prof. dr hab. inż. Andrzej Pacut
Instytut Automatyki i Informatyki Stosowanej
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

2 lutego 2018

Wstępna recenzja rozprawy doktorskiej

mgr inż. Wojciecha Wodo

Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji

dla Rady Naukowej Instytutu Podstaw Informatyki
Polskiej Akademii Nauk

Niniejsza recenzja została napisana na prośbę Rady Naukowej Instytutu Podstaw Informatyki Polskiej Akademii Nauk i dotyczy pracy doktorskiej w dziedzinie nauk technicznych w dyscyplinie informatyka, przedstawionej przez p. mgr. inż. Wojciecha Wodo.

Praca składa się z 5 rozdziałów i liczy 86 stron. We rozdziale wstępnym określono tezy pracy. Rozdział 2 (15 s.) poświęcony jest kilku różnym zagadnieniom ochrony danych biometrycznych. Zawiera on: metody personalizacji wykorzystujące funkcje haszujące i drzewa Merkla i schemat ochrony bazy danych biometrycznych w oparciu o (t, n) *threshold subset problem* którego definicja, rozwiązanie z wykorzystaniem filtrów Blooma jest dziełem doktoranta. Rozdział 3 (30 s.) dotyczy rozpoznawania na podstawie rytmu pisania na klawiaturze (*keystroking*) w aspekcie weryfikacji i identyfikacji i ukrywania oryginalnych danych biometrycznych. Jest to najbardziej rozbudowana część pracy, zawierająca zarówno nowe rozwiązania jak i projekt i implementację urządzeń. W r. 4 (12 s.) zawarte są rozważania na temat weryfikacji na podstawie obrazu twarzy. Pracę kończy podsumowanie i wykaz literatury.

1. **Zagadnienie naukowe rozpatrywane w pracy (teza rozprawy)**: jasność sformułowania, charakter rozprawy (teoretyczny, doświadczalny, inny)?

   Zagadnienia rozpatrywane w pracy znajdują się na przecięciu, ogólnie, nie-biometrycznych metod ochrony danych i biometrii. Teza pracy sformułowana jest w kontekście czterech scenariuszy ochrony danych biometrycznych, a mianowicie (kolejność doktoranta, etykiety moje)

   (a) zagadnień niezamierzonego przepływu danych biometrycznych,

   (b) ujawniania uwierzytelnionych danych biometrycznych,

   (c) ochrony biometrycznych baz danych i

   (d) ochrony upublicznionych danych biometrycznych.przed ponownym nieuprawnionym użyciem przy zastosowaniu schematów steganograficznych

   Praca nie została jednak przedstawiona w kolejności zgodnej z powyższymi scenariuszami. Rozdziały pracy pogrupowano raczej wg zastosowanych metod. I tak, zagadnienia scenariusza a) omawiane sa w r. 3, zagadnienia scenariuszy b) w r. 2 i (częściowo) r.

4, natomiast zagadnienia scenariuszy c) i d) omawiane są w r. 2. Struktura taka komplikuje sposób i śledzenie rozwiązań ochrony danych biometrycznych w omawianych scenariuszach.

Zapowiedź realizacji zagadnień wzmiankowanych w scenariuszach jest mało konkretna. Dla pierwszego scenariusza „przedstawiono koncepcję techniczną". W zakresie drugiego zagadnienia "pokazujemy, że ochrona jest możliwa", w kolejnym „wykazujemy na przykładzie (...) odcisków palców że możliwy jest wysoki stopień ochrony bazy i uwierzytelnienia", a wreszcie dla ostatniego zagadnienia „wskazujemy, że odpowiednie znakowanie danych w steganograficzny sposób (...) może pozwolić" na wyodrębnienie przypadków bezprawnego wykorzystania. Jak formułuje to doktorant, dla każdego z tych scenariuszy „wskazujemy dedykowane metody ochrony", co nie jest raczej wystarczającym wynikiem badań. W trakcie lektury rozprawy można się oczywiście zorientować, że praca zawiera nie tylko „wskazanie metod", ale że metody te zostały zaprojektowane i zbadane przez doktoranta.

Rozprawa ma szerokie spektrum metodologiczne, gdyż obejmuje zarówno teorię, jak i obliczenia numeryczne, a również (w jednym - ważnym - przypadku) konstrukcję elektronicznego urządzenia realizującego zadane cele. Pokazuje to rozległość zainteresowań, wiedzy i kompetencji doktoranta.

