

Kapitola 1

Cíl práce

Exploze stavového prostoru při model checkingu, Partial Order Redukce to řeší, LLVM model checking je super a chceme tam POR.

To, co chceme, je vygenerovaný NTS, který není paralelní. Tedy výstupem redukce není explicitní stavový prostor.

Kapitola 2

Použité technologie

Proč volíme právě NTS? Jakou podmnožinu těch jazyků podporuji?

2.1 LLVM

2.1.1 Typový systém

2.1.2 Omezení

Pole

Nepodporuji (zatím). Přidat podporu některých operací nad poli by ale nemusel být problém

Pointery

Nestaráme se o pointery

Ale s proměnnými se pracuje zkrze pointery!

Instrukce

Také nepodporuji znaménkovou aritmetiku a hromadu instrukcí. Například znaménkové přetečení při add neošetřuji, přestože je to principiálně možné. Vede to k složitým formulím. Pokud ale bude překladová utilita vyvíjena dále, je to jedna z věcí, která by mohla být volitelná (tedy jestli vracet nedefinovaný výsledek nebo přejít do err stavu nebo pokračovat jako nic). S tím souvisí Poison a Undef values.

2.2 NTS

2.2.1 Popis originální verze

Potřeba říci, že máme datové typy Int, Bool, Real. Int, Real jsou matematické, s neomezeným rozsahem a (u Real) s neomezenou přesností.

2.2.2 Rozšíření

Jak můžeme vidět v předchozích částech, typový systém jazyka NTS se od typového systému LLVM IR v mnohém liší. Zatímco v LLVM IR mohou stejný datový typ (Integer type, například i8) využívat jak aritmetické instrukce (binary instructions), tak bitové logické instrukce (bitwise binary instructions), jazyk NTS již na úrovni syntaxe podporuje logické operace pouze nad termy typu Bool. Pro překlad LLVM IR do NTS je tedy nezbytné cílový jazyk vhodným způsobem rozšířit. Možností je více:

1. Rozšířit logické operace i nad datový typ Int. Jaký by ale byl význam například negace? Totiž, pro různou bitovou šířku dává negace různý výsledek. Budeme-li mít například proměnnou typu Int s hodnotou 5 (binárně 101), pokud se na ní budeme dívat jako na 4 bitové číslo (tedy 0101), dostaneme po negaci 10 (1010), pokud se na ní budeme dívat jako na 3 bitové číslo, dostaneme po negaci hodnotu 2 (binárně 010).
2. Zavést speciální operátory, které budou v sobě obsahovat informaci o uvažované bitové šířce.
3. Zavést datový typ BitVector<k>, který je vlastně k-ticí typu Bool. Na něm můžeme dělat logické operace bitově a aritmetické operace jako na binárně reprezentovaném čísle.

Volíme třetí způsob.

V syntaxi původního nts je rozlišováno mezi aritmetickým a booleovským literálem. Chtěli bychom nějak zapisovat bitvectorové hodnoty. Ve hře jsou dvě možnosti:

1. Pro zápis konstant typu BitVector<k> bychom mohli používat speciální syntaxi. Například zápis 32x"01abcd42" může představovat konstantu typu BitVector<32>, zapsanou v hexadecimálním tvaru. Toto řešení je navíc vhodné pro zápis speciálních konstant (jako 2^{31}) v lidsky čitelném tvaru.
2. Aritmetické literály mohou mít schopnost být Int-em nebo BitVector<k>-em podle potřeby (polymorfismus?).

Použita byla druhá možnost (také si jí tu podrobněji popíšeme). V případě potřeby můžeme do jazyka přidat i tu první, ale už ne kvůli potřebě rozlišovat mezi datovými typy, ale kvůli užitečnosti zapisovat některé konstanty hezky.

Syntaxi jazyka zjednodušíme tak, že definujeme jeden neterminál <term> místo dvou neterminálů <arith-term> a <bool-term>. S každým termem bude ale spojená sémantická informace, kterou je jeho datový typ. Definujeme následující skalární datové typy:

1. Int
2. BitVector<k>, $\forall k \in \mathbb{N}^+$
3. Real

Původní datový typ Bool chápeme jako zkratku za typ BitVector<1>

$\langle literal \rangle ::= \langle id \rangle \mid \text{tid} \mid \langle numeral \rangle \mid \langle decimal \rangle$

$\langle aop \rangle = '+' | '-' | '*' | '/' | '\%'$

$\langle bop \rangle = '\&' | '|' | '->' | '<->'$

$\langle op \rangle = \langle aop \rangle | \langle bop \rangle$

$\langle term \rangle ::= \langle literal \rangle | \langle term \rangle \langle aop \rangle \langle term \rangle | \langle term \rangle \langle bop \rangle \langle term \rangle | '(' \langle term \rangle ')'$

Typovací pravidla

Abychom zajistili, že syntakticky stejný výraz může být typu BitVector nebo Int, definujeme typovou třídu Integral se členy BitVector<k> a Int. Typ výrazu je definován rekurzivně

TR1 Numerická konstanta <numeral> je libovolného typu 'a' z třídy Integral.

TR2 <decimal> je typu Real

TR3 tid je libovolného typu 'a' z třídy Integral

TR4 <id> je stejného typu, jako odpovídající proměnná

Mějme termy

$a1 :: \text{Integral } a \Rightarrow a$

$a2 :: \text{Integral } b \Rightarrow b$

$b1 :: \text{BitVector} \langle k1 \rangle$

$b2 :: \text{BitVector} \langle k2 \rangle$

$i1 :: \text{Int}$

$i2 :: \text{Int}$

Potom: $a1 \langle aop \rangle a2 :: \text{Integral } a \Rightarrow a$ $a1 \langle op \rangle b1 :: \text{Bitvector} \langle k1 \rangle$ $a1 \langle aop \rangle i1 :: \text{Int}$ $b1 \langle op \rangle b2 :: \text{BitVector} \langle \max k1, k2 \rangle$ $i1 \langle aop \rangle i2 :: \text{Int}$

Tato pravidla platí i komutativně. Co se do nich nevejde, není typově správný výraz.

Tedy přidáváme datový typ BitVector, typové třídy a implicitní typová konverze.

Btw můžeme využívat anotace. Zatím je využíváme jenom trochu, ale dají se o nich vkládat i nějaké informace z LLVM nebo další postřehy.

2.2.3 Omezení

Zatím nepodporuji složitější operace s poli. Také nechápu, jak pracovat s parametry (par).

Pole

Kapitola 3

Jak na to?

3.1 Použitá terminologie

3.2 Velký obraz

Cílem této práce tedy je vytvořit ze vstupního programu v jazyce LLVM IR, jež využívá posixová vlákna, odpovídající sekvenční přechodový systém v jazyce NTS. Tuto úlohu je možné přirozeně rozdělit na dvě části, a to na

1. přirozený překlad programu z LLVM IR do NTS se zachováním paralelismu (část dále označovaná jako překlad), a
2. převod paralelního programu v jazyce NTS na sekvenční program (část dále označovaná jako sekvencializace).

Tyto dvě části jsou poměrně nezávislé a mohou být implementovány jako rozdílné nástroje, pracující nad společnou knihovnou. Rozdělením navíc získáváme možnost ověřovat vlastnosti sekvenčních programů v jazyce LLVM IR nástrojem, který rozumí jazyku NTS .

3.2.1 Funkční volání

Chceme tedy z paralelního NTS, který vznikl překladem z LLVM IR, učinit sekvenční. Pokud o vstupním paralelním NTS nebudeme nic předpokládat, nástroj pro sekvencializaci bude muset umět korektně zacházet s funkčními voláními, protože ta mohou být součástí NTS. To zejména znamená, že nástroj si bude muset pro každý stav výstupního systému udržovat přehled o tom, jaké funkce jsou v jakém vlákne aktivní. Implementace takového nástroje nemusí být snadná, proto jsme učinili následující rozhodnutí:

Rozhodnutí 1. *Nástroj pro sekvencializaci předpokládá, že vstupní paralelní systém neobsahuje funkční volání, tedy (v termínech) NTS není hierarchický. Takové NTS budeme dále označovat jako "ploché".*

Většina skutečných programů v jazyce LLVM IR ovšem obsahují funkční volání, a rádi bychom takové programy podporovali. Nabízejí se dvě možnosti, jak takovou podporu zajistit. První spočívá v tom, že paralelní program v jazyce

Až na uvedená rozšíření. Nechtěli jsme mít možnost použít unbounded integery?

LLVM IR přeložíme na paralelní a (potenciálně) hierarchický přechodový systém v jazyce NTS, který následně převedeme na ekvivalentní paralelní plochý přechodový systém. Alternativní možností je nejprve odstranit funkční volání z paralelního LLVM IR programu a následně tento plochý program převést na plochý paralelní přechodový systém.

Tak či onak, programy s rekurzivním voláním nebudou podporovány.

Zplošťování LLVM Zploštění programu v LLVM IR nevyžaduje téměř žádnou další práci, protože již na tuto úlohu existuje llvm průchod. Jmenuje se inline a je součástí projektu LLVM. Nicméně nástroj pro překlad LLVM IR na NTS musí umět korektně přeložit volání funkce `pthread_create(3)`, které v tomto případě nemůže být přeloženo jako volání `BasicNts` na úrovni NTS. Volání této funkce by tedy muselo být vhodně přeloženo již na úrovni LLVM kódu.

