prof. dr hab. inż. Krzysztof Goczyła
Politechnika Gdańska
Wydział Elektroniki, Telekomunikacji i Informatyki
kris@eti.pg.edu.pl

Gdańsk, 15.03.2021

Recenzja rozprawy doktorskiej

mgr. inż. Michała Drozdowicza

zatytułowanej

"Semantic technologies for support of access control to data and services in the Internet of Things"

(„*Semantyczne wspomaganie kontroli dostępu do danych i usług w Internecie rzeczy*")

1. Wstęp

Niniejsza recenzja została sporządzona na podstawie pisma Zastępcy Dyrektora ds. Naukowych Instytutu Badań Systemowych PAN dr. hab. inż. Jana W. Owsińskiego z dn. 11.01.2021 r., w którym zawarta jest informacja o powołaniu mnie na recenzenta ww. rozprawy doktorskiej. Do pisma dołączono tekst rozprawy, wykaz publikacji Doktoranta oraz krótkie streszczenie rozprawy w języku polskim.

2. Problem badawczy i jego znaczenie

Tematyka rozprawy dotyczy metod kontroli dostępu do zasobów informacji zgromadzonych w tzw. Internecie rzeczy. W kontekście recenzowanej rozprawy, Internet rzeczy rozumiany jest bardzo szeroko – nie tylko jako zbiór fizycznych urządzeń dostępnych w sieci Internet poprzez adres IP, ale także jako zbiór informacji zgromadzonych i udostępnianych w Internecie. Problem ten jest niezwykle istotny i aktualny z uwagi na to, że coraz więcej zasobów, fizycznych i informacyjnych, jest bezpośrednio podłączonych do globalnej sieci, a tym samym narażonych na nieuprawniony, w tym złośliwy i szkodliwy, dostęp, który może być nawet katastrofalny w skutkach. W swojej rozprawie Autor zajął się zagadnieniami kontroli dostępu w kilku wybranych obszarach – zdrowie, logistyka, sieci energetyczne i współdzielenie danych prywatnych.

Problem ochrony informacji i innych zasobów jest dobrze rozpoznany i rozwijany w informatyce od wielu lat. Wypracowano różne schematy i metody definiowania i wymuszania kontroli dostępu, o różnym stopniu komplikacji. Schematy te dają możliwości definiowania konkretnej polityki dostępu w zależności od tego, jakie zasoby mają być chronione, kto powinien mieć i jakie uprawnienia dostępu do tych zasobów i jak zadbać o to, by te uprawnienia nie były naruszane.

Autor rozprawy skoncentrował się na dość ogólnym schemacie kontroli zwanym *Attribute Based Access Control* (ABAC), czyli na schemacie bazującym na wartościach atrybutów chronionych zasobów oraz podmiotów uczestniczących w procesie dostępu do nich. Schemat ten Autor rozszerza o warstwę semantyczną opartą na ontologiach i wnioskowaniu. Jest to podejście uzasadnione merytorycznie przynajmniej z dwóch względów. Po pierwsze – opis świata, w tym zasobów dostępnych w Internecie rzeczy, w postaci ontologii jest opisem opartym na logice, a więc precyzyjnym, a jednocześnie zrozumiałym dla komputerów w tym sensie, że podlegającym

przetwarzaniu komputerowemu. (Oczywiście, komputer nie jest w stanie dokonać merytorycznej walidacji konkretnej ontologii dziedzinowej; może ją tylko zweryfikować formalnie). Takich ontologii, zarówno wysokiego poziomu, jak i dziedzinowych, udostępnianych jest coraz więcej. Po drugie, warstwa semantyczna pozwala wzbogacić schemat ochrony zasobów o zależności i fakty niejawne, wywnioskowane z ontologii. Oczywiście, dodanie warstwy semantycznej może wiązać się z pogorszeniem efektywności systemów kontroli dostępu, m.in. z uwagi na fakt, że wnioskowanie z nietrywialnych ontologii należy do klasy problemów trudnych obliczeniowo.

