Poznań, 19.09.2024 r.

Dr hab. inż. Krzysztof Kurowski
Instytut Chemii Bioorganicznej PAN
Poznańskie Centrum Superkomputerowo – Sieciowe
Ul. Jana Pawła II 10
61-139 Poznań

WOJSKOWA AKADEMIA TECHNICZNA
im. Jarosława Dąbrowskiego
Wydział Cybernetyki
ul. Gen. Sylwestra Kaliskiego 2
00-908 Warszawa 46

Recenzja

Przedmiotem recenzji jest rozprawa mjr mgr inż. Artura Stachurskiego zatytułowana „*Model i metoda prognozowania podatności oprogramowania na zagrożenia w cyberprzestrzeni*". Promotorem rozprawy jest dr hab. inż. Andrzej Najgebauer, prof. WAT, a promotorem pomocniczym płk dr inż. Rafał Kasprzyk. Recenzja została opracowana na podstawie zlecenia i uchwały nr 29/RDN ITiT/2024 z dnia 9 lipca 2024 r. podjętej przez Radę Dyscypliny Naukowej Informatyka Techniczna i Telekomunikacja Wojskowej Akademii Technicznej.

1. Znaczenie podjętej tematyki oraz cel rozprawy

Rosnąca liczba i złożoność udanych cyberataków zaobserwowanych w ostatnich dekadach ma szereg poważnych konsekwencji dla bezpiecznego funkcjonowania wielu organizacji, poszanowania prywatności użytkowników, a w skrajnych przypadkach bezpośrednio zagraża bezpieczeństwu infrastruktury krytycznej, co może skutecznie sparaliżować kluczowe procesy w centralnych instytucjach państwa. Od wielu lat jednym z kluczowych elementów wielu cyberataków jest odpowiednie wykorzystanie podatności oprogramowania na różnego rodzaju zagrożenia w cyberprzestrzeni. Problematyka związana z modelowaniem i prognozowaniem podatności oprogramowania oraz efektywnym zarządzaniem poziomem bezpieczeństwa, co w konsekwencji pozwala wypracować nowe i skuteczne metody zmniejszające ryzyko potencjalnych zagrożeń oraz zwiększa ogólną odporność systemów informatycznych na współczesne cyberataki, stanowi bardzo ważny obszar naukowo-badawczy z zakresu cyberbezpieczeństwa oraz inżynierii oprogramowania. Dodatkowo, ze względu na złożoność i

różnorodność współcześnie występujących podatności oprogramowania na zagrożenia, problemy te nie są trywialne, zwłaszcza w kontekście ich praktycznego zastosowania. Warto również podkreślić, że omawiana problematyka wymaga znajomości wielu zagadnień z dziedziny informatyki i informatyki technicznej, takich jak inżynieria bezpieczeństwa oprogramowania, systemy wspomagania decyzji, modelowanie komputerowe, statystyka, a w kontekście stawianej w pracy doktorskiej głównej tezy, także praktycznego doświadczenia w analizie i audytach bezpieczeństwa oprogramowania.

Główna teza doktoratu mjr. mgr. inż. Artura Stachurskiego została sformułowana poprawnie i w przemyślany sposób akcentując trzy główne wyzwania badawcze, odpowiednio:

- opracowanie modelu matematycznego umożliwiającego szacowanie poziomu bezpieczeństwa oprogramowania w oparciu o wykrywanie podatności oprogramowania oraz zastosowanie działań profilaktycznych, w tym:
  - opracowanie metody wyznaczania intensywności wykrywania podatności oprogramowania na zagrożenia w cyberprzestrzeni (w danym przedziale czasu);
  - pracowanie metody szacującej skuteczność zastosowanych działań profilaktycznych;
- konstrukcja prototypu komputerowego narzędzia umożliwiającego analizę oraz prognozowanie podatności oprogramowania na zagrożenia w cyberprzestrzeni;
- weryfikacja opracowanego modelu i metody w oparciu o dane z bazy danych podatności oprogramowania na cyberzagrożenia.

Doktorant we wstępnych rozdziałach rozprawy merytorycznie charakteryzuje obszar badawczy, przeprowadzając przegląd kluczowych pojęć i definicji z zakresu cyberbezpieczeństwa. Na tej podstawie Doktorant prezentuje wybrane modele klasyfikacji zagrożeń i podatności, statystyki incydentów i zagrożeń, ataków cyberbezpieczeństwa, wykorzystując zarówno własne, jak i referencyjne opracowania literaturowe, akcentując praktyczne znaczenie omawianej problematyki. Chociaż zaprezentowane statystyki i wykresy nie obejmują ostatnich lat, Doktorant uwzględnia aktualny stan wiedzy w kontekście analizy ilościowej podatności. Pełny przegląd literatury przedstawiony jest w dalszej części rozprawy, odnosząc się do istniejących modeli i metod analizy oraz prognozowania podatności. Na podstawie obszernej analizy literatury, z zastosowaniem narzędzi matematycznych, probabilistyki i statystyki, w kolejnych rozdziałach Doktorant wprowadza podstawowe założenia niezbędne do konstrukcji bazowych modeli, ich autorskich rozszerzeń oraz proponowanych metod obliczeniowych do prognozowania podatności oprogramowania, z wykorzystaniem wybranych własności łańcucha Markowa. Bazując na tych modelach, Doktorant definiuje i proponuje nowe metody, przedstawiając przykładowe obliczenia dla szacowania i oceny poziomu bezpieczeństwa oprogramowania, intensywności wykrywania krytycznych podatności oraz skuteczności działań profilaktyczno-naprawczych. Na tej podstawie Doktorant proponuje autorskie podejście do wykorzystania mechanizmów uczenia maszynowego w celu przyporządkowania nowo wykrytej podatności do wcześniej zdefiniowanych kategorii. Do realizacji zadania klasyfikacyjnego Doktorant wykorzystuje wyniki eksperymentów obliczeniowych dla czterech modeli uczenia maszynowego: regresji logistycznej, analizy dyskryminacyjnej liniowej, analizy dyskryminacyjnej kwadratowej oraz algorytmu k najbliższych sąsiadów. Rozprawa doktorska została również uzupełniona o założenia oraz opis funkcjonalności autorskiej aplikacji wspierającej obliczenia dla zaproponowanych modeli i metod prognozowania podatności oprogramowania.
Podsumowując, praca naukowa, zarówno w teorii, jak i w praktyce, wpisuje się w nurt badań w dziedzinie Informatyki Technicznej i Telekomunikacji. Tematyka opracowania modelu matematycznego do prognozowania podatności oprogramowania na zagrożenia w

cyberprzestrzeni, w połączeniu z opracowanymi i zweryfikowanymi ilościowo metodami, ma istotne znaczenie naukowe. Wyniki prac badawczych Doktoranta stanowią znaczący i twórczy wkład w rozwój nowych modeli prognozowania podatności na zagrożenia w cyberprzestrzeni. Warto również podjąć dalsze badania, szczególnie w zakresie podatności wynikających z nieprawidłowego użytkowania systemów przez rosnącą liczbę użytkowników oraz coraz bardziej wyrafinowanych ataków z użyciem technik inżynierii społecznej.
Zakres działań naukowych przedstawionych w pracy odpowiada wymaganiom stawianym rozprawom doktorskim w dziedzinie Informatyki Technicznej i Telekomunikacji, zarówno pod względem trudności, jak i oryginalności.

2. Analiza i ocena poszczególnych fragmentów pracy

Rozprawa doktorska mjra mgra inż. Artura Stachurskiego to spójny tematycznie opis oryginalnej pracy naukowej. W rozprawie doktorskiej uwzględniono również artykuły naukowe Doktoranta [58-59][64-65] opublikowane w latach 2016 – 2019, w tym trzy artykuły pokonferencyjne.

Przedłożona rozprawa doktorska składa się z pięciu tematycznych Rozdziałów głównych oraz trzech załączników opisujących podstawowe definicje i właściwości z probabilistyki wykorzystywane w modelu matematycznym oraz przykładowe podzbiory danych wykorzystywanych w eksperymentach obliczeniowych (łącznie 126 stron wraz z bibliografią). W samej rozprawie można wyróżnić zasadniczo trzy części, odpowiednio Rozdziały 1-2 skupiające się na wprowadzeniu podstawowych założeń i zagadnień wraz z przykładowymi ilustracjami związanymi z modelami i metodami analizy i prognozowania podatności występujących w literaturze; Rozdziały 3-4 prezentujące autorskie podejście do modelowania prognozowania podatności oprogramowania oraz autorskim rozwiązywaniu problemu klasyfikacji podatności z wykorzystaniem referencyjnych rzeczywistych danych historycznych z wykorzystaniem technik uczenia maszynowego. Rozdział 5 to zwięzłe i merytoryczne podsumowanie osiągniętych wyników pracy naukowej Doktoranta wraz ze szczegółowym opisem zaimplementowanej aplikacji, wykorzystanych technologiach i narzędziach programistycznych oraz prezentacją interfejsu graficznego i zakresu funkcjonalnego.

