

Arhitektura računara 1

skripta za vežbe sa zadacima

Oktober 2023

Sadržaj

1	Uvod	1
1.1	Registri	2
1.2	Prosledjivanje argumenata	3
1.3	GDB debugger	4
1.4	Hello world program	5
1.4.1	Intelova sintaksa	5
1.4.2	Sekcija sa podacima	6
1.4.3	Tekst i global	7
1.4.4	Enter, leave i ret	7

1 Uvod

U ovoj skripti akcenat će biti na 64-bitnoj varijanti x86-arhitekture, x64 i operativnom sistemu Linux.

x64 je generičko ime za 64-bitnu ekstenziju Intelovih i AMD-ovih skupova instrukcija 32-bitne x86-arhitekture. AMD je predstavio prvu verziju x64, inicijalno nazvanu x86-64, a kasnije su je preimenovali u AMD64. Intel je svoju implementaciju nazvao IA-32e, a potom EMT64. Postoje neznatne razlike između ove dve verzije, ali većina koda radi dobro u obe.

Asembler je bilo koji programski jezik niskog nivoa sa jakim vezom između instrukcija u samom jeziku i instrukcija mašinskog jezika koje odgovaraju samoj arhitekturi. Svaki assembler ima različitu sintaksu. Čak i između assemblera koji imaju "Intelovu sintaksu" odnosno koji su kompatibilni sa Intelovim procesorima, postoji razlika. Na primer, za rad sa x86 procesorima možemo da koristimo sledeće asemblere: MASM, NASM, YASM, itd.

GNU assembler, poznatiji kao gas ili as je assembler razvijan od strane GNU projekta. On je podrazumevani "deo u pozadini" GCC kompajlera. Takođe je i cross-platform, što znači da može da se izvršava na velikom broju različitih arhitektura. Na početku je podržavao samo AT&T sintaksu, a od verzije 2.10

podržava i Intelovu sintaksu. Na ovom kursu radićemo GNU assembler i Intelovu sintaksu.

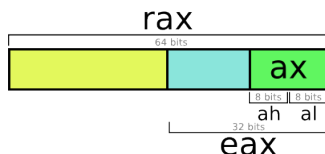
1.1 Registri

x86 ima 8 registara opšte namene (eax, ebx, ecx, edx, ebp, esp, esi, edi). Pošto je x86 32-bitna arhitektura, to znači da adrese u memoriji možemo da zapišemo u 32 bita, odnosno da ukupno možemo da zapišemo 4GB RAM-a. Takođe, veličina registara je 32 bita. U x64 ovi registri su prošireni do 64 bita. Novodobi-jeni registri imaju isti naziv osim što je umesto slova 'e' prvo slovo 'r' i njihova donja polovina (niži bitovi) su odgovarajući x86 registri. Npr. prošireni registar rax u x64 ima kao svoju donju polovinu eax, dok prošireni registar rbx ima kao svoju donju polovinu rbx.

U x86 nisu svi registri zaista bili registri opšte namene, jer su se neki im-plicitno koristili u specifične svrhe. Npr. registri ebp i esp za rad sa stekom. U x64, njima odgovarajući registri rsp i rbp zadržavaju tu namenu.

Detaljnije o registrima:

- **rax** - Povratne vrednosti funkcije se skladište u njemu, niža polovina mu je eax, dok je niža polovina eax-a ax. Registar od samo 16 bitova ax, dovoljan je da se u njemu smesti npr. podatak tipa short int, i ima dve svoje polovine - višu ah i nižu al, koje su po 8 bitova i u njima može da stane podatak tipa char.



- **rbx** - Registar koji mora da se sačuva pre upotrebe. Nakon što završimo rad sa njim, mora da ponovo ima vrednost koju je imao pre nego što smo ga koristili. Niža polovina mu je ebx, čija je niža polovina bx. Registar bx ima višu polovinu bh i nižu bl.
- **rcx** - Tipični registar opšte namene. Često se koristi kao brojač. Niža polovina mu je ecx, čija je niža polovina cx. Registar cx ima višu polovinu ch i nižu cl.
- **rdx** - Registar opšte namene. Niža polovina mu je edx, čija je niža polovina dx. Registar dx ima višu polovinu dh i nižu dl.
- **rsp** - Registar koji pokazuje na vrh steka (detalji u nastavku). Niža polovina mu je esp, čija je niža polovina sp. Niža polovina registra sp je spl. Kada koristimo ovaj registar moramo da budemo veoma pažljivi. Nije preporučljivo koristiti ga za druge namene, jer onda dolazi do grešaka u radu sa stekom.

