ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

POSUDEK OPONENTA ZÁVĚREČNÉ PRÁCE

I. IDENTIFIKAČNÍ ÚDAJE

Název práce:	**Detection of Malicious Network Behaviour in Encrypted Network Traffic**
Jméno autora:	**Bc. Pavel Potoček**
Typ práce:	Master's Thesis
Fakulta/ústav:	Faculty of Electrical Engineering
Katedra/ústav:	Dept. Of Computer Science
Oponent práce:	RNDr. Petr Somol, Ph.D.
Pracoviště oponenta práce:	Cisco Systems, ÚTIA AV ČR

II. HODNOCENÍ JEDNOTLIVÝCH KRITÉRIÍ

Zadání

Hodnocení náročnosti zadání závěrečné práce.

Zadání netriviální, ke zvládnutí vyžadovalo dlouhodobou aktivní práci na daném problému. Náročnost vyplývá zejména ze zaměření na analýzu průmyslových dat velmi obtížných vlastností. Autor čelil abnormálně velikým datovým souborům, neúplným, nevybalancovaným, s obtížně odhadnutelnými distribucemi. V takovémto kontextu typicky selhávají zaběhlé postupy a je nutno hledat postupy principiálně nové. Konečný cíl zadání – vylepšení přesnosti detekce na HTTPS telemetrii – je pak sám o sobě mimořádně ambiciózní.

Splnění zadání

Posuďte, zda předložená závěrečná práce splňuje zadání. V komentáři případně uveďte body zadání, které nebyly zcela splněny, nebo zda je práce oproti zadání rozšířena. Nebylo-li zadání zcela splněno, pokuste se posoudit závažnost, dopady a případně i příčiny jednotlivých nedostatků.

Autor navrhl a ověřil několik novátorských dílčích řešení a propojil je do prototypu funkčního systému řešícího zadání. Experimentální ověření a související analýza prokazují že zadání bylo splněno.

Zvolený postup řešení

Posuďte, zda student zvolil správný postup nebo metody řešení.

V rámci očekávaného rozsahu práce jakožto diplomové práce jsou zvolené postupy správné a pokrývají zadání v rozumné šíři.

Odborná úroveň

Posuďte úroveň odbornosti závěrečné práce, využití znalostí získaných studiem a z odborné literatury, využití podkladů a dat získaných z praxe.

Přinejmenším odpovídá očekáváním kladeným na diplomovou práci. Z pečlivé diskuse napříč textem je zřejmé, že autor domýšlel souvislosti zkoumaných problémů a možných řešení až do jinak snadno přehlédnutelných detailů. Diskuse je dobře srozumitelná, věcná a správná. Kvalita práce jakožto vědecké studie převyšuje běžnou kvalitu diplomových prací.

Formální a jazyková úroveň, rozsah práce

Posuďte správnost používání formálních zápisů obsažených v práci. Posuďte typografickou a jazykovou stránku.

Práce je napsána mimořádně kvalitní angličtinou, celkově vykazuje dobře promyšlenou strukturu jak řešení problému, tak presentace. Dle mých zkušeností je takto kvalitně presentována jen menšina prací. (Jediná drobnost které bych doporučoval se příště vyhnout je na str. 30 v prvním odstavci sekce 2.2. odkaz na specifickou rovnici zavedenou až o dvě stránky dále.)

Výběr zdrojů, korektnost citací

Vyjádřete se k aktivitě studenta při získávání a využívání studijních materiálů k řešení závěrečné práce. Charakterizujte výběr pramenů. Posuďte, zda student využil všechny relevantní zdroje. 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.

Citace týkající se klíčových přínosů práce zahrnují všechny podstatné zdroje, o nichž vím, je tedy zřejmé že autor provedl kvalitní rešerši. Jako případný námět pro rozšíření tohoto výzkumu o další alternativní směr by bylo možno přidat k diskusi na str. 28 také boostovací princip popsaný v Pevný, Tomáš. "Loda: Lightweight on-line detector of anomalies." Machine Learning 102.2 (2016): 275-304. K textu práce ještě drobná poznámka: na str.20 bych doporučil raději uvést odkaz na Evangelista a Acc@Top přímo do textu a ne jen nepřímo do popisku obrázku.

Další komentáře a hodnocení

Vyjádřete se k úrovni dosažených hlavních výsledků závěrečné práce, např. k úrovni teoretických výsledků, nebo k úrovni a funkčnosti technického nebo programového vytvořeného řešení, publikačním výstupům, experimentální zručnosti apod.

Jakožto kvalitní vědecká práce tato práce otevírá řadu možných směrů k dalšímu rozvoji i umožňuje formulovat nové otázky.

III. CELKOVÉ HODNOCENÍ, OTÁZKY K OBHAJOBĚ, NÁVRH KLASIFIKACE

Shrňte aspekty závěrečné práce, které nejvíce ovlivnily Vaše celkové hodnocení. Uveďte případné otázky, které by měl student zodpovědět při obhajobě závěrečné práce před komisí.

Předloženou závěrečnou práci hodnotím jako **vynikající**. Jednoznačně doporučuji aby práce byla uznána za diplomovou práci.

Do diskuse se nabízí otázky:

1) v sekci 2.1.2 by zřejmě bylo možno uvažovat i medián. Neuvádíte tuto možnost kvůli jeho výpočetní složitosti nebo by jeho efekt nebyl zajímavý?,

2) na str.38 navrhujete budovat model z tří dnů dat na základě předchozí diskuse – tento počet je univerzálně doporučitelný nebo závisí na daném kontextu či typu dat?,

3) str. 44 na poslední řádce uvádíte předpoklad že local i global poskytnou vyhlazovací efekt, jejich porovnání tedy ukáže rozdílný efekt sdílení informace. To je nepochybně pravda, zajímala by mě ale diskuse jestli přesto se nemůže vyhlazovací efekt u local a global lišit, a tím pádem souviset s rozdílnými výsledky local a global (navíc k rozdílu danému sdílením dat),

4) str.45 odstavec 4 uvádí hypotézu že očekávaní zlepšení na HTTPS by mělo být výraznější než u http a to v důsledku MiTM. To je intuitivně správný předpoklad, který ale experimenty zřejmě vyvrátily. Můžete tento detail okomentovat?,

5) tabulka 4.3 ukazuje „pairwise“ jakožto nejlépe fungující algoritmus. Tyto výsledky by zasluhovaly podrobnější diskusi, proč právě tento algoritmus vychází lépe.

Datum: 12. 6. 2017 Podpis: Petr Somol