# 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 podporů 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 vyvíjená 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 logcké 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 booleovském 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>

$\langle literal \rangle$ ::= $\langle id \rangle$ | tid | $\langle numeral \rangle$ | $\langle decimal \rangle$

⟨*aop*⟩ = ‘+’ | ‘-’ | ‘*’ | ‘/’ | ‘%’

⟨*bop*⟩ = ‘&’ | ‘|’ | ‘->’ | ‘<->’

⟨*op*⟩ = ⟨*aop*⟩ | ⟨*bop*⟩

⟨*term*⟩ ::= ⟨*literal*⟩ | ⟨*term*⟩ ⟨*aop*⟩ ⟨*term*⟩ | ⟨*term*⟩ ⟨*bop*⟩ ⟨*term*⟩ | ‘(’ ⟨*term*⟩ ‘)’

**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 :: Integral a => a
a2 :: Integral b => b
b1 :: BitVector<k1>
b2 :: BitVector<k2>
i1 :: Int
i2 :: Int

Potom: a1 <aop> a2 :: Integral a => a a1 <op> b1 :: Bitvector<k1> a1 <aop> i1 :: Int b1 <op> b2 :: BitVector<maxk1,k2> i1 <aop> i2 :: 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 .

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

### 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 vlákně aktivní. Implementace takového nástroje nemusí být snadná, proto jsme učinili následující rozhodnutí:

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 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 implementace 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 blocku nemění tok řízení a po jejím vykonání je vykonávána instrukce následující, můžeme každý základní blok s N vnitřními instrukcemi přeložit jako N nových řídících 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 instruce 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 překladu 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 bezznamnékové aritmetiky , výsledné formule by byly složitější. Tato instrukce přiřadí do proměnné

```
%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 $0 \ldots 9$ nebo $128 \ldots 255$.

```
%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 $y <= 127$.
2. Obě proměnné mají shodné znaménko a zároveň x je neznaménkově menší než y.

ukázat jak na to

zminit semantiku operaci nad BitVector

Někde zmínit, že llvm ir využívá právě dvojkový doplněk

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

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

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

#### 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

ne nutně řádek

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

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 POR

Pozor, POR je víc druhů. Uvézt chytrou knížku, Ample sety.

### 3.6.1 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 řídícím 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.ž $j \neq i$) 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

$$W(A_i) \cap (W_j \cup R_j) = \emptyset \wedge R(A_i) \cap W_j = \emptyset \tag{3.1}$$

Popsat tu argumentaci z knizky?

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

$$W_i = W(t_1) \cup W(t_2) \cup \ldots \cup W(t_k) \tag{3.2}$$

$$R_i = R(t_1) \cup R(t_2) \cup \ldots \cup R(t_k) \tag{3.3}$$

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

Dat to do jedne rovnice?

.

### 3.6.2 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.

Formálněji popsat, co jsou proměnné čtené / modifikované nějakým přechodem

#### Problémová situace

Uvažme například příklad . Hlavní funkce vytvoří dvě vlákna, která používají vzájemně disjunktní 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 funkce @p1 a @p2.

všechny, které lze vyjádřit v modelu "vytvoř vlákna na začátku", todo taky nekde popsat dalsi mozne 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;
}
```

reference

Program to pojmenovava '__thread_poll_routine' - prijde opravit

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

Opravit číslování vláken

```
_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

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.

**Možné řešení**

Protože každé pracovní vlákno (kromě vlastního) může potenciálně vykonávat libovolnou úlohu, tak téměř každé vlákno může použít téměř každou proměnnou. Redukce by se zredukovala na pouhý test "používám globální proměnné"? Tedy je potřeba mít rozdělené stavy / přechody do úloh. O každé úloze spočítáme, jaké globální proměnné používá, a také, jaké jiné úlohy může aktivovat. Potom, pokud budeme znát řídící stav každého vlákna, můžeme zjistit, jaké úlohy běží a tedy i jaké globální proměnné jsou důležité.

### 3.6.3 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.

### 3.6.4 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, jejihž 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 zplacatělém 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