Warszawa, dn. 12.05.2022 r.

Dr hab. Jerzy Cytowski, prof. uczelni
Instytut Informatyki
Uniwersytet Kardynała Stefana Wyszyńskiego w Warszawie

RECENZJA ROZPRAWY DOKTORSKIEJ DLA RADY NAUKOWEJ DYSCYPLINY INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA WOJSKOWEJ AKADEMII TECHNICZNEJ

Tytuł rozprawy. "Extending Selected Statistical Disclosure Control Models with Adaptive Metrics and Evaluation Framework"
Autor rozprawy: mgr inż. Olga Dzięgielewska
Promotor: dr hab. inż. Bolesław Szafrański, prof. WAT

1 Tematyka rozprawy

Recenzowana rozprawa doktorska przedstawia wyniki badań dotyczące ochrony statystycznych baz danych. Praca dotyczy więc bardzo ważnych problemów związanych z gromadzeniem i wykorzystywaniem danych osobowych przez publiczne instytucje statystyczne. Od dawna powstaje tu wiele napięć i konfliktów pomiędzy polityką publiczną a życiem prywatnym. Prowadzone są intensywne prace nad konstrukcją zabezpieczeń gwarantujących nie wykorzystywanie danych statystycznych w sposób, który mógłby mieć negatywny wpływ na jednostki. Praca Pani mgr inż. Olgi Dzięgielewskiej wpisuje się więc w nurt aktualnych i ważnych badań w zakresie konstruowania narzędzi informatycznych chroniących prywatność osób. W naszym kraju prace w zakresie ochrony przed naruszeniem danych, czy kradzieżą tożsamości są szczególnie ważne. Według rankingu Open Data Inventory (ODIN) z 2020 roku, oceniającego stopień dostępności i otwartości danych prezentowanych przez krajowe urzędy statystyczne, GUS awansował na 2 pozycję wśród 187 krajowych urzędów statystycznych na świecie. W publicznym udostępnianiu, różniących się od siebie, prywatnych danych lub statystyk w taki sposób by dawały prawidłowe wyniki, ale niwelowały ryzyko ujawnienia prywatności, bardzo trudne jest określenie uniwersalnych mierników - metryk oceniających jakość danych

statystycznych. Jeśli parametr utraty prywatności jest ustawiony tak, aby faworyzować użyteczność, korzyści związane z prywatnością są zmniejszone (mniej „szumów" jest wprowadzanych do systemu); jeśli parametr utraty prywatności jest ustawiony tak, aby faworyzować dużą prywatność, dokładność i użyteczność zbioru danych są obniżone (więcej „szumów" jest wprowadzanych do systemu). Ważne jest, aby decydenci rozważyli kompromisy wynikające z prywatności różnicowej, aby pomóc w ustaleniu odpowiednich najlepszych praktyk i standardów dotyczących korzystania z tej praktyki ochrony prywatności, zwłaszcza biorąc pod uwagę różnorodność w organizacyjnych przypadkach użycia. Warto jednak zauważyć, że zmniejszona dokładność i użyteczność jest powszechnym problemem we wszystkich metodach ograniczania ujawniania danych statystycznych i nie dotyczy wyłącznie prywatności różnicowej. Praca mgr inż. Olgi Dzięgielewskiej wpisuje się w ważny obszar badań. Przedstawiona w rozprawie metoda ochrony statystycznych baz danych oparta na mechanizmach zniekształcania (zaszumiania) danych polega na rozszerzeniu znanego z literatury modelu prywatności różnicowej o metryki adaptacyjne. Jest to rozwiązanie, którego celem jest spełnienie zarówno wymaganego poziomu poufności, jak i dokładności statystyk uzyskiwanych ze statystycznej bazy danych.

2 Ocena treści rozprawy i wkładu oryginalnego

2.1 Ocena treści rozprawy

Rozprawa jest zwięzła: składa się z 9 rozdziałów, bibliografii i ma w sumie 117 stron.

We wstępnym rozdziale Autorka przedstawiła rozszerzony spis treści, tabelę stosowanych w pracy terminów, definicji, skrótów oraz spisy rysunków i tablic zamieszczonych w pracy. Rozdział 2 prezentuje definicję statystycznej bazy danych, podkreślając specyfikę danych statystycznych w problemach bezpieczeństwa danych. Analizowane są tutaj problemy statystycznych interfejsów umożliwiających dostęp do danych, uwarunkowań prawnych w Polsce, UE , USA oraz problemy związane z zarządzaniem cyklem danych. Rozdział 3 opisuje motywacje Autorki, która skłoniły ją do podjęcia badań przedstawionych w rozprawie. Sformułowane zostały cele pracy i hipotezy badawcze. Pani mgr inż. Olga Dzięgielwska pragnie w rozprawie wykazać, że statystyczna metoda kontroli ujawniania informacji, oparta na modelu prywatności różnicowej, rozszerzona o metryki adaptacyjne (metryki dokładności ryzyka i wyniku informacyjnego) spełnia statystyczne kryteria bezpieczeństwa (zgodnie z definicją w ramach VIOLAS) i stanowi skuteczny środek zaradczy przeciwko naruszeniom prywatności przez ataki wnioskowania. Tak zmodyfikowany model prywatności różnicowej rozszerza ochronę baz statystycznych również w przypadku pojawienia się nowych danych wrażliwych w bazie. Istotnym dla wykazania przydatności rozwiązań zaproponowanych w rozprawie jest zastosowanie odpowiedniej metodologii oceny ryzyka dla danych statystycznych. W rozdziale 4 przedstawione są normy i reguły takich ocen ryzyka dla systemów IT oraz metodologia przeprowadzenia

oceny dla ataków teoretycznych. Rozdział 5, oparty na publikacji[1], definiuje strukturę VIOLAS. Strukturę oryginalnie zaprojektowaną i opracowaną w ramach badań Autorki rozprawy. VIOLAS pozwala na ustandaryzowaną ocenę statystycznej kontroli ujawnień. W rozdziale 6 przedstawiony jest rozszerzony model prywatności różnicowej oraz zdefiniowane są nowe metryki. Treść tego rozdziału oparta jest na publikacji[2]. Rozdział 7 ukazuje ocenę skuteczności nowych rozwiązań. Efektywność nowych metryk jest oceniana przy użyciu struktury VIOLAS i porównywana ze standardowymi analizami bez stosowania tych metryk. W tym rozdziale analizowane są problemy bezpieczeństwa związane z oceną ryzyka. Rozdział 8 prezentuje analizę możliwości zastosowań i rozszerzeń proponowanych w rozprawie rozwiązań. Prezentuje możliwości zastosowania tych rozwiązań w transakcyjnych bazach danych. Zasygnalizowane są również możliwości optymalizacji implementacyjnych. Rozdział 9 zawiera podsumowanie i wnioski

2.2 Ocena kompletności i redakcji pracy

Redakcja pracy jest poprawna. Praca napisana jest klarownym i przystępnym językiem angielskim. Autorka umiejętnie operuje terminologią i prowadzi wywody w logiczny i przekonujący sposób. W tym względzie należy wyróżnić jeden z ważniejszych – rozdział 6. Cele pracy zostały wyraźnie sformułowane i Autorka w przekonywujący sposób wykazała ich realizację.

Autorka nie uniknęła jednak pewnych braków i uchybień:

- wydaje się, że analiza rozprawy mogłaby być bardziej efektywna przy jawnym sformułowaniu listy deklarowanych wyników oryginalnych. Są sformułowane w zawoalowanej postaci,
- praca stałaby się bardziej kompletna, gdyby umieszczono w niej analizę bibliograficzną znanych modyfikacji modelu prywatności różnicowej,
- styl odsyłaczy bibliograficznych i zasady opisu wykazu literatury nie jest jednolity, czasami jest niekompletny i zawiera literówki,
- nie wszystkie ilustracje są wystarczająco czytelne. Na przykład na rysunku 3 (istotnym ze względu na opis dokonań Autorki) bardzo trudno jest odczytać informacje.

2.3 Ocena wkładu oryginalnego

Mimo nielicznych uchybień redakcyjnych, praca jest merytorycznie spójna, przekonująca,
i zaawansowana w sensie metodycznym. Za jej główne oryginalne wyniki uznaję:

- opracowanie struktury VIOLAS,

[1] O. Dzięgielewska, Evaluating Quality of Statistical Disclosure Control Methods – VIOLAS Framework, Privacy in Statistical Databases, UNESCO Chair in Data Privacy, International Conference, PSD2020, Springer Lecture Notes in Computer Science 12276.
[2] O. Dzięgielewska, Defeating inference threat with scoring metric. 36th IBIMA Conference.

- rozszerzenie modelu prywatności poprzez opracowanie metryk adaptacyjnych modyfikujących własności metody prywatności różnicowej i stwarzających możliwość pomiaru jakości i dokładności wyniku informacyjnego,
- wykonanie analizy jakości opracowanej metody ochrony i wykazanie jej zalet w porównaniu do bazowego modelu prywatności różnicowej,
- przedstawienie najważniejszych wyników pracy w znaczących publikacjach.

3 Konkluzja końcowa

Rozprawa doktorska mgr inż. Olgi Dzięgielewskiej zawiera, w mojej ocenie, oryginalne i wartościowe osiągnięcia, zilustrowane analizą jakościową. Nieliczne uwagi, odnoszące się głównie do redakcji rozprawy, nie podważają głównych konkluzji rozprawy i mojej pozytywnej jej oceny. Szczególnie przekonują mnie: stopień zaawansowania koncepcyjnego i metodologicznego pracy, oryginalność wyników, aktualność tematyki i sprawność stosowania narzędzi badawczych.

Wobec powyższego stwierdzam, że rozprawa doktorska mgr inż. Olgi Dzięgielewskiej pt. "Extending Selected Statistical Disclosure Control Models with Adaptive Metrics and Evaluation Framework" spełnia z nawiązką wymagania stawiane rozprawom doktorskim w Ustawie o stopniach naukowych i tytule naukowym oraz stanowi oryginalne rozwiązanie problemu naukowego. Wnoszę do Rady Dyscypliny Naukowej Informatyka Techniczna i Telekomutacja Wojskowej Akademii Technicznej w Warszawie o dopuszczenie Pani mgr inż. Olgi Dzięgielewskiej do dalszych etapów przewodu doktorskiego.