Stąd też podjęcie takiego wyzwania – rozszerzenia znanego i dość uniwersalnego schematu kontroli dostępu o warstwę semantyczną – uważam za zadanie ważne i naukowo interesujące z punktu widzenia współczesnej szeroko rozumianej inżynierii oprogramowania, w tym inżynierii systemów opartych na wiedzy.

3. Cele i tezy rozprawy

W rozdziale 1. rozprawy Autor definiuje tezę i cele rozprawy. Zgodnie z tezą, **rozszerzenie systemów kontroli dostępu o techniki wnioskowania semantycznego stanowi efektywną metodę zarządzania dostępem do zasobów i prywatnością w środowiskach Internetu rzeczy** (*tłum. moje*). Wykazaniu prawdziwości tej tezy służą szczegółowe cele badawcze. Są nimi:

- Opracowanie wielowarstwowej architektury ontologicznej pozwalającej na wykorzystanie w systemach kontroli dostępu istniejących ontologii dziedzinowych.
- Opracowanie metody reprezentacji wiedzy o podmiotach i zasobach z zastosowaniem ontologii oraz sposobów odwzorowania ontologii na atrybuty schematu ABAC.
- Opracowanie i implementacja semantycznego systemu kontroli dostępu do zasobów, rozszerzającego w sposób naturalny standardową implementację schematu ABAC opartą na języku XACML.
- Opracowanie narzędzia umożliwiającego utworzenie potrzebnych ontologii i wykazanie możliwości zastosowania utworzonego systemu w wybranych obszarach Internetu rzeczy.

Do samego sformułowania tezy, celów pracy i proponowanego podejścia badawczego nie mam zastrzeżeń, aczkolwiek trzeba sobie zdawać sprawę, że uzasadnienie tak sformułowanej tezy nie jest łatwe. Potencjalnie wymaga to złożonego i trudnego w praktycznej realizacji procesu walidacji w rzeczywistych środowiskach produkcyjnych, a nie tylko w środowiskach testowych i w hipotetycznych scenariuszach. Również pojęcie efektywności, do którego odnosi się teza, nie jest jednoznaczne. Jednak podobne problemy z tezami występują w większości prac doktorskich z obszaru inżynierii oprogramowania, które często są uzasadniane jedynie w środowiskach eksperymentalnych.

4. Układ i zawartość pracy

Rozprawa napisana jest w bardzo dobrym języku angielskim. W całym tekście natrafiłem na jedynie kilka błędów literowych. W rozdziale 1., poprzedzonym użytecznym wykazem skrótów i oznaczeń, Autor wprowadza w zagadnienia i kontekst pracy, uzasadniając podjęcie tematyki, formułując tezę i cele rozprawy. W rozdziale 2. znajduje się bardzo szczegółowy opis aktualnego stanu wiedzy w zakresie mechanizmów kontroli dostępu oraz z obszaru współczesnych technik semantycznych. W rozdziale 3. Autor opisuje cztery scenariusze, które staną się przedmiotem eksperymentów mających na celu wykazanie prawdziwości tezy. Scenariusze te dotyczą: dostępu do danych zdrowotnych pacjenta, dostępu do zasobów portowych, współdzielenia danych prywatnych

w sieciach typu Smart City oraz zarządzania uprawnieniami w sieciach energetycznych typu Smart Grid. Rozdział 4., mieszczący się aż na 65 stronach gęstego tekstu, stanowi zasadniczą, oryginalną część pracy. W tym rozdziale Autor prezentuje swój pomysł rozszerzenia schematu kontroli dostępu ABAC o warstwę semantyczną bazującą na ontologiach dziedzinowych i ontologiach odwzorowań pomiędzy dziedzinami a schematem ABAC. W końcowej części tego rozdziału prezentowane jest opracowane przez Autora narzędzie OntoPlay służące do przyjaznego dla użytkownika tworzenia ontologii odwzorowań. Moim zdaniem, ten rozdział zyskałby na czytelności, gdyby został podzielony przynajmniej na dwie części – część opisującą koncepcję proponowanego rozwiązania i część opisującą narzędzie OntoPlay. Mam też inne zastrzeżenia co do zawartości tego rozdziału, które sformułowałem poniżej, w sekcji 6. niniejszej recenzji. W rozdziale 5. Autor opisuje eksperymenty, w których zastosował proponowane podejście do poprzednio zaprezentowanych scenariuszy testowych. Rozdział 6. zawiera podsumowanie wyników rozprawy i oryginalnego dorobku Autora, a także bardzo dobrze zdefiniowane możliwe kierunki dalszych prac. Pracę wieńczy obszerna bibliografia zawierająca 152 pozycje, w tym 5 pozycji, których pierwszym współautorem jest Autor rozprawy.

Ogólnie, z powyższym zastrzeżeniem dotyczącym rozdziału 4., układ rozprawy jest logiczny i konsekwentny, służący realizacji celu rozprawy i uzasadnieniu prawdziwości postawionej tezy.

5. Oryginalny wkład Autora

Oryginalny dorobek i wkład w dziedzinę Autor zaprezentował w rozdziałach 4. i 5., w których opisano koncepcję i szczegóły połączenia schematu ABAC z warstwą semantyczną opartą na ontologiach oraz dokonano eksperymentów potwierdzających prawdziwość tezy rozprawy. Za najważniejsze elementy tego oryginalnego wkładu uważam:

1. Zdefiniowanie ogólnej metody semantycznego wzbogacania bazującego na atrybutach modelu kontroli dostępu o model dziedzinowy sformułowany w formie ontologii zapisanych w języku OWL.
2. Zdefiniowanie rozszerzonych o tę warstwę semantyczną stosownych modułów odpowiadających elementom referencyjnej architektury zdefiniowanej w standardzie XACML: tj. modułów *Policy Administration Point* (PAP), *Policy Decision Point* (PDP) i *Policy Information Point* (PIP).
3. Rozszerzenie zestawu funkcji XACML o funkcje w języku SPARQL wydobywające potrzebne wartości atrybutów z ontologii dziedzinowych.
4. Opracowanie i zweryfikowanie formalne ontologii utworzonych na bazie ontologii istniejących, a także utworzenie ontologii oryginalnych, odpowiednich do zaproponowanego podejścia i scenariuszy testowych.
5. Zdefiniowanie odwzorowań pomiędzy modelami ontologicznym a atrybutowym, wykorzystywanych w procesie wzbogacania schematu atrybutowego o warstwę semantyczną.
6. Zaprojektowanie i implementację narzędzia umożliwiającego przeprowadzenie potrzebnych eksperymentów testowych, a także – poprzez udostępnienie go na zasadach *open source* - innych, analogicznych.
7. Przeprowadzenie, za pomocą opracowanego narzędzia, różnorodnych eksperymentów wykazujących możliwość praktycznego zastosowania zaproponowanego podejścia do kilku wybranych problemów związanych z kontrolą dostępu do zasobów szeroko rozumianego Internetu rzeczy.

Podsumowując tę część recenzji, stwierdzam, że Autora rozprawy w pełni zrealizował postawione w rozprawie cele i uczynił to w sposób wyczerpujący i dojrzały z naukowego punktu widzenia. Tezę rozprawy można uznać za udowodnioną w takim zakresie, w jakim jest to możliwe w rozprawie doktorskiej z obszaru praktycznej inżynierii oprogramowania.

6. Uwagi krytyczne i komentarze

Poniżej zamieszczam uwagi i komentarze, jakie nasunęły mi się w trakcie czytania rozprawy. Niektóre z nich mają charakter dyskusyjny i mam nadzieję na podjęcie ich w trakcie publicznej obrony.

1. W pracy brakuje wysokopoziomowego opisu procesu biznesowego związanego z proponowanym podejściem. W rozdziale 4. Autor zasypuje czytelnika szczegółami technicznymi, nie prezentując żadnego ogólnego schematu, który pokazywałby kolejne etapy postępowania w proponowanej przez Autora metodzie, a także udział w tych etapach podmiotów (aktorów) ożywionych i nieożywionych. Schemat komponentowy prezentowany jest dopiero w rozdziale 5., ale ma on charakter implementacyjny, a nie procesowy. Warto by określić, kto i kiedy definiuje lub wybiera ontologie dziedzinowe, kto je waliduje, kto i kiedy definiuje ontologie odwzorowujące, a także jak przebiegają dalsze etapy tworzenia żądań i ich przetwarzania. W pracy występuje jedynie tajemniczy „użytkownik", ale wiadomo, że aktorami tego procesu są też moduły programowe określone w schemacie ABAC oraz moduły warstwy semantycznej. Można to oczywiście wywieść z, miejscami jednak dość zawikłanego, technicznego tekstu rozdziału 4., jednak moim zdaniem taki schemat procesowy powinien być przedstawiony przed prezentacją szczegółów rozwiązania. Wówczas poszczególne elementy techniczne można by w sposób naturalny wiązać z etapami procesu biznesowego. Jest to według mnie poważny mankament natury poznawczej, utrudniający śledzenie szczegółowych wywodów będących zasadniczym elementem rozprawy.

2. W rozdziale 5. Autor prezentuje wyniki eksperymentów dla czterech hipotetycznych scenariuszy, pokazujących zastosowanie proponowanego podejścia w różnych kontekstach, także dla zasobów hierarchicznych. Należy to docenić, jednak wartość tej części rozprawy znacznie by wzrosła, gdyby choć jeden, nawet nie nazbyt skomplikowany, był scenariuszem rzeczywistym. Jest to oczywiście problem natury głębszej, gdy w pracach doktorskich z inżynierii oprogramowania zazwyczaj najsłabszym elementem jest walidacja proponowanego rozwiązania w rzeczywistym środowisku przemysłowym. Jednak myślę, że w tym przypadku Autor mógłby pokusić się o kontakt z jakąś firmą i krytyczne przeanalizowanie swojego rozwiązania na rzeczywistym problemie dostępu do zasobów.

3. W tymże rozdziale 5. podawane są czasy realizacji poszczególnych operacji w scenariuszach testowych. Intencją Autora było wykazanie prawdziwości tezy rozprawy, czyli efektywności proponowanej metody, i to istotnie jest wykazane, gdyż nawet w najbardziej złożonych konfiguracjach zasobów czasy te są akceptowalne. Jednak mam tu dwa zastrzeżenia. Po pierwsze, jaki jest sens podawania czasów z dokładnością do 1 μs w sytuacji, gdy te czasy są rzędu setek milisekund, a w przypadkach najbardziej złożonych kilku tysięcy milisekund (patrz np. tabela 5.2). Czy Autor szacuje względną dokładność swoich pomiarów aż na 10^{-6}? Po drugie, wartość tych eksperymentów i samych pomiarów niepomiernie wzrosłaby, gdyby porównano je z pomiarami dla tych samych scenariuszy przy zastosowaniu „czystego" schematu ABAC, bez warstwy semantycznej. W rozprawie brak jakiejkolwiek wzmianki

na ten temat, a przecież wiadomo, że w informatyce praktycznej jest tak, że wprowadzanie kolejnych warstw w architekturze systemu wiąże się z nieuniknionymi kosztami.

4. Algorytm nr 1 „Finding attribute values" jest w kilku miejscach niejasny. W wierszu 1. operacja „load domain ontology" jest pokazana jako bezargumentowa procedura, natomiast w wierszu 2. operacja „load mapping ontology" jest opatrzona nawiasami sugerującymi, że ma jakieś – niewymienione – argumenty. Czy jest to celowe, czy niekonsekwencja? Ponadto zakładając, że strzałka w lewo oznacza instrukcję podstawienia, te dwa wiersze: 1. i 2. pozostają w sprzeczności z wierszem 3., w którym po prawej stronie występuje „new ontology". Ten fragment pseudokodu wymaga uściślenia. Podobnie jest z fragmentem od wiersza 16. do 21. Nie rozumiem podstawienia w wierszu 17. – po prawej stronie występuje zapytanie: „query ontology for attribute value". Podejrzewam, że chodzi tu nie o zapytanie, tylko o jego wyniki. Ponadto dziwnie wygląda fraza „result in sol" w instrukcji **foreach** w wierszu 17., gdyż zmienna *result* jest przecież wielozbiorem. Sądzę, że w tym miejscu powinna być zastosowana podobna konstrukcja pętli **foreach**, jak w algorytmie nr 2 „Evaluation for multiple resource class instances" na str. 95. Na marginesie – w tym algorytmie również występuje wątpliwa konstrukcja, w której po prawej stronie instrukcji podstawienia występuje instrukcja – wiersz 7. – oraz analogiczne do algorytmu nr 1 niespójności w wierszach 1. do 3.
5. Na stronie 106. podano definicję symbolu C_p raz jako „list of declaring properties", a raz jako „list of all classes for which the property can be applied". Ponadto wspomnianej definicji symbolu C_p towarzyszy nigdzie nie użyty w tej definicji symbol c_0. Ciekawe, że ten sam symbol C_p w algorytmie nr 3 na następnej stronie jest nazwany „listDeclaringClasses(p)".
7. Czy algorytm nr 3 musi być wykonywany oddzielnie dla każdej klasy c_0 ze zbioru wszystkich klas C? Warto rozważyć, czy nie można tak skonstruować tej procedury, by była wykonywana po kolei dla wszystkich właściwości (properties), a nie dla wszystkich klas. W aktualnej wersji każda właściwość jest przetwarzana wiele razy, co – jak się wydaje – stwarza niepotrzebny narzut.
8. W opisie algorytmu nr 3 na str. 106. i 107. użyto terminu „dense ranking". Warto by wyjaśnić znaczenie tego terminu.
9. W opisie ontologii prezentowanej na rys. 5.10 występuje klasa *ObservedLicense*. Nie znalazłem tej klasy na tym rysunku.

Podkreślam, że powyższe uwagi krytyczne i komentarze nie stoją w sprzeczności z moją ogólnie bardzo pozytywną opinią o recenzowanej rozprawie.

10. Podsumowanie

Stwierdzam, że mgr inż. Michał Drozdowicz w swojej rozprawie doktorskiej wykazał się bardzo dobrą znajomością problematyki kontroli dostępu w środowiskach Internetu rzeczy oraz zagadnień konstrukcji systemów informatycznych opartych na wiedzy. Przedstawił w sposób wyczerpujący aktualny stan wiedzy w tym zakresie, a dla osiągnięcia celów swojej pracy wykorzystał istniejące narzędzia informatyczne oraz opracował własne. Swoją bardzo dobrą znajomość obszaru objętego tematyką rozprawy potwierdził adekwatnym doborem literatury przedmiotu. Jako wynik swojej pracy dwa ważne i aktualne obszary informatyki praktycznej – metody kontroli dostępu i inżynierię systemów opartych na wiedzy – połączył w spójne logicznie i implementacyjnie rozwiązanie.

Oceniam, że Autor recenzowanej rozprawy jest badaczem zdolnym do prowadzenia samodzielnych badań w obszarach związanych z inżynierią oprogramowania i inżynierią wiedzy. W swojej rozprawie wykazał się wiedzą i umiejętnościami wymaganymi do uzyskania stopnia doktora nauk technicznych w dyscyplinie **informatyka techniczna i telekomunikacja**, zgodnie z obowiązującymi przepisami, i wnoszę o przekazanie recenzowanej rozprawy do dalszych etapów przewodu doktorskiego.