registers (again)
  - Memory layout
  - Stack Frame
- Helpful tools!
  - o objdump
  - Binary Ninja
  - godbolt.org

## Overview

People new to assembly

- People new to assembly
- Anyone that wants a better understanding of how computers work!

- People new to assembly
- Anyone that wants a better understanding of how computers work!
- Those that want to start playing reverse engineering / pwning challenges

- People new to assembly
- Anyone that wants a better understanding of how computers work!
- Those that want to start playing reverse engineering / pwning challenges
  - We're going to start with very abstract concepts

- People new to assembly
- Anyone that wants a better understanding of how computers work!
- Those that want to start playing reverse engineering / pwning challenges
  - We're going to start with very abstract concepts
  - Build up to what it actually looks like

• Have you ever wondered how a machine can run code?

- Have you ever wondered how a machine can run code?
  - The answer is "machine code"

- Have you ever wondered how a machine can run code?
  - The answer is "machine code"
  - Machine code is translated from Assembly

- Have you ever wondered how a machine can run code?
  - The answer is "machine code"
  - Machine code is translated from Assembly
- Machine code is what your cpu is able to actually interpret and run

- Have you ever wondered how a machine can run code?
  - The answer is "machine code"
  - Machine code is translated from Assembly
- Machine code is what your cpu is able to actually interpret and run
- The instructions can let you modify the cpu, memory and other physical devices directly

Why should we care at all?

• There are basically no rules!

- There are basically no rules!
- Bugs at this low level lead to vulnerabilities

- There are basically no rules!
- Bugs at this low level lead to vulnerabilities
- Better our understanding of how computers work!

- There are basically no rules!
- Bugs at this low level lead to vulnerabilities
- Better our understanding of how computers work!
  - Everything that runs has to be in one way or another translated back to assembly

- There are basically no rules!
- Bugs at this low level lead to vulnerabilities
- Better our understanding of how computers work!
  - Everything that runs has to be in one way or another translated back to assembly
  - There is no avoiding it!

- There are basically no rules!
- Bugs at this low level lead to vulnerabilities
- Better our understanding of how computers work!
  - Everything that runs has to be in one way or another translated back to assembly
  - There is no avoiding it!
  - The computer you're viewing this lecture from, the stream your watching it on, the servers all that data is passing through, all assembly

• Assembly can't lie....

- Assembly can't lie....
- Have you ever paid for an application that you have downloaded?

- Assembly can't lie....
- Have you ever paid for an application that you have downloaded?
  - That application will be a compiled program of "machine code"

- Assembly can't lie....
- Have you ever paid for an application that you have downloaded?
  - That application will be a compiled program of "machine code"
  - Using special tools we can figure out how the application works!

- Assembly can't lie....
- Have you ever paid for an application that you have downloaded?
  - That application will be a compiled program of "machine code"
  - Using special tools we can figure out how the application works!
  - This is called "reverse engineering"

Going into assembly

• I'm going to ask you to do something difficult during this lecture

- I'm going to ask you to do something difficult during this lecture
  - Let go of whatever understanding you have of how programs work (in python, c++, java, go, etc.)

- I'm going to ask you to do something difficult during this lecture
  - Let go of whatever understanding you have of how programs work (in python, c++, java, go, etc.)
  - Instead try to conceptualize this as a totally new way of thinking

- I'm going to ask you to do something difficult during this lecture
  - Let go of whatever understanding you have of how programs work (in python, c++, java, go, etc.)
  - Instead try to conceptualize this as a totally new way of thinking
  - Don't try to connect this new way of thinking to your current understanding (yet)

- I'm going to ask you to do something difficult during this lecture
  - Let go of whatever understanding you have of how programs work (in python, c++, java, go, etc.)
  - Instead try to conceptualize this as a totally new way of thinking
  - Don't try to connect this new way of thinking to your current understanding (yet)
  - Once you understand these systems we'll connect it back to C specifically

- I'm going to ask you to do something difficult during this lecture
  - Let go of whatever understanding you have of how programs work (in python, c++, java, go, etc.)
  - Instead try to conceptualize this as a totally new way of thinking
  - Don't try to connect this new way of thinking to your current understanding (yet)
  - Once you understand these systems we'll connect it back to C specifically
    - C is easily readable and can be basically directly connected back to assembly

• We're going to be using Intel syntax, **NOT** AT&T syntax

- We're going to be using Intel syntax, **NOT** AT&T syntax
- This lecture will focus on 32bit x86

- We're going to be using Intel syntax, **NOT** AT&T syntax
- This lecture will focus on 32bit x86
  - Almost all modern computers are 64bit now

#### Required Mindset

- We're going to be using Intel syntax, **NOT** AT&T syntax
- This lecture will focus on 32bit x86
  - Almost all modern computers are 64bit now
  - Only small differences!

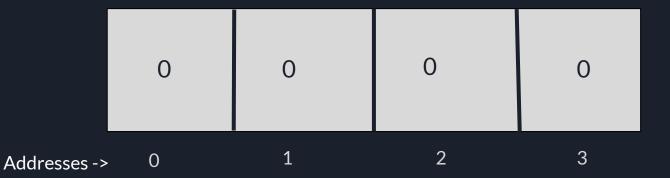
# The basics

• Think of memory as one long array...

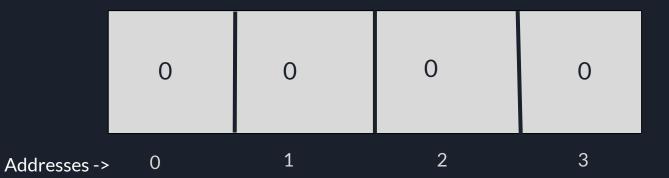
- Think of memory as one long array...
- This array has indices (called addresses)

- Think of memory as one long array...
- This array has indices (called addresses)
- Same as with any program, you can't read outside the array!

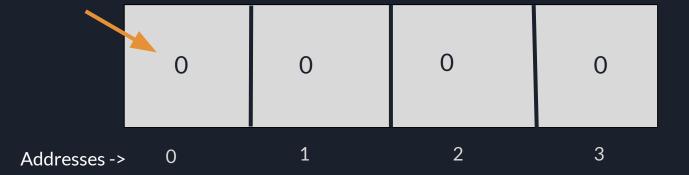
- Think of memory as one long array...
- This array has indices (called addresses)
- Same as with any program, you can't read outside the array!



• In assembly we move values in and out of memory one at a time



- In assembly we move values in and out of memory one at a time
- We can mov 1 into address 0



- In assembly we move values in and out of memory one at a time
- We can "mov" 1 into address 0

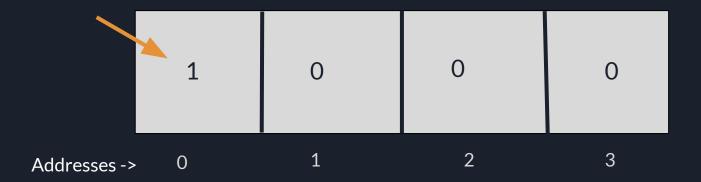


- In assembly we move values in and out of memory one at a time
- We can "mov" 1 into address 0
- In assembly this would be: mov [0], 1

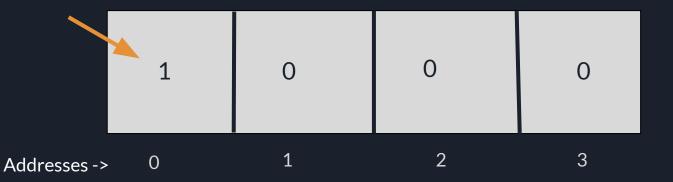


The brackets mean "dereference"

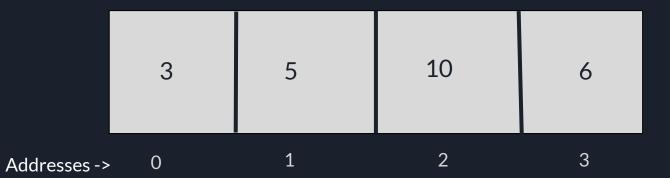
mov [destination address], source



- mov [destination address], source
- Where destination address is the index (0) and source is the value of the new number (1)



• So what will "mov [2], 3" do?



• So what will "mov [2], 3" do?



• So what will "mov [2], 3" do?



# The basics

• The biggest mistake you can make is thinking of registers as variables...

- The biggest mistake you can make is thinking of registers as variables...
- Registers are the physical memory locations on the CPU

- The biggest mistake you can make is thinking of registers as variables...
- Registers are the physical memory locations on the CPU
- Hold small amounts of data

- The biggest mistake you can make is thinking of registers as variables...
- Registers are the physical memory locations on the CPU
- Hold small amounts of data
- Perform super fast operations

- The biggest mistake you can make is thinking of registers as variables...
- Registers are the physical memory locations on the CPU
- Hold small amounts of data
- Perform super fast operations
  - Much faster than memory

• Let's say we have a register "a"

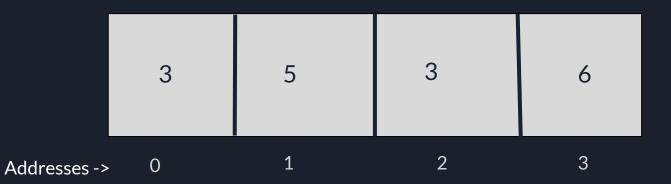
- Let's say we have a register "a"
- This register a can hold 8 bits or 1 byte of data

- Let's say we have a register "a"
- This register a can hold 8 bits or 1 byte of data
- That means we can have values between 0 and 255 in a

- Let's say we have a register "a"
- This register a can hold 8 bits or 1 byte of data
- That means we can have values between 0 and 255 in a
- So how do we move values in and out of registers?

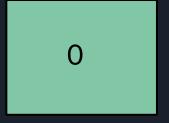
• So how do we move data in and out of registers?

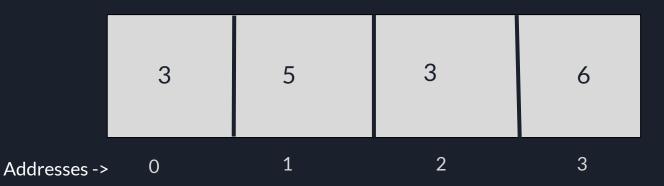




Remember its:
mov dst, src

- So how do we move data in and out of registers?
- mov a, 10



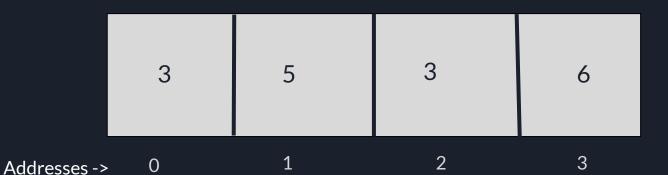


Remember its:
mov dst, src

- So how do we move data in and out of registers?
- mov a, 10

move value 10 into a





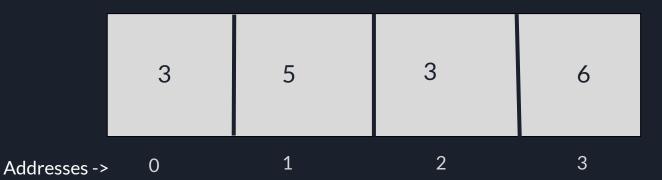
Remember its:
mov dst, src

• So how do we move data in and out of registers?

• mov a, 10

move value 10 into a

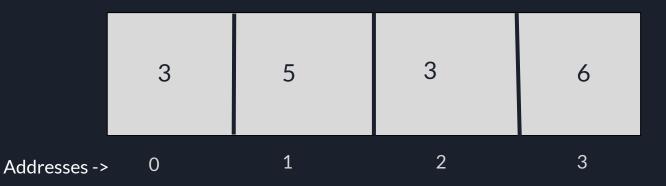




Remember its:
mov dst, src

What about moving things between memory?

10

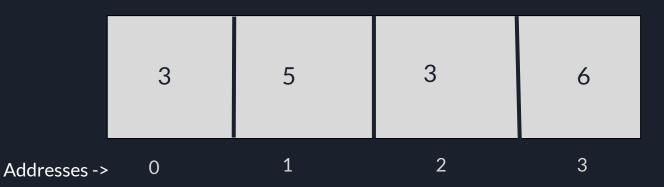


Remember its:
mov dst, src

• What about moving things between memory?

• mov a, [0]

10



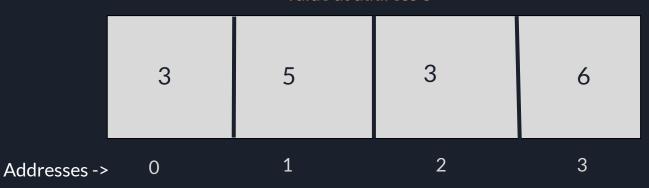
Remember its:
mov dst, src

• What about moving things between memory?

10

mov a, [0]

Square brackets means the value at address 0



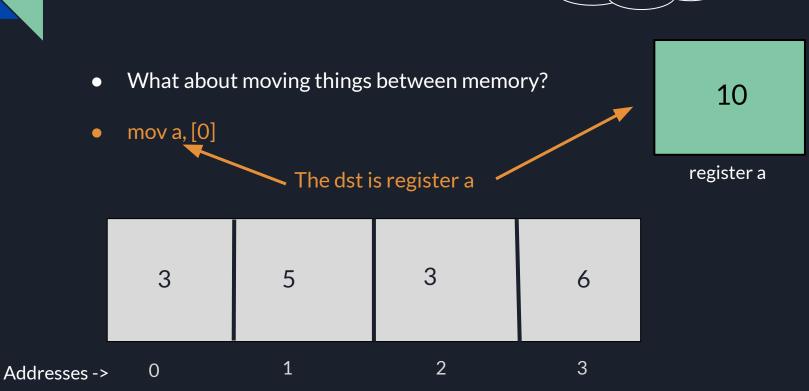
Remember its: mov dst, src

What about moving things between memory?

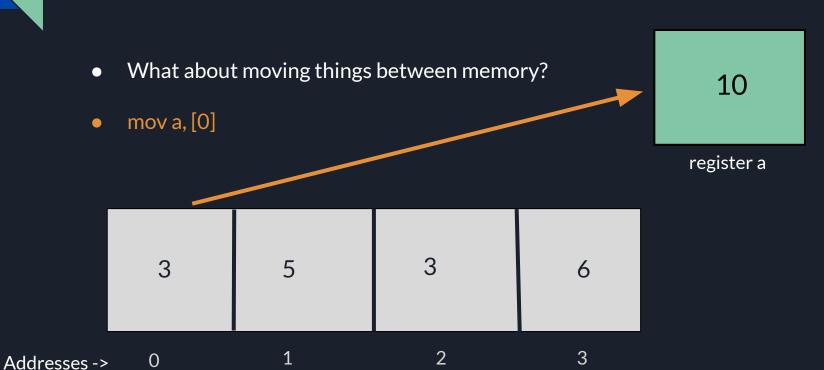
mov a, [0] The src is the value at address 0 2 3 Addresses ->

10

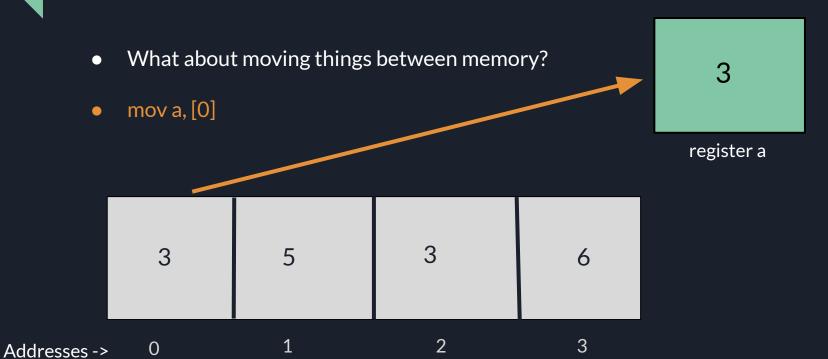
Remember its:



Remember its: mov dst, src



Remember its: mov dst, src

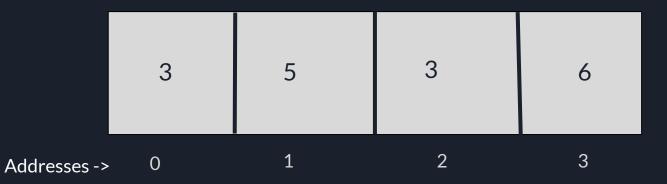


Remember its:
mov dst, src

• Ok now let's change the value...

