POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	**Password recovery job scheduling for online deep file analysis**
Jméno autora:	**Bc. Petr Kubelka**
Typ práce:	diplomová
Fakulta/ústav:	Fakulta elektrotechnická (FEL)
Katedra/ústav:	Katedra počítačů
Oponent práce:	Ing. Martin Schaefer
Pracoviště oponenta práce:	Katedra počítačů

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání **průměrně náročné**

Hodnocení náročnosti zadání závěrečné práce.

Zadání považuji za průměrně náročné, ovšem práce jde za požadavky zadání v implementační softwarové části, pokud by i toto bylo zamýšleno jako součást zadání, hodnotil bych zadání jako náročnější.

Splnění zadání **splněno**

Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

Práce je oproti zadání výrazně rozšířena v implementační softwarové části, kdy je popsán a vytvořen kompletní produkčně nasaditelný systém například včetně integrace existujících nástrojů na prolamování hesel. Zadání je zaměřeno na rozvrhování úloh, na formalizaci problému, výběr vhodné metody a její implementaci a experimentální ověření rozvrhovací metody. Všechny body jsou splněny, avšak u některých mám připomínky k jejich provedení a/nebo prezentaci v textu.

Zvolený postup řešení **částečně vhodný**

Posuďte, zda student zvolil správný postup nebo metody řešení.

Domnívám se, že student dospěl ke správnému řešení, ale text práce mne nepřesvědčuje o správném postupu ve všech krocích. Jádro problému vidím v nejasné identifikaci kritéria hodnocení kvality rozvrhování. Ve formalizaci a následně v experimentech je často zmíněn C max, tedy délka plánu rozvrhu. To ovšem plně nekoresponduje s tím, o co se student svým řešením napříč prací snaží dosáhnout. Ve finále je rozvrhování spuštěno v online režimu, a pomocí prioritizace jsou preferovány nové úlohy, aby se, pokud možno, brzy vyřešily úlohy, které jsou „rychle“ vyřešitelné. Minimalizace délky plánu, pokud vůbec dává nějaký smysl v online režimu, obecně povede k minimalizaci doby doběhnutí všech úloh, což dle mého pochopení neodpovídá požadavkům. Ač to není explicitně řečeno, z práce si odnáším poznatek, že zásadním požadavkem na řešení je schopnost preempce úloh, která umožní nejen úlohy lépe poskládat pro kratší délku plánu (což je možná nepodstatné), ale zejména umožní „rychlé“ úlohy odbavit dříve. Jde možná o problém volně formulovatelný jako maximalizace očekávaného počtu úspěšně dokončených úloh v každém čase. Pro případnou další práci studenta si dovolím zmínit, že např. pro evakuace existuje v principu podobný problém tzv. „Earliest Arrival flows“.
Celkově mám z práce dojem, že nejde o teoreticko-algoritmickou práci, ale ani softwarově-inženýrskou, výsledná práce je pak kombinací, která nemá plně dotažené náležitosti ani jedné varianty.

ČVUT
ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

Odborná úroveň **B - velmi dobře**

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Student projevil schopnost aplikovat znalosti pro tvorbu sofistikovaného softwaru. Možnost pracovat s reálnými daty je v případě této práce důležité, reálná data resp. use case nasazení implementovaného řešení je stěžejní pro samotnou formalizaci, volbu řešení i evaluaci. Je proto škoda, že v práci není dobře popsáno odkud se data vzala. Tím spíše, že data pravděpodobně nejsou veřejná, bylo by vhodné tato data popsat, aby bylo možné posoudit relevantnost experimentů, např. předpokladu, že „přichází 30 souborů“ za hodinu.

Formální a jazyková úroveň, rozsah práce **C - dobře**

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Práce je psána anglicky, text je dobře čitelný a srozumitelný. Za nedostatek považuji chybějící referencování grafů (Figures, pseudokodu) z textu. Některé obrázky stejného typu jsou ve vektorové verzi (6.6) jiné ne (6.4). Ve struktuře práce mi nesedí umístění algoritmů do kapitoly „Formalization“, algoritmům je s přihlédnutím k jejich důležitosti v práci a v poměru k jiným částem věnováno málo prostoru.

Výběr zdrojů, korektnost citací **A - výborně**

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Zdroje jsou využity korektně.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Implementovaný systém ve formě microservices v jednotlivých docker images je přiložen k práci spolu s uměle vytvořenými daty. Přílohy obsahují i skripty pro zpracování výstupů z experimentů do grafů. Bez řádné dokumentace jsem nebyl schopen implementaci spustit a otestovat. Nemám indicie se domnívat, že by řešení studentovi nefungovalo tak, jak popisuje, ocenil bych však připravenou alespoň částečnou demonstraci funkčnosti implementace.

Drobné komentáře:
Etická otázka: Je prolamování hesel za účelem ověření obsahu eticky v pořádku? Není specifikováno kdo a kdy tuto činnost má provádět. V Introduction je nastíněno několik možností, kdy je prolamování hesel využíváno. Je pouze nastíněno, že pro využití pro detekování škodlivého kódu, očekával bych však explicitní vypořádání se s etickou otázkou prolamování hesel.

Po přečtení Introduction si stále nejsem jist co je přesně úlohou. Nejspíše prolamování hesel, ale pokud chceme optimalizovat rozvrhování, potřebujem vědět, jaká jsou kritéria, co vlastně rozvrhujeme a s jakým cílem. V této souvislosti je Related Work přehledem technik bez jasné vazby na řešený problém.

Z textu nerozumím „encryption vs. hashing“. Nevím totiž, co je relevantní pro tuto práci, 4.1 říká „Since this master thesis is focused on the recovery of passwords from files, our main interest is cryptographic hashing.“ Není vysvětleno proč.

Máme k dispozici password hashes nebo nemáme? Jak funguje ověřování správnosti hesla? Password hashes se nasledně zmiňují až na str. 34 v rámci implementace.

Není vysvětleno, co jsou „rounds for password verification".

Pokud r_i je proměná, znamená to, že solver nastavuje hodnotu pravděpodobnosti a že úloha bude trvat déle? Neměla by to být vstupní konstanta?

6.6: Nejasnost v diagramech: co když úlohu "cracknu" posledním útokem? Pak nejspíš soubor nepošlu Unpackeru.

7.16 Pravděpodobně špatný čas v popisku.

7.1.1 a 7.1.2 Cracking tools mají rozdílné jednotky, ale význam je stejný - proč to není sjednoceno, když se to porovnává?

7.1.5 Zásadně chybí diskuse výsledků - co ty výsledky znamenají, co vidíme na Figure 7.6?

7.1.5 a 7.1.6. není dobře vysvětlen smysl experimentů.

7.2.1 Až v této části vysvětlen význam konstanty„weight" s kterou student prezentuje algoritmus a následně experimentuje. Nejspíše jsou prohozené popisky v legendě grafů.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Hodnocená práce přes popsané nedostatky splňuje zadání a náležitosti diplomové práce.
Mezi hlavní nedostatky považuji textovou prezentaci odvedené práce. Jazyk a forma je na dobré úrovni, nedostatky vidím zejména ve faktickém obsahu a částečně ve struktuře práce. Naopak pozitivem je rozsah odvedené práce v implementaci systému.

Otázky:
Co je podle Vás nejhodnotnějším zjistěním/výsledkem Vaší práce?
Diskutujte prosím mnou výše nastíněnou pochybnost o zvoleném optimalizačním kritériu.

Předloženou závěrečnou práci hodnotím klasifikačním stupněm **C - dobře.**

Datum: 3.6.2020 Podpis: