Częstochowa, dn. 09.09.2017

Prof. dr hab. inż. Roman Wyrzykowski
Instytut Informatyki Teoretycznej i Stosowanej
Politechnika Częstochowska
ul. Dąbrowskiego 69
42-201 Częstochowa
roman@icis.pcz.pl

RECENZJA
ROZPRAWY DOKTORSKIEJ
mgr inż. Łukasza Kufla
„Security and System Events Monitoring in Distributed Systems Environment"

Promotor: Prof. dr hab. inż. Jan Węglarz
Wydział Informatyki
Politechnika Poznańska

1. Problem badawczy i jego znaczenie

Tematyka przedłożonej do recenzji rozprawy doktorskiej mgr inż. Łukasza Kufla dotyczy obszaru badawczego z zakresu systemów rozproszonych, jednego z kluczowych dla rozwoju współczesnej informatyki. Rozwój badań w tym obszarze toczy się pewnymi falami, wśród których we wcześniejszym okresie wyróżnić można było np. **metakomputery i systemy gridowe**, zaś w ostatnich latach przebiega pod znakiem **systemów chmurowych** (ang. cloud). Swoisty boom w tej dziedzinie wymaga udzielenia szczególnej uwagi na takie aspekty, jak: doskonalenie oprogramowania warstwy pośredniej, służącego do zarządzania zasobami i usługami w heterogenicznym środowisku systemów geograficznie rozproszonych; zapewnienie bezpieczeństwa; problem przezroczystego dla użytkownika łączenia wielu domen z uwzględnieniem efektywności zarządzania zasobami dla różnych rodzajów przetwarzania, etc. Najnowsze tendencje, o których warto wspomnieć w tym kontekście, dotyczą pojęć, dla których wydaje się, że nie istnieją jeszcze ich polskie odpowiedniki, takich jak **fog computing** oraz **dew computing**. Jednym z czynników, które odegrały istotną rolę w ich pojawieniu się, jest niewątpliwie szybki rozwój Internetu Rzeczy.

W tym kontekście pozytywnie należy ocenić wybór zagadnienia badawczego rozważanego w recenzowanej rozprawie. Jest nim problem wyboru, zaprojektowania oraz wdrożenia rozwiązania do monitoringu zdarzeń systemowych i bezpieczeństwa w środowiskach systemów rozproszonych. Przeprowadzone badania objęły przy tym zarówno pojedyncze centra danych oraz wiele takich centrów, jak również chmury. Problem ten ma głównie znaczenie praktyczne, gdyż jego rozwiązanie umożliwia w pierwszej kolejności poprawę efektywności monitorowania. Traktowany bardziej ambitnie, można rzec całościowo, ma on również charakter naukowy jako określone wyzwanie poznawcze ukierunkowane na lepsze i głębsze poznanie różnorakich aspektów, które maja wpływ na uzyskaną efektywność monitorowania, w całej złożoności ich wzajemnych powiązań.

2. Wkład autora

Praca napisana jest w języku angielskim i składa się z sześciu rozdziałów oraz bibliografii, która obejmuje 29 witryn internetowych dotyczących narzędzi i środowisk monitorujących oraz 62 pozycje zasadnicze. Wersja angielska pracy liczy łącznie 108 stron.

Rozdział pierwszy zawiera wprowadzenie do zagadnień w niej poruszanych, wraz z określeniem celów prowadzonych badań.

Rozdział drugi omawia podstawy monitorowania systemów rozproszonych z punktu widzenia zarówno ich budowy, jak i funkcjonowania. W szczególności, zwrócono uwagę na konieczność uwzględnienia poziomów krytyczności systemów i zachodzących w nich zdarzeń w połączeniu z wielkością interwałów monitorowania.

W **rozdziale trzecim** przedstawiono istniejące podejścia do monitorowania systemów rozproszonych, w tym również niedawno wprowadzone do praktyki podejście oparte na wykorzystaniu strumieni danych. Rozdział ten obejmuje również dostatecznie wnikliwą analizę porównawczą istniejących podejść. Kontynuację tego rozdziału stanowi **rozdział czwarty**, który zawiera wnikliwy przegląd narzędzi i środowisk monitorujących podzielonych na dwie kategorie: pierwsza z nich dotyczy monitorowania szeroko rozumianych zagadnień bezpieczeństwa, zaś druga – odnosi się do monitorowania infrastruktury ze względu zarówno na jej dostępność, jak i charakterystyki ilościowe.

Wkład Autora skoncentrowany jest w **rozdziale piątym**, który poświecono rozwiązaniu zagadnienia zaprojektowania oraz wdrożenia efektywnego systemu do monitorowania zdarzeń systemowych i bezpieczeństwa w środowiskach systemów rozproszonych. Niezwykle istotnym elementem tego rozdziału są intensywne badania eksperymentalne potwierdzające zalety zaproponowanego przez Autora innowacyjnego podejścia hybrydowego. Pracę kończy **rozdział szósty**, zawierający podsumowanie uzyskanych w niej rezultatów, a także nakreślający niektóre kierunki przyszłych badań Autora.

Uwzględniając powyższe omówienie zawartości pracy oraz ogólną pozytywną ocenę jej zawartości merytorycznej, uważam, że za bezsporne osiągnięcia i wkład Autora należy uznać następujące rezultaty:

1. W aspekcie poznawczym za główne osiągniecie Autora należy uznać zaproponowane w pracy innowacyjne podejście do monitorowania systemów rozproszonych – tzw. order-based monitoring (OBM), bazujące na wykorzystaniu harmonogramów zadań i idei zaplanowanego zamówienia. Wśród zalet tego podejścia można w pierwszej kolejności wymienić zmniejszenie narzutów na monitorowanie przy zachowaniu możliwości otrzymywania dogłębnej informacji w procesie monitorowania, a także niezależność od wykorzystywanych platform oraz zdolność do współpracy z wbudowanymi narzędziami i protokołami monitorującymi, jak również z produktami innych dostawców.
2. Kontynuując wątek poznawczy, pragnę zwrócić uwagę na autorską propozycję koncepcji lokalnego dystrybutora jako sposobu redukcji wpływu opóźnienia sieciowego na wydajności monitorowania rzeczywistych systemów rozproszonych. Zastosowanie tej koncepcji pozwala także na uzyskanie dodatkowych korzyści dzięki wprowadzeniu takich funkcjonalności jak np. kompresja i filtracja danych, możliwość definiowania zaawansowanych polityk w zakresie ustalania wartości progowych czy też pełnienia roli weryfikującej i podejmowania autonomicznych działań naprawczych.
3. Istotnym wkładem Autora, łączącym aspekt poznawczy i praktyczny (aplikacyjny), jest także przeprowadzenie wyczerpujących badań eksperymentalnych, co umożliwiło wykazanie celowości zastosowania zaproponowanych przez Autora rozwiązań dla różnorodnych scenariuszy. obejmujących nie tylko jedno lub wiele centrów danych, lecz również systemy chmurowe. W szczególności, przeprowadzone badania umożliwiły

dogłębną analizę wpływu opóźnienia sieciowego na całkowitą wydajność procesów monitorowania dla scenariusza wielu centrów danych.

4. W aspekcie aplikacyjnym innym wkładem Autora, zasługującym na uwagę, jest zaprojektowanie i potwierdzone testami wdrożenie efektywnych systemów monitorowania zdarzeń systemowych i bezpieczeństwa dla różnych typów systemów rozproszonych, ze szczególnym uwzględnieniem chmur publicznych takich jak Amazon AWS oraz Google Cloud. Potwierdza to możliwość wykorzystania zaproponowanych rozwiązań do monitorowania najbardziej zaawansowanych typów systemów rozproszonych.

Na podkreślenie zasługuje także fakt opublikowania przez Autora trzech samodzielnych prac w czasopismach indeksowanych w bazie JCR.

