



Review of the Ph.D. Thesis *Increasing Performance and Maintaining Data Consistency in Mobile Devices for Distributed File Systems*

Submitted by Ing. Pavel Bžoch

The thesis addresses important topics of files access acceleration from mobile devices together with caches' consistency maintenance in frame of distributed file systems (DFS). The idea is to bring in these area novel efficient algorithms for maintaining cache consistency while respecting limited cache capacity available. The contribution of the thesis consists in proposing novel caching policies employing both server statistics and local statistics for replacement decisions to decrease network traffic and the cost of counting the priority of data unit whilst staying in cache. Moreover, new algorithms were developed to prevent cache inconsistent states.

The dissertation is presented in seven chapters and complemented by appendix with summary of author activities, and by bibliography. After introduction and state of the art of the discussed area with a review of four traditional DFS, their properties, expected trends and the proper motivation for research, the third chapter concisely introduces an experimental DFS, which is under development at the author's department. Evidently, this DFS is utilized for experimental evaluation of the proposed algorithms. The rest of the thesis discusses original contributions of the work presented, namely caching policies and consistency control, and evaluation of original algorithms.

I appreciate the work done by the applicant in the field of improving performance and maintaining data consistency in mobile devices for distributed file systems. His contribution appears as a novel input both from the viewpoint of presented conception and sophisticated algorithms that were elaborated in detail and tested in a running environment of the experimental DFS.

In conclusion, I can summarize that the thesis is devoted to the very important and topical subject. According to my opinion, the submitted framework represents considerable contribution to the field of DFS with respect to mobile devices. The author's approach to the problem solving, methods applied and research achievements demonstrated by this dissertation proves his qualified competence to participate successfully in current research and development activities. Concurrently, candidate's publications fulfill common requirements for the Ph.D. degree. For these reasons I can express my full confidence in the candidate's capabilities for the scientific work.

I therefore recommend the acceptance of this thesis for the fulfillment of candidate's studies.

Brno, December 10, 2014



prof. Ing. Miroslav Švéda, CSc.
Brno University of Technology
Faculty of Information Technology

Oponentský posudek dizertační práce
Ing. Pavel Bžoch, KIVT FAV ZCU
Increasing Performance and Maintaining Data Consistency in
Mobile Devices for Distributed File Systems

Hodnocení významu pro obor. Předložená práce řeší jednoznačně aktuální a důležité téma: efektivní způsoby přístupu k vzdáleným datům v cloudu pro mobilní klientská zařízení, které nemají zaručenou stabilní konektivitu a nemají ani dostatečné klientské zdroje.

Hodnocení obsahu práce, postupu řešení, použitých metod. Dizertační práce Ing. Bžocha přináší několik nových algoritmů pro zefektivnění funkce distribuovaných systémů souborů (DFS) pro mobilní platformy.

Práce obsahuje v kapitole 2 rešerši existujících tradičních i výzkumných DFS a diskutuje podrobně otázky vysoké dostupnosti, spolehlivosti a zvyšování výkonnosti pomocí replikace dat a kešování, dále otázky bezpečnosti a konzistence dat. Na pozadí této rešerše pak definuje cíl dizertační práce: výzkum metod, které umožní zvýšení efektivnosti a zajištění konzistence dat při přístupu k datům v DFS z mobilních zařízení. Tato specifikace cílů je velmi vágní a stručná.

V kapitole 3 popisuje slovně architekturu a funkčnost fakultního DFS zvaného KIV-DFS, který byl navržen a je vyvíjen s cílem poskytovat lepší podporu pro mobilní zařízení a pro dizertanta je prostředím pro experimentování se zkoumanými metodami.

Obecně konstatuji, že kapitola 4 má na dizertační práci neovyklou logiku a strukturu. Na začátku kapitoly pak dizertant podrobně analyzuje nejnovější řešení v oblasti algoritmů pro nahrazování položek při zaplnění během kešování, jak na straně klienta tak na straně serveru. V sekci 4.2. je pod názvem Caching units zdůvodněno rozhodnutí, že další výzkum bude uvažovat pouze kešování celých souborů. V sekci 4.3. pak začne dizertant náhle hovořit o svých řešeních, uvedené menší neformální analýzou, následovanou popisem dvou nových heuristik, které tvoří jeden z hlavních přínosů práce. Motivační rámec je vymezen předpokladem, že mobilní zařízení mají omezenější výpočetní kapacitu než běžné PC, aniž dizertant upřesnil, co všechno zahrnuje do kategorie mobilních zařízení. Vzhledem k tomu, že se práce zabývá DFS, měly by být zahrnuty chytré telefony, kde toto omezení neplatí. Hlavní myšlenkou navržených heuristik pro zlepšení kvality nahrazování položek během kešování je pro rozhodnutí využívat statistiku, kterou shromažďuje o daném souboru server.

Podobnou metodiku výkladu dizertant zvolil v kapitole 5 pro druhou část výsledků své dizertace, které se týkají metod udržování konzistence obsahu keší, čili zajištění, že data získaná z keší jsou stejná jako ta, která by za stejných podmínek klient získal ze vzdáleného zdroje, kdyby se kešování nepoužilo. Výklad zakládá na známé 3rozměrné taxonomii. Opět, v sekci 5.2. bez jakékoli předchozí analýzy dizertant popisuje své dva algoritmy, i když u toho prvního - pro uživatele se spolehlivým připojením - a popsaného velmi neformálně, není příliš zřejmé, zda se jedná o dizertantův přínos. Myšlenka porovnávat lokální a serverové číslo verze souboru je triviální. Přestože dizertant v diskuzi konstatuje, že se jedná o řešení s jednoduchou implementací, nediskutuje kvantitativně výkonnostní problémy takového přístupu a navíc není jasné, proč je tento jednoduchý algoritmus na str. 56 označen jako "novel algorithm for maintaining client-side cached data".

Západočeská univerzita v Plzni

Doručeno: 04.03.2015
ZCU 005851/2015

listy: 6 přílohy:
druh:



zcupese4fef5

consistency for mobile clients”, když v diskuzi na další stránce se konstatuje, že je “suitable for clients with wired connection or for clients with reliable mobile connection....”, neboť motivační rámec celé práce je absence takovýchto příhodných podmínek.

V kapitole 6 je shrnutí experimentálních výsledků, kterými dizertant srovnává vlastnosti navržených algoritmů, a to na úrovni simulací. Navrhl vlastní metodiku pro srovnávání algoritmů pro zajištění konzistence. Měření byla prováděna na základě simulací, pro ten účel si napsal vlastní simulátor a experimentoval s různými rozděleními vstupních parametrů.

Souhrnně lze obecně konstatovat, že dizertant zvolil správnou metodu, založenou na důkladném testování a srovnávacích měření svých vylepšených a stávajících algoritmů. Za slabší považuji absenci diskuze o možnostech řešení kešovacích politik a zajištění konzistence pro mobilní zařízení. Po popisech stávajících řešení rovnou popisuje svá vylepšení, aniž provede analyzuje důvody pro takovou volbu. Čtenář se o inspiracích či návrhových rozhodnutích může pouze domýšlet.

Další připomínky a komentáře

- Str. 25: Co znamená “Method pre-creating file object is similar to ...”?
- Str. 32: Chyba v označení n a N .
- Str. 36: V popisu KIV-DFS se v 3. odstavci sekce 3.6. říká, že “The database is replicated on each node for increasing availability and stability.” To je ale v rozporu s tím, že KIV-DFS je DFS. Takové řešení neškáluje. Na téže stránce se hovoří o “table replicas”, ty ale nejsou vysvětleny.
- Str. 38, začátek 2. odstavce. Není zřejmé, zda pod označením “the system” je myšlen KIV-DFS či jiný.
- Str. 44: Závěr v textu sekce 4.2 neodpovídá obr. 4.2., na který se odkazuje, kde jsou evidentně soubory nad 5MB.
- Str. 48: Výraz v definici (3) není vysvětlen. Proč je P_{LFU-SS} takto definován? Jak je zaručeno, že jmenovatel není nikdy roven nule?
- Str. 49: Přesně stejná připomínka k definici (5). Co to jsou “computation priorities”?

Zhodnocení splnění cílů, výsledků a přínosu práce Vzhledem k tomu, že cíle jsou v práci vytyčeny velmi volně (jedna prostá věta), je zhodnocení splnění či nesplnění cílů značně subjektivizovatelné. Nicméně čistě formálně vzato, dizertant cíle splnil vytvořením a otestováním několika vlastních návrhů vylepšení kešovacích politik a zajištění konzistence keší. Přínos práce spočívá v návrhu nových algoritmů a experimentální prověření, za jakých podmínek poskytují lepší výsledky než jiné stávající postupy. Zda jsou tyto výsledky přínosné i pro praktické nasazení nad současnými technologiemi a budou tedy i nasazeny komerčně, není z práce zřejmé.

Zhodnocení formální úpravě a jazykové úrovni práce Jak jsem již naznačil výše, v práci chybí podrobnější specifikace cílů a diskuze alternativ a inspiračních rámců navržených algoritmů. Práce je sepsána v angličtině, která je naneštěstí značně zatížena

častými drobnými chybami (členy, interpunkce). Několik vět je nesrozumitelných. Zdroje jsou citovány pečlivě.

Zhodnocení publikačních aktivit Publikační seznam obsahuje 2 články ve Scopus časopisech a několik příspěvků na zahraničních konferencích. Lze tedy konstatovat, že výsledky práce byly dostatečně publikovány. Dva pro první část dizertační práce stěžejní příspěvky na konferencích [75,76] mají 4 spoluautory, je třeba ujasnit autorský podíl dizertanta.

Závěr Předložená práce prokazuje dizertantovu schopnost samostatné výzkumné činnosti a obsahuje původní a uveřejněné výsledky. V souladu s čl. 47 Zákona o vysokých školách v platném znění předloženou dizertační práci **doporučuji** k obhajobě s cílem získání titulu PhD.

V Praze, 28.2.2015

Prof. Ing. Pavel Tvrdík, CSc.



