

Posudek vedoucího na práci Lukáše Folwarczného

On the Hardness of General Caching

Předložená práce se zabývá několika otázkami algoritmů pro obecné cachování. Jedná se o teoretickou práci z oblasti algoritmů, jejich návrhu a složitosti.

Cachování je klasický problém se silnou motivací dvouúrovňových paměťových systémů, původně motivovaný stránkováním vnitřní paměti počítačů a ukládáním momentálně nepotřebných stránek na disk. Současné systémy motivaci dále rozšiřují jednak na nižší úroveň na víceúrovňové paměti na úrovni procesorů a jednak na vyšší úroveň na cachování webových stránek či jiných souborů různé velikosti. Právě cachování stránek/souborů různé velikosti je hlavní motivací studovaných problémů.

Základní (a původní) verze cachování pracuje se stránkami jednotné velikosti. Pro tento případ je znám jednoduchý optimální algoritmus – tzv. Beladyho pravidlo. Tento algoritmus ovšem potřebuje znát celý vstup, tj. pro optimální zpracování daného požadavku potřebuje znát i budoucí požadavky. To samozřejmě pro praktické použití není možné, a přirozeně se od začátku studia online algoritmů problém cachování stal jedním ze základních studovaných problémů a výsledky jsou velmi dobré, jsou známy deterministické i pravděpodobnostní optimální algoritmy.

S nástupem WWW bylo proto od 90. let přirozené studovat i algoritmy pro obecné cachování, kdy různé stránky mají různou velikost a cenu. I v tomto případě jsou výsledky studia online algoritmů velmi dobré. Oproti tomu dlouhou dobu nebylo známo nic o složitosti optimálních řešení. Tento známý problém i přes velké úsilí dlouho odolával řešení a NP-úplnost základních variant byla dokázána až v průlomovém článku

M. Chrobak, G. J. Woeginger, K. Makino, and H. Xu,
Caching is hard – even in the fault model, *Algorithmica*, 63 (2012), pp. 781-794.
A preliminary version appeared at ESA 2010..

Tento článek však v důkazu NP-úplnosti podstatným způsobem používá stránky větší než polovina paměti, což neodpovídá motivaci a tedy není příliš uspokojivé.

Hlavním výsledkem první kapitoly je důkaz NP-úplnosti obecného cachování se stránkami omezené velikosti, konkrétně na stránky velikosti nejvýše 3, navíc při omezení cen na jednotkové ceny nebo na ceny rovné velikosti stránek. Tím dává definitivní odpověď na otázku výpočetní složitosti tohoto problému.

Další dvě kapitoly se zabývají otázkou cachování stránek stejné velikosti v paměti, jejíž velikost se mění. K tomuto účelu se studuje reprezentace možných konfigurací, která dává úplný popis možných konfigurací optimálního algoritmu po dané posloupnosti požadavků. Tato technika pro základní cachování byla zavedena v klasických člancích

L. A. McGeoch and D. D. Sleator, A strongly competitive randomized paging algorithm, *Algorithmica*, 6 (1991), pp. 816-825.

E. Koutsoupias and C. H. Papadimitriou, Beyond competitive analysis, *SIAM Journal on Computing*, 30 (2000), pp. 300-317.

Výsledky kapitol 2 a 3 poněkud zjednoduší prezentaci této techniky a především ji zobecňují pro proměnlivou velikost paměti.

Motivací tohoto zobecnění je snaha najít polynomiální algoritmy pro některé varianty cachování se dvěma velikostmi stránek. Některé takové speciální případy jsou popsány v poslední kapitole. Tato oblast výzkumu se jeví jako perspektivní s dobrou šancí na další výsledky.

Článek s výsledky první kapitoly byl zaslán k publikaci na konferenci ISAAC 2015. Odpovídající abstrakt byl přijat k prezentaci na workshopu MAPSP 2015.

Podíl Lukáše Folwarzného na společném výzkumu je velmi podstatný, většina klíčových nápadů v podaném článku i ve zbytku práce je jeho (s výjimkou jednoduchého dolního odhadu v první kapitole). Sám sepsal i většinu textu. Domnívám se, že prezentace důkazů je velmi slušná, včetně angličtiny.

Předložená práce tedy obsahuje kvalitní nové výsledky, které jsou také solidně prezentované. Dle mého názoru je na úrovni kvalitní diplomové práce, z čehož vyplývá jednoznačné doporučení k obhajobě jako bakalářské práce.

5. srpna 2015

prof. RNDr. Jiří Sgall, DrSc.