Posudek oponenta diplomové práce

Student: Šuhaj Peter, Bc.
Téma: Rozšíření NetFlow záznamů pro zlepšení možností klasifikace šifrovaného provozu (id 23159)
Oponent: Jeřábek Kamil, Ing., UIFS FIT VUT

1. Náročnost zadání **obtížnější zadání**

Jedná se o složitější zadání, kde student musel nastudovat techniky klasifikace šifrovaného provozu. Na základě statistické analýzy identifikovat vhodné parametry komunikace, zároveň vhodné pro rozšíření NetFlow záznamů. Implementovat rozšíření sondy o zpracování těchto záznamů. Primárně pak navrhnout a implementovat vlastní algoritmus pro klasifikaci šifrovaného provozu a o ověřit správnost algoritmu na základě dostatečného množství dat.

2. Splnění požadavků zadání **zadání splněno s vážnými výhradami**

Zadání bylo splněno, student provedl analýzu dat, implementaci rozšíření záznamů a navrhl a implementoval algoritmus, který následně ověřil. Avšak, důsledky analýzy byly vyvozeny z minimálního množství dat (pouze jednotky toků). Na množině dat, která byla použita pro trénink a testování algoritmu, není znatelné, zdali je algoritmus schopný klasifikovat správně provoz či nikoli.

3. Rozsah technické zprávy **je v obvyklém rozmezí**

Práce je v obvyklém rozmezí.

4. Prezentační úroveň předložené práce **70 b. (C)**

Práce je logicky dobře strukturovaná, jednotlivé kapitoly na sebe navazují. První část práce se dostatečně zabývá teorií problematiky práce. Na tuto část pak navazuje vlastní analýza protokolů a dat, zde jsou na základě analýzy navrženy vhodné vlastnosti pro výslednou klasifikaci. Následuje návrh a implementace jak algoritmu tak rozšíření NetFlow záznamů. Kapitola 5.1 popisující návrh klasifikačního algoritmu je místy zmatečná. V závěru práce obsahuje testování a vyhodnocení přesnosti algoritmu. Zde bych uvítal vhodnější interpretaci výsledků a výpočtu přesnosti algoritmu.

5. Formální úprava technické zprávy **80 b. (B)**

Typografická stránka práce je s drobnými chybami v pořádku.

6. Práce s literaturou **80 b. (B)**

Výběr studijních pramenů je pro práci relevantní až na některé online zdroje, kterých je však malé množství. Zmíněné online citace nejsou vždy úplné.

7. Realizační výstup **60 b. (D)**

Programové řešení je rozumně navržené a rozšiřitelné. Avšak mám výhrady k nastavení parametrů klasifikačního algoritmu, které jsou ve zdrojovém kódu zadány pevně. Zvlášť když student zmiňuje, že některé vlastnosti použité pro klasifikaci mohou být specifické pro danou síť a bude je potřeba upravit. Testování a vyhodnocení výsledků úspěšnosti algoritmu je věnována rozumně obsáhlá část práce. Nicméně z malé testovací množiny a vyhodnocených výsledků nelze jednoznačně určit správnost algoritmu pro klasifikaci.

8. Využitelnost výsledků

Práce rozšiřuje poznatky o možnosti klasifikace šifrovaného provozu. Zároveň svou implementací rozšiřuje řešení, konkrétně modul pro exportér firmy Flowmon Networks.

9. Otázky k obhajobě

- V kapitole 7.2 uvádíte, že úspěšnost algoritmu byla vypočítána jako podíl součtu jednoznačně a nejednoznačně určených toků k počtu klasifikovaných toků. Jak si tento vzorec zdůvodňujete?
- Jak jste vygeneroval a následně rozlišil DNS over HTTPS toky od normálního HTTPS provozu?

10. Souhrnné hodnocení **70 b. dobře (C)**

Práci doporučuji ohodnotit stupněm C. Zadání bylo obtížnější jelikož obnášelo nastudovat a analyzovat danou datovou problematiku pro vytvoření vhodného klasifikačního algoritmu šifrovaného provozu a rozšíření záznamů NetFlow. Zadání bylo studentem splněno, nicméně práce vychází z malého množství dat (jedná se o jednotky, maximálně pár desítek vzorků) což poměrně výrazně snižuje kvalitu práce. S přihlédnutím k tomuto faktu, nejsme schopni jednoznačně určit, zdali je navržený a implementovaný algoritmus s danými zvolenými parametry vhodný pro řešení tohoto klasifikačního problému.

Prohlášení: Uděluji VUT v Brně souhlas ke zveřejnění tohoto posudku v listinné i elektronické formě.

V Brně dne: 29. června 2020

Jeřábek Kamil, Ing.
oponent