УНИВЕРЗИТЕТ У БЕОГРАДУ
МАТЕМАТИЧКИ ФАКУЛТЕТ

Сретен Ковачевић

V8 - ИМПЛЕМЕНТАЦИЈА НЕОПТИМИЗУЈУЋЕГ WEBASSEMBLY КОМПИЛАТОРА ЗА MIPS АРХИТЕКТУРУ

мастер рад

Београд, 2018.

Ментор:

др Филип Марић, ванредни професор
Универзитет у Београду, Математички факултет

Чланови комисије:

др Милена Вујошевић-Јаничић, ванредни професор
Универзитет у Београду, Математички факултет

др Милан Банковић, доцент
Универзитет у Београду, Математички факултет

Датум одбране: _____

Наслов мастер рада: v8 - имплементација неоптимизујућег WebAssembly компилатора за MIPS архитектуру

Резиме:

Кључне речи:

Садржај

Глава 1

Увод

Глава 2

Архитектура *MIPS*

У овој глави је описана архитектура *MIPS* процесора. У поглављу 2.1 су описане архитектура процесора *CISC* (скраћено од енгл. *Complex Instruction Set Computing*) и *RISC* (скраћено од енгл. *Reduced Instruction Set Computing*) и њихове разлике, а у поглављу 2.2 *MIPS* архитектура. У поглављима 2.3 и 2.4 описане су инструкције и регистри *MIPS* архитектуре. У поглављу 2.5 описан је механизам проточне обраде и њена имплементација на *MIPS* архитектури, а у поглављлу 2.6 је представљен слот закашњења.

2.1 *CISC* и *RISC* архитектура

Архитектура у рачунарству представља спој организације (начин комуникације међу различитим деловима рачунара), хардвера (примена конкретних логичких кола) и скуп инструкција и регистара *ISA* (скраћено од енгл. *Instruction Set Architecture*) [3]. Да би разумели разлику између *CISC* (скраћено од енгл. *Complex Instruction Set Computing*) и *RISC* (скраћено од енгл. *Reduced Instruction Set Computing*) архитектуре довољно је да посматрамо последњу ставку.

Процесори дизајнирани по *CISC* архитектури карактеришу се великим бројем инструкција које су на раслопагању програмеру. Овим дизајном се смањује број наредби у програму, по цену броја циклуса по инструкцији. Смањењем броја инструкција по програму постиже се мања потрошња меморије [1]. Инструкције често имају различиту дужину записа, а могу и обављати неколико независних задатака. Оваква архитектура захтева сложен хардвер, што може довести до потешкоћа у разумевању и програмирању оваквих чипова. Пружају могућност великог броја различитих начина адресирања. Овакви процесори

се користе на личним рачунарима, радним станицама и серверима. Типичан пример је *Intel x86* серија процесора.

RISC архитектура користи високо оптимизован скуп инструкција. Мотив је супротан у односу на мотив *CISC* архитектуре. Смањује се број циклуса по инструкцији, али се зато добија мањи број инструкција, те је потребан већи број инструкција по програму. Јединствена карактеристика *RISC* архитектуре је проточна обрада (енгл. *Pipelining*). Проточна обрада се постиже преклапањем извршавања већег броја инструкција. Захваљујући томе постижу се боље перформансе у поређењу са *CISC* процесорима [1]. Како је број инструкција које су подржане мали, имплементација самог чипа је знатно једноставнија и јефтинија. Подржана су четири начина адресирања: регистарско, *PC*-релативно, псеудо-директно и базно. Процесори *RISC* архитектуре се користе у наменским уређајима (енгл. *embedded*) и примери су *ARM* и *MIPS*.

2.2 *MIPS*

MIPS је представник *RISC* архитектуре, настао средином осамдесетих година двадесетог века на Станфорд универзитету. Група студената, предвођена Џоном Хенесијем, је истраживала рад *RISC* процесора и открила да се бољом применом проточне обраде, која је до тад била недовољно искоришћена, може доћи до бржих процесора на мањем чипу. *MIPS* је током година успео да се одржи на тржишту, али и да сачува епитет једне од најједноставнијих архитектура [4].

Процесори из *MIPS* породице успели су да пронађу примену у наменским уређајима. Могу се наћи у мобилним уређајима, мрежним уређајима, сет-топ боксовима, паметним телевизорима. Све претходно наведене примене захтевају покретање апликација које захтевају интензивна израчунавања (процесирање слика и видеа, анализа података, интеракција међу субјектима, итд.).