Zplošťování NTS Přestože je pro zplošťování NTS potřeba udělat o něco více práce, napsání odpovídajícího nástroje nám může umožnit sekvencializaci libovolného NTS. Navíc nejsme omezeni na práci s plochými NTS jako v předchozím případě. Z uvedených důvodů tato práce využívá tento přístup.

3.2.2 Architektura

Celý problém se tedy rozpadá na několik částí. Nejprve potřebujeme přeložit program z jazyka LLVM IR do formalizmu NTS, poté z přeloženého programu odstranit volání subsystémů, plochý program sekvencializovat s využitím partial order redukce a nakonec ho uložit do souboru. Pro spolupráci těchto částí je třeba mít vybudovanou vhodnou paměťovou reprezentaci programu v jazyce NTS. Existence parseru by byla užitečná (zejména pro testování), ale není nutná.

3.3 Architektura libNTS

- inspirováno LLVM

Existuje
Javovská
implemen-
tace NTS

3.4 Překlad llvm na nts

Btw nepatří rozhodnutí o omezení vstupního jazyka sem?

3.4.1 Funkce

Jak jazyk NTS, tak LLVM IR podporují určitou formu hierarchického členění programu na podprogramy. Zatímco u LLVM IR je základní jednotkou tohoto členění funkce, v případě jazyka NTS jde o `BasicNts`, který reprezentuje jednoduchý přechodový systém. Jak funkce, tak `BasicNts` mohou mít libovolný počet (vstupních) parametrů zvoleného typu, ašak zatímco funkce může vracet nejvýše jednu hodnotu, `BasicNts` může definovat libovolný počet výstupních parametrů. Funkce i `BasicNts` navíc mohou definovat lokální proměnné. Vzhledem k těmto podobnostem je celkem přirozené překládat funkce z LLVM IR na `BasicNts`.

3.4.2 BasicBlocky

Každá funkce se skládá z BasicBlocků. Protože žádná instrukce uvnitř základního bloku nemění tok řízení a po jejím vykonání je vykonávána instrukce následující, můžeme každý základní blok s N vnitřními instrukcemi přeložit jako N nových řídicích stavů, kde n-tý stav odpovídá stavu programu po vykonání n instrukcí. Sémantika jednotlivých vnitřních instrukcí bude poté zachycena v přechodových pravidlech mezi jednotlivými stavy.

Terminující instrukce Nicméně poslední instrukce v základním bloku může změnit tok řízení, například ukončit vykonávání funkce nebo skočit na začátek jiného základního bloku. Třída takovýchto instrukcí se nazývá "terminační instrukce", protože tyto instrukce ukončují základní blok. Odpovídající přechody tedy nemusí vést do nového stavu, ale do již existujícího stavu jiného základního bloku.

V jazyce LLVM IR existuje instrukce Phi, reprezentující Φ uzel v SSA jazyce. Tato instrukce se nachází pouze na začátku základního bloku a její výsledek je závislý na předchozím dokončeném základním bloku. Z tohoto důvodu jsou při překládání funkce základní bloky očíslovány a každá terminující instrukce ukládá číslo svého základního bloku do speciální proměnné. Samotná phi instrukce však v době psaní tohoto textu není v překládacím nástroji implementována.

Zmínit
SSA

známe
z teorie
překladačů?

3.4.3 Znaménkovost

Jak číselné datové typy v jazyce LLVM IR, tak přidaný BitVector typ v jazyce NTS jsou bezznaménkové. Nicméně některé LLVM instrukce (například ICMP) mohou interpretovat použité proměnné znaménkově. Přestože je principiálně možné vyjádřit sémantiku znaménkových operací pomocí bezznaménkové aritmetiky, výsledné formule by byly složitější. Tato instrukce přiřadí do proměnné

ukázat jak
na to

zmínit
seman-
tiku op-
eraci nad
BitVector

Někde
zmínit,
že llvm ir
využívá
právě dvo-
jkový do-
plněk

```
%b = icmp slt i8 %x, 10
```

'b' hodnotu 1 právě pokud x je znaménkově menší (signed less then, slt) než hodnota 10, tedy právě pokud x bude mít hodnotu v rozsahu -128 až 9. Tento rozsah, vyjádřený v dvojkovém doplňkovém kódu, odpovídá neznaménkovým hodnotám v rozsahu 0 ... 9 nebo 128 ... 255.

```
%b = icmp slt i8 %x, %y
```

Pokud jsou ovšem obě hodnoty neznámé, porovnání je obtížnější. Výsledkem operace bude 1, právě pokud bude splněna jedna z následujících podmínek:

1. Hodnota proměnné x je záporná a hodnota proměnné y je kladná, tedy v unsigned aritmetice $x > 127$ a $y \leq 127$.
2. Obě proměnné mají shodné znaménko a zároveň x je neznaménkově menší než y.

