

Recenzja rozprawy doktorskiej

Andrzeja Stroińskiego

zatytułowanej:

Odkrywanie modeli procesów w reprezentacji sieci Petriego dla systemów rozproszonych typu CRS

*(Discovery of Petri Net models of distributed processes in
Communicating Resource Systems)*

1. Problem badawczy i jego znaczenie

Jednym z najczęściej wykorzystywanych stylów stosowanych przy budowaniu i eksploatacji systemów rozproszonych typowych dla bankowości, wojskowości, komunikacji pasażerskiej, instytucji zarządzania kryzysowego itp. jest tzw. architektura zorientowana na usługi (Service Oriented Architecture (SOA)). Architektura ta umożliwia projektowanie określonych funkcjonalności systemu w wyniku konkretyzacji procesów biznesowych skomponowanych z usług - komponentów tego systemu. Wśród głównych podejść stosowanych do budowy systemów SOA wyróżnia się podejście proceduralne (akcentujące wywoływanie usług) i podejście deklaratywne (akcentujące wywoływanie zasobów odpowiedzialnych za realizację usług). Każde z tych podejść ma swoje zalety i wady uwidaczniające się w kontekście specyfiki systemu, w którym jest implementowane. W kontekście rozważanej w dysertacji klasy systemów komunikujących się zasobów (Communicating Resource Systems (CRS)) obejmujących systemy implementowane zgodnie ze stylem architektonicznym REST (Representational State Transfer) czy ROA (Resource Oriented Architecture), przyjęte zastosowanie podejścia deklaratywnego pozwala, w porównaniu do podejścia proceduralnego, istotnie uprościć konstrukcję i utrzymanie systemu oraz zmniejszyć narzuty czasowe. W systemach CRS nie unika się jednak problemów zarządzania związanych m.in. z kumulowaną w czasie modyfikacją działania zasobów, skutkującą pogorszeniem lub utratą wymaganej jakości świadczonych usług. Podjęcie stosownych działań w takim wypadku wymaga szczegółowej analizy modelu opisującego obserwowane zachowanie usługi, co przy częstym w praktyce braku aktualnej dokumentacji lub błędach w implementacji, wymusza konieczność okresowej/bieżącej identyfikacji modeli realizowanych usług. Dostępność modelu procesu odpowiadającego rzeczywistemu działaniu systemu pozwala zidentyfikować miejsca wystąpienia błędów w realizowanym procesie biznesowym. Przedstawioną potrzebę bieżącej identyfikacji rzeczywistego funkcjonowania systemu zaspokajają algorytmy odkrywania modeli realizowanych w nim procesów. Pozyskiwane przez nie modele wyznaczane są na podstawie analizy wcześniejszych aktywności procesów archiwizowanych w dziennikach zdarzeń.

Specyfika rozważanej klasy systemów komunikujących się zasobów uniemożliwia wykorzystanie dostępnych algorytmów odkrywania modeli, algorytmów wykorzystujących kompletne (rejestrujące wszystkie możliwe przeploty aktywności procesów) dzienniki zdarzeń. Oznacza to, że

występowanie w tych systemach dużej liczby rozproszonych i asynchronicznie wykorzystywanych zasobów indywidualnie (zwykle w nieustrukturyzowanej formie) archiwizujących swoje dzienniki zdarzeń wymusza potrzebę rekonstrukcji stosownych dzienników realizowanych w nich procesów biznesowych.

W przedstawionym kontekście problem badawczy rozprawy zakłada konieczność opracowania procedur rekonstrukcji dzienników zdarzeń i procedur odkrywania zawartych w nich modeli procesów biznesowych. Ponadto, z uwagi na potrzebę automatycznej identyfikacji miejsc wystąpienia błędów (odchyień) w realizowanych procesach biznesowych, do reprezentacji odkrywanych modeli przyjmuje się formalizm teorii sieci Petriego, formalizm umożliwiający efektywną analizę procesów współbieżnych, m.in. pod kątem takich ich właściwości jak żywotność i ograniczoność.

Elementy nowości zaproponowanego podejścia przejawiają się, tak w oryginalnym sformułowaniu problemów rekonstrukcji procesów biznesowych, jak i autorskich, proponowanych do ich rozwiązania, metodach implementujących formalizmy teorii sieci Petriego, technikach analizy środowisk rozproszonego przetwarzania i transmisji danych oraz programowaniu w środowiskach obliczeń współbieżnych.