2.3 Инструкције

У *MIPS* асемблерском језику све инструкције су једнаке дужине 32 бита. Могу се поделити у следеће групе [5]:

- аритметичке: сабирање, одузимање, множење, дељење

- логичке: и, или, шифтовање
- приступ меморији: учитавање, записивање у меморију
- гранања и скокови
- контролне

На основу типова операнада, инструкције се могу поделити у три типа [4]:

R - Инструкције које као операнд очекују регистре. Представљају се у следећем формату:

$$OP\ rd, rs, rt$$

OP представља ознаку инструкције, rd је регистар за смештање резултата, док су регистри rs и rt операнди. Неки од примера ових инструкција су:

- *jr* - скакање на адресу смештену у регистру
- *slt* - поставља 1 у регистар уколико је први аргумент мањи од другог
- *addu* - смештање у регистар збира аргумената, посматрајући аргументе као неозначене целе бројеве

I - Инструкције које као операнде имају регистар и константну вредност, у облику специјалне вредности која је уписана у инструкцију (енгл. *Immediate*), представљају се форматом:

$$OP\ rd, rs, Imm$$

OP представља ознаку инструкције, rd је регистар за смештање резултата, rs први операнд (регистар), а Imm представља константу која је други операнд. Константа може имати највише 16 бита. Примери ових инструкција су:

- *lw* - учитавање вредности са адресе $rs + Imm$ у одредишни регистар
- *sw* - смештање вредности из одредишног регистра на адресу $rs + Imm$
- *beq* - гранање уколико је вредност у регистру једнака константи
- *addiu* - смештање у регистар збира вредности из регистра и константе

J - Инструкције које се користе при скоковима. Представљају се следећим форматом:

j label

Постоје две инструкције овог типа, а то су *j* и *jal*. Код прве инструкције (енгл. *Jump*) се ток извршавања пребацује на позицију *label*, исто се дешава и са другом инструкцијом (енгл. *Jump and link*), али се и адреса наредне инструкције уписује у *$ra* регистар. Ове инструкције прихватају највеће константе, које су дужине 26 бита, што је оправдано великим бројевима којим се представљају адресе.

2.4 Регистри

Регистри представљају малу, веома ефикасну меморију која се налази у процесору. *MIPS* инструкције могу као аргументе да примају једино регистре и специјалне константе.У *MIPS* архитектури постоје 32 регистра опште намене, од којих два имају другачије понашање [4]:

$0 - Увек враћа нулу, без обзира шта се у њега уписује

ra - Користи се за смештање адресе повратка из функције приликом коришћења *jal* инструкције Сви остали регистри могу се равноправно користити у инструкцијама (чак се и регистар $0 може користити, али ће ре резултат у том случају бити занемарен).

У наставку ће бити описани регистри, као и њихово препоручено коришћење:

at - Резервисан за псеудоинструкције које генерише асемблер.

v0, v1 - Користе се за смештање целобројних повратних вредности функција. Уколико се резултат не може сместити у два регистра, компилатор ће резултат сместити у меморију, а адресу у ове регистре.

a0 - a3 - Користе се за смештање прва четири целобројна аргумента при позиву функција. Остали се смештају на стек.

t0 - t9 - Привремени регистри, није потребно рестаурирати вредност након коришћења.

s0 - s7 - Садржај ових регистара мора остати непромењен након сваке функције, што се постиже привременим чувањем њиховог садржаја на стеку уколико се користе. Дужност да сачува њихову вредност има позвана функција (енгл. *callee saved registers*).

k0, k1 - Резервисани за системе прекида оперативног система, иначе се ретко користе.

gp - Има два начина примене. Уколико се ради о коду који не зависи од позиције (енгл. *Position Independent Code* скраћено *PIC*), овај регистар показује на табелу показивача (енгл. *Global Offset Table*). Уколико је у питању регуларан код, показује на средину у статичкој меморији. На тај начин се помоћу једне инструкције може приступити било ком податку који је 32KB лево или десно од њега. Овај регистар не користе сви системи за компилацију и сва окружења за извршавање.

sp - Показивач на стек. Стање стека је потребно експлицитно ажурирати, те се инструкције за одржавање показивача на стек углавном генеришу на почетку и на крају функција. Како стек расте надоле, на почетку функције се поставља на најнижу тачку до које ће стек расти.

fp - Показивач на стек оквир. Користи се од стране функције, за праћење стања на стеку. Уколико се при превођењу не може одредити на коју вредност да се постави **sp** регистар, променљивим на стеку се приступа помоћу овог регистра.

ra - Подразумевани регистар за смештање адресе повратка из функције. Овакво понашање је подржано кроз одговарајуће инструкције скока. Ово је разлика у односу на *x86* архитектуру, где се адреса повратка смешта на стек. Функције се углавном завршавају наредбом *jr $ra*. Иако се може користити и било који други регистар, то се не препоручује због оптимизација које врши процесор у случају коришћења овог регистра. Функције које позивају друге функције морају сачувати његову вредност.