2. **Analiza źródeł**

Bibliografia pracy, zawiera 84 pozycje. Większość z nich dotyczy zagadnień ochrony kryptograficznej. Znacznie uboższe są cytowania literatury dotyczące dziedzin biometrycznych, np. nowych wyników biometrii twarzy, będącej tematem r.4. Doktorant nie przedstawił niestety odrębnej listy własnych publikacji. Sądząc z załączonej bibliografii, jest on współautorem co najmniej 3 publikacji, w tym jednej w czasopiśmie z bazy JCR.

3. **Oryginalność i przydatność rozprawy, jej pozycja w stosunku do stanu wiedzy** (oryginalność rozprawy, samodzielny i oryginalny dorobek autora, pozycja rozprawy w stosunku do stanu wiedzy/poziomu techniki)

Rozpoznawanie biometryczne rozwija się od kilkunastu lat i dawno już przeszło od teorii do fazy implementacji rynkowych. Dość szybko jednak stwierdzono, że samodzielne stosowanie biometrii może wprowadzać niebezpieczeństwo, zamiast zwiększać bezpieczeństwo uwierzytelniania. Do sytuacji takich należy np. tzw. atak prezentacji nie omawiany w pracy, jak również szereg innych scenariuszy będących tematem pracy. Niebezpieczeństwa te w dużej mierze mogą zostać zażegnane poprzez zastosowanie specyficznych schematów kryptograficznych. Opracowanie i zastosowanie tych schematów stanowi istotne i ważne dokonanie doktoranta. Niektóre czynniki bezpieczeństwa mogą być wprowadzane dla specyficznych modalności biometrycznych. Drugim istotnym zakresem osiągnięć jest opracowanie i implementacja elektroniczna systemu ochrony i weryfikacji tożsamości na podstawie rytmu pisania na klawiaturze, poprzez przekształcanie prawdziwych danych biometrycznych wysyłanych na zewnątrz przy równoczesnej możliwości stosowania systemu do ochrony piszącego użytkownika. Powyższe dwa osiągnięcia uważam za ważny i istotny wkład pracy doktoranta w bezpieczeństwo systemów

biometrycznych.

Zakres badań doktoranta jest szeroki, ale poszczególne zagadnienia nie stanowią całości, i dotyczą dość różnych problemów ochrony danych biometrycznych. Nie widać myśli przewodniej rządzącej doborem poszczególnych tematów.

Pierwsze dyskutowane zagadnienie (r. 2.1) dotyczy **Scenariusza (b)**, tzn. personalizacji danych biometrycznych w elektronicznym dokumencie tożsamości. Przykładowo, podpis dostawcy danych w paszportach biometrycznych może zostać użyty przez nieuprawnioną osobę. W rozprawie zaproponowano rozwiązanie bazujące na drzewach Merkla polegające na tym, że dla każdego elektronicznego dowodu tożsamości zapisany jest główny sekret generowany probabilistycznie przez wystawcę dokumentu. Omówione są poszczególne kroki przedstawionej metody: personalizacja dokumentu przez wystawcę, personalizacja dokumentu przez właściciela i tworzenie tak spersonalizowanego podpisu. Opisywana jest również metoda weryfikacji tak utworzonego podpisu i wykrywania nadużyć przez użytkownika. Przeprowadzono również testy wydajnościowe przedstawionego systemu personalizacji podpisu. Przedstawiony system jest ciekawym i ważnym sposobem zabezpieczenia dowolnych danych, niekoniecznie biometrycznych.

W kolejnym rozdziale (r. 2.2) omawiany jest **Scenariusz (c)** dotyczący bezpieczeństwa biometrycznych baz danych. Doktorant wprowadza nowy system autoryzacji zapewniający anonimowość wykorzystujący obliczenia na zbiorach zachowujące prywatność. Zdefiniowany tu (t, n)-*Threshold Subset Problem* (TSP) wykorzystuje ochronę wzorców biometrycznych w oparciu o przekształcenie jednokierunkowe i filtry Blooma. Postępowanie to może dodatkowo chronić dane biometryczne, zabezpieczone w taki sposób, by można było z surowych danych biometrycznych tworzyć niezależne wzorce używane w razie kompromitacji poprzednich. Do technik takich należą odporne funkcji haszujące, trudnych do odwrócenia funkcji zachowujących prawdopodobieństwo i tzw. szkice biometryczne. Wprowadzony przez doktoranta (t, n)-TSP wraz z filtrami Blooma pozwala na utworzenie białej listy dla ograniczonej liczby użytkowników, która chroni prywatność użytkowników i uniemożliwia odtworzenie informacji które dane należą do których osób. Jądrem biometrycznym jest tu protokół Oblivious Polynomial Evaluation Naora i Pincasa, wykorzystujący wielomianową reprezentację sekretu. Doktorant szczegółowo analizuje działanie proponowanego (t, n)-TSP i jego bezpieczeństwo. Na potrzeby przykładu rozważany jest problem odcisku palca, zwykle używany do celów RBR.