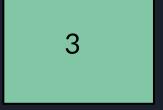
3

register a

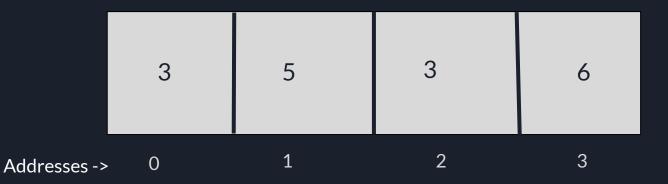


Remember its:
add dst, src

- Ok now let's change the value...
- add a, 10



register a



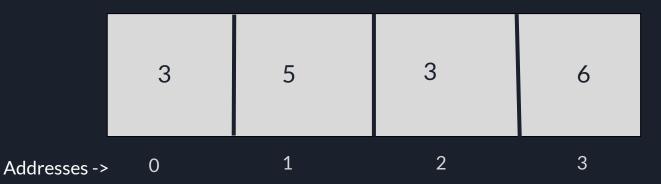
Remember its:

- Ok now let's change the value...
- add a, 10

Add 10 and the value in a

3



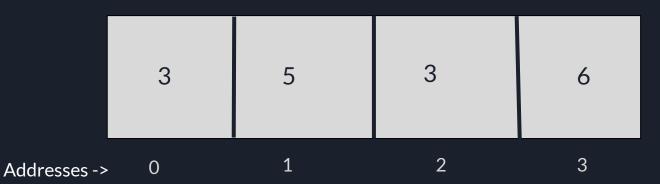


Remember its:
add dst, src

- Ok now let's change the value...
- add a, 10

Then place the result in a



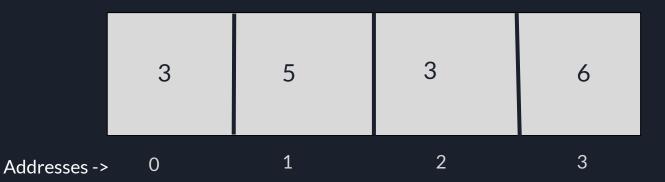


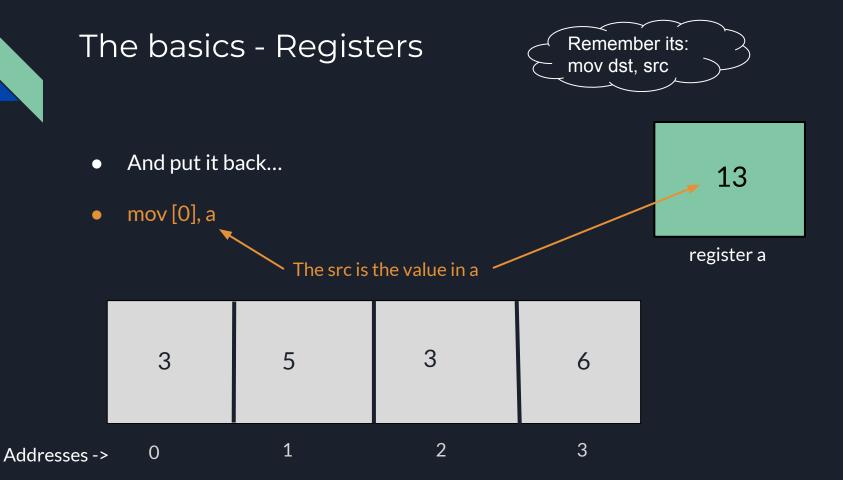
Remember its:
add dst, src

- Ok now let's change the value...
- add a, 10

Then place the result in a



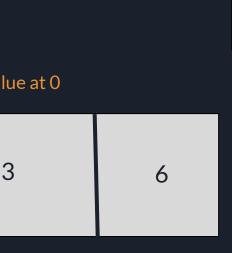




Remember its: mov dst, src

And put it back...

mov [0], a



The dst is the value at 0

2

3

13

register a

Addresses ->

0







## That was a lot of steps!

I promise it gets easier...

## The basics

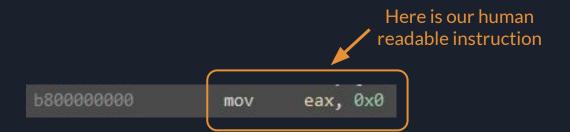
Instructions

#### Instructions

• Instructions are sometimes called "machine code" or "bytecode"

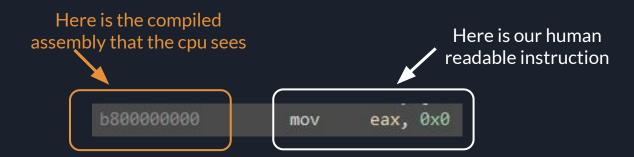
#### Instructions

- Instructions are sometimes called "machine code" or "bytecode"
- The assembly that we can read (like mov a, 10) gets translated to bytecode



#### Instructions

- Instructions are sometimes called "machine code" or "bytecode"
- The assembly that we can read (like mov a, 10) gets translated to bytecode



• mov dst, src - Moves memory from src to dst

- mov dst, src Moves memory from src to dst
- add dst, src Adds dst and src, stores result in dst
- [address] Dereference (says the value in brackets is an address)
- jmp address Jumps from one location in the code to another

- mov dst, src Moves memory from src to dst
- add dst, src Adds dst and src, stores result in dst
- jmp address Jumps from one location in the code to another

- mov dst, src Moves memory from src to dst
- add dst, src Adds dst and src, stores result in dst
- jmp address Jumps from one location in the code to another
- call address Fancy jump (more on this later)

- mov dst, src Moves memory from src to dst
- add dst, src Adds dst and src, stores result in dst
- jmp address Jumps from one location in the code to another
- call address Fancy jump (more on this later)
- push value Pushes the value to the top of the stack

- mov dst, src Moves memory from src to dst
- add dst, src Adds dst and src, stores result in dst
- jmp address Jumps from one location in the code to another
- call address Fancy jump (more on this later)
- push value Pushes the value to the top of the stack
- cmp value1, value2 Sets flags in EFLAGS

- mov dst, src Moves memory from src to dst
- add dst, src Adds dst and src, stores result in dst
- jmp address Jumps from one location in the code to another
- call address Fancy jump (more on this later)
- push value Pushes the value to the top of the stack
- cmp value1, value2 Sets flags in EFLAGS
- je, jne, jz, jnz Conditional jumps (look them up there are a lot)

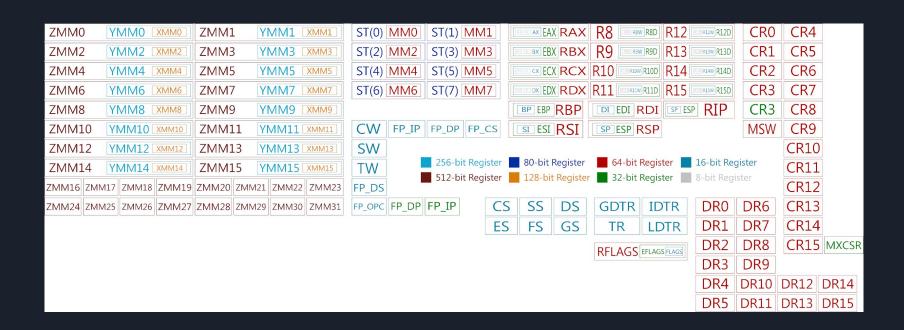
• Any time you see an instruction you don't know...

- Any time you see an instruction you don't know...
  - o Look it up!

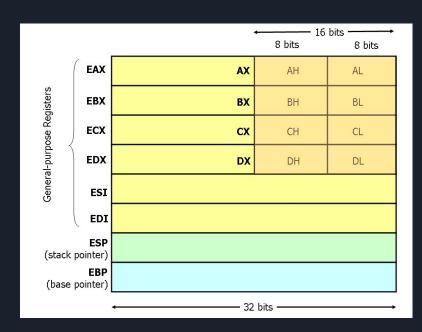
- Any time you see an instruction you don't know...
  - Look it up!
  - There are decades of forum questions on any and all assembly problems

Again but advanced...

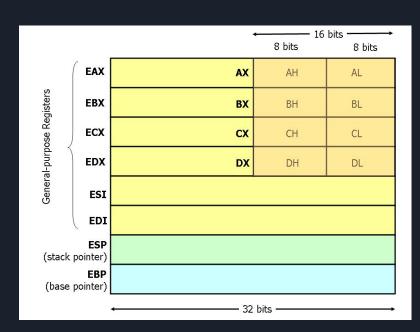
#### Registers (there's a lot of them)



• Here are the ones we care about...



- Here are the ones we care about...
- You can refer to different sections of the same space...



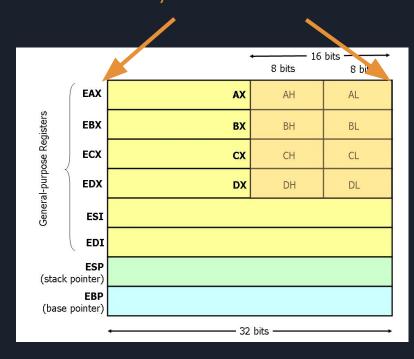
- Here are the ones we care about...
- You can refer to different sections of the same space...

# al refers to the lowest byte of the eax register

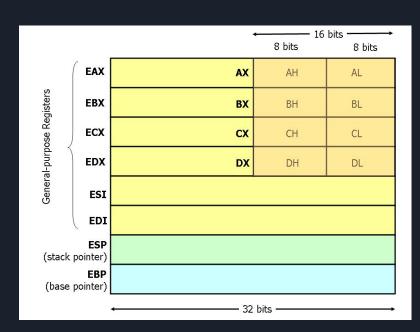


- Here are the ones we care about...
- You can refer to different sections of the same space...

# eax refers to the full 4 bytes

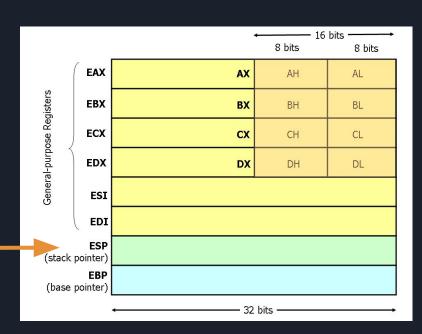


- Here are the ones we care about...
- You can refer to different sections of the same space...
- There are some registers we can use for anything, some are reserved



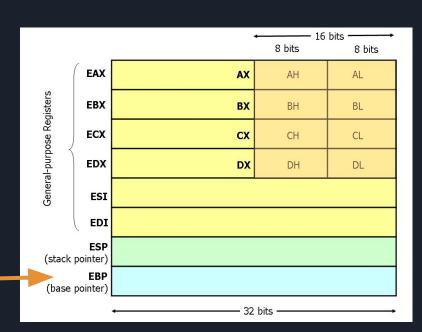
- Here are the ones we care about...
- You can refer to different sections of the same space...
- There are some registers we can use for anything, some are reserved

esp is reserved for the stack pointer

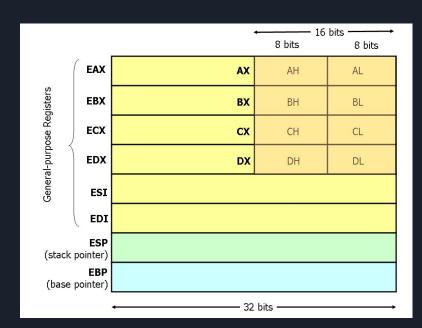


- Here are the ones we care about...
- You can refer to different sections of the same space...
- There are some registers we can use for anything, some are reserved

ebp is reserved for the stack base pointer

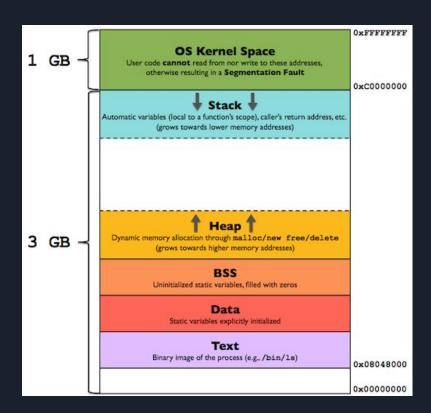


- Here are the ones we care about...
- You can refer to different sections of the same space...
- There are some registers we can use for anything, some are reserved
- More on that later...



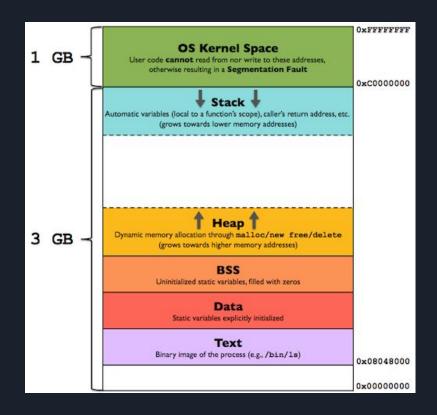
#### Memory Layout

Some will draw the stack growing up, some growing down.



Some will draw the stack growing up, some growing down.

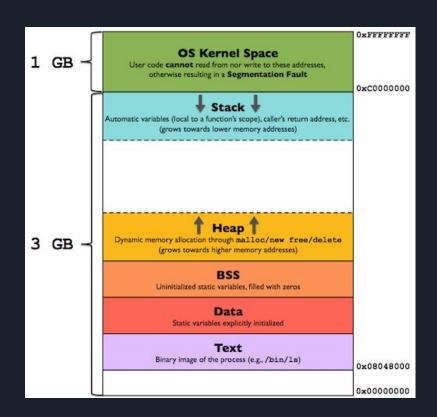
It doesn't matter which model you adopt, they are exactly the same!



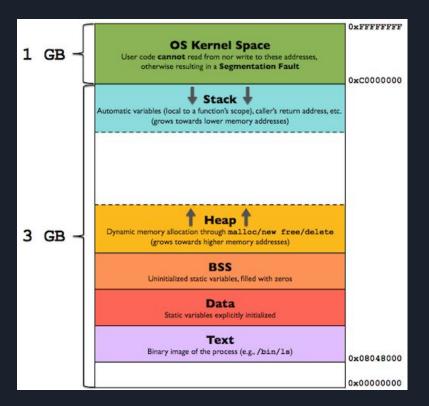
Some will draw the stack growing up, some growing down.

It doesn't matter which model you adopt, they are exactly the same!

Programs have "sections" of memory



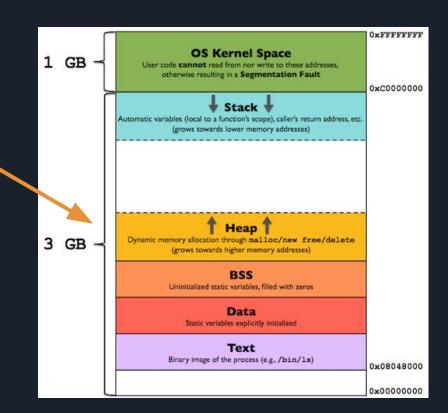
Something important:
All of this is just data! We as programmers give purpose to this data in the way we treat it!



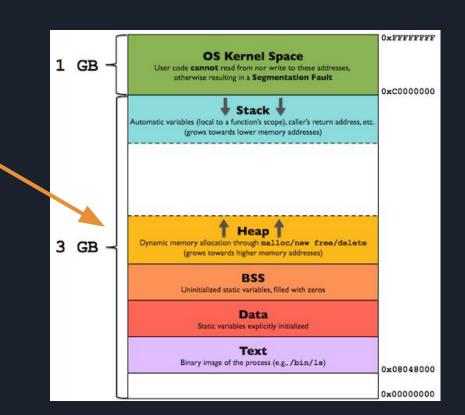
Oxffffffff **OS Kernel Space** 1 GB User code cannot read from nor write to these addresses, otherwise resulting in a Segmentation Fault 0xC0000000 Stack 1 Automatic variables (local to a function's scope), caller's return address, etc. (grows towards lower memory addresses) Stack is where the local variables for functions live 3 GB Dynamic memory allocation through malloc/new free/delete (grows towards higher memory addresses) BSS Uninitialized static variables, filled with zeros Data Static variables explicitly initialized Text Binary image of the process (e.g., /bin/ls) 0x08048000 0x00000000

Oxffffffff **OS Kernel Space** 1 GB User code cannot read from nor write to these addresses, otherwise resulting in a Segmentation Fault 0xC0000000 Stack 1 Automatic variables (local to a function's scope), caller's return address, etc. (grows towards lower memory addresses) As you call functions, the stack grows down 3 GB Dynamic memory allocation through malloc/new free/delete (grows towards higher memory addresses) BSS Uninitialized static variables, filled with zeros Data Static variables explicitly initialized Text Binary image of the process (e.g., /bin/ls) 0x08048000 0x00000000

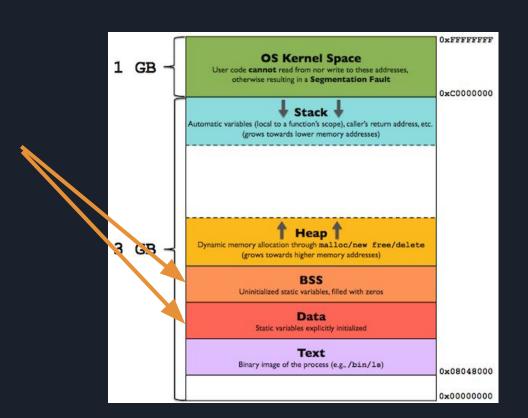
The heap is where dynamically allocated data lives



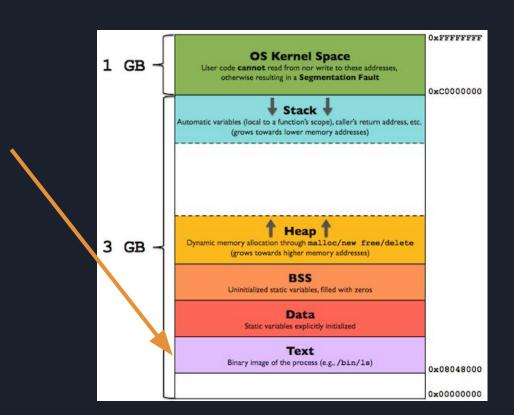
As you allocate more data, the heap grows up



The bss and data sections are where global variables live



The "text" section holds the assembly code of the program

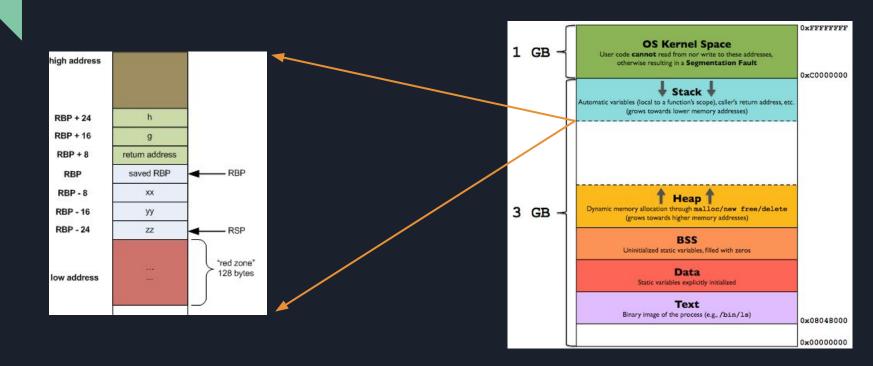


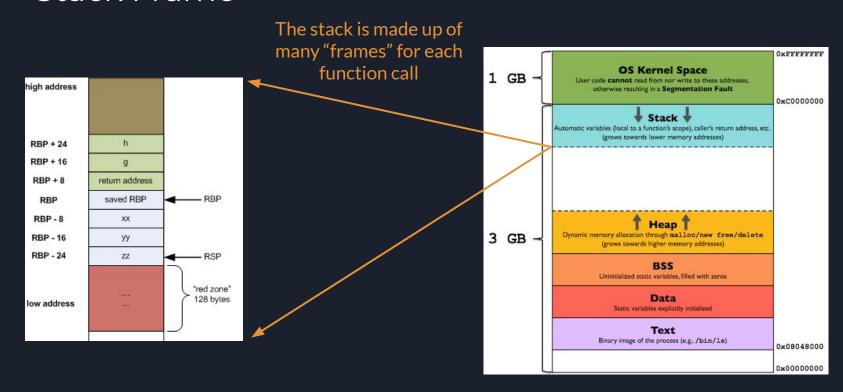
# Ok now we have a problem...

- How exactly do we organize memory?
- At this level there are basically no rules right?
- So how do we keep things straight?

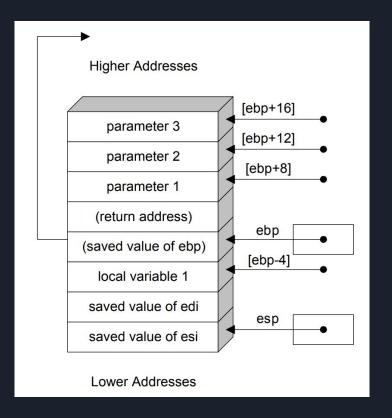
# The organization solution

The Stack Frame

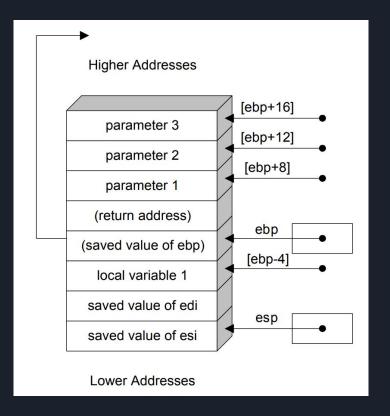




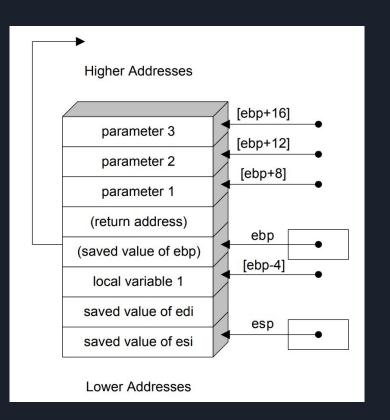
- Let's talk instructions
  - call address pushes the address
     of the next instruction then jmp



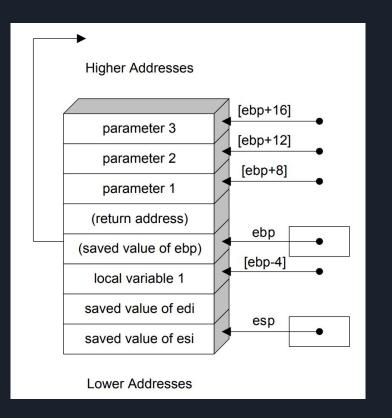
- Let's talk instructions
  - call address pushes the address
     of the next instruction then jmp
  - ret pop's top value off the stack into eip



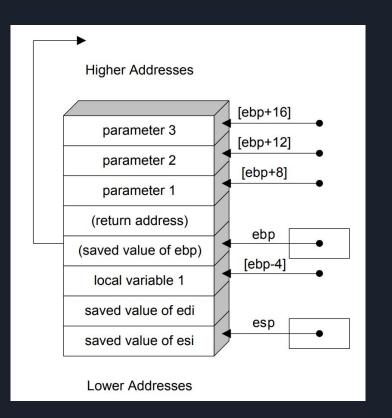
- Let's talk instructions
  - call address pushes the address
     of the next instruction then jmp
  - o ret pop's top value off the stack into eip
- When a new function is called, it will run a prolog and epilog
  - The prolog sets up the frame
  - The epilog cleans it up and ret's



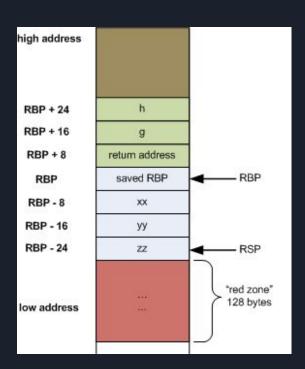
- Let's talk instructions
  - call address pushes the address
     of the next instruction then jmp
  - ret pop's top value off the stack into eip
- When a new function is called, it will run a prolog and epilog
  - The prolog sets up the frame



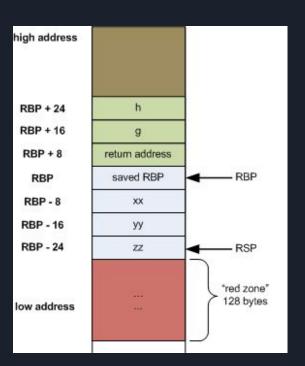
- Let's talk instructions
  - call address pushes the address
     of the next instruction then jmp
  - o ret pop's top value off the stack into eip
- When a new function is called, it will run a prolog and epilog
  - The prolog sets up the frame
  - The epilog cleans it up and ret's



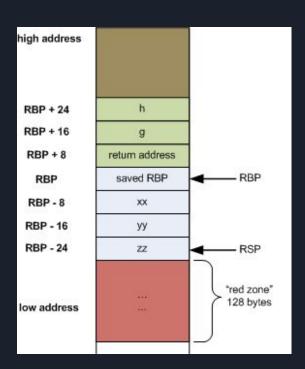
 Every function will make a new stack frame, then clean it up before returning

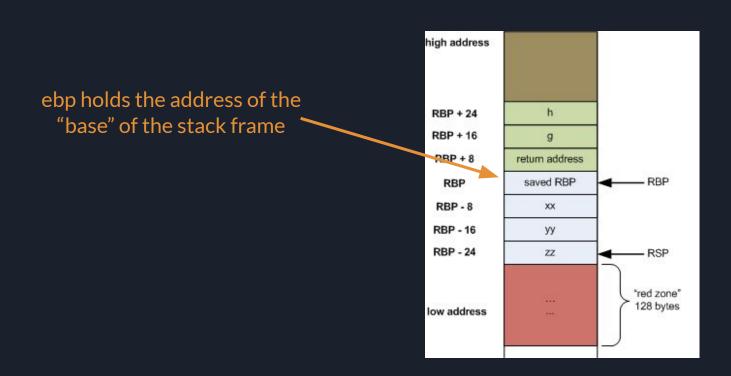


- Every function will make a new stack frame, then clean it up before returning
- The stack frame is where local variables are stored

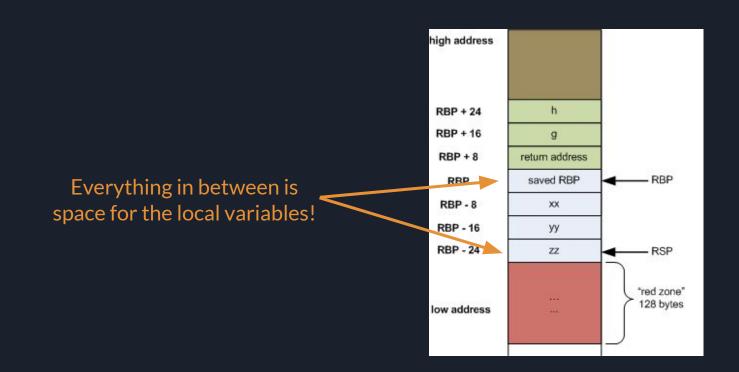


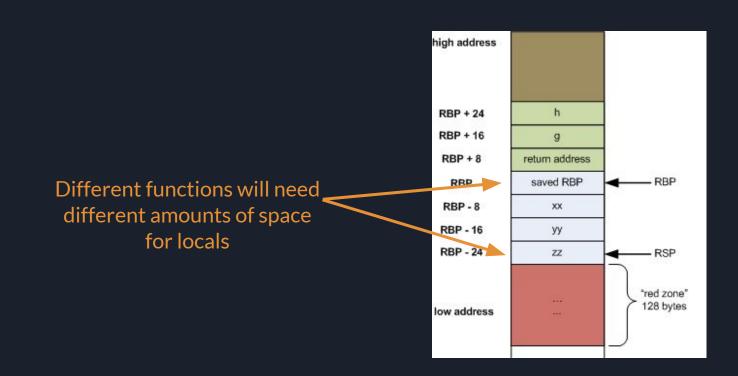
Remember ebp and esp from before?

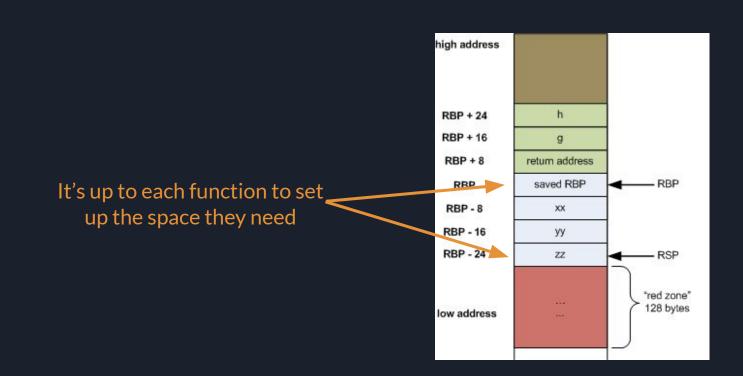




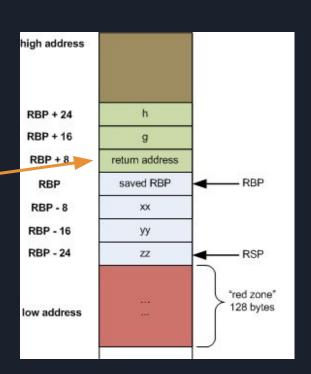


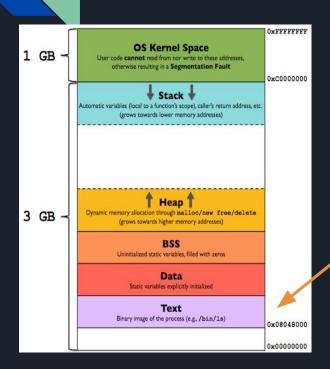




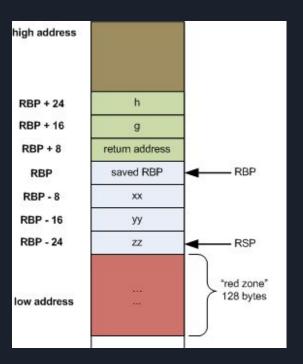


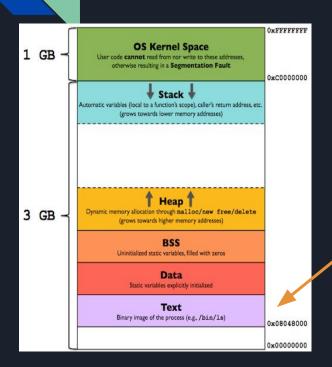
To understand the return address, let's remember how the program layout looks



