Review of Master's Thesis

Student: Sedlo Ondřej, Bc.
Title: Improvement of Adversarial Classification in Behavioral Analysis of Network Traffic Intended for Targeted Attack Detection (id 22643)
Reviewer: Malinka Kamil, Mgr., Ph.D., DITS FIT BUT

1. Assignment complexity — **more demanding assignment**

Obtížnost zadání hodnotím jako nadprůměrnou, protože pro úspěšné vyřešení všech bodů, bylo potřeba vyřešit několik navazujících a poměrně rozsáhlých komponent - nastudování a pochopení principů detekce síťových útoků, nastudování problematiky klasifikátorů a zvládnutí jejich praktické aplikace, vytvoření vlastní aktualizované datové sady (poměrně časově náročné) a vlastní provedení analýz nad novou datovou sadou a její porovnání s již existujícími sadami. Rozsah nutných prací přesahuje obvyklou hranici pro diplomovou práci.

2. Completeness of assignment requirements — **assignment fulfilled**

Práce splňuje všechny body zadání. V bodě 5 jej dokonce rozšiřuje nad rámec požadavků, tedy z původních vyžadovaných 3 klasifikátorů použitých pro porovnání student realizoval měření na 6 klasifikátorech.

3. Length of technical report — **exceeds requirements**

Práce je delší než obvykle. Je to dáno i obtížnějším zadáním, které vyžadovalo pokrýt více oblastí, nicméně zejména první půlka mohla být stručnější. Student prezentuje tří druhy taxonomií, které jsou výborně zpracovány jako kvalitní rešerše, nicméně mi chybí jejich větší aplikace v další části a alespoň nějaká snaha o interpretaci a vysvětlení, k čemu je chce dále použít (bohužel se to neobjeví ani dále v práci). Dále způsob prezentace taxonomií je až moc hutný, odkazuje se na jiné relevantní články, ale v rámci pochopení jednotlivých případů dané systematiky pak nutí čtenáře k dvojímu zanoření pro dohledání referencí, takže se ztrácí přidaná hodnota rešerše. Nicméně student dobře ukazuje svou orientaci v dané oblasti.

4. Presentation level of technical report — **100 p. (A)**

K prezentační úrovni správy nemám v podstatě žádné výhrady. Úroveň práce snese srovnání s kvalitní vědeckou publikací. Popis a realizace je extrémně systematická.
Jedinou výtku mám k jisté absenci snahy o vlastní interpretaci prezentovaných taxonomií (hlavně v kontextu k navazujícím kapitolám) a výsledků. Dále není občas jasno, co si čtenář má vlastně odnést z dané kapitoly.

5. Formal aspects of technical report — **97 p. (A)**

Velmi dobrá stylistika a jazyk (práce je navíc v angličtině). Obsahuje jen minimální množství typografických chyb. Drobné výtky mám k umístění obrázku, kdy dohledávání odkazovaného obrázku o 3 strany dál není úplně šťastné.

6. Literature usage — **100 p. (A)**

Vše řádně citováno. Práce obsahuje velké množství relevantních referencí.

7. Implementation results — **97 p. (A)**

Realizační výstup se skládá z několika částí: vytvoření databáze (což vyžadovalo implementaci všech útoků v testovacím prostředí, vytvoření parseru pro extrakci ASNM vlastností), přípravu dat pro klasifikaci a pak vlastní srovnání s již existujícími sadami.

8. Utilizability of results

Výsledky jsou využitelné bezpečností komunitou, kdy je databáze využitelná pro zlepšování výkonu klasifikátorů pro IDS. Výstupem nad rámec DP je pak plánované vědecká publikace (ve spolupráci s vedoucím práce).

9. Questions for defence

1. Na základě jakých parametrů jste vybral 6 použitých klasifikátorů?
2. Existují i jiné použitelné zdroje resp. databáze zranitelností, které byste mohl využít pro obohacení vašeho datasetu?
3. V datasetu se vyskytuje pouze 11 služeb, na které útočíte. To mi přijde málo. Z jakého důvodu je ten počet takto nízký. Proč jste vybral zrovna tyto služby? Kolik je to % z CVE dat?
4. Jaké byly časové náročnosti jednotlivých analýz z kapitol 7 a 8?

10. Total assessment — **97 p. excellent (A)**

Velmi pozitivně hodnotím kompaktnost celé práce, velmi kvalitní zpracování jak textu, tak realizačního výstupu a hlavně množství odvedené práce. Výsledky ukazují důležitost trénování klasifikátorů nad obfuskovanými daty a zároveň potřebu aktualizace trénovacích sad o nové útoky.

Práce dobře popisuje jednotlivé provedené kroky, nicméně mi často chybí lepší vysvětlení motivace, proč danou věc autor realizoval zrovna takto (což by bylo vhodné, pokud by to někdo chtěl výsledky reprodukovat). Je škoda, že student nevyužil plný potenciál dosažených výsledků, kdy dostatečně neprezentuje některé dílčí zajímavé výsledky - např. str. 55 - úspěšnosti obfuskačních technik. To je velmi zajímavý poznatek využitelný bezpečnostní komunitou, ale je dobře schován v hutném textu. Také interpretace výsledků se často uchyluje pouze k popsaní výsledných čísel bez další snahy o interpretaci.

Nicméně všechny mé připomínky jsou v podstatě jen drobnosti, jak vylepšit již tak výbornou práci. Práci doporučuji k obhajobě a hodnotím "A", dále doporučuji komisi práci navrhnout jako kandidáta na cenu děkana, případně i do vybraných soutěží diplomových prací.

In Brno 29 June 2020

Malinka Kamil, Mgr., Ph.D.
reviewer