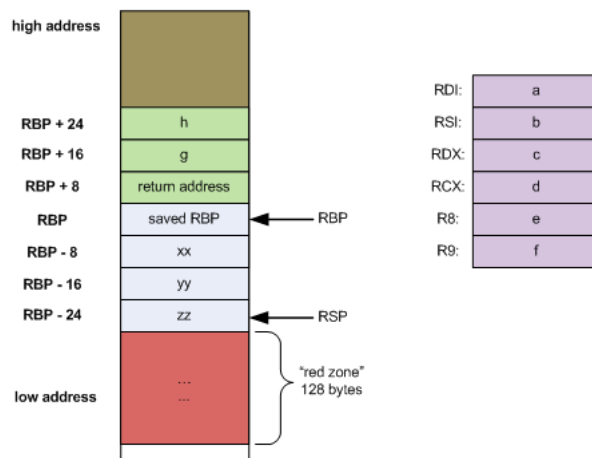
- **rbp** - Ponekad se koristi da čuva staru vrednost steka ili "bazu". Preciznije, kada smo u pozivu funkcije (što podrazumeva da je na stek stavljen okvir funkcije i da je rsp promenjen), moramo da zapamtimo gde smo stali na steku kada se funkcija završi i ukloni sa steka. Za to nam služi rbp. Niža polovina mu je ebp, čija je niža polovina bp. Niža polovina registra bp je bpl. Nije preporučljivo koristiti ga za druge namene, jer onda dolazi do grešaka u radu sa stekom.
- **rsi** - Registar opšte namene. Koristi se za prenos drugog argumenta funkcije. Niža polovina mu je esi, čija je niža polovina si. Niža polovina registra esi je sil.
- **rdi** - Registar opšte namene. Koristi se za prenos prvog argumenta funkcije. Niža polovina mu je edi, čija je niža polovina di. Niža polovina registra edi je dil.
- **r8, r9, r10, r11** - Registri opšte namene. Niža polovina r8 je r8d. Niža polovina r8d je r8w. Niža polovina r8w je r8b. Sufiksi d,w,b su skraćenice od dword, word, byte i opisuju koliko memorije zauzimaju. Slično se određuju delovi registara r9, r10 i r11.
- **r12, r13, r14, r15** - Registri koje bi trebalo sačuvati, odnosno ako radimo sa njima, trebalo bi ih nakon što završimo vratiti na vrednosti koje su imali pre nego što smo počeli rad sa njima. Njihovi delovi se određuju slično kao za registar r8.

1.2 Prosledjivanje argumenata

Prvih šest celobrojnih argumenata funkcije (od kojih su neki potencijalno pokazivači, jer pokazivači čuvaju adrese (koje su celi brojevi) i zauzimaju isti prostor kao neki celobrojni podatak) prenose se redom u registrima rdi, rsi, rdx, rcx, r8 i r9. Ako imamo više od šest ovakvih argumenata onda se oni prenose preko steka. Najbolje je da pogledamo na primeru. Za ovu funkciju stek izgleda na sledeći

```
long myfunc(long a, long b, long c, long d,
            long e, long f, long g, long h)
{
    long xx = a * b * c * d * e * f * g * h;
    long yy = a + b + c + d + e + f + g + h;
    long zz = utilfunc(xx, yy, xx % yy);
    return zz + 20;
}
```

način: Ne smemo da pokvarimo rbp zato što nam rbp određuje poziciju na steku iznad koje ne smemo da diramo podatke. Preciznije, mi znamo da se na adresi rbp+8 nalazi adresa povratka iz funkcije. Ako to "pokvarimo" nećemo znati gde treba da se vratimo iz funkcije. Takođe, pošto smo imali osam argumenata funkcije, a ne šest, dva su morala da budu preneti preko steka. Oni se redom nalaze iznad adrese povratka, na adresama rbp+16, rbp+24, itd. (**Ovde treba biti pažljiv. U pitanju su bili podaci tipa long i oni zauzimaju 8**



bajtova, da je bio u pitanju podatak tipa `int` zauzeo bi samo 4 bajta, i ne bi išao na adresu `rbp+16`, nego na `rbp+12`, jer bi bio 4 bajta udaljen od adrese povratka koja je na `rbp+8`).

