FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

# Posudek oponenta závěrečné práce

| | |
|---|---|
| **Oponent práce:** | Ing. Jaroslav Pešek |
| **Student:** | Štěpán Jílek |
| **Název práce:** | Detekce WireGuard provozu pomocí Active Learning |
| **Obor / specializace:** | Bezpečnost a informační technologie |
| **Vytvořeno dne:** | 31. 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 bakalářské práce považuji spíše za nadprůměrně náročné z důvodu nutnosti nastudovat mnoho principů a technologií, které jsou nad rámcem bakalářského studia a zároveň bylo nutné přijít s netriviálním řešení specifického problému.

Žádná datová sada není výstupem práce, tedy ani není jasné, jestli se podařilo nějaké datové sady vytvořit. Z práce není jasné, zda a jak se vydařilo nasazení.

### 2. Písemná část práce 59 / 100 (E)

Rozsah bakalářské práce je přiměřený, neobsahuje zbytečné části, ani žádné nechybí. V rešerši se objevují věcné chyby (např. definice síťového toku v kapitole 1.1.3 není správně), definice jsou často vágní (např. definice strojového učení), věcné nekonzistence (např. kapitola 1.1.7 obsahuje tvrzení, že paketová analýza je nejpřesnější, ale v kapitole 2.2.4 se hovoří o tom, že "paketová analýza není tak rychlá a optimální", minimálně v oblasti přesnosti ale optimální z prvního tvrzení je), nekonzistence v používání terminologie (např. kapitola 1.2 obsahuje pojmy "přesnost, recall, f1", ale recall má také český ekvivalent, nebo naopak přesnost se konvenčně ponechává jako "precision"), nekonzistence psaní slov (např. práce obsahuje pojmy jako ip malými písmeny i IP velkými písmeny ve stejném kontextu).

Bakalářská práce obsahuje mimořádné množství typografických i gramatických chyb. Některé reference chybí (v kapitole 1.3 se autor odkazuje na "obrázek XY") a chybí i celé věty, což v některých částech práce ztěžuje pochopení textu. Práce obsahuje pozůstatky z doby, kdy autor práci teprve psal (kapitola 4.4.4 obsahuje instrukce pro autora o tom, že nemá zapomenout text doplnit, ale zapomněl). Jazyk je často příliš neformální; autor píše

v ich formě, což nepovažuji za významný nedostatek, ale občas se sklouzává k formulacím typu "já myslím".

Práce je členěna logicky ve schématu rešerše, analýza, návrh, implementace, vyhodnocení. Významnou připomínku mám ke kapitole Vyhodnocení, jež je nedostatečná a není zcela jasné, co se nakonec povedlo a nepovedlo. Navíc, výsledky jsou popsány pouze slovně, bez použití jakékoliv vizualizace.

Byly použity adekvátní zdroje (celkem 53). Citování je v pořádku; nemám připomínek.

### 3. Nepísemná část, přílohy 75 /100 (C)

K vytvoření autorova softwaru bylo použito objektivně orientované programování v jazyce Python, který spolu s ostatními technologiemi považuji za vhodně zvolený. Chybí mi nějaká rozšířenější dokumentace. Experimenty jsou neopakovatelné, neboť chybí jejich popis a není ani nijak popsána metodologie. Testy chybí. Absenci dokumentace považuji za méně závažný nedostatek, absenci testů a potom především experimentů za závažnější. Kvalita kódu je odpovídající bakalářské úrovni.

### 4. Hodnocení výsledků, jejich využitelnost 79 /100 (C)

Dílo lze využít v momentě, kdy vlastníme nebo spravujeme nějaký WireGuard server a chceme na něm anotovat provoz. Jak autor píše v úvodu práce, podobnému tématu se nevěnovalo mnoho autorů a takto samostatně se protokolu WireGuard v součinnosti s active learning principem nezabýval nikdo. Za velký přínos považuji, že autor byl schopen rozšířit již existující softwarové dílo a zařídit, aby jeho řešení s tímto existujícím dílem bylo schopné spolupracovat. Zároveň autorovo dílo lze rozšiřovat a jeho navržená architektura je navržená vcelku dobře a po refaktorování a nutných opravách bude moci být využito v praxi a případném výzkumu. Škoda je, že výsledky nejsou lépe zpracovány a kvůli absenci datové sady nejdou verifikovat, to důvěryhodnost řešení velmi sráží.

## Celkové hodnocení 69 /100 (D)

Z bakalářské práce mám ambivalentní dojem. Autor naprogramoval a vymyslel řešení netriviálního problému, ale řešení nelze moc dobře verifikovat. Textová zpráva je velmi slabá kvůli značnému množství problémů a chyb, ale jen pouhé další přečtení autorem by velkou část problémů vyřešila. Práce byla evidentně dělaná takzvaně horkou jehlou a bohužel se to projevilo. V takovéto podobě práci navrhuji k obhajobě se známkou D, ale pokud autor u obhajoby předloží výsledky, které by svědčily o úspěchu a předložil seznam experimentů a výsledků, které provedl, potom by moje hodnocení bylo na stupnici C a vyšší.

## Otázky k obhajobě

Proč plnohodnotná datová sada není součástí přílohy bakalářské práce? Pokud je problém se soukromím, je možné datové sady anonymizovat? Plánujete sadu zveřejnit nebo alespoň dát k dispozici výzkumné komunitě?

Vyskytly se během experimentování nějaké problémy?

Ohrožuje Vaše navržené řešení soukromí uživatelů? Jak případně lze toto mitigovat?

Lze v principu Vaše řešení využít i na jiné protokoly pro VPN? Pokud ne, proč?

# 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.