FAKULTA
INFORMAČNÍCH
TECHNOLOGIÍ
ČVUT V PRAZE

Hodnocení vedoucího závěrečné práce

Vedoucí práce:	Ing. Josef Koumar
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

Student musel implementovat atributy paketových časových řad jednotlivých síťových toků do reálného exportéru ipfixprobe, který musel nastudovat, a navíc v seznamu atributů se nacházeli atributy vytvořené z Lomb-Scarglova periodogramu jehož implementace je velice náročná. Tudíž zadání práce přesahuje průměrnou náročnost bakalářské práce.

Student naimplementoval prototyp pluginu pro open source exportér síťových toků ipfixprobe pro rozšíření IP flow záznamů o atributy získané analýzou časové řady paketů. I přes to, že student ode mě obdržel seznam atributů a jejich matematické reprezentace, tak se aktivně a svévolně podílel na finální podobě a seznamu atributů. Dokonce našel a opravil chyby, kterých jsem se v seznamu atributů dopustil. Finální rozřazení atributů do jednotlivých podpluginů student navrhnul s ohledem na způsob výpočtu, díky čemuž optimalizoval jejich výpočet. Také navrhnul přidání exportování několika atributů, které vznikly mezivýpočty u jinýhc atributů. Student vhodně a detailně otestoval plugin na reálných datech i síti, přičemž odhadnul i zátěž sítě, na které plugin dokáže fungovat bez problému. Student se snažil i vyhodnotit nasaditelnost na reálné vysokorychlostní síti. Jako pomyslnou tečkou, kterou student zakončil tuto práci je pull request do ipfixprobe, čímž práci plně splnil.

2. Písemná část práce

95 / 100 (A)

Text je členěn do několika vhodně tématicky oddělených kapitol. Text obsahuje přiměřené množství drobných stylistických chyb a překlepů, které ale nijak nezhoršují kvalitu čtení a přijetí výsledků práce ze samotného textu práce, proto považuji za nesmyslné je

promítnout do hodnocení. Rozsahově je práce nadprůměrná a zajišťuje detailní popis výsledků práce. Text detailně diskutuje i volbu datových struktur a jejich efektivitu pro různé atributy. V textu jsou přiloženy části kódu vhodné pro svoji důležitost ve výsledném prototypu.

V kapitole Testování se student zaměřil na správnost implementace jak atributů, které student porovnal s výstupem Python verze, kterou ode mě student obdržel spolu se seznamem atributů, tato implementace nicméně používá pro spočtení atributů řadu knihoven, které v C++ nejsou k dispozici, a tak studentovi tento Python kód sloužil pouze k ověření správnosti implementace. Student se také zaměřil na správnost implementace Lomb-Scargle periodogramu přes NFFT. V práci tento počin vhodně prezentuje porovnáním s Python knihovnou Astropy pomocí grafu.

Student dále testoval samotný plugin jak dvěmi způsoby na anonymizovaných PCAP záchytech národní infrastruktury CESNET, které jsem mu poskytl, tak i na reálné lokální síti. Výsledky testování z pohledu časové náročnosti by mohli obsahovat více výsledků, například kolik dat bylo zpracováno a jak dlouho prototyp běžel při testování na lokální síti. Testování náročnosti na paměť študent vhodně a přehledně prezentuje prostřednictvím grafů.

Citační styl studenta je v souladu s předmětem DPR. Student citoval 36 zdrojů z toho většina jsou konferenční či časopiskecké publikace. Online zdroje jsou odkazy na dokumentaci či git repozitáře použitých knihoven. Proti citacím nemám žádné výhrady.

Celkově hodnotím text jako povedený s drobnými nedostatky a hodnotím jej 95/100.

3. Nepísemná část, přílohy 98 /100 (A)

Příloha obsahuje prototyp pluginu pro exportér síťových toků ipfixprobe, který plně funguje v nýnější formě. Implementuje všechny potřebné atributy, které jsem nadefinoval a to v nejefektivnější možné formě jak z hlediska času potřebného pro výpočet, tak pamě'tové náročnosti. Navíc lze pomocí nastavení parametrů určovat jaké skupiny atributy se mají exportovat. Díky tomu je možné přizpůsobit nasazení potřebám uživatele či rychlosti sítě.

