FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Student:	Bc. Jakub Dvořák
Oponent práce:	Ing. Ivo Petr, Ph.D.
Název práce:	Porovnání algoritmů pro faktorizaci velkých celých čísel
Obor:	Počítačová bezpečnost

Datum vytvoření: 8. 6. 2020

Hodnotící kritérium:	*Způsob hodnocení – následující škálou 1 až 4:*
1. Splnění zadání	*1=zadání splněno,* *2=zadání splněno s menšími výhradami,* ***3=zadání splněno s většími výhradami,*** *4=zadání nesplněno*

Popis kritéria:
Posuďte, zda předložená ZP dostatečně a v souladu se zadáním obsahově vymezuje cíle, správně je formuluje a v dostatečné kvalitě naplňuje. V komentáři uveďte body zadání, které nebyly splněny, posuďte závažnost, dopady a případně i příčiny jednotlivých nedostatků. Pokud zadání svou náročností vybočuje ze standardů pro daný typ práce nebo student případně vypracoval ZP nad rámec zadání, popište, jak se to projevilo na požadované kvalitě splnění zadání a jakým způsobem toto ovlivnilo výsledné hodnocení.

Komentář:
Zadání práce, má-li být splněno pečlivě, považuji za náročné, přesahující rámec diplomové práce. Student si vzal za cíl nastudovat, popsat a implementovat známé algoritmy řešící problém faktorizace složených čísel. Text práce však považuji za velmi nepovedený, popis algoritmů neúplný či zmatený. Čtenáři nepřináší vhled do problému a spíše připomíná kuchařku. Algoritmy jsou všeobecně známé, přidanou hodnotou práce tedy měla být jejich paralelizace a porovnání sekvenční a paralelizované verze. Navržená paralelizace je v lepším případě naivní, v případě kvadratického síta jde ale dokonce proti duchu celého celého algoritmu. Testování algoritmů probíhá na instancích kde sofistikovanější algoritmy s lepší asymptotickou složitostí nemají šanci porazit primitivní kolizní algoritmy a k reálnému srovnání metod tedy v práci nedošlo.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
2. Písemná část práce	*40 (F)*

Popis kritéria:
Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části. Dále posuďte, zda předložená ZP je po věcné stránce v pořádku, případně vyskytují-li se v práci věcné chyby nebo nepřesnosti. Zhodnoťte dále logickou strukturu ZP, návaznosti jednotlivých kapitol a pochopitelnost textu pro čtenáře. Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3. Posuďte, zda student využil a správně citoval relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami. Zhodnoťte, zda převzatý software a jiná autorská díla, byly v ZP použity v souladu s licenčními podmínkami.

Komentář:
Text práce je zřejmě její největší slabinou. Na první pohled je sice práce logicky dobře členěna, čtenář však velmi brzy jistí, že student při výkladu spíše používá otřepané fráze než aby problematiku vysvětlil. První kapitola má zavést potřebné algebraické pojmy a student přiznává že se jedná o výtah ze studijních materiálů. Text následujících kapitol je však tak vágní, že se k využití zavedených pojmů vůbec nedostane. Naopak student dále pracuje s pojmy které nezavedl. Druhá kapitola věnovaná systému RSA má zjevně sloužit jako motivace ke studiu samotného problému. Student ale nikde neuvede jak bezpečnost/prolomení systému souvisí s problémem faktorizace. RSA je navíc představen ve své nejelementárnější formě, v jaké jej v praxi nelze použít. Vůbec není jasné jak s prací souvisí kapitola 2.5 týkající se útoků postraními kanály, nebo úvodní zmínka o Shorově algoritmu. Kapitola 3 obsahuje velmi zevrubný popis procedur které je třeba provést v jednotlivých algoritmech, zcela ale uniká jejich podstata. Student nevysvětlí jak volit funkci f v Pollardově rho-metodě či polynomy v kvadratickém sítu, pro jaká a selhává Pollardova p-1-metoda ani společné algebraické pozadí Lenstrovy faktorizace a Pollardovy p-1 metody či princip číselného síta. Není jasné jak mají být voleny parametry algoritmů a např. způsob jakým jsou hledána kvadratická rezidua je jen tak mimochodem zmíněn až v kapitole 6.1.1. Asymptotické složitosti algoritmů jsou převzaté a čtenář nemá jakékoliv vodítko k tomu aby k těmto údajům dospěl. Následující kapitoly týkající se implementace obsahují spíše návod na použití knihoven GMP, FLINT a OpenMPI. Vzhledem k vágnosti textu nemohu zaručit že jsem zcela pochopil studentův způsob řešení problému paralelizace.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
3. Nepísemná část, přílohy	*45 (F)*

