

Posudek oponenta bakalářské práce

Student: Borbély Richard
Téma: Identifikace mobilních aplikací s využitím TLS dat (id 23693)
Oponent: Matoušek Petr, Ing., Ph.D., M.A., UIFS FIT VUT

1. **Náročnost zadání** průměrně obtížné zadání
Cílem práce bylo vytvořit klasifikátor pro identifikace mobilních aplikací na základě otisků komunikace TLS.
2. **Splnění požadavků zadání** zadání splněno
3. **Rozsah technické zprávy** je v obvyklém rozmezí
4. **Prezentační úroveň předložené práce** 60 b. (D)
V kapitole 2.1 není potřeba popisovat model OSI, který se v práci nevyužívá. V části 3.1 chybí podrobnější popis datové sady, tj. odkud ji student získal, jaký má formát, kolik dat obsahuje, jakých verzí aplikací se týká apod. Části 3.8 a 3.9 popisující vstupy a výstupy aplikace logicky patří do kapitoly 4 (Implementace). V podkapitolách 4.4-4.6 není potřeba dopodrobna popisovat jednotlivé funkce. Pro vysvětlení je vhodnější použít vývojový diagram, pseudokód či stavový automat.
5. **Formální úprava technické zprávy** 55 b. (E)
Práce obsahuje větší množství překlepů: fingerprintig, paramertrami, funckie, porovnávaní atd., což se dá odstranit automatickou kontrolou pravopisu. Některé použité výrazy jsou nespisovné (týmpádom) či nevhodně přeložené (enkryptovaná data). Při skloňování cizích slov (handshake, data mining) nelze přidávat za pomlčku slovenské koncovky (handshake-om, handshake-y, data mining-om), ale je třeba číst platná jazyková pravidla.
6. **Práce s literaturou** 55 b. (E)
Student využívá převážně internetové zdroje. U citací online zdrojů by měl být uveden autor a datum publikace (pokud jsou známy), chybí také datum přístupu. Místo odkazů na wikipedii je vhodné použít existující standardy (např. RFC pro popis TLS) či doporučenou knihu Data Mining pro definici klasifikačních metrik. U standardu OSI 7498 chybí v literatuře autor i název standardu.
7. **Realizační výstup** 50 b. (E)
Vytvořená aplikace slouží k analýze dat a vyhodnocení přesnosti klasifikace pomocí vybraných vstupních atributů. U vytvořeného klasifikátoru v jazyce Python chybí ve zdrojovém souboru jméno autora, datum a požadované parametry pro spuštění. Zcela chybí instalační manuál s požadovanou verzí Pythonu a potřebnými knihovnami. Při běhu s nesprávnými vstupy (např. chybějící dataset) aplikace padá. Výsledkem analýzy jsou matice obsahující počty správně či nesprávně identifikovaných aplikací. Pro podrobnější analýzu výsledků by bylo dobré, kdyby aplikace vypisovala průběžné výsledky zpracování, např. množinu nesprávně detekovaných otisků, apod.
8. **Využitelnost výsledků**
Vytvořená aplikace je navržena jako jednorázová - obsahuje přímo zadaný seznam testovaných mobilních aplikací a jejich počet, což omezuje její využití. Slouží spíše pro demonstrační účely. Zajímavým přínosem je výpočet podobnosti SNI a klíčových slov. Způsob výpočtu (str. 23) není z mého pohledu sémanticky vhodný, neboť například podřetězec obsažený v jiném řetězci dává různou míru podobnosti v závislosti na délce většího řetězce.
9. **Otázky k obhajobě**
 1. Příložená datová sada pro danou aplikaci obsahuje i TLS handshaky jiných aplikací, což má vliv na učení a přesnost klasifikace. Jak poznáte, které otisky patří dané aplikaci a které ne?
10. **Souhrnné hodnocení** 60 b. uspokojivě (D)
Vytvořená aplikace demonstruje možnost identifikace mobilních aplikací na základě vybraných atributů. Výběr datasetů a atributů je fixní, což omezuje její použití. Celkově hodnotím práci stupněm D.

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

V Brně dne: 3. června 2021

Matoušek Petr, Ing., Ph.D., M.A.
oponent