FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Student:	Bc. Michal Buchovecký
Oponent práce:	prof. Ing. Róbert Lórencz, CSc.
Název práce:	Semi-supervised learning pro detekci malware
Obor:	Počítačová bezpečnost

Datum vytvoření: 5. 6. 2019

Hodnotící kritérium:	*Způsob hodnocení – následující škálou 1 až 4:*
1. Splnění zadání	***1=zadání splněno,*** *2=zadání splněno s menšími výhradami,* *3=zadání splněno s většími výhradami,* *4=zadání nesplněno*

Popis kritéria:
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.

Komentář:
Zadání bylo splněno bez výhrad.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
2. Písemná část práce	*81 (B)*

Popis kritéria:
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. 26/2017, 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.

Komentář:
Práce má z větší částí popisný charakter. V této popisné částí jsou uvedeny zejména základní pojmy a přístupy z oblasti detekce malware. Popis je orientován hlavně na oblast využití strojového učení. V práci můžeme nalézt řadu překlepů a nepřesností. Například v textu je odkazováno na "grafy", ale ve skutečnosti je míněno "obr." Dále jsou obrázky ne úplně popsané nebo text v nich je nečitelný (Obr. 26, 25, 24). Obrázky mají zajímavé značení. V matematických výrazech se rovněž vyskytují chyby (co to je TPR, str 23, co sigma, vektorová funkce je reálná hodnota str. 31, atd.).

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
3. Nepísemná část, přílohy	*86 (B)*

Popis kritéria:
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

Komentář:
Implementace algoritmů byla úspěšně vykonaná v jazyce Python při využití knihovny Scikit-learn.

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Hodnocení výsledků, jejich využitelnost	*81 (B)*

Popis kritéria:
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.

Komentář:
Využitelnost výsledků je v popise Semi-superviset learning algoritmů aplikovaných na detekci malware. Samotné výsledky z experimentů nedosahuji očekáváné kvality.

Hodnotící kritérium:	*Způsob hodnocení – nehodnotí se*
5. Otázky k obhajobě	
Popis kritéria: Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).	
Otázky: Seznam příznaků není přesně specifikován. Může student uvést, jaké příznaky použil v jednotlivých metodách? Proč aplikováním modifikace na Co-Training (přidání 3. klasifikátoru) se přesnost klasifikace nezvýšila? Může autor vysvětlit neúspěšnost LLGC algoritmu v realizaci ?	

Hodnotící kritérium:	*Způsob hodnocení – bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Celkové hodnocení	*85 (B)*
Popis kritéria: 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.	
Text hodnocení: Práce je psaná čtivě, obsahuje nepřesnosti a chybí v ní některé částí související s vykládanou problematikou. Rovněž výsledky práce jsou průměrné. Práci hodnotím stupněm B (85 bodů), protože je v ní celkem souvisle a do určité míry použitelně obsažena problematika využití algoritmů strojového učení pro detekci malware.	

Podpis oponenta práce: