Wrocław, 28.11.2022

Prof. dr hab. inż. Sławomir Sujecki
Katedra Telekomunikacji i Teleinformatyki
Politechnika Wrocławska

Recenzja rozprawy doktorskiej mgr inż. Piotra Białczaka

Wykorzystanie protokołu http do identyfikacji i klasyfikacji złośliwego oprogramowania

1. Jakie zagadnienie naukowe/badawcze jest rozpatrywane w pracy (cel i teza rozprawy) i czy zostało ono dostatecznie jasno sformułowane przez autora?

W pracy sformułowano następującą **tezę**: Wykorzystanie żądań protokołu HTTP umożliwia identyfikację złośliwego oprogramowania poprzez stworzenie unikalnej reprezentacji jego ruchu sieciowego oraz skuteczne rozpoznanie rodziny złośliwego oprogramowania, a także obecności przedstawicieli rodzin dotychczas nieznanych. Sformułowano także trzy podtezy:

1. Ruch sieciowy żądań protokołu HTTP złośliwego oprogramowania zawiera cechy charakterystyczne, które umożliwiają jego odróżnienie od ruchu sieciowego aplikacji niezłośliwych.
2. Możliwe jest stworzenie unikalnej reprezentacji żądań protokołu HTTP, która umożliwia identyfikację złośliwego oprogramowania.
3. Wykorzystanie odpowiednio stworzonej reprezentacji żądań HTTP umożliwia skuteczne rozpoznanie rodzin złośliwego oprogramowania, w tym także istnienia klas dotychczas nieznanych.

Natomiast zasadniczym **celem pracy** jest stworzenie oprogramowania do detekcji i klasyfikacji złośliwego oprogramowania przy użyciu protokołu HTTP i metod uczenia maszynowego operujących w scenariuszu otwartozbiorowym. Opracowane oprogramowanie ma umożliwić klasyfikację znanych apriori klas złośliwego oprogramowania oraz wykrywać obecność klas nieznanych, tzn. takich które nie były uwzględnione w zbiorze uczącym danej metody uczenia maszynowego.

Cel i teza rozprawy zostały sformułowane jasno.

2. Czy w rozprawie przeprowadzono w sposób właściwy analizę źródeł, w tym literatury światowej, stanu wiedzy i zastosowań w przemyśle?

W spisie literatury autor rozprawy powołał się na 117 pozycji przy czym około 25 razy odniósł się do wiodących czasopism naukowych z zakresu informatyki i telekomunikacji, t.j. IEEE Access, IEEE Transactions on Network and Service Management, IEEE Transactions on Pattern Analysis and Machine Intelligence. Ponadto spis literatury zawiera odnośniki do istotnych norm RFC, zasobów udostępnionych na portalu github, konferencji naukowych z zakresu telekomunikacji

i informatyki i książek. Autor powołał się wielokrotnie na dostępne jedynie w internecie. Co zważywszy, że tematyka pracy dotyczy cyberbezpieczeństwa w mojej opinii jest uzasadnione. Stwierdzam zatem, że w mojej opinii analiza źródeł została przeprowadzona w sposób właściwy i stanowi wierne odwzorowanie obecnego stanu wiedzy w zakresie rozważanej tematyki.

3. Czy autor rozwiązał postawione zagadnienia, czy użył właściwej do tego metody i czy przyjęte założenia są uzasadnione?

Autor stworzył pakiet oprogramowania do identyfikacji złośliwego oprogramowania poprzez analizę żądań protokołu HTTP. W ramach badań opracował we własnym zakresie opracował nowatorskie narzędzie Hfinger, które tworzy reprezentacje pojedynczych żądań HTTP, a to z kolei umożliwia identyfikację złośliwego oprogramowania. Przypisanie jednoznacznych i unikalnych reprezentacji do żądań wygenerowanych przez próbkę złośliwego oprogramowania pozwala stworzyć tzw. odcisk palca (ang. fingerprint), który umożliwia identyfikację i klasyfikację złośliwego oprogramowania. W drugim kroku autor zastosował metody uczenia maszynowego do otwartozbiorowej klasyfikacji złośliwego oprogramowania.

Według mojej opinii autor rozwiązał postawione zagadnienia i użył właściwej metody. Przyjęte założenia są uzasadnione.

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

Autor wykorzystał i opisał w rozprawie szereg nowych rozwiązań, w tym projekt oraz implementację narzędzia Hfinger, metody detekcji oraz klasyfikacji złośliwego oprogramowania wykrytego w ruchu HTTP. Ponadto zastosował metody uczenia maszynowego do otwartozbiorowej klasyfikacji złośliwego oprogramowania. Są to rozwiązania oryginalne i nowatorskie. Stanowią one zatem samodzielny i oryginalny dorobek autora. Zaprezentowane rozwiązanie zastosowane do wykrywania i klasyfikacji złośliwego oprogramowania nie było według mojej wiedzy wcześniej publikowane przez innych autorów w Polsce lub na świecie.

5. Czy autor wykazał umiejętność poprawnego i przekonywującego przedstawienia uzyskanych przez siebie wyników (zwięzłość, jasność, poprawność redakcyjna rozprawy)?

W mojej opinii autor wykazał się umiejętnością poprawnego i przekonującego przedstawienia uzyskanych przez siebie wyników. Praca jest napisana starannie choć pozostało w tekście kilka literówek i możnaby jeszcze popracować nad poprawą stylu. Doktorant poprawnie wykorzystał tabele i różne rodzaje wykresów w celu poparcia wniosków. Istotną część 4 i piątego rozdziału stanowią artykuły napisane w języku angielskim. Takie podejście 'hybrydowe' trochę narusza ciągłość narracji w

i prawdopodobnie lepiej byloby dokonać tłumaczenia na język polski ale w mojej opinii jest akceptowalne w przypadku szybko rozwijających się dziedzin nauki takich jak cyberbezpieczeństwo.

6. Jaka jest przydatność rozprawy dla nauk inżynieryjno-technicznych?

Ogólnie opracowane rozwiązania mogą być wykorzystane w narzędziach monitorujących ruch sieciowy. Moim zdaniem zaletą zaproponowanej metody jest to, iż nie spowalnia ona przesyłu informacji w sieci gdyż polega na jedynie biernej analizie żądań protokołu HTTP.

Ponadto należy zauważyć, że bardzo ważnym aspektem rozprawy jest jej wartość praktyczna. Opracowane oprogramowanie zostało bowiem przetestowane w ramach projektu Unii Europejskiej Horyzont dla CERT Polska. Świadczy to o tym, że opracowane oprogramowanie stanowi praktyczne narzędzie do wykrywania i klasyfikacji złośliwego oprogramowania i w związku z tym są duże możliwości jego potencjalnego wykorzystania w systemach cyberbezpieczeństwa.

Dodatkowo autor umieścił własne kody źródłowe w otartych repozytoriach (github.com). Umożliwia to innym specjalistom z tej dziedziny weryfikacje wyników uzyskanych przez autora oraz wykorzystanie opracowanego oprogramowania przez innych naukowców we własnych pracach badawczych.

Biorąc pod uwagę przedstawioną przez Doktoranta rozprawę 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.

Sławomir Kujawa