

Dr hab. inż. Roman Stanisław Deniziak, prof. uczelni
Katedra Systemów Informatycznych
Politechnika Świętokrzyska
ul. 1000-lecia Państwa Polskiego 7
25-314 Kielce

Kielce, 12.09.2022

Recenzja rozprawy doktorskiej dla Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej

Tytuł rozprawy: Ewaluacja systemów wbudowanych poprzez monitorowanie programowe

Autor rozprawy: mgr inż. Kazimierz Krosman

Promotor rozprawy: Prof. dr hab. inż. Janusz Sosnowski

1. Cel, zakres, charakter i teza rozprawy

Rozprawa dotyczy problematyki monitorowania pracy systemów wbudowanych poprzez analizę danych zebranych w trakcie działania systemu. Rozpatrywane są trzy rodzaje informacji charakteryzujących przebieg wykonania oprogramowania budowanego: przebieg wykonywania instrukcji tzw. ślad instrukcji, przebieg wybranych sygnałów w postaci szeregów czasowych oraz logi programowe. Analiza tych danych umożliwia wykrywanie specyficznych sytuacji takich jak błędy wykonania, anomalie, nieefektywność działania systemu. Głównym celem pracy doktorskiej było opracowanie efektywnych metod analizy śladu programu, przebiegów sygnałów oraz logów programowych. Metody te zwiększą skuteczność monitorowania wykonywania oprogramowania wbudowanego pod kątem poprawności i wydajności. Cel pracy został sformułowany w postaci następującej tezy rozprawy:

„Analiza i monitorowanie pracy systemów wbudowanych wymaga stworzenia odpowiednich modeli dekompozycji, agregacji i korelacji danych z procesów monitorowania”.

O ile istota tezy w kontekście przedstawionej motywacji i opisu celu pracy jest zrozumiała to jednak tak sformułowana teza wydaje się zbyt ogólna i powinna być doprecyzowana poprzez dodanie czynnika jakościowego w odniesieniu do oczekiwanych wyników analizy, wskazującego na uzyskanie lepszych wyników niż za pomocą istniejących metod. Takie stwierdzenie jest już podane w dalszej części opisującej cel pracy, gdzie jest mowa o tym, że stworzenie odpowiednich modeli jest niezbędne do tego aby analiza była efektywna, co można rozumieć jako dająca większe możliwości niż dostępne rozwiązania. Przy takim uzupełnieniu tezę można uznać za poprawną.

Cel pracy został jasno sformułowany. Autor przedstawił motywację do podjęcia tematyki badawczej, wskazując na słabe strony i braki istniejących metod analizy wyników monitorowania systemów wbudowanych, mających na celu optymalizację, wykrycie błędów i poprawę jakości oprogramowania wbudowanego. Na tej podstawie określił cel pracy jako opracowanie odpowiednich modeli reprezentacji danych uzyskanych w wyniku monitorowania pracy systemu wbudowanego oraz algorytmów analizy bazujących na tych modelach.

Praca ma charakter projektowo-doświadczalny. Autor zaproponował modele do reprezentacji danych z monitoringu przebiegu wykonywania programów wbudowanych. Następnie zaproponował metody analizy dedykowane do opracowanych modeli. Opracowane rozwiązania zostały zweryfikowane i zademonstrowane na przykładowych zestawach danych pozyskanych dla rzeczywistych systemów.

2. Analiza źródeł

Bibliografia obejmuje 95 pozycji w tym 6 publikacji naukowych własnego autorstwa lub współautorstwa. Zdecydowana większość pozycji bibliograficznych obejmuje publikacje z czasopism zagranicznych oraz materiałów konferencji międzynarodowych. Autor w większości korzystał z literatury wydanej w ostatnich latach, około 80% bibliografii obejmuje publikacje z ostatnich 10 lat, znacząca liczba źródeł pochodzi z ostatnich 3 lat. Zatem Autor korzystał z najnowszej literatury i wykaz źródeł jest aktualny. Tematyka wykorzystywanej literatury jest ściśle związana z tematyką pracy i dotyczy metod monitorowania, analizy i testowania a także metod wykrywania anomalii i analizy szeregów czasowych.

Przegląd literatury został przedstawiony w podrozdziale 1.1. Autor przedstawił kolejno znane z literatury różne podejścia do monitorowania pracy systemów wbudowanych na poziomie śledzenia przebiegu wykonywania instrukcji, monitorowania sygnałów oraz monitorowania zdarzeń. Następnie przedstawił przegląd metod analizy zebranych danych mającej na celu stworzenie modeli behawioralnych systemu. Znacząca część analizy źródeł poświęcona jest metodom analizy szeregów czasowych w tym z wykorzystaniem metod uczenia maszynowego. Metody te służą wykrywaniu właściwości periodycznych oraz trendów w zachowaniu się systemu. Inną grupą metod w tej grupie stanowią metody klasyfikacji szeregów czasowych. Przegląd literatury obejmuje również metody analizy poboru energii w systemach wbudowanych oraz analizy i korelacji logów zdarzeniowych.

Analiza źródeł stanowiła punkt wyjścia do sformułowania tezy rozprawy doktorskiej. Na podstawie swojego doświadczenia praktycznego Autor zidentyfikował potrzeby opracowania metod umożliwiających wykonanie efektywnej analizy danych pozyskiwanych w trakcie monitorowania pracy systemu wbudowanego. Przegląd literatury światowej potwierdził brak zadowalających rozwiązań w tym zakresie.

Analiza aktualnego stanu wiedzy wykonana przez Autora jest właściwa. Autor przeanalizował zarówno potrzeby wynikające z zastosowań w przemyśle jak i aktualne rozwiązania przedstawione w literaturze światowej, opierając się na najnowszych publikacjach związanych z tematyką pracy.

3. Metodyka badań

Przyjęta metodyka badań zakładała opracowanie 3 głównych algorytmów: algorytm analizy śladu instrukcji, algorytm analizy szeregu czasowego oraz algorytm korelacji logów tekstowych z szeregiem czasowym. Opracowanie tych algorytmów wymagało opracowania odpowiednich modeli danych, umożliwiających zastosowanie efektywnych metod przetwarzania. Następnie opracowane modele i metody zostały zweryfikowane na praktycznych przykładach. Takie podejście jest poprawne i właściwe z punktu widzenia celu pracy.

Autor formułując tezę rozprawy założył, że środkiem do uzyskania efektywnej metody monitorowania i analizy systemów wbudowanych będzie opracowanie odpowiednich modeli i metod dekompozycji, agregacji i korelacji danych pozyskanych w procesie monitorowania systemu. W ramach realizacji pracy doktorskiej Autor rozwiązał postawione zagadnienia poprzez opracowanie tych modeli i wykazanie skuteczności zaproponowanych metod. Tym samym potwierdził, że przyjęte założenia były słuszne.

Autor uzyskał zamierzony cel pracy, co stanowiło jednocześnie dowód tezy. Dla zaproponowanych modeli danych opracował odpowiednie algorytmy analizy. Wykonane analizy praktycznych przykładów potwierdziły skuteczność i przydatność opracowanych rozwiązań.

4. Oryginalność uzyskanych wyników

Najważniejsze oryginalne osiągnięcia autora przedstawione w pracy to:

1. Opracowanie modelu danych reprezentującego ślad wykonania programu oraz opracowanie algorytmu wykrywania anomalii bazującego na tym modelu.
2. Opracowanie hierarchicznego, obiektowego modelu danych reprezentujących szeregi czasowe oraz opracowanie algorytmu analizy szeregów czasowych operującego na zaproponowanym modelu danych.
3. Opracowanie modelu danych reprezentującego logi tekstowe oraz opracowanie algorytmu korelacji logów tekstowych z szeregiem czasowym.

Ważnym rezultatem pracy jest również implementacja opracowanych metod w postaci oryginalnych narzędzi wspomagających monitorowanie i analizę systemów wbudowanych.

