dr hab. inż. ***Maciej Walkowiak***
prof. uczelni
Politechnika Bydgoska im. Jana i Jędrzeja Śniadeckich w Bydgoszczy
Wydział Telekomunikacji, Informatyki i Elektrotechniki

Bydgoszcz, 24 kwietnia 2024 r.

RECENZJA ROZPRAWY DOKTORSKIEJ

mgr. inż. Mariusza SZARKA

Budowanie mechanizmu obrony przed dedykowanymi kampaniami phishingowymi

Podstawą do przygotowania recenzji rozprawy Pana mgr. inż. Mariusza Szarka jest pismo Pana dr. hab. inż. Zbigniewa Tarapaty, profesora WAT, przewodniczącego Rady Dyscypliny Naukowej *Informatyka Techniczna i Telekomunikacja* Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego w Warszawie z 13 grudnia 2023 roku, informujące o uchwałach Rady Dyscypliny Naukowej *Informatyka Techniczna i Telekomunikacja* Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego.

Opiniowana praca doktorska Pana Mariusza Szarka powstała pod naukowym kierunkiem Pana dr. hab. inż. Ryszarda Antkiewicza, profesora WAT.

Praca została przedstawiona Radzie Dyscypliny Naukowej *Informatyka Techniczna i Telekomunikacja* Wojskowej Akademii Technicznej im. Jarosława Dąbrowskiego w trybie przewidzianym w Ustawie z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce* (Dz. U. 2018 poz. 1668 z późn. zm.). Rada Dyscypliny Naukowej podjęła uchwałę nr 54/RDN ITiT/2023 powierzając mi wykonanie recenzji pracy doktorskiej.

1 *Zgodność sporządzonej recenzji z zaleceniami Rady Doskonałości Naukowej*

Rada Doskonałości Naukowej, na podstawie obowiązujących przepisów oraz prawomocnych orzeczeń sądowych wydała zalecenia dotyczące trybu oraz zawartości recenzji w postępowaniach o awans naukowy. W przypadku recenzji rozpraw doktorskich wspomniane zalecenia sugerują, aby recenzje zawierały trzy elementy:

1) ocenę wraz z uzasadnieniem, iż rozprawa doktorska prezentuje ogólną wiedzę teoretyczną osoby ubiegającej się o nadanie stopnia doktora w określonej dyscyplinie;
2) ocenę wraz z uzasadnieniem, że rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej lub artystycznej przez osobę ubiegającą się o nadanie stopnia doktora oraz
3) ocenę wraz z uzasadnieniem, iż rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego, oryginalne rozwiązanie w zakresie zastosowania wyników własnych badań naukowych w sferze gospodarczej lub społecznej albo oryginalne dokonanie artystyczne.

Jednocześnie Rada doskonałości Naukowej stwierdza, że brak jest podstaw, by recenzenci wyrażali w swych recenzjach opinie odnoszące się do innych kwestii niż te, które zostały przedstawione, a które to wynikają z przepisów ustawy z dnia 20 lipca 2018 r. *Prawo o szkolnictwie wyższym i nauce*. Co więcej, Rada sugeruje, iż opinia, która odnosi się do innych elementów niż wskazane, może budzić wątpliwości odnośnie do jej prawidłowości.

Biorąc powyższe pod uwagę, ograniczę elementy oceniające tej recenzji do wymienionych.

Recenzję przygotowałem na podstawie rozprawy doktorskiej.

2 Ogólna charakterystyka i konstrukcja rozprawy

Treść rozprawy Mariusza Szarka pt.: *Budowanie mechanizmu obrony przed dedykowanymi kampaniami phishingowymi* jest wydana przez Wydział Cybernetyki Wojskowej Akademii Technicznej w Warszawie. Praca ma 274 strony.

Jak wskazuje tytuł, praca dotyczy zmory informatycznej wielu ostatnich lat, zmory, która zapewne jeszcze przez kolejne lata będzie męczyć użytkowników i administratorów. Phishing jest wykorzystywany w różnych celach, wszystkie jednak wyrządzają użytkownikowi zło. Zjawisko to ewoluuje i przybiera rozmaite formy, a za tą ewolucją podążają modele ataków phishingowych oraz próby budowy mechanizmów przeciwdziałających. W obecnej sytuacji cyberwojny zagadnienie to jawi się jako bardzo ważne, narzędzia ochrony zaś – niezmiernie potrzebne.

Doktorant w swej rozprawie prezentuje – jak pisze – nowe podejście do zagadnienia wykrywania i przeciwdziałania atakom phishingowym, głównie poprzez identyfikację większej niż dotąd liczby cech mogących być objawem ataku.

Pracę rozpoczyna spis skrótów oraz wybranych pojęć, po czym następuje krótki wstęp. Merytoryczną całość pracy tworzy pięć numerowanych rozdziałów. Po wspomnianych rozdziałach mamy podsumowanie, spis źródeł, rysunków i tabel oraz cztery dodatki.

Rozdział pierwszy wprowadza podstawowe pojęcia oraz przedstawia ogólne modele identyfikacji różnych rodzajów ataków phishingowych w cyberprzestrzeni.

W drugim rozdziale opisano znane z literatury metody identyfikacji ataku. Jest to rozdział analityczny, w którym różne rodzaje narzędzi są porównane z wskazaniem na słabe i silne strony wykrywania.

Trzeci obszerny rozdział (około 50 stron) poświęcony jest wskaźnikom ataku phishingowego. Wskaźniki są propozycją doktoranta, powstałą w wyniku analizy wiadomości phishingowych. Wskaźniki podzielono na dwie grupy: wskaźniki mogące być wykrywane przez oprogramowanie oraz wskaźniki wykrywane dzięki wiedzy eksperckiej.

W czwartym rozdziale jest przedstawiona nowatorska metoda detekcji phishingu w oparciu o zidentyfikowane wskaźniki. Metoda działa w oparciu o moduły uczenia maszynowego. Rozdział zawiera opis pozyskania i przygotowania danych w oparciu o analizę wiadomości email.

Weryfikację działania metody opisano w rozdziale piątym. Weryfikowaniu poddano dwie rzeczy: jakość klasyfikacji cech phishingowych w wiadomościach pocztowych oraz ocenę klasyfikacji samych wiadomości na trzy przyjęte grupy (spam, wiadomości normalne, wiadomości phishingowe).

Zwrócić muszę uwagę na pewien szczególny fakt, różniący ocenianą rozprawę od wielu innych. Mianowicie, autor przedstawia cele oraz tezę rozprawy nie na początku pracy, a dopiero na końcu rozdziału pierwszego, na dalekiej stronie 81. Ten niespotykany manewr oceniam jako zdecydowanie korzystny. W ten sposób, autor przedstawia tezę i cele rozprawy dopiero wówczas, gdy wprowadził już czytelnika nie tylko w podstawową terminologię, ale również w obszerny materiał związany z *state-of-the-art* badanych zagadnień. Sprawia to, że treść rozprawy przestaje być hermetyczną dla wąskiego kręgu zajmujących się taką tematyka badaczy, ale także staje się dostępna dla wielu czytelników będących na o wiele niższym od eksperta poziomie wiedzy na temat phishingu.

3 *Ocena teoretycznej wiedzy doktoranta w zakresie właściwej dziedziny wiedzy*

Na podstawie tak bardzo szczegółowo ukierunkowanej pracy jest niesłychanie trudno ocenić teoretyczną wiedzę autora. Posługiwanie się zwyczajowo przyjętymi kryteriami niewiele tutaj daje. Dlatego do standardowych kryteriów oceny dołączę jedno dodatkowe, a mianowicie kryterium braku błędów i pomyłek mogących świadczyć o niespełnieniu owych kryteriów standardowych.

I tak, autor nie popełnił żadnego błędu ani nawet niedopowiedzenia w posługiwaniu się podstawowymi koncepcjami informatyki, takimi jak: algorytmy, struktury danych, sieci komputerowe, systemy operacyjne czy inżynieria oprogramowania. Posługuje się tymi pojęciami w razie potrzeby.

Doktorant ma szeroką wiedzę o podstawowych obszarach badawczych dziedziny, w tym o: sztucznej inteligencji, sieciach neuronowych, bazach danych, cyberbezpieczeństwie czy interakcji człowiek-komputer. Zna postępy wiedzy w tych obszarach i zdaje sobie sprawę z ograniczeń.

Potrafi analizować i oceniać prace badawcze z zakresu informatyki, w tym identyfikować problem badawczy, określać metodologię, czytać wyniki i rozumieć wnioski. Potrafi też krytycznie ocenić badania, uwzględniając ich mocne i słabe strony.

Z całą pewnością solidnie rozumie zasady interakcji człowiek-komputer, w tym projektowanie interfejsu użytkownika, testowanie użyteczności i rolę czynnika ludzkiego, w szczególności wpływ tego ostatniego na bezpieczeństwo danych i systemu.

Jasno i skutecznie przekazuje złożone koncepcje z zakresu informatyki. Potrafić pisać raporty techniczne i prezentować wyniki badań w sposób jasny i zwięzły. Ta akurat cecha wydaje mi się ograniczająco wpłynęła na ostateczną postać pracy doktorskiej.

Również swobodnie doktorant porusza się po algorytmach i skryptach realizujących działanie tych algorytmów.

Mogę więc stwierdzić, że zakresie wiedzy właściwej dla informatyki technicznej i telekomunikacji, doktorant porusza się swobodnie.

W ostatnich latach w mechanizmach obrony poczty pojawił się protokół DMARC. Doktorant nie wspomina o tym protokole. Chciałbym, aby w trakcie obrony wyjaśnił, przyczyny tego: czy to była świadoma decyzja i czym motywowana.

Formalizm matematyczny wykorzystywany w pracy nie jest wysokich lotów, bowiem rozważane zagadnienia nie wymagają wyższej matematyki. Niemniej oczekiwałbym od przyszłego doktora sprawniejszej i bardziej eleganckiej edycji wzorów matematycznych, o czym piszę dalej.

4 *Ocena umiejętności samodzielnego prowadzenia pracy naukowej*

Kryteria oceny umiejętności różnią się zdecydowanie od kryteriów oceny wiedzy. Na pierwszym miejscu wymieniłbym dobrą znajomość metodologii badań, na co składa się zarówno znajomość różnorodnych metod badawczych, jak i umiejętność stosowania takich metod. Doktorant udowadnia dużą znajomość odpowiedniej metodologii, w końcu jeden z podstawowych problemów w jego rozprawie polega na rozróżnieniu wskaźników technicznych (identyfikowalnych automatycznie) i nie technicznych (wymagających wiedzy eksperckiej).

Kolejnym kryterium może być umiejętność krytycznego i analitycznego rozumowania. W znacznej mierze dotyczy to krytycznego czytania dostępnej literatury, dostrzegania luk w publikowanych rozwiązaniach, a także wyciągania wniosków z osiągniętych wyników badawczych. I tu także widzę dużą sprawność doktoranta. Analiza cytowanych źródeł jest dogłębna i krytyczna. Wnioski są formułowane sprawnie i czytelnie.

Jednym z ważnych elementów samodzielnego prowadzenia badań naukowych jest informowanie świata inżynierskiego i naukowego o swoich osiągnięciach i własnym wkładzie w rozwój dyscypliny. A tu pojawia się problem języka, w naszym przypadku – języka polskiego. Rozprawa jest napisana dość poprawnym językiem polskim. Mało jest w treści wyrażeń gwarowych. Zdania są dość rozbudowane, widać, że zdania te są przemyślane.

Niestety, w treści zauważam sporo błędów, głównie interpunkcyjnych, ale także stylistycznych i składniowych. W samym krótkim tekście streszczenia (polskiego) mamy już błędy. Bierze się to najczęściej z dziwnego przekonania wordopisarzy, że sam moment wydruku poprawia ewentualne błędy. Tak oczywiście nie jest; co więcej, sporo błędów ma to do siebie, że staję się widoczne dopiero w trakcie czytania wydrukowanych kartek. W przyszłej pracy badawczej jest to element do poprawienia.

Nie sposób nie wspomnieć o kontekście etycznym. W przypadku recenzowanej rozprawy jest to szczególni ważne, gdyż phishing można określić jako sposób kradzieży, co dotyka poszczególnych ludzi. Osobiste straty – w tym finansowe – mogą tu być znaczne. Autor doskonale rozumie ten kontekst i skutecznie próbuje przeciwdziałać kradzieżom.

Autor przytacza równo 100 pozycji źródłowych, z których każda jest cytowana w pracy. Jak na rozprawę doktorską, liczba źródeł jest skromna. Dodam jednocześnie, że źródła są raczej różnorodne. Mamy tu źródła pisane: artykuły, książki i doniesienia konferencyjne. Mamy też sporo źródeł sieciowe.

Jak już napisałem, liczba źródeł jest skromna i zastanawiałem się, czy wystarczająca. Zdecydowałem się na uznanie zbioru źródeł za wystarczający z powodu, o którym piszę dalej.

Jako wieloletni – teraz już dożywotni – członek IEEE, zawsze zwracam uwagę na

cytowanie publikacji IEEE. Ta organizacja wydaje około 70% literatury światowej rocznie (dotyczy języka angielskiego) i wydawałoby się, że podobny udział powinny mieć źródła IEEE w każdej większej bibliografii. Krótkie sprawdzenie głównego hasła recenzowanej rozprawy w przeglądarce Xplore daje kilkaset wskazań na artykuły oraz sporo ponad 1000 cytowań materiałów konferencji pod auspicjami IEEE. Kto nie cytuje z tak obszernego zbioru, nie może być uznany za dobrze poinformowanego. Jest jednak w spisie źródeł doktoranta parę materiałów opublikowanych przez IEEE.

Języczkiem u wagi stała się dla mnie obecność w spisie źródeł słynnego artykułu przeglądowego Khanji, co prawda nieco starego, bo z 2013 roku, ale mającego ponad 300 cytowań; w rozprawie jest to pozycja [55]. Ten artykuł zdecydował, że mogę uznać źródła za wystarczające.

Inną kwestią jest, że źródła wydają mi się starszawe. Przy tak aktualnej tematyce widziałbym więcej źródeł z datami publikacji bliższymi teraźniejszości.

Oddzielną sprawą jest wspomniana wcześniej forma prezentacji wzorów matematycznych. Trudno mi wymienić jakąś zasadę edycji równań i wzorów matematycznych, która tutaj została zachowana. Z przyjemnością w trakcie obrony zobaczyłbym slajdy z poprawnie zapisanymi wzorami; doktorant może to zrobić, korzystając – jako z wzorca – choćby z prac swojego promotora. Dodam tylko, że wzory są zwykle częścią zdania i podlegają takim samym zasadom interpunkcyjnym, jak zwyczajne obiekty języka polskiego.

5 *Ocena oryginalności rozprawy*

Phishing to jedna z najczęstszych, ale i najgroźniejszych form cyberataków, w której oszuści próbują wyłudzić poufne informacje, takie jak dane logowania, numery kart kredytowych czy inne wrażliwe dane, podszywając się pod zaufane instytucje lub osoby. Takie działania często mają charakter zmasowany, a ich celem jest nie pojedynczy obiekt, a szeroka grupa osób, obiektów o pewnych wspólnych cechach.

Ochrona przed phishingiem jest niezwykle ważna, gdyż atak prowadzić może m.in. do: zmniejszenia lub utraty bezpieczeństwa danych, utraty prywatności, znacznych szkód finansowych, utraty reputacji osób fizycznych bądź prawnych. Może też prowadzić do utraty integralności systemu informatycznego, co często jest pierwszym krokiem do późniejszego ataku na cyfrowe zasoby osoby lub firmy.

Z przestępczego punktu widzenia, phishing jest metodą bardzo atrakcyjną, ze względu na stosunkowo niski koszt przeprowadzenia takiego ataku. Wykradzione dane służą zwykle do właściwej kradzieży.

Złodzieje nie śpią. Każda nowa przeszkoda po jakimś czasie zostaje zwykle przełamana lub znajdowane jest dogodne obejście.

Między innymi z wymienionych powodów, ochrona przed atakami phishingowymi jawi się jako działanie pierwszej potrzeby. I tu pomysły doktoranta jawią się jako nowe, niespotykane do tej pory narzędzie przeciwdziałania atakom.

Doktorant skupił się na wiadomościach pocztowych. Po skrupulatnej analizie charakteryzujących phishing cech, zaproponował sprawdzanie cech nowych, do tej pory nie stosowanych. Cechy te podzielił na możliwe do narzędziowej identyfikacji oraz na możliwe do prawdopodobnej identyfikacji przez odpowiednio wyuczoną maszynę ekspercką. Jeśli powstanie komercyjne narzędzie na bazie pomysłu doktoranta, to może stać się istotnym, sprawnym i skutecznym.

Do najważniejszych i mających zarówno naukowy, jak i praktyczny aspekt osiągnięć zaliczam:

- definicję nowych cech wektora identyfikacji, co znacząco wydłuża ów wektor, w oczywisty sposób wpływając na poprawne wykrycie ataku phishingowego,
- opracowanie algorytmu odczytującego nowe cechy z nagłówków wiadomości;
- zaprojektowanie modułów uczenia maszynowego bazującego na metodach heurystycznych analizy odnośnika URL, z niezależnym badaniem domeny zawartej w tym odnośniku.

Biorąc to wszystko pod uwagę, nie mam wątpliwości, że nowatorska zawartość pracy jest duża i potencjalnie nadaje się do komercjalizacji.

PODSUMOWANIE

Reasumując – uważam, iż recenzowana rozprawa doktorska jest ciekawą i oryginalną pracą badawczą. Przeprowadzając swoje wywody, autor wykazał się dobrą znajomością wiedzy teoretycznej i praktycznej w reprezentowanej przez niego dyscyplinie, a także umiejętnością samodzielnego prowadzenia pracy naukowej. Doktorant osiągnął założone w pracy cele i wykazał dużą przydatność prezentowanych metod.

Jestem przekonany, że praca spełnia wymogi stawiane rozprawom doktorskim w obowiązujących przepisach i wnoszę o dopuszczenie pracy do obrony.

W trakcie postępowania **będę głosował za przyznaniem Mariuszowi Szarkowi stopnia doktora nauk inżynieryjno-technicznych w dyscyplinie informatyka techniczna i telekomunikacja.**