

dr hab. inż. Rafał Kozik, profesor uczelni
Politechnika Bydgoska
im. Jana i Jędrzeja Śniadeckich w Bydgoszczy,
Wydział Telekomunikacji, Informatyki i Elektrotechniki,
Al. prof. S. Kaliskiego 7,
85-796 Bydgoszcz

Bydgoszcz, 18.12.2022

RECENZJA ROZPRAWY DOKTORSKIEJ WYKONANA DLA RADY NAUKOWEJ DYSCYPLINY INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ

Tytuł rozprawy: Wykorzystanie protokołu HTTP do identyfikacji i klasyfikacji złośliwego oprogramowania

Autor rozprawy: mgr inż. Piotr Biańczak

- 1. Jakie zagadnienie naukowe jest rozpatrzone w pracy (teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez Autora? Jaki charakter ma rozprawa (teoretyczny, doświadczalny, itd.)?**

Zakres rozprawy, której Autorem jest Pan mgr inż. Piotr Biańczak, dotyczy wykorzystania protokołu HTTP do identyfikacji i klasyfikacji złośliwego oprogramowania. Autor niniejszej pracy zakłada również, iż możliwym jest rozpoznanie rodziny tego oprogramowania oraz wskazanie obecności przedstawicieli tych rodzin, których (w danej chwili) pochodzenie nie jest znane. Wymaga to istnienia pewnych cech w żądaniu HTTP, które występują dla oprogramowania złośliwego, ale są niezwykle rzadkie dla oprogramowania niezłośliwego.

W rozdziale pierwszym, Autor rozprawy precyzyjnie wskazał kluczowe wyzwania i problemy związane z tematyką rozprawy. W szczególności, Autor słusznie wskazuje jak istotnym zagrożeniem dla użytkowników Internetu jest złośliwe oprogramowanie i jednym ze skutecznych narzędzi do walki z nim jest efektywna jego identyfikacja.

Ponadto, w rozdziale 1.2 Autor rozprawy bardzo szczegółowo nakreślił motywację i cel badań. Zidentyfikowane zostały kluczowe problemy dotychczasowych badań nad detekcją i klasyfikacją złośliwego oprogramowania przy użyciu protokołu HTTP. W szczególności Autor rozprawy słusznie wskazuje na brak możliwości pracy w tzw. scenariuszu otwarto-zbiorowym.

W świetle niniejszego wprowadzenia, określenia celu badań oraz motywacji, postawiona zostaje teza rozprawy. Jest ona sformułowana przez Autora w sposób jasny.

W mojej ocenie rozprawa ma charakter teoretyczno-eksperymentalny. W rozdziale 1.3, Autor rozprawy definiuje kilka szczegółowych zagadnień badawczych (tzw. podtezy), dla których jasno zostaje zdefiniowany sposób rozwiązania (udowodnienia). Dla każdego z zagadnień zaplanowany i wykonany zostaje odpowiedni eksperyment.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł (w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle) świadczącej o dostatecznej wiedzy Autora. Czy wnioski z przeglądu źródeł sformułowano w sposób jasny i przekonujący?

W mojej ocenie analiza literaturowa podzielona została na dwie części: ogólny zarys problematyki związanej z rozprawą doktorską (rozdział 1) oraz właściwa analiza i przegląd aktualnego stanu wiedzy (rozdział 3).

Pierwsza część analizy, pozwala czytelnikowi zrozumieć kluczowe problemy i ograniczenia dotychczasowych badań nad detekcją i klasyfikacją złośliwego oprogramowania przy użyciu protokołu HTTP. W szczególności Autor rozprawy nakreśla tło badań, powód i słuszność wyboru protokołu HTTP do detekcji złośliwego oprogramowania. Pozwala to w jasny i czytelny sposób określić cel niniejszej rozprawy.

Druga część analizy (zawarta w rozdziale 3) stanowi analizę najważniejszych prac i badań, które związane są z problematyką rozprawy. Bibliografia niniejszej rozprawy zbudowana jest z ponad 100 pozycji. Blisko połowa z nich została opublikowana w latach 2017-2022. W szczególności Autor przeanalizował sporą część źródeł, które opublikowane zostały w latach 2020 oraz 2021. Niniejsza analiza przeprowadzona jest w sposób właściwy i nawiązuje do charakterystyki ruchu sieciowego protokołu HTTP, problematyki identyfikacji żądań oraz wykrywania złośliwego oprogramowania.

W mojej ocenie wnioski z analizy literatury są sformułowane w sposób jasny i są trafne. W szczególności, w świetle przedstawionego aktualnego stanu wiedzy, słuszne jest stwierdzenie, że niewiele rozwiązań bazujących na protokole HTTP ma możliwości pracy w tzw. scenariuszu otwarto-zbiorowym.

3. Czy Autor rozwiązał postawione zagadnienia? Czy użył do tego właściwych metod i czy przyjęte założenia są uzasadnione?

W mojej ocenie Autor w sposób właściwy rozwiązał postawione zagadnienia. W procesie tym użyte zostały odpowiednie narzędzia i metody. W szczególności badania i eksperymenty osadzone zostały na obiektywnej analizie literaturowej.

Autor rozprawy odniósł się także do aktualnych trendów i ograniczeń metod typowo wykorzystywanych przy klasyfikacji złośliwego oprogramowania. Ponadto, w rozdziale 4 przeprowadzono szczegółową analizę ruchu sieciowego protokołu HTTP. Pozwoliło to Autorowi niniejszej pracy na sformułowanie i skonstruowanie odpowiednich cech żądań HTTP, które docelowo umożliwiły stworzenie unikatowej reprezentacji żądań złośliwego oprogramowania oraz modeli klasyfikacji rodzin złośliwego oprogramowania.

W mojej ocenie, zaplanowane i przeprowadzone eksperymenty są wiarygodnym dowodem zrealizowania przez Autora celów badawczych. Tym samym Autor dowiódł, że posiadał umiejętności związane z metodyką prowadzenia badań.

4. Na czym polega oryginalność rozprawy, co stanowi samodzielny i oryginalny dorobek Autora, jaka jest pozycja rozprawy w stosunku do stanu wiedzy czy poziomu techniki reprezentowanej przez literaturę światową?

Oryginalność rozprawy oraz samodzielny dorobek Autora, stanowi propozycja wykorzystania analizy żądań HTTP do identyfikacji i klasyfikacji złośliwego oprogramowania. W szczególności, w trakcie prowadzonych prac badawczych, Autor rozprawy:

- Eksperymentalnie udowodnił, iż ruch sieciowy protokołu HTTP złośliwego oprogramowania zawiera specyficzne cechy, które umożliwiają odróżnienie ruchu sieciowego aplikacji złośliwych i niezłośliwych.
- Opracował nowatorskie narzędzie Hfinger, które umożliwia tworzenie unikalnej reprezentacji żądań protokołu HTTP w efektywniejszy sposób niż inne popularne narzędzia.
- Eksperymentalnie dowiódł, że reprezentacja żądań HTTP tworzona przez narzędzie Hfinger może być wykorzystana do skutecznej klasyfikacji otwarto-zbiorowej złośliwego oprogramowania.

Oceniając pozycję niniejszej rozprawy w odniesieniu do innych prac, można stwierdzić, iż jest ona zgodna z aktualnym stanem wiedzy oraz poziomem techniki reprezentowanej przez literaturę światową. Punktowane publikacje, w których Autor niniejszej rozprawy jest pierwszym autorem, pozwalają stwierdzić, iż zademonstrowane osiągnięcia naukowe stanowią istotny wkład w literaturze międzynarodowej.

5. Czy Autor wykazał umiejętności poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

Nie mam wątpliwości, że Autor rozprawy posiada dużą wiedzę dotyczącą zagadnień związanych z tematem rozprawy doktorskiej a w szczególności w obszarze cyberbezpieczeństwa. Doktorant wykazał się przede wszystkim umiejętnościami dotyczącymi prowadzenia badań naukowych, eksperymentów oraz projektowania algorytmów.

Również zrealizowane przez Autora rozprawy eksperymenty są zgodne z dobrymi praktykami w zakresie pomiaru skuteczności systemów klasyfikacji. Autor w tym zakresie użył znanych metryk (np. F1-score) co pozwala odnieść się do innych rozwiązań dostępnych w literaturze.

Na zwrócenie uwagi zasługuje także fakt, iż Autor rozprawy w poprawny sposób posługuje się różnego rodzaju diagramami i wykresami (np. strona 72 rozprawy) do analizy danych. Pozwala to skondensować stosunkowo dużą ilość informacji na diagramie o małej powierzchni, co docelowo umożliwia wizualną eksplorację żądań HTTP.

6. Jakie są wady i słabe strony rozprawy?

Rolą recenzenta jest zauważenie ewentualnych niedociągnięć i mankamentów przedstawianej pracy, oraz zgłoszenie uwag, które mogą być pomocne i przydatne w dalszych pracach. W szczególności:

- Przy opracowywaniu wyników, Autor mógł przedstawić wartości średnie i odchylenia standardowe z uwzględnieniem istotności statystycznej przedstawianych różnic. Przykładowo dla tabeli 6.3 łatwiej byłoby porównać czy poszczególne metody różnią się między sobą pod względem wartości F1 oraz MCC.
- Algorytm 1 przedstawia autorską propozycję protokołu badań, której podstawą jest losowy podział danych na testowe i treningowe. Zastanawiające jest dlaczego Autor nie dokonał modyfikacji 5-krotnej walidacji krzyżowej dla problemu otwarto-zbiorowego. Pozwoliłoby to zapewnić, że przy każdej próbie testowania wykorzystywane byłyby inne próbki.
- W przypadku analizy żądań HTTP, jego ciało (payload) reprezentowane jest tylko przez trzy wartości (obecność znaków specjalnych, wartość entropii Shannona oraz długość). Zastanawiające jest, czy Autor rozważał bardziej szczegółowy zapis cech, który pozwoliłby uchwycić więcej detali (argumenty, format, itd.).
- W pracy można także dopatrzyć się kilku, mało istotnych pomyłek edytorskich, takich jak: „wykorzystanietcyh cech” czy „sandboksach”.

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Pomimo przedstawionych powyżej uwag, rozprawa mgr inż. Piotra Biańczaka posiada wiele silnych stron. Przede wszystkim rozprawa dotyczy ważnej tematyki z zakresu cyberbezpieczeństwa i wnosi interesujący wkład do zagadnień związanych z klasyfikacją

złośliwego oprogramowania. W szczególności, zawiera ona autorską propozycję narzędzia Hfinger. Jak pokazały wyniki z eksperymentów, proponowane rozwiązanie można wykorzystać do efektywnego identyfikowania żądań HTTP wysyłanych przez złośliwe oprogramowanie.

Na podkreślenie zasługuje również fakt udostępnienia przez Autora rozprawy kodu źródłowego, który powstał w trakcie badań. Z jednej strony pozwala to innym badaczom na przejrzyste odtworzenie najważniejszych wyników. Z drugiej strony stanowi to przyczynek do dalszych eksperymentów, proponowania innych algorytmów klasyfikacji, itd.

Również stworzony (zgromadzony) przez Autora zbiór danych ruchu sieciowego HTTP złośliwego oprogramowania (oraz aplikacji niezłośliwych), może stanowić interesujący wkład do dalszych badań i z powodzeniem powinien być wykorzystane przez innych badaczy.

Warto także zauważyć, że Autor ma udokumentowany dorobek publikacyjny w zakresie rozprawy. W publikacjach Doktorant jest pierwszym autorem, więc jego wkład można uznać za znaczący.

Wniosek

Niewątpliwie recenzowana rozprawa doktorska dowodzi dużej wiedzy Autora dotyczącej zagadnień związanych z rozprawą doktorską. Liczne i rozbudowane eksperymenty dowodzą skrupulatności i pozwalają stwierdzić, że Autor rozprawy opanował technikę planowania i prowadzenie badań naukowych. Ponadto, recenzowana praca jasno formułuje tezę, która została udowodniona poprzez badania eksperymentalne i realizację wszystkich celów badawczych.

Wobec powyższego stwierdzam, że recenzowana praca **spełnia wymagania stawiane rozprawom doktorskim** przez obowiązujące przepisy. Dlatego wnoszę o przyjęcie niniejszej rozprawy i **dopuszczenie mgr inż. Piotra Białczaka do publicznej obrony.**