Przedmiot badań, treść rozprawy oraz wykorzystane metody badawcze pozwalają stwierdzić, że praca ma charakter teoretyczny i interdyscyplinarny. Zagadnienia, które przedstawiono w rozprawie doktorskiej są związane z obszarami konstrukcji i zarządzania systemami rozproszonymi, przetwarzania rozproszonego, modelowania i analizy procesów, eksploracji procesów, a także złożoności obliczeniowej, są też dobrze ugruntowane w literaturze przedmiotu i licznych eksperymentach komputerowych.

Reasumując, opiniowana rozprawa, koncentrując się na zagadnieniach budowy i oceny efektywności funkcjonowania algorytmów odkrywania modeli procesów biznesowych reprezentowanych w formalizmach wybranych klas sieci Petriego, podejmuje ważny i aktualny problem wyznaczania warunków determinujących możliwości eksploracji procesów w systemach komunikujących się zasobów. Uważam, że podjęcie przedstawionej problematyki jest w pełni uzasadnione zarówno ze względów poznawczych, jak i możliwości wielu praktycznych zastosowań m.in. związanych z opracowaniem technik eksploracji procesów biznesowych do rekomendowania i usprawniania procesów realizowanych w różnych obszarach zastosowań rozproszonych systemów informatycznych.

2. Wkład autora

Zmierzając do osiągnięcia wyznaczonego przez siebie celu jakim jest opracowanie nowych metod odkrywania modeli sieci Petriego dla procesów biznesowych wykonywanych w systemach CRS, Doktorant uzyskał szereg nowych rezultatów. Do ważniejszych z nich, wyróżniających je spośród dostępnych w literaturze przedmiotu, można zaliczyć:

1. Model formalny autorskiej koncepcji systemu komunikujących się zasobów stanowiącego uogólnienie systemów klasy REST. Istotą tego uogólnienia jest odejście lub osłabienie niektórych wymogów stawianych systemom REST odnoszących się na przykład do własności protokołów komunikacyjnych czy zbioru dostępnych operacji.
2. Model formalny sieci komunikujących się zasobów (wywodzącej się z koncepcji sieci przepływu) oraz zgodnej i ustrukturyzowanej sieci przepływu; sieci umożliwiających modelowanie systemu CRS dopuszczające kompozycję procesu biznesowego z wielu niezależnych procesów lokalnych, synchroniczny charakter komunikacji zasobów oraz hierarchiczny charakter architektury zasobów.
3. Opracowanie algorytmu rekonstrukcji kompletnego logu procesu biznesowego stanowiącego swoisty dziennik zdarzeń procesu realizowanego w CRS, dziennik składający się ze zdarzeń niezależnie zbieranych z poszczególnych zasobów CRS.
4. Opracowanie algorytmów mapujących kompletny dziennik zdarzeń procesu biznesowego na zgodny i ustrukturyzowany model procesu biznesowego realizowanego w CRS i odpowiadającego

modelowi rzeczywistemu tego procesu. W szczególności, opracowane zostały cztery algorytmy odkrywania modelu procesu biznesowego w CRS, algorytm: odkrywający model zgodnej i ustrukturyzowanej sieci przepływu (CRS miner), odkrywający model sieci komunikujących się zasobów (hCRS miner), umożliwiający odkrywanie hierarchii zasobów (dhCRS miner), umożliwiająca odkrywanie kolorowanej sieci komunikujących się zasobów (edhCRS miner).

5. Zaplanowanie oraz przeprowadzenie bardzo wielu eksperymentów i analiz porównawczych, weryfikujących wzajemną efektywność opracowanych algorytmów.

Lapidarny i sformalizowany styl narracji przyjętej w rozprawie (73 definicje, 6 twierdzeń, 17 listingów) z jednej strony bardzo ułatwia śledzenie logiki prezentowanych argumentacji, z drugiej zaś mocno ogranicza semantyczny charakter prowadzonych badań. Przykładem tego jest formalna definicja systemu CRS, której nie towarzyszy jednak żadna ilustracja czy też szerszy komentarz przybliżający jej znaczenie do przypadków spotykanych w praktyce.

Rozwijając ten wątek, formalne ujęcie wykorzystywanych pojęć pozwala wskazać warunki, spełnienie których gwarantuje poprawność przedstawianych algorytmów odkrywania modeli procesów. Warto zauważyć, że postulowane twierdzenia mając charakter warunków wystarczających nie ograniczają możliwości budowy innych, być może jeszcze bardziej efektywnych metod eksploracji.

Pewien niedosyt, w rozważanym kontekście, rodzi brak stosownego uzasadnienia warunku wykorzystywanego w algorytmie rekonstrukcji kompletnego logu procesu biznesowego w CRS (Listing 4, linia 15). Podobny niedosyt budzi również brak odniesień wyznaczonych oszacowań złożoności obliczeniowych podanych algorytmów do innych, aktualnie wykorzystywanych procedur eksploracji procesów.

3. Poprawność

Z nasuwających się uwag warto zwrócić na te o charakterze redakcyjnym oraz te akcentujące słabsze i mocniejsze aspekty przedstawionych badań. Spośród pierwszej grupy uwag warto podkreślić brak listy stosowanych w tekście rozprawy skrótów, symboli i oznaczeń. Jej brak utrudnia lekturę pracy i skutkuje, stosunkowo nielicznymi, błędami formalnymi oraz typograficznymi. Przykładowe z nich:

- błędy formalne, tekst ze str. 106₁₃ „...resource $r \in R \setminus r' \dots$ ” winien być zapisany jako „...resource $r \in R \setminus \{r'\} \dots$ ”, zapis IWFnetr:CRPLr \ IWFnetr:IRPLr \ IWFnetr:ORPLr ze str. 105₃ jest niejednoznaczny - operator różnicy mnogościowej nie jest łączny $(A \setminus B) \setminus C \neq A \setminus (B \setminus C)$,
- usterki typograficzne, np.: w różnych miejscach tekstu symbol konkatencji jest różnie notowany: na stronie 13₉ jako”.”, z kolei na stronie 88₂ jako” \oplus ”; ta sama uwaga dotyczy nawiasów wykorzystywanych do oznaczenia tupli: np. w Def.73 „...CT = $\langle a, b, c \rangle$...”, a na str. 73 106₁₀ „...tuple (a,B)...” ; str. 152 podobnie 92₁₄ „...dotted...” winno być zastąpione przez „...dashed...”
- mało przyjazny sposób dokumentacji źródeł bibliograficznych, np. pomijając brak opisu przyjętego sposobu kodowania cytowanych pozycji widać brak konsekwencji w jego stosowaniu np. [GVDA07] i [MWVdAvdB02], różne kody oznaczają te same pozycje bibliograficzne np. [DSB15a] i [DSB15b], pozycje netograficzne nie są wydzielone i nie każda z nich opisana jest datą dostępu. Warto zauważyć, że kod [DSB16] odnosi się nie do raportu technicznego jak to ma miejsce w dysertacji, a do pozycji: Dwornikowski, D., Stroiński, A., Brzeziński, J., Conformance Checking of Communicating Resource Systems with RAs Calculus, IEEE Int. Conf. on Services Computing (SCC 2015), IEEE, pages 759-764, 2015. Warto również dodać, że Doktorant nie zacytował paru swoich wcześniejszych publikacji m.in.: Stroiński, A.; Dwornikowski, D., Brzeziński, J., RESTful Web Service Mining: simple algorithm supporting resource-oriented systems, In Proc. of 21th IEEE Int. Conf. on Web Services, ICWS 2014, IEEE, pages 694-695, 2014., Stroiński, A.; Dwornikowski, D., Brzeziński, J., Resource Mining: applying process mining

to resource-oriented systems, In Proc. of 17th Int. Conf. on Business Infor. Systems, Springer Int. Services, pages 217-228, 2014.

Większość tego typu uwag przekazałem bezpośrednio Autorowi. Oprócz wcześniej już wspomnianych wartościowych aspektów rozprawy warto również wskazać na:

- uzasadniony wybór formalizmu sieci Petriego jako reprezentacji modelu procesu biznesowego, spośród innych aktualnie dostępnych takich jak np. BPMN, notacja i diagramy EPC, diagramy aktywności, diagramy stanów UML, sieci Petriego są jednym z najbardziej popularnych formalizmów służących do modelowania i analizy systemów współbieżnych. Używa się w nich intuicyjnego języka graficznego, który ułatwia modelowanie systemu, oraz zaawansowanych metod analizy formalnej własności skonstruowanego modelu,
- propozycję metody rozproszenia obliczeń związanych z odkrywaniem modelu procesu biznesowego uruchamianego w systemie CRS, metody umożliwiającej współbieżne odkrywanie procesów składających się z wielu niezależnych lokalnych procesów realizowanych przez jego zasoby i skutkującej znacznym skróceniem narzutów czasowych,
- wyczerpujące i zarazem krytyczne przedstawienie aktualnego stanu badań prowadzonych w obszarze modelowania i eksploracji procesów biznesowych.

Do słabszych aspektów rozprawy należą moim zdaniem:

- praktycznie pominięte, kwestie oceny podatności opracowanych sieciowych modeli procesów biznesowych widziane w kontekście zwyczajowo wyróżnianych perspektyw eksploracji procesów obejmujących perspektywy przepływu sterowania, organizacyjną, instancji procesu, oraz czasu,
- brak dyskusji i/lub przykładów ilustrujących możliwości wykorzystania przyjętych reprezentacji modeli procesów biznesowych, ukazujących potencjał tych reprezentacji np. w zakresie możliwości wykrywania odchyśleń od nominalnych przebiegów procesów i/lub możliwości ich doskonalenia,
- podobny niedosyt budzi brak dyskusji nakładów obliczeniowych ponoszonych w wybranym podejściu proceduralnym i zaproponowanym podejściu deklaratywnym, pozwalający na ich porównanie np. w kontekście do jednej, wybranej instancji procesów biznesowych.

4. Wiedza kandydata

Doktorant wykazał się znajomością podstawowej literatury przedmiotu rozprawy, metod modelowania i eksploracji procesów biznesowych, środowisk rozproszonego przetwarzania i transmisji danych, umiejętnością modelowania w wybranych reprezentacjach sieci Petriego, programowania w środowiskach obliczeń rozproszonych, a także zdolnościami praktycznego wykorzystania tych narzędzi w odkrywaniu modeli procesów biznesowych. Potwierdzeniem tych ocen jest wyczerpujący, krytyczny opis stanu badań prowadzonych w obszarach wytwarzania i utrzymywania systemów rozproszonych zgodnie z SOA, formalne ujęcia problemów rekonstrukcji dzienników zdarzeń w systemach CRS i odkrywania związanych z nimi (reprezentowanych w formalizmie sieci Petriego) modeli procesów biznesowych, a także liczne eksperymenty umożliwiające ocenę efektywności obliczeniowej komputerowych implementacji opracowanych algorytmów eksploracji.

Uważam, że uzyskane rezultaty potwierdzają wysokie kwalifikacje Doktoranta umożliwiające Mu swobodne poruszanie się zarówno w obszarach zagadnień z zakresu modelowania problemów eksploracji procesów, metod ich rozwiązywania, a także technik programowania i planowania eksperymentów komputerowych. Warto dodać, że Doktorant jest współautorem 9 publikacji oraz 2 raportów technicznych, a także zwrócić uwagę na fakt, że opiniowana rozprawa wykonana została w ramach projektu badawczego NCN, No. DEC-2012/05/N/ST6/03051 pt.: „Metody okrywania procesów biznesowych i rachunek procesów dla systemów zgodnych ze stylem architektonicznym REST” przyjętego do realizacji w dyscyplinie informatyka (ST6_2, ST6_4). Wymienione fakty

potwierdzają, że Doktorant potrafi podejmować i samodzielnie realizować zaplanowane cele badawcze.

Biorąc pod uwagę opinie zaprezentowane w poprzednich punktach i wymagania zdefiniowane przez artykuł 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym (z późniejszymi zmianami)¹ moja ocena rozprawy pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problem naukowego? (wybierz jedną opcję stawiając znak **X**)

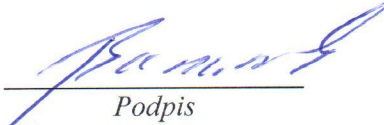
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

B. Czy po przeczytaniu rozprawy zgadzasz się, że kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie Informatyka lub Automatyka i Robotyka?

<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE

C. Czy kandydat ma umiejętność samodzielnego prowadzenia pracy naukowej?

<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Zdecydowanie TAK	Raczej TAK	Trudno powiedzieć	Raczej NIE	Zdecydowanie NIE



Podpis

¹ http://www.nauka.gov.pl/g2/oryginal/2013_05/b26ba540a5785d48bee41aec63403b2c.pdf