Główne motywacje oraz założenia rozprawy doktorskiej, w tym kluczowa teza naukowa oraz najistotniejsze cele dla podjętych badań naukowych wraz z krótkim podsumowaniem oryginalnego wkładu Doktoranta zostały przedstawione w Rozdziale 1. W tym wprowadzającym rozdziale zabrakło jednak pogłębionego przeglądu literaturowego, w szczególności w odniesieniu do statystyk i analiz cyberataków zebranych w ostatnich latach oraz wykorzystania inżynierii społecznej przez cyberprzestępców do ciągłego atakowania użytkowników systemów komputerowych, aplikacji społecznościowych, komunikatorów, itp. oraz wykradania ich danych. Relatywnie bardziej obszerny przegląd literaturowy został przedstawiony przez Doktoranta w odniesieniu do istniejących podejść oraz narzędzi wspierających rożne analizy ilościowe podatności oraz modele i metody analiz oraz prognozowania podatności. Zaprezentowane opisy wraz licznymi rysunkami i referencjami odpowiednio akcentują różne istotne zagadnienia oraz założenia, które będą podstawą do sformułowania autorskiego modelu matematycznego i zaproponowania nowych metod uwzględniających szereg dodatkowym wymagań omawianych w rozprawie. Doktorant nie

ustrzegł się jednak błędów redakcyjno-edycyjnych oraz braku spójności pomiędzy definicjami i miarami, które były wykorzystywane do oceny i ewaluacji różnych przytaczanych z literatury metod. Przykładem może być definicja *Precision*, *Recall*, ale już brak definicji *Accuracy* (*wykorzystanego np. w ocenie modelu VPM Tabela 4*). Warto również w tym miejscu zwrócić uwagę na poprawę sposobu prezentacji diagramów i wykresów w rozprawie doktorskiej. Niektóre diagramy i histogramy są bardzo duże, np. rys. 35-38, a niektóre wręcz przeciwnie są niestety mało nieczytelne np. rys 16,17, 30, itd. (zbyt mała czcionka, niedobrany kolor tła, itp.) co utrudnia szczegółowe analizy zaprezentowanych wyników prac eksperymentalnych i autorskich metod. Rozdział 3 dotyczy autorskich modeli prognozowania podatności oprogramowania, choć Doktorant znaczną jego część poświęca na szczegółowe omówienie dobrze znanych pojęć i narzędzi z obszaru probabilistyki i statystyki. Bardzo szczegółowo zaprezentowane i omówione zostały również zagadnienia związane z bazowym, dobrze znanym w literaturze modelem łańcucha Markowa oraz w kolejnych podrozdziałach autorskie rozszerzenia do tego modelu z zastosowaniem działań profilaktyczno-naprawczych i z uwzględnieniem przejść wielokrokowych. Rozdział 4 odnosi się już bezpośrednio do istotnego autorskiego rozwiązania problemu klasyfikacji podatności na podstawie danych historycznych z rzeczywistych bazy danych podatności z wykorzystaniem mechanizmów uczenia maszynowego. Doktorant wyraźnie zaznacza, iż w praktyce każda zarejestrowana podatność jest opisana za pomocą licznych parametrów i metadanych, obejmujących m. in. wartości pojedynczych składowych jej wektora oraz oszacowany stopień zagrożenia, określony w skali od 0-10, wyliczony na bazie gotowych formuł w systemie scoringowym CVSS, co można wykorzystać w celu przydzielenia podatności do ustalonych wcześniej kategorii. W zaprezentowanej w dalszej części analizie ilościowej Doktorant wybrał jednak tylko jedną, relatywnie małą próbkę danych o bardzo ograniczonym wektorze cech (*Załącznik 3*), a sam wektor cech został również ograniczony w interfejsie graficznym autorskiej aplikacji (*Rys. 70, Access, Complexity, Authentication, Availability*). Warto w tym kontekście rozważyć inne, być może bardziej elastyczne podejścia bazujące na zaawansowanych technikach uczenia maszynowego uwzględniające np. dodatkowe rzeczywiste metadane, opisy tekstowe, itp. dotyczące gromadzonych podatności oraz sposobów ich wykorzystania w cyberatakach opisanych bardziej szczegółowo w biuletynach bezpieczeństwa. Doktorant bardzo zwięźle na samym końcu rozprawy odnosi się do tej kwestii i zaznacza, iż wykorzystane w pracy modele i metody prognozowania podatności, z racji ograniczonego dostępu do rzetelnych danych rzeczywistych, mają zastosowanie przede wszystkim dla oprogramowania dedykowanego konkretnemu użytkownikowi końcowemu i takiego, które nie jest udostępniane do użytku zewnętrznego. Nie zmienia to jednak faktu, iż do analiz zaproponowanego modelu i metod można było wykorzystać dane syntetyczne lub też pogłębić analizy związane z rzeczywistymi danymi gromadzonymi w wielu w bazach danych podatności i biuletynach bezpieczeństwa wskazywanych w rozprawie doktorskiej (*Tabela 2. Źródła danych podatności. Opracowanie własne*). W ostatnim Rozdziale 5 szczegółowo zaprezentowano autorskie środowisko obliczeniowe wykorzystane do realizacji założeń eksperymentów obliczeniowych zaprezentowanych w rozdziale 3 i 4. Rozdział ten jest jednocześnie krótkim podsumowaniem podjętych tematów w rozprawie doktorskiej, w tym zrealizowanych przedsięwzięć oraz wyników badań, które potwierdzają trafność postawionej tezy, iż „***opracowanie modelu i metody prognozowania podatności oprogramowania na zagrożenia w cyberprzestrzeni pozwala na racjonalne kształtowanie poziomu bezpieczeństwa oprogramowania***". Ponadto, w ostatnim rozdziale Doktorant wskazuje potencjalne możliwości wykorzystania wyników prac w różnych i bardzo praktycznych obszarach problemowych, w tym:

- możliwość prognozowania oceny bezpieczeństwa oprogramowania użytkowanego przez użytkownika końcowego oraz wewnątrz organizacji;

- możliwość oceny oraz automatyzacji i usprawnienia procesu projektowania potencjalnych działań profilaktyczno-naprawczych na potrzeby wytwarzania kolejnych produktów oprogramowania;
- możliwość udoskonalenia strategii zarządzania bezpieczeństwem informacji wewnątrz organizacji;
- możliwość udoskonalenia programów szkolenia z wytwarzania bezpiecznego oprogramowania (SDL – Secure Development Lifecycle) przeznaczonego dla zespołów projektowych.

Konkludując, pomimo zwięzłej formuły opisowej nawiązującej do dobrze znanych probabilistycznych podejść i technik w zakresie modelowania matematycznego, wybranych i przeanalizowanych pod kątem zastosowań algorytmów uczenia maszynowego, pewnych drobnych braków i pomyłek redakcyjno-edytorskich, zawężonego zakresu rozważań i eksperymentów obliczeniowych, **rozprawa doktorska jest merytorycznie przekonująca**. Przedstawione wyniki eksperymentów obliczeniowych dla testowych próbek danych potwierdzają zakładane kluczowe tezy w zakresie poprawności, skalowalności oraz możliwość zastosowania opracowanego przez Doktoranta modelu i metod prognozowania podatności oprogramowania na zagrożenia cyberbezpieczeństwa.

3. Ocena końcowa pracy

Kandydat posiada dostateczną wiedzę w dyscyplinie Informatyka Techniczna i Telekomunikacja oraz nabył wymagane doświadczenie i umiejętność samodzielnego prowadzenia pracy naukowej.

Stwierdzam, że rozprawa doktorska Pana mjr mga inż. Artura Stachurskiego pt.: *„Model i metoda prognozowania podatności oprogramowania na zagrożenia w ceberprzestrzeni"* **spełnia wymogi określone w Ustawie o stopniach i tytule naukowym**, w tym: art. 14 ust. 2 pkt 2 i art. 20 ust. 5 ustawy z dnia 14.03.2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z 2017 r. poz. 1789 z późn. zm.).

W związku z tym **wnoszę o dopuszczenie Pana mjr mgr inż. Artura Stachurskiego do publicznej dyskusji nad Jego rozprawą doktorską** w Radzie Dyscypliny Informatyki Technicznej i Telekomunikacji Wojskowej Akademii Technicznej w Warszawie.