3. Poprawność

Poprawność treści pracy nie wzbudza moich istotnych zastrzeżeń, a stwierdzenia w niej zawarte wydają się być godne zaufania, co wynika w szczególności z dosyć szczegółowych uzasadnień, popartych wynikami przeprowadzonych badań eksperymentalnych. Generalnie sposób i jakość przeprowadzenia badań eksperymentalnych stanowi bardzo wartościowy element pracy i zasługuje na podkreślenie. W szczególności, dotyczy to dostatecznie wyczerpujących eksperymentów przeprowadzonych w wiodących chmurach publicznych, jakimi są Amazon AWS oraz Google Cloud. Z doświadczeń moich współpracowników wiem bowiem, z jakimi trudnościami trzeba się niekiedy zmierzyć, aby dostosować chmurę publiczną do określonych wymagań i wręcz zmusić ją do stabilnego funkcjonowania dla niestandardowych zastosowań i aplikacji, jakie stanowią w szczególności rozpatrywane w pracy warianty konfiguracyjne systemów monitorujących.

Jednocześnie Autor nie ustrzegł się pewnych braków i słabości. Wśród uwag o charakterze krytycznym, a po trosze dyskusyjnym, wymienić należy :

1. Jednym z celów zaproponowanych w pracy rozwiązań jest zwiększenie wydajności monitorowania z uwzględnieniem np. opóźnienia w warstwie sieciowej. W celu potwierdzenia ich skuteczności Autor stosuje wyłącznie badania eksperymentalne i nie próbuje w ogóle budować jakichkolwiek modeli, aby później wykorzystać je do analizy wydajności w sposób mniej lub bardziej ogólny.

2. Generalnie oparcie się wyłącznie na podejściu eksperymentalnym jako jedynym stosowanym w pracy sposobie weryfikacji uzyskanych wyników stanowi jej słabość. Brak rozważań o charakterze bardziej teoretycznym, z wykorzystaniem odpowiedniego aparatu formalnego, pozbawia Autora możliwości głębszego zbadania charakterystyk procesów monitorowania i wyciągnięcia na tej podstawie wielu istotnych konkluzji.
3. W swojej rozprawie Autor zajmuje się miedzy innymi projektowaniem systemów monitorowania z uwzględnieniem sformułowanych kryteriów i ograniczeń. W tym kontekście wydaje się, iż można byłoby pokusić się o sformułowanie mniej lub bardziej ścisłej metodyki projektowania. Niestety takiej próby w pracy nie podjęto.
4. Wariant wdrożenia systemu monitorującego w chmurach publicznych stanowi istotne wyzwanie dla efektywności zaproponowanych w pracy rozwiązań. Autor podjął to wyzwanie, przezwyciężając z powodzeniem napotkane przy tym trudności. Określony niedosyt wywołuje jednak bardzo lakoniczny i moim zdaniem wręcz nieprzekonywujący sposób przedstawienia architektury i konfiguracji systemu monitorującego w tym przypadku.
5. Opracowane przez Autora rozwiązania służą w szczególności do monitorowania zdarzeń bezpieczeństwa. W związku z tym, ale oczywiście nie tylko z tego powodu, bardzo istotne jest, aby one same nie wnosiły dodatkowych zagrożeń pod względem cyberbezpieczeństwa. Wydaje się, że temu zagadnieniu poświęcono zbyt mało uwagi w rozprawie.

4. Wiedza kandydata

Z omówienia treści pracy, które przytoczono w punkcie 2 niniejszej recenzji, wynika, iż cztery pierwsze rozdziały rozprawy poświęcone są głównie krytycznemu przedstawieniu stanu wiedzy w zakresie tematyki pracy, potwierdzając w ten sposób ogólny stan wiedzy w zakresie dyscypliny Informatyka, ze szczególnym uwzględnieniem wybranych zagadnień budowy i funkcjonowania systemów rozproszonych. Jakość tych rozdziałów nie budzi moich zastrzeżeń. Świadczą one o dużej wiedzy Autora w zakresie tematyki badań, popartej szerokim doświadczeniem praktycznym związanym z wdrożeniem i wykorzystaniem systemów monitorujących. Również bibliografia

zawarta i wykorzystana w pracy nie budzi zastrzeżeń, a moja opinia o jej kompletności jest pozytywna.

5. Podsumowanie

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) moja ocena pod względem trzech podstawowych kryteriów jest następująca:

A. Czy rozprawa zawiera oryginalne rozwiązanie problemu naukowego ?

   Zdecydowanie TAK

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

   Zdecydowanie TAK

C. Czy kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej ?

   Zdecydowanie TAK

Prof. dr hab. inż. Roman Wyrzykowski