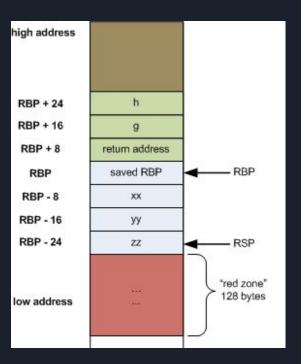


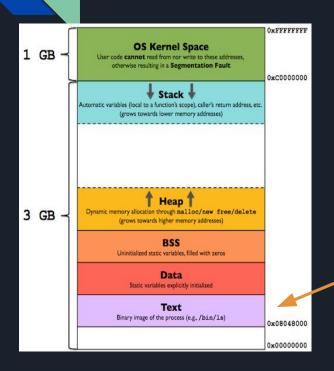
The compiled assembly lives in memory



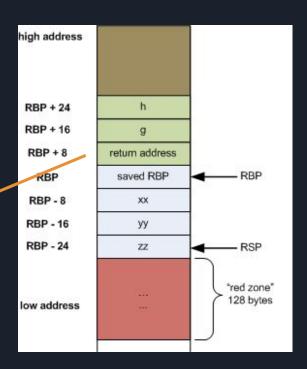


That means it has addresses!





So the value in the return address is the address of where the function should return to



### 32 vs 64 bit

Calling conventions are the main difference.

- 32 bit uses stack
- 64 bit uses registers
  - Don't memorize, just look them up!

# Tools

We'll take all the help we can get....

Objdump
For when you have no other option...

# Objdump - As basic as it gets

```
jc@aion < master > : ~/osiris/git/Hack-Night/Rev/static
[1] % objdump -M intel --disassemble= start chal
chal:
         file format elf64-x86-64
Disassembly of section .init:
Disassembly of section .plt:
Disassembly of section .plt.got:
Disassembly of section .text:
0000000000000760 <_start>:
       31 ed
 760:
                                      ebp,ebp
                               xor
       49 89 d1
                                     r9,rdx
                               mov
       5e
                                     rsi
                               pop
      48 89 e2
                                     rdx,rsp
 766:
                               mov
 769:
      48 83 e4 f0
                                     rsp,0xffffffffffffff0
                               and
 76d:
                               push
                                     rax
 76e:
       54
                               push
                                     rsp
                                     r8,[rip+0x27a]
 76f:
      4c 8d 05 7a 02 00 00
                                                           # 9f0 <__libc_csu_fini>
                               lea
      48 8d 0d 03 02 00 00
                               lea rcx,[rip+0x203]
                                                            # 980 <__libc_csu_init>
 776:
                                     rdi,[rip+0x150]
                                                            # 8d4 <main>
 77d:
      48 8d 3d 50 01 00 00
                               lea
 784:
      ff 15 56 08 20 00
                               call
                                     QWORD PTR [rip+0x200856]
                                                                     # 200fe0 <__libc_start_main@GLIBC_2.2.5>
 78a:
Disassembly of section .fini:
```

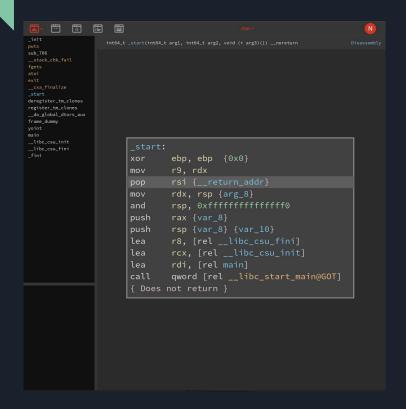
# Objdump - As basic as it gets

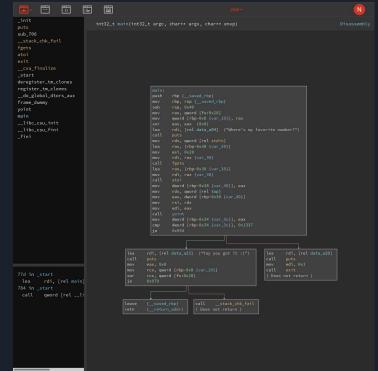
```
jc@aion < master > : ~/osiris/git/Hack-Night/Rev/static
[1] % objdump -M intel --disassemble= start chal
chal:
          file format elf64-x86-64
                                                              This is the executable file
Disassembly of section .init:
                                                              Im asking to disassemble just
Disassembly of section .plt:
                                                              the start function
Disassembly of section .plt.got:
                                                              And I'm specifying Intel
Disassembly of section .text:
                                                              syntax
00000000000000760 <_start>:
 760:
        31 ed
                                      ebp,ebp
                               xor
       49 89 d1
                                      r9, rdx
                               mov
       5e
                                      rsi
                               pop
                                      rdx,rsp
 766:
       48 89 e2
                               mov
 769:
       48 83 e4 f0
                                      rsp,0xfffffffffffff0
                               and
 76d:
       50
                               push
                                      rax
 76e:
       54
                               push
                                      rsp
       4c 8d 05 7a 02 00 00
                                     r8,[rip+0x27a]
 76f:
                                                           # 9f0 <__libc_csu_fini>
                               lea
                                                            # 980 <__libc_csu_init>
       48 8d 0d 03 02 00 00
                                     rcx.[rip+0x203]
 776:
                               lea
                                      rdi,[rip+0x150]
 77d:
       48 8d 3d 50 01 00 00
                               lea
                                                            # 8d4 <main>
 784:
       ff 15 56 08 20 00
                               call
                                      QWORD PTR [rip+0x200856]
                                                                     # 200fe0 <__libc_start_main@GLIBC_2.2.5>
 78a:
Disassembly of section .fini:
```

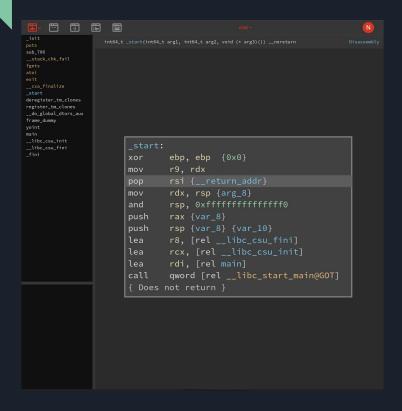
# Objdump - As basic as it gets

```
jc@aion < master > : ~/osiris/git/Hack-Night/Rev/static
[1] % objdump -M intel --disassemble= start chal
chal:
         file format elf64-x86-64
                                         The compiled bytecode
Disassembly of section .init:
Disassembly of section .plt:
                                                                      The human
Disassembly of section .plt.got:
                                                                 readable assembly
Disassembly of section /text:
00000000000000760 <__tart>:
       31 ed
 760:
                               xor
                                      ebp,ebp
       49 89 d1
                                      r9, rdx
                               mov
 765:
       5e
                                      rsi
                                pop
       48 89 e2
                                      rdx,rsp
 766:
                               mov
 769:
       48 83 e4 f0
                                      rsp,0xfffffffffffff0
                               and
 76d:
                               push
                                      rax
 76e:
                               push
                                      rsp
                                      r8,[rip+0x27a]
                                                            # 9f0 <__libc_csu_fini>
 76f:
       4c 8d 05 7a 02 00 00
                                lea
       48 8d 0d 03 02 00 00
                                      rcx,[rip+0x203]
                                                              980 <__libc_csu_init>
 776:
                               lea
 77d:
       48 8d 3d 50 01 00 00
                               lea
                                      rdi,[rip+0x150]
                                                              8d4 <main>
                                      QWORD PTR [rip+0x200856]
 784:
       ff 15 56 08 20 00
                               call
                                                                      # 200fe0 <__libc_start_main@GLIBC_2.2.5>
 78a:
Disassembly of section .fini:
```

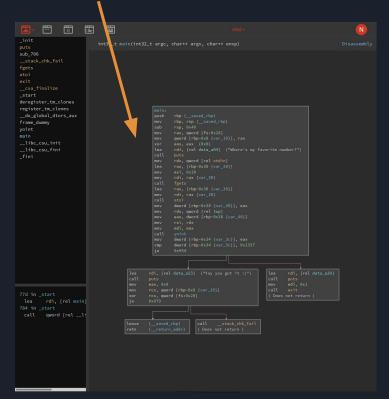
## Binary Ninja cloud.binary.ninja

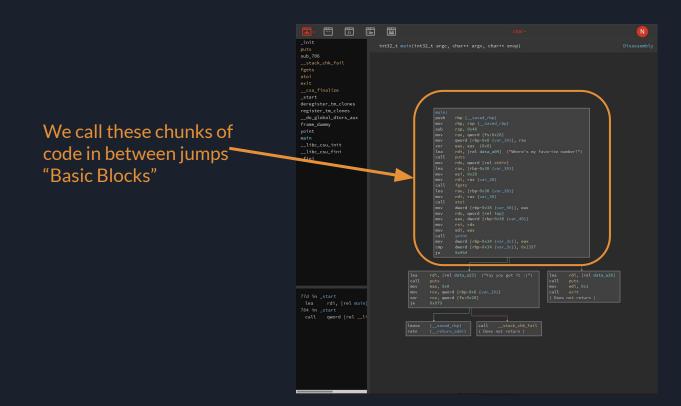


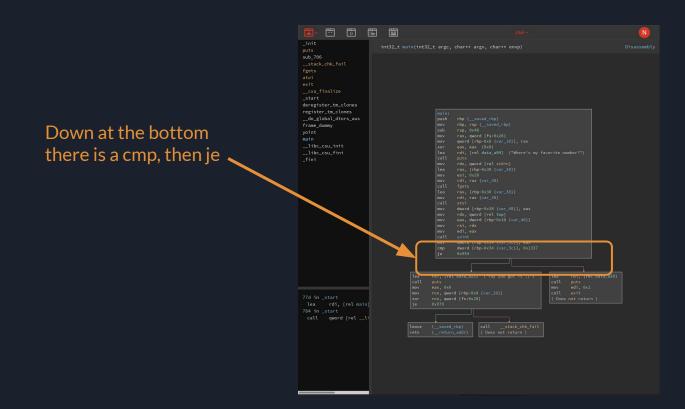


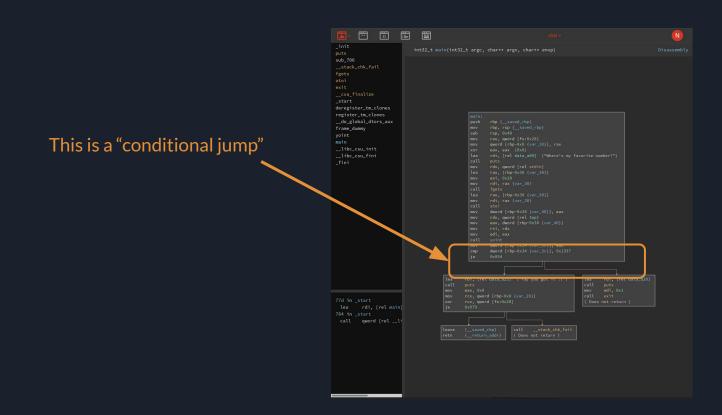


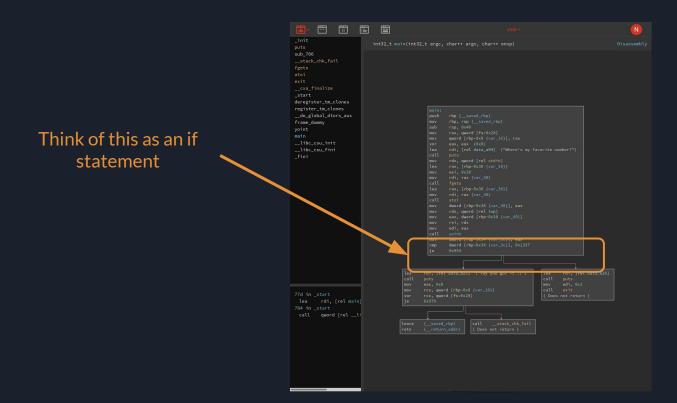
#### Graphs!

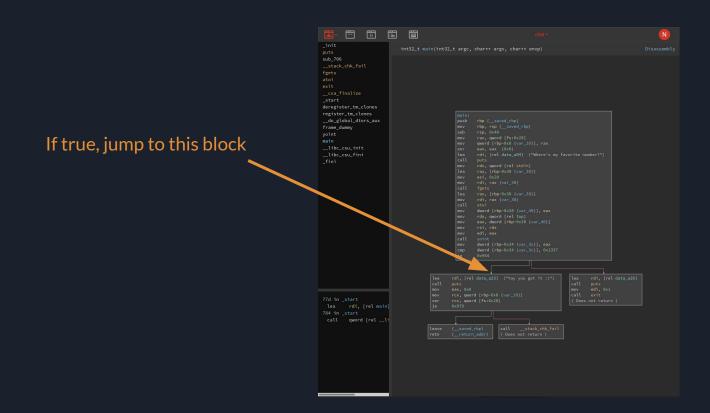


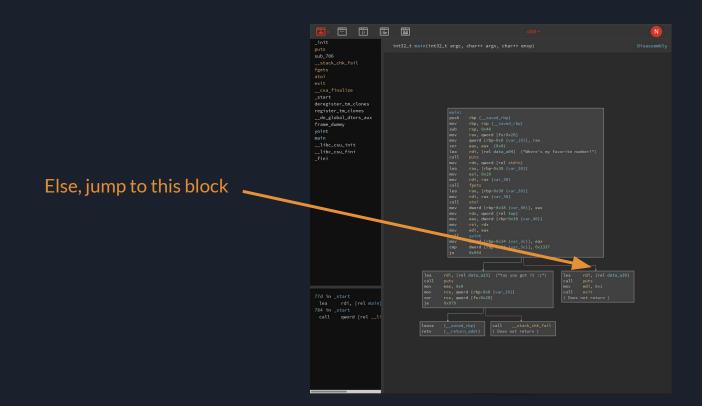




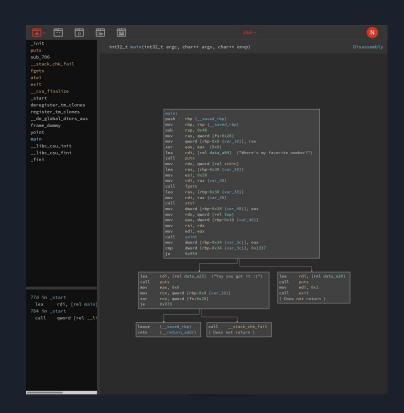








This type of "graph" view is the most peoples prefered way to read assembly



# Godbolt.org

Connecting it back to C/C++

Questions?