Jsou i jiné způsoby, jak v neznaménkové aritmetice vyjádřit chování této instrukce, ale nenašel jsem žádný elegantnější nebo jednodušší. Z tohoto důvodu současná verze překládacího nástroje nepodporuje překlad znaménkových operací do aritmetiky neznaménkových bitvektorů.

3.4.4 Model paralelismu

Jazyky LLVM IR a NTS implementují paralelismus velice odlišně. Během návrhu překladu jsme preferovali pokrytí možných použití pthreads v LLVM před jednoduchostí výsledného NTS.

Možná, že popis paralelních modelů LLVM a NTS nepatří do této sekce. Někde to ale vysvětlit musím.

LLVM, pthreads

Jazyk LLVM IR sám neposkytuje podporu pro explicitní paralelismus. Proto podporujeme posixovou knihovnu pro vlákna, pthreads (resp. její část). Nejzajímavější funkcí z této knihovny je `pthread_create`, která jako jeden ze svých argumentů přijímá pointer na funkci, jež následně spustí v novém vytvořeném vlákně. Volání této funkce se může ve zdrojovém programu vyskytnout kdekoliv, a na umístění a četnost těchto výskytů neklademe žádná omezení. Omezujeme ale možnosti použití této funkce, a to následujícím způsobem:

1. Argument parametru `start_routine` musí být přímo funkce, nikoliv jiný funkční ukazatel.
2. Všechny ostatní argumenty funkce `pthread_create` musí být null.
3. Návratová hodnota funkce `pthread_create` nesmí být použita.
4. Funkce, předaná jako argument parametru `start_routine`, nesmí používat svůj parametr.
5. Návratová hodnota této funkce musí být vždy null.

```
int pthread_create ( pthread_t *thread ,  
                    const pthread_attr_t *attr ,  
                    void *(*start_routine) (void *),  
                    void *arg  
                    );
```

NTS

Jazyk NTS řeší paralelismus odlišným způsobem. Počet vláken je po celou dobu běhu programu stejný, tedy každé vlákno běží již od začátku. Každý paralelní NTS obsahuje řádek ve tvaru

```
instances basic_nts_1[2] , basic_nts_2[3*N+4] , ... , main_nts[1];
```

ne nutně
řádek

který určuje, jaké vlákno bude vykonávat jaký přechodový (sub)systém (BasicNts). Uvedený příklad specifikuje, že dvě vlákna (s id 0 a 1) budou vykonávat BasicNts `basic_nts_1`, dalších několik vláken (s id 2, 3, 4, ... $3*N+5$) bude vykonávat `basic_nts_2` a vlákno s nejvyšším id bude vykonávat `main_nts`.

Překlad

V případě paralelního LLVM překládací program vygeneruje NTS, jež vytváří 1 hlavní vlákno a N vláken pracovních. Každé pracovní vlákno má přiřazeno stav a v každém čase se nachází buď ve stavu nečinném, nebo ve stavu činném. Protože na začátku běhu programu v jazyce LLVM IR je spuštěno pouze jedno vlákno, zatímco v přechodovém systému NTS jsou spuštěna všechna vlákna, chceme, aby běžící pracovní vlákna neměla žádný *efekt*. Tedy pracovní vlákno bude ve výchozím stavu v režimu nečinném. Veškerá volání funkcí `pthread_create` nahradíme kódem, který způsobí, že se nějaké nečinné pracovní vlákno stane činným a začne vykonávat kód odpovídající funkce. Po ukončení běhu této funkce vlákno opět přejde do stavu nečinnosti a je tak možné ho znovu využít.

Vložit
někam
výsledný
NTS kód

Implementace

3.5 Inlining

Jo teda nepodporuju rekurzi, a proto si můžu dovolit to, co dělám v inlineru - prostě tak dlouho zainlinovávám jednotlivé BasicNts, až mi nezůstané žádné volání.

3.6 Sekvencializace

V rámci zachování jednoduchosti v této části textu používáme o něco jednodušší definice než v [NTSref], zejména vypouštíme definici valuace proměnných.

Definice 1. Řídící stav plochého paralelního NTS s k vlákny je k -tice $q = \langle q_0, \dots, q_{k-1} \rangle$, kde q_i je řídicí stav BasicNts, který je vykonáván ve vlákne s $tid = i$.

S každým řídicím stavem q paralelního NTS je spojena množina přechodů $\text{Outgoing}(q)$, vedoucích z tohoto stavu.

$$\text{outgoing}(q) = \text{outgoing}(q_0) \cup \dots \cup \text{outgoing}(q_{k-1}) \quad (3.1)$$

Definice 2. Stav s paralelního přechodového systému je dvojice $s = \langle q, v \rangle$, kde v je řídicí stav daného systému a v značí valuaci (globálních i všech lokálních) proměnných.

Referenční manuál označuje právě definovaný stav jako "konfiguraci".

Formálněji,
odkaz do
manuálu

Definice 3. Pro libovolný přechod t a stavy s_1, s_2 paralelního systému, zápisem $s_1 \xrightarrow{t} s_2$ rozumíme, že je možné dostat se ze stavu s_1 do stavu s_2 pomocí přechodu t . Zápisem

$$s_1 \xrightarrow{t_1} s_2 \xrightarrow{t_2} \dots \xrightarrow{t_{m-1}} s_m$$

zkracujeme konjunkci

$$s_1 \xrightarrow{t_1} s_2 \wedge \dots \wedge s_{m-1} \xrightarrow{t_{m-1}} s_m$$

3.6.1 Jednoduchá sekvencializace

3.6.2 Partial Order Reduction

Partial Order Reduction je technika pro redukci stavového prostoru (paralelních) programů, která využívá komutativity paralelně vykonávaných přechodů. Pokud lze efektivně ověřit, že stav programu po vykonání přechodů t_1 a t_2 nezávisí na jejich pořadí a ověřovaná vlastnost není citlivá na jednotlivé přechody, není třeba uvažovat některé možné běhy. Technika použitá v této práci vychází z techniky popsané v [CLARKE].

Nezávislost a komutativita přechodů

Na rozdíl od [CLARKE] nevyžadujeme, aby přechodová pravidla byla (datově) deterministická. Všechny přechody v NTS již jsou kontrolně deterministické. Nezávislost přechodů tedy definujeme mírně odlišně.

Definice 4. *Binární relace I na přechodech paralelního NTS nazýváme relací nezávislosti, pokud je symetrická, antireflexivní a splňuje následující požadavek: Pro libovolné přechody $(t_1, t_2) \in I$ a libovolné stavy (konfigurace v NTS) s_1, s_2, s_3 takové, že $s_1 \xrightarrow{t_1} s_2 \xrightarrow{t_2} s_3$ existuje stav s'_2 takový, že $s_1 \xrightarrow{t_2} s'_2 \xrightarrow{t_1} s_3$.*

Pro NTS bez nedeterministických přechodů je tato definice shodná s definicí v [CLARKE].

Definice 5. *Přechod t je povolený ve stavu (konfiguraci) s (píšeme $t \in \text{enabled}(s)$) právě tehdy, když existuje stav s' tak že $s \xrightarrow{t} s'$*

Je zřejmé, že pokud pro stav $s = \langle()q, v\rangle$ a přechod t platí $t \in \text{enabled}(s)$, pak $t \in \text{outgoing}(q)$. Stavy takové, že vybraný $\text{ample}(s) = \text{enabled}(s)$, označujeme jako plně expandované.

Technika, použitá v [CLARKE], tedy partial order reduction s využitím dostatečných (ample) množin, spočívá v tom, že pro každý stav (konfiguraci) paralelního systému je spočítána množina přechodů $\text{ample}(s) \subseteq \text{enabled}(s)$, která má tu vlastnost, že přechody mimo $\text{ample}(s)$ nejsou "důležité". Formálněji řečeno je třeba, aby ke každé cestě v úplném explicitním přechodovém grafu paralelního systému existovala (určitým způsobem) ekvivalentní cesta v tomtéž grafu, která nevyužívá žádné přechody mimo ample . Při konstrukci odpovídajícího sekvenčního systému pak není třeba uvažovat přechody mimo ample , což zmenšuje velikost výsledného sekvenčního systému.

Nutné podmínky

Kniha [CLARKE] uvádí několik podmínek, postačujících pro zachování korektnosti. Pro připomenutí, ta kniha pracuje pouze s deterministickými přechody.

$$C0 \quad \text{ample}(s) = \emptyset \Rightarrow \text{enabled}(s) = \emptyset$$

- C1 Každá cesta v úplném explicitním přechodovém grafu paralelního systému, začínající ve stavu s , má následující vlastnost: pokud přechod t_2 , který se vyskytuje po cestě, závisí na nějakém přechodu $t_1 \in \text{ample}(s)$, pak se před jeho výskytem vyskytuje nějaký přechod $t'_1 \in \text{ample}(s)$.

C2 Pokud s není plně expandovaný, pak každý přechod z $\mathbf{ample}(s)$ je neveditelný .

zadefinovat

C3 Nesmí existovat cyklus, který obsahuje přechod povolený v některém z jeho stavů a neobsažený v žádném z \mathbf{ample} stavů cyklu.

Uvedené podmínky se vztahují k explicitnímu CFG, zatímco náš nástroj generuje symbolický CFG a o proměnných si mnoho informací neudrží. Protože $\mathbf{ample}(s) \subseteq \mathbf{enabled}(s)$, nemůžeme generovat ample sety přímo: museli bychom totiž rozhodnout, zda je nějaký z přechodů v daném stavu povolený. Náš nástroj namísto toho pro *řídící* stav (tedy nikoliv konfiguraci) q paralelního systému spočítá množinu $A(q) \subseteq \mathbf{outgoing}(q)$ tak, že pro libovolný stav (konfiguraci) $s = (q, v)$ můžeme položit $\mathbf{ample}(s) = A(q) \cap \mathbf{enabled}(s)$.

Problém s C0 Pokud zvolíme A neprázdné tak, že je dosažitelný stav, kdy žádný přechod z A nebude proveditelný, ale nějaký přechod mimo A ano, pak dojde k porušení C0. Toto je ale problém pouze pro obecné NTS, protože přechody překládané z LLVM IR jsou vždy proveditelné.

Jednoduchá heuristika

Nezávislost Definice relace nezávislosti umožňuje jistou flexibilitu při výběru vhodné relace. Za nezávislé budeme považovat ty páry přechodů (t_1, t_2) , které splňují obě následující podmínky:

- Přechody t_1 a t_2 patří ke dvěma různým vláknům.
- Neexistuje proměnná, sdílená oběma přechody t_1 i t_2 a modifikovaná alespoň jedním z nich.

Ample sety Pro řídící stav $q_0 = \langle q_{0,0}, \dots, q_{0,k-1} \rangle$ spočítáme $A(q_0)$ následujícím způsobem. Pro každé $0 \leq i < k$ zkusíme položit $A(q_0) = \mathbf{outgoing}(q_{0,i})$. Pokud zkoušená množina splňuje podmínky C0 až C3, použijeme ji. Pokud žádná z vyzkoušených množin není použitelná, položíme $A(q_0) = \mathbf{outgoing}(q_0)$.

Jak uvádí [CLARKE], mohou nastat dvě situace, kdy dojde k porušení podmínky C1. V úplném explicitním přechodovém grafu musí existovat posloupnost $s_0 \xrightarrow{\alpha_0} \dots \xrightarrow{\alpha_{m-1}} s_m \xrightarrow{\beta} s_{m+1}$, $s_j = (q_j, v_j)$ taková, že $\beta \in \mathbf{enabled}(s_l)$, β je závislý na nějakém přechodu ve zvoleném $\mathbf{ample}(s_0)$ a všechny α_j jsou nezávislé na všech přechodech v $\mathbf{ample}(s_0)$.

- Přechod β pochází z vlákna i (tedy $\beta \in \mathbf{outgoing}(q_{m,i})$). Protože α_i jsou nezávislé na $\mathbf{ample}(s_0)$, musí pocházet z jiného vlákna než i , tedy $q_{0,i} = q_{m,i}$. Protože $\beta \notin \mathbf{ample}(s_0)$, tak $\beta \notin \mathbf{enabled}(s_0)$. V tom případě se ale stavy s_0 a s_m liší ve valuaci některých globálních proměnných, které jsou používány přechodem β , a tedy nějaké vlákno s $tid \neq i$ obsahuje přechod, modifikující proměnnou, která je používána přechodem β .
- Pokud $\beta \notin \mathbf{outgoing}(q_{m,i})$, tedy přechod β pochází z vlákna jiného než i , nějaké jiné vlákno než i obsahuje přechod, který je závislý na nějakém přechodu z $\mathbf{ample}(s_0)$.

V obou případech tedy musí nějaké jiné vlákno obsahovat přechod, který sdílí proměnnou s nějakým přechodem z `ample(s0)`, která je modifikována jedním z nich.

3.6.3 Jak to má fungovat

Nakonec jde jenom o to, zda proměnnou, kterou nějaký přechod používá, používá i jiné vlákno

Možná
trocha
teorie

Použitá heuristika

Pro každý stav sekvencializovaného systému se snažíme najít dostačující množinu jeho následníku tak, že postupně zkoušíme volit pro každý proces i množinu A_i přechodů v jeho řídicím stavu povolených. Definujeme množiny $W(A_i)$ a $R(A_i)$ proměnných, které mohou být některým z přechodů v A_i změněny nebo čteny. Pro každý proces P_i dále definujeme množinu R_i globálních proměnných takovou, že pokud P_i obsahuje přechod, který čte proměnnou v , pak $v \in R_i$. Také potřebujeme znát množinu W_i globálních proměnných takovou, že pokud proces P_i může změnit hodnotu proměnné v , pak $v \in W_i$.

Pokud existuje, zvolíme takovou A_i , pro kterou platí, že pro libovolný jiný proces P_j (t.j. $j \neq i$) platí, že žádný z přechodů tohoto procesu nemění ani nečte proměnné změněné některým z přechodů v A_i , ani nemění žádou z proměnných, čtenou některým z vybraných přechodů. Formálněji

$$W(A_i) \cap (W_j \cup R_j) = \emptyset \wedge R(A_i) \cap W_j = \emptyset \quad (3.2)$$

Popsat
tu argu-
mentaci z
knížky?

Výpočet Mějme vlákno P_i , které instancuje plochy BasicNts B_i . Pokud B_i obsahuje k přechody t_1, \dots, t_k , potom W_i i R_i můžeme spočítat jako sjednocení

$$W_i = W(t_1) \cup W(t_2) \cup \dots \cup W(t_k) \quad (3.3)$$

$$R_i = R(t_1) \cup R(t_2) \cup \dots \cup R(t_k) \quad (3.4)$$

kde $W(t_j)$, $R(t_j)$ reprezentují množinu proměnných, které jsou modifikovány / čteny přechodem t_j .

Dat to
do jedné
rovnice?

3.6.4 Problém velkého procesu

Ukazuje se, že pro zvolený model paralelismu 3.4.4 uvedená heuristika není příliš efektivní. Pro velkou množinu programů totiž napočítané množiny W_i a R_i obsahují mnoho proměnných, které ve skutečnosti vlákno i nepoužívá.

Problémová situace

Uvažme například příklad . Hlavní funkce vytvoří dvě vlákna, která používají vzájemně disjunktí množiny globálních proměnných. Po přeložení do jazyka NTS (se zvoleným modelem paralelismu) výsledné NTS obsahuje vygenerované BasicNts "`__thread_pool_routine`", které obsahuje přechody, volající přeložené

Formálněji
popsat,
co jsou
proměnné
čtené /
modifiko-
vané ně-
jakým pře-
chodem

všechny,
které lze
vyjádřit
v modelu
"vytvoř
vlákna na
začátku",
todo taky
nekde pop-
sat další
možné
modely

```

@G1 = global i32 16;
@G2 = global i32 0;

define i8* @p1 ( i8* %x ) {
    ... ; some local calculations
    store i32 42, i32* @G1;
    ret i8* null;
}

define i8* @p2 ( i8* ) {
    ... ; some more local calculations
    store i32 14, i32* @G2;
    ret i8* null;
}

define void @main() {
    call i32 @pthread_create ( ... , @p1, ... );
    call i32 @pthread_create ( ... , @p2, ... );
    ret void;
}

```

Obrázek 3.1: Vlákna, využívající odlišnou sadu proměnných. Zjednodušeno.

funkce @p1 a @p2.

Po zploštění tedy `thread_pool_routine` obsahuje přechody, které odpovídají původním funkcím `p1` a `p2`. Protože tento `BasicNts` je vykonáván v každém pracovním vlákně, v každém pracovním vlákně jsou také obsaženy všechny přechody přeložených funkce `p1` a `p2`. Napočítané množiny W_i (a R_i) všech pracovních vláken jsou pak stejné a tedy zbytečně velké: některé z přechodů se v daném vlákně totiž nikdy nevykonají.

```

_thread_pool_routine {
    initial si;
    si -> sr1 { ( ( tps[tid] = 1 ) && havoc ( ) ) }
    sr1 -> ss { p2 ( ) }
    ss -> si { ( tps'[tid] = [0] && havoc ( tps ) ) }
    si -> sr2 { ( ( tps[tid] = 2 ) && havoc ( ) ) }
    sr2 -> ss { p1 ( ) }
    ss -> si { ( tps'[tid] = [0] && havoc ( tps ) ) }
}

```

Obrázek 3.2: Přechodový systém pracovního vlákna

Opravit
číslování
vláken

Možné řešení

Pro zvolený model paralelismu se zdá být rozložení programu na vlákna příliš hrubým, protože mnohá vlákna budou mít totožné řízení (BasicNts). Rozložme tedy program jemněji na jednotky, kterým budeme říkat "úlohy".

Definice 6. Pro plochý přechodový systém N , množina úloh $\mathit{Tasks}(N)$ je množina taková, že libovolný řídicí stav nějakého BasicNts z N leží v právě jedné úloze $T \in \mathit{Tasks}(N)$.

Přechodové systémy, vzniklé překladem pomocí našeho nástroje z LLVM IR a následným zploštěním, obsahují právě dva použité BasicNts: první z nich odpovídá funkci `main` původního programu, druhý pak kódu vykonávaným vláknem z thread poolu. Pro účely této práce můžeme rozdělit jejich stavy do úloh následujícím způsobem.

1. Máme právě jednu hlavní úlohu T_{main} , právě jednu nečinnou úlohu T_{idle} a jednu úlohu T_i pro každou funkci f_i , která je někde v původním LLVM IR programu předávána jako parametr funkce `pthread_create`.
2. Každý stav s hlavního BasicNts (tj. toho, který vznikl překladem funkce `main`) leží v T_{main} .
3. Všechny stavy, odpovídající stavům nezploštělého `__thread_pool_routine`, leží v T_{idle} .
4. Všechny ostatní řídicí stavy jsou řídicími stavy zploštělého `__thread_pool_routine`, do kterého se dostaly během fáze zplošťování. Každý z těchto stavů přiřadíme úloze T_i takové, že daný stav pochází z funkce f_i .

Koncept úloh přirozeně rozšiřujeme i na přechody:

Definice 7. Přechod t náleží úloze T právě tehdy, když zdrojový řídicí stav přechodu t náleží úloze T .

Protože každý přechod má jeden zdrojový řídicí stav, každý přechod náleží právě jedné úloze. Tímto způsobem můžeme o každé úloze říci, jaké globální proměnné používá.

Definice 8. Úloha T čte globální proměnnou v právě tehdy, pokud existuje přechod t , který čte v a patří T . Podobně, T mění v právě pokud nějaký přechod t patří T mění v .

3.6.5 Problém velkého pole

V případě, že bychom měli pole takové, že by na každou jeho pozici přistupoval nejvýše jeden proces, a procesů bychom měli mnoho, vyplatilo by se sledovat jeho jednotlivé buňky zvlášť. To ale neděláme. Btw jedno takové pole máme.

Jak může úloha aktivovat jinou a co z toho plyne. Je ale třeba mít popsané zákony POR.

3.6.6 Problém závislosti na datech

Zda může nějaké vlákno běžet, závisí na datech. My se ale o data moc nestaráme (TODO: tohle je třeba ujasnit na začátku). Tedy nemůžeme vědět, že na začátku poběží jenom hlavní vlákno. Tedy zeserializovaný systém bude obsahovat běhy, jejichž podmínka cesty bude nesplnitelná. Obecně tohle řešit snadno nelze, ale pokud se omezíme na zjištění informace, zda nějaké vlákno z thread poolu může začít vykonávat nějakou úlohu, stačí nám sledovat pár zvolených proměnných. Na to máme dvě možnosti:

a) Analyzovat vykonávané přechody, zda modifikují naše vybrané proměnné. Předpokládáme, že většina formulí bude mít hezký tvar, a že tedy nemusíme vědět všechno na to, abychom některé mohli rovnou označit za nesplnitelné a o jiných prohlásit, že modifikují námi vybranou proměnnou jednoduchým způsobem.

b) Umět rozpoznat původně existující struktury i v přeloženém a zplacatěném přechodovém systému. Tedy musíme vědět, co jsou pracovní vlákna, co dělá `__thread_create` (a jak jí poznám) a další věci.

Kapitola 4

Závěr

4.1 Možná budoucí rozšíření

4.1.1 Znaménkové datové typy v NTS

Petr psal autorům NTS, a pokud si dobře vzpomínám, jeden z nich projevil přání, abychom mohli překládat llvm do stávající podmnožiny NTS - tedy do intů. Imho to není dost dobře možné.

Obsah

1	Cíl práce	1
2	Použité technologie	2
2.1	LLVM	2
2.1.1	Typový systém	2
2.1.2	Omezení	2
2.2	NTS	2
2.2.1	Popis originální verze	2
2.2.2	Rozšíření	3
2.2.3	Omezení	4
3	Jak na to?	5
3.1	Použitá terminologie	5
3.2	Velký obraz	5
3.2.1	Funkční volání	5
3.2.2	Architektura	6
3.3	Architektura libNTS	6
3.4	Překlad llvm na nts	6
3.4.1	Funkce	6
3.4.2	BasicBlocky	7
3.4.3	Znaménkovost	7
3.4.4	Model paralelismu	8
3.5	Inlining	9
3.6	Sekvencializace	9
3.6.1	Jednoduchá sekvencializace	10
3.6.2	Partial Order Reduction	10
3.6.3	Jak to má fungovat	12
3.6.4	Problém velkého procesu	12
3.6.5	Problém velkého pole	14
3.6.6	Problém závislosti na datech	14
4	Závěr	16
4.1	Možná budoucí rozšíření	16
4.1.1	Znaménkové datové typy v NTS	16