2.5 Проточна обрада

Проточна обрада (енгл. *pipelining*) почива на чињеници да различите фазе извршавања користе различите ресурсе. Уколико имамо систем у ком је свака фаза једнаке дужине, добили би систем код ког би на крају завршетка фазе за једну инструкцију у ту фазу ушла следећа инструкција [4]. Да би овакав систем био могућ, процесори *RISC* архитектуре бирају минималан скуп инструкција које имају приближно исто време извршавања у свакој фази. Такође, инструкције су исте дужине како би се осигурало да је фаза декодирања идентична у свакој фази. Оваква конфигурација може се видети и у *MIPS* архитектури [4].

Како би проточна обрада била ефикасна користи се кеш меморија, чиме се убрзавају приступи меморији. Кеш меморија је мала, веома брза, локална меморија у којој се налази копија података из меморије [4]. У кешу се чувају подаци које је процесор најскорије користио, док се најстарији подаци преписују (уколико је кеш попуњен). Када процесор у кешу не пронађе потребне податке („промашај кеша", дешава се у 10% случајева), тада се приступа меморији.

Код *RISC* архитектуре, кеш је уско повезан за процесор и активно се користи за имплементацију проточне обраде, док се код *CISC* архитектуре кеш посматра као део меморијског система. *MIPS* има одвојен кеш података и инструкцијски кеш, што омогућава симултано читање инструкција и уписивање или читање података.

MIPS инструкције су подељене у пет фаза, и трајање сваке фазе је фиксирано. Прва, трећа и четврта фаза трају по један такт процесора, док друга и пета захтевају пола такта за своје извршење [4]. У наставку је описана свака од фаза.

1. Дохватање инструкције из инструкцијског кеша и њено декодирање
2. Читање садржаја наведених регистара
3. Извршавање аритметичко/логичких операција у једном такту (операције у покретном зарезу, множење и дељење су сложеније и раде се другачије)
4. Дохватање и уписивање у меморију. У 75% случајева инструкције не раде ништа у овој фази, али она постоји да не би више инструкција чекало на приступ кешу података
5. Резултат операције се уписује у одредишни регистар

2.6 Слот закашњења

Слот закашњења (енгл. *Delay slot*) је најосетнији ефекат проточне обраде из угла програмера. Због структуре проточне обраде на *MIPS* архитектури (која је описана у поглављу 2.5), у тренутку када наредбе гранања или скока дођу до фазе извршавања, рад на наредној инструкцији ће већ бити започет, иако је ток извршавања потенцијално промењен. Започета инструкција се извршава без обзира на исход наредбе промене тока извршавања и на тај начин се започети посао не одбацује.

Како би се постигло да се у слоту закашњења не појави више од једне инструкције, наредбе гранања имају посебно понашање при ком се већ после пола такта у фази извршавања аритметичко/логиких операција зна где ће се извршавање наставити. Како и друга фаза траје пола такта, овим смо обезбедили да само једна инструкција може доспети до прве фазе, за чије обављање је потребан један такт. Оваква конфигурација пружа могућност програмеру или компилатору да промени редослед инструкција у програму и тако неко израчунавање смести у слот закашњења [4].

Упркос уштеди коју слот закашњења може да донесе, он представља и потенцијални ризик. Посебно треба истаћи условна гранања, у којима нека операција не треба да буде извршена у оба случаја. Некад је безбедније (или једино исправно) оставити у слоту закашњења инструкцију *nop*.

Још једна последица проточне обраде је и слот закашњења учитавања (енгл. *load delay slot*). Подаци дохваћени *load* инструкцијом постају расположиви тек након инструкције која следи иза ње. Стога се њен резултат не може користити у следећој инструкцији. Модерни процесори имају механизам блокирања резултата *load* инструкције. Уколико резултат проба да се искористи у следећој инструкцији, процесор ће зауставити извршавање док резултат не буде спреман. На ранијим верзијама такав код је имао недефинисано понашање [4].

