FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Hodnocení vedoucího závěrečné práce

Vedoucí práce:	Ing. Josef Koumar
Student:	Vojtěch Chvojka
Název práce:	Detekce těžby kryptoměn na základě periodického chování síťové komunikace
Obor / specializace:	Bezpečnost a informační technologie
Vytvořeno dne:	15. 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

Tato práce je ryze postavená na aplikování existující metody detekce periodicity na síťovém provozu na problém detekce kryptoměn a experimentálnímu vyhodnocení zda je pro detekci vhodná. V rámci bakalářské práce měl student vytvořit datovou sadu periodických atributů síťového provozu těžby kryptoměn. Tento cíl se podařilo splnit pomocí aplikování modelu detekce periodicty, který ode mě student obdržel, na veřejně dostupnou datovou sadu těžby kryptoměn ze sítě CESNET vytvořenou Richardem Plným et al.. Student následně na datovou sadu aplikoval strojové učení dle zadání a zhodnotil výsledky v práci. Dále student předložil softwarový prototyp, který s drobnými úpravami by mohl rozšířit stávající architekturu detekce kryptoměn na síti CESNET. Nicméně student v textu příliš nediskutuje nasaditelnost do reálných sítí, pouze zmíňuje, že výsledky práce jsou vhodné pro rozšíření DeCrypto systému.

Student splnil zadání, ale bohužel s minimálním úsilím a konečný výsledek práce nenese žádné znaky zlepšení nad rámec zadání a podobě mnou definovaných experimentů.

2. Písemná část práce 65 / 100 (D)

Text závěrečné práce obsahuje kapitoly Úvod, Teoretická část, Data, Detekce těžby kryptoměn, a Závěr. Celkově student obsáhnul 30 stran, nicméně z toho téměř 5 stran je nepopsaných. Rozsahově je tedy text práce spíše slabý a vzhledem k tématu práce by mohla být práce mnohem rozsáhlejší.

Kapitola Teoretická část obsahuje směs popisu co jsou kryptoměny, relevantních prací pro detekci kryptoměn a teoretického popisu stojového učení. Celá teoretická část obsahuje

občasné používání hovorových výrazů jako třeba "spousta těžebního softwaru". Část zabývající se relevantními existujícími pracemi pro detekci kryptoměn sice obsahuje 6 citací konferenčních článků, ale je velice strohá a jelikož se práce zabývá detekcí, tak by si zasloužila být obsáhlejší. Rozdíly motod detekce z relevantních prací není porovnávaná s metodou aplikovanou v rámci této práce.

Kapitola Data obsahuje popis vstupní datové sady, ze které vznikla datová sada periodických chování rovněž popsaná v této části. Dále text obsahuje hlubší popis IPFIX protokolu, který by mohl být lépe umístěn v kapitole Teoretická část. Část popisující vytváření časových řad je mnohdy psaná počeštěnou angličtinou, příkladem je hojné používání slova "datapointů". V práci je napsáno, že nejlépe pro metodu vychází rozdělení časových řad po 6 hodinách dle nastavení časového okenka sbírání časových řad, bohužel práce nebosahuje hlubší popis experimentů jak k této hodnotě student došel. V textu je také detailně popsána metoda detekce periodického chování, u které se uvádí, že je nutné nastavit přesnost detekce pomocí 3 parametrů, které student ponechal na defaultních hodnotách nastavené při předání kódu vedoucím. Výsledná datová sada obsahuje téměř 11 tisíc záznamů, což hodnotím kladně vhledem k 6 hodinovému časovému okénku. V této části je také řada tabulek, které obsahují stylistické nedokonalosti.

Kapitola Detekce těžby kryptoměn obsahuje aplikování strojového učení včetně použité metodologie. Student správně aplikoval strojové učení a udělal vše pro vyhnutí se přeučení (angl. overfitting) včetně použití K-fold validace. Kladně hodnotím, že z textu je zřejmé jak replikovat výsledky experimentu, čímž student do jisté míry překonal nýnější výzkum v této oblasti, protože většina článků je publikovaná tak, že experimenty nelze zreplikovat.

Citační styl studenta je v souladu s předmětem DPR. Student citoval 48 zdrojů z toho většina jsou konferenční či časopiskecké publikace. Dále obsahují citace RFC dokumentů a online zroje v oblasti kryptoměn. Nicméně online zdroje jsou v oblasti kryptoměn podstatné, protože se jedná o velice rychle se vyvíjející oblast, u které publikované práce na konferencích a časopisech zaostávají za realitou kvůli dlouhým recenzním řízením.

Celkově text hodnotím 65/100, jelikož text práce nedosahuje doporučeného rozsahu a obsahuje nedokonalosti, kterých si čtenář všimne na první pohled.

3. Nepísemná část, přílohy 85 /100 (B)

Výsledkem je prototyp, který je obsažen v několika Python programech, Bash skriptech a Jupyter noteboocích, které jsou dobře popsány a dokumentovány. Mnohé výstupy jsou v popisku označeny mým autorstvím, nicméně v nich student provedl drobné změny, zejména pak co se týče podpory časového okenka popsaného v textu. Bohužel oněch 6 hodin je vloženo konstantou v kódu a není vyvedeno argumentem pro nastavování a experimenty.

Celkévě vytvořený prototyp hodnotím 85/100, jelikož není zřejmé jak prototyp použít pro detekci těžby kryptoměn v reálném čase na síti.

4. Hodnocení výsledků, jejich využitelnost 90 /100 (A)

Jak jsem se již zmínil výsledky práce silně závisí na několika nastavitelných parametrech a to i) tří parametrů detekce periodicity a ii) nastavení časového okénka sbírání časových

řad. Student v práci definuje nastavení hodnot, které označuje za nejlepší pro výsledky klasifikátoru, ale neadresuje jak se výsledky změní pokud by se nastavily jiné hodnoty. Bylo by na místě otestovat vliv nastavování parametrů na výsledky detekce či použití tkz. posuvného okénka pro vytváření časových řad.

Výsledky detekce těžby kryptoměn pomocí periodických chování jsou velice kladné a vhodné pro publikování na mezinárodní konferenci, jelikož podobnou detekci těžby kryptoměn nikdo zatím neadresoval v dosavadních publikací. Nicméně samotné publikování bude možné až po doplnění výše zmíněných experimentů, které práce bohužel postrádá.

Celkově hodnotím díky vysoké přesnosti klasifikátoru a publikačnímu potencionálu práce 90/100, přičemž penalizuji kvůli chybějícím experimentům.

5. Aktivita studenta

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

Bohužel musím konstantovat, že student se mnou komunikoval málo a text psal na poslední chvíli, což mělo vliv na výslednou podobu práce. Proto hodnotím jako slabou aktivitu, ale ještě dostatečnou pro splnění práce.

6. Samostatnost studenta

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

Student samostatně provedl experimenty a vyhodnotil výsledky. Prokázal samostatnost při práci s cizým kódem (modelem detekce periodicity a modulem pro vytváření časové řady) a jejich drobnými úpravami. Nicméně tvůrčí práce studenta nad rámec pokynů vedoucího je minimální. Proto hodnotím samostatnost stupněm 3. Tedy průmerná samostatnost.

Celkové hodnocení 78 /100 (C)

Celkově doporučuji studentovi za tuto práci udělit finální hodnocení lepší C (78/100), jelikož jak výsledky tak text práce silně ovlivnila aktivita studenta. Text práce je slabý a provedené experimenty jsou minimalistické, z tohoto důvodu i přes vysokou přesnost výsledného klasifikátoru jsem se přiklonil ke klasifikačnímu stupni C.

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.