W kolejnym rozdziale (r. 3) dyskutowany jest **Scenariusz (a)**, dotyczący ochrony niezamierzonego przepływu danych biometrycznych, dla systemu biometrycznego wykorzystującego rytm pisania na klawiaturze (*keystroking*). W odróżnieniu od poprzednich rozdziałów i scenariuszy, badana metoda dotyczy w zasadzie tylko konkretnej modalności biometrycznej. Doktorant zaprojektował i utworzył oprogramowanie pozwalające na rejestrację zdarzeń klawiatury, ich analizę i badanie profili użytkowników. Doktorant zaimplementował również algorytmy ochronne pozwalające na zmianę prawdziwych charakterystyk użytkowników. Zbudowano także urządzenie elektroniczne i zaimplementowano opracowane wcześniej algorytmy. Urządzenie przedstawiono do opatentowania w Polsce i uzyskano oznakowanie CE. Omawiane urządzenie jest wszechstronne. W trybie konfiguracji pozwala na dobór algorytmu ochrony danych. W trybie ochrony dokonuje modyfikacji danych biometrycznych. W trybie uczenia tworzy wzorzec bio-

metryczny osoby. W trybie weryfikacji wreszcie pozwala na zabezpieczenie komputera przed użyciem go przez nieuprawnione osoby, na podstawie ciągłej weryfikacji tożsamości.

Wydaje mi się, że kolejny rozdział (r. 4) nie wnosi pozytywnych efektów do pracy. Stwierdzenie to omówię w następnej części recenzji.

Ku mojemu zaskoczeniu, **Scenariusz (d)**, zapowiedziany w tezie pracy i wzmiankowany w kilku miejscach, nie jest nigdzie analizowany w pracy. Sądząc z wykazu literatury, której cytowanie rozdziale 2.2 kończy się na pozycji [22] a rozpoczyna w r.3 od pozycji [35], i faktu, że prace od [24] do [34] dotyczą steganografii, wydaje się, że autor przez pomyłkę (?) usunął cały podrozdział dotyczący jednej z tez. Brak analizy jednej z tez jest poważnym brakiem i wymaga uzupełnienia lub korekty.

4. **Poprawność metodologiczna rozwiązań, osiągnięcie celu** (czy rozwiązano postawione zagadnienia, czy użyto właściwe metody, czy przyjęte założenia są uzasadnione?)

   Badania przedstawione w r. 2.2 (scenariusz (c) dotyczący bezpieczeństwa biometrycznych baz danych) budzą mój niedosyt ze względu na brak analizy biometrycznej badanego systemu. Problem z proponowanym rozwiązaniem polega m.in. na tym, że konstrukcja koła, którego wycinki są podstawą zliczania minucji może być obciążona błędem. Argument, że od kilku do kilkunastu minucji w tych samych „miejscach" jednoznacznie identyfikuje osobę nie jest wystarczający, gdyż „miejsca" te są wycinkami koła, a nie współrzędnymi minucji. Ponadto, chodzi o współrzędne i kąty minucji a nie same współrzędne. Jest to jeden z ważnych problemów biometrii anulowalnych, który nie został jednak w pracy zbadany. Wartość proponowanych rozwiązań zależy silnie od jakości oryginalnych danych i doboru charakterystyk „mało zależnych" od wewnątrz-klasowego rozrzutu pomiarów. To podstawowe zagadnienie nie zostało w pracy zbadane.

   Do badań przeprowadzonych w r. 3 (scenariusz (a), dotyczący rytmu pisania na klawiaturze) włączono badanie rozkładów czasów „flight time" i testowanie ich normalności. Testowanie normalności nie zostało wykorzystane. Co więcej, mimo odrzucenia hipotezy o normalności, stosowano charakterystyki rozrzutu dla danych normalnych (rys. 12). Wykorzystywanie własności rozkładu normalnego nie było jednak potrzebne, gdyż można stosowana regułę 3-σ wynikającą bezpośrednio z nierówności Czebyszewa bez założenia normalności. Warto jeszcze zauważyć, że przy wielkiej liczbie próbek (jak u doktoranta) praktycznie każda hipoteza o rozkładzie zostanie odrzucona. Z drugiej strony, z inżynierskiego punktu widzenia „niewielkie" odchyłki od normalności są dopuszczalne. Problem ten jest szeroko dyskutowany w literaturze. Analizując metodę *keystroking*, doktorant podaje szereg wyników literaturowych FRR i FAR wziętych z (dość skąpego) przeglądu literatury na ten temat, nie znalazłem jednak nigdzie wartości FAR i FRR otrzymanych dla urządzenia zaprojektowanego przez doktoranta. Mam ponadto wątpliwości dotyczące przydatności „zamazywania", jeśli jego celem ma być uniemożliwienie rozpoznania osoby piszącej. Rytm pisania zostanie bowiem zastąpiony innym rytmem pisania, a wówczas rozpoznanie może nastąpić na podstawie rytmu zmodyfikowanego.

   Kolejny rozdział (r. 4) w zamierzeniu omawia zapewne zagadnienia związane ze scenariuszem (b) dla systemu rozpoznawania twarzy, ale sam system zabezpieczeń nie został

przedyskutowany. Przedstawiono system rozpoznawania twarzy na podstawie obrazu RGB i - być może - obrazu głębi IR (być może - gdyż schemat systemu nie zawiera elementów IR). Dziełem doktoranta jest budowa i testowanie tego systemu, budzą one jednak szereg zastrzeżeń. Zbudowany system dokonuje detekcji twarzy przy użyciu metody Violi-Jonesa (opis algorytmu Violi-Jonesa jest jednak błędny: działanie algorytmu na pierwszym stopniu kaskady nie zależy od „prostoty próbki", ale wynika z niskiego progu akceptacji twarzy na tym stopniu). Następnie dokonano weryfikacji twarzy przy użyciu kilku klasycznych algorytmów (twarze własne, twarze Fishera, histogramy LBP). Testowanie (s. 69) jest przeprowadzane na danych uczących (optymalizacja parametrów algorytmu V-J, parametrów filtrów etc.), w związku z tym wyniki są obciążone. Otrzymane rezultaty, szczególnie biorąc pod uwagę, że tabela dotyczy twarzy frontalnych, są żenująco słabe i - wbrew stwierdzeniu doktoranta - nie są obiecujące, szczególnie w kontekście ostatniego skoku jakościowego metod rozpoznawania twarzy. U doktoranta tylko połowa nieautoryzowanych prób dostępu została wykryta i stało się to kosztem ogromnej liczby fałszywych odrzuceń. Co więcej, wyniki testowania (tab. 3 i 4) są niejasne lub błędne, gdyż definicje używanych tam pojęć (s. 71) są błędne: zdarzenie błędnego odrzucenia przy weryfikacji jest związane z przekroczeniem progu, a nie z największym podobieństwem. Brak treningu detektora i klasyfikatorów został uzasadniony stwierdzeniem, że do uczenia należałoby dysponować dostatecznie dużą bazą próbek pozytywnych. Jest to argument połowicznie słuszny, gdyż np. dla algorytmu Violi-Jonesa istotnym elementem działania jest również zbiór „nie-twarzy", a także zbiór tła, na którym twarze się pokazuje, specyficzny dla przeprowadzanego eksperymentu. Argumentem przemawiającym przeciwko stosowaniu uczenia nie może być również - jak u doktoranta - chęć uniknięcia zbytniego dopasowania („przeuczenia"), gdyż zjawisko to może być kontrolowane na wiele sposobów, w tym przez stosowanie walidacji, regularyzację, tzw. *dropout* etc.

Ponadto, zastosowanie wyników do utworzenia tymczasowych dokumentów tożsamości, choć ciekawe, wydaje się jeszcze nieprzemyślane. Koncepcja, w której bez dotarcia do najbliższej jednostki konsularnej możliwe jest utworzenie dokumentu tymczasowego w oparciu o kamerę, mikrofon (i kamerę głębi ?) z użyciem oświetlaczy podczerwieni (?) jest wątpliwe. Cały rozdział różni się wyraźnie od poziomu pozostałych części pracy.

5. **Sposób przekazania wyników (jasność wywodów, poprawność redakcyjna, inne uwagi)**

   Sposób przedstawiania wyników jest nierówny. W zakresie kryptografii język jest zwięzły, zakładający wysoki poziom profesjonalizmu czytelnika, a w wywodach niektóre prostsze fragmenty sa pomijane. Wywody biometryczne są jednak zasadniczo inne, często dotyczą spraw podstawowych, a na dodatek bywają mylnie przedstawiane.

6. **Przydatność rozprawy dla nauk technicznych**

   Przedstawiona do recenzji praca doktorska jest wartościowa i przydatna dla nauk technicznych. Jednak brak analizy zapowiedzianej tezy dotyczącej steganografii musi zostać uzupełniony lub tezy pracy przeformułowane Równocześnie powinny być usunięte błędy metodologiczne w r. 4.

Powyższe zmiany mogą być wykonane w trybie „uzupełnienia" pracy doktorskiej.

Podsumowując stwierdzam, że przedstawiona przez pana mgr inż. Wojciecha Wodo rozprawa doktorska pt. *Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji*

wymaga wprowadzenia poprawek i uzupełnień wyjaśniających zarzuty, lub komentujących zarzuty, z którymi Autor się nie zgadza. Konieczne jest, w szczególności, wyjaśnienie dotyczące (omyłkowo?) brakującego uzasadnienia i dyskusji jednej z tez rozprawy. Sądzę, że wyjaśnienia te mogą być złożone w postaci uzupełnienia, bez konieczności przedkładanie nowej rozprawy.

prof. dr hab. inż. Andrzej Pacut
Instytut Automatyki i Informatyki Stosowanej
Wydział Elektroniki i Technik Informacyjnych
Politechnika Warszawska

30 października 2018

Recenzja rozprawy doktorskiej

mgr inż. Wojciecha Wodo

„Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji"

wraz z jej Uzupełnieniem

dla Rady Naukowej Instytutu Podstaw Informatyki Polskiej Akademii Nauk

Na prośbę Rady Naukowej Instytutu Podstaw Informatyki Polskiej Akademii dokonałem (2 lutegp 2018) recenzji rozprawy doktorskiej w dziedzinie nauk technicznych w dyscyplinie informatyka, przedstawionej przez p. mgr. inż. Wojciecha Wodo pt. „Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji". Recenzja ta zawierała szereg zastrzeżeń. W konkluzji recenzji stwierdziłem, że

> *przedstawiona do recenzji praca doktorska jest wartościowa i przydatna dla nauk technicznych. Jednak brak analizy zapowiedzianej tezy dotyczącej steganografii musi zostać uzupełniony lub tezy pracy przeformułowane Równocześnie powinny być usunięte błędy metodologiczne w r. 4. Powyższe zmiany mogą być wykonane w trybie „uzupełnienia" pracy doktorskiej.*

W wyniku powyższej recenzji otrzymałem od Doktoranta wyjaśnienie pt. „Uzupełnienie rozprawy doktorskiej pt. 'Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji' " (4 strony). We wstępie Uzupełnienia doktorant stwierdza, że „Niniejsze uzupełnienie rozprawy doktorskiej powstało na prośbę recenzenta rozprawy - prof. Andrzeja Pacuta *i stanowi integralną część rozprawy.*

Poniższe moje uwagi dotyczą więc całości Rozprawy, łącznie z jej integralnym uzupełnieniem.

Stwierdzam, że mój zasadniczy zarzut, dotyczący brak jednej z tez pracy, został całkowicie wyjaśniony. Jak pisze Doktorant, znaleziony został krytyczny atak na system bezpieczeństwa przedstawiony we wstępnej wersji rozprawy. W rezultacie, odpowiedni fragment pracy został z pracy usunięty, nie usunięto jednak omyłkowo zapowiedzi rozwiązania omawianego problemu. Wytłumaczenie powyższe w całości akceptuję.

Mój drugi istotny zarzut, który mógł wpłynąć na inne wyniki pracy, dotyczył błędnego sformułowania definicji pewnych parametrów jakości systemów biometrycznych. Poprawne definicje zostały przytoczone w Uzupełnieniu.

Uzupełnienie zawiera również wyjaśnienia związane z wieloma mniej istotnymi zarzutami recenzji. Odpowiedzi na te zarzuty traktuję jako element dyskusji naukowej nad wynikami pracy i jako takie w całości je akceptuję.

Podsumowując stwierdzam, że w związku z moją wcześniejszą recenzją i jej zaleceniami, przedstawiona przez pana mgr inż. Wojciecha Wodo rozprawa doktorska pt. „Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji" wraz z jej „Uzupełnieniem"

spełnia wymagania stawiane pracom doktorskim

i wnioskuję o jej przyjęcie i dopuszczenie do publicznej obrony .

Do mojej recenzji załączam recenzję z dn. 2 lutego 2018 pt "Wstępna recenzja rozprawy doktorskiej" mgr inż. Wojciecha Wodo "Zastosowanie i ochrona danych biometrycznych przy autoryzacji i identyfikacji"

Andrzej Pacut