FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Student:	Bc. Marek Jílek
Oponent práce:	Ing. Lukáš Machlica, Ph.D.
Název práce:	Využití strojového učení pro extrakci kontextu síťových incidentů
Obor:	Počítačová bezpečnost

Datum vytvoření: 30. 5. 2018

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 5:*
1. Náročnost a další komentář k zadání	*1=mimořádně náročné zadání,* ***2=náročnější zadání,*** *3=průměrně náročné zadání,* *4=lehčí, ale ještě dostatečně náročné zadání,* *5=nedostatečně náročné zadání*

Popis kritéria:
Podrobněji charakterizujte diplomovou (bakalářskou) práci a její případné návaznosti na předchozí nebo běžící projekty. Dále posuďte, čím je zadání této ZP náročné. (U obtížnější ZP lze dále tolerovat některé nedostatky, které by u ZP standardní obtížnosti tolerovány nebyly; a naopak u jednoduché ZP mohou být zjištěné nedostatky hodnoceny přísněji.)

Komentář:
Student se v práci věnuje problému klasifikace síťových útoků a extrakce kontextu, ve kterém k útoku došlo. Problém automatického vysvětlení klasifikací je v úloze detekce bezpečnostních incidentů často opomíjený, ale velice důležitý, protože incidenty jsou ve většině případů dále analyzovány lidmi. Není tedy důležité jenom jestli k incidentu došlo, ale stejně důležité je i jak a proč.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
2. 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 splňuje zadání. V komentáři uveďte body zadání, které nebyly zcela splněny, případně rozšíření ZP oproti původnímu zadání. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

Komentář:
Práce je rozdělena do tří částí. V první části jsou podrobně analyzována data, následně jsou představeny metody pro extrakci kontextu a optimalizace těchto metod. Poslední část práce obsahuje srovnání klasifikačních technik jak z pohledu jejich přesnosti a časové náročnosti trénovaní a testování, tak z pohledu možnosti využití jednotlivých metod pro vysvětlení kontextu klasifikace.

Velice kladně hodnotím také experimenty, které se věnují analýze systému v čase a dávají informaci o časovém intervalu, po kterém je nutno systém přetrénovat, aby nedošlo k výraznému sníženi jeho účinnosti.

Hodnotící kritérium:	*Způsob hodnocení - následující škálou 1 až 4:*
3. Rozsah písemné zprávy	***1=splňuje požadavky,*** *2=splňuje požadavky s menšími výhradami,* *3=splňuje požadavky s většími výhradami,* *4=nesplňuje požadavky*

Popis kritéria:
Zhodnoťte přiměřenost rozsahu předložené ZP vzhledem k obsahu, tj. zda všechny části ZP jsou informačně bohaté a ZP neobsahuje zbytečné části.

Komentář:
Práce splňuje požadavky kladené na rozsah a formu diplomové práce.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Věcná a logická úroveň práce	*95 (A)*

Popis kritéria:
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.

Komentář:
Práce je dobře strukturovaná. Jednotlivé kapitoly na sebe logicky navazují. Algoritmy a experimenty jsou srozumitelně a jasně popsány. Předpoklady a závěry práce odvozené z experimentů jsou logické a legitimní.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
5. Formální úroveň práce	*90 (A)*

Popis kritéria:
Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 26/2017, článek 3.

Komentář:
Práce obsahuje několik překlepů, které ale nijak zvlášť nesnižují čitelnost a přehlednost práce. Jedinou výtkou je používání anglické terminologie i když je termín dobře zaveden v českém jazyce, např.: features = příznaky, olabelovat = anotovat, confidence = důvěra, request = požadavek

V některých případech také dochází k používaní jak české, tak anglické varianty jednoho pojmu, např.: cross-validace a křížová validace

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Práce se zdroji	*100 (A)*

Popis kritéria:
Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení ZP. Charakterizujte výběr studijních pramenů. Posuďte, zda student využil všechny relevantní zdroje nebo zda se pokoušel řešit již vyřešené problémy. Ověřte, zda jsou všechny převzaté prvky řádně odlišeny od vlastních výsledků a úvah, zda nedošlo k porušení citační etiky a zda jsou bibliografické citace úplné a v souladu s citačními zvyklostmi a normami.

Komentář:
Práce obsahuje 36 relevantních zdrojů, které jsou v práci správně citovány.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
7. Hodnocení výsledků, publikační výstupy a ocenění	*100 (A)*

Popis kritéria:
Vyjádřete se k úrovni dosažených hlavních výsledků ZP, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, apod. Případně také zhodnoťte, zda software nebo zdrojové texty, které nevytvořil sám student, byly v ZP použity v souladu s licenčními podmínkami a autorským právem. Popište případnou publikační činnost a získaná ocenění související s řešením této ZP.

Komentář:
Student prokázal schopnost pracovat s literaturou, která je v práci vhodně citována. Vlastní přínosy jsou v práci zřejmé. Práce obsahuje velký počet relevantních experimentů, které jsou dobře strukturovány a popsány.

Hodnotící kritérium:	*Způsob hodnocení - nehodnotí se*
8. Komentář o využitelnosti výsledků	

Popis kritéria:
Uveďte, zda hlavní výsledky ZP rozšiřují již publikované známé výsledky a/nebo přinášející zcela nové poznatky. Uveďte možnosti využití výsledků ZP v praxi.

Komentář:
Práce je součástí výzkumného záměru v rámci komerčního produktu CTA, Cisco Systems a týká se relevantní částí extrakce kontextu při detekci síťových útoků. Výstupy a experimenty v diplomové práci jsou zaměřeny na reálné problémy a reálna data.

Hodnotící kritérium:	*Způsob hodnocení - nehodnotí se*
9. Otázky k obhajobě	

Popis kritéria:
Uveďte případné dotazy, které by měl student zodpovědět při obhajobě ZP před komisí (body oddělte odrážkami).

Otázky:
1. V případe Sekce 4.1. uvádíte, že hodnoty pro \phi_{consy} byly v rozsahu 1-3, ale jak kosínová tak Jaccardova vzdálenost jsou z intervalu <0, 1>, můžete upřesnit jakým spúsobem jste vzdálenosti počítal?

2. Jakým spůsobem lze SVM využít jako klasifikátor pro klasifikaci do více tříd?

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
10. Celkové hodnocení	*95 (A)*

Popis kritéria:
Shrňte stránky ZP studenta, které nejvíce ovlivnily Vaše celkové hodnocení. Celkové hodnocení **nemusí** být aritmetickým průměrem či jinou hodnotou vypočtenou z hodnocení v předchozích jednotlivých kritériích 1 až 9.

Text hodnocení:
Student v rámci práce prokázal schopnost samostatně nastudovat problematiku, problém rozvinout, navrhnout řešení a experimenty a ty následně srozumitelně popsat.

Podpis oponenta práce: