

Posudek diplomové práce

předložené na Matematicko-fyzikální fakultě Univerzity Karlovy

Posudek oponenta

Autor:	Bc. Karel Král
Název práce:	Data structure behavior with variable cache size
Stud. program a obor:	informatika, diskrétní modely a algoritmy
Rok odevzdání:	2017
Jméno a tituly oponenta:	Mgr. Martin Mareš, Ph.D.
Pracoviště:	Katedra aplikované matematiky
Kontaktní e-mail:	mares@kam.mff.cuni.cz

Chování algoritmů na počítačích se složitější pamětovou hierarchií (tvořenou typicky hlavní pamětí a několik úrovněmi cache) se obvykle popisuje pomocí cache-oblivious modelu. Tradičně tento model předpokládá, že parametry hierarchie zůstávají po celou dobu výpočtu konstantní. To nemusí být pravda – například je-li cache sdílena více procesory. Předložená práce se studuje rozšíření cache-oblivious modelu o proměnlivou velikost cache.

Práce nejprve seznamuje čtenáře s architekturou cache a způsoby jejího modelování. Představuje problém proměnlivé cache a dosud známé výsledky. Ukazuje, že předchozí výsledky pracují s proměnlivostí nejednotně: někteří autoři předpokládají, že se velikost cache může měnit pouze při načítání bloku z hlavní paměti (cache miss), jiní dovolují změnu při každém přístupu do paměti (cache miss i cache hit).

Následující část práce buduje nový model proměnlivé cache. Ten umožňuje libovolně zvolit dobu trvání cache missu i cache hitu a profil cache jako funkci určující velikost cache v závislosti na čase. Různou volbou parametrů pak získáme mimo jiné dříve uvažované modely. (V klasickém cache-oblivious modelu nemá tato parametrizace zajímavý efekt, jelikož změny parametrů se projeví nejvýše konstantním zpomalením. Zde ovšem může drobná úprava parametrů způsobit posunutí přístupů do paměti vůči skokovým změnám profilu cache.)

V novém modelu pak autor dokazuje analogie tvrzení známých pro pevně velkou cache. Speciálně ukazuje, že strategie LRU (least-recently used) je stále c -kompetitivní vůči offline optimální strategii pro vhodnou konstantu c . Z toho přímo plynou i některá dříve známá tvrzení o jiných modelech.

Autor také zavádí způsob adresace paměti pomocí hloubkových indexů (pořadí v LRU hypotetické nekonečně velké cache). Dokazuje, že tento způsob je ekvivalentní s klasickou adresací pořadovým číslem buňky. Přesněji že ke každému algoritmu s jedním způsobem adresace existuje algoritmus s druhým způsobem, který počítá totéž a vykoná stejnou posloupnost přístupů do paměti. Jeho časová složitost se nicméně může lišit. Pomocí hloubkové adresace lze pak snadno zkonstruovat algoritmy s určeným chováním ke cachí.

Poslední část práce se věnuje modelování programů běžících paralelně nad sdílenou cachí. Dokazuje, že dva programy „přetahující se“ o cache velikosti M provedou dohromady nejvýše tolik cache missů, jako kdyby každý z nich pracoval samostatně nad cachí

velikosti $M/2$. Pokud funkce popisující počty cache missů splňují tzv. podmínku regularity, způsobuje tedy sdílení cache s jiným algoritmem nejvýše konstantní zvýšení počtu cache missů. Regularita je přitom splněna u prakticky všech známých cache-oblivious algoritmů; z předchozí části práce ovšem plyne snadná konstrukce neregulárních algoritmů (nicméně neúčinných).

Práce přináší nové výsledky a zajímavým způsobem zobecňuje předchozí modely proměnlivé cache. Je psána příjemnou, čtivou angličtinou s naprostým minimem jazykových chyb. Použité zdroje jsou korektně citovány.

Práci proto doporučuji přijmout jako diplomovou.

V Praze dne 30. května 2017
Martin Mareš