Глава 3

WebAssembly

WebAssembly (скраћено *Wasm*) је безбедан и преносив код ниског нивоа. Дизајниран је са идејом да обезбеди ефикасно извршавање и компактну репрезентацију. Главни циљ му је да омогући функционисање апликација високог нивоа на Вебу, али и да не прави никакве претпоставке о окружењу нити уводи нове функционалности, што би га чинило погодним за коришћење у оквиру других окружења [2]. У овој глави ће бити описан *WebAssembly* и његове карактеристике. У поглављу 3.1 ће бити приказан дизајн *WebAssembly*-ја, као и карактеристике које са њим долазе, а у поглављу 3.2 је описана његова структура. У поглављу 3.3 је дат приказ семантичких фаза.

3.1 Дизајн

Пред *WebAssembly* су постављени одређени циљеви у погледу дизајна. Два основна циља су брза, безбедена и преносива семантика и ефикасна и преносива репрезентација. Оба циља са својим испуњењем језику пружају карактеристике неопходне да се постигне жељени ниво перформанси.

Из угла семантике *WebAssembly* је [2]:

- Брз - извршава се ефикасношћу која тежи ефикасности језика нижег нивоа (енгл. *native*), користећи предности савременог хардвера
- Безбедан - код се валидира и извршава меморијском сефу, заштићеном окрружењу које спречава угрожавање података или упаде

- Добро дефинисан - потпуно и прецизно дефинише исправне програме и њихово понашање на начин који је лако разумети, како формално тако и неформално
- Хардверски независтан - може се превести на свим архитектурама, персоналним и преносивим рачунарима и наменским уређајима
- Језички независтан - не фаворизује одређени језик вишег нивоа, стил програмирања или објектни модел
- Платформски независтан - може бити уграђен у прегледач, представљати самосталну виртуелну машину или представљати део неког већег окружења
- Отворен - програми могу комуницирати са окружењем користећи једноставне методе

Посматрајући репрезентацију *WebAssembly*-ја, можемо закључити да је [2]:

- Компактан - бинарни формат се брзо преноси захваљујући запису који је краћи и од обичног текста и од кода на језику ниског нивоа
- Модуларан - програм се може поделити на мање целине које се могу слати, кеширати и користити независно
- Ефикасан - може се декодирати, валидирати и компилирати у једном пролазу, како са *JIT* (скраћено од енгл. *Just-in-time*) тако и са *AOT* (скраћено од енгл. *Ahead-of-time*) компилацијом.
- Проточан (енгл. *streamable*) - омогућује да декодирање, валидација и компилација почну пре него што су сви подаци на располагању
- Паралелизабилан (енгл. *parallelizable*) - допушта да декодирање, валидација и компилација буду издељени у више независних паралелних задатака
- Преносив - не прави претпоставке о архитектури које нису широко распрострањене међу модерним хардвером

3.2 Структура

WebAssembly кодира језик ниског нивоа, који је налик на асемблер. Његову структуру чине [2]:

- Вредности
- Инструкције
- Замке (енгл. *Traps*)
- Функције
- Табеле
- Линеарна меморија
- Модули
- Уграђивач (енгл. *Embedder*)

У наставку ће бити описан сваки од елемената структуре.

Вредности

У оквиру *WebAssembly*-ја постоје четири типа вредности. То су целобројне вредности и бројеви у покретном зарезу (имплементирани по стандарду *IEEE 754-2008)*, оба у 32-битној и 64-битној варијанти. 32-битне целобројне вредности се користе и за репрезентацију истинитосних вредности (енгл. *boolean*) и меморијских адреса. На располагању су све уобичајене операције над овим типовима, као и конверзије међу њима. Не постоји разлика између означених и неозначених целих бројева, већ се на основу конкретне операције одлучује како ће се број посматрати [2].

Инструкције

Рачунски модел је заснован на принципима стек машине. Код се састоји од низа инструкција које се редом извршавају. Инструкције врше промене над подацима који се налазе на имплицитном стеку операнада и могу се поделити у две основне категорије. Једноставне инструкције са стека узимају аргумент

и резултат смештају назад на стек. Контролне инструкције мењају ток извршавања програма. Програм је добро структуиран, односно подељен у блокове, петље и условне кодове и наредбе гранања могу да гађају само неке од ових структура [2].

Замке

Неке инструкције, под одговарајућим условима, могу изазвати замку (енгл. *trap*), које прекидају извршавање програма. *WebAssembly* не поседује механизам за обраду замки, већ се оне прослеђују окружењу, где се хватају и обрађују на одговарајући начин [2].

Функције

Код је подељен у одвојене функције. Свака функција прима низ вредности као параметре и враћа низ вредности као резултат[1]. Функције се могу међусобно позивати, укључујући и рекурзивне позиве. Функције могу декларисати локалне променљиве које се могу користити попут виртуелних регистара [2].

Табеле

Табела представља низ вредности неког типа. На тај начин се допушта програму да помоћу индекса индиректно приступи елементу. Тренутно, једини подржани тип је референца на функцију. Захваљујући томе, програм може позивати функције користећи само индекс табеле. Ово опонаша показиваче на функције [2].

Линеарна меморија

Линеарна меморија је непрекидан, променљив низ сирових бајтова. Таква меморија има иницијалну величину, али се може динамички проширити. Програм може учитати или уписати вредност у меморију на адресу било ког бајта (укључујући и непоравнату). Уколико се покуша приступ ван тренутних граница меморије, замка ће бити активирана [2].

[1] У тренутној имплементацији може се вратити само један резултат.

Модули

WebAssembly у свом бинарном запису узима облик модула. Модул садржи дефиниције функција, табела, и линеарне меморије. Такође, може садржати и глобалне променљиве и константе. Дефиниције могу бити увезене тако што ћемо навести модул из ког увозимо и име дефиниције коју увозимо заједно са одговарајућим типом. Опционо, неке дефиниције могу бити извезене под једним или више различитих имена. Осим дефиниција, могу се додати иницијализациони подаци за меморије и табеле. Могу садржати и почетне функције, чије извршавање се одвија аутоматски [2].

Уграђивач

Имплементација *WebAssembly*-ја је углавном уграђена (енгл. *embedded*) у окружење домаћина. То окружење одређује како ће модули бити учитани, како су увози доступни и како се приступа извезеним дефиницијама. Детаљи зависе од окружења и нису одређени структуром самог језика [2]. Конкретни примери уграђивача биће приказани касније.

3.3 Семантичке фазе

Семантика је подељена у три фазе. За сваки део језика постоји одговарајућа фаза, а оне су [2]:

Декодирање - *WebAssembly* модули се шаљу у бинарном облику. Декодирање је процес који форматира и конвертује бинарни облик у интерну репрезентацију модула. Интерна репрезентација може бити у облику апстрактне синтаксе, али и конкретан машински код.

Валидација - Декодирани модул мора бити валидан. Ова фаза проверава услове добре дефинисаности како би се осигурало да је модул исправан и безбедан. Прецизније, врши се провера типова функција, као и низ инструкција које јој припадају како би се утврдило да је стек операнада конзистентно коришћен.

Извршавање - Уколико су прве две фазе успешно окончане, модул се може извршити. Сама фаза извршавања се састоји од две подфазе:

- Инстанцирање (енгл. *Instantiation*)- Инстанца модула је његова динамичка репрезентација, са сопственим стањима и стеком извршавања. У

овој фази се извршава тело модула, све увезене дефиниције, иницијализују се глобалне променљиве, меморије, табеле и активира се почетна функција (уколико је дефинисана). Враћа примерке извоза модула.

- Позивање (енгл. *Invocation*) - Једном када је фаза инстанцирања завршена, њен резултат се може користити да се позивају извезене функције из претходно инстанцираног модула. Функцијама се прослеђују одговарајући аргументи, а као резултат се добија резултат њиховог извршавања.

Фазе инстанцирања и позивања су операције које одређује окружење домаћина.

Глава 4

v8

Глава 5

Имплементација

Глава 6

Закључак

Литература

[1] Tarun Agarwal. *What is RISC and CISC Architecture with Advantages and Disadvantages*. 2017. URL: `http://www.edgefxkits.com/blog/what-is-risc-and-cisc-architecture/` (посећено 07/23/2018).

[2] WebAssembly Community Group. *WebAssembly Specification, release 1.0*. 2018.

[3] John L. Hennessy и David A. Peterson. *Computer Architecture - A Quantitive Approach, Fourth Edition*. Morgan Kaufmann, 2007.

[4] Dominic Sweetman. *See Mips Run, Second Edition*. Morgan Kaufmann, 2007.

[5] MIPS Technologies. *MIPS Architecture For Programmers Volume II-A: The MIPS 32 Instruction Set*. MIPS Technologies, 2013.