Popis kritéria:
Dle charakteru práce se případně vyjádřete k nepísemné části ZP. Například: SW dílo – kvalita vytvořeného programu a vhodnost a přiměřenost technologií, které byly využité od vývoje až po nasazení. HW – funkční vzorek – použité technologie a nástroje, Výzkumná a experimentální práce – opakovatelnost experimentů

Komentář:

Způsob paralelizace algoritmů je velmi neurčitě popsán v kapitole 4. Pokud popisu dobře rozumím, jedná se u prvních třech algoritmů o velmi naivní paralelizaci, kdy je randomizovaný algoritmus spuštěn paralelně v několika vláknech a ta spolu soupeří. Tato myšlenka skutečně může vést ke zrychlení, neboť algoritmus nemusí nutně dát výsledek. Nerozumím ale proč vstupní hodnoty (třeba eliptickou křivku) generuje hlavní proces a pak čeká až doběhnou všechna vlákna (žádné nemusí dát výsledek) než pošle novu sadu stupních hodnot. To paralelizaci naopak činí neefektivní. Dle popisu v kapitole 4.4 student paralelizuje kvadratické síto tak, že rozdělí faktorizační bázi. To je ale zcela proti smyslu celého algoritmu. GNFS je "paralelizováno" podobně jako předchozí algoritmy, tedy každé vlákno počítá samo se svým polynomem. Nejedná se tedy o paralelizaci v pravém slova smyslu. Ta by snad nastala kdyby proces generování kongruencí probíhal paralelně a po nalezení patřičného počtu kongruencí by hlavní proces provedl další kroky. Samostatnou kapitolou je porovnání výkonu algoritmů, které porbíhá v oblasti, kde primitivní algoritmy zdaleka předčí sofistikované algoritmy. Student to uvádí v závěru práce, taková rešerše ale měla proběhnout hned ze začátku.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Hodnocení výsledků, jejich využitelnost	*50 (E)*

Popis kritéria:
Dle charakteru práce zhodnoťte možnosti nasazení výsledků práce v praxi nebo uveďte, zda výsledky ZP rozšiřují již publikované známé výsledky nebo přinášející zcela nové poznatky.

Komentář:

Práce uvádí základní varianty zkoumaných algoritmů. Při řádném provedení rešerše a kvalitní diskuzi by mohla sloužit jako studijní text. K tomu by ale nutné text silně revidovat. Úspěšná paralelizace algoritmů by mohla mít praktický přínos. Studentem navržená paralelizace je ale v lepším případě naivní a neefektivní, využití v praxi je velmi nepravděpodobné.

Hodnotící kritérium:	*Způsob hodnocení – nehodnotí se*
5. Otázky k obhajobě	

Popis kritéria:
Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:

- Objasněte algebraický princip GNFS a rozdíly mezi bázemi které v algoritmu využíváte.
- Objasněte vámi navržený způsob paralelizace u Pollardovy rho-metody, kvadratického síta a GNFS. Jak se váš návrh liší od implementací zmiňovaných v kapitole 6.7?

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Celkové hodnocení	*45 (F)*

Popis kritéria:
Shrňte stránky ZP, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení nemusí být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích. Obecně platí, že bezvadně splněné zadání je hodnoceno klasifikačním stupněm A.

Text hodnocení:

Vzhledem k náročnosti tématu by bylo vhodnější, aby se student zaměřil na menší množství algoritmů, kterým by se v práci řádně věnoval, než aby se povrchně věnoval všem. Doplnění základů jednotlivých metod a řádné prozkoumání možností paralelizace by práci velmi pomohlo.

Vzhledem k velkému množství závažných nedostatků nedoporučuji práci v současném stavu k obhajobě.

Podpis oponenta práce: