Prof. dr hab. inż. Robert Burduk
Politechnika Wrocławska
Wydział Informatyki i Telekomunikacji
Ul. Wybrzeże Stanisława Wyspiańskiego 27
50-370 Wrocław

Wrocław, dnia 15.02.2024 r.

RECENZJA

rozprawy doktorskiej mgra inż. Tomasza Wesołowskiego
zatytułowanej: **„Biometryczna weryfikacja użytkownika systemu komputerowego z automatyczną aktualizacją profilu aktywności"**

Recenzja została sporządzona w związku z powołaniem przez Radę Naukową Instytutu Informatyki Uniwersytetu Śląskiego w Katowicach w dniu 13.12.2023 r., piszącego niniejszą recenzję, jako recenzenta rozprawy doktorskiej mgra inż. Tomasza Wesołowskiego.

Kryteria oceny dysertacji wynikają z przepisów zawartych w art. 187 ustęp 1 oraz 2 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce.

Problem badawczy i jego znaczenie

Zakres recenzowanej rozprawy dotyczy wykorzystania biometrii behawioralnej w procesie weryfikacji osób korzystających z systemu komputerowego. W szczególności Doktorant koncentruje się na zagadnieniu analizy dynamiki pisania na klawiaturze. Dane gromadzone podczas korzystania z klawiatury są wykorzystywane do zdefiniowania profilu użytkownika, który jest elementem składowym systemu weryfikacji behawioralnej użytkownika systemu komputerowego. Problem badawczy recenzowanej rozprawy dotyczy zdefiniowania poszczególnych części składowych systemu weryfikacji behawioralnej, opracowania metod analizy oraz klasyfikacji danych pochodzących ze strumienia tekstu pisanego na klawiaturze wraz z badaniami porównawczymi odnoszącymi się do wyników innych metod referencyjnych. Opracowane przez mgra inż. Tomasza Wesołowskiego metody weryfikacji behawioralnej użytkownika mogą działać w trybie na bieżąco (*on-line*) jak i w trybie po zdarzeniu (*off-line*).

W rozprawie sformułowano hipotezę badawczą oraz pięć celów szczegółowych. Hipoteza badawcza zakłada, że opracowany system ciągłej weryfikacji behawioralnej wykorzystujący automatyczną aktualizację profilu użytkownika uzyska lepszą skuteczność identyfikacji intruza niż wybrane metody referencyjne. Postawione cele szczegółowe dotyczą gromadzenia danych rzeczywistych zawierających aktywność użytkownika podczas korzystania z klawiatury, opracowania algorytmów będących elementami behawioralnego systemu weryfikacji użytkownika, wykorzystaniu metod klasyfikacji nadzorowanej w procesie weryfikacji użytkownika wraz z testowaniem opracowanego przez Doktoranta systemu. Eksperymentalna weryfikacja postawionej hipotezy badawczej została wykonana z wykorzystaniem bazy danych zgromadzonej podczas pracy nad dysertacją oraz dwóch publicznie dostępnych zbiorów danych (*Clarkson II, Bufallo*).

Lektura dysertacji pozwala stwierdzić, że została potwierdzona teza badawcza oraz zostały osiągnięte wszystkie postawione cele szczegółowe.

Tematyka rozprawy jest interesująca, w pełni uzasadniona i wpisuje się w nurt aktualnych zagadnień cyberbezpieczeństwa, w szczególności dotyczy nieautoryzowanego dostępu lub korzystania z systemu komputerowego. Przedstawione i opracowane przez Doktoranta autorskie metody biometrycznej weryfikacji użytkownika systemu komputerowego mają duży potencjał wdrożeniowy, a zagadnienia poruszane w dysertacji mieszczą się w dziedzinie Nauk Ścisłych i Przyrodniczych w dyscyplinie Informatyka.

Struktura, treść pracy oraz wiedza Autora

Recenzowana praca została napisana w języku polskim i liczy 124 strony maszynopisu. Składa się z czterech rozdziałów merytorycznych, wstępu, podsumowania, bibliografii, spisu rysunków oraz tabel, wykazu skrótów oraz oznaczeń stosowanych w rozprawie. Motywacja dotycząca podjęcia tematyki badawczej, teza badawcza, cele szczegółowe oraz struktura dysertacji zostały przedstawione w rozdziale wstępnym. Rozdział nr 2 przedstawia zagadnienia dotyczące biometrii behawioralnej, weryfikacji oraz identyfikacji użytkownika systemu komputerowego, metryk stosowanych do oceny systemów weryfikacji oraz systemów klasyfikacyjnych. Przedstawiona w rozdziale nr 2 treść wskazuje, że Autor rozprawy posiada ogólną wiedzę teoretyczną, która dotyczy problematyki cyberbezpieczeństwa, a w szczególności biometrycznej weryfikacji użytkownika systemu komputerowego wraz z oceną systemu do weryfikacji. Treść rozdziału nr 2 mieści się w zakresie dyscypliny

naukowej Informatyka, a jego lektura pozwala stwierdzić, że mgr inż. Tomasz Wesołowski posiada ogólną wiedzę teoretyczną z zakresu wymienionej dyscypliny.

W rozdziale 3 Autor przedstawił koncepcję dwóch typów profilu użytkowania systemu komputerowego. Profil lokalny reprezentuje bieżącą aktywność użytkownika, natomiast profil rozmyty reprezentuje aktywność użytkownika w szerszym horyzoncie czasowym. Oba profile wykorzystują dane jakimi są polecenia systemowe wpisywane za pomocą klawiatury. Rozdział 3 zawiera również wyniki badań eksperymentalnych, które miały na celu wyznaczenie optymalnych parametrów dla zaproponowanego przez mgra inż. Tomasza Wesołowskiego profilu rozmytego oraz porównanie zaproponowanej metody wykrywania intruzów z innymi metodami znanymi z literatury. Porównania dokonano z wykorzystaniem publicznie dostępnego zbioru danych SEA. Należy zaznaczyć, że wyniki przedstawione w rozdziale 5 odnoszą się do jednego z celów szczegółowych rozprawy. Jednak jak wskazuje Autor w zakończeniu rozdziału, analiza ilościowa poleceń systemowych wpisywanych na klawiaturze nie stanowi podstawy do budowy systemu weryfikacji behawioralnej użytkownika systemu komputerowego. Zastosowanie natomiast analizy uwzględniającej zależności czasowe dotyczące wpisywania znaków na klawiaturze pozwala na budowę systemu weryfikacji behawioralnej użytkownika.

Rozdział 4 zawiera opis między innymi autorskiej metody profilowania użytkowników, która wykorzystuje dynamikę pisania na klawiaturze. Opisana metoda bazuje na zależnościach czasowych którymi są czas trzymania klawisza oraz czas wciśnięcie-wciśniecie. W rozdziale 4 przestawiono również algorytm określający cechy użytkownika, które są cechami biometrycznymi wynikającymi ze sposobu pisania na klawiaturze. Zidentyfikowane cechy posłużyły Doktorantowi do zdefiniowania biometrycznego profilu użytkownika.

Rozdział 5 przedstawia autorskie propozycje części składowych sytemu uwierzytelniania behawioralnego, które dotyczą problemu klasyfikacji. W szczególności Autor opisał sposób wyznaczania optymalnych wartości parametrów zaprojektowanego systemu, którymi są: maksymalna pojemność kontenera, minimalna liczba następujących po sobie zdarzeń, maksymalny dozwolony czas między dwoma zdarzeniami oraz liczba wektorów cech wchodzących w skład profilu użytkownika. Optymalna wartość wymienionych parametrów ustalana jest przy wykorzystaniu algorytmu wykorzystującego pojęcie odległości, który został zaproponowany przez Doktoranta. Problem klasyfikacyjny rozważany jest przez mgra inż. Tomasza Wesołowskiego w trzech koncepcjach. Punkt 5.2 przedstawia wyniki eksperymentalne uzyskane przy wykorzystaniu wybranych pojedynczych (bazowych) klasyfikatorów. W punkcie 5.3 Autor dysertacji zaproponował strukturę stanowiącą zespół

równoległych klasyfikatorów heterogenicznych oraz przedstawił wyniki odpowiednich badań eksperymentalnych. Punkt 5.4 przedstawia zagadnienie klasyfikacji, w którym z zespołu klasyfikatorów wybierany jest jeden klasyfikator bazowy o najwyższej wartości kompetencji. Postać funkcji kompetencji wykorzystanej w badaniach eksperymentalnych została zaproponowana przez Autora dysertacji. Punkt 5.5 przedstawia autorską strategię aktualizacji profilu użytkownika, która umożliwia dostosowanie się opracowanego przez Doktoranta modułu sytemu uwierzytelniania behawioralnego do zmieniającej się w czasie charakterystyki pisania tekstu na klawiaturze przez użytkownika. Kolejne punkty rozdziału 5 przedstawiają wyniki badań eksperymentalnych wykonanych przy wykorzystaniu danych zgromadzonych podczas pracy nad dysertacją (baza *realKDD*) oraz dwóch publicznie dostępnych zbiorów danych (*Clarkson II, Bufallo*) wraz z interpretacją otrzymanych wyników.

Spis literatury liczy 197 pozycji. Cytowane prace dobrane są prawidłowo i odnoszą się do omawianych problemów. Praca napisana jest bardzo starannie pod względem językowym oraz edycyjnym. W dysertacji znajdują się jedynie niewielkie uchybienia redakcyjne (np. rys. 5.15 – opis osi w języku angielskim, str. 101 – „[99, 100]Niestety", czy też kolejność cytowań „[19, 121, 138, 94]").

Wkład Autora — oryginalne osiągnięcia

Wkład Autora w rozwój dyscypliny Informatyka polega na:

1. opracowaniu profilu rozmytego użytkownika systemu komputerowego wraz z algorytmem jego wzmocnienia bazującego na poleceniach systemowych wpisywanych z klawiatury,
2. opracowaniu profilu użytkownika systemu komputerowego wykorzystującego dynamikę pisania na klawiaturze,
3. wyznaczeniu optymalnych wartości parametrów zaprojektowanego systemu uwierzytelniania behawioralnego z wykorzystaniem opracowanego algorytmu klasyfikacji minimalnoodległościowej,
4. opracowaniu zespołu klasyfikatorów o strukturze równoległej złożonej z heterogenicznych klasyfikatorów bazowych w celu wyznaczenia etykiety klasy: użytkownik uprawniony lub intruz,
5. opracowaniu algorytmu selekcji klasyfikatora bazowego o największej wartości kompetencji wykorzystywanego do wyznaczenia etykiety klasy: użytkownik uprawniony lub intruz,

6. opracowaniu modułu sytemu uwierzytelniania behawioralnego dostosowanego do zmieniającej się w czasie charakterystyki pisania tekstu przez użytkownika na klawiaturze,
7. eksperymentalnej weryfikacji opracowanej metody wykorzystującej profil rozmyty wraz z porównaniem jej skuteczności z metodami referencyjnymi,
8. eksperymentalnej weryfikacji opracowanej metody uwierzytelniania behawioralnego wykorzystującej charakterystykę pisania tekstu przez użytkownika na klawiaturze wraz z porównaniem jej skuteczności z metodami referencyjnymi.

Wymienione osiągnięcia świadczą o umiejętności samodzielnego prowadzenia pracy naukowej przez mgra inż. Tomasza Wesołowskiego oraz o oryginalnym rozwiązaniu problemu naukowego jakim jest behawioralna weryfikacja użytkownika systemu komputerowego mieszcząca się w zakresie cyberbezpieczeństwa.

Uwagi krytyczne

Optymalne wartości parametrów opracowanej metody biometrycznego profilowania użytkowników zostały wyznaczone eksperymentalnie (rozdział 5.1). W jaki sposób zostały wyznaczone optymalne wartości hiperparametrów algorytmów wykorzystywanych w komitetach klasyfikatorów? Wyniki przedstawione w tabeli 5.2 uzasadniają wybór czterech klasyfikatorów wykorzystywanych w komitetach ale bez odniesienia się do możliwych wartości ich hiperparametrów.

Zaproponowany i przedstawiony w rozdziale 5.3 moduł klasyfikacji komitetowej w ostatniej warstwie wykorzystuje metodę głosowania większościowego. Warstwa ta jest nadmiarowa, ponieważ w warstwie klasyfikatorów bazowych wyznaczane są wartości wsparcia dla każdej etykiety klasy i każdego równoległego bloku danych uczących zgodnie z regułą *arg max*. Rozszerzenie tej reguły na wszystkie równoległe bloki danych pozwala na zrezygnowanie z ostatniej logicznej warstwy opracowanego modułu klasyfikacji komitetowej. Opisując zaproponowany moduł Doktorant odnosi się tylko do trzech równoległych bloków danych uczących (np. wzory 5.7 oraz 5.9). W tabelach 5.3 i 5.11 oraz na rysunku 5.8 przedstawione są również wyniki dla innych liczb równoległych bloków danych. Opis w rozdziale 5.3 powinien być uniwersalny, tzn. odnosić się do dowolnej liczby równoległych bloków danych uczących (dowolnej liczby podziału pierwotnego zbioru uczącego).

Zaproponowana przez Doktoranta funkcja wyznaczająca kompetencję (wzór 5.17) wykorzystuje funkcję ekspotencjalną. Jakie było kryterium, które zostało wykorzystane w celu wyboru tej postaci funkcji?

Doktorant analizując wykres pudełkowy (rysunek 5.8) stwierdza o istnieniu różnic statystycznych między poszczególnymi wariantami klasyfikatora zespołowego. Aby jednoznacznie stwierdzić o istnieniu różnicy statystycznej powinny zostać przedstawione wyniki odpowiedniego testu statystycznego.

W pracy nie zamieszczono komentarza dotyczącego wyników przedstawionych na rysunku 5.12. Czy otrzymana liczba alertów przedstawia informację o skuteczności zaproponowanej strategii automatycznej aktualizacji profilu użytkownika?

Podsumowanie

Reasumując stwierdzam, iż mgr inż. Tomasz Wesołowski posiada ogólną wiedzę teoretyczną z dyscypliny Informatyka. W szczególności wiedza Doktoranta dotyczy biometrii behawioralnej oraz metod weryfikacji oraz identyfikacji użytkownika systemu komputerowego. Recenzowana praca zawiera sformułowaną tezę badawczą, która została udowodniona doświadczalnie przy wykorzystaniu danych dostępnych w publicznych repozytoriach oraz autorskiego zbioru danych. Lektura dysertacji pozwala stwierdzić, że Autor zaprezentował na jej łamach umiejętność samodzielnego prowadzenia pracy naukowej. Recenzowana rozprawa stanowi oryginalne rozwiązanie problemu naukowego jakim jest opracowanie oraz testowanie systemu weryfikującego użytkownika komputera. Opracowany oraz testowany przez mgra inż. Tomasza Wesołowskiego system wykorzystuje metodę biometrii behawioralnej jaką jest analiza dynamiki pisania na klawiaturze.

Wobec powyższego, recenzowana praca spełnia wymagania zdefiniowane w art. 187 ustęp 1 oraz 2 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce. Konkludując, wnoszę o przyjęcie rozprawy oraz dopuszczenie mgra inż. Tomasza Wesołowskiego do publicznej obrony.

Dodatkowo biorąc pod uwagę aktualność tematyki badawczej mieszczącej się w zakresie cyberbezpieczeństwa, szeroki zakres zaproponowanych i opracowanych metod składających się na cały system behawioralnej weryfikacji użytkownika komputera oraz jakość pracy badawczej, której wyrazem jest między innymi publikacja w czasopiśmie *Applied Soft Computing* składam wniosek o wyróżnienie rozprawy doktorskiej mgra inż. Tomasza Wesołowskiego.

R. Burduk