W odniesieniu do aktualnego stanu wiedzy rozwiązania zaproponowane przez Autora są innowacyjne i stanowią samodzielny i oryginalny dorobek Autora. Wyniki prac zostały opublikowane w czasopiśmie zagranicznych i materiałach konferencji międzynarodowej.

5. Organizacja i redakcja rozprawy

Rozprawa składa się z 6 rozdziałów, bibliografii i dwóch dodatków. Rozdział pierwszy stanowi wprowadzenie do tematyki pracy i obejmuje motywację do podjęcia tematyki badań, przegląd literatury oraz sformułowanie tezy i celu pracy. W rozdziale 2 Autor przedstawił proponowane rozwiązania z zakresu analizy śladu instrukcji. Rozdział 3 zawiera opis opracowanej metodyki analizy szeregów czasowych. W rozdziale 4 opisano metodę korelacji

logów tekstowych z szeregiem czasowym. Rozdział 5 przedstawia opis badań eksperymentalnych. Rozdział 6 stanowi podsumowanie pracy. W dodatkach przedstawiono opisy opracowanych narzędzi. Taka organizacja pracy jest poprawna i stanowi logiczny układ treści.

Autor opisał omawiane zagadnienia w sposób zwięzły i zrozumiały. Praca jest starannie zredagowana, tekst pracy jest przejrzysty, chociaż niektóre rysunki nie są zbyt czytelne. Praca zawiera też nieliczne błędy redakcyjne i stylistyczne takie jak:

- str. 36: na rys 3 występuje anomalia nr 0, nie ma jej w Tab 3.
- str. 39: zamiast „urządzenia” jest „urządzenie”,
- str. 44: czy pojęcie „pojedynczy stan bazowy” oznacza „jeden stan bazowy”?
Pojęcia grupy i stanu są używane zamiennie, nie jest to poprawne.
- str. 67 „... są zdefiniowane są ...”,
- str. 69: „włączy on zegar”, zamiast „liniami” jest „piniami”, „bez kosztowe”,
- str. 79: „mogą być obserwowany”
- str. 82: zamiast OI pojawia się skrót IO,
- str. 97: „czas logowani”.

Czasem też autor niepotrzebnie opisuje szczegóły, które są powszechnie znane np. opis znaczenia standardowych notacji matematycznych (str. 82), czy opis notacji pseudokodu (p.3.1.2). Na poziomie rozprawy doktorskiej należałoby założyć, że tak podstawowe konstrukcje są znane i nie wymagają wyjaśnień.

W pracy Autor wprowadza wiele formalnych pojęć oraz symboli. Jest to jak najbardziej poprawne, ponieważ pozwala w zwięzły i jednoznaczny sposób zdefiniować poszczególne pojęcia. Jednak definicje oznaczeń i pojęć są podawane w tekście, wydaje się, że praca byłaby bardziej czytelna, gdyby wyodrębniono z tekstu wprowadzane pojęcia w postaci numerowanych definicji formalnych. Przydatny byłby też wykaz oznaczeń i symboli używanych w pracy.

Pomimo ww. drobnych uwag rozprawę należy uznać za poprawnie zredagowaną i uważam, że Autor potrafi w sposób poprawny i przekonujący przedstawić uzyskane wyniki.

6. Uwagi krytyczne, wady i słabe strony rozprawy

Oprócz wcześniej przedstawionych uwag dotyczących redakcji rozprawy oraz sformułowania tezy pracy wydaje się, że niektóre zagadnienia powinny być dokładniej przedyskutowane lub szerzej opisane. Dotyczy to następujących problemów:

1. W rozdziałach 3 i 4 Autor najpierw wprowadził formalny model danych i formalne definicje stosowanych pojęć, co ułatwiło zrozumienie prezentowanych dalej algorytmów. Brakuje takiego podejścia w p. 2. W szczególności brakuje formalnych definicji kluczowych pojęć stosowanych w algorytmie takich jak: anomalia, ślad instrukcji, segment itp.
2. W przypadku algorytmu z p. 3 wykonano analizę złożoności obliczeniowej. Dlaczego takiej analizy nie wykonano w przypadku algorytmów z p. 2 oraz p. 4?

3. Część eksperymentalna dla algorytmu z p. 2 jest dosyć uboga. Ponieważ skuteczność algorytmu zależy od przyjętych parametrów, interesujące byłoby pokazanie zachowania się algorytmu w zależności od wartości tych parametrów.
4. Na str. 42 wprowadzona jest funkcja stanowiąca metrykę segmentu szeregu czasowego. Następnie tę samą funkcję stosuje się w odniesieniu do grupy segmentów. Jaka jest definicja metryki grupy segmentów?
5. Algorytm 3 został zilustrowany jedynie dla danych zebranych dla bliżej nieokreślonego urządzenia. Natomiast w 3.2 przedstawiono opis analizy szeregu czasowego poboru prądu dla praktycznego urządzenia Holter, jednak analiza ta nie wykorzystuje opracowanych metod. Jedynie w podsumowaniu Autor wskazał, że wykorzystanie opracowanej metody pozwoliłoby na zautomatyzowanie procesu optymalizacji energetycznej urządzenia. Powstaje zatem pytanie dlaczego w tym rozdziale autor nie zademonstrował tych możliwości przedstawionej metody dla przedstawionego, praktycznego problemu?
6. W pracy brakuje oceny uzyskanych wyników w odniesieniu do istniejących metod znanych z literatury. W analizie źródeł Autor przedstawił przegląd publikacji opisujących podejścia związane z tematyką pracy, natomiast w dalszej części pracy nie odniósł się tych rozwiązań, na przykład poprzez porównanie, wskazanie przewag, wad i zalet opracowanych rozwiązań w porównaniu z podobnymi metodami znanymi z literatury i/lub stosowanymi w praktyce.

Wyżej wymienione uwagi nie podważają poprawności metodologii badawczej i wartości uzyskanych wyników badań. Ale omówienie tych zagadnień przyczyniłoby się do lepszego zrozumienia, kompletności opisywanych zagadnień i podkreślenia praktycznego znaczenia opracowanych rozwiązań.

7. Znaczenie uzyskanych wyników

Praktycznym efektem pracy są opracowane narzędzia:

- *TraceAnalyzer* – program do agregacji i analizy śladu instrukcji,
- *StateEventAnalyzer* – program do detekcji sekwencji stanów i klas cykli oraz korelacji logów tekstowych.

Narzędzia te mogą mieć zastosowanie w praktyce inżynierskiej do monitorowania i analizy działania systemów wbudowanych. Co pozwoli na wykrywanie anomalii i błędów w tych systemach a także na optymalizację oprogramowania wbudowanego np. pod względem minimalizacji zużycia energii.

Opracowane rozwiązania mogą stanowić również punkt wyjścia do dalszych badań. Na bazie zaproponowanych modeli i metod mogą być opracowywane kolejne rozwiązania zwiększające możliwości analizy danych uzyskanych podczas monitorowania pracy systemów wbudowanych.

8. Ocena końcowa rozprawy

Podsumowując, uważam że rozprawa doktorska przedstawia oryginalne rozwiązanie zaprezentowanego w niej zagadnienia naukowego i projektowego. Autor podjął w niej problem, który ma istotne znaczenie w dziedzinie informatyki. Trafnie określił założenia dotyczące jego analizy i uzyskane wyniki potwierdził wykonanymi eksperymentami. Wykazał się dobrą znajomością ogólnej wiedzy teoretycznej i praktycznej z zakresu tematyki pracy, a także umiejętnością samodzielnego prowadzenia pracy naukowej. Stwierdzam, że recenzowana praca spełnia wymogi stawiane rozprawom doktorskim i niniejszym wnoszę o dopuszczenie jej do publicznej obrony.

