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

Bydgoszcz, 16.08.2021

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

Tytuł rozprawy: **Wieloskładnikowe i transparentne uwierzytelnianie użytkownika z wykorzystaniem technik uczenia maszynowego,**

Autor rozprawy: **mgr inż. Paweł Łąka**

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.)?**

Rozprawa, której Autorem jest Pan mgr inż. Paweł Łąka, dotyczy wieloskładnikowego uwierzytelniania użytkowników, które u swych podstaw wykorzystuje niezależne metody potwierdzania tożsamości. Zrealizowane jest to na podstawie tego co użytkownik wie, posiadania, czy też na podstawie cech biometrycznych oraz behawioralnych.

W swych badaniach, Pan mgr inż. Paweł Łąka, duży nacisk kładzie na wygodę użytkownika. Jest to uzasadnione na podstawie analizy literaturowej, która wykazała, że aktualnie brakuje kompleksowego rozwiązania biorącego pod uwagę zarówno bezpieczeństwo samego procesu uwierzytelniania jaki i wygodę użytkownika końcowego.

Proponowane przez Autora pracy podejście do uwierzytelniania użytkownika nie wymaga od niego żadnych wymuszonych (dodatkowych) czynności, co sprawia iż cały proces uwierzytelniania jest dla niego mniej uciążliwy. W tym świetle jasno sformułowana została teza rozprawy.

Jeśli chodzi o charakter rozprawy to jest on w mojej ocenie koncepcyjno-doświadczalny. Autor pracy stworzył prototyp proponowanego systemu oraz zebrał dane niezbędne do przeprowadzenia badań. Ponadto, zaplanował i wykonał różne eksperymenty, które pozwoliły ocenić proponowane rozwiązanie pod względem skuteczności i niezawodności.

Dodatkowym elementem, jaki można było wykonać w ramach pracy nad rozprawą, to ankieta oceny transparentności (czy uciążliwości) proponowanych przez Autora modalności (metod). Można było to wykonać niskim nakładem pracy w trakcie tworzenia profili użytkowników.

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

Niniejsza rozprawa obejmuje najnowsze osiągnięcia nauki. Autor przedstawił bardzo rozległy rozdział omawiający wcześniejsze dokonania oraz aktualną literaturę dotyczącą zagadnień związanych z rozprawą doktorską. W szczególności dokonany został przegląd różnych metod uwierzytelniania z uwzględnieniem metod wieloskładnikowych.

Autor rozprawy, przedstawił również stan badań dotyczący metod uznawanych potencjalnie za wygodne i bezinwazyjne. Również przegląd literatury dotyczący wykorzystania uczenia maszynowego w procesach logowania jest przeprowadzony i przedstawiony w sposób właściwy.

Wnioski i spostrzeżenia z przeglądu źródeł przedstawiono bardzo szczegółowo i jasno. Pozwoliło to dobrze podkreślić sposób w jaki niniejsza rozprawa wyróżnia się na tle innych rozwiązań w literaturze.

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 postawiony problem. W procesie tym użyte zostały właściwe narzędzia i metody. Tym samym Autor dowiódł, że posiadł umiejętności związane z metodyką i metodologią prowadzenia badań.

W szczególności badania i eksperymenty osadzone zostały na solidnej i głębokiej analizie literaturowej. Autor rozprawy odniósł się także do aktualnych trendów rozwojowych dotyczących uwierzytelniania i kontroli dostępu. W tym świetle zaprojektowany i zaimplementowany został prototyp proponowanego rozwiązania.

Ponadto, zaplanowane i przeprowadzone eksperymenty są wiarygodnym dowodem zrealizowania przez Autora celu pracy. Dodatkowo, udokumentowane w rozprawie wyniki świadczą o efektywności proponowanego rozwiązania w analizowanych scenariuszach badawczych.

Dodatkowym elementem, jaki można było w ramach prac na rozprawą zrealizować, to interpretacja opracowanych modeli z użyciem istniejących narzędzi taki jak ELI5, LIME, czy SHAP. Pozwoliłoby to szerzej ocenić wpływ poszczególnych modalności (metod uwierzytelniania) na finalne odpowiedzi poszczególnych klasyfikatorów.

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 reprezentowanych przez literaturę światową?**

W raz ze postępującą informatyzacją wielu procesów obecnych w naszym codziennym życiu, coraz częściej spotykamy się z różnymi mechanizmami kontroli dostępu i autoryzacji. Oprócz docelowego poziomu bezpieczeństwa jakie nowe mechanizmy powinny gwarantować, ważnym elementem jest także zadowolenie użytkownika, gdyż w wielu przypadkach dodatkowe elementu uwierzytelniania mogą stać się bardzo uciążliwe. Dlatego w mojej ocenie Autor pracy dobrze wpasował się w aktualne trendy rozwojowe, co sprawia, że tematyka pracy jest aktualna i ważna.

Ponadto, aktualnie trudno jest zapewnić połączenie wysokiego zadowolenia (czy tez wygody) użytkownika z rygorystycznymi standardami bezpieczeństwa. W mojej ocenie w tym obszarze pozycjonuje się główny dorobek doktoranta. W szczególności Autor pracy proponuje nowatorskie podejście do procesu uzyskiwania dostępu w oparciu o uwierzytelnianie wieloskładnikowe i uczenie maszynowe. Jednoczenie w całym tym procesie brany jest pod uwagę i analizowany poziom wygody oraz zadowolenia potencjalnego użytkownika.

W trakcie rozprawy powstał też prototyp systemu potwierdzania tożsamości. Ponadto, doktorant zaproponował też własną biometryczną metodę weryfikacji tożsamości użytkownika w oparciu o tak zwany skład ciała.

W tym świetle rozprawa ma znaczenie dla nauki i techniki, gdyż Autor eksperymentalnie udowodnił, że można stworzyć kompleksowy system uwierzytelniania oparty o algorytmy uczenia maszynowego oraz transparentne dla użytkownika metody potwierdzania tożsamości. Ponadto, otrzymane wyniki, jak zauważył Autor rozprawy, stanowią podstawę do dalszego rozwijania systemu w kierunku jego komercjalizacji.

5. **Czy Autor wykazał umiejętności poprawnego i przekonywującego 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. Doktorant wykazał się przede wszystkim umiejętnościami dotyczącymi prowadzenia badań naukowych, eksperymentów, projektowania systemów informatycznych oraz implementacji i programowania.

Również zrealizowane przez Autora rozprawy eksperymenty nie odbiegają od obowiązujących dobrych praktyk w zakresie pomiaru skuteczności systemów weryfikacji. Autor w tym zakresie użył znanych metryk co pozwala odnieść się do innych rozwiązań dostępnych w literaturze.

Ponadto, eksperymenty zostały dodatkowo rozszerzone o scenariusze, które pozwalają ocenić skuteczność proponowanego rozwiązania w momencie gdy wybrane funkcjonalności systemu zostają celowo zaatakowane.

6. **Jakie są wady i słabe strony rozprawy?**

Rolą recenzenta jest zauważenie i wskazanie pewnych niedociągnięć (czy mankamentów) ocenianej pracy, oraz zgłoszenie tych uwag, które mogą być potencjalnie przydatne w dalszych pracach badawczo-rozwojowych. Dlatego też do wad i słabych stron rozprawy zaliczam:

- Prezentacja graficzna wyników w rozdziale 7. W mojej ocenie wykresy zaprezentowane w rozdziale 7 powinny wykorzystywać osobne skale dla błędu FRR i FAR (np. lewa i prawa oś). Pozwoliłoby to lepiej ocenić różnice w wartościach, gdyż w aktualnej formie przebiegi FRR są „płaskie" i ściśnięte do wartości zerowych. Ponadto, prezentowanie wykresów dla zerowych wartości błędów (np. „pusty" wykres 7.10) nie ma większego sensu. Choć przykuwa to uwagę czytającego i podkreśla dobrą efektywność proponowanego systemu to w mojej ocenie taki zabieg jest niepotrzebny. Ponadto, umieszczenie punktowych wartości błędów na wykresach słupkowych również polepszyłoby czytelność prezentowanych wyników.
- Symulacja wystąpienia pewnych zdarzeń celem oceny ich wpływu na odpowiedź systemu IdP. W mojej ocenie, zastosowanie takich narzędzi jak SHAP czy LIME pozwoliłoby głębiej zrozumieć jaki jest wpływ poszczególnych serwisów (metod) identyfikacji w wypracowywaniu finalnej odpowiedzi przez poszczególne algorytmy uczenia maszynowego. Idąc dalej, Autor mógł, przy tak małej liczbie cech (6 metod identyfikacji) i danych uczących, wytrenować modele dla wszystkich kombinacji zaistniałych sytuacji (czyli brak odpowiedzi z poszczególnych serwisów). Ponadto, jeżeli mowa o systemach autoryzacji użytkownika, to (w mojej ocenie) system nie

powinien sam „radzić" sobie w momencie gdy brakuje odpowiedzi (czy też informacji) dla poszczególnych serwisów (metod).

- Odrzucenie w rozdziale 7.2.1 celowości analizy współczynnika FAR. Zasadniczo powód dlaczego FAR nie jest analizowany został dobrze uzasadniony przez Autora, ale istnieje pewne ryzyko, że atakujący poprzez modyfikacje odpowiedzi jednej z metod (przykładowo poprzez blokowanie sygnału Bluetooth) będzie próbował podnieść poziom swoich uprawnień poprzez podszycie się pod innego użytkownika.
- Celowość demonstracji i analizy wykresów 7.1-7.6. Początek rozdziału 7 pozwala głębiej zrozumieć rozkład wartości dla poszczególnych cech, jednak jego forma w mojej ocenie jest dość nietypowa. Dla czytającego prawdopodobnie bardziej zrozumiała mogłaby być typowa analiza EDA (ang. Exploratory Data Analysis) wykonana dla poszczególnych cech (metod identyfikacji użytkownika). Jedną z popularnych bibliotek wspierających ten proces analizy dla języka Python jest Sweetviz.
- Celowość prezentowania rozdziału 5.6.3 a w szczególności detali związanych ze sposobem działania algorytmów uczenia maszynowego. W mojej ocenie opisy te mają charakter bardzo ogólny i są zbędne, gdyż nie wnoszą wiele do lepszego zrozumienia rozprawy przez czytającego.
- Pewne niedomówienia jeżeli chodzi o fizyczną realizacje całego systemu. Przykładowo w rozdziale 5.5 mowa jest o trzymaniu kierownicy przy badaniu składu ciała. Natomiast w rozdziale 6.1 mowa jest o staniu na urządzeniu. Rodzi to pytanie czy możliwym jest badanie składu ciała poprzez trzymanie urządzenia (nawet odpowiednio przystosowanego czy zmodyfikowanego) i jak wiarygodny będzie taki pomiar.
- Patrząc na tabele 5.6 a w szczególności na czasy akwizycji z poszczególnych modułów (serwisów) można odnieść wrażenie, że niektóre z pomiarów są znacznie rozciągnięte w czasie. Rodzi to pytanie w kontekście wygody użytkownika w momencie gdy proces akwizycji danych niezbędnych do uwierzytelnienia zajmuje ok. 18 sekund.
- Język w jakim napisano rozprawę. Pewne fragmenty niniejszej rozprawy są napisane językiem dość nieformalnym z pewnymi usterkami językowymi:
    - „Mierzenie systemu" czy „algorytmy uczące" – prawdopodobniej lepiej byłoby napisać „mierzenie skuteczności systemu" oraz „algorytmy uczenia maszynowego".
    - Również stwierdzenie, że system „samodzielnie radzi sobie" jest dość potoczne i nieprecyzyjne.
    - Powstały system „nie jest progowy" (str. 42) – zwrot ten jest również nieprecyzyjny. Ponadto, jakby się przyjrzeć dokładniej klasycznym metodom uczenia maszynowego (np. drzewo decyzyjne) to metody te w

trakcie uczenia ustalają jednak pewien próg decyzyjny na wybranym atrybucie.

- W niektórych rozdziałach również można wskazać pewne ogólnikowe zwroty nacechowane emocjonalnie (np. „działają całkiem dobrze" (str.95), „trudne do przewidzenia" (str.95), „są łatwe w szkoleniu" (str.95))
- Za pewne niedopatrzenie edytorskie można też uznać odsyłacz do nieistniejącego rozdziału 0 (str. 105)

7. Jaka jest przydatność rozprawy dla nauk technicznych?

Pomimo przedstawionych powyżej uwag, rozprawa mgra inż. Pawła Łąki posiada wiele mocnych stron.

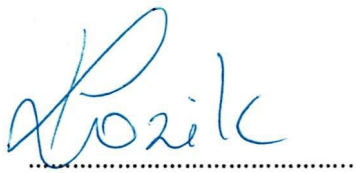
Przede wszystkim zawiera ona oryginalną autorską propozycję systemu uwierzytelniania użytkownika. Rozprawa dotyczy ważnego zagadnienia, jest aktualna i wnosi interesujący wkład do biometrii, systemów rozpoznawania i zagadnień związanych z uwierzytelnianiem wieloskładnikowym.

Pomimo wymienionych usterek praca jest ciekawie napisana, czyta się ją bardzo dobrze i przyjemnie. Niewątpliwie recenzowana praca dowodzi dużej wiedzy Autora dot. zagadnień uwierzytelniania wieloskładnikowego.

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

Wniosek

Biorąc pod uwagę przedstawioną przez Doktoranta rozprawę stwierdzam, że spełnia **ona wymagania stawiane rozprawom doktorskim** przez obowiązującą Ustawę o stopniach i tytule naukowym, i wnioskuję o **dopuszczenie** jej do publicznej obrony.



.....

Podpis