FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Josef Koumar
Student:	Abdulaziz Ismoilov
Název práce:	Analýza síťových toků pro detekci proxy serverů
Obor / specializace:	Informační bezpečnost 2021
Vytvořeno dne:	11. června 2025

Hodnotící kritéria

1. Splnění zadání

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

Zadání bylo splněno.

2. Písemná část práce 80 /100 (B)

Klíčová část teoretické kapitoly věnovaná tématu Residential Proxies je zpracována pouze ve dvou odstavcích, což považuji za výrazný nedostatek. Téma by si zasloužilo hlubší rozpracování. Sekce Related Work Analysis, zaměřená na analýzu existujících přístupů k detekci residential IP adres v síťovém provozu, se podrobně věnuje pouze dvěma článkům, přestože je dostupná výrazně širší literatura.

Vizuální stránka práce trpí technickým nedostatkem a to tím, že všechny obrázky obsahují velmi malé popisky, které jsou bez výrazného přiblížení obtížně čitelné. Pozitivně hodnotím použití vektorové grafiky, díky které je tento problém alespoň částečně kompenzován při elektronickém prohlížení. Nicméně při tisku budou tyto texty prakticky nečitelné, což je pro tištěnou verzi práce nevhodné.

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

Složka src/models, která je součástí příloh, je prázdná. Datová sada, se kterou autor v práci pracuje, není součástí příloh. Zdrojový kód postrádá strukturu, komentáře i dokumentační řetězce, což ztěžuje jeho pochopení i údržbu. V několika funkcích se opakují bloky kódu o desítkách řádků, které by bylo možné elegantně nahradit několika řádky pomocí vhodnějších konstrukcí jazyka Python.

Experimenty prováděné v prostředí Jupyter Notebooku jsou zpracovány nečitelně. Chybí v nich jakékoli popisné markdown bloky či komentáře, které by vysvětlovaly průběh a smysl jednotlivých kroků. Kódové buňky jsou často extrémně dlouhé (několik stovek řádků), což zásadně snižuje přehlednost a znemožňuje efektivní orientaci v experimentální části práce.

4. Hodnocení výsledků, jejich využitelnost 60 /100 (D)

Student se dopustil kritické metodologické chyby, když provedl undersampling majoritní třídy na celém datasetu ještě před jeho rozdělením na trénovací a testovací část. V důsledku toho nejsou prezentované výsledky na testovací sadě reprezentativní a nelze z nich spolehlivě odvodit výkonnost modelu v reálném nasazení.

Navíc vytvořená datová sada neodpovídá reprezentativnímu vzorku běžného (benigního) síťového provozu. Skládá se ze směsi různorodých datasetů, které byly původně určeny pro odlišné úlohy v oblasti síťové bezpečnosti. Tyto datasety navíc pocházejí z různých časových období, často s odstupem několika let, což je zvláště problematické vzhledem k dynamické povaze síťového provozu. V důsledku toho není možné považovat použitá data za adekvátní reprezentaci současné běžné komunikace v sítích.

Celkové hodnocení 70 /100 (C)

Práce se zaměřuje na aktuální téma v oblasti síťové bezpečnosti. Je vhodně uchopeno, ale metodologické chyby v praktické části a kvalita odevzdaného kódu výrazně snižují hodnotu díla. Proto navrhuji celkově klasifikační stupeň C.

Otázky k obhajobě

1) Datová sada vznikla sloučením různých typů síťového provozu. Který typ provozu byl pro klasifikátor nejproblematičtější, tedy vykazoval nejvyšší míru chybovosti?
2) Model využíval určité vstupní příznaky častěji než jiné. Jaký je podle Vás důvod, že právě tyto příznaky byly pro model nejvýznamnější?

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.