FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Jaroslav Pešek
Student:	David Kežlínek
Název práce:	Analýza časových řad v exportéru síťových toků
Obor / specializace:	Bezpečnost a informační technologie
Vytvořeno dne:	22. května 2023

Hodnotící kritéria

1. Splnění zadání

- ▸ [1] **zadání splněno**
  [2] zadání splněno s menšími výhradami
  [3] zadání splněno s většími výhradami
  [4] zadání nesplněno

Zadání práce považuji spíše za náročnější, neboť se autor musel vypořádat s netriviální statistickou teorií, musel implementovat kód v jazyce C++ do existujícího open source projektu a pak dílo otestovat a vyhodnotit.

2. Písemná část práce 85 / 100 (B)

Rozsah práce je přiměřený. Práce je logicky členěná do schématu rešerše a teorie – analýza – implementace – testování/vyhodnocení, což je pro tento typ práce ideální. Práce obsahuje minimum gramatických chyb, několik málo překlepů (iplementace, bitflow místo biflow), ale nepřehlédnutelné množství typografických chyb a nekonzistencí (přetékající řádky, popisky tabulek jsou jednou nad tabulkou, jindy pod tabulkou, nekonzistentní reference na zdroje aj.).

Citování je bez výhrad, použity byly relevantní zdroje.

Analýza a rešerše jsou poctivé, sice je občas podána neúplná informace (například v kapitole 1.2.1 je popsáno pouze několik možných vstupů exportéru, ne všechny), s přihlédnutím k účelu práce je to ale zanedbatelný problém. Kapitola Analýza časových řad obsahuje vyčerpávající seznam vzorečků pro výpočet atributů – příště by stálo za úvahu takový seznam dát spíše do přílohy. Kapitola Implementace je bez výhrad. Kapitola věnovaná testování je stručnější než by bylo vhodné. K tvrzení o správnosti řešení bych očekával více dat, která podpoří důvěru k autorově implementaci. V kapitole věnující se vyhodnocování a experimentování bych také nečekal spekulativní výroky, jako například "Po nasazení do reálně sítě by plugin dosahoval lepších výsledků." – lepší by bylo toto tvrzení podložit daty. Jedním z cílů práce bylo vyhodnotit nasaditelnost do

vysokorychlostní sítě, ale tento cíl je odbytý jednou větou, která je navíc poměrně vágní. Oceňuji měření režie, což je nad rámec zadání.

3. Nepísemná část, přílohy 89 /100 (B)

Výsledkem práce je plugin do open source exportéru síťových toků ipfixprobe. Kód je členěn na dva soubory, hlavičkový a implementační. Kód je poměrně rozsáhlý, proto by bylo na místě jej pro lepší čitelnost rozdělit do více souborů.
Mimo malý počet nedobrých konstrukcí nejzávažnější problém je, že kód postrádá robustnější memory management – jsou zde alokace objektů, které však nejsou dealokovány a chybí kontroly, jestli alokace proběhla úspěšně a nebyl z nejrůznějších důvodů navrácen nulový ukazatel, což může v takovém mezním případě vést k nestabilitě celého programu.

Čitelnost kódu není optimální, doporučil bych dlouhé metody rozdělit do více menších a lépe čitelnějších celků. Jsou zde drobné prohřešky vůči kultuře kódu, například používání velkých písmen pro členské proměnné. Dokumentace je přímo v kódu, což je zcela v pořádku, ale u takto netriviální implementace by budoucí vývojáři jistě ocenili poněkud extenzivnější formu.

Verifikace funkcionality byla ověřována proti již existujícím implementacím v oboru astrofyziky (knihovna AstroPy).

Až na zmiňované prohřešky je kód adekvátní bakalářské úrovni. Nutno ocenit, že se autor dobře vypořádal s přispěním do open source projektu, který samozřejmě musel nastudovat a je na místě uvést, že projekt není nekomplikovaný.

4. Hodnocení výsledků, jejich využitelnost 95 /100 (A)

Autorovi se povedlo implementovat výpočet specifických atributů síťových toků, které v jiných exportérech nejsou k dispozici. Nasaditelnost takového řešení ve vysokorychlostních sítí je spekulativní, neboť se jedná o výpočetně náročné operace, to však plyne z charakteru řešeného problému. Uplatnitelnost pro offline exporty toků nebo v sítích o malých rychlostech je reálná možnost. Atributy, které jsou exportovány jsou dobře využitelné pro analýzu i jako vysvětlující proměnné do různých klasifikačních problémů.
Byl vytvořen pull request pro merge do hlavní větve open source projektu, ale pro schválení bude potřeba provést refactoring. Věřím, že pro výzkum v oblasti síťového provozu autorova práce nalezne uplatnění.

Celkové hodnocení 90 /100 (A)

Bylo zmíněno několik nedostatků, kterými práce trpí, ale ve výsledku student vykonal mnoho kvalitní práce. Za nejvážnější nedostatek považuji slabší kapitolu o testování, kterou bych uvítal obsáhlejší a závěry, které autor činí, více explicitní. Nedostatky v implementaci jsou napravitelné snadno a jsou způsobeny spíše nezkušeností.
Výsledkem však je kvalitní dílo.

Otázky k obhajobě

Jak přesně proběhla verifikace správnosti výsledků pluginu?

Porovnejte výkonnost exportéru ipfixprobe se spuštěným pluginem a bez pluginu. Jaký pokles výkonnosti pozorujeme?

V jakých sítích je vaše řešení nasaditelné? Jaké problémy bychom museli řešit, pokud bychom jej chtěli nasazovat do vysokorychlostní sítě?

Instrukce

Splnění zadání

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.

Písemná část práce

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. 52/2021, 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.

Nepísemná část, přílohy

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.

Hodnocení výsledků, jejich využitelnost

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.

Celkové hodnocení

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.