Ostaje da kažemo nešto i o crvenoj zoni koja je jedan vid optimizacije. Možemo da pretpostavimo da 128 bajtova ispod `rsp` neće biti narušene raznim signalima i prekidima, te možemo da koristimo taj prostor za međjuračunanja, bez da eksplicitno pomeramo vrh steka. Treba imati na umu da ova zona može da bude pogodjena pozivima drugih funkcija, te je najbolje da je koristimo u funkcijama koje ne sadrže pozive drugih funkcija u sebi. Pogledajmo kako "radi" crvena zona na primeru funkcije koja u sebi nema pozive drugih funkcija: Izgled

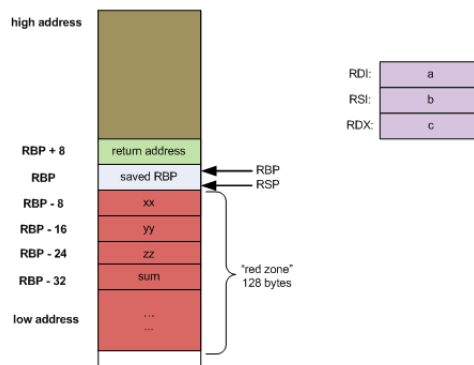
```
long utilfunc(long a, long b, long c)
{
    long xx = a + 2;
    long yy = b + 3;
    long zz = c + 4;
    long sum = xx + yy + zz;

    return xx * yy * zz + sum;
}
```

steka za ovu funkciju dat je na slici ispod. U funkciji `utilfunc`, nema prenosa promenljivih po steku, jer argumenata funkcije ima manje od šest. Takođe, zbog nedostatka poziva ka drugim funkcijama, moguće je da se rezultati nekih izračunavanja (ovde sve promenljive za međjukurake zauzimaju manje jednako od 128 bajtova) sačuvaju u crvenoj zoni.

1.3 GDB debugger

Ako smo napisali naš program 1.s koristeći assembler Intel64 arhitekture (x64), taj program ćemo prevesti sledećim dvema naredbama:



```
as -gstabs -o 1.o 1.s
gcc -g -Wall -pedantic -no-pie -o 1 1.o
```

Opcije -g i -gstabs odnose se na formate za debugovanje. Opcija -g podrazumeva naivni format za debugovanje koji pruža operativni sistem i ovakav format za debugovanje podržan je od strane GDB debagera koji ćemo koristiti. Što se tiče -gstabs opcije, na nekim sistemima GNU asembleri zahtevaju ovu opciju. Opcija -pedantic uključuje dodatne ekstenzije i generiše detaljnija upozorenja. Opcija -no-pie omogućava da se ne generiše poziciono nezavisan izvršni fajl. Pie je bezbednosni preduslov da svaki put kernel učitava binarni fajl i zavisnosti u proizvoljnu adresu virtuelne memorije. Često zahteva dodatno programiranje.

GDB debugovanje pokrećemo u terminalu sa: `gdb ./1` gde je 1 generisani izvršni program. Ako želimo da pokrenemo program potrebno je da unesemo komandu `start`. Ako želimo da se pomerimo na sledeću mašinsku instrukciju koristimo komandu `stepi`. Ako želimo da udjemo u poziv funkcije i proverimo šta se u njoj dešava koristimo instrukciju `step`. Za štampanje vrednosti registara npr. registra `rax` kucamo `info registers $rax`.

1.4 Hello world program

Započnimo naš put ka učenju GNU assemblera najjednostavnijim programom - Hello world. U stanju smo da sa slike 1 identifikujemo nekoliko važnih činjenica koje ćemo postupno izložiti.

1.4.1 Intelova sintaksa

U prvoj liniji hello world programa vidimo da stoji navedeno **.intel_syntax noprefix**. Ova jednostavna komanda nam govori da je sintaksa koju ćemo koristiti Intelova sintaksa bez prefiksa, a ne AT&T sintaksa koju podrazumevano koristi GCC kompajler. Neke od razlika su sledeće:

- U AT&T na nazive registara dodaje se prefiks %, a na nazive konstanti prefiks #
- Redosled operandu u instrukcijama je različit. U Intelovoj sintaksi prvi operand je odredišni (engl. *destination*), a drugi izvorni (engl. *source*), dok je u AT&T sintaksi obrnuto.
- Kada u Intelovoj sintaksi želimo da pristupimo vrednosti koja se nalazi na adresi zapisanoj u registru rax, to činimo na sledeći način sa uglastim zagradama: [rax]. U AT&T sintaksi, zapis bi bio (%rax).
- U AT&T sintaksi, nazivi instrukcija dobijaju odgovarajući sufiks, koji zavisi od veličine operandu. Npr. ako želimo da u Intelovoj sintaksi prebacimo neku vrednost iz registra al u bl, ili iz registra ax u bx, bez obzira na to što al i bl zauzimaju različit broj bitova od ax i bx, naziv instrukcije bi bio isti - mov. Zapisali bismo te instrukcije na sledeći način:

mov bl, al

mov bx, ax

Sa druge strane, u AT&T sintaksi, prva instrukcija bi bila:

movb al, bl

jer su al i bl veličine jednog bajta, i kao što smo ranije naučili, redosled operandu je obrnut od onog u Intelovoj sintaksi. Druga bi bila:

movw ax, bx

jer su ax i bx veličine 16 bitova. Oznaka 'w' je za broj bitova, odnosno za word - 4 bajta. Za long vrednosti - 8 bajtova, koristi se sufiks 'l'.

1.4.2 Sekcija sa podacima

U jednostavnom hello world programu prikazana su dva načina da se definiše promenljiva koja će u sebi sadržati tekst "Hello world". Drugi način je stavljen pod komentar koristeći oznaku #. U zavisnosti od toga da li dozvoljavamo menjanje promenljivih koje definišemo, možemo da se opredelimo za stavljanje te promenljive u sekciju za read-only podatke (podatke koji se samo mogu čitati, ali se njihove vrednosti ne mogu menjati), datu sa **.section .rodata** ili u sekciju sa podacima koji se mogu menjati **.data**. Podaci u read-only odeljku se moraju inicijalizovati, dok je za podatke u odeljku .data, preporuka da se inicijalizuju. Promenljive se definišu imenom iza koga sledi dvotačka, tip i vrednost. Tip promenljive može da bude .string, .ascii, .asciiz, .double, .float, .byte, .word, .int, .long, itd. Direktive .ascii, .asciz i .string na prvi pogled mogu delovati identično, ali to nije u potpunosti tačno. Naime, intuitivno, .ascii je slično char[], te se kada mu dodelimo npr. reč "hello" to prevodi na sledeći kod 0x48

```

1.intel_syntax noprefix
2
3.section .rodata
4Hello: .string "Hello world"
5
6# .data
7# x:      .int 0
8# message: .asciz "Hello world!\n"
9
10.text
11
12.global main
13
14main:
15    enter 0,0
16
17    lea rdi, Hello
18
19    call printf
20
21    leave
22    ret

```

Figure 1: Caption

0x65 0x6c 0x6c 0x6f. Sa `.asciz` je isto, osim što se dodaje nul na kraj, pa dobijamo sekvencu 0x48 0x65 0x6c 0x6c 0x6f 0x00. Što se tiče `.string`, kao i većina drugih gas direktiva, može da ima različita značenja u zavisnosti od toga na kom sistemu radimo. Tako na nekim sistemima `.string` je isto što i `.ascii`, a na nekim isto što i `.asciz`.

1.4.3 Tekst i global

Kao i `.data` i `.section .rodata`, `.text` i `.global` su asemblerske direktive zadužene za kontrolu sekcija. Direktiva `.text` nam govori da počevši od nje kreće kod. Direktiva `.global` zadužena je da u fazi linkovanja, ukoliko budemo asemblersku funkciju `main` pozivali iz nekog drugog programa (npr. c-ovskog 1.c) taj program zna da je definicija funkcije `main` u našem fajlu 1.s, odnosno `.global` čini funkciju `main` vidljivu drugim fajlovima koji se linkuju sa našim. Nakon **`.global main`** vidimo labelu **`main`** koja nam govori odakle počinje naša funkcija `main`. Labele možemo koristiti i za druga mesta koja po našem mišljenju čine neku celinu i na njih možemo skakati naredbama uslovnih i bezuslovnih skokova.

Ono što je posebno važno da imamo na umu jeste da unutar jednog asemblerskog fajla možemo imati i više funkcija, i svaku od njih možemo "proglasiti" za `global`.

1.4.4 Enter, leave i ret

Svaka funkcija koju budemo napisali u GNU assembleru počinje instrukcijom `enter`, a završava se instrukcijama `leave` i `ret`. Zato ćemo u ovom odeljku posebnu

```

push rbp
mov rbp, rsp
sub rsp, numVars

```

Figure 2: Caption

pažnju posvetiti objašnjenju kako one rade. Krenimo od `enter`. Instrukcija **`enter varNum, 0`** je zamena za niz instrukcija na slici 2.

U registru `rbp` čuva se adresa na steku iznad koje ne smemo ništa da diramo, jer ta memorija ne pripada nama. Kada neka funkcija (recimo `main`) poziva našu funkciju (recimo `prime`), na steku se formira novi stek okvir za nas. Funkcija `main` ima svoju vrednost `rbp-a`, odnosno ona "zna" gde se nalazi memorija koja njoj ne pripada. U trenutku kada ona pozove `prime`, `prime` ne sme da menja memoriju koja pripada `main-u` i ne sme da izgubi informacije koje su zbog nje stavljene na stek (parametre koji nisu mogli da se prenesu preko registara, adresu povratka...). Tako da i `prime` treba da postavi svoju vrednost `rbp-a` na trenutnu vrednost vrha steka koja se nalazi u `rsp`. Medjutim, ako to odmah uradi, izgubiće vrednost `rbp-a` funkcije `main`, i po povratku u funkciju `main` nakon što se `prime` završi može se dogoditi da funkcija `main` pristupi memoriji kojoj ne bi smela. Zato je važno da pre nego što promenimo vrednost registra `rbp` u adresu na steku koja je nama potrebna, sačuvamo stari `rbp` na stek. Nakon što ga stavimo na stek, `rsp` se automatski ažurira i pokazuje na novi vrh steka. Nakon toga možemo da `rbp-u` dodelimo tu vrednost, što je za nas signal da funkcija `prime` neće promeniti vrednosti na steku i u memoriji koje ne bi smela.

Treća instrukcija na slici 2 služi da "odvoji" memoriju na steku za naše lokalne promenljive. Izrazito je korisna ukoliko nam je potrebno da imamo niz fiksne veličine. Konstanta `numVars` je broj bajtova koji nam je potreban da se odvoji i važno je da na x64 arhitekturi to bude broj koji je parni umnožak od osmice, kako bi stek bio poravnat (o tome u nastavku skripte).

Pogledajmo primer na slici 3. Kada u funkciju udjemo sa **`enter 16,0`**, to znači da smo alocirali 16 bajtova za naše promenljive. Ovde ćemo da stavimo niz od dva elementa. Naš `rsp` je bio jednak `rbp-u`, a potom se od njega oduzelo 16, te je on ustvari postao `rbp-16`. Nulti član niza se upisuje na adresu `rsp`, prvi član niza na adresu `rsp+8`. Da smo ih imali više upisivali bismo ih redom na adrese `rsp+16`, `rsp+24`, `rsp+32`, itd. Ovo može delovati da je u suprotnosti sa pravilom opadajućeg steka gde novi podaci dobijaju manje adrese, medjutim to nije tačno. Naime, moramo ispoštovati princip stavljanja nizova u memoriju, a to je da se niz pakuje u memoriju u jednom komadu, i da je element `a[i]` na adresi `a+i`. Odnosno, nulti član je na adresi `a`, prvi na `a+1`, drugi na `a+2`, itd. gde je uvećavanje za 1 zapravo pokazivačka aritmetika, odnosno uvećavanje za jedan zapravo predstavlja uvećavanje za broj bajtova koji promenljiva zauzima što je u našem slučaju 8. Više o tome možemo videti na slici 4.

Objasnimo detaljnije kod sa slike 3. Čitaocu savetujemo da se na ovo objašnjenje vrati nakon što pročita deo o `leave` i `ret`. Kako će se vrednost registra `rsp` menjati u toku programa, najbolje bi bilo da tu vrednost zapam-

```

1.intel_syntax noprefix
2
3.section .rodata
4Hello: .string "Hello world\n"
5
6.section .data
7    x: .int 0
8    message: .asciz "Hello world!\n"
9    fmt: .asciz "%d\n"
10.text
11
12.global main
13
14main:
15    enter 16,0
16
17    mov r8, rsp
18
19    mov qword ptr [r8],100
20    mov qword ptr [r8+8], 2
21
22    push r8
23    push r8
24    lea rdi, Hello
25    call printf
26    pop r8
27    pop r8
28    mov rax, [r8]
29
30
31    lea rdi, fmt
32    mov rsi, rax
33    call printf
34
35    mov rax, [rsp+8]
36
37|
38    lea rdi, fmt
39    mov rsi, rax
40    call printf
41
42    leave
43    ret

```

Figure 3: Caption

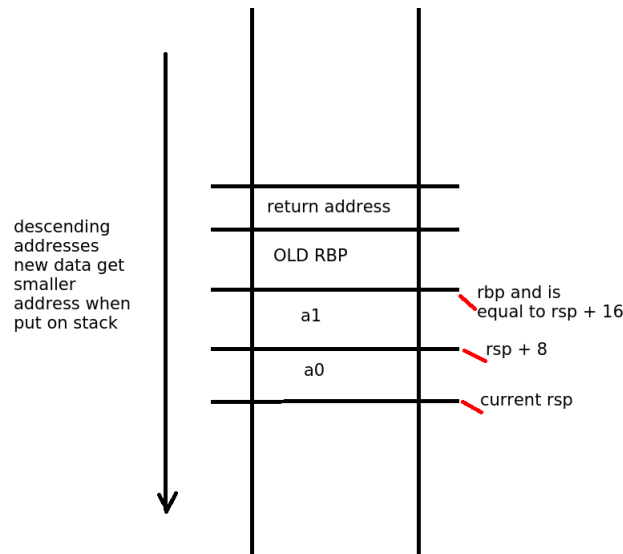


Figure 4: Caption

timo u nekom drugom registru, npr. `r8` i to činimo u 17. liniji instrukcijom `mov`. Instrukcije u linijama 19. i 20. su karakteristične zato što opisuju donošenje podatka u memoriju koji može da bude zapisan sa više ili manje bajtova (kada napišemo broj 100, možemo da mislimo na `int` ili na `long` ili na `short`), te je zato potrebno da preciziramo tačan broj bajtova koji naš podatak zauzima. Ukoliko to ne uradimo dobićemo sledeći tip greške: *ambiguous operand size for instruction mov*. Problem se može rešiti na dva načina. Možemo da pre 19. i 20. linije stavimo 100 i 2 u registre, čime se oni zapisuju u onoliko bajtova koliko taj registar zauzima, i da onda na adresu zapisanu u `r8` donesemo vrednost tog registra. Npr.

```
mov rax,100
mov [r8],rax
```

Drugi način je da se ispred adrese precizira i veličina podatka koji ćemo doneti. U ovom slučaju želimo da podatak zapišemo na osam bajtova pa ćemo koristiti `qword`. Za 4 bajta bi bilo `dword`, za dva bajta `word`, i za jedan byte. Ne zaboravimo da nakon toga napišemo `ptr`, što je skraćenica za pokazivač. Odnosno, `r8` je pokazivač na taj memorijski blok, veličine koju smo odredili.

Može se dogoditi da se u pozivu funkcije `printf` pokvari vrednost registra `r8`, te je zato potrebno sačuvati taj registar pre ovog poziva i dohvatiti ga nakon poziva. Stavili smo dva `push-a`, jer želimo da nam adrese na steku stalno budu poravnate (u nastavku skripte više o tome), odnosno da stalno budu parni umnožak osmice.

`Leave` i `ret` su zamena za instrukcije na slici 5. Prva komanda na toj slici nam govori o tome da mi najpre brišemo sve lokalne promenljive. Potom vraćamo

stari rbp jer se pripremamo da se vratimo u funkciju čiji je to rbp bio i želimo da ona ima informaciju o tome čiju memoriju ne sme da pokvari. Nakon što smo stari rbp skinuli sa steka, rsp nam pokazuje na adresu povratka. Instrukcija ret uzima adresu sa steka na koju pokazuje rsp i vodi nas na nju. Nakon toga se nastavlja izvršavanje funkcije koja je pozvala našu.

```
mov rsp, rbp
pop rbp
ret
```

Figure 5: Caption