

Posudek vedoucího na diplomovou práci Martin Jambor: Optimalizace vědeckých výpočtů pro GNU Compiler Collection

Diplomová práce se věnuje problému optimalizace v prostředí produkčního překladače. GCC je přibližně 20 let trvající projekt a jeho vývoj sám o sobě vyžaduje velké zkušenosti a dobrou znalost zdrojového kódu (vlastní back-end překladače je přibližně 500 000 řádek kódu v jazyce C). Pro diplomovou práci bylo zvoleno téma optimalizace vědeckého kódu, které bylo v GCC dlouhou dobu podceňováno a teprve nedávno byly do GCC přidány některé základní vlastnosti běžné u překladačů určených pro vědecké výpočty (jako autovektorizace, paralelizace či podpora Fortranu 95).

Řešitel se rozhodl věnovat optimalizaci vědeckých programů v C++. Jiné týmy (zejména IBM Haifa) se už začaly věnovat optimalizaci klasických programů psaných ve Fortranu a vědecké výpočty v C++ přinášejí zcela nové problémy, které zatím nejsou dostatečně zvládnuty.

Nové programy jsou navrhovány s intenzivním využitím šablon, které sice teoreticky generují dobrý kód (po provedení běžných optimalizací), v praktickém nasazení se však brzy narazí na limity dnešních překladačů a tak se k výkonu stejných algoritmů psaných v C či Fortranu jen vzdáleně blíží. V GCC byl za poslední dva roky dosažen značný pokrok zejména zlepšením inlineru, alias analýzy a přepsáním interprocedurálních optimizérů nad SSA formu. GCC se zdá být v této oblasti pokročilejší, než jiné překladače, jako Intel C++, Open64 či XLC a často generuje rychlejší i kratší kód.

První část diplomové práce podle zadání věnuje rozboru optimalizace několika dostupných softwarových balíčků (Blitz, Pooma, TraMP-3D atd.).

Druhá část diplomové práce implementuje optimalizace na základě rozboru v první části. Řešitel se rozhodl věnovat dataflow analýze s přihlédnutím na objektově orientovaný styl programování. Analýza je v praktickém nasazení funkční i dostatečně rychlá pro produkční překladač a dosahuje výrazně lepších výsledků, než stávající dataflow analýza na skalárních proměnných, kterou implementovala IBM Haifa. Na základě této dataflow analýzy byla implementována propagace konstant, která sice přináší jen střední zlepšení reálných benchmarků, je ale důkazem, že vlastní dataflow funguje. Aby bylo dosaženo výrazných zlepšení benchmarků je nutné vyřešit ještě několik dalších problémů, které ale přesahují rozsah diplomové práce. Práce patří k úspěšnějším projektům optimalizace v GCC.

Řešitel tedy prokázal schopnost samostatně pracovat na problému a zorientovat se v rozsáhlých projektech jakým je jak GCC tak i samotné analyzované knihovny v C++. Implementace je napsána pěkně a odpovídá většině podmínek na zařazení do oficiální distribuce GCC. Vlastní zařazení se plánuje pro příští verzi, GCC 4.4. Práci tedy doporučuji přijmout jako diplomovou a návrh klasifikace přikládám.


Jan Hrbicka