FAKULTA INFORMAČNÍCH TECHNOLOGIÍ ČVUT V PRAZE

Posudek oponenta závěrečné práce

Oponent práce:	Ing. Tomáš Čejka, Ph.D.
Student:	Zdena Tropková
Název práce:	Klasifikace akcí přenášených skrz šifrované TLS spojení
Obor / specializace:	Bezpečnost a informační technologie
Vytvořeno dne:	7. června 2021

Hodnotící kritéria

1. Splnění zadání

▸ [1] **zadání splněno**
[2] zadání splněno s menšími výhradami
[3] zadání splněno s většími výhradami
[4] zadání nesplněno

Tato bakalářská práce se zabývá výzkumným tématem analýzy a klasifikace šifrovaného provozu, konkr. provozu protokolu TLS. Odevzdaná práce zkoumá aplikaci metod strojového učení na tuto oblast s využitím tzv. IP flow dat obohacených o posloupnosti paketových charakteristik. Touto problematikou se zabývají současné publikované vědecké práce, což svědčí o aktuálnosti tohoto tématu. Podle provedené rešerše existujících řešení se zdá, že odevzdané řešení představuje významný prvek inovace. Prezentované výsledky vypadají velice slibně, výstupem práce je vytvořený funkční prototyp, čili zadání považuji za splněné.

2. Písemná část práce 79 /100 (C)

Písemná část práce je na dobré úrovni, avšak některé formulace jsou poměrně krkolomné a tudíž hůře srozumitelné. Vedle toho mám malé výhrady i proti formátu citovaných zdrojů v seznamu referencí, některé zdroje jsou uvedeny nestandardním způsobem, např. [5] a [7]. V textu práce jsem postrádal podrobnosti týkající se fungování prototypu klasifikačního modulu, např. vysvětlení, co konkrétně dělají zmiňované funkce.

3. Nepísemná část, přílohy 90 /100 (A)

Podle textu práce jsou výstupem bakalářské práce datové sady síťového provozu a sada experimentů zkoumající úspěšnost klasifikátorů založených na různých modelech strojového učení. Obojí je přínosné pro mezinárodní vědeckou komunitu. Zdrojové kódy experimentů jsou přílohou práce a umožňují tak tyto experimenty zopakovat a případně

použít na dalších nových datech. Příloha práce obecně je vypracovaná pečlivě a nenarazil jsem na nedostatky.

4. Hodnocení výsledků, jejich využitelnost 95 /100 (A)

Výsledky této bakalářské práce vypadají zajímavě a vytvořený softwarový prototyp dosahuje vysoké úspěšnosti klasifikace. Vzhledem k tomu, že analýza šifrovaného provozu je nyní aktuální výzkumný problém, přínos této závěrečné práce je nezpochybnitelný. Prezentované charakteristiky síťového provozu i navržený klasifikátor mají publikační potenciál.

Celkové hodnocení 95 /100 (A)

V rámci práce se podařilo vytvořit prototyp klasifikačního modulu pro rozpoznávání typu šifrované komunikace přes TLS. Tento klasifikační modul podle provedených experimentů dosahuje vysoké úspěšnosti. Celkově vypadá bakalářská práce velmi zajímavě a výsledky jsou slibné. V textu chybí informace o reálném nasazení v síťové infrastruktuře, ale popisované experimenty a datové sady se zdají být dostatečným otestováním vyvinutého výsledku. Pro praktické použití v reálné síti většího rozsahu by bylo potřeba vylepšit implementaci prototypu a tím dosáhnout vyšší rychlosti. Vzhledem k tomu, že hlavní přínos této práce jsou primárně výzkumné úspěchy kolem přesné klasifikace, nepovažuji rychlost prototypu za závažný nedostatek odevzdané práce.

Otázky k obhajobě

Jak přesně probíhá zpracování vstupních dat?
Je možné transformaci charakteristik pomocí metody PCA uplatnit i pro jiné klasifikátory síťového provozu než je vytvořený tls_classifier.py?

Instrukce

Splnění zadání

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.

Písemná část práce

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.

Nepísemná část, přílohy

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.

Hodnocení výsledků, jejich využitelnost

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.

Celkové hodnocení

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.