Za velice kladný výsledek také hodnotím, že se studentovi podařilo naimplementovat v jazyce C++ nejefektivnější metodu pro výpočet Lomb-Scargle periodogramu přes NFFT (Nonequispaced Fast Fourier Transforms). Díky čemuž je možné frekvenční atributy získat nejefektivněji jak dosavadní vědecké poznání umožňuje. Navíc student v práci diskutuje použití knihovny, která by periodogram počítala na grafických kartách čímž by se ještě zrychlil výpočet. Nicméně monitorovací sondy nyní nepříliš často mají grafické karty a proto tato implementace nakonec nevznikla.

Prototyp pluginu proto hodnotím za velice povedený a student provedl pull request tohoto pluginu, takže prototyp či jeho části bude vývojáři exportéru ipfixprobe postupně zaintegrován do hlavního vlákna. Kód je vhodně strukturovaný, ale mohl by být více dokumentovaný komentáři než je nyní. Proto tuto část hodnotím 98/100.

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

Výsledek této práce je připraven na reálné nasazení dle mého názoru na malých až středně velkých sítích a určitě se nejedná o šuplíkovou závěrečnou práci, protože kód či

jeho podstatné části bude využit v rámci monitorování národní infrastruktury CESNET2 a detekci bezpečnostních hrozeb na této síti, protože atributy, které výsledný prototyp obsahuje jsou vhodné pro klasifikaci šifrovaného provozu. Navíc většina těchto atributů není v implementována existujících exportérech díky čemuž získá open source exportér ipfixprobe konkurenční výhodu oproti jeho komerčním či open source konkurentům.

Proto jsem se rozhodl udělit studentovi za využitelnost výsledků 100/100.

5. Aktivita studenta

▸ [1] **výborná aktivita**
[2] velmi dobrá aktivita
[3] průměrná aktivita
[4] slabší, ale ještě dostatečná aktivita
[5] nedostatečná aktivita

Student byl nadprůměrně aktivní. Účastnil se pravidelných schůzek a svými návrhy a nápady nad rámec původního zadání vylepšil výsledné dílo do nynější podoby.

6. Samostatnost studenta

▸ [1] **výborná samostatnost**
[2] velmi dobrá samostatnost
[3] průměrná samostatnost
[4] slabší, ale ještě dostatečná samostatnost
[5] nedostatečná samostatnost

Student byl velice samostatný. Dokázal si nastudovat architekturu ipfixprobe, matematické náležitosti analýzy časových řad, atributů jenž implementoval a všech potřebných knihoven. Samostatně vytvořit výsledný prototyp pluginu ipfixprobe exportéru.

Celkové hodnocení 98 /100 (A)

Celkově je tato práce velice povedená. Student splnil všechny body zadání. Kód i text práce jsou kvalitně zpracované a výsledky jsou použitelné již ve stávající podobě, nehledě na to v jaké podobě bude zakomponován do exportéru ipfixprobe pro monitorování ISP vysokorychlostní sítě CESNET2.

Práce navíc obsahuje další benefit, protože kód Lomb-Scargle periodogramu je vhodně využit v rámci této práce a výsledného pluginu pro ipfixprobe, ale může být využit v řadě dalších oblastech. Kód tohoto periodogramu v jazyce C++ je prostřednictvím této práce dán volně k dispozici, čímž může být užitečný pro vědeckou i nevědeckou komunitu.

Proto jsem se rozhodl dát práci finální hodnocení A (98/100).

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.

Aktivita studenta

V souvislosti s průběhem a výsledkem práce posuďte, zda byl student během řešení aktivní, zda dodržoval dohodnuté termíny, jestli své řešení průběžně konzultoval a zda byl na konzultace dostatečně připraven.

Samostatnost studenta

V souvislosti s průběhem a výsledkem práce posuďte schopnost studenta samostatné tvůrčí práce.

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.