Posudek oponenta závěrečné práce

České vysoké učení technické v Praze **Fakulta informačních technologií**

Student: Bc. Adam Plánský
Oponent práce: Ing. Josef Kokeš
Název práce: Automatická analýza nahlášených bezpečnostních incidentů
Obor: Počítačová bezpečnost

Datum vytvoření: 24. 5. 2017

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:
Účelem práce bylo vytvořit filtr, který vybere ze zachycených bezpečnostních incidentů takové, které je vhodné dále podrobněji analyzovat. To je ve vysokorychlostních sítích velmi potřebné, protože nefiltrovaných incidentů v nich je nezvladatelně mnoho. Náročnost zadání je průměrná.

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:
Zadání bylo splněno. Filtrační algoritmus byl vytvořen a úspěšně nasazen.

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:
Rozsah práce splňuje požadavky.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
4. Věcná a logická úroveň práce	*85 (B)*

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:
Po věcné stránce je práce v pořádku, nacházím pouze drobnější chyby: V algoritmu z kapitoly 4.2.5 se zdá, že jsou 1., 3. a 4. bod nadbytečné. V kapitole 5.5 mi chybí srovnání dosaženého výkonu 2300 událostí za sekundu s očekávaným počtem událostí v reálném systému; soudě podle vytížení systému v kapitole 5.6 jde zřejmě o mnohem větší výkon, než je vyžadován. Nenašel jsem detailní diskusi vlivu změn nastavení na fungování algoritmu, kapitola 5.7 se zabývá pouze metrikou redukce bezpečnostních událostí, a to bez zohlednění toho, jak důležité události ve výstupu zůstaly - a co je ještě důležitější, které důležité události byly vyfiltrovány.

Výhrady mám k logické struktuře textu, zejména k sekci 3.2. Na straně 21 je uveden výsledný algoritmus, který je teprve následně vysvětlován, jak funguje. To je nešťastné už proto, že používá symboly a značení, které v tomto místě dosud nebyly zavedeny. Opačný přístup (popsat komponenty a následně jejich spojení do celku) by pravděpodobně byl srozumitelnější. Velmi bych se přimlouval k zápisu algoritmu v podobě více IFů (i vnořených) a méně komplikovaných (a zbytečných) matematických výpočtů, které snižují srozumitelnost. Každopádně by práci pomohlo tuto část napsat podrobněji, ani po prostudování práce mi není jasné, co přesně zachycují obrázky 3.3 až 3.8.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
5. Formální úroveň práce	*70 (C)*

Popis kritéria:
Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku ZP, viz Směrnice děkana č. 14/2015, článek 3.

Komentář:
Formální zápisy jsou v pořádku, kombinování algoritmu a výpočtů v Algoritmu 1 je však nešťastné, protože velmi snižuje přehlednost.

Typografická stránka je dobrá, zůstávají ovšem jednopísmenkové předložky a spojky na koncích řádků a na straně 20, 23, 40 i přetečení textu mimo oblast stránky. Seznam použitých zkratek není seřazen podle abecedy a zápis jejich vysvětlení není konzistentní. Klíčová slova anglického abstraktu jsou až na další stránce.

V práci bohužel zůstalo značné množství jazykových chyb. Jedná se zejména o pravopisné chyby, shodu podmětu s přísudkem, čárky. Také překlepů je mnoho, a to i na velmi viditelných místech (např. v popisech obrázků).

Nerozumím míchání češtiny a angličtiny v názvech filtrů (sekce 2.1.3) - např. "Black List Filter" vs. "Vertical Port Sken (!) Detektor (!)".

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
6. Práce se zdroji	*80 (B)*

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:
Zdrojů je uvedeno přiměřené množství, nejsem ale stoprocentně přesvědčen o jejich vhodnosti. Některé zdroje se zdají být duplicitní nebo vzájemně závislé (3 a 5, 15 a 18), případně irelevantní (20, 21). Odkazem jsou často označeny konkrétní termíny nebo technické specifikace spíše než převzaté myšlenky. I v případě převzatých definic je zdrojován termín, ne jeho definice. Naopak u tvrzení občas citace, která by je podpořila, chybí (např. druhá věta v sekci 1.2.1, první věta na straně 5, předposlední věta na str. 15).

Seznam zdrojů není seřazen podle jména autora. Zápis jednotlivých zdrojů není konzistentní (autoři u 11, 14 a 19). Autoři u 14 a 22 by jistě mohli být uvedeni s diakritikou.

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
7. Hodnocení výsledků, publikační výstupy a ocenění	*80 (B)*

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:
Výsledky požadované zadáním byly dosaženy a úspěšně nasazeny. Navržená hodnotící funkce, hlavní teoretický výsledek, zjevně plní svůj účel eliminovat příliš četná hlášení o incidentech, bohužel ale nebyla zkoumána její účinnost ve smyslu false-positive a false-negative rate. Bez toho je obtížné rozlišit, zda jde o geniálně jednoduchou funkci nebo o funkci příliš triviální.

Programové řešení se zdá být funkční a licenčně v pořádku. Konkrétní realizace ale zanedbává řadu podstatných rysů, nenalezl jsem například žádný kód řešící neočekávané vstupy, tzn. aplikace přijme cokoliv a doufá, že ji to nezboří. Umístění souboru load_static_prices.py s programovým kódem mezi konfigurační soubory je z hlediska bezpečnosti hrubá chyba. Synchronizace přístupů ve vícevláknovém zpracování závisí na tom, že se používají pouze atomické operace pop a append, není žádný explicitní synchronizační mechanismus. Program se bohužel velmi obtížně čte, hlavní soubor aplikace má 12 tříd a 900 řádků Pythonového kódu téměř bez komentářů. Navíc účel jednotlivých souborů není až na výjimky zdokumentován.

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:
Vytvořený filtr může výrazně pomoci s analýzou útoků v rozsáhlých sítích, v tomto jde o velmi užitečný nástroj. Byl úspěšně nasazen v síti CESNET, kde skutečně vedl k redukci počtu hlášení; praktické využití tedy už nastalo. Obávám se však, že chybějící otestování false-positive a false-negative rate může ještě přinést značné komplikace.

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) Bylo nějak testováno false-positive a false-negative rate, tzn. to, zda program a) nezahazuje události, které by měl ponechat, nebo b) neponechává události, které by bylo lepší zahodit? S jakými výsledky?

2) Uvádíte, že základní konfigurace ponechala v proudu incidentů jen 7 incidentů za minutu. Není to z hlediska zpracování analytiky stále příliš mnoho?

Hodnotící kritérium:	*Způsob hodnocení - bodové hodnocení 0 až 100 bodů (známka A až F):*
10. Celkové hodnocení	*80 (B)*

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:
Práce řeší velmi důležitou problematiku výběru potenciálně zajímavých bezpečnostních incidentů ve vysokorychlostních sítích. Byla úspěšně nasazena a splnila očekávání do ní vkládaná. Bohužel jí škodí slabší jazyková stránka, formální nedostatky a hlavně ne zcela srozumitelný popis v kritické části práce. Vytvořené programy jsou funkční, kvalita kódu ale nesplňuje docela to, co bych od absolventa FIT - a zvlášť oboru Počítačová bezpečnost - očekával. Z těchto důvodů navrhuji hodnotit práci známkou B - velmi dobře.

Podpis oponenta práce: